Mehdi Khosrow-Pour

Encyclopedia of Information Science and Technology

Volumes 1-5

Encyclopedia of Information Science and Technology

Volume I

Mehdi Khosrow-Pour, D.B.A.
Information Resources Management Association, USA

Idea Group REFERENCE

IDEA GROUP REFERENCE
Hershey • London • Melbourne • Singapore

Acquisitions Editor: Mehdi Khosrow-Pour
Asst. Acquisitions Editor: Renée Davies
Senior Managing Editor: Jan Travers
Managing Editor: Amanda Appicello
Development Editor: Michele Rossi
Copy Editors: Amanda Appicello, Elizabeth Arneson, Maria Boyer, Alana Bubnis, Jane Conley, Renée Davies, Lori Eby, Bernard J. Kieklak, Jr., Shanelle Ramelb, Sue VanderHook and Ingrid Widitz
Typesetters: Sara Reed and Rachel Shepherd
Support Staff: Corrina Chandler, Amanda Phillips, Michelle Potter, Marko Primorac, Jennifer Wetzel, Tara White, Chris Yura and Larissa Zearfoss
Cover Design: Lisa Tosheff
Printed at: Yurchak Printing Inc.

Published in the United States of America by
Idea Group Reference (an imprint of Idea Group Inc.)
701 E. Chocolate Avenue, Suite 200
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@idea-group.com
Web site: http://www.idea-group-ref.com

and in the United Kingdom by
Idea Group Reference (an imprint of Idea Group Inc.)
3 Henrietta Street
Covent Garden
London WC2E 8LU
Tel: 44 20 7240 0856
Fax: 44 20 7379 3313
Web site: http://www.eurospan.co.uk

Library of Congress Cataloging-in-Publication Data

Encyclopedia of information science and technology / Mehdi Khosrow-Pour, editor.
p. cm.
Includes bibliographical references and index.
ISBN 1-59140-553-X — ISBN 1-59140-794-X (ebook)
1. Information science—Encyclopedias. 2. Information technology—Encyclopedias. I. Khosrowpour, Mehdi, 1951-
Z1006.E566 2005
004'.03—dc22

2004025544

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

All work contributed to this book is new, previously-unpublished material. The views expressed in this book are those of the authors, but not necessarily of the publisher.

Editorial Advisory Board

List of Contributors

(listed alphabetically by first name)

A. Vadivel / *Indian Institute of Technology, Kharagpur, India*
A.C. Gómez Lora / *Universidad de Málaga, Spain*
A.K. Majumdar / *Indian Institute of Technology, Kharagpur, India*
Abdulbaset Gaddah / *Carleton University, Canada*
Abdulwahed Mohammed Khalfan / *Public Authority Applied Education & Training, Kuwait*
Aberdeen Leila Borders / *University of New Orleans, USA*
Ada Scupola / *Roskilde University, Denmark*
Adolfo Sánchez de Miguel / *Cronos Ibérica, S.A., Spain*
Agustinus Borgy Waluyo / *Monash University, Australia*
Ai Cheo Yeo / *Monash University, Australia*
Airi Salminen / *University of Jyväskylä, Finland*
Ajantha Dahanayake / *Delft University of Technology, The Netherlands*
Ala Abu-Samaha / *Amman University, Jordan*
Albert L. Ingram / *Kent State University, USA*
Alex A. Freitas / *University of Kent, UK*
Alexa Heucke / *Munich University of Applied Sciences, Germany*
Alexander Markhasin / *Siberian Telecommunications University, Russia*
Alexandra Galatescu / *National Institute for R & D in Informatics, Romania*
Alexandros Nanopoulos / *Aristotle University of Thessaloniki, Greece*
Ali Arya / *Simon Fraser University, Canada*
Ali Metwalli / *Western Michigan University, USA*
Alina Lazar / *Youngstown State University, USA*
Amaryllis Raouzaiou / *National Technical University of Athens, Greece*
Amita Goyal Chin / *Viginia Commonwealth University, USA*
Amr El Abbadi / *University of California, Santa Barbara, USA*
Amy B. Woszczynski / *Kennesaw State University, USA*
Ana C. Andrés del Valle / *VICOMTech Research Center, Spain*
Ana Funes / *Universidad Nacional de San Luis, Argentina*
Ana Tereza Vasconcelos / *LNCC, Brazil*
Anabela Sarmento / *ISCAP/IPP, Portugal and University of Minho, Portugal*
Anand Iyer / *PRTM, UK*
Anatália Saraiva Martins Ramos / *Universidade Federal do Rio Grande do Norte, Brazil*
Anders Avdic / *Örebro University, Sweden*
André de Carvalho / *USP São Carlos, Brazil*
Andrea De Lucia / *University of Salerno, Italy*
Andreas L. Opdahl / *University of Bergen, Norway*
Andreas Mitrakas / *Ubizen, Belgium*
Andreas Rülke / *PRTM, UK*
Andrés Coratella / *Universidad de Buenos Aires, Argentina*
Andrew Borchers / *Kettering University, USA*

Andrew Laghos / *City University, London, UK*
Andrew S. Targowski / *Western Michigan University, USA*
Andrew Urbaczewski / *University of Michigan, USA*
Andrey Naumenko / *Triune Continuum Enterprise, Switzerland*
Andries Burger / *University of the Free State, South Africa*
Anesh Maniraj Singh / *University of Kwazulu Natal, South Africa*
Anet Potgieter / *University of Cape Town, South Africa*
Angel D. Sappa / *Computer Vision Center, Spain*
Ann Frances Cameron / *Queens University, Canada*
Anna Malina / *e-Society Research, UK*
Anne Honkaranta / *University of Jyväskyä, Finland*
Anthony Scime / *State University of New York College at Brockport, USA*
Antonio Díaz-Andrade / *Escuela de Administración de Negocios para Graduados – ESAN, Peru*
Antonio P. Braga / *UFMG, Brazil*
Antonios Danalis / *University of Delaware, USA*
Ardis Hanson / *University of South Florida, USA*
Aristides Dasso / *Universidad Nacional de San Luis, Argentina*
Ari-Veikko Anttiroiko / *University of Tampere, Finland*
Arlyn J. Melcher / *Southern Illinois University at Carbondale, USA*
Arnor Solberg / *SINTEF ICT, Norway*
Arthur R. Edwards / *Erasmus University, The Netherlands*
Arthur Tatnall / *Victoria University, Australia*
Arun Rai / *Georgia State University, USA*
Aryya Gangopadhyay / *University of Maryland Baltimore County, USA*
Athanasios Drosopoulos / *National Technical University of Athens, Greece*
Athanasis Karoulis / *Aristotle University of Thessaloniki, Greece*
Athena Vakali / *Aristotle University of Thessaloniki, Greece*
Audun Jensvoll / *EDB Telesciences, Norway*
Awie Leonard / *University of Pretoria, South Africa*
Axel Schmetzke / *University of Wisconsin-Stevens Point, USA*
Aybüke Aurum / *University of New South Wales, Australia*
Babita Gupta / *California State University Monteray Bay, USA*
Badrul M. Sarwar / *University of Minnesota, USA*
Bala Srinivasan / *Monash University, Australia*
Beatriz de la Iglesia / *University of East Anglia, UK*
Ben Martz / *University of Colorado at Colorado Springs, USA*
Bernard Blandin / *CESI SAS, France*
Bernard C.Y. Tan / *National University of Singapore, Singapore*
Beverly Shattuck / *Hinks and Elaine Shimberg Health Sciences Center Library, USA*
Bill Davey / *RMIT University, Australia*
Bill Karakostas / *City University, London, UK*
Bin Zhang / *Louisiana State University, USA*
Birgit Jaeger / *Roskilde University, Denmark*
Bishwajit Choudhary / *Information Resources Management Association, USA*
Bob Justis / *Louisianna State University, USA*
Bob Ritchie / *University of Central Lancashire, UK*
Bosede Aworuwa / *Texas A&M University- Texarkana, USA*
Brian Dobing / *University of Lethbridge, Canada*
Bruce Lubotsky Levin / *University of South Florida, USA*
Bruce Rollier / *University of Baltimore, USA*
Bruce Vanstone / *Bond University, Australia*
Burak Ozer / *Verificon Corp., USA*
C.L. Vidal-Rodeiro / *Norman J. Arnold School of Public Health, USA*
C. Makatsoris / *Orion Logic Ltd., UK*

Carla Wilkin / *The University of Melbourne, Australia*
Caroline Howard / *Touro International University, USA*
Catherine M. Beise / *Salisbury University, USA*
Catherine M. Ridings / *Lehigh University, USA*
Cathy Cavanaugh / *University of North Florida, USA*
Celeste See-pui NG / *Yuan Ze University, Taiwan, R.O.C.*
Cesar Alexandre de Souza / *University São Judas Tadeu, Brazil*
Cesare Alippi / *Politecnico Di Milano, Italy*
Chad Lin / *Edith Cowan University, Australia*
Chalermsak Lertwongsatien / *Ministry of Finance, Thailand*
Chanda Elbert / *Texas A&M University, USA*
Chandra S. Amaravadi / *Western Illinois University, USA*
Charlene A. Dykman / *University of St. Thomas, USA*
Charles C. Hinnant / *The University of Georgia, USA*
Charles H. Mawhinney / *Metropolitan State College of Denver, USA*
Charles Møller / *Aarhus School of Business, Denmark*
Charles R. Graham / *Brigham Young University, USA*
Chen Way Siew / *National University of Singapore, Singapore*
Chi Kin Chan / *The Hong Kong Polytechnic University, Hong Kong*
Chin-Chen Chang / *National Chung Hsing University, Taiwan, R.O.C.*
Chong W. Kim / *Marshall University, USA*
Chris Chatwin / *University of Sussex, UK*
Chris Coulston / *Penn State Erie, USA*
Chris Kimble / *University of York, UK*
Christine Ashley / *University of Minnesota, USA*
Christoph Richter / *University of Hannover, Germany*
Christophe Nicolle / *Université de Bourgogne, France*
Christopher Klassen / *University of Dallas, USA*
Chung Hsun Hsieh / *Chaoyang University of Technology, Taiwan, R.O.C.*
Chung S. Kim / *Southwest Missouri State University, USA*
Chung-Shing Lee / *Pacific Lutheran University, USA*
Chung-wei Lee / *Auburn University, USA*
Clare Brindley / *Manchester Metropolitan University, UK*
Clare M. Pollock / *Curtin University of Technology, Australia*
Clarence N.W. Tan / *Bond University, Australia*
Clarence Tan / *Bond University, Australia*
Claudia Cevenini / *University of Bologna, Italy*
Claudia Pereira / *Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina*
Clint A. Bowers / *University of Central Florida, USA*
Colin C. Venters / *University of Manchester, UK*
Concetta M. Stewart / *Temple University, USA*
Conrad Shayo / *California State University, San Bernardino, USA*
Coral Calero / *University of Castilla-La Mancha, Spain*
Craig Van Slyke / *University of Central Florida, USA*
Cristiano Castelfranchi / *Institute for Cognitive Sciences and Technology, National Research Council of Italy, Italy*
Cristóbal Pareja-Flores / *Universidad Complutense de Madrid, Spain*
Cynthia LeRouge / *Saint Louis University, USA*
Cynthia Ruppel / *The University of Alabama in Huntsville, USA*
D. Bruce Sun / *California State University at Long Beach, USA*
D. Chakraborty / *Tohoku University, Japan*
D. Theotokis / *University of Peloponnese, Greece*
D.I. Martakos / *National and Kapodistrian University of Athens, Greece*
D.M. Gorman / *Texas A&M University, USA*
D. Sandy Staples / *Queens University, Canada*

D.S. Stamoulis / *National and Kapodistrian University of Athens, Greece*
Dale H. Shao / *Marshall University, USA*
Damianos Chatziantoniou / *Athens University of Economics and Business, Greece*
Dan S. Petty / *North Texas Commission, USA*
Dane Peterson / *Southwest Missouri State University, USA*
Daniel Brandon, Jr. / *Christian Brothers University, USA*
Daniel J. Buehrer / *National Chung Cheng University, Taiwan, R.O.C.*
Daniel J. Power / *University of Northern Iowa and DSSResources.com, USA*
Daniel Riesco / *Universidad Nacional de San Luis, Argentina*
Daniel Romero / *Universidad Nacional de Rio Cuarto, Argentina*
Daniel Schwabe / *Departamento de Informática, PUC-Rio, RJ, Brazil*
Danny A. Samson / *University of Melbourne, Australia*
Darleen M. DeRosa / *Right Management Consultants, USA*
Darren Dalcher / *Middlesex University, UK*
David A. Banks / *University of South Australia, Adelaide, Australia*
David Bryde / *Liverpool John Moore University, UK*
David Gefen / *Drexel University, USA*
David J. Pauleen / *Victoria University of Wellington, New Zealand*
David Knight / *Brunel University, UK*
David Petie / *Petie Ltd, UK*
David Robinson / *St. Cloud State University, USA*
David Ruppel / *The University of Toledo, USA*
David Sammon / *University College Cork, Ireland*
David Taniar / *Monash University, Australia*
Davood Askarany / *University of South Australia, Australia*
Deborah Bunker / *University of New South Wales, Australia*
Deborah Bursa / *Georgia Institute of Technology, USA*
Dedric A. Carter / *Nova Southeastern University, USA*
Denis Trcek / *University of Primorska, Koper, Slovenia*
Denise Fukumi Tsunoda / *CEFET-PR, Brazil*
Denise Johnson McManus / *Wake Forest University, USA*
Dennis E. Tucker / *Yaasuo Consulting Company, USA*
Dennis Guster / *St. Cloud State University, USA*
Derrie Perez / *University of South Florida, USA*
Detmar Straub / *Georgia State University, USA*
Diane H. Sonnenwald / *Goteborg University, Sweden*
Dianne Willis / *Leeds Metropolitan University, UK*
Dimitrios Katsaros / *Aristotle University of Thessaloniki, Greece*
Dimitris Plexousakis / *Institute of Computer Science, FORTH, Greece*
Dirk Trossen / *Nokia Research Center, USA*
Dirk Vriens / *Radbound University of Nijmegen, The Netherlands*
Dirk Werth / *German Research Center for Artificial Intelligence, Germany*
Divyakant Agrawal / *University of California, Santa Barbara, USA*
Dohoon Kim / *Kyung Hee University, Korea*
Dolores Cuadra / *Universidad Carlos III de Madrid, Spain*
Donald A. Hantula / *Temple University, USA*
Donald R. Morris-Jones / *SRA, USA*
Donaldo de Souza Dias / *Federal University of Rio de Janeiro, Brazil*
Donna Ure / *Brigham Young University, USA*
Dorothy G. Dologite / *Baruch College, USA*
E. Kapetanios / *Semantic Technologies, Switzerland*
E. Vázquez-Fernández / *Public Health Department of Galicia, Spain*
E.V. Krishnamurthy / *Australian National University, Australia*
Earl Woodruff / *OISE - University of Toronto, Canada*

Eberhard Stickel / *Hochschule der Sparkassen – Finanzgruppe and University of Applied Sciences-Bonn GmbH, Germany*

Edilberto Casado / *Gerens Escuela de Gestion y Economia, Peru*

Edilberto Magalhães Silva / *Brazilian Public News Agency, Brazil*

Edilson Ferneda / *Catholic University of Brasília, Brazil*

Eduardo Rodriguez / *Universidad de Buenos Aires, Argentina*

Eduardo Salas / *University of Central Florida, USA*

Edward J. Szewczak / *Canisius College, USA*

Edward Szczerbicki / *The University of Newcastle, Australia*

Eileen M. Trauth / *The Pennsylvania State University, USA*

Eitel J.M. Lauría / *Marist College, USA*

Elaheh Pourabbas / *National Research Council, Rome, Italy*

Elaine R. Winston / *Hofstra University, USA*

Elayne Coakes / *University of Westminster, UK*

Elena Castro / *Universidad Carlos III de Madrid, Spain*

Elena Karahanna / *University of Georgia, USA*

Elizabeth Furtado / *Universidade de Fortaleza, Brazil*

Elizabeth Smythe / *Concordia University College of Alberta, Canada*

El-Sayed Abou-Zeid / *Concordia University, Canada*

Emma Tonkin / *University of Karlsruhe, Germany*

Emmanuelle Vaast / *Long Island University, USA*

Eric Pardede / *La Trobe Univresity, Australia*

Erick Slazinski / *Purdue University, USA*

Erik Buelen / *Tilburg University, The Netherlands*

Erik Molenaar / *Technical University of Aachen, Germany*

Erwin Van den Enden / *Institute of Tropical Medicine, Belgium*

Eshaa M. Alkhalifa / *University of Bahrain, Bahrain*

Esperanza Marcos Martínez / *Universidad Rey Juan Carlos, Spain*

Ester Zumpano / *DEIS Università della Calabria, Italy*

Fabio A. Schreiber / *Politecnico di Milano, Italy*

Fadi P. Deek / *New Jersey Institute of Technology, USA*

Fahui Wang / *Northern Illinois University, USA*

Fang Zhao / *RMIT University, Australia*

Farhad Daneshgar / *University of New South Wales, Australia*

Fay Cobb Payton / *North Carolina State University, USA*

Fazli Erbas / *University of Hanover, Germany*

Felix B. Tan / *Auckland University of Technology, New Zealand*

Fergal Carton / *University College Cork, Cork, Ireland*

Fernando José Barbin Laurindo / *University of São Paulo, Brazil*

Floriana Grasso / *Liverpool University, UK*

Folker den Braber / *SINTEF ICT, Norway*

France Bélanger / *Virginia Polytechnic Institute and State University, USA*

Francisco Cervantes / *Mexico Autonomous Institute of Technology (UNAM), Mexico*

Francisco-Javier Díaz-Pernas / *University of Valladolid, Spain*

Françoise Preteux / *Institut National des Télécommunications, France*

Frank E. Ritter / *The Pennsylvania State University, USA*

Fred L. Kitchens / *Ball State University, USA*

Fred Niederman / *Saint Louis University, USA*

Frederic Adam / *University College Cork, Ireland*

Frederick Ericsson / *Örebro University, Sweden*

Fredrik Vraalsen / *SINTEF ICT, Norway*

Freek Ebeling / *Telematica Instituut, The Netherlands*

G. Chakraborty / *Iwate Prefectural University, Japan*

G. Eleftherakis / *City College, Greece*

G. Peter Zhang / *Georgia State University, USA*
Gabor Hosszu / *Budapest University of Technology and Economics, Hungary*
Gaelle Picherit-Duthler / *University of North Carolina Charlotte, USA*
Galen Sevcik / *Georgia State University, USA*
Garp Choong Kim / *Inha University, Korea*
Gary F. Kohut / *University of North Carolina at Charlotte, USA*
Gee-Woo (Gilbert) Bock / *National University of Singapore, Singapore*
Georg Disterer / *University of Applied Sciences and Arts, Hanover/Germany*
Georg Peters / *Munich University of Applied Sciences, Germany*
George Doukidis / *Athens University of Economics and Business, Greece*
George M. Giaglis / *Athens University of Economics and Business, Greece*
George Pallis / *Aristotle University of Thessaloniki, Greece*
Geraldine Clarebout / *University of Leuven, Belgium*
Gerardo Canfora / *University of Sannio, Italy*
German Montejano / *Universidad Nacional de San Luis, Argentina*
Giancarlo Fortino / *Università della Calabria, Italy*
Gillian H. Wright / *Manchester Metropolitan University Business School, UK*
Gina Green / *Baylor University, USA*
Giorgos Zacharia / *MIT, USA*
Giovanni Vanini / *Politecnico di Milano, Italy*
Giuseppi Pozzi / *Politecnico di Milano, Italy*
Glenn Lowry / *United Arab Emirates University, UAE*
Glenn Russell / *Monash University, Australia*
Goran D. Putnik / *University of Minho, Portugal*
Göran Pulkkis / *Arcada Polytechnic, Helsinki, Finland*
Gordana Jovanovic-Dolecek / *National Institute of Astrophysics Optics and Electronics, INAOE, Mexico*
Graeme Richards / *University of East Anglia, UK*
Graham Pervan / *Curtin University of Technology, Australia*
Greg Chiasson / *PRTM, UK*
Grigoris Antoniou / *Institute of Computer Science, FORTH, Greece*
Guadalupe Salazar-Zárate / *Technical University of Catalonia, Spain*
Guisseppi Forgionne / *University of Maryland Baltimore County, USA*
Gustavo Rossi / *UNLP, La Plata, Argentina*
Ha Sung Hwang / *Temple University, USA*
Hamada H. Ghenniwa / *University of Western Ontario, Canada*
Hanni Muukkonen / *University of Helsinki, Finland*
Hanns-Christian Hanebeck / *GlobeRanger Corporation and Univerity of Dallas, USA*
Hannu Salmela / *Institute of Information Systems Science, Finland*
Hao Lou / *Ohio University, USA*
Hareton Leung / *Hong Kong Polytechnic University, Hong Kong*
Harold Webb / *The University of Tampa, USA*
Hassan A. Karimi / *University of Pittsburgh, USA*
Haydee M. Cuevas / *University of Central Florida, USA*
Hector Perez-Meana / *National Polytechnic Institute of Mexico, Mexico*
Heidrun Allert / *University of Hannover, Germany*
Heitor S. Lopes / *CEFET-PR, Brazil*
Helen Mitchell / *Unitec, Auckland, New Zealand*
Helen Thompson / *University of Ballarat, Australia*
Hendrik Decker / *Instituto Tecnológico de Informática, Spain*
Henry Aigbedo / *Oakland University, Michigan, USA*
Herbert J. Mattord / *Kennesaw State University, USA*
Herbert Kotzab / *Copenhagen Business School, Denmark*
Herberto Ochoa-Morales / *University of New Mexico, USA*

Hércules Antonio do Prado / *Brazilian Enterprise for Agricultural Research and Catholic University of Brasília, Brazil*
Hock Chuan Chan / *National University of Singapore, Singapore*
Holly Yu / *California State University, USA*
Hong Va Leong / *The Hong Kong Polytechnic University, Hong Kong*
Houston Hume Carr / *Auburn University, USA*
H. D. Richards / *Orion Logic Ltd, UK*
Hsien-Chu Wu / *National Taichung Institute of Technology, Taiwan, R.O.C.*
Hung-Jen Yang / *University of North Dakota, USA*
I. Navas Delgado / *Universidad de Málaga, Spain*
I.M. Jawahar / *Illinois State University, USA*
Ian K. Wong / *Queens University, Canada*
Ido Millet / *Penn State Erie, USA*
Ikuo Kitagaki / *Hiroshima University, Japan*
Imed Kacem / *Université de Technologie de Troyes, France*
In Lee / *Western Illinois University, USA*
Ionnis Tarnanas / *Kozani University of Applied Sciences, Greece*
Irena Ali / *DSTO Fernhill, Australia*
Irvine Clarke III / *James Madison University, USA*
Isabel Ramos / *Universidade do Minho, Portugal*
Isola Ajiferuke / *University of Western Ontario, Canada*
István Mezgár / *Budapest University of Technology and Economics, Hungary and Hungarian Academy of Sciences, Budapest, Hungary*
J. Ángel Velázquez-Iturbide / *Universisdad Rey Juan Carlos, Spain*
J. Wenny Rahayu / *La Trobe University, Australia*
J. William Holland / *Georgia Bureau of Investigation, USA*
J.F. Aldana Montes / *Universidad de Málaga, Spain*
J.H. Erik Andriessen / *Delft University of Technology, The Netherlands*
J.K. Suresh / *Infosys Technologies Ltd., India*
Jae-Woo Chang / *Chonbuk National University, South Korea*
Jakov Crnkovic / *University of Albany, USA*
James Griffin / *Tipperary Institute, Ireland*
James R. Lindner / *Texas A&M University, USA*
Jan Achterbergh / *Radboud University of Nijmegen, The Netherlands*
Jan Elen / *University of Leuven, Belgium*
Jana Polgar / *Monash University, Australia*
Jane E. Klobas / *Bocconi University, Italy and University of Western Australia, Italy*
Jane L. Garb / *Baystate Medical Center, USA*
Janet Toland / *Victoria University of Wellington, New Zealand*
Janette Moody / *The Citadel, USA*
Janice C. Sipior / *Villanova University, USA*
Janos Fustos / *Metropolitan State College of Denver, USA*
Janse Tolmie / *University of the Free State, South Africa*
Jari Veijalainen / *University of Jyväskyä, Finland*
Jason C.H. Chen / *Ganzaga University, USA*
Jatinder N.D. Gupta / *University of Alabama in Huntsville, USA*
Javier Díaz-Carmona / *National Institute of Astrophysics Optics and Electronics, INAOE and Technology Institute of Celaya (ITC), Celaya, Mexico*
Jean-Charles Pomerol / *Université Pierre et Marie Curie, France*
Jean-Claude Simon / *Université de Bourgogne, France*
Jef Van den Ende / *Institute of Tropical Medicine, Belgium*
Jeffrey L. Brewer / *Purdue University, USA*
Jeffrey Parsons / *Memorial University of New Foundland, Canada*

Jeffrey Stanton / *Syracuse University, USA*
Jennifer Paige Nightingale / *Duquesne University, USA*
Jenq-Foung Yao / *Georgia College and State University, USA*
Jerome Darmont / *Universite Lumiere Lyon 2, France*
Jerzy Kisielnicki / *Warsaw University, Poland*
Jilin Tu / *University of Illinois at Urbana-Champaign, USA*
Jim O. Otieno / *Middlesex University, UK*
Jinyoul Lee / *State University of New York at Binghamton, USA*
Jo Ann Oravec / *University of Wisconsin-Whitewater, USA*
Jo Ellen Moore / *Southern Illinois University Edwardsville, USA*
Joan Pastor / *Universidad Politecnica de Catalunya, Spain*
Joan Pelland / *New College of Florida, USA*
João Álvaro Carvalho / *Universidade do Minho, Portugal*
Joaquim Armando Pires Jorge / *Technical University if Lisbon, Portugal*
Joe Gilbert / *University of Nevada, USA*
Johan M. Karlsson / *Lund Institute of Technology, Sweden*
Johanna Lammintakanen / *University of Kuopio, Finland*
John A. Hoxmeier / *Colorado State University, USA*
John A. Nicholson / *Utah State University, USA*
John B. Nash / *Stanford University, USA*
John C. McIntosh / *Bentley College, USA*
John Day / *Ohio University, USA*
John H. Nugent / *University of Dallas, USA*
John Krogstie / *NTNU and SINTEF, Norway*
John M. Artz / *The George Washington University, USA*
John M. Carroll / *The Pennsylvania State University, USA*
John Mendonca / *Purdue University, USA*
John R. Durrett / *Texas Tech University, USA*
John S. Edwards / *Aston Business School, UK*
John T. Riedl / *University of Minnesota, USA*
John W. Priest / *University of Texas at Arlington, USA*
Jon Dron / *University of Brighton, UK*
Jon Oldevik / *SINTEF ICT, Norway*
Jon Patrick / *University of Sydney, Australia*
Jonathan Frank / *Suffolk University, USA*
Joost Lowyck / *K.U. Leuven, Belgium*
Jorge H. Doorn / *Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina*
Jörgen S. Svensson / *University of Twente, The Netherlands*
Joris P.G. de Rooij / *Delft University of Technology, The Netherlands*
José Esteves / *Instituto de Empresa, Spain*
Jose M. Framinan / *University of Seville, Spain*
José María Cavero Barca / *Universidad Rey Juan Carlos, Spain*
José Palazzo Moreira de Oliveira / *Federal University of Rio Grande do Sul, Brazil*
José-Fernando Díez-Higuera / *University of Valladolid, Spain*
Joseph A. Konstan / *University of Minnesota, USA*
Joseph Bradley / *Central Washington University, USA*
Joseph Sarkis / *Clark University, USA*
Jounghae Bang / *University of Rhode Island, USA*
Juan M. Ale / *Universidad de Buenos Aires, Argentina*
Judith Bishop / *University of Pretoria, South Africa*
Judith V. Boettcher / *University of Florida, USA*
Judith C. Simon / *The University of Memphis, USA*
Judith Jeffcoate / *University of Buckingham, UK*
Judith Kabeli / *Ben-Gurion University, Israel*

Juergen Seitz / *University of Cooperative Education Heidenheim, Germany*
Juha Pärnistö / *Fujitsu Services, Finland*
Julie Mackey / *Christchurch College of Education, New Zealand*
Jyh-haw Yeh / *Boise State University, USA*
Kai Hakkarainen / *University of Helsinki, Finland*
Kaija Saranto / *University of Kuopio, Finland*
Kaj Grahn / *Arcada Polytechnic Helsinki, Finland*
Kalpdrum Passi / *Laurentian University, Canada*
Kambiz E. Maani / *University of Auckland, New Zealand*
Kan Hou Vat / *University of Macau, China*
Karen D. Schenk / *K.D. Schenk and Associates Consulting, USA*
Karen Neville / *National University of Ireland, Cork, Ireland*
Karen S. Nantz / *Eastern Illinois University, USA*
Karim K. Hirji / *AGF Management Ltd., Canada*
Karl Kurbel / *European University Viadrina Frankfurt (Oder), Germany*
Kate Kaiser / *Marquette University, USA*
Kathryn A. Marold / *Metropolitan State College of Denver, USA*
Kathy Arsenault / *University of South Florida St. Petersburg, USA*
Katy Campbell / *University of Alberta, Canada*
Keith C.C. Chan / *The Hong Kong Polytechnic University, Hong Kong*
Keng L. Siau / *University of Nebraska-Lincoln, USA*
Kenneth J. MacGregor / *University of Cape Town, South Africa*
Kerrie Mengersen / *The University of Newcastle, Australia*
Ketil Stølen / *SINTEF ICT, Norway*
Kevin E. Voges / *University of Canterbury, New Zealand*
Kevin Johnston / *University of Cape Town, South Africa*
Khamis Al-Gharbi / *Sultan Qaboos University, Oman*
Kim E. Dooley / *Texas A&M University, USA*
Kim Man Lui / *The Hong Kong Polytechnic University, Hong Kong*
Kin Choong Yow / *Nanyang Technological University, Singapore*
Kirk St.Amant / *Texas Tech University, USA*
Kirsimarja Blomqvist / *Lappeenranta University, Finland*
Kokou Yétongnon / *Bourgogne University, France*
Konrad J. Peszynski / *Deakin University, Australia*
Konstantina Stoupa / *Aristotle University of Thessaloniki, Greece*
Kostas Karpouzis / *National Technical University of Athens, Greece*
Kovach Dragan / *SUPER-King Inc., Croatia*
Kristiina Häyrinen / *University of Kuopio, Finland*
Kristina Setzekorn / *University of Evansville, USA*
Kristine F. Dery / *University of Sydney, Australia*
Krzystof Wecel / *The Poznan University of Economics, Poland*
Kuldeep Kumar / *Bond University, Australia*
Kurt April / *University of Cape Town, South Africa*
L. E. Moser / *University of California, Santa Barbara, USA*
Lakshmi S. Iyer / *University of North Carolina, USA*
Laura C. Rivero / *Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina and Universidad Nacional de La Plata, Buenos Aires, Argentina*
Laura Felice / *Universidad Nacional del Centro de la Provincia de Beunos Aires, Argentina*
Lawrence L. Schkade / *University of Texas at Arlington, USA*
Leandro Krug Wives / *Centro Universitário Feevale and Federal University of Rio Grande do Sul, Brazil*
Leandro Nunes de Castro / *Catholic University of Santos, Brazil*
Lei Dang / *University of Manchester, UK*
Leigh M. Smith / *Curtin University of Technology, Australia*
Lejla Vrazalic / *University of Wollongong, Australia*

Len Asprey / *Practical Information Management Solutions Pty Ltd, Australia*
Leo Tan Wee Hin / *Singapore National Academy of Science and Nanyang Technological University, Singapore*
Leon Urbas / *Berlin University of Technology, Germany*
Leonardo Tininini / *Istituto di Analisi dei Sistemi e Informatica "Antonio Ruberti", Rome, Italy*
Leoni Warne / *DSTO Fernhill, Australia*
Leopoldo E. Colmenares / *University Simon Bolivar, Venezuela*
Lerina Aversano / *University of Sannio, Italy*
Letizia Tanca / *Politecnico di Milano, Italy*
Liliana Favre / *CIC, Argentina*
Liliana Martinez / *Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina*
Linda V. Knight / *DePaul University, USA*
Lisa A. Burke / *Louisianna State University Shreveport, USA*
Lisa Burnell / *Texas Christian University, USA*
Liyang Zhang / *Baidu.Com (Beijing) Co., Ltd., China*
Ljiljana D. Mili / *University of Belgrade, Serbia and Montenegro*
Lloyd D. Brooks / *The University of Memphis, USA*
Lorraine Justice / *Georgia Institute of Technology, USA*
Louise Lane / *Laurentian University, Canada*
Luca Tosi / *Comune di Milano, Italy*
Luiz A. DaSilva / *Virginia Polytechnic Institute and State University, USA*
Luiz Antonio Joia / *Brazilian School of Public and Business Administration of Getulio Vargas Foundation and Rio de Janeiro State University, Brazil*
Luiz Augusto Machado Mendes-Filho / *Faculdade Natalense para o Desenvolvimento do Rio Grande do Norte, Brazil*
Lutz Hamel / *University of Rhode Island, USA*
Luyin Zhao / *Philips Research, USA*
Lynette Kvasny / *The Pennsylvania State University, USA*
Lynne D. Roberts / *University of Western Australia, Australia*
M. Asim Qayyum / *University of Toronto, Canada*
M. Gheorghe / *University of Sheffield, UK*
M. Gordon Hunter / *University of Lethbridge, Canada*
M. Holcombe / *University of Sheffield, UK*
M. Spinu / *EXITECH S.r.L., Italy*
M. Zahirul Huq / *ICDDRB, Centre for Health and Population Research, Bangladesh*
M.E. López-Vizcaíno / *Galician Statistical Institute, Spain*
M.I. Santiago-Pérez / *Public Health Department of Galicia, Spain*
M.M. Roldan Garcia / *Universidad de Málaga, Spain*
Maciej Zakrzewicz / *Poznan University of Technology, Poland*
Maggie McPherson / *University of Sheffield, UK*
Mahboob ur Rahman / *Gasmin Ltd., Bangladesh*
Mahesh S. Raisinghani / *University of Dallas, USA*
Maija-Leena Huotari / *University of Helsinki, Finland*
Majed Z. Al-Hajery / *College of Business Studies, Kuwait*
Malcom King / *Loughborough University, UK*
Manoj K. Singh / *i2 Technologies, USA*
Manuel Mora / *University of Aguascalientes(UAA), Mexico*
Manuel Serrano / *University of Castilla-La Mancha, Spain*
Maosheng Lai / *Peking University, China*
Marc Holzer / *Rutgers University, Newark, USA*
Marcela Daniele / *Universidad Nacional de Rio Cuarto, Argentina*
Marcelo Schneck de Paula Pessôa / *University of São Paulo, Brazil*
Marco Roccetti / *University of Bologna, Italy*
Marek Wojciechowski / *Poznan University of Technology, Poland*
Maria Alexandra Rentroia-Bonito / *Technical University of Lisbon, Portugal*
María Carmen Leonardi / *UNCPBA, Argentina*

Maria Eugenia Iacob / *Telematica Instituut, The Netherlands*
Maria Manuela Cunha / *Polytechnic University of Cávado and Ave, Portugal*
María Virginia Mauco / *Universidad Nacional del Centro de la Pcia de Buenos Aires, Argentina*
Marijn Janssen / *Delft University of Technology, The Netherlands*
Mariko Nakano-Miyatake / *National Polytechnic Institute of Mexico, Mexico*
Mario Bochicchio / *University of Lecce, Italy*
Mario G. Piattini / *University of Castilla-La Mancha, Spain*
Marion Cottingham / *The University of Western Australia, Australia*
Marios C. Angelides / *Brunel University, UK*
Marius Preda / *Institut National des Télécommunications, France*
Mark E. Nissen / *Naval Postgraduate School, USA*
Mark Griffiths / *Nottingham Trent University, UK*
Mark I. Hwang / *Central Michigan University, USA*
Mark Kieler / *Carnegie Mellon University, USA*
Mark Snyder / *Clemson University, USA*
Mark Srite / *University of Wisconsin - Milwaukee, USA*
Mark Toleman / *University of Southern Queensland, Australia*
Mark Xu / *University of Portsmouth, UK*
Marly Monteiro de Carvalho / *University of São Paulo, Brazil*
Martha Myers / *Kennesaw State University, USA*
Martin C. Kindmüller / *Berlin University of Technology, Germany*
Martin D. Crossland / *Oklahoma State University, USA*
Martin Doerr / *Institute of Computer Science, FORTH, Greece*
Martin Gaedke / *University of Karlsruhe, Germany*
Martin Nussbaumer / *University of Karlsruhe, Germany*
Mary Elizabeth Brabston / *The University of Manitoba, Canada*
Mary Jane Lenard / *University of North Carolina - Greensboro, USA*
Mass Soldal Lund / *SINTEF ICT, Norway*
Mathias Weske / *University of Potsdam, Germany*
Matthew W. Guah / *Warwick University, UK*
Mauel Mora / *Autonomous University of Aguascalientes, Mexico*
Maureen C. Minielli / *Indiana University-Purdue University at Indianapolis, USA*
Mauricio Minuto Espil / *Universidad Católica Argentina, Argentina*
Maurizio Rafanelli / *Italian National Research Council, Italy*
Melanie Misanchuk / *University of Guelph, Canada*
Merilyn Burke / *University of South Florida, USA*
Michael Dixon / *Murdoch University, Australia*
Michael E. Whitman / *Kennesaw State University, USA*
Michael J. Mullany / *Northland Polytechnic, New Zealand*
Michael J. West / *Carnegie Mellon University, USA*
Michael Middleton / *Queensland University of Technology, Australia*
Michael N. Huhns / *University of South Carolina, USA*
Michael W. Dixon / *Murdoch University, Australia*
Michele Zappavigna-Lee / *University of Sydney, Australia*
Michelle M. Ramim / *Nova Southeastern University, USA*
Miguel Felder / *Universidad de Buenos Aires, Argentina*
Mihir A. Parikh / *University of Central Florida, USA*
Mikael Wiberg / *Umea University, Sweden*
Mikolaj Morzy / *Poznan University of Technology, Poland*
Mila Gascó-Hernández / *Open University of Catalonia, Spain*
Mimi Kessler / *Georgia Institute of Technology, USA*
Minaz J. Parmar / *Brunel University, UK*
Ming Yung Tsai / *Chaoyang University of Technology, Taiwan, R.O.C.*
Minna Lakkala / *University of Helsinki, Finland*

Mirek Riedewald / *Cornell University, USA*
Mirja Iivonen / *University of Tampere, Finland*
Mizue Kayama / *Senshu University, Japan*
Moez Limayem / *City University of Hong Kong, Hong Kong*
Mohammad A. Rashid / *Massey University - Albany, New Zealand*
Mohammad Ali / *International Vaccine Institute, Korea*
Molly Wasko / *Florida State University, USA*
Monica T. Whitty / *Queen's University Belfast, Northern Ireland*
Monideepa Tarafdar / *University of Toledo, USA*
Morgan Jennings / *Metropolitan State College of Denver, USA*
Mukesh Mohania / *IBM India Research Lab, India*
Murray E. Jennex / *San Diego State University, USA*
Murray Turoff / *New Jersey Institute of Technology, USA*
N. Moreno Vergara / *Universidad de Málaga, Spain*
N. Shiratori / *Tohoku University, Japan*
Nabeel A.Y. Al-Qirim / *United Arab Emirates University, UAE*
Nadia N. Moertiyoso / *Hewlett-Packard Singapore, Singapore*
Ned Kock / *Texas A&M International University, USA*
Neeraj Parolia / *University of Central Florida, USA*
Neil F. Doherty / *Loughborough University, UK*
Nicholas Beaumont / *Monash University, Australia*
Nicola Fiore / *University of Lecce, Italy*
Nicolas Tsapatsoulis / *National Technical University of Athens, Greece*
Nikhilesh Dholakia / *University of Rhode Island, USA*
Niki Aifanti / *Informatics & Telematics Institute, Greece*
Niki Panteli / *University of Bath, UK*
Nikos Grammalidis / *Informatics & Telematics Institute, Greece*
Nikos Karacapilidis / *University of Patras, Greece*
Nilmini Wickramasinghe / *Cleveland State University, USA*
Nir Kshetri / *The University of North Carolina at Greensboro, USA*
Nitaya Wongpinunwatana / *Thammasat University, Thailand*
Norm Archer / *McMaster University, Canada*
Norman F. Schneidewind / *Naval Postgraduate School, USA*
Norman Pendegraft / *University of Idaho, USA*
Nory B. Jones / *University of Maine Business School, USA*
O. Kao / *University of Paderborn, Germany*
Olena Daly / *Monash University, Australia*
Olga De Troyer / *WISE Research Group, Vrije Universiteit Brussel, Belguim*
Ovsei Gelman / *National Autonomous University of Mexico (UNAM), Mexico*
P. Kefalas / *City College, Greece*
P.M. Melliar-Smith / *University of California, Santa Barbara, USA*
P. Nesi / *University of Florence, Italy*
Pallab Saha / *National University of Singapore, Singapore*
Paloma Martínez / *Universidad Carlos III de Madrid, Spain*
Panayiotis Zaphiris / *City University, London, UK*
Pasi Tiitinen / *University of Jyväskylä, Finland*
Pasi Tyrväinen / *University of Jyväskylä, Finland*
Patricia L. Rogers / *Bemidji State University, USA*
Patrick R. Mullen / *U.S. General Accountability Office, Washington, USA*
Paul Darbyshire / *Victoria University, Australia*
Paul Gruenewald / *Prevention Research Center, USA*
Paul Hildreth / *K-Now International Ltd., UK*
Paul Johannesson / *Stockholm University and Royal Institute of Technology, Sweden*
Paul Lajbcygier / *Monash University, Australia*

Paul Leng / *Liverpool University, UK*
Paul M. Leonardi / *Stanford University, USA*
Paul Mulligan / *Babson College, USA*
Paul N.J. Hendricks / *University of Nijmegen, The Netherlands*
Paul Rippon / *The University of Newcastle, Australia*
Paul Safonov / *St. Cloud State University, USA*
Pawel Jan Kalczynski / *University of Toledo, USA*
Peik Åström / *Arcada Polytechnic, Helsinki, Finland*
Pengyu Hong / *Harvard University, USA*
Penny Baillie-de Byl / *University of Southern Queensland, Australia*
Pere Botella / *Technical University of Catalonia, Spain*
Peretz Shoval / *Ben-Gurion University, Israel*
Pervaiz Alam / *Kent State University, USA*
Petek Askar / *Hacettepe University, Turkey*
Peter Aiken / *Defense Information Systems Agency, USA*
Peter Fettke / *Johannes Gutenberg-University Mainz, Germany*
Peter Goldschmidt / *The University of Western Australia, Australia*
Peter J. Smith / *Athabasca University, Canada*
Peter Kim Streatfield / *Center for Population and Health Research, Bangladesh*
Peter Murphy / *Victoria University of Wellington, New Zealand*
Peter Rittgen / *University College of Borås, Sweden*
Petia Todorova / *Fraunhofer FOKUS, Germany*
Petros A.M. Gelepithis / *Kingston University, UK*
Petros Theodorou / *Technological Education Institution of Kavala, Greece*
Philip Duchastel / *Information Design Atelier, Canada*
Philip J. Dobson / *Edith Cowan University, Australia*
Philip L. Powell / *University of Bath, UK*
Phyllis Ruscella / *University of South Florida, USA*
Piet Boekhoudt / *Telematica Instituut, The Netherlands*
Pieter Blignaut / *University of the Free State, South Africa*
Pombortsis Andreas / *Aristotle University of Thessaloniki, Greece*
Pouwan Lei / *University of Sussex, UK*
Pruthikrai Mahatanankoon / *Illinois State University, USA*
Purnendu Mandal / *Lamar University, USA*
R. Kelly Rainer / *Auburn University, USA*
R. Subramaniam / *Singapore National Academy of Science, Singapore and Nanyang Technological University, Singapore*
R.A. Haraty / *Lebanese American University, Lebanon*
R.P. Sundarraj / *University of Waterloo, USA*
Rafael S. Parpinelli / *CEFET-PR, Brazil*
Rajesh Chandrashekaran / *Fairleigh Dickinson University, USA*
Ramesh Subramanian / *Quinnipiac University, USA*
Ratchata Peachavanish / *University of Pittsburgh, USA*
Rebecca Watson / *Deloitte, Australia*
Reima Suomi / *Turku School of Economics and Business Administration, Finland*
Rezaul Begg / *Victoria University, Australia*
Richard B. Wait / *Baystate Medical Center, USA*
Richard C. Hicks / *Texas A&M University, USA*
Richard Chbeir / *Bourgogne University, France*
Richard Discenza / *University of Colorado at Colorado Springs, USA*
Richard J. Hartley / *Manchester Metropolitan University, UK*
Richard P. Bagozzi / *Rice University, USA*
Richard Schwester / *Rutgers University, Newark, USA*
Richard T. Herschel / *St. Joseph's University, USA*

Richi Nayak / *Queensland University of Technology, Australia*
Rick Gibson / *American University, USA*
Rino Falcone / *Institute for Cognitive Sciences and Technology, National Research Council of Italy, Italy*
Robert A. Schultz / *Woodbury University, USA*
Robert Barone / *Buffalo State College, USA*
Robert D. Sprague / *University of Wyoming, USA*
Robert G. Dyson / *University of Bath, UK*
Robert Gay / *Nanyang Technological University, Singapore*
Robert J. Mockler / *St. John's University, USA*
Robert Joseph Skovira / *Robert Morris University, USA*
Robert Lipton / *Prevention Research Center, USA*
Robert M. Verburg / *Delft University of Technology, The Netherlands*
Robert Mac Gregor / *University of Wollongong, Australia*
Robert S. Owen / *Texas A&M University- Texarkana, USA*
Robert W. Holt / *Ganzaga University, USA*
Roberto Evaristo / *University of Illinois, Chicago, USA*
Roberto Hirsh / *Universidad de Buenos Aires, Argentina*
Roberto Paiano / *University of Lecce, Italy*
Robin Teigland / *Stockholm School of Economics, Sweden*
Rodney Nason / *Queensland University of Technology, Australia*
Rodney Turner / *Victoria University of Technology, Australia*
Roger McHaney / *Kansas State University, USA*
Roger Tagg / *University of South Australia, Australia*
Ron Craig / *Wilfrid Laurier University, Canada*
Ronald B. Wilkes / *The University of Memphis, USA*
Ronald Dattero / *Southwest Missouri State University, USA*
Ronaldo Zwicker / *University São Judas Tadeu, Brazil*
Ruidong Zhang / *University of Wisconsin-Eau Claire, USA*
Rung Ching Chen / *Chaoyang University of Technology, Taiwan, R.O.C.*
Rupert Young / *University of Sussex, UK*
Ruth A. Guthrie / *California State Polytechnic University, Pomona, USA*
Ryan R. Peterson / *Information Management Research Center, Spain*
S. Geisler / *University of Clausthal, Germany*
Sagarmay Deb / *University of Southern Queensland, Australia*
Sai Ho Kwok / *California State University, USA*
Salvatore Valenti / *Università Politecnica delle Marche, Italy*
Sam Lubbe / *University of KwaZulu-Natal, South Africa*
Sandro Leuchter / *Berlin University of Technology, Germany*
Sanjay Madria / *University of Missouri-Rolla, USA*
Sara Dexter / *University of Virginia, USA*
Sari Rissanen / *University of Kuopio, Finland*
Schahram Dustdar / *Vienna University of Technology, Austria*
Scott Nicholson / *Syracuse University, USA*
Sergio Greco / *DEIS Università della Calabria, Italy*
Seung-Kyoon Shin / *University of Rhode Island, USA*
Shamik Sural / *Indian Institute of Technology, Kharagpur, India*
Shannon Howle Schelin / *University of North Carolina at Chapel Hill, USA*
Shantha Fernando / *University of Moratuwa, Sri Lanka*
Sharmila Pixy Ferris / *William Patterson University, USA*
Shawn D. Long / *University of North Carolina at Charlotte, USA*
Sheng-Uei Guan / *National University of Singapore, Singapore*
Sherif Kamel / *The American University at Cairo, Egypt*
Sherry Y. Chen / *Brunel University, UK*
Sheryl Burgstahler / *University of Washington, USA*

Shin-Ping Liu / *Jarvis Christian College, USA*
Shirley Ann Becker / *Florida Institute of Technology, USA*
Simon Dixon / *Austrian Research Institute for Artificial Intelligence, Austria*
Simpson Poon / *IVE, Hong Kong and Charles Sturt University, Australia*
Sotiris Malassiotis / *Informatics & Telematics Institute, Greece*
Spiros Ioannou / *National Technical University of Athens, Greece*
Stacie Petter / *Georgia State University, USA*
Stanley Loh / *Catholic University of Pelotas and Lutheran University of Brazil, Brazil*
Stefano Ferretti / *University of Bologna, Italy*
Stefano Renzi / *Bocconi University, Italy*
Stefanos Kollias / *National Technical University of Athens, Greece*
Stephan Olariu / *Old Dominion University, USA*
Stephanie Allen / *Brigham Young University, USA*
Stephen Burgess / *Victoria University, Australia*
Stephen Hawk / *University of Wisconsin-Parkside, USA*
Stephen M. Fiore / *University of Central Florida, USA*
Steve Sawyer / *The Pennsylvania State University, USA*
Steven A. Morris / *Middle Tennessee State University, USA*
Steve Gordon / *Babson College, USA*
Steven Utsi / *K.U. Leuven, Belgium*
Stu Westin / *University of Rhode Island, USA*
Stuart D. Galup / *Florida Atlantic University, USA*
Stuart J. Barnes / *Victoria University of Wellington, New Zealand*
Subhasish Dasgupta / *The George Washington University, USA*
Subramanian Sivaramakrishnan / *University of Manitoba, Canada*
Sue E. Kase / *The Pennsylvania State University, USA*
Supawadee Ingsriswang / *BIOTEC and NSTDA, Thailand*
Susan A. Baim / *Miami University Middletown, USA*
Susan E. George / *University of South Australia, Australia*
Sushil K. Sharma / *Ball State University, USA*
Susy S. Chan / *DePaul University, USA*
Suzanne M. Embury / *University of Manchester, UK*
Suzanne T. Bell / *University of Guam, Guam*
Sven A. Carlsson / *Lund University, Sweden*
T.T. Wong / *The Hong Kong Polytechnic University, Hong Kong*
Tadeusz Morzy / *Poznan University of Technology, Poland*
Tamara Babaian / *Bentley College, USA*
Tamio Shimizu / *University of São Paulo, Brazil*
Tanya J. McGill / *Murdoch University, Australia*
Ta-Tao Chuang / *Gonzaga University, USA*
Terence Cavanaugh / *University of North Florida, USA*
Teresa Ludermir / *UFPE, Brazil*
Teresa Torres-Coronas / *Universitat Rovira i Virgili, Spain*
Terry Halpin / *Northface University, USA*
Theo McDonald / *University of the Free State, South Africa*
Theresa A. Steinbach / *DePaul University, USA*
Theresa Flaherty / *James Madison University, USA*
Theresa M. Vitolo / *Gannon University, USA*
Theresa P. Murphrey / *Texas A&M University, USA*
Thomas E. Marshall / *Auburn University, USA*
Thomas F. Stafford / *University of Memphis, USA*
Thomas Kunz / *Carleton University, Canada*
Thomas S. Huang / *Beckman Institute, USA*
Tiehan Lv / *Princeton University, USA*

Till Hänisch / *University of Cooperative Education, Heidenheim, Germany*
Tim Coltman / *University of Wollongong, Australia*
Timothy H. Murphy / *Texas A&M University, USA*
Tiok-Woo Teo / *Bond University, Australia*
Tony Polgar / *Coles Myer Pty. Ltd., Australia*
Tor Guimaraes / *Tennessee Technological University, USA*
Toshio Okamoto / *University of Electro-Communications, Japan*
Tzu-Chuan Chou / *University of Bath, UK*
Ugur Halici / *Middle East Technical University, Turkey*
Utpal M. Dholakia / *Rice University, USA*
V. P. Kochikar / *Infosys Technologies Ltd., India*
V.K. Murthy / *University of New South Wales, Australia*
V.K. Narayanan / *Drexel University, USA*
Valentina Plekhanova / *University of Sunderland, UK*
Valeria De Antonellis / *University of Brescia, Italy*
Valerie Baker / *University of Wollongong, Australia*
Vassilios Kikis / *Kozani University of Applied Sciences, Greece*
Vassilis Christophides / *Institute of Computer Science, FORTH, Greece*
Vassilis Serafeimidis / *Atos KPMG Consulting, UK*
Venkat Reddy / *University of Colorado at Colorado Springs, USA*
Vicki L. Gregory / *University of South Florida, USA*
Victoria Youngohc Yoon / *University of Maryland, USA*
Vincent Yen / *Wright State University, USA*
Virpi Lyytikäinen / *University of Jyväskylä, Finland*
Vladimir A. Kulyukin / *Utah State University, USA*
Vladimír Modrák / *Technical University of Košice, Slovakia*
W. Andrew Taylor / *University of Bradford, UK*
W. Zhao / *University of California, Santa Barbara, USA*
Wai Law / *University of Guam, Guam*
Waleed E. Farag / *Indiana University of Pennsylvania, USA*
Walter D. Fernández / *Australian National University, Australia*
Wayne W. Huang / *Ohio University, USA*
Wayne Wolf / *Princeton University, USA*
Wei K. Lou / *Northern Illinois University, USA*
Wei-Ping Tan / *National University of Singapore, Singapore*
Wen-Chen Hu / *University of North Dakota, USA*
Wendy L. Currie / *Warwick University, UK*
Wendy Lucas / *Bentley College, USA*
Wesley J. Johnston / *Georgia State University, USA*
Wilfried Lemahieu / *Katholieke Universiteit Leuven, Belgium*
William F. Wieczorek / *Buffalo State College-State University of New York, USA*
William H. Friedman / *University of Central Arkansas, USA*
William J. Switala / *Duquesne University, USA*
William K. Holstein / *The College of William and Mary and*
Graduate School of Business Administration-Zurich, USA/Switzerland
William R. Gates / *Naval Postgraduate School, USA*
William S. Lightfoot / *International University of Monaco, Monaco*
William T. Hewitt / *University of Manchester, UK*
Witold Abramowicz / *The Poznan University of Economics, Poland*
Wole Olatokun / *University of Ibadan, Nigeria*
Woojong Suh / *Inha University, Korea*
X. Hervada-Vidal / *Public Health Department of Galicia, Spain*
Xiannong Meng / *Bucknell University, USA*
Xiaohui Liu / *Brunel University, UK*

Xiaotong Li / *University of Alabama in Huntsville, USA*
Xiaowen Fang / *DePaul University, USA*
Xin Fu / *University of North Carolina at Chapel Hill, USA*
Xue Li / *University of Queensland, Australia*
Y. S. Chang / *HanKuk Aviation University, Korea and Orion Logic Ltd., UK*
Yair Levy / *Nova Southeastern University, USA*
Yangjun Chen / *University of Winnipeg, Canada*
Yannis Manolopoulos / *Aristotle University of Thessaloniki, Greece*
Yanqing Duan / *University of Luton, UK*
Yanyan Yang / *University of California, Davis, USA*
Yaolin Gu / *Southern Yangtze University, China*
Ye Diana Wang / *University of Maryland Baltimore County, USA*
Yen-Ping Chu / *National Chung Hsing University, Taiwan, R.O.C.*
Ye-Sho Chen / *Louisianna State University, USA*
Yining Chen / *Ohio University, USA*
Yin-Leng Theng / *Nanyang Technological University, Singapore*
Young-Gul Kim / *KAIST, Korea*
Yuan Gao / *Ramapo College of New Jersey, USA*
Yu-An Ho / *National Chung Hsing University, Taiwan, R.O.C.*
Yue "Jeff" Zhang / *California State University, USA*
Yu-Jin Zhang / *Tsinghua University, Beijing, China*
Yukiko Inoue / *University of Guam, Guam*
Yung-Kuan Chan / *National Chung Hsing University, Taiwan, R.O.C.*
Zaiyong Tang / *Louisianna Tech University, USA*
Zhen Wen / *University of Illinois at Urbana-Champaign, USA*
Zhensen Huang / *University of Maryland Baltimore County, USA*
Zhixiang Chen / *University of Texas-Pan American, USA*
Zhonghua Yang / *Nanyang Technological University, Singapore and Southern Yangtze University, China*
Zongqiao Xiao / *Georgia College and State University, USA*
Zoran Stojanovic / *Delft University of Technology, The Netherlands*

Contents
by Volume

VOLUME I

VOLUME II

VOLUME III

VOLUME IV

VOLUME V

Contents
by Subject

Business/IT Alignment

Data Mining

Electronic Government

End User Computing

ERP

Geographic IS

Global IT Management

Human Side of IT

Information Modeling

Intelligent IS

IS Research

IS Security

IT Education

IT Evaluation

IT in Developing Countries

IT in Healthcare

IT in Libraries

IT in Small Business

IT in Personnel Management

IT Project Management

Knowledge Management

Mobile Technologies

MultiMedia

Object Oriented Technologies

Software Engineering

Strategic IT

Telecommunications and Networking

UML

Virtual Organizations

Web Applications

Preface

The information technology revolution is for all practical purposes a second industrial revolution in the history of mankind, impacting every facet of business, society, and life worldwide with a speed beyond imagination. As in the case of the industrial revolution, the growth and expansion of information technology began during the early 1950s but continued on through the following decades. During the 1960s and ‘70s, computer technology was mainly utilized in business data processing and scientific applications of a mostly number crunching nature. As a result, the use of this technology was limited to those who had a good command of these systems and computer programming languages. With the merge of computers and telecommunications in the 1980s, a new information technology was born with a strong focus on the management and dissemination of information by both information providers and users. However, the advent of personal computers and the ability of users to communicate with each other regardless of their locations directed information technology into the lives of all society.

The most noticeable explosion in the information technology revolution was the creation of the World Wide Web (WWW) and its potential in the early 1990s. During the past two decades, WWW technologies have become the driving force in allowing people worldwide to communicate and exchange information in ways that have created a totally new dimension known as a virtual world. In recent years, through the use of web-enabled technologies, organizations of all types and sizes around the world have managed to utilize these technologies to conduct both information processing and dissemination with their prospective customers, suppliers, students, and governments. These technologies, now allowing readily available information for everyone regardless of their geographic location, bring the true meaning of the information age to its full realization.

In recent years, the science of understanding the nature of information processing and management combined with computer and telecommunication technologies to process, disseminate, and manage information has become known as “Information Science and Technology.” This new science includes many resources and components, including: 1) Type of Information; 2) Computer Hardware Technologies; 3) Computer Software Technologies; 4) Telecommunication Technologies; 5) Technology Applications; 7) Information Processing Systems Structures; 8) Systems Personnel and Management; 8) End Users; and 9) Management Skills and Programs (Khosrow-Pour & Yaverbaum, 1990).

Because information science and technology has profoundly impacted science, business, society, and every other aspect of life on our planet, numerous researchers around the world have focused on accumulating knowledge on this discipline. The volume and intensity of research in the field of information science and technology has by far exceeded many other fields of science, and the sheer impact of research discoveries in this area has become the driving force of many emerging technologies and applications. No longer is this discipline limited to a few technology-related areas, but, similar to the field of medicine, the field of information science and technology today is a collection of many specific disciplines researchers have created. This collection process has been accomplished by producing research results to understand the potentials, problems, and challenges of each individual discipline and by trying to expand the body of literature related to the topics of that discipline.

To access the latest research related to the many disciplines of the information science and technology field, I decided several years ago to launch an encyclopedia project where researchers from all over the world would assist me in providing the necessary coverage of each respective discipline in information science and technology. The primary objective of this project was to assemble as much research coverage related to the disciplines selected for this encyclopedia by defining the technologies, terms, and acronyms related to each discipline, and providing the most comprehensive list of research references related to each discipline. I had hopes of creating a single comprehensive reference source on all related discipline topics of information science and technology.

To provide the most comprehensive, in-depth, and recent coverage of information science and technology, disciplines carefully selected for this encyclopedia project included: Accounting Information Systems, Computing History, Database Management and Technologies, Data Warehousing and Mining, Decision Support Systems Technologies, Distance Education Technologies, E-collaboration, Electronic Commerce Technologies Management, End User Computing, Enterprise Resource Planning, Expert Systems, Geographical Information Systems, Global IT Management, Human Computer Interaction, Human Side of IT, Information Resources Management, Information Security Management, Information Systems Research, Information Technology Education, IT Evaluation Methods and Management, IT Management in Libraries, IT Management in Healthcare, IT in Small Business, IT Personnel, Professional IT Association, Intelligent Information Systems, Knowledge Management, Minorities in Information Technology, Mobile Computing and Commerce, Multimedia Information Management, Objected Oriented Technologies, Open Source Technologies and Systems, Social Responsibility in the Information Age, Software Engineering, Strategic IT Management, Telecommunications and Networking Technologies, Unified Modeling Languages and Unified Process, and Virtual Communities and IT.

In order to provide the best balanced coverage of concepts and issues related to the selected topics of this encyclopedia, researchers from around the world were asked to submit proposals describing their proposed coverage and the contribution of such coverage to the encyclopedia. All proposals were carefully reviewed by the editor-in-chief in light of their suitability, researcher's records of similar work in the area of the proposed topics, and the best proposal for topics with multiple proposals. The goal was to assemble the best minds in the information science and technology field from all over the world to contribute entries to the encyclopedia. Upon the receipt of full entry submissions, each submission was forwarded to at least three expert external reviewers on a double-blind, peer review basis. Only submissions with strong and favorable reviews were chosen as entries for this encyclopedia. In many cases submissions were sent back for several revisions prior to final acceptance.

As a result, this five-volume encyclopedia includes more than 550 entries highlighting current concepts, issues and emerging technologies. All entries are written by knowledgeable, distinguished scholars from many prominent research institutions around the world. Five thousand technical and managerial terms enhance these entries. These terms each have a 5-50 word description that allows the readers of this extensive research source to learn the language and terminology of the field. In addition, this five-volume set offers a thorough reference section with over 11,500 sources of additional information for scholars, students, and researchers in the field of information science and technology to access.

To assist readers in navigating and identifying needed information, this five-volume encyclopedia has been organized by listing all entries in alphabetical order by title throughout the five volumes, and by including the title in the regular "Table of Contents" in the beginning of each volume. Furthermore, all entries are organized under their prospective discipline/area and placed in a separate table of contents at the beginning of each volume. In addition, a comprehensive glossary of all terms will direct readers to the short definition on the subject. A keyword index is also available, listing important keywords included throughout all entries.

To keep up with emerging technologies and research discoveries, regular entry additions will be made to the online version of the encyclopedia. I am pleased to add that complimentary online access to this encyclopedia for the life of the edition will be provided to any library with the purchase of the print copy. This complimentary online availability will allow students, faculty, researchers, and corporate managers to access the latest contents of this comprehensive and in-depth encyclopedia regardless of their location. This particular feature will prove to be an extremely valuable resource for distance learning educational programs worldwide.

The diverse and comprehensive coverage of multiple disciplines in the field of information science and technology in this five-volume authoritative encyclopedia will contribute to a better understanding all topics, research, and discoveries in this evolving, significant field of study. Furthermore, the contributions included in this encyclopedia will be instrumental in the expansion of the body of knowledge in this vast field. The coverage of this five-volume encyclopedia provides strength to this reference resource for both information science and technology researchers and also decision makers in obtaining a greater understanding of the concepts, issues, problems, trends, challenges and opportunities related to this field of study. It is my sincere hope that this publication and its vast amount of information and research will assist my research colleagues, all faculty, their students, and our organizational decision makers in enhancing their understanding of this discipline. Perhaps this publication will even inspire its readers to contribute to the current discoveries in this immense field, tapping possibilities to assist humankind in making the world a better place to live for all inhabitants of this universe.

REFERENCES

Khosrow-Pour, M. & Yaverbaum, G.J. (1990). *Information technology resources utilization and management: Issues and trends*. Hershey, PA: Idea Group Publishing.

Khosrow-Pour, M. (2000). *Managing Web-enabled technologies in organizations: A global perspective*. Hershey, PA: Idea Group Publishing.

Acknowledgements

Putting together a comprehensive publication of this magnitude requires a tremendous contribution and much assistance from everyone involved. The most important goal of editing this encyclopedia, as with most publications, was identifying and selecting quality contributions. This goal could not have been achieved without the valuable contributions of more than 1,000 external reviewers who have provided me with their expert, rigorous, unbiased blind review and assessment of the manuscripts assigned to them. I am extremely grateful for their contribution to this publication. I am also indebted to all the authors for their excellent contributions. I would also like to express my gratitude and deep appreciation to the members of the Editorial Advisory Board of this publication for their wisdom, guidance, and assistance in deciding on different issues of this publication.

In terms of editorial assistance, I would like to convey my deep appreciation and gratefulness to Idea Group Reference Senior Managing Editor, Ms. Jan Travers, for all her wisdom and encouragement, Ms. Mandy Appicello, Managing Editor, for her countless attention to detail and keeping this project on schedule, and Ms. Renee Davies, Assistant Acquisitions Editor, for her immeasurable assistance during the review process. I would like to express my many thanks to Ms. Sara Reed, Editorial Assistant, Ms. Michelle Potter, Editorial Assistant, Ms. Rachel Shepherd, Editorial Assistant, numerous copy editors, particularly Ms. Ingrid Widitz, and Ms. Lisa Tosheff, Graphic Artist, for their involvement and assistance with this publication. My warmest thanks go to my wife, Beth Peiffer, for her support, wisdom, encouragement, understanding, patience, and love. My heart also goes to my two young girls, Basha and Anar, for the joys that they have brought to my life. Finally, much gratitude goes to all those who have taught me immeasurable amounts during the past three decades.

Mehdi Khosrow-Pour, D.B.A.
Editor-In-Chief

About the Editor

Mehdi Khosrow-Pour, D.B.A. is currently the Executive Director of the Information Resources Management Association (IRMA) and Senior Academic Editor for Idea Group Reference. Previously, he served on the faculty of the Pennsylvania State University as an Associate Professor of Information Systems for twenty years. He has written or edited over twenty books in Information Technology Management, and he is also the editor of the Information Resources Management Journal, Journal of Electronic Commerce in Organizations, Journal of Cases on Information Technology, and International Journal of Cases on Electronic Commerce.

A Duplicate Chinese Document Image Retrieval System

A

Yung-Kuan Chan
National Chung Hsing University, Taiwan, R.O.C.

Yu-An Ho
National Chung Hsing University, Taiwan, R.O.C.

Hsien-Chu Wu
National Taichung Institute of Technology, Taiwan, R.O.C.

Yen-Ping Chu
National Chung Hsing University, Taiwan, R.O.C.

INTRODUCTION

An optical character recognition (OCR) system enables a user to feed an article directly into an electronic computer file and translate the optically scanned bitmaps of text characters into machine-readable codes; that is, ASCII, Chinese GB, as well as Big5 codes, and then edits it by using a word processor. OCR is hence being employed by libraries to digitize and preserve their holdings. Billions of letters are sorted every day by OCR machines, which can considerably speed up mail delivery.

The techniques of OCR can be divided into two approaches: template matching and structure analysis (Mori, Suen & Yamamoto, 1992). The template matching approach is to reduce the complexity of matching by projecting from two-dimensional information onto one; the structure analysis approach is to analyze the variation of shapes of characters. The template matching approach is only suitable for recognizing printed characters; however, the structure analysis approach can be applied to recognize handwritten characters.

Several OCR techniques have been proposed, based on statistical, matching, transform and shape features (Abdelazim & Hashish, 1989; Papamarkos, Spilioties & Zoumadakis, 1994). Recently, integrated OCR systems have been proposed, and they take advantage of specific character-driven hardware implementations (Pereira & Bourbakis, 1995). OCR generally involves four discrete processes (Khoubyari & Hull, 1996; Liu, Tang & Suen, 1997; Wang, Fan & Wu, 1997):

1. separate the text and the image blocks; then finds columns, paragraphs, text lines, words, and characters;
2. extract the features of characters, and compare their features with a set of rules that can distinguish each character/font from others;
3. correct the incorrect words by using spell checking tools; and
4. translate each symbol into a machine-readable code.

The duplicate document image retrieval (DDIR) system transforms document formatted data into document images, then stores these images and their corresponding features in a database for the purpose of data backup. The document images are called duplicate document images. When retrieving a duplicate document image from the database, users input the first several text lines of the original document into the system to create a query document image. Then the system figures out the features of the image, and transmits to the users the duplicate document image whose image features are similar to those of the query document image (Nagy & Xu, 1997).

Some approaches have been proposed for the DDIR system. Doermann, Li, and Kia (1997) classified and encoded character types according to the condition that four base lines cross each text line, and uses the codes as the feature of the document image. Caprari (2000) extracted a small region from one document, assigned this region to the template (signature generation), and then scanned this template over a search area in another document. If the template also appears in the second document (signature matching), the two documents are classified as duplicates. Angelina, Yasser, and Essam (2000) transformed a scanned form into a frameset composed of a number of cells. The maximal grid encompassing all of the horizontal and vertical lines in the form is generated; meanwhile, the number of cells in the frameset, where each cell was created by the maximal grid, was calculated. Additionally, an algorithm for similarity matching of document framesets based on their grid representations is proposed too. Peng, Long, Chi, and Siu (2001) used the size of each component block containing a paragraph text image in a duplicate document image and its relative location as the features of the duplicate document image.

The approaches mentioned previously are only suitable for stating the characteristics of an English document image. The characteristics of Chinese characters are quite different from those of English ones, and the strokes and shapes of Chinese characters are much more complicated than those of English characters. Chan, Chen, and Ho (2003) provided a line segment feature to represent a character image block and presented a duplicate Chinese document image retrieval (DCDIR) system based on this feature. The purpose of this short article is to give a brief overview of the duplicate Chinese DDIR systems.

BACKGROUND

Traditional information retrieval methods use keywords for textual databases. However, it is difficult to describe an image using exact information, and defining manually keywords is tedious or even impossible for a large image database. Moreover, some non-text components cannot be represented in a converted form without sufficient accuracy. One solution is to convert a document into digital images; meanwhile, some methods are applied to extract the features of the images. Based on the feature, some document images with database satisfying query requirements are returned.

A duplicate document image retrieval (DDIR) system has to own the following properties (Doermann, Li, & Kia, 1997):

- **Robust:** The features should be reliably extracted even when the document becomes degraded.
- **Unique:** The extracted features can distinguish each document image from others.
- **Compact:** The storage capacity required to hold the features should be as small as possible.
- **Fast:** The system needs a quick response with an answer to the query.
- **Scalable:** As more documents are processed, the size of the database could grow to tens of millions.
- **Accurate:** The system should accurately response with an answer, which satisfies the query requirement.

Unfortunately, many DDIR systems are vulnerable to poor qualities of document images, such as the scale, translation, rotation, and noise variants. Because of different resolution setup of a scanner, the same image may be scanned to become two images with different sizes. We call this phenomenon the scale variant. When an image is added with a great amount of noises, it may be regarded as a different image from the original one. It is named a noise variant image of the original one. In a particular document, images with rotation and translation variants may be generated owing to placing the document on different orientation angles or on different positions on a scanner. The variants mentioned previously will cause many troubles in feature extracting and image matching stages. They should be removed in advance.

A CHINESE DDIR SYSTEM

Many techniques about the DDIR system have been proposed (Caprari, 2000; Doermann, Li, & Kia, 1997; Peng, Chi, Siu, & Long, 2000; Peng, Long, Chi, & Siu, 2001). Since an English document mostly consists of approximately **70** commonly-used characters which contain **52** uppercase as well as lowercase English letters and punctuation marks, the classification and encoding procedure based on the feature of these characters' font types are possible. However, these techniques are only suitable for duplicate English document images, but not for duplicate Chinese document image retrieval (DCDIR) because the number of different Chinese characters is about **45,000**. What is more, the shapes of Chinese characters are complex, and many different characters have similar shapes to each other. Hence, there are several major problems with Chinese character recognition, that is, Chinese characters are distinct and ideographic, the size of a character is large, and there exist many structurally similar characters (Amin & Singh, 1996; Chan, Chen, & Ho, 2003).

It is necessary to develop a feature offering an excellent identification capability to classify Chinese characters by only using a little extra memory space. To reduce the extra memory space, it is feasible to segment a duplicate document image into blocks, each of which contains a set of adjacent characters, and then to extract the features from the blocks. Since the number of the blocks in a duplicate document image is much smaller than that of the characters in an identical duplicate document image, the feature dimensions are reduced greatly; however, its identification capability is lessened.

I. DCDIR System

The proposed duplicate document image retrieval system approximately includes three parts — image preprocessing, database creation, and document retrieval. This section will introduce these three parts in details.

A. Image Preprocessing

When scanning a document to generate a duplicate document binary image, the position of the document on the scanner may be misplaced so that the duplicate document image may become inclined. Figure 1(a) shows an original image and Figure 1(b) is its duplicate document image that appears inclined. The inclined condition of a document image may lead to inconvenience to users and cause the errors in extracting its image features. Peng et al. (2000) used a correlation-based method to detect the inclination of an image, and then applied an interpolation technique to turn the image back according to the detected inclination. The DCDIR system will use this technique to turn the inclined document image back. Figure 1(c) is the duplicate document image after adjusting the inclination.

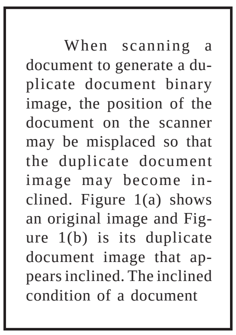
As in Figure 1(c), after turning back the duplicate document image, the frame of the duplicate document image will become inclined. It is necessary to cut off the border blank of the document image. While removing the border blank, the system starts scanning the duplicate document image from the left-top pixel. Then, in the order from left to right and top to bottom, each pixel is scanned until one certain black pixel ***P*** is found. Finally, all pixels locating on the lines, which are prior to the line containing ***P*** are removed to cut off the top border blank of the document image. By using the same method, the bottom, left, and right border blanks of the document image are removed as well. Figure 1(d) demonstrates the final duplicate document image after cutting off the border blanks of the document image as illustrated in Figure 1(c).

B. Database Creation

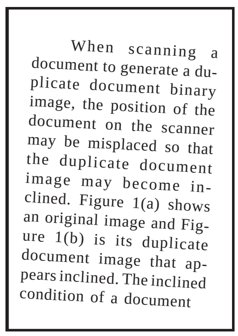
After that, the DCDIR system extracts the character image features from the duplicate document image ***I*** in which its border blanks have been cut off and the system stores the features in the database. Before extracting the character image features of ***I***, the system first performs the text line segmentation on ***I*** to make every line image

Figure 1. Normalization of an inclined document image

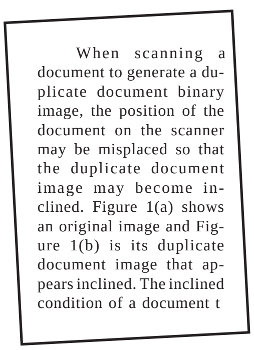
(a) Original document



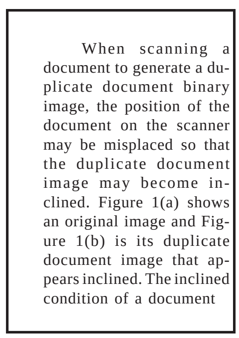
(b) Duplicate document image after scanning



(c) Duplicate document image after inclination adjusting



(d) Duplicate document image after cutting off the border blanks



block contain only the complete image of one certain line text in the original document. Then, the system segments out all character image blocks from each previously segmented line image block so that every character image block contains only one Chinese character. Finally, the feature of each character image block is then extracted.

Concerning the steps of segmenting line image blocks from a duplicate document image, first, all of the black pixels in the duplicate document image are projected in horizontal direction onto a projection vertical axis. The length of a black section on the projection vertical axis is just the height of the corresponding text line containing those character images whose black pixels are projected onto the black section.

Next, all of the black pixels in every line image block is projected in vertical direction onto a certain projection horizontal axis. In this case, the distribution borders of the black pixels and the white pixels on the projection horizontal axis are the locations of the left and the right boundaries of the character image blocks. On the projection horizontal axis, the length of a black section is just the width of the corresponding character image block whose black pixels are projected onto the black section.

The sizes of most Chinese characters are close. When the height of a certain character image block ***CB*** is smaller than three-fourths of the average height of all character image blocks in the document image, and the width of ***CB*** is also smaller than three-fourths of the average width of all character image blocks, the system will then regard the character in ***CB*** as a noise, and then remove it.

After that, three horizontal scanning lines are drawn on each character image block. These three horizontal scanning lines are respectively located at $1/4\times H$, $2/4\times H$ and $3/4\times H$ character heights in the block. Here H represents the height of the character image block. According to the ratio of the total number of the black pixels to that of the white pixels, which the scanning line goes through, an encoding process is executed to reduce the memory space required to store the feature of the character image block. The way of encoding is shown as follows:

$$X_i = \begin{cases} 0, \text{ if } D_{i,b} \times m > D_{i,w} \\ 1, \text{ if } D_{i,b} \times m \le D_{i,w} \end{cases}, \text{ for } i = 0, 1, \text{ and } 2.$$

In this equation, $D_{i,w}$ and $D_{i,b}$ are respectively the total numbers of the white pixels and the black pixels that the i-th scanning line passes through, and m is the weight (a given constant value) for the ratio from the total numbers of the black pixels and the white pixels on the scanning line. Thus, each character image block can be represented by a three-bit ($X_0X_1X_2$) code; we name the code the feature code of the character image block. There are ***8*** different binary codes ***000***, ***001***, ***010***, ***011***, ***100***, ***101***, ***110***, and ***111*** corresponding to decimal feature codes ***0, 1, …,*** and ***7*** respectively.

Because the resolution setup of a scanner may be different, the same original document may be scanned to become duplicate document images of different sizes. This proposed feature adopts the ratio from the number of black pixels and white pixels, so the feature encoding will not be affected due to the scale variant of images. Moreover, the desired duplicate document image and the query document image are both from the same original document. Therefore, the problem that the font type and style of the characters in the query document are different from those in the desired duplicate document image will not occur in this system.

C. Document Retrieval

Let $Q = q_1q_2\ldots q_l$ be the feature code of query document image I_q, and the length of the feature code be l. Next, the system extracts the first feature codes with the length l from every duplicate document image I_d in the database. Let the extracted feature codes be $D = d_1d_2\ldots d_l$. Then, the system compares the corresponding bit pair q_i and d_i between Q and D from left to right, respectively. When $q_i = d_i$, the system adds ***1*** to the value of S. The final value of S is the similarity between I_q and I_d. Finally, the duplicate document image with the largest similarity value is found out.

II. Experiments

Experiment 1 is to explore the constant weight m. As for different values of m, the character image blocks of ***5401*** in commonly used Chinese characters among Big5 codes, are categorized into eight groups each of which corresponds to one feature code. Table 1 shows the number of members in each group for $m = 2, 3$, and 4, where σ_c^2 is the variance of the number of members of the eight groups. The experimental results shows that when $m = 3$, σ_c^2 is minimal. This means, when $m = 3$, all Chinese character image blocks are most uniformly mapped to various kinds of feature codes. The next experiment will set $m = 3$.

Experiment 2 is to investigate the performance of the DCDIR system. This experiment scans each page of the book "朝花夕拾、吶喊" with ***336*** sheets to become images by a scanner. This experiment rescans ***101*** sheets of the book to generate the query document images. Here, the first L text lines of the ***101*** document sheets are respectively used as the contents of the query document images. Table 2 shows the experimental results. The average searching time is approximately ***8*** seconds for each query.

Table 1. Results of the first experiment

Feature code \ m	2	3	4
000	372	1253	2266
001	596	773	763
010	262	387	402
011	813	525	312
100	337	564	628
101	817	591	324
110	390	374	262
111	1798	918	428
σ_c^2	220549	74523	387785

FUTURE TRENDS

After a paper document is used over a long period, the document may be stained or worn out, so that its contents may be indistinct. How to develop an effective image feature insensitive to the rotation, scale, translation, and noise variations is an important task in the future. Many documents are printed or handwritten texts, are multilingual, or are composed of basic composition style and font for text. A future document image indexing method should own the robustness of the above variants.

Moreover, the following topics have been explored, and will continue to be researched. The first is to find or to match the instances of a document image with known content in a database. The techniques can be applied to maintaining database integrity by eliminating duplicates and retrieval itself. The second is to index image captions and to establish a relationship between the content and the images they describe. Then the caption can be a valuable tool of their duplicate document images. Ideally, a duplicate detection algorithm can find both exact duplicates which have just the same content, and partial duplicates, which have a large percentage of their text in common. Locating exact duplicates could reduce the storage required for a large database. Finding partial duplicates will allow users to easily find other versions of a given document.

REFERENCES

Abdelazim, H.Y., & Hashish, M.A. (1989). Automatic reading of bilingual typewritten text. *Proceeding of VLSI and Microelectronic Applications in Intelligent Peripherals and their Application Network*, 2.140-2.144.

Amin, A., & Singh, S. (1996). Machine recognition of hand-printed Chinese characters. *Intelligent Data Analysis, 1*, 101-118.

Angelina, T., Yasser E.S., & Essam A.E.K. (2000). Document image matching using a maximal grid approach. *Proceedings of SPIE on Document Recognition and Retrieval IX*, 4670, 121-128.

Caprari, R.S. (2000). Duplicate document detection by template matching. *Image and Vision Computing*, *18*(8), 633-643.

Chan, Y.K., Chen, T.S., & Ho, Y.A. (2003). A duplicate Chinese document image retrieval system based on line segment feature in character image block. In S. Deb (Ed.), *Multimedia systems and content-based image retrieval, (pp.14*-23). Hershey, PA: Idea Group Publishing.

Doermann, D., Li, H., & Kia, O. (1997). The detection of duplicates in document image databases. *Image and Vision Computing*, *16*(12-13), 907-920.

Khoubyari, S., & Hull, J.J. (1996). Font and function word identification in document recognition, *Computer Vision and Image Understanding*, *63*(1), 66-74.

Liu, J., Tang, Y.Y., & Suen, C.Y. (1997). Chinese document layout analysis based on adaptive split-and-merge and qualitative spatial reasoning. *Pattern Recognition*, *30*(8), 1265-1278.

Mori, S., Suen, C., & Yamamoto, K. (1992). Historical review of OCR research and development. *Proceedings of the IEEE*, *80*(7), 1029-1058.

Nagy, G., & Xu, Y. (1997). Bayesian subsequence matching and segmentation. *Pattern Recognition Letters*, *18*(11-13), 1117-1124.

Papamarkos, N., Spiliotıes, I., & Zoumadakis, A. (1994). Character recognition by signature approximation. *Inter-*

Table 2. Results of Experiment 2

experimental results \ Text lines	*L*=3	*L*=5	*L*=6	*L*=7	*L*=10
Correctly finding out desired images	99	100	100	101	101
Accuracy rate (%)	98.0	99.0	99.0	100	100

national Journal of Pattern Recognition and Artificial Intelligence, *8*(5), 1171-1187.

Peng, H., Chi, Z., Siu, W.C., & Long, F. (2000). PageX: An integrated document processing software for digital libraries. *Proceedings of the International Workshop on Multimedia Data Storage, Retrieval, Integration, and Applications, Hong Kong*, (pp.203-207).

Peng, H., Long, F., Chi, Z., & Siu, W.C. (2001). Document image template matching based on component block list. *Pattern Recognition Letters*, *22*(9), 1033-1042.

Pereira, N., & Bourbakis, N. (1995). Design of a character driven text reading system. *Proceedings of "SPIE", San Jose, California*, (p.6-9).

Wang, A.B., Fan, K.C., & Wu, W.H. (1997). Recursive hierarchical radical extraction for handwritten Chinese characters. *Pattern Recognition*, *30*(7), 1213-1227.

KEY TERMS

Character Segmentation: The technique that partitions images of lines or words into individual characters.

Content-Based Image Retrieval (CBIR): The technique of image retrieval based on the features automatically extracted from the images themselves.

Duplicate Document Detection: The technique to find the exact duplicates, which have exactly the same content, or partial duplicates which have a large percentage of their text in common.

Duplicate Document Image Retrieval (DDIR): A system for finding the image-formatted duplicate of documents from a database.

MFU: Most frequently used characters in a character set with enormous members.

Optical Character Recognition (OCR): The technique of automatically translating the content of an image formatted document into text-formatted materials.

Skew Correction: The technology detects and compensates for the inclination angle of a document image.

Template Matching: The approach involves designing template masks, which are capable of detecting incidences of the relevant feature at different orientations.

A Language/Action Based Approach to Information Modelling

Paul Johannesson
Stockholm University/Royal Institute of Technology, Sweden

INTRODUCTION

There are several different views of the role of information systems. Two of the most important are the data view and the communicative view. According to the data view, the primary purpose of an information system is to provide a model of a domain, thereby enabling people to obtain information about reality by studying the model. In this respect, an information system works as a repository of data that reflects the structure and behaviour of an enterprise, and the system provides data that can be used for decisions about the enterprise. In contrast, the communicative view states that the major role of an information system is to support communication within and between organisations by structuring and coordinating the actions performed by organisational agents. The system is seen as a medium through which people can perform social actions, such as stating facts, making promises, and giving orders.

The data and communicative views of information systems are mirrored by two different views of organisations: the functional view and the constructional view (Dietz, 2003a). The functional view focuses on the functions of an organisation with respect to its environment, in particular, the resources that the organisation consumes and produces. A model of an organisation from a functional perspective is a black-box model, as it shows the interactions with the environment but not the internal mechanisms. The constructional view, on the other hand, focuses on how behaviour and function are brought about by the operations and structure of an organisation. A model of an organisation from a constructional perspective is a white-box model as it shows the inner workings of the organisation.

In information systems design, the starting point has often been based on the data view and the functional view, though frequently augmented by concepts like reasoning and monitoring. However, these views easily lead to a computer- and technology-biased management of the communication taking place in an organisation, and they benefit from being complemented by the communicative and constructional views. A promising theoretical foundation for these views is the language/action approach, which is based on theories from linguistics and the philosophy of language. In the language/action approach, business actions are modelled on the notions of speech acts and discourses, which provide a basis for distinguishing between different communication phases, such as preparation, negotiation, and acceptance. The purpose of this chapter is to outline how the language/action approach can be used as a basis for the information modelling of communicative aspects in organisations.

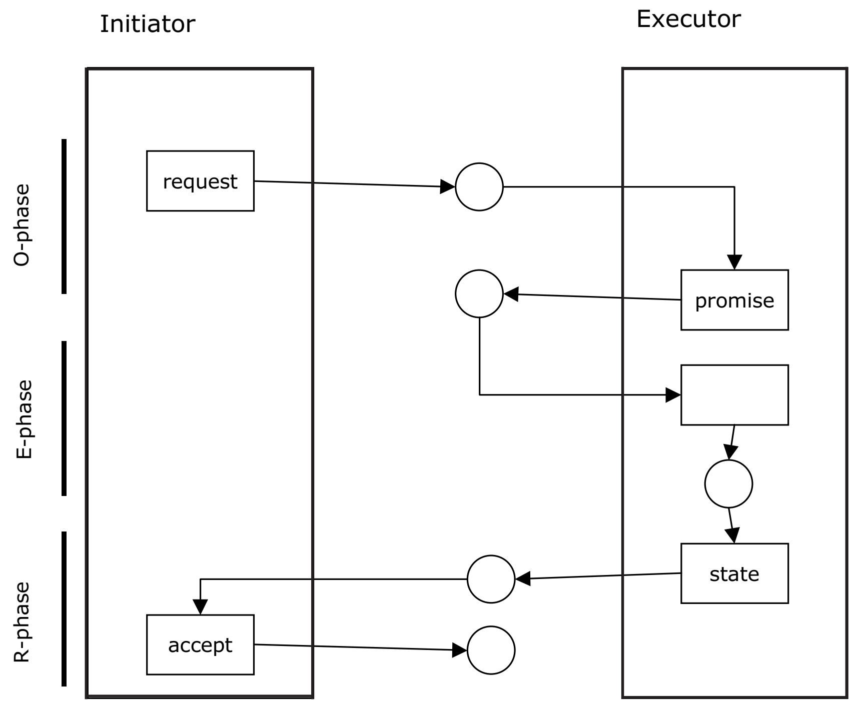
BACKGROUND

One important foundation of the language/action approach is speech act theory (Austin, 1962; Searle, 1969). The basic insight of speech act theory is that language can serve purposes other than that of representing the states of affairs of the world. Certain statements are equivalent to actions. For example, when someone says "I apologise," "I promise…," or "I name this ship…," the utterance changes the psychological or social reality. Statements such as these are called *speech acts*, and they enable people to use language as a means for acting as well as coordinating action.

In Searle (1969), a classification of speech acts is proposed based upon the way in which a speech act affects the social world. Searle identified five classes: assertive, commissive, directive, declarative, and expressive. An *assertive* is a speech act, the purpose of which is to convey information from a speaker to a hearer, e.g., "the cat is on the mat." A *commissive* is a speech act, the purpose of which is to commit the speaker to carry out some action or bring about some state of affairs, e.g., "I promise to bring it back." A *directive* is a speech act, where the speaker requests that the hearer carry out some action or bring about some state of affairs, e.g., "Please bring me the salt." A *declarative* is a speech act, where the speaker brings about some state of affairs by the mere performance of the speech act, e.g., "I hereby baptise you Samuel." An *expressive* is a speech act, the purpose of which is to express the speaker's attitude, e.g., "I like coffee."

In order to understand the role of speech acts, it is helpful to view human communication as taking place in three different worlds:

Figure 1. OER pattern



- The physical world—In this world, people carry out message actions. They utter sounds, wave their hands, send electronic messages, etc. Furthermore, other instrumental acts may take place in the physical world, such as repairing equipment.
- The communicative world—In this world, people express their intentions and feelings. They tell other people what they know and try to influence the behaviour of others through communication, i.e., they perform speech acts. These speech acts are brought about by means of message actions in the physical world. Note that a message action does not need to be verbal, as it can also be expressed by body language.
- The social world—In this world, people change the social and institutional relationships among them. For example, people become married or acquire possession of property. People perform such social actions by performing speech acts in the communicative world.

LANGUAGE/ACTION FOR BUSINESS PROCESS MANAGEMENT

The most important applications of the language/action approach have been made in the area of business process management (Lehtinen, 1986; Weigand, 2003). A language/action perspective provides a clear and well-founded basis for identifying and modelling recurring patterns in business processes. One such pattern is the order–execution–result (OER) pattern (Dietz, 2003b), which models a basic form of interaction that occurs in every business process (Figure 1). The interaction takes place between two parties—the initiator and the executor—and governs how they coordinate their actions. The interaction starts in the order phase by the initiator making a directive speech act, namely, a request to carry out some action (shown by a rectangle), which results in a state (shown by a circle in Figure 1), where there is an order from the initiator to the executor. The executor accepts the order by performing a commissive speech act (a rectangle labelled "promise" in Figure 1), resulting in a state where there is a commitment for the executor to carry out the action. This concludes the order phase, which is followed by the execution phase, where the executor actually performs the action (shown by an unlabelled rectangle in Figure 1) he or she is committed to. This action may be an instrumental action, e.g., delivering a package, or a declarative speech act, e.g., grading an exam. However, the execution phase is always concluded by a declarative speech act, where the executor states that he or she has carried out the committed action. The final phase is the result phase, where the initiator performs a declarative speech act and acknowledges that the executor has carried out the requested action in a satisfactory way.

The pattern introduced above shows the success path, where the initiator and executor agree to each other's actions. However, the pattern needs to be extended to handle cases where the parties do not agree, e.g., when the initiator does not accept the result of the execution phase. Larger business processes are typically built up by combinations of the OER pattern, e.g., when three or more parties are involved in a process or when reciprocal commitments are created in the process. The latter case occurs, for example, in e-commerce, as discussed in Lind (1997, 2003). Several variants of the OER pattern have been proposed in the literature, e.g., the conversation for basic action in Winograd (1986).

FUTURE TRENDS

The language action approach influenced another recent trend in the information systems community, agent-oriented information systems (AOIS Workshops, 2004). The language/action approach can be seen as one of the most active traditions within agent-oriented information systems. Agent concepts offer high-level abstractions addressing issues such as knowledge representation, communication, coordination, and cooperation among heterogeneous and autonomous parties. One precursor of agent-oriented information systems was the REA framework (McCarthy, 1982), which was designed for representing and reasoning about economic exchanges. The basic REA pattern models an exchange by three components: the *events* of the exchange, the *resources* that are exchanged, and the participating *agents*. In order to obtain a resource, an agent must give up some other resource. Therefore, an economic exchange always consists of two corresponding events, e.g., a purchase and a payment. Extended versions of the REA pattern also include commitments that are established, where a commitment is an obligation to perform a resource transfer some time in the future. These commitments are created through applications of the OER patterns.

One of the most comprehensive approaches to agent-oriented information systems is agent-object-relationship (AOR; Wagner, 2003), which is based on the notions of agents, events, actions, claims, and commitments. These notions form the basis for a general meta-model that can be used as a foundation for any information system. Another related work is that by Weigand et al. (1998), which proposes a layered architecture for information systems. At the lowest layer, we find the elementary acts—speech acts as well as instrumental acts. The next layer contains the business transactions. A business transaction is the smallest sequence of speech acts that results in a new deontic state, i.e., a state in which an obligation to carry out some action has been created, or a state where an authorisation for certain actions has been established, or a state where some action has been accomplished so that a previous obligation has become fulfilled. A single speech act is, in most cases, not sufficient to achieve a deontic effect. In general, at least two messages are required. For example, a customer requests that a supplier deliver a product, and the supplier promises to do so. These speech acts will result in two obligations: one for the supplier to deliver and one for the customer to pay upon delivery. The top layer is the workflow level, where a workflow is a set of linked business transactions that realise a business objective within an organisational structure.

CONCLUSION

The language/action approach to information modelling and systems provides a solid theoretical framework for analysing and designing communicative action in organisations. One foundation of the language/action approach is speech act theory, which investigates how language can be used to perform actions. The main application of the language/action approach has been in the area of business process design and management, where the approach can assist in creating complete and well-functioning processes. A recent trend is to apply the language/action approach for agent-oriented information systems, including applications to e-commerce and e-business (Bergholtz, 2003).

REFERENCES

AOIS Workshops. (2004). Retrieved from *www.aois.org*

Austin, J. L. (1962). How to do things with words. Oxford.

Bergholtz, M., Jayaweera, P., Johannesson, P., & Wohed, P. (2003). Reconciling physical, communicative and social/institutional domains in agent oriented information systems—A unified framework. In *Proceedings of the International Workshop on Agent Oriented Information Systems, ER 2003*, Chicago. Heidelberg: Springer.

Dietz, J. (2003a). Generic recurrent patterns in business processes. Business Process Management 2003, Eindhoven, LNCS 2678. Heidelberg: Springer.

Dietz, J. (2003b). The atoms, molecules and fibers of organizations. *Data and Knowledge Engineering, 47*(3).

Lehtinen, E., & Lyytinen, K. (1986). Action based model of information system. *Information System, 11*(4), 299–317.

Lind, M., & Goldkuhl, G. (1997). Reconstruction of different business processes—A theory and method driven analysis. *Conference on Language/Action Perspective '97*, Veldhoven.

Lind, M., & Goldkuhl, G. (2003). The constituents of business interaction—Generic layered patterns. *Data and Knowledge Engineering, 47*(3).

McCarthy, W. (1982). The REA accounting model: A generalized framework for accounting systems in a shared data environment. *The Accounting Review*, (July), 554–78.

Searle, J. R. (1969). *Speech acts—An essay in the philosophy of language*. London; New York: Cambridge University Press.

Wagner, G. (2003). The agent-object-relationship metamodel: Towards a unified view of state and behavior. *Information Systems, 28*(5).

Weigand, H., & de Moor, A. (2003). Workflow analysis with communication norms. *Data and Knowledge Engineering, 47*(3).

Weigand, H., & v d Heuvel, W. (1998). Meta-patterns for electronic commerce transactions based on FLBC. In *Proceedings of the Hawaii International Conference on System Sciences (HICSS'98)*. Los Alamitos, CA: IEEE Press.

Winograd, T., & Flores, F. (1986). Understanding computers and cognition: A new foundation for design. Norwood, NJ: Ablex.

KEY TERMS

Business Process: This is comprised of business transactions that realise a business objective.

Business Transaction: This consists of speech acts that result in a deontic effect.

Deontic Effect: This is the establishment of an obligation or the fulfilment of an obligation.

Information System: This is a system for supporting communication within and between organisations.

Instrumental Act: This is an act performed in order to change the physical world.

OER Pattern: This is a basic pattern for business interaction based on order–execution–result phases.

Speech Act: This is a linguistic act performed with the intention of changing the social relationships between agents, in particular, creating deontic effects.

A Primer on E-Government

Shannon Howle Schelin
University of North Carolina at Chapel Hill, USA

OVERVIEW

E-government (electronic government) has become a mainstay in local, state, and federal government. The era of e-commerce and e-business began with the widespread adoption of the Internet in the mid-1990s and today many citizens expect the same responsiveness and access to government services as found in the private sector. According to the 2002 International City/County Managers Association e-government survey, over 73 percent of municipalities with populations larger than 2,500 have Web sites. The 2002 Pew Internet and American Life Project indicates that 58 percent (68 million people) of American Internet users have accessed at least one governmental Web site (Larson and Rainie, 2002).

Although there is widespread interest in the topic, e-government lacks a consistent, widely accepted definition. It is often related to revolutionizing the business of government through the use of information technology (IT), particularly Web-based technologies, which improve internal and external processes, efficiencies, and service delivery. The American Society for Public Administration (ASPA) and United Nations Division for Public Economics and Public Administration (UNDPEPA) have defined e-government as "utilizing the Internet and the World Wide Web for delivering government information and services to citizens" (UN & ASPA, 2001, p. 1). Based on this working definition of e-government, this article seeks to examine the historical premises and typologies of e-government.

HISTORICAL PREMISES

E-government has evolved from the information technology revolution. Information technology enables new methods of production, increases the flow and accuracy of information, and even may replace traditional standard operating procedures (Landsbergen & Wolken, 2001). Information technology in government has long been acknowledged as a method for improving efficiency and communication (Kraemer & King, 1977; Norris & Kraemer, 1996). Now, IT developments such as electronic mail (e-mail) have changed interpersonal communications to eliminate the constraints of geography, space, and time with profound organizational consequences (Rahm, 1999). The ability to buy and sell goods and services via the Internet has led to new private sector industries, constituting a new business model that the public sector now seeks to emulate. In addition, IT has promoted globalization, which also changes the environment within which public agencies function (Kettl, 2001).

The main concerns of e-government focus not only on the electronic dissemination of public information arising from traditional agency functions, but even more on reinventing agency processes to fully exploit the potential of information technology. As Fountain (2001) has noted, the reinvention process requires overcoming the rigidities and limits of traditional bureaucratic forms. Specific objectives may include the centralization of public data and the improvement of internal processes and communications (Alexander & Grubbs, 1998).

One of the first comprehensive visions of e-government is found in the 1993 National Performance Review report, *Creating a Government that Works Better and Costs Less: Reengineering Through Information Technology* (Gore, 1993; Kim & Wolff, 1994). This report laid the groundwork for new customer- and client-oriented ways for agencies to engage citizens via technology, involving both improved agency processes and improved methods of delivery. Most reinventing government literature has cited the need to rely on information technology to improve citizen-centric government services (Gore, 1993; Osborne & Gaebler, 1992). Although the reinventing government and e-government movements are related, the prospects are that the focus of public administration on e-government will endure for the foreseeable future, outlasting the reinventing government movement.

The 1995 amendment of the 1980 Paperwork Reduction Act (PRA) was another important milestone in the history of e-government. This amendment offered specific policies on issues associated with managing electronic information, including the establishment of standards, mandates for cross-agency information technology initiatives, and technology investment guidelines (Relyea, 2001). By outlining guidelines for information technology, the amended PRA solidified the government's commitment to improving citizen services via new channels based on technology.

The 1996 Electronic Freedom of Information Act (EFOIA) amendments added a new level of clarity to the issue of electronic records. This amendment extended the right of citizens to access executive agency records to include access to electronic formats and online information (Relyea, 2001). EFOIA also extended the time limits from 10 to 20 days to prohibit the use of agency backlogs as an excuse for noncompliance with information requests (Hammitt, 1999).

The 1996 Clinger-Cohen Act further heightened the role of information technology in government. It established a chief information officer (CIO) in every agency, making agencies responsible for developing an IT plan. Later, as the e-government movement gained momentum, the existence of the CIO strategic planning structure was an important element facilitating e-government implementation at the federal level.

Also in 1996, Congress passed the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA). This act, also known as the Welfare Reform Act, represented one of the first national pushes to incorporate the rhetoric of e-government with the routine services of agencies, specifically the administration of Temporary Aid to Needy Families (TANF). The act required interagency, interstate, and intergovernmental coordination of information technology systems to ensure that no individual exceeded the allotted five-year lifetime cap on assistance (Scavo & Shi, 2000).

In July 1996, President Clinton issued Executive Order 13011, which sought to improve management of information technology at the federal level. It also provided broad support for coordinated approaches to technology application in the executive office (Relyea, 2001). Although this executive order mandated implementation of and adherence to the PRA Act and the Clinger-Cohen Act, it also focused on the alignment of technology goals with strategic organizational goals. The support for interagency coordination of technology is codified in Executive Order 13011. Mandated goal-alignment and technology-investment reviews are included in the directive as a method for reducing the failure rates and cost overruns associated with federal technology initiatives.

More recently, in 2001 the E-Government Act was offered for consideration in the U.S. Senate. This act, approved by the Senate in June 2002, mandated the establishment of an e-government administrator in the Office of Management and Budget, and also provided for considerable financial assistance to spur interagency e-government initiatives. Each of these legislative actions has strengthened the federal government's commitment to e-government.

One of the most significant information technology developments at the federal level occurred after the tragedies of September 11. The attacks against America forced government officials to reexamine their information technology policies, infrastructure, and systems. The newly established Office of Homeland Security and its associated directives comprise the largest centralization and consolidation effort involving governmental databases in the history of the US. A recent General Accounting Office report (2002) highlights this effort and its challenges by examining the types of necessary data, the amount of data, and the transmission format of the data across vertical and horizontal governmental lines. The report also notes the challenges associated with traditional agency "stovepipes" and the need to move beyond this approach toward a centralized enterprise initiative (p. 8). The lack of connectivity and interoperability between databases and agency technologies is another crucial challenge that must be overcome, it is argued, in order to create a comprehensive infrastructure to deal with issues of terrorism.

Another example of the critical need to centralize and consolidate government information in order to mitigate future terrorist attacks is found in the Chambliss-Harman Homeland Security Information Sharing Act (HR 4598), passed by the U.S. House of Representatives in June 2002. This act mandates the dissemination of critical intelligence information from federal agencies, such as the CIA and FBI, to state and local governments (GCN, 2002). The goal of this act is to further reduce the vertical stovepipes that exist between federal, state, and local governments with respect to information access, and to encourage data sharing across all branches and levels of government in order to foster coordination and collaboration.

Although the effects of September 11 have impacted the use of information technology in the public sector in a variety of ways, there is little doubt that citizen demand for electronic information and services is likely to continue the trend of e-government adoption and expansion. According to the 2002 Pew Internet and American Life Project, Americans continue to use the Internet to access government information, research policy issues, contact elected officials, and participate in e-democracy in increasing numbers (Larson & Rainie, 2002). The number of Americans who have accessed government information online was 68 million in January 2002, compared with 40 million in March 2000. This marked increase further supports the idea that citizen demand for and use of e-government will continue to expand in the future.

E-GOVERNMENT TYPOLOGIES

Although several typologies have been developed to explain the progression of e-government (Layne & Lee, 2001; Moon, 2002), the UN and ASPA definition of the stages of e-government maintains continuity with the working definition set forth at the outset of this essay. It is also important to note that the stages to be discussed do not represent a true linear progression, nor are they specific block steps. Rather, the stages are a continuum in which governments can be within the same stage with very different functionalities and service offerings.

According to the UN and ASPA (2001), there are five main stages of e-government. Figure 1 offers a side-by-side comparison of the major components of the five stages of e-government. Each of the stages is discussed in further detail below.

The lack of an organizational Web site is not defined by a stage, but may be considered Stage 0. Stage 1 is the emerging Web presence, which involves static information presented as a type of online brochure (UN & ASPA, 2001). The main goal of the emerging Web stage is to provide an online mechanism for communicating key general information about the government to interested citizens and entities. The Web site lacks information about services and is not organized in a citizen-focused manner. Typically, the government has used a “go it alone” approach, which visually represents the stovepipes that exist within agencies—there is little coordination across agencies and levels of government in Stage-1 Web sites.

In Stage 2, enhanced Web presence, the role of the Web site becomes associated with information on services, although it is still organized by departments rather than by user groups. Enhanced-Web-presence sites typically have e-mail as a means of two-way communication. However, rarely are there available forms for download. Stage 2 offers limited communication and greater information about the services of the government, but it does not meet the citizen-centric approach that has been advocated for e-government (UN & ASPA, 2001).

Stage 3, interactive Web presence, begins to move into the citizen-centric realm of e-government (UN & ASPA, 2001). Typically, the information is portrayed by intuitive groupings that cross agency lines. For example, the Web site might use a portal as the single point of entry into various departments and service areas. The portal would offer major groupings like business, new-resident, seniors, children, or other standard groups. Then, the end user would select the grouping that applies and be launched into a new section of the portal where the most common services requested for the group are located. The services would not be listed by departmental areas, but rather by functional areas. Stage-3 sites have downloadable forms with online submissions, e-mail contacts for various governmental employees, and links to other governmental Web sites.

Stage 4, transactional Web presence, offers the ability to conduct secure online transactions (UN & ASPA, 2001). This stage is also organized by user needs and contains dynamic information. The Web site may offer a variety of transactions, including paying for services, paying bills, and paying taxes. Transactional Web presence includes online submission of forms, many downloads, e-mail contacts, and several links to other government sites. The use of digital signatures also falls under Stage 4.

The final stage, Stage 5, involves seamless government. Although this stage represents an ideal, there is no real example of its application. Stage 5 involves a cross-agency, intergovernmental approach that only displays one front, regardless of service area (UN & ASPA, 2001). For example, a seamless Web site would offer local, state, and federal government services via the state portal without the end user recognizing what level of government provides the service. A Stage-5 site would offer vertical and horizontal integration and would require true organizational transformation with respect to adminis-

Figure 1. E-government typology

Stage	Orientation	Services	Technology	Citizens
Stage 1: Emerging Web Presence	Administrative	Few, if any	Only Web	Going it alone
Stage 2: Enhanced Web Presence	Administrative, Information	Few forms, no transactions	Web, e-mail	Links to local agencies
Stage 3: Interactive Web Presence	Information, Users, Administrative	Number of forms, online submissions	Web, e-mail, portal	Some links to state and federal sites
Stage 4: Transactional Web Presence	Information, Users	Many forms and transactions	Web, e-mail, digital signatures, PKI, portals, SSL	Some links to state and federal sites
Stage 5: Seamless Web Presence	Users	Mirror all services provided in person, by mail, and by telephone	Web, e-mail, PKI, digital signatures, portal, SSL, other available technologies	Crosses departments and layers of government

trative boundaries (UN & ASPA). The closest example to seamless government is the FirstGov Web site created by the federal government of the United States. However, even this site lacks the coordination and cooperation required to make seamless government a reality.

With a working knowledge of the typology associated with e-government, it is easy to assess the current status of the concept. Much of the literature indicates that Stage 2, enhanced Web presence, is the typical placement of an American local government on the continuum. Alexander and Grubbs (1998, p. 3) note, "Few sites capitalized on the interactive nature of the Internet to conduct public discussions, maintain bulletin boards, or provide data and information available for download."

In his analysis of the 2000 International City/County Managers Association's (ICMA) E-Government Survey, Moon (2002) finds that a majority of municipalities with populations over 10,000 are not offering transactional Web sites. Furthermore, based on the 2002 ICMA E-Government Survey, only 62 (1.7 percent) municipalities offer online payment of taxes, 113 (3.1 percent) offer online payment of utility bills, and 91 (2.5 percent) offer online payment of fines and fees. The percentages for counties are only slightly higher, with 69 (16.3 percent) offering online payment of taxes, 17 (4.0 percent) offering online payment of utility bills, and 21 (5.0 percent) offering online payment of fines and fees. The state and federal governments offer more robust transactional services, which is to be expected based on information technology diffusion patterns.

CONCLUSION

The role of the Internet continues to increase as more citizens use it to find pertinent information, purchase goods and services, and to participate in virtual communities. By capitalizing on the Internet revolution, governments can create new channels of communication and new methods for participation via e-government. The changing information environment and the movement toward a knowledge economy, juxtaposed against citizen and business demands, mandate that government become involved in e-government initiatives and related uses of information technologies (Ho, 2002). Furthermore, success of existing e-government efforts provides increased legitimacy for further information technology adoption (Norris, 1999).

Although e-government has been offered as a panacea to the ills of the public sector, it has not yet achieved the massive reform and reorientation promised by its proponents. In fact, the e-government movement, similar to other management movements such as total quality management (TQM), new public management, and the reinventing government efforts, has not been successful in altering the core structures of government or avoiding its associated failures (Scavo & Shi, 2000). Given the documented lag time for technology adoption in the public sector, public sector realization of the power of the Internet and e-government must be seen as still nascent at best. By understanding the theoretical premises that underlie e-government and public information technology, programs and processes can be designed to assist government in reaping the benefits associated with the strategic use of information technology and e-government.

REFERENCES

Alexander, J. H., & Grubbs, J. W. (1998). Wired government: Information technology, external public organizations, and cyberdemocracy. *Public Administration and Management: An Interactive Journal, 3*(1). Retrieved from http://www.pamij.com/

Fountain, J. (2001). *Building the virtual state: Information technology and institutional change*. Washington, DC: Brookings Institution.

General Accounting Office. (2002). *National preparedness: Integrating new and existing technology and information sharing into an effective homeland security strategy* (GAO-02-811IT). Washington, DC: Government Printing Office.

Gore, A. (1993). *Creating a government that works better and costs less: Reengineering through information technology* (Report of the National Performance Review). Washington, DC: Government Printing Office.

Hammitt, H. (1999). The legislative foundation of information access policy. In G. D. Garson (Ed.), *Handbook of public information systems* (pp. 27-40). New York: Marcel Dekker, Inc.

Ho, A. T. (2001). Reinventing local government and the e-government initiative. *Public Administration Review, 62*(4), 434-444.

Kettl, D. (2000). *The transformation of governance: Globalization, devolution, and the role of government*. Paper presented at spring meeting of National Academy of Public Administration. Retrieved from *http://www.lafollette.wisc.edu/research/publications/transformation.html*

Kim, P. S., & Wolff, L. W. (1994). Improving government performance: Public management reform and the National

Performance Review. *Public Productivity & Management Review, 18*(1), 73-87.

Kraemer, K. L., et al. (1978). Local government and information technology in the United States. *Local Government and Information Technology, 12*.

Landsbergen, D., & Wolken, G., Jr. (2001). Realizing the promise: Government information systems and the fourth generation of information technology. *Public Administration Review, 61*(2), 206-220.

Larsen, E., & Rainie, L. (2002). The rise of the e-citizen: How people use government agencies' Web sites. *Pew Internet and American life project*. Retrieved from http://*www.pewinternet.org/reports/pdfs/PIP_Govt_Website_Rpt.pdf*

Layne, K., & Lee, J. (2001). Developing fully functional e-government: A four stage model. *Government Information Quarterly, 18*(2), 122-136.

Moon, M. J. (2002). The evolution of e-government among municipalities: Rhetoric or reality? *Public Administration Review, 62*(4), 424-433.

Norris, D. F. (1999). Leading edge information technologies and their adoption: Lessons from US cities. In G. D. Garson (Ed.), *Information technology and computer applications in public administration: Issues and trends* (pp. 137-156). Hershey, PA: Idea Group Publishing.

Norris, D. F., & Kraemer, K. (1996). Mainframe and PC computing in Americancities: Myths and realities. *Public Administration Review, 56*(6), 568-576.

Osborne, D., & Gaebler, T. (1992). *Reinventing government: How entrepreneurial spirit istransforming the public sector*. Reading, MA: Addison-Wesley.

Rahm, D. (1999). The role of information technology in building public administration theory. *Knowledge, Technology, and Policy, 12*(1), 74-83.

Relyea, H. C. (2001). E-gov: The federal overview. *The Journal of Academic Librarianship, 27*(2), 131-148.

Scavo, C., & Shi, Y. (2000). The role of information technology in the reinventing government paradigm: Normative predicates and practical challenges. *Social Science Computer Review, 18*(2), 166-178.

UN & ASPA. (2001). *Benchmarking e-government: A global perspective—Assessing the UN member states.* Retrieved from *http://www.unpan.org/egovernment2.asp*

KEY TERMS

Citizen-Centric: A new approach to organization of government information and dissemination of government services that focuses on citizen needs and desires instead of traditional bureaucratic functions. For example, a citizen-centric Web site may combine various services, provided by different departments or agencies, under a common heading based on life events.

E- (Electronic) Democracy: A new method of engaging citizens in political participation, including e-rulemaking, e-voting, and various means of communication with elected and appointed government officials.

E-Government: A strategy for revolutionizing the business of government through the use of information technology (IT), particularly Web-based technologies, which improve internal and external processes, efficiencies, and service delivery.

E-Procurement: The online provision of goods and services by government agencies. E-procurement includes online requests for proposal (RFPs), online requests for bids (RFB), online bid acceptance, and online monitoring of contracts.

E-Services: The provision of government services via the Internet, including online information, online forms, online transactions, and online referral processes.

Portal: A one-stop, cross-department/business unit and cross-jurisdictional Web site that serves as an aggregation of government services in a given functional area or based on user needs.

"Stovepipe": The traditional orientation of government departments or business units, in which each department or business unit acts independently and autonomously. Lack of integration and interoperability and issues of duplication of efforts characterize stovepipe organizations.

A Socio–Technical Case Study of Bangladesh

Mahesh S. Raisinghani
University of Dallas, USA

Mahboobur Rahman
Gasmin Ltd., Bangladesh

EXECUTIVE SUMMARY

The revolutionary advent of IT has accelerated the pace of day-to-day office, personal and corporate communication. However, even now, Bangladesh, despite its natural resources and easily trainable workforce, cannot cope with the evolution of technology when compared to most of the developed and developing countries. Of course, there are reasons behind the slow growth of technology in Bangladesh. Despite this IT implementation here has begun, albeit a little slowly in different private and public sectors. In Bangladesh, IT use is still in a backward stage in terms of information generation, utilization and applications. In the absence of any private entrepreneur, the government initiates and funds as many of these projects as possible.

Constructing a case study on Bangladesh IT scenario is an impetus to depict the real picture of its obstacles and pitfalls, and success as well, if any.

INTRODUCTION

This write-up topic presents the IT scenario of Bangladesh as a whole, not an organization in particular. Readers will be interested in knowing facts like:

(1) Is there any historical background of computerization in Bangladesh?
(2) Was there any chronological follow-up to next generation of computerization?
(3) How did the private, autonomous and public agencies play roles in different periods of time for computerization?
(4) What are the hindrances that prevent Bangladesh from attaining a sufficient pace for IT advancement?
(5) What is the current status of IT as a whole in the country?

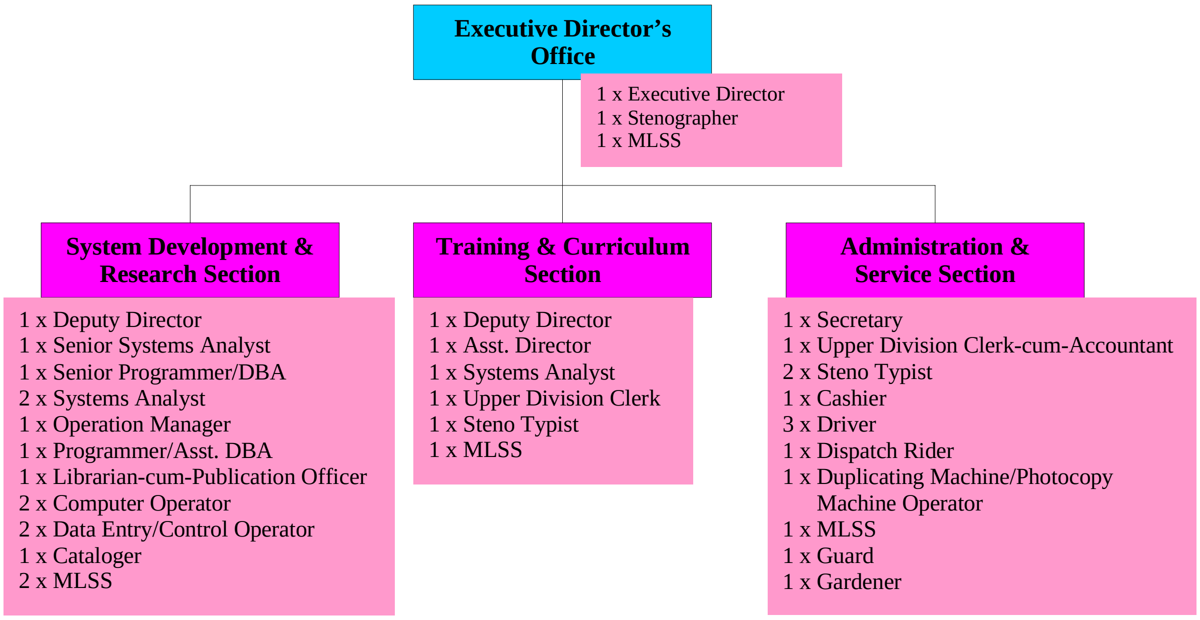
Readers will only get an organizational view as the spokesman for the IT organization (Bangladesh Computer Council) will offer his views during the case deliberations.

BACKGROUND

Bangladesh gained independence from Pakistan in 1971, missing out on much of technological wealth from its predecessor. To begin, the atomic energy commission acquired an IBM 1620 in 1964. This was the first computer in the country. Next, in line were Admajee Jute Mills (the biggest Jute mill in the world) and United Bank Limited (now Agrani Bank) acquiring an IBM 1400 and 1900 series. In early 1970s Bangladesh Bureau of Statistics (BBS) and Bangladesh University of Engineering and Technology (BUET) also acquired mainframe and mini computers. Until 1970s, use of computers was limited to research and calculations. As the boom of microcomputers occurred throughout the world, Bangladesh was merely touched by it. Gradually, the non-government sectors providing support to socio-economic development in the country took lead to accept the microprocessor-based standalone computers (XT and AT type) for their operations, research and office management. This revolution led to the slow but steady use of Microprocessors and Pentium series by government and non-government offices. Also, the installation of UNIX and CP/CMS operating systems on most Mainframe and mini computers in the 1970s made the punch cards and reading systems obsolete. This was followed by the quick adoption of operating systems such as DOS, OS/2, Windows NT, XENIX, AIX, and Mac.

Network technology in Bangladesh started in early 90s with banks and international organizations establishing LAN to meet business and office requirements. Very few government departments opted to have their computer under network environment. First Internet provider began in 1996 as opposed to the traditional dial-up Internet service provider from outside the country. There are more than 50 ISPs in the country, 8-10 more prominent than the others. Bangladesh had missed the opportunity to set up fiber optic infrastructure: first in 1992 when it passed up an almost free of cost opportunity to be connected with an underground cable SEAMEU-2; and second in 1998 when it passed up a minimum cost opportunity with SEAMEU-3. However, recently, there has been an initiation for a connection to the undersea cable

Figure 1. BCC's organizational structure



network APCN-2 and a tri-party contract agreement is under process.

However, very few firms are capable of customizing software; hence this is a small industry in Bangladesh. The necessary ingredients to become a potential exporter of computer software and data processing services do not currently exist in the required quantity in Bangladesh. If Bangladesh wants to enter into this market, it needs to cross a number of hurdles within a very short period. The National Computer Committee offered some ideas for the enhancement of the IT sector, and therefore the BCC designed an organizational structure to carry its functionalities with humanpower strength of 37 as shown in Figure 1.

BCC proposed a budget of US $5 million in their fourth 5-year plan (1990-1995) for the following five proposed IT development projects:

(1) BCC's Infrastructure Facilities
(2) National Information Technology Training Institute (NITI)
(3) National Data Network Authority (NDNA)
(4) National Software Export Bureau (NSEB)
(5) System Integrator Company Limited (SICOM)

Budget allocation was made for three of these projects for financial years 1991-92 and 1992-93. BCC also proposed a new project for establishment of a "data entry industry" in the country in 1992. However, none of the BCC's projects materialized due to some political controversy.

In fact, development of the IT sector and the role of BCC seemed to fade during the period between mid-1991 to mid-1996. In late 1996, a 14-member expert committee headed by a professor of BUET and members from different business and government bodies was formed by the government to study the feasibility of software export from Bangladesh. The committee submitted a report in September 2000 to the government with some very positive recommendations to boost the IT sector in order to create a favorable environment for software industry in the country and to export software. Following the recommendations of the expert committee, the government has taken some favorable decisions from 1998 and it is being continued to date, such as the following:

(1) Declaring IT industry as a thrust sector
(2) Patronizing of IT industry by different financial sectors, especially the banks
(3) Full exemption of import tax for all IT products (e.g., hardware, software, network, peripherals and accessories)
(4) Minimizing VSAT usage charge
(5) Increase of budget allocation for IT sector

In formulating an IT policy in the country, the government has formed a probe committee. The body is active and arranges workshops to draft policies on IT related HRD, procurement and usage guidelines, creating markets and developing infrastructure. The information age has also raised the importance of IT skills both in the workplace and civil society. A lack of such skills on a national level will pose a barrier to Bangladesh competing effectively in the global economy. Therefore, any IT human resource development policy will encourage widespread IT literacy and an adequate supply of IT professionals to support economic goals. The computer literacy rate in Bangladesh is very low due to lack of widespread IT education within the primary and secondary schools. The country needs to produce a large amount of skilled IT humanpower immediately to meet this challenge.

In the context of the world scenario of human resource, Bangladesh's strength is relatively poor. At the university level all national public universities and in particular new universities have started producing computer science graduates through opening computer science departments. Steps have been taken to implement computer education in the secondary and higher secondary levels. Since the last few years vigorous efforts have been made to improve the quality of IT education and research. The massive and strong participation by young people and their urge to become computer literate and develop professional competence is the most encouraging observation.

BCC conducts short-duration professional efficiency enhancement training program in IT. It has started a standard IT training program in Rajshahi division. The Bangladesh Institute of Communication and Information Technology (BICIT) has been included in the Annual Development Program in the 1999-2000 fiscal year. This institute will impart standard IT education, training, standardization and certification of IT related course curricula and products.

BCC, since its formation in 1990 to mid-1992, and some other private firms provided some training on few programming languages and databases, but not at a great extent. Taking this opportunity, several foreign franchisers started appointing their franchisees in Bangladesh. Presently approximately 20 foreign franchisers are operating their IT training centers in Bangladesh.

ANALYSIS

As mentioned, the computer revolution in Bangladesh began in 1960s. Mini (IBM 360, 370), mainframe (IBM 4300, 4361) and microprocessor-based computers were used and programming languages such as FORTRAN, COBOL and RPG remained popular till 1980s. Database management systems used here were SQL/DS, DB2 and earlier version of Oracle. Some organizations still follow the old system, while others have upgraded from mainframe and standalone environment. The current IT usage scenario at a glance can be viewed as shown in Table 1:

Compared to India or Malaysia the IT growth in the country still has a far way to travel. India has attained this tremendous achievement in the past two decades, with liberal policies, patronizing IT entrepreneurs and dependable infrastructure. Government changes neither dampen nor discourage the policies adopted by their predecessor.

Conversely in Bangladesh, government has the sole monopoly. Policy makers came into effect with NCC, NCB and BCC amidst controversy. Different governments modify the IT policies affecting the growth of the IT industry. During 1991-1996, the IT sector was totally ignored. With no follow-up, it may be a while before

Table 1. Current IT usage scenario in Bangladesh

Technology Usage Type		**Technology Product Model**
Computer	Server	IBM RS/4000 and RS/6000, Sun Enterprise, HP Netra, Compaq Proliant, Gateway ALR, Mac
	Client	Pentium (I, II & III), 486 processors, Mac
Operating Systems	Server	UNIX, AIX NT, LINUX, Mac OS, Novel Netware, OS/400
	Client	Windows 9x & NT, Mac OS (Client)
DBMS	Back Engine	Oracle, Sybase, SQL, Access, Informix
	Front Tools	Developer 2000/6i, Delphi, Power Builder, Power Designer
	Language	PL/SQL, T/SQL, Visual Basic
Programming Languages		Visual Basic, C, C++, Java, COBOL, RPG, OOP, J++
Web Design & Development Tools		FrontPage, Dream weaver, HTML Editor, XML, CSS, Java Script, ASP, Hot Dog
Drawing, Animation & Multimedia Tools		Adobe Photoshop, Visio Drawing, Illustrator, Freehand, Corel Draw, Quark Express, Macromedia Director/Flash, Auto Cad, Sound Forge
Back Office Mail Server		MS Exchange Server, Linux, IBM AX

Bangladesh can be at par with the rest of the world. Also, the acceptance attitude and capability to upgrade their knowledge and skills is always under question. There is no customized software due to thin response and lack of enthusiasm from the HRD and the entrepreneurs. The inherent bureaucratic and traditional attitude of the government has caused the major impediment in infrastructure.

DISCUSSION

The Committee formed by the Ministry of Commerce of the People's Republic of Bangladesh has identified a few major problems that threaten the growth of software industry:

- Fiscal
- Human Resource Development
- Infrastructure
- Marketing

In order to make Bangladesh a potential offshore source of software and data processing services, this case makes short-term and medium-term recommendations for each of the above category, depending on the priority and importance.

Short Term

Fiscal

a) To exempt computer hardware and software from import duty, VAT, infrastructure development surcharge, import license fee, and advance income tax.
b) To allow a tax holiday for the export-oriented software and data processing services industry, for 10 (ten) years.
c) To give a 15% price advantage to local software developers over import of the same products.
d) To allow export of software and data processing services through sales contract, instead of letters of credit.
e) To bring the bank interest rate on loans/advances/ overdraft down to the level applicable to other export-oriented thrust sectors.
f) To create a special fund to be administered by the Ministry of Science & Technology for giving interest-free loans to teachers and students for purchase of computers and related equipment.
g) To create a venture capital fund of at least Tk.10 Crore at eExport Promotion Bureau for equity participation in export-oriented software and data processing services companies.

Human Resource Development

a) To give BCC necessary authority to function as the primary facilitator to help growth of private sectors IT industry (such as National Computer Board of Singapore and Department of Electronics of India).
b) To introduce "basic computer skills as a mandatory subject and to set up Computer Science Department in all universities.
d) To ask BCC to update computer science course curriculum.
e) To empower BCC to develop IT related certification exams and to encourage employers of both government and private sectors to give preference to such certificate holders for jobs.

Infrastructure

a) To enact appropriate laws for the protection of intellectual property rights, as required under the WTO Charter.
b) To set up low-cost high-speed data and voice communication link with the USA and the UK, with a minimum speed of 2 MBPS. Private sector should be allowed to provide such service along with BTTB.
e) To make video conferencing facility available through VSAT.
f) To allow private sector to set up their own satellite communication links in order to obtain competitive price advantage and greater availability of communication facilities.
g) To create separate customs houses to handle all incoming and outgoing equipment/ documents so as to ensure clearance of such equipment/documents within 24 hours.
h) To create a central resource center at BCC with current books, magazines, periodicals, software, and user/system manuals on IT related subjects.
i) To encourage software firms to form an association primarily to protect the interest of the software and data processing services sectors, in line with NASSCOM, ASOCIO, WITSA, and JISA.

Marketing

a) To mobilize non-resident Bangladeshi IT professionals and inform them about the incentives being provided by GOB and encourage their support to help Bangladeshi entrepreneurs.

b) To ban use of all pirated software in all organizations, both in the public and private sectors.

c) To encourage all government, semi-government, autonomous organizations and private sector companies to replace the manual system of documentation and records by computerized system through the use of locally developed customized application software.

d) To send marketing missions to North America/E.U. consisting of members from IT Associations and EPB, on a regular basis, with a view to publicizing Bangladesh software and data processing services capabilities as well as establishing personal contacts with the prospective customers.

e) To create a database and a Web site of all major organizations/institutions engaged in out-sourcing of software and data processing services.

f) To explore the possibility of obtaining business on sub-contract basis from the suppliers of software and data processing services in India, Sri Lanka and the Philippines.

g) To ask the concerned trade associations to organize international exhibitions/fairs in Bangladesh for IT products and services, in collaboration with Export Promotion Bureau.

Medium Term

Fiscal

a) To create a market promotion fund to be administered by Export Promotion Bureau for meeting the expenses of promoting Bangladesh as a potential source of software and data processing services to the overseas markets.

b) To create a special fund for supporting industry oriented IT research and development activities, to be administered by BCC.

Human Resource Development

a) To strengthen BCC and make it responsible for imparting higher-level special need-based training to the IT professionals who are graduating from the universities.

b) To incorporate industry placement program in final year of a computer science degree program.

Infrastructure

a) To set up an Information Technology Village (ITV) and to equip the same with all necessary facilities, such as high-speed communication, special custom bonded warehouse, resource center (hardware, software, manuals, book), power and water supplies, and telecom facilities.

b) To ask BTTB to set up ISDN/HDSN/ADSL lines all over the country, and a fibre optic backbone.

c) To set up a communication hub in Bangladesh.

d) To form a standing committee to formulate and implement policies, strategies and action plans for promotion of export of software and data processing services.

Marketing

a) To ask Export Promotion Bureau to set up permanent liaison offices in the USA and the UK to be manned by professional marketers of IT field, who should perform and achieve definite quantified business objectives.

b) To ask BCC to create a database of Bangladesh IT professionals working at home and abroad, in order to obtain their help when needed.

d) To encourage IT industry members to take steps for ISO-9000 certifications and eventually ratings from Software Engineering Institute (SEI) of Carnegie-Mellon University of USA.

Other Recommendations

Bangladesh can also learn from the experience of its neighboring country – India, and can adopt the measures that helped India.

(1) Take advantage of the opportunities available in the conversion market of US $650 billion and the Euro-currency conversion.

(2) Focus on multimedia market that is experiencing the maximum growth.

(3) Influence Microsoft Corporation of USA to involve Bangladesh in their plan to introduce Bangla as a language for Windows NT 5.0.

CONCLUSION

The fact that Bangladesh is far behind local and global IT status is established. Currently, a lot of decisions and steps have been taken in favor to uphold and patronize IT sector. Many entrepreneurs and investors are attracted to the IT sector. The need for an IT policy for the development of the IT sector within the framework of overall national development is now well recognized. The massive and strong participation by young people and their urge to become computer literate and develop professional competence is the most encouraging.

REFERENCES

Bangladesh Computer Council. (2000). Ministry of Science and Technology, Bangladesh. Retrieved November 6, 2000, from *http://www.bccbd.org/*

Heeks, R. (1999). Reinventing government in the information age. In R. Heeks (Ed.), *International practices in IT-enabled public sector reform.* Routledge Publishers.

Kirkpatrick, D. (2001, February 5). Looking for profits in poverty. *Fortune*, 174-176.

NASSCOM. (2000). National Association of Software and Service Companies. Retrieved November 6, 2000, from *http://www.nasscom.org/*

KEY TERMS

Export-Oriented Software: A unit will be considered export-oriented if at least 70% of its revenue comes from export.

Long Term: Timeframe of three to four years to accomplish the recommendations.

Medium Term: Timeframe of two to three years to accomplish the recommendations.

NASSCOM: National Association of Software and Service Companies.

Short Term: Timeframe of a year to accomplish the recommendations.

Taka (Tk): Bangladesh currency is called Taka (1 US Dollar = 60.552 Bangladeshi Taka, 1 Bangladeshi Taka (BDT) = 0.01651 US Dollar (USD) as of March 14, 2004).

Tax Holiday: No taxes for export-oriented software and data processing services industry every 10 years.

A Systemic Approach for Information Systems Evaluation

Ala Abu-Samaha
Amman University, Jordan

INTRODUCTION

Traditionally, information technology (IT) evaluation, pre-implementation appraisals, and post-implementation reviews have been characterised as economical, tangible, and hard in nature. The literature review on IT evaluation shows a great bias towards using economical and tangible measures that represent the management's view of what is 'good' and 'bad', which had been described as narrow in scope and limited in use. Smithson and Hirschheim (1998) explain that "there has been an increasing concern that narrow cost benefit studies are too limited and there is a need to develop a wider view of the impact of a new system." Ezingeard (1998) emphasises the importance of looking at the impact of IS on both the overall system and the whole organisation.

The concern of IT evaluation is to measure whether the IT solution meets its technical objectives and to what extent. In such activity, instrumentation is highly appreciated and particularly chosen. Product oriented, instrument led, and unitary are the main characteristics of such a perception. Mainly evaluation was seen as a by-product of the decision-making process of information systems development and installation. Most of the evaluation tools and techniques used were economic based, geared to identifying possible alternatives, weighing the benefits against the costs, and then choosing the most appropriate alternative. St. Leger, Schnieden, and Walsworth-Bell (1992) explain that evaluation is "the critical assessment, on as objective a basis as possible, of the degree to which entire services or their component parts fulfil stated goals." Rossi and Freeman (1982) advocate that "evaluation research is the systematic application of the practice of social research procedures in assessing the conceptualisation and design, implementation, and utility of social intervention programs."

Post-implementation evaluation has been described by Ahituv, Even-Tsur, and Sadan (1986) as "probably the most neglected activity along the system life cycle." Avison and Horton (1993) report that "evaluation during the development of an information system, as an integral part of the information systems development process, is even more infrequently practised." In acknowledging all of the above concerns about the evaluation of IT interventions, the author presents in this article a measures identification method that aims at identifying those measures or indicators of performance that are relevant to all the stakeholders involved in such interventions.

POST IMPLEMENTATION REVIEW

Many researchers concerned with IT evaluation, mainly post-implementation reviews, have identified an urgent need to migrate from this 'traditional' and economical view towards using a mixed approach to IT evaluation. Such an approach will allow IT evaluators to mix between 'hard' and 'soft' measures, as well as economical and non-economical measures (Chan, 1998; Ezingeard, 1998; Bannister, 1998; Smithson & Hirschheim, 1998; Avison 7 Horton, 1993). Furthermore, there is a need to shift towards utilising an approach that reflects the concerns of all involved stakeholders rather than a Unitarian approach (Smithson & Hirschheim, 1998). Any systemic approach to IS evaluation must take into account two main issues regarding the collective nature of IS: choosing the relevant measures of performance, and equal account for economical as well as non-economical measures.

Choosing the Relevant Measures of Performance

Abu-Samaha and Wood (1999b) show that:

"the main methodological problem in evaluating any project is to choose the right indicators for the measurement of success, or lack of it. These indicators will obviously be linked to aims but will also be relevant to the objectives chosen to achieve these aims, since if the wrong objectives have been chosen for the achievement of an aim, then failure can be as much due to inappropriate objectives as to the wrong implementation of the right objectives."

Willcocks (1992), in a study of 50 organisations, gives 10 basic reasons for failure in evaluation practice; amongst these reasons are inappropriate measures and neglecting intangible benefits. Ezingeard (1998) shows that "…it is difficult to decide what performance measures should be used." On the other hand, a different set of indicators or measures of performance will be chosen at each level or layer of the IT intervention (product, project, and programme), which adds more to the relevance of the chosen measures.

Another important aspect of choosing indicators or measures of performance is to choose the relevant measures that add value to a particular person or group of persons. Smithson and Hirschheim (1998) explain:

"There are different stakeholders likely to have different views about what should be the outcome of IS, and how well these outcomes are met. Who the different stakeholders are similarly need[s] to be identified."

The measures identification method proposed by the author in this article provides such relevance through the identification of stakeholders and the subsequent 'human activity system' analysis. This is done by exploring the particular worldview, which is unique for each stakeholder, and the identification of the relevant criteria for *efficacy*, *efficiency*, and *effectiveness* of each stated transformation process. Such investigation would allow for the identification of the most relevant measures or indicators of performance for the stated stakeholder(s).

Equal Account for Economical as well as Non-Economical Measures

Post-implementation reviews have had a tendency to concentrate on 'hard', 'economical', and 'tangible' measures. Chan (1998) explains the importance of bridging the gap between 'hard' and 'soft' measures in IT evaluation, realising that "this in turn requires the examination of a variety of qualitative and quantitative measures, and the use of individual, group, process, and organisation-level measures." Avison and Horton (1993) warn against confining post-implementation reviews to monitoring cost and performance and feasibility studies on cost-justification, saying that "concentration on the economic and technical aspects of a system may cause organisational and social factors to be overlooked, yet these can have a significant impact on the effectiveness of the system." Fitzgerald (1993) suggests that a new approach to IS evaluation, which addresses both efficiency and effectiveness criteria, is required.

The approach described in this article gives an equal account to tangible as well as intangible benefits of IT intervention by identifying efficacy and effectiveness measures along with efficiency measures. The measures identification method proposed by the author provides a better understanding of the context of evaluation which would give a better account of the content of evaluation.

SOFT EVALUATION

The approach advocated here brings together formal work in evaluation (Patton, 1986; Rossi & Freeman, 1982) with a qualitative process of investigation based on Soft Systems Methodology (Checkland & Scholes, 1990) in order to allow us to make judgements about the outcomes of an implementation from a number of different viewpoints or perspectives. The performance measures identification method proposed in this article operates through three stages.

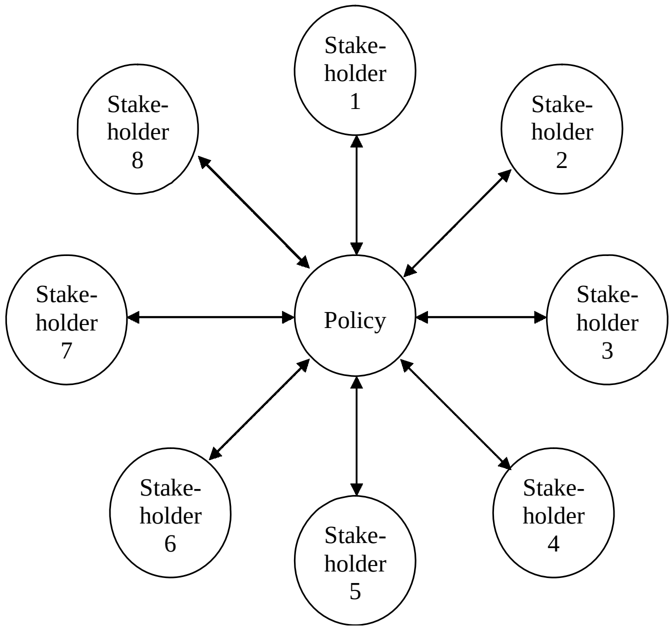
Stage One: Stakeholder Analysis

The first stage of the proposed method is to identify the intra- and inter-organisational stakeholders involved in the intervention. A stakeholder, as defined by Mitroff and Linstone (1993), is any "individual, group, organisation, or institution that can affect as well as be affected by an individual's, group's, organisation's, or institution's policy or policies." Mitroff and Linstone (1993) explain that an "organisation is not a physical 'thing' per se, but a series of social and institutional relationships between a wide series of parties. As these relationships change over time, the organisation itself changes." Mitroff and Linstone's view of an organisation is synonymous to Checkland and Howell's (1997), which negates the 'hard goal seeking machine' organisation.

Stakeholder analysis can be seen as a useful tool to shed some light on the subjective process of identifying relevant measures of performance for evaluation. A number of questions can be asked at this stage such as where to start, who to include, and who to leave out. The value of the investigation will be of greater importance if all relevant stakeholders are identified and included in the evaluation effort. It is obvious at this stage that some stakeholders will be of greater importance than others because of the power base that they operate from, and such stakeholders are to be acknowledged. At the same time, however, other relevant stakeholders should not be undermined for lack of such power.

While Mitroff and Linstone (1993) do not describe the process through which stakeholders may be identified, they recommend the use of a *stakeholder map*, as shown in Figure 1. They explain that "a double line of influence extends from each stakeholder to the organisation's policy or policies and back again-an organisation is the entire set of relationships it has with itself and its stakeholders." On

Figure 1. Generic stakeholder map (Adopted from Mitroff & Linstone, 1993, p. 141)



the other hand, Vidgen (1994) recommends the use of a rich picture, being a pictorial representation of the current status, to support this activity.

Stage Two: Human Activity Systems Analysis

The second stage of the proposed method constitutes the systemic organisational analysis of the problem situation or subject of evaluation. SSM is based upon systems theory, and provides an antidote to more conventional, 'reductionist' scientific enquiry. Systems approaches attempt to study the wider picture: the relation of component parts to each other, and to the whole of which they are a part ('holism') (Avison & Wood-Harper, 1990; Checkland, 1981; Checkland & Scholes, 1990; Lewis, 1994; Wilson, 1990). SSM uses systems not as representations of the real world, but as epistemological devices to provoke thinking about the real world. Hence SSM is not an information systems design methodology. It is rather a general problem-solving tool that provides us with the ability to cope with multiple, possibly conflicting, viewpoints.

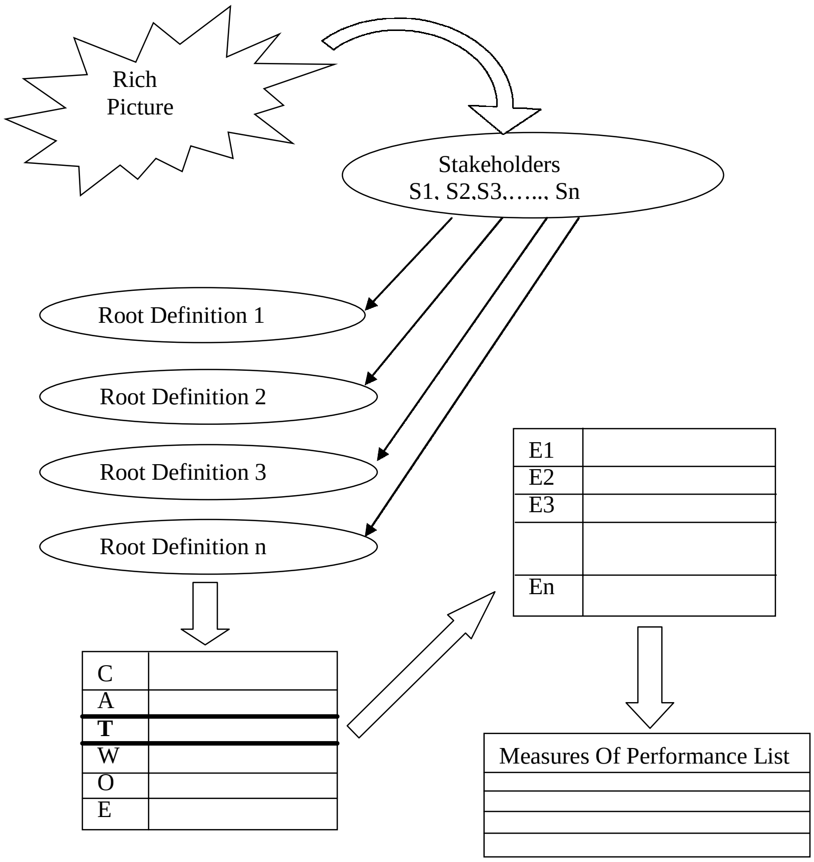
Within this stage the aim is to determine a *root definition* for each stakeholder, which includes their "worldview," a transformation process, together with the definition of the criteria for the efficacy, efficiency, and effectiveness of each stated transformation process (Wood-Harper, Corder, Wood & Watson, 1996). A root definition is a short textual definition of the aims and means of the relevant system to be modelled. The core purpose is expressed as a transformation process in which some entity 'input' is changed or transformed into some new form of that same entity 'output' (Checkland & Scholes, 1990). Besides identifying the world-view and the transformation process, there are the three E's to consider as stated by Checkland and Scholes (1990):

- *E1: Efficacy.* Does the system work? Is the transformation achieved?
- *E2: Efficiency.* A comparison of the value of the output of the system and the resources needed to achieve the output; in other words, is the system worthwhile?
- *E3: Effectiveness.* Does the system achieve its longer-term goals?

Stage Three: Measures of Performance Identification

According to Checkland's Formal Systems Model, every human activity system must have some ways of evaluating its performance, and ways of regulating itself when the desired performance is not being achieved. This final stage aims at identifying appropriate Measures of Performance related to the criteria identified in the earlier stage.

Figure 2. Generic process of soft evaluation



For each criterion of the three Es identified in the earlier stage, it is important to identify the measure of performance that would allow us to judge the extent to which the system has achieved its objectives.

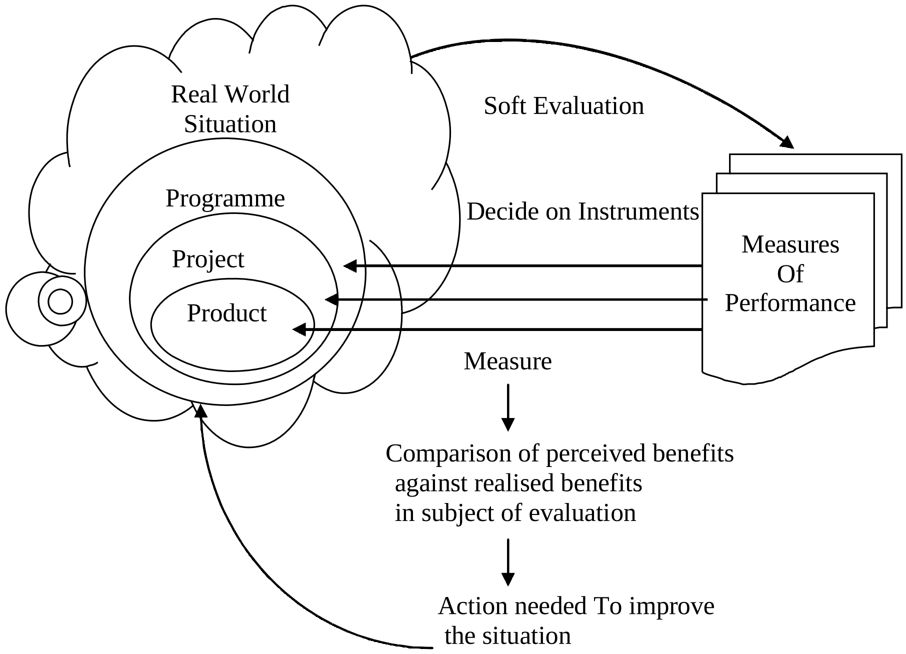
Figure (2) provides a diagrammatic representation of the generic process of soft evaluation, which shows the relationship between the different tools used to investigate the success/failure of an IT intervention.

CONCLUSION

This proposed IT evaluation framework is refined from an earlier work by Abu-Samaha and Wood (1999) to reflect the layered nature of IT interventions. The approach advocated here is founded on the interpretive paradigm (Burrell & Morgan, 1979) and brings together formal work in evaluation (Patton, 1986; Rossi & Freeman, 1982) with a qualitative process of investigation based on Soft Systems Methodology (Checkland & Scholes, 1990) in order to allow us to make judgements about the outcomes of an implementation from a number of different viewpoints or perspectives. The literature survey has revealed three levels of IS/IT intervention evaluation. Measures would differ at each level of analysis, prompting a different instrument as well as a different process of evaluation. Figures 3 and 4 provide a diagrammatic representation of the soft evaluation method and framework, where the three layers of IT intervention-product, project, and programme-have been included in the evaluation process.

Relevance is of paramount importance for IT and IS evaluation. The main methodological problem in evaluating any project is to choose the right indicators for the measurement of success, or lack of it. The earlier classification of product, project, and programme evaluation shows that a different set of indicators or measures of performance will be chosen at each level or layer of the IT

Figure 3. Soft evaluation framework in use



intervention which adds more to the relevance of the chosen measures. Another important aspect of choosing indicators or measures of performance is to choose the relevant measures that add value to a particular person or group of persons. Non-economic as well as economic measures are of equal importance. The proposed approach gives an equal consideration to the tangible as well as the intangible benefits of IT intervention by identifying both efficacy and effectiveness measures along with efficiency measures.

REFERENCES

Abu-Samaha, A. & Wood, J.R.G. (1998, June 7-9). Evaluating inter-organisational systems: The case of (EDI) in general practice. *Proceedings of the 11th Bled Electronic Commerce Conference* (pp. 174-190), Bled, Slovenia.

Abu-Samaha, A. & Wood, J.R.G. (1999a, March 6-9). GP/provider links: Who benefits? *Healthcare Computing Conference* (pp. 114-120), Harrogate, UK.

Abu-Samaha, A. & Wood, J.R.G. (1999b, November 4-5). Soft evaluation: A performance measures identification method for post-implementation reviews. *Proceedings of the 6th European Conference on Information Technology Evaluation* (pp. 221-228), Brunel University, Uxbridge, London.

Ahituv, N., Even-Tsur, D. & Sadan, B. (1986). Procedures and practices for conducting post-evaluation of information systems. *Journal of Information Systems Management, 3*(2).

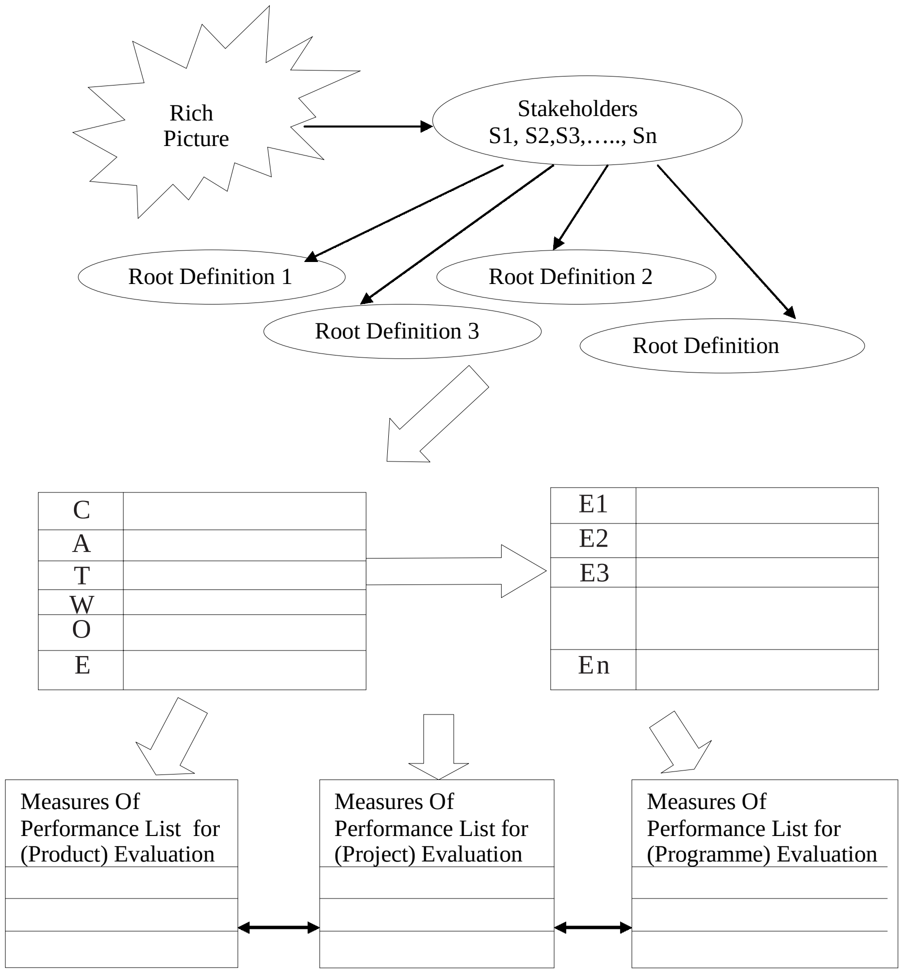
Avison, D.E. & Horton, J. (1993). Evaluation and information systems development. In R. Arduini (Ed.), *Investimenti in information technology nel settore bancario* (pp. 248-279). Milan: Franco Angeli.

Avison, D.E. & Wood-Harper, A.T. (1990). *Multiview: An exploration in information systems development*. Oxford: Blackwell Scientific Publications.

Bannister, F. (1998). In defence of instinct: IT value and investment decisions. *Proceedings of the 5th European Conference on the Evaluation of Information Technology*, Reading University Management Unit, Reading, UK.

Chan, Y.E. (1998). IT value-the great divide between qualitative and quantitative, and individual and organisational, measures. *Proceedings of the 5th European Conference on the Evaluation of Information Technology*, Reading University Management Unit, Reading, UK.

Figure 4. Soft evaluation method in use



Checkland, P. (1981). *Systems thinking, systems practice.* Chichester: John Wiley & Sons.

Checkland, P. & Howell, S. (1997). *Information, systems and information systems: Making sense of the field.* Chichester, UK: John Wiley & Sons.

Checkland, P. & Scholes, J. (1990). *Soft Systems Methodology in action.* Chichester, UK: John Wiley & Sons.

Ezingeard, J.-N. (1998). Towards performance measurement process for manufacturing information systems. *Proceedings of the 5th European Conference on the Evaluation of Information Technology*, Reading University Management Unit, Reading, UK.

Farhoomand, A.F. & Drury, D.H. (1996). Factors influencing electronic data interchange success. *Data Base Advances, 27*(1), 45-57.

Howles, J. & Wood, M. (1995). Diffusion and management of electronic data interchange: Barriers and opportunities in the UK pharmaceutical and health-care industries. *Technology Analysis and Strategic Management, 7*(4), 371-386.

Mitroff, I.I. & Linstone, H.A. (1993). *Unbounded mind: Breaking the chains of traditional business thinking.* Oxford: Oxford University Press.

Patton, M.Q. (1986). *Utilisation-focused evaluation* (2nd edition). London: Sage Publications.

Rossi, P.H. & Freeman, H.E. (1982). *Evaluation: A systematic approach* (2nd edition). London: Sage Publications.

St. Leger, A.S., Schnieden, H. & Walsworth-Bell, J.P. (1992). *Evaluating health services effectiveness.* Open University Press, Milton Keynes.

Symons, V.J. (1991). A review of information systems evaluation: Content, context and process. *European Journal of Information Systems, 1*(3), 205-212.

Willcocks, L. (1992). Evaluating information technology investments: Research findings and reappraisal. *Journal of Information Systems*, *2*(4).

Wilson, B. (1990). *Systems: Concepts, methodologies and applications*. Chichester: John Wiley & Sons.

Wood-Harper, A.T., Corder, S., Wood, J.R.G., & Watson, H. (1996). How we profess: The ethical systems analyst. *Communications of the ACM*, *39*(3), 69-77.

Vidgen, R. (1994, August 31-September 2). Research in progress: Using stakeholder analysis to test primary task conceptual models in information systems development. *Proceedings of the 2nd International Conference of the British Computer Society Information Systems Methodologies Specialist Group*, Heriot-Watt University, Edinburgh, UK.

KEY TERMS

Effectiveness: A measure of performance that specifies whether the system achieves its longer-term goals or not.

Efficacy: A measure of performance that establishes whether the system works or not-if the transformation of an entity from an input stage to an output stage has been achieved?

Efficiency: A measure of performance based on comparison of the value of the output of the system and the resources needed to achieve the output; in other words, is the system worthwhile?

IS Success: The extent to which the system achieves the goals for which it was designed. In other words, IS success should be evaluated on the degree to which the original objectives of a system are accomplished (Oxford Dictionary).

Post-Implementation Evaluation: The critical assessment of the degree to which the original objectives of the technical system are accomplished through the systematic collection of information about the activities, characteristics, and outcomes of the system/technology for use by specific people to reduce uncertainties, improve effectiveness, and make decisions with regard to what those programs are doing and affecting (Patton, 1986).

Product Evaluation: The emphasis at this level is on the technical product, software, IT solution, or information system. The concern at this level is to measure to what extent the IT solution meets its technical objectives. The emphasis here is on efficacy: Does the chosen tool work? Does it deliver?

Programme Evaluation: The concern at this level is to measure whether the programme aims have been met and whether the correct aims have been identified.

Project Evaluation: The emphasis at this level is on the project, which represents the chosen objective(s) to achieve the aims of the programme. The concern at this level is to measure whether the project meets these objectives and to what extent it matches user expectations. The emphasis here is on efficiency. Does the project meet its objectives using minimal resources?

Soft Systems Methodology: A general problem-solving tool that provides us with the ability to cope with multiple, possibly conflicting viewpoints based on systems theory, which attempts to study the wider picture; the relation of component parts to each other. It uses systems as epistemological devices to provoke thinking about the real world.

Stakeholder: Any individual, group, organisation, or institution that can affect as well as be affected by an individual's, group's, organisation's, or institution's policy or policies (Mitroff & Linstone, 1993).

A University/Community Partnership to Bridge the Digital Divide

David Ruppel
The University of Toledo, USA

Cynthia Ruppel
The University of Alabama in Huntsville, USA

INTRODUCTION

Companies want employees with core values who ascribe to corporate values. Emotional intelligence (EQ) is used by companies in recruitment (Foote, 2001), and guides managers in dealing with team performance problems. Similarly, leadership requires refocusing on core values, which over time builds character (Badaracco, 1998). Thus, educational institutions should devote considerable attention to character building (Foote, 2001).

Service-learning is designed to help. Jacoby (1996a, p. 5) has defined service-learning as "…a form of experiential education in which students engage in activities that address human and community needs together with structured opportunities intentionally designed to promote student learning and development".

Service-learning is important in information technology where students need technical skills and experience, and a strong ethical foundation. Legal aspects of technology have not kept pace with technology; often IT people are confronted with complex ethical decisions. It has been argued that service-learning represents a "unique pedagogy…that enhances the ability of private sector managers to be effective stewards of society's moral authority" (Godfred, p. 364). Service-learning in colleges is tightly linked with K-12 education (Jacoby, 1996B) due to the growing number of at-risk children, a vested interest for colleges to improve the future students, and because students will view service-learning as an appropriate college activity if they benefited from it prior to college (Jacoby, 1996b).

A policy concern in the information age is the "digital divide," a gap between those who have easy access to technology and those who do not. References are made to information "haves" and "have-nots" in an age where information is equivalent to wealth (Holloway, 2000). The "have-nots" are in danger of exclusion from the new economy and marginalization into low-wage jobs (Dunham, 1999). In 2000, the President of the United States asked the IT community to help close this digital divide for moral reasons and to ensure that the economy flourishes with the availability of skilled workers (Shewmake, 2000).

This overview summarizes a five-phase service-learning project accomplished through a partnership between the University of Toledo and a local K-8 parochial/non-profit school. The students were primarily enrolled in a Systems Analysis, Design and Implementation course (SAD). This longitudinal project was undertaken to plan, design, and wire a network for the school and to assess and implement continuing and future computer needs. It allowed students to gain "real-life" experience while contributing to the growth of IT among children in a non-profit setting.

BACKGROUND

The school is a parochial school enrolling approximately 200-250 students. All grades have a dedicated classroom; a computer lab and library are also provided.

Existing computers consisted of a classroom set of older Macintosh computers in the 8th grade room. Each classroom had an older Macintosh computer for the teacher, all with unused LAN capability. The computer lab contained older Apple computers used in the primary grades for computer literacy and keyboarding skills.

Phase 1

The school had accumulated technology funds and hired a teacher with a Master's degree in Educational Technology. The teacher and principal agreed to participate in the project since an estimate from a local company exceeded the funds accumulated. While the teacher had pedagogic knowledge of computers, he did not possess the expertise to evaluate the quotation or analyze the technical aspects of the network. The school indicated that it hoped to apply for a grant, but needed technical information.

Students self-selected into the project: *The goal of the project was to educate themselves as to alternatives, costs and provide background information concerning networking to prepare a grant application.* They had the opportunity to examine the existing environment and interview stakeholders.

The instructor and students toured the building, including attic and closet locations where existing asbestos could not be disturbed. The stakeholders were asked to determine the number of “drops/connections” required in each room based on immediate and future use. Two drops were requested in each classroom — one for the teacher’s computer and another for a classroom network. The group submitted a plan to the school including alternatives, costs, and technical information for the design of the campus network to be used for grant preparation.

Phase 2

Phase 2 included completing a grant proposal and the physical networking of the building. The wiring project was popular among students and required instructor selection to participate. Two students had experience, while others were chosen based on enthusiasm and desire to learn the “hands-on” aspects of networking.

Using the plan, a grant proposal was submitted providing evidence of the school’s commitment and a plan for the educational use of the network. The university students’ involvement was documented, and the authors were listed as consultants.

The grant writing was divided among the authors, the teacher, and the principal. Guidelines required a technology plan and a specific grant request. The accumulated funds were sufficient to wire the building without grant funding. Subsequently the maximum grant was awarded for continuation of the project.

The wiring plan included an Ethernet LAN in the 8th grade room and a campus LAN with connections in all classrooms and offices. Microsoft Windows NT Server 4.0 with Services for Macintosh (SFM) was chosen as the network operating system (NOS) based on the requirements of the heterogeneous network. With SFM, the Windows server appears as an Apple server to Macintosh clients.

The NOS was installed on a computer with a dial-up Internet connection, a temporary arrangement until high-speed access was available. Proxy server and content-filtering services were installed, providing low-bandwidth Internet access to all clients in the 8th grade classroom. Secure storage for network users was provided.

In wiring the building, a storage closet was used as the wiring closet where hubs were installed and all building wiring runs terminated. Since the computers were regularly used, work was partitioned into elements that could be done over a weekend to maximize availability for students and teachers. A file server for administrative applications and intranet e-mail was installed in the wiring closet with a tape drive to provide network backup.

Phase 3

The next step was to install high-speed Internet access for which the school receives state funding. The authors recommended the installation of a T-1 line (1.544 Mbps).

In 2001 the network consisted of three servers, each running Microsoft NT Server 4.0. A laser printer was available to over 50 network clients running various operating systems.

Phase 4

We now needed to recommend replacements for outdated equipment and provide information for equipment grant applications. Students self-selected into this project, and met with the stakeholders. The main issue was whether the replacement equipment should be Apple or Windows-based. Pertinent factors were educational needs, existing equipment, and available expertise. The group analyzed these factors together with their survey of surrounding schools to determine local norms. Their recommendation was Apple equipment.

After network completion, it was obvious that the computers were outdated. An early goal was to minimize expenses by leveraging existing equipment. An obsolete Web browser was installed since the machines’ capabilities prohibited current versions. While slow, the clients provided Internet access where none existed before. In today’s world, fast access is a necessity. Revenaugh (2000) suggests that access does not equal equity; just connecting a wire to a school does not mean it will be used well. Also, the browser security certificates had expired, so secure Web sites could not be accessed.

Phase 5

Using this recommendation, a second grant proposal was prepared requesting the maximum award to purchase new computers for the teachers. The replaced computers will be placed in the classrooms in small networks for student use.

Another group of students worked on designing a Web site for the school. A preliminary school Web site, created 2 years earlier, was out-of-date and the students were asked to produce an up-to-date, easily maintainable Web site.

CONCLUSIONS AND LESSONS LEARNED

There must be a close working relationship between the community organization and the university faculty. Trust was necessary to order materials, remain in the building after-hours, and obtain quotations on behalf of the school.

Since this is a functioning organization, work cannot be left unfinished. Work must be analyzed and divided into tasks that can be completed within the available time. Student self-selection should be combined with faculty approval since the faculty member(s) have the ultimate responsibility for project completion.

Students appreciate being part of a project that benefits others. Food was provided on Saturdays, motivating the students and keeping them on site. The students had time to relax and discuss the project among themselves and with the instructors informally as recommended by the service-learning literature (Godfrey, 1999; Kenworthy-U'ren, 1999). During this time it was interesting to note their perceptions of how elementary education has changed since "their day". They even became attached to the classroom pets – feeding them, and so forth.

The school purchased materials on time and provided significant access to the building. They valued the students' work, since they had received a commercial bid. We recommend this procedure where applicable to eliminate the perception that free means valueless.

Choose the students both to ensure project completion and to make it a learning experience. From the volunteers, students were selected to ensure some students had necessary skills and also included inexperienced students. This facilitated a sharing of knowledge and maximized learning, giving the experienced students leadership opportunities.

Expect unanticipated problems. Since the building had been built in stages (to our surprise), we needed to wire through old exterior walls. Also, the 8th grade class rabbit ate through the cables and had to be sent to a new home. We recommend exploring insurance coverage for students involved in labor with the potential for injury.

It is important to seek knowledge from a variety of sources. Experience with Apple computers was limited among the students and the instructors. A student, not part of this project, was familiar with Apples and made relevant suggestions. Another former student, who works with Apple computers in his employment, supplied information, agreed to troubleshoot computers, and created a CD containing the software set-up for the computers. He showed the students how to install the set-up and was so impressed with one of the students that he arranged a job interview for the student, resulting in a job offer.

In the service-learning literature, one caveat is the time required of the instructor(s), particularly relative to the current faculty reward systems (Godfrey, 1999; Kenworthy-U'ren, 1999; Kolenko, Porter, Wheatley & Colby, 1996). The amount of instructor time and nature of instructor tasks varied according to the type of student projects undertaken. This is consistent with the findings of Bush-Bacelis (1998), who suggests that service-learning projects do not require more work, but a different type of work both for the students and the instructor. The wiring project required a significant amount of dedicated instructor time, as did the grant writing. However, students indicated they found the wiring project particularly fulfilling, especially since the feedback was a working network and a tangible skill set was learned.

Student Assessment

In group projects, students are asked to assess each member's contribution to the group. To assess the quality grade for those completing the physical wiring, the instructors were present in the building. For the more traditional projects, the instructor assessed quality and adherence to SAD principles after obtaining "perceived value and usefulness" feedback from stakeholders.

Benefits to the School

The age of the existing computers limited their usability, and the school did not have the funds to replace them. By placing them on a network, their useful life was extended. Due to limited funds and lack of technical knowledge the school was prevented from completing these projects.

Benefits to the Students

Kenworthy-U'ren (1999) suggests that service-learning projects should be grounded in experiential learning, not merely a service-related activity. These projects allowed the students' involvement in projects related to their career objectives. The students appeared determined to provide a quality product because it was being given to a "real" organization for decision making and to the next group who would use their work and attempt to "better it". These projects, if carefully planned, are win-win situations for the students and the community partner.

FUTURE TRENDS

The necessity of networking schools is becoming increasingly important and at lower levels such as primary school also. To continue to eliminate the digital divide, students, even grade school students, must be able to access and evaluate information to learn to make informed decisions. This project was designed to both give students "hands-on" experience and to provide access to young students so they can become familiar and comfortable with this access.

The technology in providing this access will continue to change. Many devices today are wireless devices (as schools can afford to purchase them), which would eliminate the need to wire the building and would use more current technology. However, in a school setting, due to cost constraints, the available technology appears to change more slowly than a business environment would. The issues involved in setting up a wireless network may seem less involved since no physical wiring must take place, but different complex issues arise, such as providing coverage to all areas and methods of extending the signal on large campuses. So the exercise for the students becomes no less important.

REFERENCES

Badaracco, J.L. (1998). The discipline of building character. *Harvard Business Review, 76*(2), 114-124.

Dunham, R.S. (1999, August 2). Across America, a troubling 'digital divide.' *Business Week, n3640,* 40.

Foote, D. (2001, February 12). What's your 'emotional intelligence'? *Computerworld*.

Godfrey, P.C. (1999). Service-learning and management education: A call to action. *Journal of Management Inquiry, 8*(4), 363-378.

Holloway, J.H. (2000). The digital divide. *Educational Leadership, 58*(2), 90-91.

Jacoby, B. (1996a). Service-learning in today's higher education. In B. Jacoby (Ed.), *Service learning in higher education: Concepts and practices* (pp. 3-25). San Francisco: Jossey-Bass Publishers.

Jacoby, B. (1996b). Securing the future of service-learning in higher education: A mandate for action. In B. Jacoby (Ed.), *Service learning in higher education: Concepts and practices* (pp. 317-335). San Francisco: Jossey-Bass Publishers.

Kenworthy-U'ren, A.L. (1999). Management students as consultants: An alternative perspective on the service-learning "call to action." *Journal of Management Inquiry, 8*(4), 379-387.

Kolenko, T.A., Porter, G., Wheatley, W., & Colby, M. (1996). A critique of service learning projects in management education: Pedagogical foundations, barriers, and guidelines. *Journal of Business Ethics, 15,* 133-142.

Shewmake, B. (2000). Clinton to IT execs: Help close digital divide. *InfoWorld, 22*(7), 12.

KEY TERMS

Bandwidth: A measure of the data transmission capacity of a communications link.

Digital Divide: The term digital divide describes the fact that the world can be divided into people who do and people who do not have access to - and the capability to use - modern information technology, such as the telephone, television, or the Internet.

Ethernet: A communications standard (IEEE 802.3) for Local Area Networks (LAN). When a device wishes to transmit, it waits until the link is empty and then transmits. In the event that two or more devices transmit simultaneously (collision), all devices stop transmitting and wait a random time period before attempting to re-transmit.

Heterogeneous Network: A network where network clients run a variety of operating systems. An example would be a network of machines running Windows, Unix, and Mac OS X.

Intranet: Computer network contained entirely within an organization.

LAN: Local Area Network. Group of computers and other devices sharing a single communications link. Typically, the longest distance between any two connections is a few kilometers. Usually owned by a single organization.

Network Operating System: An operating system for a computer designated as a server in a LAN. Manages network resources available to network clients such as printers, files and databases. Responsible for maintaining a database of names and addresses of clients attached to the network.

Service-Learning: A form of experiential education in which students engage in activities that address human and community needs together with structured opportunities intentionally designed to promote student learning and development.

Systems Analysis and Design Course (SAD): A course in which the student learns to understand how an information system can support organizational needs, how to design the system, build it and deliver it to users.

T-1 Line: A communications link that can be used for digital data transmission. Provides a data rate (bandwidth) of 1.544 million bits per second (Mbps).

A Web-Geographical Information System to Support Territorial Data Integration

V. De Antonellis
Universitá di Brescia, Italy

G. Pozzi
Politecnico di Milano, Italy

F.A. Schreiber
Politecnico di Milano, Italy

L. Tanca
Politecnico di Milano, Italy

L. Tosi
Comune di Milano, Italy

INTRODUCTION

The design of a Web-geographical information system, Web-GIS (Worboys & Duckham, 2004; Zhong Ren & Ming Hsiang, 2003), strongly requires methodological and operational tools for dealing with information distributed in multiple, autonomous and heterogeneous data sources, and a uniform data publishing methodology and policy over Internet Web sites. In this article we describe the experience of the Politecnico di Milano group in the activities of requirement analysis and conceptual design of the DEAFIN Web-GIS (Schreiber et al., 2003), whose objective is to provide a common environment for comparison of information about available vacant industrial sites coming from different regional data sources. Heterogeneity and Web availability requirements have been taken into account in the system architecture design; the system is thus conceived as a federated Web-based information system, apt to manage and provide access to all the regional relevant information in an integrated and complete fashion. Furthermore, since the data available by a given region partner can be both spatial and alphanumeric, a Web-GIS is defined for each regional component system.

BACKGROUND

The DEAFIN (development agencies and their impact on foreign direct investments) project has been launched with the purpose of allowing companies and investors to get a comprehensive information framework about areas located in European regions suited for potential investments. The aim is to make the regional data about possible investment areas homogenous and comparable, and internationally accessible. Potential investors need both a survey and a detailed view of vacant sites in different locations in order to compare different opportunities and decide their convenience. Quite naturally, such requirements call for a federated information system (FIS), which grants local sites a great deal of autonomy while enabling interoperation by means of a global integrated conceptual schema, that is, the federated data schema. Furthermore, owing to the capillarity of the end-user locations and to the need of a simple and widely known interface, Web-based access is mandatory. To define the functional specification of the system, the following activities have been carried out:

- *analysis of the requirements* of a distributed Web-based information system relying on a common conceptual database schema of the regional information that was initially (almost completely) available on paper support;
- *conceptual design* of the DEAFIN FIS, centered on the conceptual design of the federated conceptual database schema. The regional databases must be built according to the federated schema and then made accessible via the Web. The availability of data on the Web allows potential investors to navigate in the DEAFIN site according to various and customizable criteria, based on a benchmarking model developed within the project.

INFORMATION REQUIREMENTS

Three regional administrations from Germany, Italy, and Sweden were involved. The project started with a data-gathering phase, aimed at collecting requirements about data and processes managed at the partner Public Administrations. A questionnaire-based tool was circulated to collect common information to be stored in the FIS.

The basis of the questionnaire is a set of the data categories managed in Public Administration information systems. The relevant data categories concern land use plans (master and regional or specific), territorial services, industrial vacant sites, mobility data, statistical and social-economic data, base cartography data, and information on cadastral units data. Information on vacant industrial sites is the main focus of the investigation. For each category, the questionnaire collected the data characteristics reported in Table 1.

In general, the data collected at the sites show uniformity with respect to the attention paid to cartographic availability, regulations and laws about reuse of vacant areas, and centralization of resources. In particular, the need exists at each regional site to introduce tools able to treat heterogeneous data, since these data more and more intensively are to come from various data sources, to be mapped into the federated schema. Also, the degree of automation is similar, since cartographic systems and basic data management tools are available at the three partners' sites.

Several ongoing local projects concern the digital acquisition of land use plans, the automation of document management, and the development of various thematic databases and Web sites. What is required is a unified common schema for the regional databases. Moreover, the need of a uniform data publishing methodology and policy over Internet Web sites emerges clearly from the participants and involves almost all data categories.

Table 1. Summary of data characteristics collected in the data-gathering phase

Data Characteristics
Availability of the data category
Location of the data source
Support technology and name of products/tools
Data format
Data owner
User roles involved in data access
Restrictions applying to data access
Performances
Maintenance policies
Availability on the Web

USAGE MODES

The data-gathering phase has also detailed a set of user profiles, which specify how different access demands can be supported by the system towards a variety of user groups. The design and implementation of profiles have obviously a deep impact on the usability of the system. Hence, a careful analysis of user typologies and profiles has been performed during the specification phase, while an enrichment of profiles and access modes has been planned in a post-implementation follow-up. The purpose is to have the system start with a set of pre-configured access typologies, and then tune the access modes and user profiles against the most typical uses observed for a fixed period on the DEAFIN pilot implementation. The first broad difference is between the profile of public administrations and that of private users, due to different data needs. Moreover, two basic interaction modes must be provided: browsing (using thematic areas and other refined search parameters) and querying (using simple and user-friendly interfaces). The system can be regarded as: a passive tool, when responding to user questions; an active tool, when used as a decision support system or when used as a standard application, allowing new European partners to join the DEAFIN consortium.

As a passive tool, the contents of the federated database can show the advantages and disadvantages of an area: the information provided can be related to the specific search needs of a specific user. These searches are different according to user types, but apart from the function of locating (or re-locating) business activities, the motivation probably exists to retrieve general information about a region.

SYSTEM ARCHITECTURE

In the literature, several approaches and tools for handling heterogeneous data sources have been developed, and standards for distributed information systems have been defined (Mylopoulos & Papazoglou, 1997; Wiederhold, 1992). For these systems, the use of multiple layer, mediator-based architectures, and of a common data model have been employed (Garcia Molina et al., 1997). Wrapper/extractor and mediator tools (see Key Terms section) are proposed to obtain a uniform data representation (abstracting from the formats of the original data sources) and to facilitate federated access. Following this direction, in order to handle heterogeneity at each regional site, an architecture has been defined where extractors translate data coming from local heterogeneous data sources to a common reference format defined at the mediator level (Schreiber et al., 2003).

The aspect of data publishing over Internet Web sites has emerged clearly as a need from the partner regions. The system is thus conceived as a Web-based federated information system, apt to manage and provide access to all the regional relevant information in an integrated and complete fashion, thus satisfying the Web availability need.

Generally, data made available by a given partner region can be both spatial and alphanumeric, requiring each regional component system to be a GIS. Each regional site is in charge of data publishing and certification over the system according to the common federated schema. Each regional Web site contains a specific interface and stores regional data in a database by means of a Web-GIS technology, and operates as a central server that contains the common interface and some aggregate data. The global system is composed of Web-interconnected sites: the central Web site is the reference site for aggregate data and for requirements about new regional sites to be added in the federation or new applications to be developed in the component systems.

FEDERATED DATABASE CONCEPTUAL DESIGN

The federated database conceptual schema provides an integrated high-level representation of all data to be handled by the FIS. In order to achieve the highest readability and the widest possible diffusion, a common set of terms is needed. We adopted the dictionary of terms as defined by the EuroDicAutom (EuroDicAutom, 2001) of the EU; the international system of measurements, and commonly agreed upon data categories were adopted.

The database conceptual design produced a set of conceptual views, represented in terms of ER diagrams (see Key Terms section), that provide schematic requirement representations, each related to a specific information set. The following views were defined: vacant site description view, land use plan view, administration view, eco/reclamation view, transportation view, service view, document view, procedure and task view; the global conceptual data schema is a merge of these different views.

The design identified 36 entities (including entities with spatial representation that describes their shape, extension and location on the earth surface, marked as geo-referenced), with a total number of about 200 attributes. The XML (eXtensible Markup Language) language was chosen for logical specification, providing a common reference format to facilitate data customization, translation and mediation.

The FIS should enable the user to dynamically query the databases and in particular the geographical data they contain; these requirements, as we discussed previously, are fulfilled by a Web-GIS system able to perform advanced operations guided by the client side.

THE WEB-GIS

The interface of the central Web site allows the users to access data in two search modes: direct search, which leads directly to the vacant area of interest, including also information about the surroundings; and navigation, which presents a list of vacant sites to be filtered progressively, according to various criteria. Indeed, a static benchmarking table presents a list of site features, along with the corresponding values for each partner: from this table the user can choose a partner and obtain pages that describe the partner's area, highlighting its most attractive features. Some aggregate data about the vacant sites of each region are also displayed for subsequent browsing operations, where the user can execute queries in order to filter the vacant sites on the basis of preferences or specific criteria, such as: the type of usage, area size, costs, surrounding services, and accessibility. Finally, links to the regional Web sites are provided.

The Web interfaces of regional sites support more specific functions: the exploration of the partner area by choosing a specific theme such as land planning, business and market activities, mobility system, services, demographic and macroeconomic data and development strategies; the access to business information such as contact offices and particular business opportunities that are currently available for some vacant sites of the area; the access to aggregated data about vacant sites of the area and to a benchmarking table where the user can apply some selection conditions based on the available comparison parameters; the search of vacant sites based on forms that guide the user in the specification of the query, including geography related conditions (SQL with geographic extensions) like "search the vacant sites with a railway station within a radius of <parameter value> km"; the selection of specific features of the chosen vacant site, for example, buildings, eco quality, themes presenting the characteristics technological networks, land planning, mobility system, territorial services, procedures and tasks and documents.

FUTURE TRENDS

Web-GIS applications are becoming more and more important to a growing number of activities. Geography, geology, environmental studies, business marketing, and other disciplines have gained benefits from GIS tools and methods. Continual improvements in GIS hardware and software will lead to a much wider application of this technology throughout government, business, and industry. In particular, integrated geodata services based on data format standardization will increasingly facilitate the exchange of information among users of different systems by allowing data sharing and improving cooperation and communication among the organizations involved in environmental protection, planning, and resource management.

CONCLUSION

The development of the DEAFIN Federated Information System started from an analysis of the requirements and of the locally available information, performed through a data-gathering phase. The federated database was modeled in the form of a set of conceptual views, both as ER diagrams and XML specifications. The architecture of the overall system is in the form of a federated Web-GIS: at the main Web site a high-level view over the entire federated system is allowed, while local specializations are allowed at the level of every regional Web site adhering to the recommendations for the entire database.

The proposed system offers a marketing and information platform in a European context based on advanced Web functions, integrating geographic information systems (GIS), decision support tools, and database federation features. Analogous projects in the European area have been developed, mainly in the field of tourism (Pühretmair & Wöß, 2001), where integration of data sources and access to GIS in a graphical context are needed.

Further details about the project can be found in Schreiber et al. (2003).

REFERENCES

EuroDicAutom. (2001). Automatic translation of terms, European community. *http://eurodic.ip.lu:8086/cgi-bin/edicbin/EuroDicWWW.pl?page= expert*

Garcia Molina, H., Papakonstantinou, Y., Quass, D., Rajaraman, A., Sagiv, Y., Ullman, J., Vassolos, V., & Widom, J. (1997). The TSIMMIS approach to mediation: Data models and languages. *Journal of Intelligent Information Systems, 8*, 117-132.

Mylopoulos, J., & Papazoglou, M. (Eds.). (1997). Cooperative information system *IEEE Expert* [Special issue], *12*(5).

OpenGIS Consortium. (2001, February 20). *Geography Markup Language (GML)*. OGC Document Number: 01-029.

OpenGIS Consortium. (2001, June 21). *Web map service implementation specification.* OGC Document Number: 01-047r2.

Pühretmair, F., & Wöß, W. (2001). XML-based integration of GIS and heterogeneous tourism information. In K. Dittrich, A. Geppert & M. Norrie (Eds.), *Advanced information systems engineering* (pp. 346-358). Berlin: LNCS Springer Verlag.

Schreiber, F.A., Belussi, A., De Antonellis, V., Fugini, M.G., Pozzi, G., Tanca, L., & Tosi, L. (2003). The design of the DEAFIN Web-geographical information system: An experience in the integration of territorial reclamation support services. In A. Dahanayake & W. Gerhardt (Eds.), *Web-enabled systems integration: Practice and challenges* (pp. 142-168). Hershey, PA: Idea Group Publishing.

Wiederhold, G. (1992). Mediators in the architecture of future information systems. *IEEE Computer, 25,* 38-49.

Worboys, M.F., & Duckham, M. (2004). *GIS: A computing perspective* (2nd ed.). Boca Raton: CRC Press

The World-Wide Web Consortium. *http://www.w3.org*

Zhong Ren, P., & Ming Hsiang, T. (2003). *Internet GIS.* New York: John Wiley and Sons Inc.

KEY TERMS

Conceptual Schema of a Database: A semi-formal high-level description of the database, independent of its implementation.

ER (Entity-Relationship) Diagrams: The most widely used model to express the database conceptual schema.

eXtensible Markup Language (XML): Markup language proposed by the World Wide Web Consortium [W3C] for data and documents interchange.

Federated Information System (FIS): An information system is named federated when it supports interoperation among several autonomous and possibly heterogeneous information systems, by means of a shared global data schema.

Geographical Information System (GIS): Information system storing geographical data along with alphanumeric and spatial components. GIS systems also provide the data structures and algorithms to represent and efficiently query a collection of geographical data.

Mediator: A software component providing a uniform integrated interface to process and execute queries over data stored in multiple, heterogeneous data sources.

Web-GIS: A GIS system empowered with a Web-based interface.

Wrapper: A software tool to extract content from data sources and perform data format translation.

Actor-Network Theory and Adoption of E-Commerce in SMEs

Arthur Tatnall
Victoria University, Australia

Stephen Burgess
Victoria University, Australia

INTRODUCTION

Just because e-commerce technologies seems like useful tools that may assist a small to medium enterprise (SME) in doing its business better, it does not necessarily follow that these technologies will be *adopted* by this business. The implementation of an e-commerce system in an SME necessitates change in the way the business operates, and so should be considered as an innovation and studied using innovation theory.

Electronic commerce (e-commerce) is concerned with how computers, information systems and communications technologies can be used by people to improve the ways in which they do business. As e-commerce necessarily involves interactions of people and technology, any study of how it is used by a small business[1] must be considered in a socio-technical context. Although there is no universal consensus on what constitutes e-commerce, we believe that it contains elements of information systems, business processes and communications technologies. The complexity of studies in e-commerce is due, to a considerable degree, to the interconnected parts played by human actors and by the multitude of non-human entities involved. Small business managers, sales people, staff involved in procurement and warehouse operations, computers, software, Web browsers, Internet service providers (ISP), modems and Web portals are only some of the many heterogeneous components of an e-commerce system.

BACKGROUND

Adoption of E-Commerce by SMEs

In this article we will argue that the decision to adopt, or not to adopt a new technology, has more to do with the interactions and associations of both human and non-human actors involved in the project than with the characteristics of the technology. Information systems are complex socio-technical entities and research into their implementation needs to take account of this complexity, which will only be seen if it is reported in all its "messy reality" (Hughes, 1983). Research into the implementation and operation of these systems needs to take this heterogeneity into account and to find a way to give due regard to both their human and non-human aspects.

One view of the adoption of an electronic commerce innovation by a small business suggests that decisions are made primarily based on their perceptions of the characteristics of the technology concerned. Innovation diffusion (Rogers, 1995) uses this approach, and is based on the following elements: characteristics of the innovation itself, the nature of the communications channels, the passage of time, and the social system. Using this sort of approach, the researcher would probably begin by looking for characteristics of the specific e-commerce technology to be adopted, and the advantages and problems associated with its use. The next step would be to suggest that the adoption, or rejection, of this technology by the small business was due largely to these characteristics. We contend that while there may be some validity in such an approach, it is unlikely to provide the complete explanation, as it would miss other influences due to interpersonal and inter-business interactions, and to the backgrounds of the people involved.

Innovation Translation

We argue that actor-network theory (ANT) has much to offer in a situation like this. A researcher using an actor-network approach to study innovation would concentrate on issues of network formation, investigating the human and non-human actors and the alliances and networks they build up. They would investigate how the strength of these alliances may have enticed the small business to make the adoption or, on the other hand, to have deterred them from doing so (Tatnall, 2000, 2002; Tatnall & Gilding, 1999). While some research approaches to technological innovation treat the social and the technical in entirely different ways, actor-network theory proposes instead a socio-technical account in which neither social nor technical positions are privileged.

Actor-network theory argues that interactions between actors are heterogeneous and denies that purely technical or purely social relations are possible. It considers the world to be full of hybrid entities (Latour, 1993) containing both human and non-human elements. Change, in the ANT view, results from decisions made by actors, and involves the exercise of power. Latour (1986) argues that the mere possession of power by an actor does not automatically lead to change unless other actors can also be *persuaded* to perform the appropriate actions for this to occur.

In our experience it is often the case that when a small business is considering a technological innovation it is interested in *only some aspects* of this innovation and not others (Tatnall, 2002; Tatnall & Burgess, 2002). In actor-network terms it needs to *translate* (Callon, 1986) this piece of technology into a form where it can be adopted, which may mean choosing some elements of the technology and leaving out others. What results is that the innovation finally adopted is not the innovation in its original form, but a translation of it into a form that is suitable for use by the recipient small business (Tatnall, 2002).

In many instances a small business proprietor will adopt e-commerce because a friend is using it, or because they know a competitor is using it, or because a son or daughter learned about it at school (Burgess, 2002; Tatnall, 2002). The nature and size of each small business, the inter-business interactions in which they engage, and the backgrounds and interests of particular individuals in each are also likely to have had an important effect that would, most likely, have been ignored by the essentialist approach offered by innovation diffusion. Actor-network theory, in examining alliances and networks of human and non-human actors, provides a good foundation from which small business adoption and use of e-commerce can be researched. The ANT approach will be further amplified in the case studies that follow, particularly in respect of the identification of actors and networks.

FUTURE TRENDS

Actor-Network Theory and E-Commerce Innovation

The theory of innovation diffusion (Rogers, 1995) is well established and has been used as the framework of many studies. In most cases, however, the success of the diffusion model has been in explanation of innovation "in the large," when the statistical effects of big numbers of organisations and individuals involved come into play. It has, typically, been less successful in explaining how particular individuals or specific organisations make their adoption decisions, and it is in situations like this that an innovation translation approach, using actor-network theory, is especially useful.

In offering a socio-technical approach to theorising innovation, ANT provides a particularly useful tool to the study of innovations in which people and machines are intimately involved with each other. The adoption of e-commerce technologies certainly involves a consideration of the technologies themselves, but also of business organisations, business processes, and the needs and likes of individual humans. ANT, we suggest, is especially useful in researching innovations like these, and in particular, when considering individual adoption decisions.

The main use made of any research approach such as ANT is in the study of past events, and ANT makes no claim to be able to predict what may happen in the future. We suggest, however, that ANT analysis can identify some pointers towards the successful introduction of an innovation, and the change management associated with this. ANT argues that it is not the characteristics of either the innovation itself or the potential adopter acting alone that are important, but rather the interactions of many actors. The key to successful change management, it would thus seem, involves allowing for these interactions and for the socio-technical nature of the process.

Case Studies

This article now offers several brief case studies in which actor-network theory has provided a means by which adoption (or non-adoption) of technology can be explained. In each case, data for the study were obtained through semi-structured interviews with the proprietors and personnel of the businesses involved, between September 2001 and August 2002.

1. *Adoption of a Portal by a Storage and Transport Company*

The business to be considered in this study is a medium-sized Melbourne company, with about 50 employees, that stores frozen food and transports it to supermarkets and other locations around the country. It became clear from the study that the transport company had "not really been into computers," and had only recently started coming to grips with this technology.

Although the manager had some idea of the benefits to his company of using the portal, he had no clear plan for using it. It was just "a really good idea". The reasons he adopted this innovation thus had little to do with the characteristics of this technology, and much more to do with his involvement with the local business community and because of his belief that the portal had the potential to improve business in the region.

2. *Adoption of Electronic Commerce by a Small Publishing Company*

The second case concerns a small publishing company where four people work on the production of textbooks and research publications. The company is a very small business with a relatively low turnover but a well-established market. The business has adopted some e-commerce technologies, but not others. Some time ago, it registered a domain name and set up its own Web page, but only for informational purposes. The site shows details of the company's products and indicates how orders can be placed, but does not support sales or other business-consumer (B-C) transactions.

When asked why the company had not considered instituting a B-B e-commerce system with the printer, the director replied that they would like to do so, but that the printer was not really interested or geared up to get into this type of order. Adoption decisions of the company thus had more to do with the technological status of their suppliers and their customers than with the e-commerce technology itself. The company was interested in using those e-commerce technologies that its business partners used and those it considered useful or timesaving. What was adopted could be seen as a translation of an e-commerce innovation to include just these things.

3. *Non-Adoption of E-Commerce by a Small Chartered Accountancy Firm*

The last case is of a small chartered accountancy firm which is a family business in the western suburbs of Melbourne. Employees of the business are the main accountant, who has a degree in accounting and is a Certified Practicing Accountant (CPA); the main accountant's father, who previously ran the business but, as he is not qualified, is limited these days mainly to supporting the taxation side of the business; another accountant (CPA); and a full-time secretary. The firm offers services that are typical of a small accounting business and its clients include both individuals and small businesses.

For some time members of the business had been debating whether to set up a Web site. The decision seemed to come down to two major opposing viewpoints. The first, held by the main accountant, was that a Web site was "the way of the future," that customers would expect it and that some competitors already had one. The opposite viewpoint, held by the father, was that it is a waste of time and money, that there was nothing you could put on a Web site that customers would want anyway, and "who is going to do it?" Other members of the business seemed quite apathetic about the whole matter.

The upshot was that the Web site could not be "translated" into a form where it could be adopted by this business. The main accountant was the prime motivator in attempting to define the problem and to discover the solution: he did not do this well enough, however, to gain acceptance of the innovation.

CONCLUSION

All of the situations described point to the fact that decisions on the adoption of electronic commerce technologies are often made on the basis of more than just the characteristics of the technology, and that in many cases these characteristics are not especially significant in the decision making process.

In each of these instances, the actor-network approach therefore offers a useful explanation of why a particular e-commerce initiative was or was not adopted. On the other hand, an innovation diffusion approach to investigating each of the potential adoptions would have looked for explanations for the uptake, or lack of uptake, primarily in the characteristics and properties of the technology in each case. It would not have considered, as particularly important, the human and non-human interactions described here. In our view, the decision to adopt, or not to adopt, has more to do with the interactions and associations of both human and non-human actors involved in the project than with characteristics of the technology.

REFERENCES

Burgess, S. (2002). Information technology in small business: Issues and challenges. In S. Burgess (Ed.), *Information technology and small business: Issues and challenges*. Hershey, PA: Idea Group Publishing.

Callon, M. (1986). Some elements of a sociology of translation: Domestication of the scallops and the fishermen of St Brieuc Bay. In J. Law (Ed.), *Power, action & belief. A new sociology of knowledge?* (pp. 196-229). London: Routledge & Kegan Paul.

Hughes, T.P. (1983). *Networks of power: Electrification in Western society, 1880-1930*. Baltimore: Johns Hopkins University Press.

Latour, B. (1986). The powers of association. In J. Law (Ed.), *Power, action and belief. A new sociology of* *knowledge? Sociological Review monograph 32* (pp. 264-280). London: Routledge & Kegan Paul.

Latour, B. (1993). *We have never been modern*. Hemel Hempstead, Harvester Wheatsheaf.

Rogers, E.M. (1995). *Diffusion of innovations*. New York: The Free Press.

Tatnall, A. (2000). *Innovation and change in the information systems curriculum of an Australian university: A socio-technical perspective.* PhD thesis. Education. Rockhampton, Central Queensland University.

Tatnall, A. (2002). Modelling technological change in small business: Two approaches to theorising innovation. In S. Burgess (Ed.), *Managing information technology in small business: Challenges and solutions* (pp. 83-97). Hershey, PA: Idea Group Publishing.

Tatnall, A., & Burgess, S. (2002). Using actor-network theory to research the implementation of a B-B portal for regional SMEs in Melbourne, Australia. *15th Bled Electronic Commerce Conference - eReality: Constructing the eEconomy*, Bled, Slovenia, University of Maribor.

Tatnall, A., & Gilding, A. (1999). *Actor-network theory and information systems research*. 10th Australasian Conference on Information Systems (ACIS), Wellington, Victoria University of Wellington.

KEY TERMS

Actor: An entity that can make its presence individually felt by other actors. Actors can be human or non-human, non-human actors including such things as computer programs, portals, companies and other entities that cannot be seen as individual people. An actor can be seen as an association of heterogeneous elements that constitute a network. This is especially important with non-human actors as there are always some human aspects within the network.

Actor-Network Theory (ANT): An approach to research in which networks associations and interactions between actors (both human and non-human) are the basis for investigation.

E-Commerce Studies: Contain elements of information systems, business processes and communications technologies.

Innovation Diffusion: A theory of innovation in which the main elements are: characteristics of the innovation itself, the nature of the communication channels, the passage of time, and the social system through which the innovation diffuses.

Innovation Translation: A theory of innovation in which, instead of using an innovation in the form it is proposed, potential adopters *translate* into a form that suits their needs.

Small to Medium Enterprise (SME): For the purpose of this article SMEs are considered to be those businesses that have from 1-20 employees – small, and 21-50 employees – medium.

Technological Innovation: The introduction or alteration of some form of technology (often information technology) into an organisation.

Technology Adoption: The decision, by an organisation or individual, to utilise and implement a technology.

ENDNOTES

1 For the purpose of this article SMEs are considered to be those businesses that have from 1-20 employees – small, and 21-50 employees – medium. Burgess, S. (2002). Information technology in small business: Issues and challenges. *Information technology and small business: Issues and challenges*. Hershey, PA: Idea Group Publishing.

Actor-Network Theory in Information Systems Research

Arthur Tatnall
Victoria University, Australia

INTRODUCTION

Building an information system is a difficult task, partly due to the problem of ascertaining the requirements of the intended users, but also because of the complexity of the large number of human-machine interactions (Banville, 1991). This complexity is reflected in the difficulty of building these systems to operate free from error and to perform as intended. The dictionary defines innovation as "the alteration of what is established; something newly introduced". As the introduction or improvement of an information system in an organisation *necessarily* involves change, information systems research often involves research into technological innovation.

BACKGROUND

Information Systems as a Socio-Technical Discipline

The discipline of information systems (IS) is concerned with the ways people build and use computer-based systems to produce useful information and so has to deal with issues involving both people and machines; with the multitude of human and non-human entities that comprise an information system (Checkland, 1981). Information systems is neither merely a technical discipline nor a social one, but one that is truly socio-technical. Researchers in information systems face the problem of how to handle complexities due to interconnected combinations of computers, peripherals, procedures, operating systems, programming languages, software, data and many other inanimate objects; how they all relate to humans and human organisations, and how humans relate to them (Longenecker, Feinstein, Couger, David & Gorgone, 1994).

This article will outline a socio-technical approach, based on actor-network theory (ANT), to researching how people interact with and use information systems (Tatnall & Gilding, 1999). In actor-network theory the key is in using an approach that is neither purely social nor purely technical, but socio-technical.

Qualitative Research Traditions in Information Systems

Each field of academic inquiry is characterised by its own preferred and commonly used research approaches and traditions. In information systems research, Myers (1997) outlines four qualitative traditions as being particularly significant: case study research, ethnography, grounded theory and action research.

Case study research is the most commonly used qualitative approach in information systems. As IS research topics commonly involve the study of organisational systems, a case study approach is often appropriate. Ethnography has grown in prominence as a suitable approach to information systems research after work such as that undertaken by Suchman (1987) and Zuboff (1988). It has been used especially in research where the emphasis is upon design, computer-supported cooperative work, studies of Internet and virtual communities, and information-related policies (Star, 1995). Grounded theory is an "an inductive, theory discovery methodology" (Martin & Turner, 1986) that seeks to develop theory that is grounded in data that are systematically gathered and analysed and involves "continuous interplay" between data and analysis (Myers, 1997). Orlikowski (1993) argues that in information systems research situations involving organisational change, a grounded theory approach can be useful as it allows a focus on "contextual and processual" elements as well as on the actions of key players.

Action research has been described as proceeding in a spiral of steps where each step consists of planning, action and evaluation of the result of the action. It is seen as aiming "... to contribute both to the practical concerns of people in an immediate problematic situation and to the goals of social science by joint collaboration within a mutually acceptable ethical framework" (Rapoport, 1970, p. 499). A variant of action research that is slowly gaining acceptance in information systems is soft systems methodology (SSM), developed by Peter Checkland (1991). SSM attempts to give due recognition to both the human and technological aspects of a system. It acknowledges both human and non-human aspects of IS, but considers these to be entirely separate types of entities.

FUTURE TRENDS

ANT and Socio-Technical Research

Actor-network theory considers both social and technical determinism to be flawed and proposes instead a socio-technical account (Latour, 1996) in which nothing is purely social and nothing is purely technical (Law, 1991). ANT deals with the social-technical divide by denying that purely technical or purely social relations are possible.

To see better how this works, suppose that an IS researcher was investigating the uptake of a business-to-business e-commerce portal developed by a local government authority for use within a regional area, with an Internet service provider (ISP) and a software company engaged to build the portal, and a bank to provide a payment gateway. ANT asserts that the world is full of hybrid entities (Latour, 1991) containing both human and non-human elements and offers the notion of heterogeneity to describe projects such as this. The project will involve not just these entities, but also non-human entities such as computers, computer programs, data storage devices, modems and telephone lines, and human entities including local business proprietors from small and large businesses, customers, programmers and local council staff. The utilisation of heterogeneous entities (Bijker, Hughes & Pinch, 1987) then avoids questions of: "is it social?" or "is it technical?" as missing the point, which should be: "is this association stronger or weaker than that one?" (Latour, 1991).

Information systems researchers using an ANT approach would concentrate on issues of network formation, investigating the human and non-human alliances and networks built up by the actors involved. They would concentrate on the negotiations that allow the network to be configured by the enrolment of both human and non-human allies. Interactions and associations between actors and networks are all important, and actors are seen simply as the sum of their interactions with other actors and networks.

In the case of the portal an actor-network researcher would begin by identifying some of the important actors, starting perhaps with the local government portal project manager. An interview with the project manager would reveal why the project was instigated and identify some of the other actors. The main advice on method suggested by the proponents of actor-network theory is to "follow the actors" (Latour, 1996) and let them set the framework and limits of the study themselves, and one line of inquiry resulting from the interview with the project manager might be to approach the portal software designer and programmers. Another set of actors is the proprietors of the local businesses themselves, and the project manager may suggest some "business champions" to interview first. At least some of these business people might then point to the influence exerted by the computer hardware or software as a significant factor, so identifying some non-human actors. Negotiations between actors must be carefully investigated. Apart from the obvious human-to-human kind of negotiation, also included must be human-to-non-human interactions such as the business people trying to work out how the portal operates, and how to adapt this technology to be most suitable for their own business purposes. The process of adopting and implementing the portal can now be seen as the complex set of interactions that it is, and not just the inevitable result of the innate characteristics of this technology.

How Actor-Network Theory Handles Complexity

Longenecker et al. (1994) suggest that computer-based information systems should be regarded as complex socio-technical entities, begging the question of how this complexity should be handled. A common method of handling complexity in all subject areas lies in simplification, but the danger with simplification is that it runs the risk of removing just those things that constitute a useful description of the phenomenon under investigation by concealing the parts played by many of the actors (Suchman, 1987). The question here is which details to include and which to leave out, and who is to decide. In this respect, an appropriate research approach needs to ensure that complexities are not lost "in the process of labelling" (Law, 1991).

In actor-network theory the extent of a network is determined by actors that are able to make their presence *individually felt* by other actors. The definition of an actor requires this and means that, in practice, actors limit their associations to affect only a relatively small number of entities whose attributes are well defined within the network. An actor is not just a point object but an association of heterogeneous elements, themselves constituting a network. An actor can, however, in many ways also be considered as a black box (Callon, 1986), and when we open the lid of the box to look inside it will be seen to constitute a whole network of other, perhaps complex, associations. In many cases details of what constitutes an actor - details of its network - are a complication we can avoid having to deal with all the time.

When investigating the e-commerce portal it might be convenient, most of the time, to consider both the ISP and the portal software to constitute a black box. This would mean that this aspect of the technology could then be considered as just a single actor; the portal, and its interactions with other actors investigated on this basis. At other times it might be necessary to lift the lid of the black box and investigate the enclosed network of the ISP, telephone lines, computers, data storage, programmers, and interface designers it contains. The advantage of black-boxing though is that most of the time, however, the portal can be regarded as just another actor. The important thing to note about the use of black-boxing for simplification is that the complexity is not just put into the black box and lost, as it is always possible, and indeed necessary, to periodically reopen the black box to investigate its contents.

The portal black box could be considered as shown in Figure 1. This black box itself also contains several other black boxes (portal software, portal hardware and payment gateway) that we do not need to consider in detail until such consideration becomes necessary. Figure 1 also shows some of the other main actors that interact with the portal. Until (and possibly after) the portal is fully operational, of course, these actors will also interact directly with each other.

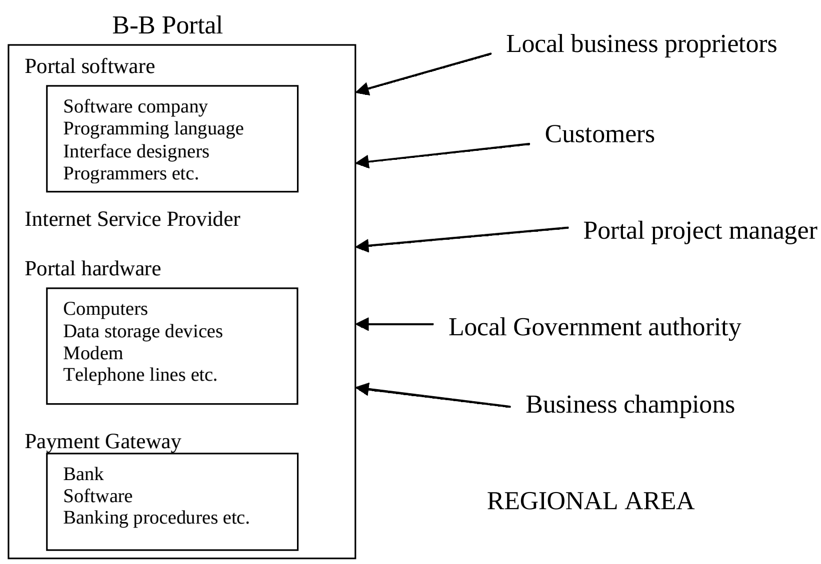
Limitations and Criticisms of Actor-Network Theory

There are several main criticisms of actor-network theory. To begin, there is the criticism by Grint and Woolgar (1997) that it is not always sufficiently clear where the boundaries of a network lie or whose account of a network is to be taken as definitive. They note that the analyst's story seems to depend on a description of the "actual" network, as if this was objectively available.

A second criticism relates to ANT's treatment of non-human actors. A critique by Collins and Yearley (1992) claims that in giving an autonomous voice to "things," ANT concedes too much to realist and technical accounts. In reply, Callon and Latour (1992) claim that technological artefacts are implicated in the very fabric of the social and are "social relations viewed in their durability and cohesion".

Thirdly, Grint and Woolgar (1997) argue that ANT retains a degree of residual technicism in its need to sometimes refer to "actual" technical capacities of a technology. They quote Callon's (1986) analysis of the attempts at building a French electric car, in which they claim that he makes reference to the "unfortunate tendency" of the catalysts to become quickly contaminated. They note that the anti-essentialist approach of actor-network theory would point to this "actual property" being treated as a construction. Despite these minor reservations, however, Grint and Woolgar note that actor-network theory points to the possibility of an understanding of technology that does not rely on the presence of a "god within the machine".

Figure 1. Actors, interactions and the portal black box



CONCLUSION

In this article I have argued that information systems is a socio-technical discipline involving both human and non-human entities, and that information systems implementations are complex activities inevitably involving some form of technological innovation. I have also argued that simplistic views of how information systems are built, implemented and used often conceal important interactions between human and non-human actors and so give a less than complete picture of what has happened.

An actor-network approach avoids the need to consider the social and the technical, and thus human and non-human actors, in different ways. Highlighting how the human and non-human actors involved in socio-technical situations, such as the building and use of information systems, interact with each other is an important benefit of adopting an ANT research framework. In showing how these interactions may lead to the formations of stable networks, actor-network theory offers a useful way to handle the complexity of such studies.

REFERENCES

Banville, C. (1991). A study of legitimacy as a social dimension of organizational information systems. In H.E. Nissen, H.K. Klein & R. Hirschheim (Eds.), *Information systems research: Contemporary approaches and emergent traditions* (pp. 107-129). Amsterdam: Elsevier Science Publications.

Bijker, W.E., Hughes, T.P., & Pinch, T.J. (Eds.). (1987). *The social construction of technological systems: New directions in the sociology and history of technology*. Cambridge, MA: MIT Press.

Callon, M. (1986). The sociology of an actor-network: The case of the electric vehicle. In M. Callon, J. Law & A. Rip (Eds.), *Mapping the dynamics of science and technology* (pp. 19-34). London: Macmillan Press.

Callon, M., & Latour, B. (1992). Don't throw the baby out with the bath school: A reply to Collins and Yearley. In A. Pickering (Ed.), *Science as practice and culture* (pp. 343-368). Chicago University Press.

Checkland, P. (1981). *Systems thinking, systems practice*. Chichester: Wiley.

Checkland, P., & Scholes, J. (1991). *Soft systems methodology in action*. Chichester: Wiley.

Collins, H.M., & Yearley, S. (1992). Epistemological chicken. In A. Pickering (Ed.), *Science as practice and culture* (pp. 301-326). Chicago University Press.

Grint, K., & Woolgar, S. (1997). *The machine at work - technology, work and organisation*. Cambridge: Polity Press.

Latour, B. (1991). Technology is society made durable. In J. Law (Ed.), *A sociology of monsters. Essays on power, technology and domination* (pp. 103-131). London: Routledge.

Latour, B. (1996). *Aramis or the love of technology*. Cambridge, MA: Harvard University Press.

Law, J. (Ed.). (1991). *A sociology of monsters. Essays on power, technology and domination*. London: Routledge.

Longenecker, H.E.J., Feinstein, D.L., Couger, J.D., David, G.G., & Gorgone, J.T. (1994, Winter). Information Systems '95: A summary of the collaborative IS curriculum specification of the Joint DPMA, ACM, AIS Task Force. *Journal of Information Systems Education,* 174-186.

Martin, P.Y., & Turner, B.A. (1986). Grounded theory and organizational research. *The Journal of Applied Behavioral Science, 22*(2), 141-157.

Myers, M.D. (1997, May 20). Qualitative research in information systems. *MIS Quarterly*. http://misq.org/misqd961/isworld/

Orlikowski, W.J. (1993). CASE tools as organizational change: Investigating incremental and radical changes in systems development. *Management Information Systems Quarterly, 17*(3), 1-28.

Rapoport, R.N. (1970). Three dilemmas in action research. *Human Relations, 23*(4), 449-513.

Star, S.L. (1995). *The cultures of computing*. Oxford: Blackwell Publishers.

Suchman, L.A. (1987). *Plans and situated actions. The problem of human-machine communication*. Cambridge University Press.

Tatnall, A., & Gilding, A. (1999). Actor-network theory and information systems research. *10th Australasian Conference on Information Systems (ACIS)*, Wellington, Victoria University of Wellington.

Zuboff, S. (1988). *In the age of the smart machine*. New York: Basic Books.

KEY TERMS

Actor: An entity that can make its presence individually felt by other actors. Actors can be human or non-human, non-human actors including such things as computer programs, portals, companies and other entities that cannot be seen as individual people. An actor can be seen as an association of heterogeneous elements that constitute a network. This is especially important with non-human actors, as there are always some human aspects within the network.

Actor-Network Theory (ANT): An approach to research in which networks associations and interactions between actors (both human and non-human) are the basis for investigation.

Black Boxing: A technique used for simplification. Multiple actors can be put into a black box so that it is not necessary to look at them in detail. The portal mentioned in this article could be considered as a black box containing the ISP, portal software, data storage devices, modems, telephone devices and so on. Black-boxing is done for convenience, as it means that an entity can then be seen as just another actor, and saves looking at the detail until necessary. The black box can later be reopened to investigate its contents.

Socio-Technical Research: Involving both social and technical interactions, occurring is such a way that it is not easily possible to disentangle them.

Technological Innovation: The introduction or alteration of some form of technology (often information technology) into an organisation.

A

Adaptive Mobile Applications

Thomas Kunz
Carleton University, Canada

Abdulbaset Gaddah
Carleton University, Canada

INTRODUCTION

The convergence of two technological developments has made mobile computing a reality. In the last few years, developed countries spent large amounts of money to install and deploy wireless communication facilities. Originally aimed at telephone services (which still account for the majority of usage), the same infrastructure is increasingly used to transfer data. In parallel, wireless LAN technologies are providing hotspot coverage in many high-traffic locations. The second development is the continuing reduction in size of computer hardware, leading to portable computation devices such as laptops, palmtops, or functionally enhanced cell phones. Given current technology, a user can run a set of applications on a portable device and communicate over a variety of communication links, depending on his/her current location.

As will be explained in more detail later on, the mobile computing environment is highly dynamic. Available bandwidth changes by orders of magnitudes, based on the selected wireless access technology. Also, portable devices differ in processing power, memory, display capabilities, and other characteristics. It is generally argued that applications should "adapt" to the current environment, for example by filtering and compressing data or by changing the functionality offered to the user. Some researchers even argue that all future applications, not just the ones intended for execution on mobile devices, will have to be able to adapt to changing requirements and changing implementation environments on time scales from microseconds to years (Kavi, 1999). This article reviews the work on adaptive mobile applications and provides an outlook on future trends.

The alternative to adaptive applications is to either implement a single application that is designed for the lowest common denominator (in terms of resource availability) or multiple functionally identical or similar binaries, tuned for specific environments. The former will needlessly sacrifice application features when running in more resource-rich environments. The latter approach is an inferior solution as well, for a number of reasons. The user of a portable device has to install and maintain multiple applications, which is a drain on the limited storage capabilities typically found on those devices. It also potentially results in different user interfaces and causes high software development overheads when developing the "same" mobile application multiple times. Finally, it forces the user to identify the current execution conditions and select the "right" application.

The next section will review the motivation for adaptive approaches towards mobile application design. We will then briefly review traditional approaches to adaptive mobile applications, followed by a discussion of mobile middleware that is intended to support adaptive mobile applications. The article finishes with a brief conclusion of the state-of-the-art and identifies areas of future work.

BACKGROUND

Wireless communication and portable devices make it possible for mobile users to have access to information anywhere and anytime. Designing, implementing and deploying applications that work well across all portable devices and across a wide range of wireless access networks is non-trivial.

There are at least three common factors that affect the design of mobile applications: portable devices, network connection, and mobility. *Portable devices* vary from one to another in term of resource availability. Devices like laptops can offer fast CPUs and large amount of RAM and disk space while others like pocket PCs and phones usually have scarce resources. It is either impossible or too expensive to augment the resource availability. Hence, applications should be designed to achieve optimal resource utilization. In general, the design of portable devices strives for properties such as size, weight, durability and long battery life. Different devices will emphasize different trade-offs between CPU speed, memory, I/O capabilities, and power consumption, providing very heterogeneous execution environments for mobile applications.

Network connection in mobile scenarios is characterized by limited bandwidth, high error rate, higher cost, and frequent disconnections due to power limitations, available spectrum, and mobility. Wireless communication is more difficult to achieve than wired communication because the surrounding environment interacts with the signal, blocking signal paths and introducing noise and echoes. Therefore, mobile application designs need to be more concerned about the network conditions than applications designed for fixed networks. Many wireless and mobile networks such as WaveLAN are organized into geographically defined cells, with a control point called a base station in each of the cells. Devices within the same cell share the network bandwidth; hence, the bandwidth rapidly decreases whenever a new device joins the cell. Portable devices may move around different areas with no coverage or high interference that cause a sudden drop in network bandwidth or a loss of connection entirely. Unpredictable disconnection is also a common issue that frequently occurs due to the handoff process or shadowed areas. Most wireless network services charge a flat fee for their service, which usually covers a fixed number of messages. Additional charges are levied on per packet or per message basis. In contrast, the cost for sending data over cellular networks is based on connection time instead. This forces mobile users to connect for short periods of time.

Physical device mobility can greatly affect network connection, which accordingly has to adapt to user mobility by reconnecting the user with respect to a new location. Portable devices may interact with different types of networks, services, and security policies as they move from one area to another. This requires applications to behave differently to cope with dynamic changes of the environment parameters. As a consequence, mobile applications also have to cope with a much greater variation in network bandwidth: bandwidth can shift from one to six orders of magnitude between being plugged into the wired network versus using (different) wireless access networks.

The constraints and limitations mobile applications face are not a product of current technology, but they are related naturally to mobility. Together, they complicate the design of mobile applications and require rethinking traditional approaches to application design. Any feasible approach to mobile computing must strike a balance between conflicting demands. This balance cannot be static, as the execution environment of mobile applications varies; it must react, or in other words, the applications must be adaptive.

ADAPTIVE MOBILE APPLICATIONS: TRADITIONAL APPROACHES

Designing adaptive applications is an active research area. Traditionally, most work focused on the wireless link(s). Early work provides general solutions that do not change the TCP semantics but focus on improving TCP throughput over wireless links; see for example Balakrishnan (1995). While this addresses issues such as high link error rates and spurious disconnections, it does not address the low and highly variable bandwidth characteristic of mobile computing.

A second group of approaches adapts to the scarce and varying wireless link bandwidth by filtering and compressing the data stream between a client application on a portable device and a server executing on a stationary host. Data compression is done at one of two places. Bolliger (1998) and Seshan (1997) enhance the server to generate a data stream that is suited for the currently available bandwidth. This typically represents an end-to-end approach, which is well known in the networking and system literature. Most other proposals (Angin, 1998; Fox, 1998) extend the client-server structure to a client-proxy-server structure, where a proxy executes in the wireless access network, close to the portable device. This proxy-based approach filters and compresses the data stream originating from the server to suit the current wireless bandwidth. Joshi (1997) incorporates both end-to-end and proxy-based approaches, using each as appropriate, to support Web access from mobile platforms. For example, tasks such as complex filtration, which require significant computational resources, are done in an end-to-end manner. The proxy-based approach, on the other hand, is used when the server is not able to generate the appropriate data stream.

A third, complementary approach, focuses on the computational effort (Kunz, 2000). Mobile applications, especially ones that require intensive computation (for example, video decoding), can be divided dynamically between the wired network and the portable device according to the mobile environment and to the availability of the resources on the portable device, the wireless link, and the wired network. The access network supports the mobile application by providing proxy servers that can execute parts of the application code. This may increase the performance of applications and reduce the power consumption on portable devices since offloading computation to the proxies in the wired network will reduce the CPU cycles and memory needed to achieve certain tasks at portable devices.

FUTURE TRENDS: MOBILE MIDDLEWARE

The early approaches reviewed in the previous section typically provide toolkits that support specific adaptation ideas. To generalize this effort, support for adaptive mobile applications should be embedded into appropriate mobile middleware. Traditional middleware systems, like CORBA and DCOM, have achieved great success in dealing with the heterogeneity in the underlying hardware and software platforms, offering portability, and facilitating development of distributed applications. However, these systems are based on the assumptions that the distributed applications will run in a static environment; hence, they fail to provide the appropriate support for mobile applications. Therefore, mobile applications need a middleware that facilitates adapting to environment variations.

Based on the mobile computing challenges reviewed previously, mobile middleware should meet the following requirements: *Asynchronous interaction* tackles the problems of high latency and disconnected operations. It allows mobile clients to issue a request for a service, disconnect from the network, and collect the result later on. This type of interaction model reduces bandwidth consumption, achieves decoupling of client and server, and elevates system scalability. *Reconfigurability* is the process of adding a new behavior or changing an existing one during system runtime. Dynamic reconfiguration of system behavior and operating context at runtime may require reevaluation and reallocation of resources. Reconfiguration could be based on context information. *Adaptivity* allows applications to modify their behavior instead of providing a uniform interface in all situations. The middleware needs to monitor the resource supply/demand, compute adaptation decisions, and notify applications about changes. *Context-awareness* of client capabilities, changes to network conditions, and the ability to change the behavior of the system as circumstances warrant are required to build an effective and efficient adaptive system. The context includes device characteristics, user's location, user's activities, and other resources. The system performance can be increased when information context is disclosed to the application to assist middleware in making the right decision. Finally, *lightweight load* should be considered when constructing middleware for mobile computing. Middleware should not increase the computational load on the most power-consuming components such as processor, network interface, and so forth. Middleware implementations often include a number of unused features that can be entirely omitted to reduce the computational load.

We identified four categories of mobile middleware: reflective, tuple space, context-aware, and event-based middleware. *Reflective middleware* like DynamicTAO (Kon, 2000) and Open-ORB (Blair, 2001) are built around the concept of component frameworks (CF). Components can be developed independently, distributed in binary form, and combined at run time. Reflection provides a *meta-interface* to inspect the internal behavior of the middleware and, if it is necessary, alter its behavior to better match the system's current operating environment. The main motivation of this approach is to make the middleware more adaptable to its environment and better able to cope with changes. Open problems are consistent dynamic reconfiguration and performance. There is some early work in this area that has focused on developing reconfiguration models and algorithms that enforce well-defined consistency rules while minimizing system disturbance (Kramer & Magee, 1990). In addition, all reflective systems impose a heavy computational load that causes significant performance degradation on portable devices.

Tuple-space systems such as LIME (Picco, 1999) and TSpaces (Wyckoff, 1998) exploit the decoupled nature of tuple spaces for supporting disconnected operations in a natural manner. A tuple space is a globally shared, associatively addressed memory space that is used by processes to communicate. Client processes create tuples and place them in the tuple space using a *write* operation. Also, they can concurrently access tuples using *read* or *take* operations. This communication paradigm fits well in a mobile setting where logical and physical mobility is involved. By default they offer an asynchronous interaction paradigm that appears to be more appropriate for dealing with intermittent connection of mobile devices, as is often the case when a server is not in reach or a mobile client requires to voluntary disconnect to save battery and bandwidth. Using a tuple-space approach, we can decouple the client and server components in time and space. In other words, they do not need to be connected at the same time and in the same place. Tuple-space systems support the concept of a space or spaces that offer the ability to join objects into appropriate spaces for ease of access. This opens up the possibility of constructing a dynamic super space environment to allow participating spaces to join or leave at arbitrary times. The ability to use multiple spaces will elevate the overall throughput of the system. One problem with tuple-space middleware systems is their excessive memory requirements, making them impractical for most portable devices available to-date.

Context-aware systems provide mobile applications with the necessary knowledge about the execution context in order to allow applications to adapt to dynamic changes in mobile host and network condition. The execution context includes but is not limited to: mobile user location, mobile device characteristics, network condition, and user activity (i.e., driving or sitting in a room). However, most context-aware applications are only focusing on a user's location. Nexus (Fritsch, 2000), for example, is designed to be a generic platform for location-aware applications. Reflective middleware may also improve the development of context-aware services and applications. For example, Capra (2003) has suggested the use of metadata and reflection to support context-aware applications. However, overall, limited attention has been given to contexts other than location. It is necessary to take into account other types of context awareness such as internal resources (i.e., memory size, battery and processor power) or external resources (i.e., network bandwidth and connectivity quality). More efforts need to be directed towards an easy context representation and simple interfaces that enable the applications to interact with the underlying middleware.

In *event-based systems*, clients first announce their interest in receiving specific events and then servers broadcast events to all interested clients. Hence, the event-based model achieves a highly decoupled system and many-to-many interaction style between clients and servers. Examples are JEDI (Cugalo, 2001) and STEAM (Meier, 2002). Most existing systems do not combine traditional middleware functionality (i.e., security, QoS, transactions, reliability, access control, etc.) with the event-based paradigm. In addition, the developers are responsible for handling the low-level event transmission issues. Current publish/subscribe systems are restricted to certain application scenarios such as instant messaging and stock quote dissemination. This indicates that such systems are not designed as general middleware platforms. The majority of event-based middleware architectures are based on a logically centralized component called event dispatcher or broker. This component acts as a gateway between interacting components and hence has global knowledge about all the generated events and subscription requests. However, this centralized design often results in performance bottlenecks. Furthermore, not all event brokers provide a persistent buffer that can store all events for the provision of a reliable event service. There is also no support for the notion of composite events. Composite-event services allow clients to sign up with several event sources and receive event notifications in form of composite events. A special mechanism is needed to model event arrival from different sources and to specify composite events. This however may complicate the system architecture and incur extra cost.

CONCLUSION

Mobile computing is a relatively new field. While the challenges arising from mobility and the limitations of the portable devices are relatively well understood, there is no consensus yet as to what should be done to address these challenges. A comprehensive solution has to address many different aspects, such as the issue of dynamically changing bandwidth, the power, computational, and other limitations of the portable devices, or the varying availability of services in different environments. Traditional approaches to these challenges involved the design and implementation of proxies to either transparently intercept the data stream or to cooperate with the client application in the portable device in offloading computational tasks. To generalize this work, such services are expected to be embedded in middleware for mobile applications. Traditional middleware systems, such as CORBA and Java RMI, are based on the assumption that applications in distributed systems will run in a static environment; hence, they fail to provide the appropriate support for mobile applications. This gives a strong incentive to many researchers to develop modern middleware that supports and facilitates the implementation of mobile applications. To date, there is no single middleware that can fully support the requirements for mobile applications. Several solutions have considered one aspect or another; however, the door for further research is still wide open.

Table 1 provides a simple comparison of mobile middleware solutions with respect to the identified requirements for mobile computing. As can be seen, no single middleware paradigm covers all requirements. Based on our analysis, an evolution of reflective mobile

Table 1. Requirements analysis for modern middleware

	Reflective	Tuple-space	Context-aware	Event-based
Asynchronous		✓		✓
Reconfigurability	✓			
Adaptivity	✓		✓	
Awareness	✓		✓	
Lightweight				✓

middleware appears most promising, however. The key challenges are to reduce the complexity and size of such middleware for thin clients and to efficiently support asynchronous interaction styles.

REFERENCES

Angin, O. et al. (1998). The Mobiware toolkit: Programmable support for adaptive mobile networking. *IEEE Personal Communications, 5*(4), 32-43.

Balakrishnan, H. et al. (1995). Improving TCP/IP performance over wireless networks. *Proceedings of the 1st Annual International Conference on Mobile Computing and Communications*, Berkeley, CA, USA (pp. 2-11).

Blair, G.S. et al. (2001). The design and implementation of Open ORB 2. *IEEE Distributed Systems Online, 2*(6).

Bolliger, J., & Gross, T. (1998). A framework-based approach to the development of network-aware applications. *IEEE Transactions on Software Eng., 24*(5), 376-390.

Capra, L., Emmerich, W., & Mascolo, C. (2003). CARISMA: Context-aware reflective middleware system for mobile applications. *IEEE Transactions on Software Engineering, 29*(10), 929–945.

Cugola, G., Nitto, E.D., & Fuggetta, A. (2001). The JEDI event-based infrastructure and its applications to the development of the OPSS WFMS. *IEEE Transactions on Software Engineering, 27*(9), 827–850.

Fox, A. et al. (1998). Adapting to network and client variation using infrastructure proxies: Lessons and perspectives. *IEEE Personal Communications, 5*(4), 10-19.

Fritsch, D., Klinec, D., & Volz, S. (2000). NEXUS positioning and data management concepts for location aware applications. *Proceedings of the 2nd International Symposium on Telegeoprocessing*, Nice-Sophia-Antipolis, France (pp. 171-184).

Joshi, A., Weerawarana, S., & Houstis, E. (1997). Disconnected browsing of distributed information. *Proc. Seventh IEEE Intl. Workshop on Research Issues in Data Engineering* (pp. 101-108).

Kavi, K., Browne, J.C., & Tripathi, A. (1999). Computer systems research: The pressure is on. *IEEE Computer, 32*(1), 30-39.

Kon, F. et al. (2000). Monitoring, security, and dynamic configuration with the dynamicTAO reflective ORB. *Proceedings of the IFIP/ACM International Conference on Distributed Systems Platforms and Open Distributed Processing (Middleware'2000),* Heidelberg, Germany (pp. 121–143).

Kramer, J., & Magee, J. (1990). The evolving philosophers problem: Dynamic change management. *IEEE Transactions on Software Engineering, 16*(11), 1293-1306.

Kunz, T., Omar, S., & Zhou, X. (2000). Mobile agents for adaptive mobile applications. *Networking and Information Systems Journal, 3*(5/6), 709-723.

Meier, R., & Cahill, V. (2002). STEAM: Event-based middleware for wireless ad hoc networks. *Proceedings of the International Workshop on Distributed Event-Based Systems (ICDCS/DEBS'02),* Vienna, Austria (pp. 639-644).

Picco, G., Murphy, A., & Roman, G.-C. (1999). LIME: Linda meets mobility. *Proceedings of the 21st Int. Conference on Software Engineering* (pp. 368–377).

Seshan, M., & Katz, R. (1997). Spand: Shared passive network performance discovery. *Proc. 1st Usenix Symposium on Internet Technologies and Systems.*

Wyckoff, P. et al. (1998). T spaces. *IBM Systems Journal, 37*(3), 454-474.

KEY TERMS

Context-Awareness: Makes applications aware of the dynamic changes in execution environment. The execution context includes but is not limited to: mobile user location, mobile device characteristics, network condition, and user activity.

Event-Based Systems: Systems in which clients (*subscribers*) have to express (*subscribe*) their interest in receiving particular events. Once clients have subscribed, servers (*publishers*) publish events, which will be sent to all interested subscribers.

Middleware: An enabling layer of software that resides between the application program and the networked layer of heterogeneous platforms and protocols. It decouples applications from any dependencies on the plumbing layer that consists of heterogeneous operating systems, hardware platforms and communication protocols.

Portable Device: Computational device that is small and can be carried by its user, such as smart cell phones, PDAs, and laptops. Unlike stationary devices, the design of portable devices typically trades-off CPU speed, memory, I/O facilities, and so forth for reduced power consumption and size.

Reflective Software: Computational process that reasons about itself, comprising an ingredient process (interpreter) formally manipulating representations of its own operations and structures.

Tuple Space: A globally shared, associatively addressed memory space that is used by processes to communicate asynchronously, based on the notion of tuples and tuple matching.

Wireless Communication: Data communication that does not require a wired connection between communicating peers, typically using radio or infrared transmissions.

A

Adaptive Playout Control Schemes for Speech over the Internet

Marco Roccetti
University of Bologna, Italy

Stefano Ferretti
University of Bologna, Italy

INTRODUCTION

The transmission of speech over the Internet is often dismissed as an impractical application because of the poor quality experienced by many users of Internet audio tools. In fact, while Internet audio services are required to operate in a bandwidth-, delay-, and packet loss-constrained environment, the actual best-effort service offered by the Internet architecture does not provide any guarantee on the delivery of data packets. Thus, it often occurs that very high packet delay, delay jitter and packet loss are experienced over many congested Internet links.

To ameliorate the effect of network delay, delay jitter and packet loss rate, novel protocol suites and networking technologies (e.g., RSVP, ATM networks) have been devised that provide users with quality of service (QoS) guarantees (Zhang, Deering, Estrin, Shenker & Zappala, 1993). However, these approaches are not yet widely used and, as a consequence, in the absence of network support to provide guarantees of QoS to users of Internet voice tools, an interesting alternative amounts to the use of adaptive playout control mechanisms at the receiver side. Those approaches are adopted by the majority of the existing Internet audio tools, such as *NeVot* (Schulzrinne, 1992), *vat* (Jacobson & McCanne, n.d.), *rat* (Hardman, Sasse & Kouvelas, 1998), *FreePhone* (Bolot & Vega Garcia, 1996) and *BoAT* (Roccetti, Ghini, Pau, Salomoni & Bonfigli, 2001).

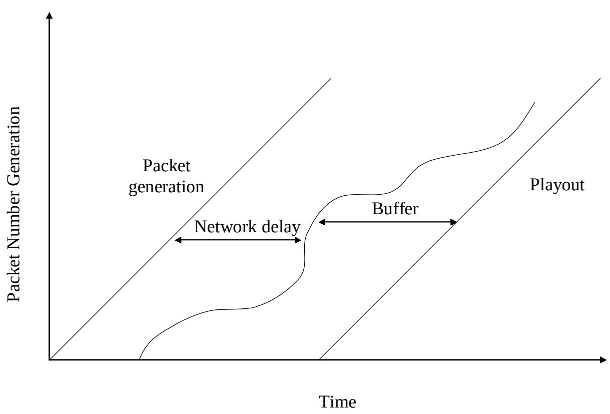
In essence, to compensate for variable network delays and remove jitter from audio packet streams, adaptive playout control mechanisms use a voice reconstruction buffer at the receiver and add artificial delay to the audio stream to smooth out the delay jitter. Simply put, received audio packets are first queued into the buffer and then the periodic playout of each packet is delayed for some quantity of time. This buffering policy must be adaptive, since jitter on the Internet may vary significantly with time. In this way, dynamic playout buffers can hide at the receiver packet delay variance at the cost of additional delay (see Figure 1).

It goes without saying that, even if this approach permits the removal of jitter from audio packet streams, and guarantees the speech intelligibility, a critical trade-off exists between the amount of delay that is introduced in the buffer and the percentage of late packets that are not received in time for playout (and are consequently lost). In fact, the longer the additional delay, the more likely it is that a packet will arrive before its scheduled playout time. Summing up, on one side, a too large percentage of audio packet loss may impair the intelligibility of an audio transmission, but, on the other side, too large playout delays (due to buffering) may disrupt the interactivity of an audio conversation (Kostas, Borella, Sidhu, Schuster, Grabiec & Mahler, 1998; Panzieri & Roccetti, 1997).

In conclusion, Internet playout control mechanisms adaptively adjust the playout delay of audio packets in order to keep this additional buffering delay as small as possible, while minimizing the number of packets delayed past the point at which they are scheduled to be played out (Boutremans & Le Boudec, 2003; Fujimoto, Ata & Murata, 2004; Liang, Farber & Girod, 2003; Sreenan, Chen, Agrawal & Narendran, 2000).

With this in view, in the remainder of this article we survey three different adaptive playout delay control mechanisms that have been designed to support speech

Figure 1. Smoothing out jitter delay at the receiver



transmission over the Internet. In particular, all three different approaches adopt adaptive control mechanisms that keep the same playout delay constant throughout a given talkspurt, but permit different playout delays in different talkspurts.

BACKGROUND

In this section we survey three different adaptive playout delay control schemes designed to transport speech over the Internet.

Playout Control Mechanism - #1

The adaptive playout delay adjustment algorithm proposed in Ramjee, Kurose, Towsley, and Schulzrinne (1994) is used in several Internet audio tools, such as *NeVoT* (Schulzrinne, 1992), *rat* (Hardman et al., 1998) and *FreePhone* (Bolot & Vega Garcia, 1996). This algorithm assumes that an external mechanism exists that keeps synchronized the two system clocks at both the sending and the receiving sites and that the delays experienced by audio packets on the network follow a Gaussian distribution. The mechanism works as follows.

The receiver buffers packets and delays their playout for a time quantity that is usually adaptively adjusted from one talkspurt to the next one. In essence, this mechanism works by calculating the playout time p_i for the first packet i of a given talkspurt as:

$$p_i = t_i + \underline{d}_i + k * \underline{v}_i,$$

where t_i is the time at which the audio packet i is generated at the sending site, d_i is the average value of the playout delay, $k \in [1, 2, 4]$ is a variation coefficient (whose effect can be enforced through shift operations) and v_i is the average variation of the playout delay. The playout point p_j for any subsequent packet j of that talkspurt is computed as an offset from the point in time when the first packet i in the talkspurt was played out:

$$p_j = p_i + t_j - t_i.$$

The estimation of both the average delay and the average delay variation are carried out using a stochastic gradient algorithm.

This strategy is also equipped with a delay spike detection and management mechanism. In essence, the algorithm works as described here, but when a spike is detected the mechanism uses the delay of the first packet of the talkspurt as the estimated playout delay for each packet in the talkspurt, in order to effectively react to very large change in transmission delays.

Playout Control Mechanism - #2

Another adaptive delay adjustment algorithm for speech transmission has been presented in Moon, Kurose, and Towsley (1998). The main idea behind this algorithm is to collect statistics on packets already arrived and then use them to calculate the playout delay.

In essence, the value of each packet delay is recorded and the distribution of packet delays is updated with each new arrival. When a new talkspurt starts, this mechanism calculates a given percentile point for an established amount of last arrived packets, and uses it as the playout delay for the new talkspurt.

In the presence of delay spikes the algorithm stops collecting packet delays as soon as a delay spike is detected, and starts following the spike by using as playout delay the delay value experienced by the packet that commenced that spike.

Playout Control Mechanism - #3

Another mechanism designed to dynamically adapt the talkspurt playout delays to the network traffic conditions has been proposed in Roccetti et al. (2001). This mechanism is at the basis of an Internet audio tool, called *BoAT*.

The mechanism is able to dynamically adjust the talkspurt playout delays to the network traffic conditions without assuming either the existence of an external mechanism for maintaining an accurate clock synchronization between the sender and the receiver, or a specific distribution of the end-to-end transmission delays experienced by the audio packets.

Succinctly, the technique for dynamically adjusting the talkspurt playout delay is based on obtaining, in periodic intervals, an estimation of the upper bound for the packet transmission delays experienced during an audio communication. Such an upper bound is periodically computed using round-trip time values obtained from packet exchanges of a three-way handshake protocol performed between the sender and the receiver of the audio communication. Then, the upper bound is used to dynamically adjust the playout delay from one talkspurt to the next, with the introduction of artificially elongated or reduced silence periods.

CRITICAL ISSUES

The need of silent intervals for allowing a playout delay control mechanism to adjust to the fluctuating network conditions is common to all three Internet audio mechanisms described in the previous section. This renders all the described control schemes particularly suitable for voice-based conversational audio with intervening silence periods between subsequent talkspurts. In fact, the presence of silent intervals permits to dynamically adjust the playout delay from one talkspurt to the next.

Thus, in order to assess the efficacy and the viability of these mechanisms, an accurate model of the talkspurt/silence characteristics of conversational speech is necessary for understanding whether sufficient (and sufficiently long) silent intervals occur in typical human conversations.

With this in view, Roccetti, Ghini and Pau (2001) conducted a simulation study showing that a sufficient total amount of silent intervals occur in human conversational speech to be used at the receiver to accommodate changes of the audio packet transmission delay, while maintaining speech intelligibility.

In particular, it is shown that the total quantity of silent intervals within a simulated two-party one-hour-long packetized conversation amounts to about 63-66%, depending on the voice packetization interval (see Figure 2).

Further, another important result obtained in that study is that the smaller the chosen packet voice size, the more the total number of silence periods increase and the average talkspurt length decreases. Instead, the larger the packet size, the larger the average duration of the talkspurt.

The main results concerning the quantity and the duration of the talkspurts are summarized in Figures 3 and 4 where, respectively, the total quantity of packetized talkspurts (with duration smaller than a fixed amount) and the percentage of talkspurts (with length smaller than a fixed amount) are reported.

As the duration of intervening silence periods in human speech may be artificially reduced or elongated to accommodate at the receiver changes of the audio packet transmission delays, another important question to be

Figure 2. Total amount of silence periods

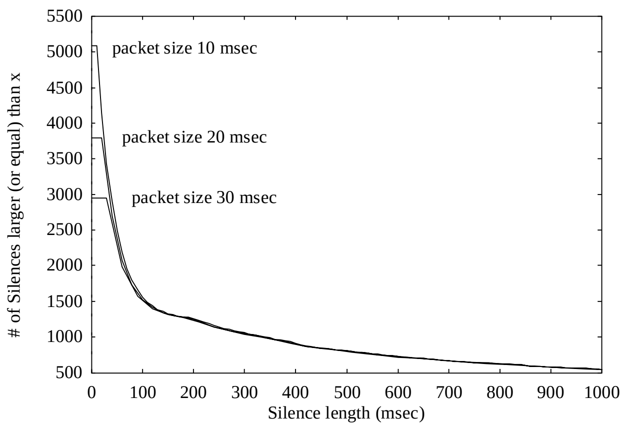


Figure 3. Total amount of talkspurts

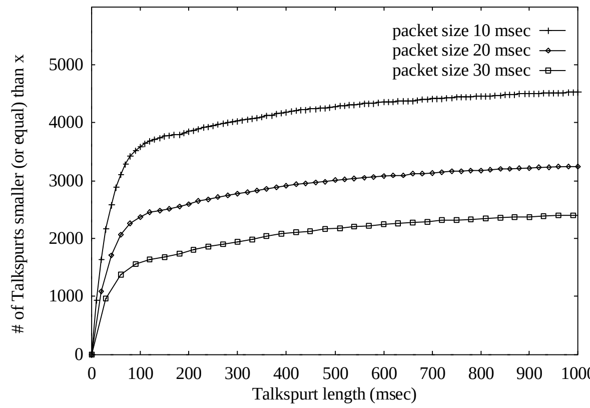


Figure 4. Percentage of talkspurts (with duration smaller than x) with regard to the total number of talkspurts

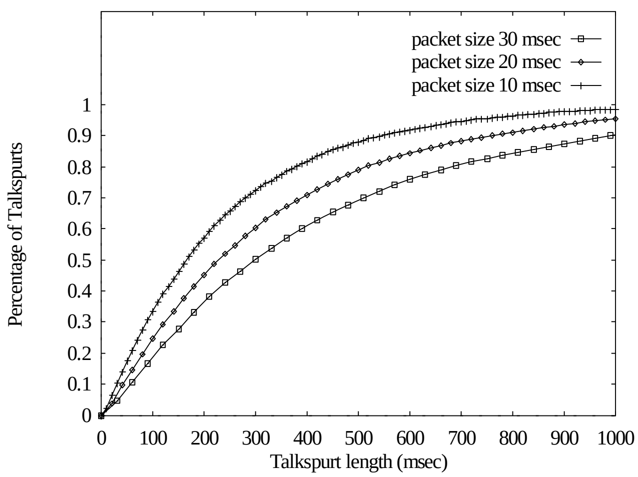
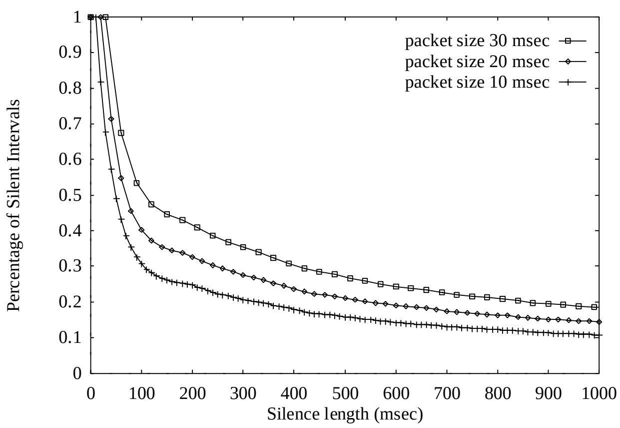


Figure 5. Percentage of silent intervals (with duration larger than x) with regard to the total number of silence periodsh



answered concerns the average length of the intervening silence periods in a human conversation. To this aim, in Figure 5 the percentage of silent periods (of length larger than a fixed amount) out of the total quantity of all the obtained silence periods are reported. Of particular interest is the fact that the larger the packet size, the larger the average silence duration.

In conclusion, the results we have reported show that there is room in human speech (i.e., silent intervals) to successfully exploit sophisticated Internet audio control mechanisms.

FUTURE TRENDS

A recent review on Internet telephony (Bradner, 2002) shows that despite important research, its actual deployment is really limited due to quality of service (QoS) concerns. Recent studies summarize the key issues relating to the design of adaptive receiver-based buffer strategies for Internet speech transmission as follows: i) Internet packet loss, delay jitter and speech distortion effects over best-effort networks remain a serious problem that needs deeper modeling efforts. To this aim, further analysis of Internet delay studies is being undertaken; ii) more investigation is needed on the optimal implementation of application delay or buffering delay within adaptive receivers and, finally, iii) more analysis is required of the uncertainties caused by hardware/software within Internet speech endpoints.

CONCLUSION

Delay adaptation in the presence of fluctuant network delays is a crucial issue in determining the audio quality for real-time speech transmission over the Internet. With this in view, the typical approach is that of dynamically adapting the audio application to the network conditions so as to minimize the trade-off between packet playout delay and packet playout loss. The performance figures illustrate the adequacy of those mechanisms for human speech transmission across the Internet.

REFERENCES

Bolot, J., & Vega Garcia, A. (1996). Control mechanism for packet audio in the Internet. *Proceedings of IEEE SIGCOMM '96*, San Francisco (pp. 232-239).

Boutremans, C., & Le Boudec, J.Y. (2003). Adaptive joint playout buffer and FEC adjustment for Internet telephony. *Proceedings of IEEE INFOCOM'03*, San Francisco (pp. 652-662).

Bradner, S. (2002). Internet telephony-progress along the road. *IEEE Internet Computing, 6*(3), 37-38.

Fujimoto, K., Ata, S., & Murata, M. (2004). Adaptive playout buffer algorithm for enhancing perceived quality of streaming applications. *Telecommunications Systems, 25*(3), 259-271.

Hardman, V., Sasse, M.A., & Kouvelas, I. (1998). Successful multiparty audio communication over the Internet. *Communications of the ACM, 41*(5), 74-80.

Jacobson, V., & McCanne, S. (n.d.). vat. Retrieved April 2004, from *ftp://ftp.ee.lbl.gov/conferencing/vat/*

Kostas, T.J., Borella, M.S., Sidhu, I., Schuster, G.M., Grabiec, J., & Mahler, J. (1998). Real-time voice over packet-switched networks *IEEE Network, 12*(1), 18-27.

Liang, Y.J., Färber, N., & Girod, B. (2003). Adaptive playout scheduling and loss concealment for voice communications over IP networks. *IEEE Transactions on Multimedia, 5*(4), 532-543.

Moon, S.B., Kurose, J., & Towsley, D. (1998). Packet audio playout delay adjustment: Performance bounds and algorithms. *ACM Multimedia Systems, 6*(1), 17-28.

Panzieri, F., & Roccetti, M. (1997). Synchronization support and group-membership services for reliable distributed multimedia applications. *ACM Multimedia Systems, 5*(1), 1-22.

Ramjee, R., Kurose, J., Towsley, D., & Schulzrinne, H. (1994). Adaptive playout mechanisms for packetized audio applications in wide-area networks. *Proceedings of IEEE INFOCOM'94*, Montreal, Canada (pp. 680-688).

Roccetti, M., Ghini, V., & Pau, G. (2001). Simulative experimental analysis of an adaptive playout delay adjustment mechanism for packetized voice across the Internet. *International Journal of Modelling and Simulation, 21*(2), 101-106. ACTA Press.

Roccetti, M., Ghini, V., Pau, G., Salomoni, P., & Bonfigli, M.E. (2001). Design and experimental evaluation of an adaptive playout delay control mechanism for packetized audio for use over the Internet. *Multimedia Tools and Applications, 14*(1), 23-53. Kluwer Academic Publishers.

Schulzrinne, H. (1992). *Voice communication across the Internet: A network voice terminal.* Department of ECE and CS, University of Massachusetts, Amherst (MA).

Sreenan, C.J., Chen, J.C., Agrawal, P., & Narendran, B. (2000). Delay reduction techniques for playout buffering. *IEEE Transactions on Multimedia*, *2*(2), 100-112.

Zhang, L., Deering, S., Estrin, D., Shenker, S., & Zappala, D. (1993). RSVP: A new resource ReSerVation Protocol. *IEEE Network Magazine*, *7*(5), 8-18.

KEY TERMS

Audio Packet: Packet encoding an audio sample in digital form. Each audio packet has a timestamp and a sequence number as additional information. Timestamps are used to measure the end-to-end delay (jitter) experienced during the communication, and sequence numbers are used to detect packet losses. Typically, during an audio communication, audio packets are transmitted over the network, received in a playout buffer, decoded in sequential order and, finally, played out by the audio device.

Audio Sample: The amplitude of a waveform is measured (sampled) at regular time intervals and converted into an integer value. Each of these instantaneous measurements is an audio sample.

Delay Jitter: Variance of the network delay computed over two subsequent audio packets.

Delay Spike: Sudden, large increase in the end-to-end network delay, followed by a series of audio packets arriving almost simultaneously.

Network Delay: Time needed for the transmission of a data packet from the source to the destination over the underlying transport network.

Playout Buffer: Buffer used at the receiver side to store received audio packets in interactive real-time applications in order to compensate for variable network delays.

Playout Control Mechanism: Adaptive mechanism that fetches audio packets from the playout buffer and sends them to the audio device for immediate playout.

Playout Delay: Total amount of time that is experienced by an audio packet from the time instant it is generated at the source and the time instant it is played out at the destination. Such a delay consists of: i) the collection time needed for the transmitter to collect an audio sample and to prepare it for transmission, ii) the network delay, and finally iii) the buffering time, which is the amount of time that a packet spends queued in the destination buffer before it is played out.

Talkspurt: Short burst of energy, in an audio communication, during which the audio activity is carried out. A packetized audio segment may be considered as being constituted of several talkspurts separated by silent periods (during which no audio packet is generated). During speech transmission activities over the Internet, in order for the receiving site to reconstruct the audio conversation, the audio packets constituting a talkspurt must be played out in the order they were emitted at the sending site.

Addressing the Central Problem in Cyber Ethics through Stories

John M. Artz
The George Washington University, USA

INTRODUCTION

The central problem in cyber ethics is not, as many might think, how to address the problems of protecting individual privacy, or preventing software piracy, or forcing computer programmers to take responsibility for the systems that they build. These are, of course, legitimate concerns of cyber ethics, but the central problem is how you decide what the right thing to do is with regard to these issues when the consequences of any responses cannot be known in advance. Stated more clearly, the central problem in cyber ethics is - how do you establish ethical standards in a professional field that is defined by a rapidly evolving technology where the consequences of the technology and the impact of any ethical standards cannot be known in the time frame in which the standards must be established? Stories play a very important role in addressing this issue. Specifically, stories provide a means of exploring ethical issues for which the full range of consequences is not currently known. But, in order to justify this claim, a few words of explanation are in order.

BACKGROUND

The word "story" means many different things to different people. For example, if one of your children tells you that your dog ate the neighbor's cat, you might challenge the veracity of this claim by asking – "Is that true, or is that just a story?" The implication is that there is truth and there are stories. And if it is a story, it cannot be true. But true versus fictitious is not the same as true versus false; and a story can contain important truths while still being wholly fictitious. If we are looking for precise intellectual truths, then perhaps stories are not the best medium for exploration. However, in areas where our understanding is unclear, either because we do not fully understand a phenomenon, or the phenomenon is not available for study because it exists in a possible world, stories play a very important role in advancing our understanding. To put a finer point on this argument, science and logic fail miserably at telling us what could be, or more importantly, what should be. In these two areas stories are powerful vehicles for intellectual explorations. A story, for the purposes of the current discussion, is a rendition or a telling of a series of true or fictitious events, connected by a narrative in which a set of characters experience and react to a set of actions or events and in doing so reveal something about the human character or condition. In order to see the value of stories for the exploration of issues in cyber ethics, three prior arguments must be made.

NARRATIVE VS. LOGICAL THINKING

Narrative and logical reasoning represent two distinct methods of making sense out of the world around us. They are both legitimate and both can be very rigorous (Bruner, 1986). Sometimes they provide alternative paths to truth and understanding. Sometimes one or the other provides the only path. Logical reasoning is general, context independent, objective and leads to a single conclusion. Narrative reasoning is specific, context dependent, open to subjective interpretation, and potentially leads to multiple conclusions. The characteristics of narrative reasoning are considered flaws when applied to logical reasoning. But the reverse applies also. A story that has only one interpretation and means the same to everyone is not much of a story. While narrative and logical reasoning are different kinds of reasoning, they are not mutually exclusive. A good narrative is also often quite logical in structure, and a good logical argument can often be better understood with a good narrative example. But for the most part, they are complimentary, alternative modes of thinking that provide different paths to truth and understanding.

To some extent, logical and narrative reasoning address different domains. Logic is well suited to mechanistic processes that can be reduced to logical description. Logic is good for articulating general principles and deductive reasons. Logic is useful for describing and explaining. While logic is good for describing "what is," narrative is good for exploring "what could be" and figuring out "what should be". Narratives are a useful means for understanding the complex and ambiguous issues in human affairs. They allow us to explore possibilities and experience situations vicariously. Narrative reasoning is particularly well suited to cyber ethics because many issues are not well understood and the goal of cyber ethics is not to discover truth about the physical world, but truth about human nature. Narrative fiction gives us a means to explore and discover truths about what could be and what should be. Through narratives we can explore possible consequences of technology, construct alternative worlds and select the one in which we would like to live.

Critics of the use of narrative in ethics point out that after exploring narratives you always have to come back to principles. Ethics, they argue, is too messy without principles and discussion of narratives does not lead to consistent conclusions. This view misses the point of narratives. First, principles are developed by extracting the principles from experience. Narratives provide some of these experiences vicariously. Hence, narratives can be used in the development of principles. Second, it is often unclear which principles apply in given situations. Narrative explorations provide insight into situations, allowing us to determine the governing principles. And narratives can be used to explore the consequences of principled decisions to determine if the outcomes are indeed what are intended. Finally, narrative reasoning does lead to conclusions - very specific conclusions about very specific situations. Narrative reasoning is lacking in generality, as was mentioned before, not lacking in conclusions.

THE ROLE OF EMOTION IN REASON

Most people believe that emotions have no role in logical reasoning. After all, reasoning should be dispassionate and free from emotional influences that may cloud our reasoning. And there is some basis for this. For example, if you lose your temper in the middle of an argument and start flinging personal insults at your opponent, rational people would not consider you as having advanced your position. Most would say that you lost the argument when you lost your temper. Yet emotions play an important role in reasoning and in order to understand this, we need to better understand exactly what emotions are.

There is considerable debate about the exact nature of emotions. The philosopher Robert Solomon (Solomon, 1994) offers one very useful observation that "emotions are judgments about the world". If you are walking down a path in the woods and it is getting dark, you might start to get a little nervous and walk a little faster. If you hear an unfamiliar noise or a rustling in the leaves your heart may begin to beat a little faster as you experience the emotional reaction of fear. This fear is a judgment about the world in which you have judged your current situation as unsafe. You did not arrive at this judgment through a rational process. Specifically, you did not think – "It is dark and hungry animals or possibly monsters come out when it is dark. I just heard a noise that I cannot identify and therefore there could be a hungry animal near me. If I walk a little faster, I might get away before the animal gets me. If I am wrong then all I have done is walked a little faster. If I am right, I might avoid being eaten. Hence, it is logical and reasonable for me to walk faster." In fact, you probably did not think at all. You just felt scared and increased your pace. If asked later why you were walking so quickly you might come up with a reasonable explanation. But that reasonable explanation is certainly constructed after the fact.

Perhaps we have conceded at this point that emotions are judgments about the world and that they can play an important role in reasoning. The obvious question is "So what?" Why do we care and why should we bother to make an effort to incorporate our emotional judgments into our reasoning? Damsio (1994) describes the case of a young man who after suffering damage to part of his brain was no longer able to feel emotions. The unexpected side effect of this malady was that he was also unable to make good decisions or assign importance to competing tasks. He seemed normal in every other way and seemed to have his intellectual facilities fully intact. Yet he seemed no longer able to feel emotions and as a result he was unable to function as a normal person. When we make a decision we evaluate alternatives. If we are unable to feel emotions we are unable to place values on the different alternatives. If we cannot place values on the different alternatives then there is no difference between the alternatives and decision-making becomes seriously flawed. Hence, without emotions rational decision-making may not be possible.

A good story about an ethical issue is much more likely to draw an emotional response that an intellectual one, whereas an abstract analysis is more likely to yield an intellectual response. Ultimately, ethical decisions are emotional decisions because they embody human values. For this reason, examining ethics from a purely rational perspective completely misses the point.

IMAGINATION AND POSSIBLE CONSEQUENTIALISM

One of the problems in establishing standards of ethical behavior in a field driven by technology is that the consequences of the technology and reactions to the technology often cannot be known. Looking to the past to provide guidance is ineffective because the past provides few clues. Marshall McLuhan is often attributed with the famous observation that looking to the past to provide guidance for the future is like driving by looking in the rear-view mirror. Although it is disputed as to whether he ever said that or not, it is a rich metaphor for understanding how we should think about the future in times of rapid technological change.

Imagination is the key to understanding the future. The problem though, in using imagination to understand the future, is that we have a cognitive bias against understanding the future. We feel quite comfortable that we understand the past, but the future is the domain of prophesies. Yet assertions about the past are never testable because the past is gone, never to return, while assertions about the future are testable. So one could argue, on the basis of the testability criterion, that the future is more knowable than the past. However, that discussion is for another time.

Consider imagination as the creative capacity to think of possibilities. Imagination lets us see the world not as it is, but as it could be. Seeing the world as it could be allows us to make choices about how it should be. It is this ability to see possibilities that drives us to build technologies to bring about, or implement our preferences about possible worlds. Stories are both a product and a tool of our imaginations. Using stories in moral reasoning provides a means for a slightly different view of ethics that could be called "possible consequentialism". Whereas the consequentialist evaluates actions based upon their consequences, the possible consequentialist evaluates actions based upon their possible outcomes. The possible outcomes are described in stories and the likelihood of the outcome is determined by the believability of the story given our understanding of current conditions and human nature. As the literary critic Northrop Frye points out, "The fundamental job of the imagination in ordinary life, then, is to produce, out of the society we have to live in, a vision of the society we want to live in" (Frye, 1964, p. 140).

When we examine issues in cyber ethics, we cannot examine them in terms of consequentialist ethics because the consequences are not known. However, through the use of stories we can construct imaginative scenarios and examine possible consequences. Possible consequentialism may be a preferable approach to computer ethics because we can look at possible outcomes, assess the likelihood of each, and select the outcome we prefer. Imagination provides us with a means of fully examining possible outcomes and stories provide us with the means of sharing our imaginings. By writing stories and sharing them we can explore possible consequences and, through social debate, derive imaginary truths. These imaginary truths allow us to choose the kind of world that we would like to live in.

FUTURE TRENDS

While stories play an important role in the exploration of problems in cyber ethics, there is still a serious barrier to using them. That barrier is the fact that precious few stories have been written. Current work is addressing that problem (Artz, 2004). As researchers and teachers in cyber ethics become more comfortable with writing stories, the value of stories for exploring possible worlds will make this approach increasingly more attractive. Further, as the implications of information technology and the ethical problems they bring move further into the future, the ability of stories to capture possible worlds will make them increasingly more compelling.

CONCLUSION

The central problem in cyber ethics is not, as many may suppose, how to prevent software piracy or how to protect privacy on the Internet. It is instead - how do you establish ethical standards in a professional field that is defined by a rapidly evolving technology where the consequences of the technology and the impact of any ethical standards cannot be known in the time frame in which the standards must be established? Stories play an important role in addressing this problem by providing a means of exploring ethical issues for which the full range of consequences are not known. Stories allow us to construct narrative arguments to explore issues that we do not fully understand. They allow us to explore the emotional as well as rational aspects of a situation. Stories allow us to explore worlds that do not currently exist, which, in turn, allows us to examine possible consequences and make choices about the world in which we would like to live.

REFERENCES

Artz, J. (2003). The central problem in cyber ethics and how stories can be used to address it. In L. Brennan & V. Johnson (Eds.), *Social, ethical and policy implications of information technology*. Hershey, PA: Information Science Publishing.

Artz, J. (2004). Using a story to explore an ethical dilemma. *Computers and Society E-Journal*.

Bruner, J. (1986). Actual minds, possible worlds. *Harvard University Press*.

Damsio, A. (1994). *Descartes' error: Emotion, reason, and the human brain*. Avon Books.

Edgar, S. (2002). *Morality and machines: Perspectives on computer ethics* (2nd ed.). Jones & Bartlett Pub.

Frye, N. (1964). *The educated imagination*. Indiana University Press.

Goleman, D. (1995). *Emotional intelligence*. Bantam Books.

Johnson, D. (2000). *Computer ethics* (3rd ed.). Prentice-Hall.

Solomon, R. (1994). *Love and Vengeance: A course on human emotion*. The Teaching Company Superstar Teachers Series.

KEY TERMS

Cyber Ethics: A branch of ethics that focuses on behaviors that are specifically related to information technology.

Ethics: A branch of moral philosophy that examines the standards for proper conduct.

Imagination: The creative capacity to think of possibilities.

Logical Reasoning: General, context independent, objective reasoning that leads to a single conclusion.

Narrative Reasoning: Specific, context dependent reasoning that is open to subjective interpretation, and potentially leads to multiple conclusions.

Possible Consequentialism: Evaluating an ethical position in terms of its possible consequences.

Story: A rendition or a telling of a series of true or fictitious events, connected by a narrative, in which a set of characters experience and react to a set of actions or events and in doing so reveal something about the human character or condition.

Adoption of E-Commerce in the Value Chain by SMEs

Judith Jeffcoate
University of Buckingham, UK

INTRODUCTION

This article describes an analytical framework to identify the triggers for value chain transformation that will encourage small and medium enterprises (SMEs) to adopt e-commerce. The framework was adopted in the KITS project as part of a qualitative approach, using multiple case studies of SMEs that were involved in business-to-business (B2B) e-commerce (Jeffcoate, Chappell & Feindt, 2004). Results from this project are used to illustrate the use of the framework.

According to Yin (1994), one of the areas in which case research is most useful is the relationship between information technology and corporate strategy—precisely the area that is crucial to the development of e-commerce. Earlier researchers also adopted multiple cases for studies of SMEs in e-commerce or related areas. For example, Poon and Swatman (1997) studied 23 small businesses in Australia that were active Internet users in order to establish the pre-conditions for success and the strategic effects from its use. Other researchers in this area have relied on much smaller numbers of case studies, typically fewer than 10. Iacovou, Benbasat and Dexter (1995) carried out interviews on the adoption of electronic data interchange (EDI) with the managers of seven SMEs that were suppliers to the British Columbia government. Another study of the effect of EDI on a sample of eight small businesses in the UK (Chen & Williams, 1998) was based on in-depth interviews with key personnel.

BACKGROUND: THE KITS PROJECT

Knowledge Information Transfer Systems (KITS) was designed as a support action for SMEs in Europe and funded by the European Commission. It included a research programme whose aim was to analyse the activities of SMEs engaged in industry value chain relationships with suppliers and customers in order to establish best practice. An SME in this context is a company that employs fewer than 250 people. The research programme for KITS included two main work packages: the first looked at individual companies, whilst the second looked at evaluation across the value chain. This article, which is based on the first of these packages, describes the analytical framework that was adopted by this project.

The KITS sample consisted of 43 European SMEs that were actively involved in B2B e-commerce across industry and country boundaries (Table 1). The research focussed primarily on traditional SMEs with some level of automation, with a secondary focus on e-commerce start-ups (the 'dot.coms'). The interview programme was car-

Table 1. Companies by sector and country

	Manufacturing	Transport and logistics	Retail/wholesale	Business services	Total
Germany	7	3	1	3	14
Austria		1			1
Switzerland				1	1
Belgium			1		1
Italy	12	1	1		14
Sweden	1				1
UK	5	1	3	2	11
Total	**25**	**6**	**6**	**6**	**43**

ried out between September 2000 and January 2001. All of the respondents had board-level responsibility. They varied from managing directors in the smaller companies to IT, operations, and marketing directors in the larger organisations. It was felt that, in an SME, only one point of view is necessary, providing it is at a high enough level within the organisation.

The interviews were conducted using a semi-structured questionnaire that included both open and closed questions. The initial conceptual framework was adapted from one proposed by Ward and Griffiths (1996) for determining the strategic potential of information systems and establishing priorities for investment. Their framework takes a structured analytical route through the upper levels of the organisation. It is closely linked to the industry value chain, as well as to the organisational value chain, and was thus particularly appropriate for KITS. The following topics were covered: company background; strategy, objectives, and critical success factors; value chain partnerships and activities; and use of technology.

THE ANALYTICAL FRAMEWORK

The framework adopted for the analysis of the KITS case studies draws on key concepts proposed by a number of authors. It consists of the following elements:

- Type of industry value chain
- Stability of position within industry value chain
- Level of penetration of electronic links with customers and suppliers
- Complexity of automation of the internal IT environment
- Key value activities and interactions between value activities
- Complexity of automation of value activity interactions
- Level of achievable impact on the industry value chain

These elements are discussed in turn in the following sections, with illustrations from the KITS project.

CURRENT POSITION: THE SME IN THE VALUE CHAIN

A key element in the process by which SMEs adopt e-commerce is through the automation of the value chain, defined by Porter (1984) as a collection of activities that are performed by a company to design, produce, market, deliver, and support its product. Other authors refer to this as the internal value chain of a company, or organisational value chain, in contrast to the industry value chain. The latter consists of the organisational value chain, together with the value chains of the organisation's competitors, suppliers, and customers. It represents the movement of goods and services from the source of raw materials through to the final customer (Benjamin & Wigand, 1995).

Porter identified nine value activities, the physically and technologically distinct activities that the company performs and which add value to a product or service. More recently Chu (1995) defined critical value activities as those activities that an organisation must execute satisfactorily to ensure successful performance.

Type of Value Chain

An SME's perception of its value chain is important because it influences the company's strategy and therefore its e-commerce strategy. Companies can be classified as participating in one of four types of industry value chain defined by Baldock (1999):

- *Customer-Centric Value Chain:* The seller tailors its products to meet fast-changing consumer needs.
- *Seller-Driven Value Chain:* The seller presumes toknow what the market might want to buy.
- *Buyer-Driven Value Chain:* The customer states what he or she wants and sets out the terms and conditions that the supplier should meet.
- *Fragmented Value Chain:* Neither the buyer nor the seller business model dominates.

The majority of the KITS SMEs, who were in customer-centric value chains, believed that they should become more customer focused to meet the needs of such a value chain. They wanted to understand customers' needs better and more quickly, and, if possible, be proactive in meeting them.

Stability of Position within Value Chain

SMEs have upstream and downstream links with their business partners in the industry value chain. The strength, intensity, and permanence of these links are important in determining how deeply these SMEs are embedded in their networks (Yli-Renko & Autio, 1998). The majority of the KITS SMEs reported stable relationships with customers that had lasted for a number of years. They supported similarly stable relationships with suppliers, to whom they were highly loyal. However, they typically had far fewer suppliers than customers.

Level of Penetration of Electronic Links

The SME may support the links with its business partners by using a variety of technologies including network connections, support for information exchange, the use of applications (e.g., messaging, marketing), and of external services (e.g., value-added network service suppliers, application service providers, or Web hosting companies).

The penetration of electronic links into the supplier and customer bases of the KITS companies was low. Where such links existed, they were typically with a fraction of an SME's customer base: with key resellers or distribution partners, with one or two IT-literate customers. The electronic links were also typically being developed to secure and strengthen these relationships. As they proliferate through the customer and supplier base, however, the opposite effect may be achieved, and more *ad hoc* relationships with value chain partners may develop.

Amongst the KITS companies, electronic links with customers were more common and more extensive than electronic links with suppliers, confirming that, in most cases, customers tend to be larger and more e-commerce literate than suppliers. This observation is also consistent with the experience of larger companies that are automating their value chain activities. Fawcett and Magan (2001) report that most large U.S. companies surveyed are dedicating more resources to building strong customer relationships than to selecting and developing a world-class supply base.

Complexity of Automation of IT Environment

It is widely agreed that SMEs are being held back from participating in B2B e-commerce by the fact that they lack suitable IT infrastructures to support it (Lockett & Brown, 2001). They may have Internet access for e-mail and hosted Web sites, but it is thought that few SMEs engage beyond this in B2B e-commerce and other more complex applications.

The complexity of the IT environment can be assessed, using the following factors:

- The type of applications and IT infrastructure the SME has in place, including legacy applications and network environment
- Whether applications are packages or proprietary developments
- The level of application complexity
- The nature of the links between applications (e-mail, intranet, extranet)

The level of complexity, as defined by Lockett and Brown (2001), may range from very low for simple communications (e.g., e-mail, Web access) to very high for collaborative enterprise applications such as supply chain management (SCM) and customer relationship management (CRM). Hawkins (2001) points out that "home-made systems appear not to damage the e-commerce prospects of firms and may even improve them by ensuring that technology investment and roll-out is linked to actual signals from the market."

More than a third of the KITS companies conform to the general perception that the lack of IT infrastructure is a barrier to SME e-commerce, with a complexity rating of very low. Most of these do not have any electronic links with customers or suppliers. A further third of the sample have a complexity ranking of low. These companies do not necessarily automate more activities, but there is generally more integration between their activities. At least a third of the KITS companies have an IT infrastructure that can support relatively complex B2B e-commerce. A number of companies with lower complexity rankings are moving quickly towards this position.

Key Value Activities and Value Activity Interactions

The electronic links discussed above can link critical value activities both internally and between partners in the industry value chain. The relative importance of these activities to the SMEs can be established by asking respondents to rank each activity on a scale from 1 to 5, where 1 is unimportant and 5 is critical to the success of the company. It may be assumed that if an SME ranks a value activity as critical, then inputs and/or outputs to this activity that involve value chain partners will be a priority for automation. E-commerce solutions automate value activity interactions (VAIs) between value chain partners—each SME's customers and suppliers. Most of the VAIs link an SME's primary value activities, such as logistics, operations, or sales and marketing, with customer/supplier primary activities.

In the KITS sample the most widely supported value activity interaction is sales and marketing to procurement. All the SMEs in the sample have implemented a Web site with details of the company and its products. The second most supported VAI is outbound logistics to (customer) procurement. The fact that so many companies have implemented online order processing suggests that they perceive this to be important to customers, and that they will be better regarded by customers if they invest in it.

Complexity of Automation of Value Activities

The two most complex e-commerce solutions described by KITS respondents (vendor-managed inventory between two Web-enabled enterprise resource management systems and e-marketplaces) were not fully operational at the time of interview. It may be expected that, over time, it will be possible to track the development and impact of the different types of e-commerce solution. It may also be assumed that there will be wider adoption of more complex solutions over time.

Impact on the Value Chain

The complexity of an e-commerce solution affects an SME's ability to increase its value and even transform its role within its value chain. High-complexity e-commerce solutions have more potential to support value chain transformation than low-complexity solutions. Position in the value chain does seem to affect the complexity of the e-commerce solution adopted. Some SMEs, most notably those in the middle tiers, such as wholesalers (including e-marketplaces) and logistics companies, are able to provide additional intangible value to their value chain partners through the embryonic creation of new, virtual value chains (Rayport & Sviokla, 1995).

Within the KITS sample there appear to be four ways in which automating VAIs has an impact on an SME's value chain. VAI automation may:

1. Streamline the supply chain, taking out risk and cost. The electronic channel to customers/suppliers replaces existing methods of carrying out the value activity interaction (e.g., online ordering replacing phone/fax).
2. Transform the way a value activity is carried out, removing cost and adding new value. The way in which a value activity is automated changes the roles of customer and supplier (e.g., suppliers carry out inbound logistics on behalf of their customers, matching information about their own capability to deliver with information about the customer's production plans and sales through vendor-managed inventory).
3. Capture information (intangible good/service) that adds value to internal processes. For example, a company may gather information on customers through its Web site and organise, select, and distribute it to improve internal processes, such as sales and marketing.
4. Extend new information (intangible good/service) to external value chain partners' processes. A small portion of this information is being used to generate a revenue stream in its own right (for example, by e-marketplaces, which are synthesising and distributing information on a subscription basis).

FUTURE TRENDS: TRANSFORMING THE VALUE CHAIN

The KITS project aimed to research and analyse the activities of SMEs engaged in industry value chain relationships with suppliers and customers, in order to identify the triggers for value chain transformation that will encourage them to adopt e-commerce. A number of conclusions were drawn from the application of the analytical framework described above to the case studies in this project.

Is the Type of Industry Value Chain a Factor in Encouraging SME E-Commerce?

The impetus for adopting e-commerce solutions does not seem to depend on the industry sector or type of value chain in which the SME is engaged. It is expected that, over time, more value chains will become buyer-driven; this would have an impact on SMEs' roles and e-commerce strategies.

How Does the Stability of the Value Chain Affect SME E-Commerce?

The value chains studied are relatively stable, and the penetration of electronic linkages up and down the chains is low. However, the low level of value chain automation is a brake to change. Early adopter SMEs cannot capitalise on their innovation and leadership until the whole value chain is at a critical point for change.

Is the Level of Penetration of Electronic Links with Customers and Suppliers Important?

Once SMEs have established the benefits of e-commerce in a few key relationships, they may be motivated to extend its deployment fairly rapidly, as they can achieve better return on investment when the solution is extended to a critical mass of value chain partners. Strength in the value chain comes from partnership. Two interlinked SMEs that have synchronised their e-commerce strategies and solutions are more powerful than two individual companies. Business partnerships and alliances are not new, but e-commerce solutions make them easier and less costly to manage and more likely, other factors being equal, to succeed.

Is Complexity of Automation a Barrier for SME E-Commerce?

It is clear that companies with a vision of how they would like to participate in their value chain in the future are more likely to adopt medium- or high-complexity e-commerce solutions. These companies have considerably more power to effect change in the value chain than might have been expected, even in value chains with powerful customers downstream or dominant suppliers upstream. SMEs can be equally as innovative at driving value chain change as larger companies, and they can be faster and more flexible in doing so. The factors holding them back are not e-commerce related (such companies have a great deal of competency in this area), but have to do with risk, timing, and lack of access to the 'visionaries' in suppliers and customers.

Are Value Activity Interactions Important for SME E-Commerce?

Those SMEs that are applying e-commerce to value activity interactions with most determination are those that are proactively creating new, intermediary roles in the value chain, such as marketplaces that would not exist without the Internet. Investing in transformation is a high-risk strategy for a small company. The development of new services, either as add-ons to or as a revenue-generating replacement for an existing physical value chain role, is expensive and the outcome is uncertain.

Which Value Activity Interactions Should SMEs Automate?

SMEs need to select for automation value activity interactions that will support their business strategy, support their e-commerce strategy, and are appropriate for the level of IT literacy in their value chains. They therefore need to understand the impact particular automated value activities interaction will have on their value chain relationships and the kind of information (intangible good/service) VAIs will yield. E-commerce solutions should be selected based on the complexity level of partners' IT environments and how critical the VAI is to the business.

Can SME E-Commerce Transform Industry Value Chains?

Where e-commerce solutions support the 'bundling' and/or interlinking of VAIs (very high-complexity solutions), a virtual value chain of information-driven intangible goods/services may emerge. In certain circumstances, SMEs can use such intangible goods/services to generate the benefit of a completely new revenue stream.

CONCLUSION

In summary, although the use of this framework has already produced interesting insights into the objectives and practice of SMEs introducing e-commerce in different types of value chains, more work is needed to extract benefits and verify best practices. Such validation can only be achieved over time, when the effects of e-commerce solutions can be compared with anticipated impacts and benefits.

REFERENCES

Baldock, R. (1999). *The last days of the giants? A route map for business survival.* New York: John Wiley & Sons.

Benjamin, R. & Wigand, R. (1995). Electronic markets and virtual value chains on the Information Superhighway. *Sloan Management Review,* (Winter), 62-72.

Chen, J. & Williams, B. (1998). The impact of EDI on SMEs: Summary of eight British case studies. *Journal of Small Business Management, 36*(4), 68-72.

Chu, P. (1995). Conceiving strategic systems: What are critical value activities…and how can they help your company? *Journal of Systems Management,* (July/August), 36-41.

Fawcett, S.E. & Magan, G.M. (2001). *Achieving world-class supply chain alignment: Benefits, barriers and bridges.* Center for Advanced Purchasing Studies. Retrieved from *www.capsresearch.org*

Hawkins, R. (2001). The business model as a research problem in electronic commerce. *STAR Issue Report No.4.* Retrieved from *www.databank.it/star*

Iacovou, C., Benbasat, I. & Dexter, A. (1995). Electronic data interchange and small organisations: Adoption and impact of technology. *MIS Quarterly, 19*(4).

Jeffcoate, J., Chappell, C. & Feindt, S. (2004). Assessing the impact of electronic commerce on SMEs in value chains: A qualitative approach, In N. Al-Qirim (Ed.), *Electronic commerce in small to medium-sized enterprises: Framework, issues and implications* (pp. 180-198). Hershey, PA: Idea Group Publishing.

Lockett, N. & Brown, D. (2001). A framework for the engagement of SMEs in e-business. *Proceedings of the Americas Conference on Information Systems.*

Poon, S. & Swatman, P. (1997). Small business use of the Internet: Findings from Australian case studies. *International Marketing Review, 14*(5), 385-402.

Porter, M.E. (1984). *Competitive advantage*. New York: The Free Press.

Rayport, J. & Sviokla, J. (1995). Exploiting the virtual value chain. *Harvard Business Review*, (November-December), 75-85.

Ward, J. & Griffiths, P. (1996). *Strategic planning for information systems* (2nd edition). Chichester: John Wiley & Sons.

Yin, R. (1994). *Case study research: Design and methods* (2nd edition). Thousand Oaks, CA: Sage Publications.

Yli-Renko, H. & Autio, E. (1998). The network embeddedness of new technology-based firms: Developing a systematic evolution model. *Small Business Economics, 11*, 253-267.

KEY TERMS

Buyer-Driven Value Chain: The customer states what he or she wants and sets out the terms and conditions that the supplier should meet.

Critical Value Activities: The *value activities* that an organisation must execute satisfactorily to ensure successful performance.

Customer-Centric Value Chain: The seller tailors its products to meet fast-changing consumer needs.

Fragmented Value Chain: Neither the *buyer-driven* nor the *seller-driven* model of the *value chain* dominates.

Industry Value Chain: Consists of the *organisational value chain*, together with the *value chains* of the organisation's competitors, suppliers, and customers. It represents the movement of goods and services from the source of raw materials through to the final customer.

Organisational Value Chain: The internal *value chain* of a company.

Seller-Driven Value Chain: The seller presumes to know what the market might want to buy.

Value Activity: A physically and technologically distinct activity that the company performs and which adds value to a product or service.

Value Activity Interaction (VAI): Interaction between two value activities performed by partners in a value chain.

Value Chain: A collection of activities that are performed by a company to design, produce, market, deliver, and support its product.

Advanced Techniques for Object-Based Image Retrieval

Y. J. Zhang
Tsinghua University, Beijing, China

INTRODUCTION

Along with the progress of imaging modality and the wide utility of digital images (including video) in various fields, many potential content producers have emerged, and many image databases have been built. Because images require large amounts of storage space and processing time, how to quickly and efficiently access and manage these large, both in the sense of information contents and data volume, databases has become an urgent problem. The research solution for this problem, using content-based image retrieval (CBIR) techniques, was initiated in the last decade (Kato, 1992). An international standard for multimedia content descriptions, MPEG-7, was formed in 2001 (MPEG). With the advantages of comprehensive descriptions of image contents and consistence to human visual perception, research in this direction is considered as one of the hottest research points in the new century (Castelli, 2002; Zhang, 2003; Deb, 2004).

Many practical retrieval systems have been developed; a survey of near 40 systems can be found in Veltkamp (2000). Most of them mainly use low-level image features, such as color, texture, and shape, etc., to represent image contents. However, there is a considerable difference between the users' interest in reality and the image contents described by only using the above low-level image features. In other words, there is a wide gap between the image content description based on low-level features and that of human beings' understanding. As a result, these low-level feature-based systems often lead to unsatisfying querying results in practical applications.

To cope with this challenging task, many approaches have been proposed to represent and describe the content of images at a higher level, which should be more related to human beings' understanding. Three broad categories could be classified: synthetic, semantic, and semiotic (Bimbo, 1999; Djeraba, 2002). From the understanding point of view, the semantic approach is natural. Human beings often describe image content in terms of objects, which can be defined at different abstraction levels. In this article, objects are considered not only as carrying semantic information in images, but also as suitable building blocks for further image understanding.

The rest of the article is organized as follows: in "Background," early object-based techniques will be briefly reviewed, and the current research on object-based techniques will be surveyed. In "Main Techniques," a general paradigm for object-based image retrieval will be described; and different object-based techniques, such as techniques for extracting meaningful regions, for identifying objects, for matching semantics, and for conducting feedback are discussed. In "Future Trends," some potential directions for further research are pointed out. In "Conclusion," several final remarks are presented.

BACKGROUND

Early Object-Based Techniques in Content-Based Image Retrieval

CBIR techniques are distinguished from traditional retrieval techniques by many aspects. Two of the most pertinent are that CBIR is a somehow subjective process, as for a given image, its means may have different interpretations for different users; and image retrieval is often a computationally expensive process, as the image database is often large in size and contains heterogeneous information. Due to these particular aspects, the results of CBIR could not be judged objectively—human perception should be considered. In addition, performing an exhaustive search for finding optimal solutions in CBIR is not feasible, and therefore, some suboptimal solutions will be chosen.

Because of the unique aspects of CBIR, object-based representation and description must be used even in so-called low-level feature-based image retrieval, though in these works, object recognition is not evidently performed and semantic information is not explicitly searched.

One typical example is in shape-based retrieval, as the shape features are generally extracted from individual objects (Latecki, 2002). In contrast, color features and textural features are often obtained by taking the whole image as a unit. From this point of view, shape-based retrieval is already at some higher level than color-based retrieval and texture-based retrieval (Zhang, 2003).

Structural query model is another instance in which partial matches are allowed and outputs related to the score of similarity can be provided. This type of retrieval is based on the relations between the individual objects and components in images (Zhou, 2001). In query by visual sketch, users sketch a scene by drawing a collection of objects. (It is assumed that these objects could fully define a scene.) For example, the objects are first identified and then used in a search (Chang, 1998).

Current Object-Based Techniques in Content-Based Image Retrieval

Currently, researchers seek explicit semantics and use the high-level descriptions that are common to humans, such as articles, people, places, and things. It is generally accepted that high-level features are crucial to improve the performance of CBIR up to so-called semantic-based querying. For this purpose, object-based content analysis, especially segmentation that segments the semantically meaningful objects from images, is an essential step (Zhang, 2001).

Complete image understanding should start at interpreting image objects and their relationships. Objects can be further identified in line with appropriate knowledge. For example, some object grammars based on rules for concept inference have been proposed (Petkovic, 2003). When domain knowledge is available, objects can be classified even without the explicit determination of object regions (Li, 2002b).

To extract high-level descriptions from images and to fill the gap between the low-level features and human beings' understanding of image contents, techniques to describe the whole image with a hierarchical structure to reach progressive image analysis are proposed (Castelli, 1998; Jaimes, 1999; Hong, 1999). The contents of images can be represented in different levels (Amir, 1998), such as the three-level content representation, including feature level content, object level content, and scene level content (Hong, 1999); and the five-level representation, including region level, perceptual region level, object part level, object level, and scene level (Jaimes, 1999). The problem here is how to implement these levels efficiently and effectively.

Another direction for extracting semantics information from an image is to map low-level visual features to high-level semantics. In other words, to fill the semantic gap, one makes the retrieval system work with low-level features, while the user puts in more high-level knowledge (Zhou, 2002). Two typical methods are to optimize query requests by using relevance feedback and semantic visual templates (Chang, 1998) and to interpret progressively the content of images by using interactive interfaces (Castelli, 1998). In both approaches, relevance feedback plays an important role, as humans are much better than computers at extracting semantic information from images (Rui, 1998; Ciocca, 1999).

MAIN TECHNIQUES FOR OBJECT-BASED IMAGE RETRIEVAL

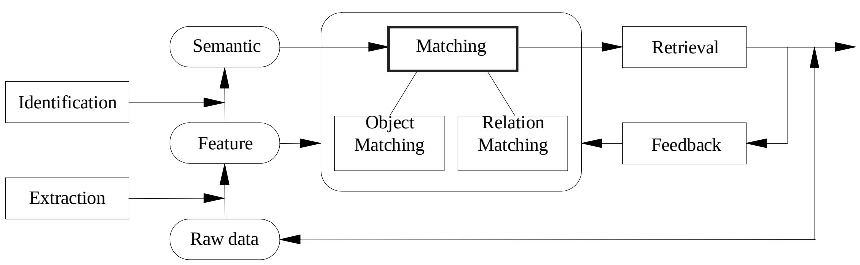
A General Paradigm

In general, people distinguish three levels of abstraction when talking about image databases: raw data level, feature level, and semantic level. The raw data are original images in the form of a pixel matrix. The feature level shows some significant characteristics of the pixel patterns of the image. The semantic level describes the meanings of identified objects in images. Note that the semantic level should also describe the meaning of an image as a whole. Such a meaning could be obtained by the analysis of objects and the understanding of images.

According to the above discussions, a multilayer approach should be used for efficiently treating image data. Though the number of layers and the definitions and functions of these layers could have some variations in different approaches, some principle steps are common for object-based image retrieval. A general paradigm is shown in Figure 1.

First, objects should be determined. Two important tasks are as follows:

Figure 1. A general paradigm for object-based image retrieval



1. Extract meaningful regions: To be able to base the image retrieval on objects, the interesting regions related to objects should be extracted first. This process relates the raw data level to the feature level.
2. Identify interesting objects: Based on the extracted regions, (perceptual) features should be taken out, and those required objects could be identified. This corresponds to the step from feature level to object level.

Once the objects in images are identified, further retrieval can be carried out by using objects as primitives. Two tasks are as follows:

1. Matching identified objects: For each identified object, suitable properties and attributes should be selected for proper description. The matching between images, such as object matching and relation matching, is then carried out.
2. Performing feedback in retrieval: This is to introduce human intelligence and to incorporate human semantics into the retrieval process.

The following sections introduce some techniques developed for each task.

Techniques for Extracting Region of Interest

Extraction of interesting regions from an image is, in general, called image segmentation (Zhang, 2001). Image segmentation is one of the most critical tasks in automatic analysis of image contents. A great variety of segmentation algorithms has been proposed in the literature. One should note that none of the proposed segmentation algorithms is generally applicable to all images, and different algorithms are not equally suitable for a particular application. This is the reason that though several thousands of algorithms have been developed, much attention and new efforts are continuously made on improving and perfecting them.

With the progress in segmentation techniques, people also realized that precise segmentation of objects in many cases is still beyond the capability of current computer techniques. On the other side, compared to some image analysis tasks that aim to obtain accurate measurements from the segmented objects, the requirement for precise segmentation of objects can be somehow relaxed in the context of image retrieval. Image retrieval is a subject-oriented process in which the precise object measurement is not the goal. In addition, for object-based image retrieval, the purpose of segmentation is for identifying the objects.

One idea derived from the above considerations is to extract approximately the so-called "meaningful region," instead of to segment the object accurately (Luo, 2001). The "meaningful region" provides an effective visual representation of objects from the point of view of object recognition. Though the "meaningful region" is not an exact representation of the objects, however, based on some domain knowledge, the semantic meaning of objects can still be recovered. On the other side, robust extraction of "meaningful regions" is easy to accomplish. This makes the object-based image retrieval with the extraction of "meaningful region" a feasible approach (Gao, 2000).

Another idea derived from the above considerations uses a particular matching procedure to reduce the requirement for precise object segmentation (Dai, 2004). The images are segmented both in a rough version and in a detailed version. The rough one is the merge result of several detailed ones and is less spatial-constrained by either oversegmentation or undersegmentation.

Techniques for Identifying Objects

From extracted regions, some perceptual features could be obtained. This can help to "recognize" what they represent in terms of human beings' perceptions. One iterative procedure uses this principle and transforms the recognition to a training-and-testing procedure (Gao, 2000). To make the problem simpler, it is assumed that there are finite types of interesting objects in a given image database. In fact, this requirement could often be satisfied in practice, as only limited image contents are considered in one application. The object recognition is then performed in an iterative way. Context-based information would be obtained during this process, helping to reach the correct recognition result.

To capture different aspects of images, multiple features are often used. The proportion of each feature in a description would be determined by training. Due to the lighting conditions and variety of object appearance, the objects belonging to the same category can have different visual aspects. To solve this problem, the most significant objects in the training set would be selected, and the trivial ones would be discarded. This recognition process can be iterated until the final recognition result is acceptable (in terms of the image composition, according to some *a priori* knowledge). With this iterative procedure, the context-based knowledge is gradually improved, and the correct recognition result is gradually approached.

Techniques for Matching Objects

Matching is one important task in image retrieval, which consists of comparison and judgment. As retrieval is a subjective process, so the decision is often made according to similarity (the distance to be reasonably small) but not according to equivalence (identical). Based on matching score, a database search can be carried out, and required images can be retrieved. Object match is more direct than feature match is in image retrieval.

One procedure is to describe the object in an image by an $M \times M$ matrix, where M is the number of all objects in the image database. It is a diagonal matrix, with each entry indicating the attribute of every meaningful region in the image. In addition, a relation matrix is also defined. It is a $K \times K$ matrix to indicate the spatial relationship between every two meaningful regions, with K representing the number of meaningful regions in the whole image. The object-matching procedure is dependent on a decision function. This decision function is determined based on the correlation among all content description matrices. It reflects whether all the relevant images have common contents, that is, the same objects. In the case that the relevant images have common objects, the match will be based on the objects of images. In the case that the relevant images do not have common objects, the match will be based on the features of images. For object matching, the similarity information from the common objects in all relevant images will be extracted to perform the matching between the relevant images and candidate images. Details can be found in Zhang (2004).

For compounded objects, matching might be accidentally performed among different parts. To solve this problem, a two-level matching is proposed (Dai, 2004). The principal idea is to describe the query images at a relatively rough scale and to describe the database images at some more detailed scales. As the rough description is based on the merging of detailed descriptions, the matching process will be carried on in an uneven way, and the minor errors caused by segmentation will be recompensed by the approximate matching procedure.

Techniques for (Interactive) Feedback

Feedback plays an important role, especially in high-level retrieval. As indicated above, retrieval is a subjective process, so feedback is required to combine the information from users, or in other words, to incorporate human knowledge and requirements. Retrieval is also a progressive process, so feedback is required to introduce interaction and to turn the search direction to follow the user's intention.

A self-adaptive relevance feedback technique has been used in an object-based image retrieval system (Gao, 2001). In such a system, objects are first identified, the relevance feedback relying on the high-level attributes could better catch image semantics, and the retrieval results are refined according to users' wishes in an explicit manner. In practice, to make the querying more convenient for the user, the procedure of feedback could be directed, also based on high-level information, without memory or with memory to make the feedback mechanism more flexible. In the former case, each feedback is an independent procedure, in which all of the relevant images selected in previous iterations would be ignored. In the latter case, the relevant image selected in previous iterations would be taken into account in the current iteration. A time-delay curve has also been proposed to simulate human beings' memory mechanisms in feedback with memory (Gao, 2001). The main idea of the proposed scheme is to analyze the fed-back relevant images marked by the user in different levels to reach comprehensive similarity analysis.

Another approach called association feedback has been proposed (Xu, 2001b). Feature elements are first defined that can be considered a type of perceptual primitives with abstraction levels located between that of raw images and that of objects in images (Xu, 2001a). These feature elements, different from commonly used feature vectors, have obvious intuitive visual senses and are relatively independent from each other physically. A selection mechanism called feature element evaluation is also proposed, which tries to find those feature elements that are closer to the interest of people by visual meaning. A group of feature elements can be used to represent compound objects. Based on feature elements, association feedback can be applied. In contrast to the weighting adjustment in relevance feedback, here the associated relations between different feature elements are counted. New sets of feature elements can thus be formed during the retrieval process, and this property is suitable for handling the so-called "interest switch" cases. In other words, the search for images can be controlled by users with the introduction of new feature elements according to the change of interest, and the search direction will be guided toward new goals.

FUTURE TRENDS

Further research can be considered in the concrete techniques for advanced image analysis and along the general research movements for image understanding.

To perform object-based image retrieval, different image analysis techniques are to be enhanced:

1. Improving the robustness of meaningful region extraction, especially with complicated images, by taking more characteristics of images into consideration

2. Describing objects as congruous to humans' sense as possible—as human beings are still far from knowing all the cognitive details from the real world, how to automatically form semantic objects is a challenging task
3. Using more efficient feedback procedures to make the search process fast following users' aspirations in the course of retrieval

In the "Background" section, two generations of object-based techniques are discussed. From the point of view of image understanding, the next stage would go beyond objects, though the third generation will still be based on objects. The actions and interactions of objects and thus generated events (or scenes) are important to fully understand the contents of images. The images would be, in this case, described by some metadata. The event detection and event retrieval have already played an important role in many applications (e.g., surveillance, war, etc.). However, only a few particular works are made now, and they are mainly based on audio and multiple frames (Li, 2002a). Further research in this direction, taking more advantage of human knowledge for constructing more intelligent image data would be promising.

CONCLUSION

Object-based techniques can fulfill many roles and tasks required by CBIR. Three generations of object-based techniques are reviewed and discussed in this article. Some of them have already made their contributions to the advancement of CBIR, and some of them need to be improved, and developing new object-based techniques for CBIR is even required.

The object-based techniques discussed here are mainly focused on CBIR. As content-based video retrieval (CBVR) appears like a natural combination of CBIR and content-based audio retrieval (CBAR), as well as some extensions along the temporal axis, many of these techniques would also be applicable for CBVR.

REFERENCES

Amir, A., & Lindenbaum, M. (1998). A generic grouping algorithm and its quantitative analysis. *IEEE PAMI, 20*(2), 168–185.

Bimbo, A. (1999). *Visual information retrieval*. San Francisco, CA: Morgan Kaufmann (Elsevier).

Castelli, V., Bergman, L. D., & Kontoyiannis, I. et al. (1998). Progressive search and retrieval in large image archives. *IBM J. Res. Develop.*, *42*(2), 253–268.

Chang, S. F., Chen, W., & Sundaram, H. (1998). Semantic visual templates: Linking visual features to semantics. In *Proceedings ICIP'98* (pp. 531–535).

Ciocca, G., & Schettini, R. (1999). Using a relevance feedback mechanism to improve content-based image retrieval. In *Proceedings of the Third International Conference, VISUAL'99* (pp. 107–114).

Dai, S. Y., & Zhang, Y. J. (2004). Unbalanced region matching based on two-level description for image retrieval. *Pattern Recognition Letters*.

Deb, S. (2004). *Multimedia systems and content-based image retrieval*. Hershey, PA: Idea Group Publishing.

Djeraba, C. (2002). Content-based multimedia indexing and retrieval. *IEEE, Multimedia*, (2), 18–22.

Gao, Y. Y., Zhang, Y. J., & Merzlyakov, N. S. (2000). Semantic-based image description model and its implementation for image retrieval. In *Proceedings of the First International Conference on Image and Graphics* (pp. 657–660).

Gao, Y. Y., Zhang, Y. J., & Yu, F. (2001). Self-adaptive relevance feedback based on multi-level image content analysis. In *SPIE Proceedings Storage and Retrieval for Media Databases 2001* (Vol. 4315, pp. 449–459).

Hong, D. Z., Wu, J. K., & Singh, S. S. (1999). Refining image retrieval based on context-driven method. In *SPIE Proceedings Storage and Retrieval for Image and Video Database VII* (Vol. 3656, pp. 581–593).

Jaimes, A., & Chang, S. F. (1999). Model-based classification of visual information for content-based retrieval. In *SPIE Proceedings on Storage and Retrieval for Image and Video Database VII* (Vol. 3656, pp. 402–414).

Kato, T. (1992). Database architecture for content-based image retrieval. *SPIE* (Vol. 1662, pp. 112–123).

Latecki, L. J., Melter, R., & Gross, A. (2002). Shape representation and similarity for image database [Special Issue]. *Pattern Recognition*, *35*(1), 1–297.

Li, B. X., & Sezan, I. (2002a). Event detection and summarization in American football broadcast video. In *SPIE Proceedings Storage and Retrieval for Media Databases 2002* (Vol. 4676, pp. 202–213).

Li, Q., Zhang, Y. J., & Dai, S. Y. (2002b). Image search engine with selective filtering and feature element based classification. In *SPIE Proceedings Internet Imaging III* (Vol. 4672, pp. 190–197).

Luo, Y., Zhang, Y. J., & Gao, Y. Y. et al. (2001). Extracting meaningful region for content-based retrieval of image and video. In *SPIE, Proceedings Visual Communications and Image Processing* (Vol. 4310, pp. 455–464).

MPEG. Retrieved from *http://www.cselt.it/mpeg/*

Petkovic, M., & Jonker, W. (2003). *Content-based video retrieval: A database perspective*. Dordrecht: Kluwer.

Rui, Y., Huang, T. S., & Mehrotra, S. (1998). Relevance feedback techniques in interactive content-based image retrieval. In *SPIE Proceedings Storage and Retrieval for Image and Video Database V* (Vol. 3312, pp. 25–34).

Veltkamp, R. C., & Tanase, M. (2000). Content-based image retrieval systems: A survey. Technical Report, UU-CS-2000-34, Utrecht University, The Netherlands.

Xu, Y., & Zhang, Y. J. (2001a). Image retrieval framework driven by association feedback with feature element evaluation built in. In *SPIE Proceedings Storage and Retrieval for Media Databases 2001* (Vol. 4315, pp. 118–129).

Xu, Y., & Zhang, Y. J. (2001b). Association feedback: A novel tool for feature elements based image retrieval. In *Lecture Notes in Computer Science 2195* (pp. 506–513).

Zhang, Y. J. (2001). *Image segmentation*. Beijing: Science Publisher.

Zhang, Y. J. (2003). *Content-based visual information retrieval*. Beijing: Science Publisher.

Zhang, Y. J., Gao, Y. Y., & Luo, Y. (2004). Object-based techniques for image retrieval. In *Multimedia systems and content-based image retrieval* (Chap. 7, pp. 154–179). Hershey, PA: Idea Group Publishing.

Zhou, X. S., & Huang, T. S. (2001). Edge-based structural features for content-based image retrieval. *Pattern Recognition Letters, 22*(5), 457–468.

Zhou, X. S., & Huang, T. S. (2002). Unifying keywords and visual contents in image retrieval. *IEEE, Multimedia*, (2), 23–33.

KEY TERMS

Content-Based Image Retrieval (CBIR): A process framework for efficiently retrieving images from a collection by similarity. The retrieval relies on extracting the appropriate characteristic quantities describing the desired contents of images. In addition, suitable querying, matching, indexing, and searching techniques are required.

Feature-Based Image Retrieval: A branch of CBIR that is based on specific visual characteristics called "features" and is considered at a low abstraction level. Features are commonly refereed to perceptive attributes of images, such as color, texture, shape, etc., of images.

Intelligent Image Data: A data format that embeds pixel information of images as well as higher-level information, such as indices and semantic information. This format is self-descriptive in the sense that data could explain by themselves what contents are inside and present and retrieve the related and interested portions for the users.

Metadata: A data format that may contain numerous information, including information obtained indirectly from the image, as well as information related to the actual description of the image content. At the highest level, images are often accompanied and associated by metadata.

MPEG-7: This is an international standard named "multimedia content description interface" (ISO/IEC 15938). It provides a set of audiovisual description tools, descriptors, and description schemes for effective and efficient access (search, filtering, and browsing) to multimedia content.

Query Model: An abstraction model for image querying in the context of CBIR. In this model, a submitted query would specify both a filter condition and a ranking expression, while the query result will be a rank of the images that satisfies the filter condition, according to grade of match for the ranking expression.

Semantic-Based Image Retrieval: A branch of CBIR based on descriptions with semantic meaning and considered at a high abstraction level. Semantic descriptions are more closely related to the human interpretation and understanding of images.

Semantic Gap (SG): The discrepancy between the perceptual property and semantic meaning of images in the context of CBIR. As the perceptual properties are usually described by low-level visual features that can be easily treated by computers, and the semantic meanings are commonly related to high-level object-based descriptions that are familiar to human beings, the semantic gap is also considered a gap between current techniques and human requirements.

Semiotics: The science that analyzes signs and sign systems and puts them in correspondence with particular meanings. It provides formal tools for image knowledge acquisition, generation, representation, organization, and utilization in the context of CBIR.

Agent- and Web-Based Employment Marketspaces in the U.S. Department of Defense

William R. Gates
Naval Postgraduate School, USA

Mark E. Nissen
Naval Postgraduate School, USA

INTRODUCTION

Two modes of matching people with jobs prevail at present: hierarchical planning and distributed markets. Patterned after centrally planned (e.g., former Soviet-style) economies, hierarchical planning remains prevalent for matching job candidates to jobs internally *within* an organization. With evolving information technology, the internal job-matching process could be accomplished far more equitably and efficiently using Web-based markets. Intelligent agents offer excellent potential to help both employees and employers find one another in a distributed electronic marketspace. But realizing this potential goes well beyond simply changing the rules of internal job matching or making agent technology available to job searchers. Rather, the corresponding markets and technologies must be designed together to mutually accomplish the desired results (e.g., efficient and effective matching) and conform to necessary properties (e.g., market clearing, employer policies).

BACKGROUND

As an example, the U.S. Navy currently uses a centralized, hierarchical labor market to match enlisted sailors to jobs (U.S. Navy, Bureau of Naval Personnel, 2000; U.S. Navy, Commander in Chief, United States Pacific Fleet, n.d.). Navy commands (e.g., ships, bases) identify open positions, and the Navy prioritizes job vacancies based on each command's mission, current staffing levels and several other relevant characteristics. Available sailors are categorized according to their qualifications, including skills (ratings), experience, education/training, career path, and so forth. Similar skill groups are arranged in "communities". Each community has a detailer charged with matching sailors to jobs. Sailors seeking job assignments express their personal preferences to the detailer. The detailer is responsive to job priority ratings, but there is some room for discretion to tailor job assignments to the sailors' personal preferences. This centralized detailing process reassigns approximately one-third of the enlisted force every year (i.e., 100,000 sailors). This labor-intensive process often leaves both sailors and commands dissatisfied.

Competitive labor markets can be used as the model to identify the desirable characteristics of internal labor markets. Competitive labor markets are both effective and efficient (Ehrenberg & Smith, 1997). They are effective because markets clear. The wage rate adjusts to its market clearing value, where the quantity of labor that employers hire willingly exactly equals the quantity of labor that employees supply willingly. Competitive labor markets are also efficient. On the demand side, markets allocate labor to its highest valued uses. As the market wage increases, lower valued uses drop out while higher valued uses remain filled. On the supply side, markets ensure that the employees hired are the most willing to work in the industry. As market wages decrease, individuals less willing to work voluntarily leave; those most attracted to the industry remain. Thus competitive labor markets balance labor supply and demand, and ensure that the most valuable jobs are filled with the most willing employees.

Unfortunately, the information requirements to ensure effectiveness and efficiency are extensive. To operate efficiently, the assignment process must have complete information regarding the employees' preferences over available jobs, and the employers' preferences over potential employees. To establish their preferences, employees must have complete information about all relevant job opportunities, including salary, benefits and job amenities (e.g., work environment and content, promotion potential, commute), and employers must have complete information about the employees' relevant job qualifications. If labor assignments do not factor in this information, the system wastes labor by applying it to less valuable jobs, and it reduces job satisfaction, morale and retention by assigning labor to jobs that are relatively less desirable. Clearly, some kind of technology could help people manage the required abundance of labor-market information.

INTELLIGENT AGENT TECHNOLOGY

Work in the area of software agents has been ongoing for some time, and it addresses a broad array of applications. Building upon research in the supply chain domain (Mehra & Nissen, 1998; Nissen & Mehra, 1998; Nissen, 2000), agent technology appears to have particular promise to automate and support electronic labor markets. As computational artifacts, they can help overcome human cognitive limitations (e.g., in terms of memory and processing speed), supporting rapid search and effective filtering through huge numbers of available jobs and potential employees. Further, agents possessing artificial intelligence (AI) can employ inferential mechanisms (e.g., rules, cases, scripts) to reflect and observe diverse individuals' preferences. In a domain with over a 100,000 available jobs and prospective employees (e.g., the Navy)—in which both employers and potential employees have specific, idiosyncratic needs and preferences—no other extant information technology offers the same level of automation and support capability as software agents. Here we discuss representative, extant agent technologies, and outline key capabilities of the Personnel Mall, a proof-of-concept multi-agent system developed to enact, automate and support an electronic employment market.

Extant Agent Applications

Following the literature survey and classification system of Nissen (2000), Table 1 summarizes extant agent applications into four classes: 1) information filtering agents, 2) information retrieval agents, 3) advisory agents, and 4) performative agents. Other groupings from the agents' literature could be used as well (e.g., Bradshaw, 1997; Franklin & Graesser, 1996; Nwana, 1996; White, 1997), but the classification scheme used here is useful to compare agent capabilities applicable to markets and matching processes.

Information filtering agents apply user-input preferences passively to screen and sort e-mail, network news groups, frequently asked questions and arbitrary text. Information retrieval agents collect information pertaining to commodities, such as compact disks and computer equipment, and services, such as advertising and insurance. This class also includes Web indexing robots and Web-based agents for report writing, publishing, assisted browsing, and so forth. Advisory agents provide intelligent advice and decision-support (e.g., CD and movie recommendations, electronic concierges, agent "hosts" for college campus visits, etc.). These agents also match buyers with sellers directly in market and matching domains. Agents in this third class decide what information is needed, and seek out and use this information to make recommendations. Finally, performative agents change the state of the external world through autonomous, deliberate action (e.g., binding commercial transactions). Performative agents include marketspaces in which agents conduct business transactions, auction environments in which agents buy and sell for their users, and several agent system designs for negotiation. Performative agents also automate knowledge work, such as scheduling, autonomously provide a cooperative-learning environment and provide digital library services.

The Personnel Mall

As noted above, the Personnel Mall is a proof-of-concept multi-agent system developed to enact, automate and support a Web-based marketspace for employee/job matching (Gates & Nissen, 2001). Like its predecessor system, the Intelligent Mall (which was developed for matching buyers with vendors in products and services markets), the Personnel Mall employs a shopping mall metaphor for employee-job matching. In a mall, shoppers are not expected to know in advance which shops exist or what products they offer for sale. Similarly, shops are not expected to know which other shops are selling like products or with which shoppers they will interact. But these agent-enabled marketspaces provide a qualitative contrast to their physical marketplace counterparts; that is, instead of *people* searching, matching, buying and selling within physical markets, such market and matching activities are performed by *software agents* representing people (i.e., the sailors and commands). The agents—implemented through software objects and methods—exchange messages with one another to communicate and coordinate their activities.

The Personnel Mall is probably categorized best as a performative agent. It can represent a variety of different users—on both the demand and supply sides—to find quickly, retrieve and organize large amounts of market information. Its conformance to market and organizational rules, established for a particular enterprise or circumstance, enables this multi-agent system to automate and support commerce in a broad diversity of electronic markets, including regulation-laden, hierarchical systems. Such ability suggests the Personnel Mall offers good potential to enact, automate and support the kinds of electronic labor markets addressed through this research.

In the Navy's assignment context, command agents can be specialized to reflect the preferences and priorities

Table 1. Agent taxonomy (Adapted from Nissen, 2000)

Information Filtering Agent Applications	
Filter e-mail messages	Maes, 1994; Malone et al., 1987
Filter network newsgroup postings	Sycara & Zeng, 1996
Filter frequently asked questions (FAQs)	Whitehead, 1994
Filter arbitrary text messages	Verity, 1997
Information Retrieval Agent Applications	
Collect product/service information	Krulwich, 1996; Insurance, 1997; PriceWatch, 1997; uVision, 1997
Web robots & publication tools	Amulet, 1997; Chen et al., 1998; Etzioni & Weld, 1995; InterAp, 1995
Assisted Web browsing	Burke et al., 1997
Advisory Agent Applications	
Recommend compact discs & movies	Maes, 1997; Nguyen & Haddawy, 1999
E-concierge services	Etzioni & Weld, 1995
Campus visit "host"	Zeng & Sycara, 1995
Planning support	Maturana & Norrie, 1997; Pinson et al., 1997
Project coordination advice	Johar, 1997
Computer interface assistance	Ball et al., 1997
Military reconnaissance support	Bui et al., 1996
Financial portfolio advice	Sycara et al., 1996
Buyer/seller matchmaking advice	Freuder & Wallace, 1999
Supply chain decision support	Goodwin et al., 1999
Performative Agent Applications	
Business marketspace	Chavez & Maes, 1996; Fox and Barbuceanu, 2000; Hu et al., 1999; Mehra & Nissen, 1998; Nissen & Mehra, 1998; Preece et al., 1999
Auction marketspace	Hu et al., 1999; Rodriguez-Aguilar et al., 1998; Sandholm, 1999
Agent negotiation	Bui, 1996; Collins et al., 1998; Guttman et al., 1998; Maes et al., 1999; Sun et al., 1999; Tesauro & Kephart, 2000
Scheduling	Sen, 1997; Walsh et al., 1998
Cooperative learning	Boy, 1997
Digital library services	Mullen & Wellman, 1996

of their principal (e.g., the Commanding Officer). For example, a command agent can be specialized to search and match based on one officer's preferences for education and training over rank and experience, whereas another agent can be specialized to reflect different (even opposing) preferences. Each command agent can be instantiated with a unique shopping list of job openings. Other knowledge and information—such as user preferences, budget restrictions, personnel requirements and need dates—can be formalized through rules for the agents.

Likewise, sailor agents can be designed by individual sailors to reflect their particular preferences. Like their command counterparts, sailor agents use messages to communicate individual sailors' attributes (e.g., rank, education, training, job skills, availability, community), and sailors can search the Mall to view a complete, current listing of job openings. Sailors use such listings to identify a specific subset of assignments in which they are particularly interested, and they use an input form to convey their relative preferences among this subset of job assignments.

With these capabilities, the Personnel Mall can be used at present to enact, automate and support search and matching in electronic labor markets, just as intelligent malls perform such functions for products and services along the supply chain. Relative to archaic centralized planning systems or crude information retrieval "bots" used in labor markets today, this innovative application of agent capability is expected to improve performance considerably (e.g., in terms of increased speed, greater information gathering, improved preference matching). For instance, the Personnel Mall can perform automatically most process activities required to match sailors with jobs—through its federation of intelligent agents—which eliminates the need for an organization of detailing personnel and associated staff. To provide a sense of magnitude, the Navy currently employs roughly 450 detailers and support personnel to perform distribution and assignment activities; some $30M could be saved by implementing a multi-agent application such as the Personnel Mall (Schlegel, 2000).

Additionally, Mall agents perform their search and matching tasks very quickly, gathering and considering substantial information in the process. Agents can also handle information overload, effectively processing much larger sets of job/sailor alternatives than their human counterparts. This suggests the lead-time required for assigning sailors to jobs can be reduced, perhaps to an interval so small it approaches just-in-time distribution and assignment. Alternatively, the period between assignments could be extended, simultaneously matching much larger groups of sailors and jobs, to improve the quality of fit. In either case, an intelligent multi-agent system eliminates many constraints currently limiting the detailing process (i.e., process time, information overload). Future analysis will determine the detailing process design that exploits these improvements best.

Redesigning the detailing process offers additional benefits. The Personnel Mall makes all job information available to sailors—and all sailor information available to commands—so personnel can search and rank alternatives from the complete listing of available Navy jobs. This addresses a common complaint from sailors: detailers fail to tell them about more than a handful of the job opportunities that the Navy needs filled. Exit interviews with sailors leaving the Navy suggest the distrust stemming from such behavior represents a significant factor in the decision not to re-enlist (Schlegel, 2000; Short, 2000).

Another putative benefit involves sailors' relative job preferences. Because the Personnel Mall matches sailors with job assignments explicitly based on preference information, personnel are more likely to be assigned to their preferred commands. Ostensibly this promotes better job performance than expected from sailors assigned to less-preferred jobs. Better sailor job matches also offer the potential for improved personnel retention, providing a better return on the Navy's investment in sailors' education and training and increasing the Navy's overall level of experience. Clearly, increased personnel retention decreases recruiting demands.

FUTURE TRENDS

The Personnel Mall—and any similar emerging agent technology—is still experimental in nature, and not yet mature or ready for "industrial strength" application (e.g., to Navy-wide distribution and assignment). One limitation, as currently designed, is that the Personnel Mall is effectively limited to *one-sided matching*; job assignments can reflect either the sailors' or commands' preferences, but do not consider both simultaneously. This suggests strong potential for incorporating within the Personnel Mall either an optimization algorithm or two-sided matching market (as used in the U.S. medical residency market (Demange et al., 1985; Gale & Shapley, 1962; Roth, 1984; Roth & Sotomayor, 1990).

Research comparing sailor and command social welfare for human detailers (the *status quo*) versus optimization or two-sided matching suggests significant increases in social welfare by moving to an algorithm-based assignment process (Gates & Nissen, 2002, 2004). Experimental results indicate that a shift away from unassisted matching can increase labor welfare by up to 20% and total welfare by 10 -14%. According to the U.S. Department of Defense (DoD) FY2002 budget, the U.S. Navy, which is currently moving from a hierarchical to a more automated assignment process, spent over $20 billion on labor in FY2002; the DoD spent over $82 billion (U.S. DoD, 2002). If these comparative results hold as we move from an experimental setting to application in the U.S. Navy or DoD, the absolute increase in welfare could be dramatic. This would represent a dramatic gain in employee morale and retention, and a significant improvement in the efficiency with which such public-sector enterprises use taxpayer dollars. Moreover, to the extent that the kinds of electronic employment market designs discussed here can be extended further to the private sector, the potential economic impact in terms of social welfare is staggering.

CONCLUSION

Two modes of matching people with jobs prevail at present: 1) hierarchical planning and 2) distributed markets. Each has strengths and limitations, but few systems have been designed to take advantage of strengths corresponding to both. With evolving information technology, however, the job-matching process could be accomplished far more equitably and efficiently using Web-based markets within the firm. Intelligent agents offer excellent potential to help both potential employees and employers find one another in a distributed electronic marketplace. But realizing this potential goes well beyond simply changing the rules of internal job matching, or making agent technology available to job searchers. Rather, the corresponding markets and technologies must be designed together to accomplish mutually the desired results (e.g., efficient and effective matching) and conform to necessary properties (e.g., market clearing). This article integrates the key economic and technological elements required to design robust electronic employment markets. The corresponding research provides new knowledge and insight into co-development of the requisite economic markets and agent technologies.

There is clearly much research yet to be accomplished before robust, electronic employment markets are ready for "industrial strength" implementation in the enterprise. Following the lead of this discussion, such research should necessarily be multi-disciplinary and integrative in nature. As noted in the introduction, economic and technological elements associated with labor-market design are inextricably intertwined, and a robust design cannot be achieved without addressing both in an integrated manner. The research described in this article has taken a first step along these lines and hopefully will stimulate further research.

REFERENCES

Amulet. (1997). Amulet online description. *http://www.amulet.com*

Ball, G. et al. (1997). Lifelike computer characters: The persona project at Microsoft. In J. Bradshaw (Ed.), *Software agents*. Menlo Park, CA: AAAI Press.

Boy, G.A. (1997). Software agents for cooperative learning. In J. Bradshaw (Ed.), *Software agents*. Menlo Park, CA: AAAI Press.

Bradshaw, J. (Ed.). (1997). *Software agents*. Menlo Park, CA: AAAI Press.

Bui, T. (1996). *Intelligent negotiation agents for supporting Internet-based competitive procurement*. Working paper.

Bui, T., Jones, C., Sridar, S., & Ludlow, N. (1996). *Decision support for reconnaissance using intelligent software agents*. Monterey, CA: Naval Postgraduate School Research Proposal.

Burke, R.D., Hammond, K.J., & Young, B.C. (1997). The FindMe approach to assisted browsing. *IEEE Expert, 12*(4), 32-40.

Chavez, A., & Maes, P. (1996). Kasbah: An agent marketplace for buying and selling goods. *Proceedings of the First International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology*, London.

Chen, H., Chung, Y.M., Ramsey, M., & Yang, C.C. (1998). A smart itsy bitsy spider for the Web. *Journal of the American Society for Information Science and Technology, 49*(7), 604-618.

Collins, J., Youngdahl, B., Jamison, Sc., Mobasher, B., & Gini, M. (1998). A market architecture for multi-agent contracting. In K. Sycara & M. Wooldridge (Eds.), *Proceedings of the Second International Conference on Autonomous Agents*, Minneapolis, MN (pp. 285-292).

Demange, G., & Gale, D. (1985). The strategy structure of two-sided matching games. *Econometrica, 53*(4), 873-88.

Ehrenberg, R.G., & Smith, R.S. (1997). *Modern labor economics: Theory and public policy*. Reading, MA: Addison-Wesley.

Etzioni, O., & Weld, D.S. (1995). Intelligent agents on the Internet: Fact, fiction, and forecast. *IEEE Expert, 10*(4), 44-49.

Fox, M., & Barbuceanu, M. (2000). *The integrated supply chain management project*. Enterprise Integration Laboratory, Department of Industrial Engineering, University of Toronto. *http://www.eil.utoronto.ca/iscm-descr.html*

Franklin, S., & Graesser, A. (1996). Is it an agent or just a program? A taxonomy for autonomous agents. *Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages*. NY: Springer-Verlag.

Freuder, E.C., & Wallace, R.J. (1999). Matchmaker agents for electronic commerce. In T. Finin & B. Grosof (Eds.), *Proceedings of American Association for Artificial Intelligence*, Orlando, FL, Workshop on AI for Electronic Commerce.

Gale, D., & Shapley. (1962). College admissions and the stability of marriage. *American Mathematical Monthly, 69*(1), 9-15.

Gates, W.R., & Nissen, M.E. (2001). Designing agent-based electronic employment markets. *Electronic Commerce Research Journal: Special Issue on Theory and Application of Electronic Market Design, 1*(3), 239-263.

Gates, W.R., & Nissen, M.E. (2002). *Two-sided matching agents for electronic employment market design: Social welfare implications*. Technical Report NPS-GSBPP-02-005. Naval Postgraduate School, Monterey, CA.

Gates, W.R., & Nissen, M.E. (2004). Experimental analysis of e-employment market design. *Journal of Organiza-*

tional Computing And Electronic Commerce. Forthcoming.

Goodwin, R., Keskinocak, P., Murthy, S., Wu, F., & Akkiraju, R. (1999). Intelligent decision support for the e-supply chain. In T. Finin & B. Grosof (Eds.), *Proceedings of American Association for Artificial Intelligence*, Orlando, FL, Workshop on AI for Electronic Commerce.

Guttman, R.H., Moukas, A.G., & Maes, P. (1998). Agent-mediated electronic commerce: A survey. *Knowledge Engineering Review*. *http://ecommerce.media.mit.edu/*

Hu, J., Reeves, D., & Wong, H.S. (1999). Agent service for online auctions. In T. Finin & B. Grosof (Eds.), *Proceedings of American Association for Artificial Intelligence*, Orlando, FL, Workshop on AI for Electronic Commerce.

Hu, J., Yen, J., & Chung, A. (1999). A virtual property agency: Electronic market with support of negotiation. In T. Finin & B. Grosof (Eds.), *Proceedings of American Association for Artificial Intelligence*, Orlando, FL, Workshop on AI for Electronic Commerce.

Insurance. (1997). Insurance online description. *http://www.dmatters.co.uk*

InterAp. (1995, June 12). InterAp assigns intelligent agents to the Web. *PCWeek*.

Johar, H.V. (1997). SoftCord: An intelligent agent for coordination in software development projects. *Decision Support Systems, 20*(1), 65-81.

Krulwich, D. (1996). *An agent of change*. Andersen Consulting Center for Strategic Technology Research.

Maes, P. (1994). Agents that reduce work and information overload. *Communications of the ACM, 37*(7), 30-40.

Maes, P. (1997, July/August). Pattie Maes on software agents: Humanizing the global computer. *Internet Computing*.

Maes, P., Guttman, R.H., & Moukas, A.G. (1999, March). Agents that buy and sell. *Communications of the ACM, 42*(3), 81-87.

Malone, T.W., Yates, J., & Benjamin, R.I. (1987). Electronic markets and electronic hierarchies. *Communications of the ACM, 30*(6), 484-497.

Maturana, F.P., & Norrie, D.H. (1997). Distributed decision-making using the contract net within a mediator architecture. *Decision Support Systems, 20*(1), 53-64.

Mehra, A., & Nissen, M.E. (1998). Case study: Intelligent software supply chain agents using ADE. *Proceedings of AAAI Workshop on Software Tools for Developing Agents*.

Mullen, T., & Wellman, M.P. (1996). Market-based negotiation for digital library services. *Second USENIX Workshop on Electronic Commerce*.

Nguyen, H., & Haddawy, P. (1999, July). DIVA: Applying decision theory to collaborative filtering. In T. Finin & B. Grosof (Eds.), *Proceedings of American Association for Artificial Intelligence*, Orlando, FL, Workshop on AI for Electronic Commerce.

Nissen, M.E. (2000). Agent-based supply chain integration. *Journal of Information Technology Management*.

Nissen, M.E., & Mehra, A. (1998). Redesigning software procurement through intelligent agents. *Proceedings of AAAI Workshop on AI in Reengineering and Knowledge Management*.

Nwana, H.S. (1996). Software agents: An overview. *Knowledge Engineering Review, 11*(3), 205-244.

Pinson, S., Louca, J.A., & Moraitis, P. (1997). A distributed decision support system for strategic planning. *Decision Support Systems, 20*(1), 35-51.

Preece, A., Hui, K., & Gray, P. (1999, July). KRAFT: Supporting virtual organisations through knowledge fusion. In T. Finin & B. Grosof (Eds.), *Proceedings of American Association for Artificial Intelligence*, Orlando, FL, Workshop on AI for Electronic Commerce.

PriceWatch. (1997). PriceWatch online description. *http://www.pricewatch.com*

Rodriguez-Aguilar, J.A., Martin, F.J., Noriega, P., Garcia, P., & Sierra, C. (1998). Competitive scenarios for heterogeneous trading agents. In K. Sycara & M. Wooldridge (Eds.), *Proceedings of Second International Conference on Autonomous Agents*, Minneapolis, MN (pp. 293-300).

Roth, A.E. (1984). The evolution of the labor market for medical interns and residents: A case study in game theory. *Journal of Political Economy, 92*(6), 991-1016.

Roth, A.E., & Sotomayor, M.A.O. (1990). *Two-sided matching: A study in game-theoretic modeling and analysis*. Cambridge, UK: Cambridge University Press.

Sandholm, T. (1999). eMediator: A next generation electronic commerce server. In T. Finin & B. Grosof (Eds.), *Proceedings of American Association for Artificial Intelligence*, Orlando, FL, Workshop on AI for Electronic Commerce.

Schlegel, R.J., LCDR, USN. (2000). *An activity based costing analysis of the Navy's enlisted detailing process*. Masters thesis. Naval Postgraduate School, Monterey, CA.

Sen, S. (1997). Developing an automated distributed meeting scheduler. *IEEE Expert, 12*(4), 41-45.

Short, M.M., Lt., USN. (2000). *Analysis of the current Navy enlisted detailing process.* Masters thesis. Naval Postgraduate School, Monterey, CA.

Sun, R., Chu, B.T., Wilhelm, R., & Yao, J. (1999). A CSP-based model for integrated supply chains. In T. Finin & B. Grosof (Eds.), *Proceedings of American Association for Artificial Intelligence*, Orlando, FL, Workshop on AI for Electronic Commerce.

Sycara, K., & Zeng, D. (1996). Coordination of multiple intelligent software agents. *International Journal of Cooperative Information Systems.*

Sycara, K., Pannu, A., Williamson, M., & Zeng, D. (1996). Distributed intelligent agents. *IEEE Expert, 11*(6), 36-46.

Tesauro, G.J., & Kephart, J.O. (2000). Foresight-based pricing algorithms in an economy of software aAgents. IAC Reports, IBM. *http://www.ibm.com/iac/papers/ice98_fs/fs_public.html*

U.S. Department of Defense. (2002). *http://www.dtic.mil/comptroller/fy2002budget/*

U.S. Navy, Bureau of Naval Personnel Homepage. *www.persnet.navy.mil*

U.S. Navy, Commander in Chief, United States Pacific Fleet. *Enlisted Distribution System Overview.* Undated Power Point presentation.

uVision. (1998). *http://www.uvision.com/*

Verity. (1997). Verity online description. *http://www.verity.com*

Walsh, W.E., Wellman, M.P., Wurman, P.R., & MacKie-Mason, J.K. (1998). Some economics of market-based distributed scheduling. *Proceedings of International Conference on Multi-Agent Systems.*

White, J. (1997). Mobile agents. In J. Bradshaw (Ed.), *Software agents.* Menlo Park, CA: AAAI Press.

Whitehead, S.D. (1994) Auto-faq: An experiment in cyberspace leveraging. *Proceedings of the Second International WWW Conference, 1*, 25-38.

Zeng, D., & Sycara, K. (1995). *Cooperative intelligent software agents.* Technical report no. CMU-RI-TR-95-14. Carnegie Mellon University.

KEY TERMS

Advisory Agents: Intelligent agents that provide intelligent advice and decision support autonomously. These agents decide what information is needed, seek it out, and use it to make recommendations.

Information Filtering Agents: Intelligent agents that apply user-input preferences autonomously to screen passively and sort information.

Information Retrieval Agents: Intelligent agents that search for and collect information autonomously based on user prescribed criteria.

Intelligent Agents: Computer software that employs artificial intelligence to make decisions and perform actions autonomously.

Performative Agents: Intelligent agents that change the state of the external world through autonomous, deliberate action (e.g., make binding market transactions). Performative agents seek out and act on information, according to user prescribed criteria.

Social Welfare: The amount by which the value society receives from an activity (e.g., product or service) exceeds the cost to society of providing that activity. Social welfare is used to measure the net benefit society receives from any activity.

Two-Sided Matching Markets: Markets involving two groups of participants, in which a member of one group is uniquely matched to one and only one member of the other group (e.g., two partners in a marriage). Matches are stable if no two participants, who are from opposite groups in the market and are not matched to one another, would prefer one another to their current partner.

Agent-Based Intelligence Infrastructure

Zaiyong Tang
Louisiana Tech University, USA

Subramanian Sivaramakrishnan
University of Manitoba, Canada

INTRODUCTION

Today's enterprises must go beyond traditional goals of efficiency and effectiveness; they need to be intelligent in order to adapt and survive in a continuously changing environment (Liebowitz, 1999). An intelligent organization is a living organism, where all components and subsystems work coherently to enable the enterprise to maximize its potential in its goal-driven endeavors. Stonier (1991) suggested that intelligent organizations must have not only intelligent individuals, but also "collective intelligence" that is created through integration of intelligence from sub-units of the organization. Researchers have developed frameworks for building organizations around intelligence, as opposed to traditional approaches that focus on products, processes, or functions (e.g., McMaster, 1996; Liang, 2002). Analogous to intelligent biological life, an intelligent organization has a life of its own. An intelligent enterprise understands its internal structure and activities, as well as external forces such as market, competition, technology, and customers. It learns and adapts continuously to the changing environment. The learning and adaptation are achieved through real-time monitoring of operations, listening to customers, watching the markets, gathering and analyzing data, creating and disseminating knowledge, and making intelligent decisions.

Building an intelligent enterprise requires an intelligent foundation that supports intelligent reasoning and behavior at all levels of the organization. Modern information and communications technologies, combined with artificial intelligence (AI) research, provide necessary tools to create and sustain intelligence in the organizational infrastructure. Artificial intelligence has found wide applications in business, ranging from production planning and scheduling to data mining and customer relationship management. However, traditional AI systems have focused on domain-specific problem solving. Simple job shop scheduling and product malfunction diagnosis do not lend themselves well to enterprise-wide management. A more recent branch of artificial intelligence research is distributed artificial intelligence (DAI). DAI helps "far-flung, often stand-alone, application components work toward a common goal" (Chaib-draa, 1998, p. 31). Another development in AI research is the study of agent-based systems, in which autonomous software agents are deployed to carry out various tasks that are traditionally performed by humans (Jennings & Wooldridge, 1998). Recent advancements in distributed artificial intelligence, multi-agent systems, and networking technology have laid the foundation for real-world deployment of intelligent agents (Martin, Cheyer & Moran, 1999; Moukas, Zacharia & Maes, 2000).

BACKGROUND

Intelligent agents go by various names such as software agents, softbots, autonomous agents, or simply, agents (Huhns & Singh, 1998). Russell and Norvig (1995, p. 33) defined an agent as "anything that can be viewed as perceiving its environment through sensors and acting upon that environment through effectors." An intelligent agent is, according to Franklin and Graesser (1996), "a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future." In practical applications, intelligent agents are named according to their main functions. Examples include news agents, e-mail agents, shopping agents, search agents, information brokering agents, personal assistant agents, collaborative agents, negotiation agents, transaction agents, security agents, and so on.

Intelligence infrastructures have not been well studied. However, many successful businesses have already built intelligence infrastructures, enabling them to become intelligent enterprises. Broadly speaking, an intelligence infrastructure includes all basic facilities, services, and installations needed for the functioning of an intelligent enterprise. Intelligence infrastructure is defined as information technology-based facilities, systems, and services that support holistic intelligent behaviors throughout an organization. In the 1970s and 1980s, the information infrastructure of a business consisted of database and database management systems that supported various business processes. During the 1990s, businesses began the move from information age organizations to learning organizations, characterized by an information infrastructure based on data mining and knowledge management systems. Intelligent agents have emerged in recent years as key components in the organizational information infrastructure.

The evolution of modern organizations and their information infrastructure can be classified into three distinctive phases, as depicted in Figure 1. Information age organizations emphasize data and information processing. Transaction processing systems (TPSs), management information systems (MISs), decision support systems (DSSs), and executive information systems (EISs) are traditional business information systems used at different managerial levels. Database (DB) management and knowledge management (KM) provide the infrastructure support for learning organizations. A learning organization emphasizes knowledge creation, sharing, and dissemination. Supply chain management (SCM), enterprise resource planning (ERP), business process redesign (BPR), and total quality management (TQM) are commonly used in learning organizations. Multi-agent systems (MASs) can be designed to create dynamic connections to various information systems. MASs, DBs, and KM provide the foundation for distributed intelligence. Intelligence infrastructure (II), global resource management (GRM), and customer relationship management (CRM) are indispensable components of an intelligent enterprise.

Intelligent organizations require seamless integration of all systems in the organization through the intelligence infrastructure. Global resource management systems allow the intelligent organization to scan, analyze, and integrate global resources. The existence of advanced information and communication systems such as MASs, IIs, and GRMs does not automatically guarantee the success of the organization. It is a necessary condition for intelligent enterprises. Success of those enterprises depends on the assimilation of advanced information technologies into their organizational design.

A key feature of the intelligence infrastructure is the integration of all components and subsystems within the enterprise. Those components and subsystems include not only various information and knowledge systems, but also management control, human resource management, and environment management. Intelligent agents automate key operational processes, monitor operations, schedule activities, coordinate tasks, process data, anticipate needs, deliver proactive information and services, negotiate and collaborate with other agents, and intimate with their masters—the knowledge users. What distinguishes intelligence infrastructure from other types of systems is its ability to continuously capture and integrate business process knowledge, hence improving the organization's ability to learn and adapt while the changes occur, not after.

Although intelligent agents have found many successful applications, the research on integrating intelligent agents with traditional systems to create intelligence infrastructure is limited. Much of the research on intelligent agents has focused on technology issues and spe-

Figure 1. Evolution of organizations and their information infrastructure

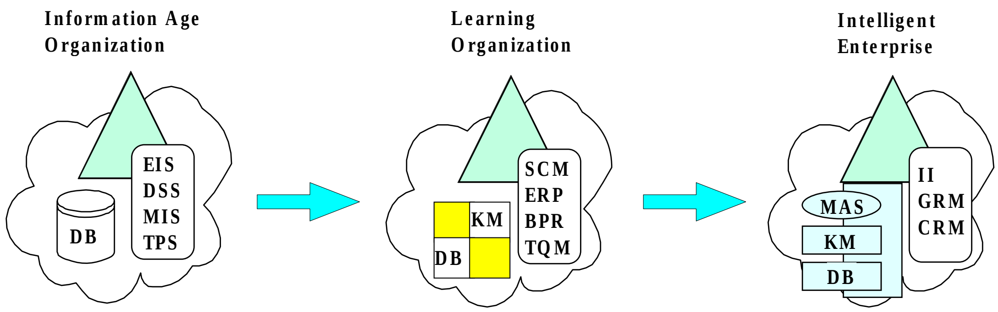


Table 1. Specialized intelligent agents in intelligence infrastructure

Intelligence Infrastructure Component	Intelligent Agents
Mission and goals	Environment scanning Executive support Business intelligence Strategic planning
Policies and procedures	Planning and scheduling Information retrieving and filtering Resource management Monitoring Diagnosis
Processes and control	Process monitoring and control Procurement Enterprise resource planning Supply chain management Data mining Network security E-commerce E-business
Information and knowledge management	Information collection Information retrieving and filtering Intranet and Web crawler Knowledge codification Knowledge distribution Directory and category Ontology creation and support
Human resource	Interface Communication assistant Personal assistant Collaborative work Decision support Project management Community support Virtual mentors Training and tutoring
Products and services	Business intelligence R&D subsystem Computer-aided design Visualization Simulation Contract negotiation Customer relationship management

cific subsystems of intelligent enterprises. Zhu, Perietula, and Hsu (1997) explored building a link between information technology and organizational learning, focusing on creating business processes that can learn in a manufacturing environment. Peng et al. (1998) created a multi-agent framework that supports the integration of various types of expertise in an intelligent enterprise. Jain, Aparicio, and Singh (1999) introduced intelligent agents for streamlining processes in virtual organizations. Intelligent agents are particularly effective in providing a reliable and flexible foundation for virtual organizations through facilitating knowledge query and communications. Sikora and Shaw (1998) proposed a multi-agent framework for integrating various traditional information systems. In business-to-business e-commerce applications, Papazoglou (2001) advocated using intelligent business agents to support a natural merging of object orientation and knowledge-based technologies.

FUTURE TRENDS

A successful intelligence infrastructure not only provides support for intelligent decision making of the organization, but also evolves and improves autonomously. This kind of self-adaptation has been observed in complex adaptive systems (CASs), in which intelligent agents adapt by changing the rules that control their behaviors as the agents gain more knowledge and experience in the system. Sutherland and van den Heuvel (2002) termed those intelligent agent-based subsystems as *agentified enterprise components*. They also provided case studies of enterprise application integration with complex adaptive systems. Rather than relying on a few highly sophisticated intelligent agents to support decision making throughout the organization, intelligence at the infrastructure level arises from the interaction of a large number of heterogeneous intelligent agents. Those agents specialize in certain functional areas such as planning, information gathering, and process monitoring and control. Each agent may have a different level of learning and self-improvement capabilities at both infrastructure and application levels. Table 1 lists specialized intelligent agents that can be deployed in an intelligence infrastructure that supports organizational functions in various capacities.

As shown in Table 1, there is a wide spectrum of tasks that can be performed and/or assisted by intelligent agents. Furthermore, intelligent agents can be used as the next-generation software components or building blocks for creating intelligent subsystems for intelligent organizations (Griss, 2001). As part of the intelligence infrastructure, modern speech and vision technologies can be used to create an intelligent environment that supports natural human interactions and provides both proactive and reactive services to the knowledge users. With a semantic network for knowledge representation, organizational knowledge can be recorded and retrieved using complex queries. Hanssens, Kulkarni, Tuchinda, and Horton (2002) described agent-based intelligent workspaces that provide a multi-layered infrastructure capable of supporting an intelligent environment.

As intelligent organizations vary in size and structure—from project teams, to business entities, industries, and nations—so will the intelligence infrastructure. Although the capability of the intelligence infrastructure covers a wide spectrum, to allow intelligent enterprises to reap the desired benefits of an organic structure and spontaneity in responding to environmental changes, the intelligence infrastructure must meet the fundamental requirements summarized in Table 2.

In traditional systems, many of the functions and processes may not be autonomous. Furthermore, interoperability may be limited due to different platforms and standards used by various subsystems. Recent developments in standards-based networking and data communications, such as transactions based on SOAP (Simple Object Access Protocol) and XML (Extensible Markup Language) Web services, promise drastic improvements in machine-to-machine communications. These developments benefit standards-based intelligent agents' communications and the integration of intelligent agent systems with traditional systems. The inter-agent communication and negotiation can be facilitated by standard protocols such as the Knowledge Query and Manipulation Language (KQML). Although much progress has been made in recent years in creating intelligence infrastructure, there remain many open issues. Table 3 summaries some of the key research issues in designing and developing a successful agent-based intelligence infrastructure.

CONCLUSION

An intelligence infrastructure is created by integrating existing information/knowledge systems through the introduction of an encompassing, efficient, flexible, and intelligent communication system that greatly enhances the integration, coordination, and collaboration of people and resources in an organization. This built-in intelligence at the infrastructure level makes the organization agile and robust. Although traditional information systems infrastructure offers a certain degree of intelligence, it is in general limited, fractural, and static in nature. An intelligent organization needs an intelligence infrastructure that is integrated, comprehensive, dynamic, and adaptive.

The intelligence infrastructure supports and enhances the complex and dynamic functionalities of an intelligent organization. Through open standards and intelligent agents, traditional systems will be able to cooperate and provide coherent information and services to end users. Intelligent agents serve as intermediaries that not only

Table 2. Requirements for agent-supported intelligence infrastructure

• A distributed information/knowledge system with central coordination
• A distributed multi-agent system integrated with traditional systems
• Open standards-based technologies to ensure a high degree of interoperability
• A common ontology that enables agents to understand one another
• A high level of self-regulation in system operation, monitoring, and adoption
• A high degree of reliability and accessibility
• Secure and trusted agent-user and agent-agent relationships
• User-friendly interface and agent-supported learning
• Improved value to the end users
• Easy to develop, deploy, and maintain
• Integrating and updating existing assets
• Leveraging external resources
• Contingency management

Table 3. A summary of research issues in agent-based intelligence infrastructure

• Multi-agent communication and collaboration
• Intelligence infrastructure and knowledge management
• Agent and intelligence infrastructure standards
• Anthropomorphic agent (human-computer interface)
• Intelligence infrastructure and XML Web services
• Integration of intelligent agents with legacy systems
• Integration of heterogeneous systems
• Security of intelligence infrastructure
• Adaptation and diffusion of intelligence infrastructure
• Control and management of intelligence infrastructure
• Performance measures of intelligence infrastructure

make vast resources available to the users, but also empower the users by automating, consolidating, and simplifying routine tasks, and allowing the users to explore and perform tasks that were not possible in the past.

REFERENCES

Chaib-draa, B. (1998). Industrial applications of distributed AI. In M.N. Huhns & M.P. Singh (Eds.), *Readings in agents* (pp. 31-35). San Francisco: Morgan Kaufmann.

Franklin, S. & Graesser, A. (1996). Is it an agent, or just a program? A taxonomy for autonomous agents. *Proceedings of the 3rd International Workshop on Agent Theories, Architectures, and Languages.* New York: Springer-Verlag.

Griss, M.L. (2001). Software agents as next generation software components. In G.T. Heineman and W. Council (Eds.), *Component-based software engineering: Putting the pieces together.* Boston: Addison-Wesley.

Hanssens, N., Kulkarni, A., Tuchinda, R. & Horton, T. (2002). Building agent-based intelligent workspaces. *ABA Conference Proceedings*. Retrieved October 22, 2003, from *citeseer.ist.psu.edu/hanssens02building.html*

Huhns, M.N. & Singh, M.P. (Eds.). (1998). *Readings in agents*. San Francisco: Morgan Kaufmann.

Jain, A.J., Aparicio IV, M. & Singh, M.P. (1999). Agents for process coherence in virtual enterprises. *Communications of the ACM, 2*(3), 62-69.

Jennings, N.R. & Wooldridge, M.J. (Eds.). (1998). *Agent technology, foundations, applications, and market*. Berlin: Springer-Verlag.

Liang, T.Y. (2002). The inherent structure and dynamic of intelligent human organizations. *Human Systems Management, 21*(1), 9-19.

Liebowitz, J. (1999). Building organizational intelligence: A knowledge primer. New York: CRC Press.

Martin, D., Cheyer, A. & Moran, D. (1999). The open agent architecture: A framework for building distributed software systems. *Applied Artificial Intelligence, 13*(1), 91-128.

McMaster, M.D. (1996). *The intelligence advantage: Organizing for complexity*. Burlington, MA: Butterworth-Heineman.

Moukas, A., Zacharia, G.R. & Maes, P. (2000). Agent-mediated electronic commerce: An MIT media laboratory perspective. *International Journal of Electronic Commerce, 4*(3), 5-22.

Papazoglou, M. (2001). Agent-oriented technology in support of e-business: Enabling the development of "intelligent" business agents for adaptive, reusable software. *Communications of ACM, 44*(4), 71-77.

Peng, Y., Finin , T.Y., Labrou B., Chu, J., Long ,W.J. & Boughannam, T.A. (1998). A multi-agent system for enterprise integration. *Proceedings of the 3rd International Conference on the Practical Applications of Agents and Multi-Agent Systems* (pp. 155-169), London.

Russell, S.J. & Norvig, P. (1995). *Artificial intelligence: A modern approach*. Englewood Cliffs, NJ: Prentice-Hall.

Sikora, R. & Shaw, M.J. (1998). A multi-agent framework for the coordination and integration of information systems. *Management Science, 44*(12), 65-78.

Stonier, T. (1991). Towards a new theory of information. *Journal of Information Science, 17*(5), 257-263.

Sutherland, J. & van den Heuvel, W. (2002). Enterprise application integration and complex adaptive systems. *Communications of ACM, 45*(10), 59-64.

Zhu, D., Prietula, M. & Hsu, W. (1997). When processes learn: Steps toward crafting an intelligent organization. *Information Systems Research, 8*(3), 302-317.

KEY TERMS

Distributed Artificial Intelligence (DAI): A subset of artificial intelligence that is concerned with the study of issues related to knowledge distribution and problem solving involving a society of decentralized but connected entities.

Intelligence Infrastructure: The organizational foundation that consists of information technology-based facilities, systems, and services that support holistic intelligent behaviors throughout an organization.

Intelligent Agent: An autonomous software program that performs certain tasks delegated to it by its master. The program is able to learn and adapt to its environment in order to perform the tasks better. Task-specific intelligent agents are given different names based on their functionalities, such as interface agent, information agent, e-commerce agent, and personal assistant agent.

Intelligent Organization: An organization is a living organism where all components and subsystems work coherently together to enable the organization to maximize its potential in its goal-driven endeavors. It is characterized by learning and adapting to the changing environment.

Multi-Agent Systems (MASs): Distributed systems with a group of intelligent agents that communicate, bargain, compete, and cooperate with other agents and the environment to achieve goals designated by their masters. Some MASs exhibit collective emerging behaviors that cannot be predicted from individual agent behaviors.

Organizational Intelligence: The collective organizational capacity of achieving goals and effectively responding to environmental changes.

Web Services: The programmatic interfaces that support application-to-application communication on the World Wide Web. Standards-based Web services use XML to interact with each other.

Agent-Based Negotiation in E-Marketing

V.K.Murthy
University of New South Wales, Australia

E.V.Krishnamurthy
Australian National University, Australia

INTRODUCTION

This article describes in brief the design of agent-based negotiation system in e-marketing. Such a negotiation scheme requires the construction of a suitable set of rules, called protocol, among the participating agents. The construction of the protocol is carried out in two stages: first expressing a program into an object-based rule system and then converting the rule applications into a set of agent-based transactions on a database of active objects represented using high-level data structures.

BACKGROUND

An agent is a code-containing object, that along with data and execution context can migrate autonomously and purposefully within a computer network. Thus an agent knows what to do with the information obtained from its environment. Agents behave like actors and have intentions and actions. In addition, agents are flexible, proactive and have multithreaded control. In this overview, we describe in detail how a set of agents can be used for negotiation in e-marketing. For this purpose we need to have a model of the multi agent-based paradigm for executing the negotiation process analogous to what we humans do. Negotiation is an interactive process among a number of agents that results in varying degrees of cooperation, competition and ultimately to commitment that leads to a total agreement, consensus or a disagreement. It has many applications, including economics, psychology, sociology and computer science.

MAIN THRUST OF THE ARTICLE

The following subsections bring out the main thrust of this chapter, namely: what is a multi-agent system, what is planning, reasoning and negotiation, and how agents can be useful in modeling e-market and e-auction. Also we will briefly describe how a coalition among agents can cause a speculation bubble or a crash in e-share market.

A MULTI-AGENT SYSTEM

A simple model of an agent that is suitable for our purpose is shown in Figure 1. This is a unified model based on several important contributions made by the following authors: (Chen & Dayal, 2000; Dignum & Sierra, 2001; Fisher, 1995; Genesereth & Nilsson, 1987; Ishida, 1994; Murthy, 2002; Woolridge, 2002).

As shown in Figure 1, an agent consists of the following subsystems:

(1) Worldly states or environment U: Those states which completely describe the universe containing all the agents.

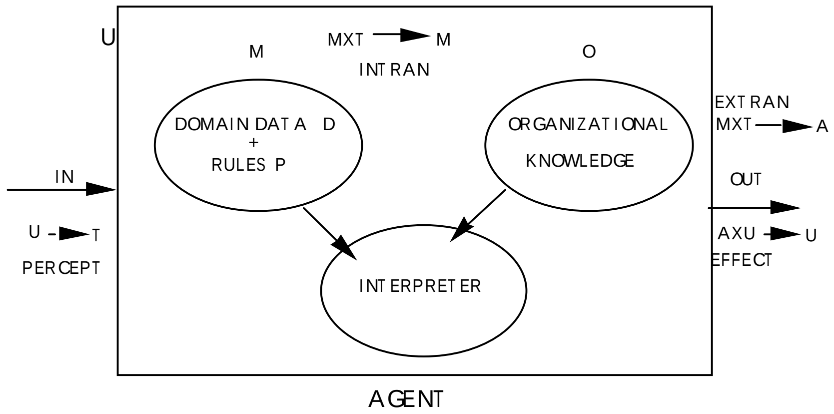
(2) Percept: Depending upon the sensory capabilities (input interface to the universe or environment), an agent can partition U into a standard set of messages T, using a sensory function Perception (PERCEPT): PERCEPT : $U \rightarrow T$.

PERCEPT can involve various types of perception: see, read, hear, smell. The messages are assumed to be of standard types based on an interaction language that is interpreted identically by all agents.

(3) Epistemic states or Mind M: We assume that the agent has a mind M (that is essentially a problem domain knowledge consisting of an internal database for the problem domain data and a set of problem domain rules) that can be clearly understood by the agent without involving any sensory function. The database D sentences are in first order predicate calculus (also known as extensional database) and agents' mental actions are viewed as inferences arising from the associated rules that result in an intentional database, whcih changes (revises or updates) D.

The agent's state of belief, or a representation of an agent's state of belief at a certain time, is represented by an ordered pair of elements (D, P). D is a set of beliefs about objects, their attributes and relationships stored as an internal database and P is a set of rules expressed as preconditions and consequences (conditions and actions). When T is input, if the conditions given in the left-hand side of

Figure 1.



P match T, the elements from D that correspond to the right-hand side are taken from D and suitable actions are carried out locally (in M) as well as on the environment.

(4) Organizational Knowledge (O): Since each agent needs to communicate with the external world or other agents, we assume that O contains all the information about the relationships among the different agents. For example, the connectivity relationship for communication, the data dependencies between agents, and interference among agents with respect to rules. Information about the location of different domain rules is in O.

(5) INTRAN: M is suitably revised or updated by the function called internal transaction (INTRAN).

Revision: Revision means acquisition of new information about the environment that requires a change in the rule system P. This may result in changes in the database D.

Example: The inclusion of a new tax rule in the tax system.

Update: Update means adding new entries to the database D; the rules P are not changed.

Example: Inclusion of a new tax-payer in the tax system.

Both revision and update can be denoted in set-theoretic notation by: INTRAN: $M \ X \ T \rightarrow M$

(6) EXTRAN: External action is defined through a function called global or external transaction (EXTRAN) that maps an epistemic state and a partition from an external state into an action performed by the agent. That is: EXTRAN: $M \ X \ T \rightarrow A$
This means that the current state of mind and a new input activates an external action from A.

(7) EFFECT: The agent also can affect U by performing an action from a set of actions A (ask, tell, hear, read, write, speak, send, smell, taste, receive, silent), or more complex actions. Such actions are carried out according to a particular agent's role and governed by an etiquette called protocols. The effect of these actions is defined by a function EFFECT, which modifies the world states through the actions of an agent:

EFFECT: $A \ X \ U \rightarrow U$; EFFECT can involve additions, deletions and modifications to U.

Thus an agent is defined by a set of nine entities, called a 9-tuple:

(U,T,M(P,D),O,A,PERCEPT,INTRAN,EXTRAN,EFFECT).

The interpreter repeatedly executes selected rules in P, until no rule can be fired.

We can interpret all the abstract machine models (such as a finite state machine or a Turing machine) and parallel computational models (such as classifier systems) as subclasses of the agents, by suitably formulating the definitions.

The nature of internal production rules P, their mode of application, and the action set A determines whether an agent is deterministic, nondeterministic, probabilistic or fuzzy. Rule application policy in a production system P can be modified by:

(1) Assigning probabilities/fuzziness for applying the rule
(2) Assigning strength to each rule by using a measure of its past success
(3) Introducing a support for each rule by using a measure of its likely relevance to the current situation

The preceding three factors provide for competition and cooperation among the different rules. Such a model is useful for negotiation in learning, as well as in e-marketing that involves interactions between many agents.

WHAT IS NEGOTIATION?

A negotiation protocol is viewed as a set of public rules that dictate the conduct of an agent with other agents to achieve a desired final outcome.

A negotiation protocol among agents involves the following actions or conversational states:

(1) Propose: one puts forward for consideration a set of intentions called a proposal.
(2) Accept: The proposal is accepted for execution into actions.
(3) Refuse: The proposal is rejected for execution into actions.
(4) Modify: This alters some of the intentions of the proposer and suggests a modified proposal that is at the worst, a refuse and a new proposal; or a partial acceptance and new additions.
(5) No proposal: No negotiation.
(6) Abort: Quit negotiation.
(7) Report agreement: This is the termination point for negotiation in order to begin executing actions.
(8) Report failure (agree to disagree): Negotiation breaks down.

Note that the previous actions are not simple exchange of messages but may involve some intelligent or smart computation.

Multiagents can cooperate to achieve a common goal to complete a transaction to aid the customer. The negotiation follows rule-based strategies that are computed locally by its host server. Here competing offers are to be considered; occasionally cooperation may be required. Special rules may be needed to take care of risk factors, domain knowledge dependencies between attributes, and positive and negative end conditions. When making a transaction several agents have to negotiate and converge to some final set of values that satisfies their common goal. Such a goal should also be cost effective so that it is in an agreed state at the minimum cost or a utility function. To choose an optimal strategy each agent must build a plan of action and communicate with other agents.

PLANNING, REASONING AND NEGOTIATION

The negotiation process is usually preceded by two cooperating interactive processes: planning and reasoning (Woolridge, 2000). The ability to plan ahead for solving a problem is the key aspect of intelligent behavior. To solve a problem through negotiation, we start with a set of desired properties and try to devise a plan that results in a final state with the desired properties. For this purpose, we define an initial state where we begin an operation and also define a desirable goal state or a set of goal states. Simultaneously, we use a reasoning scheme and define a set of intended actions that can convert a given initial state to a desired goal state or states. Such a set of intended actions called the plan exists if and only if it can achieve a goal state starting from an initial state and moving through a succession of states. Therefore to begin the negotiation process, we need to look for a precondition that is a negation of the goal state and look for actions that can achieve the goal. This strategy is used widely in AI and forms the basis to plan a negotiation (Genersereth & Nilsson, 1987). Such a planning is possible for clear-cut algorithmic problems. For general AI problems, however, we can only generate a plan that may or may not work; if the plan does not work we need to either modify the plan or devise a new plan. The same approach is used for devising a multi-agent negotiation protocol that is useful in e-auction, e-marketing and in other applications (Horlait et al., 2003; Marik et al., 2003; Schillo et al., 2003; Woolridge, 2000).

We now describe how to carry out distributed multi-agent negotiation by sending, receiving, handshaking and acknowledging messages and performing some local computations.

A multi-agent negotiation has the following features (Dignum & Sierra, 2001; Murthy, 2002):

(1) There is a seeding agent who initiates the negotiation and coordinates the negotiation.
(2) Each agent can be active or inactive.
(3) Initially all agents are inactive except for a specified seeding agent that initiates the computation.
(4) An active agent can do local computation, send and receive messages and can spontaneously become inactive.
(5) An inactive agent becomes active, if and only if it receives a message.
(6) Each agent may retain its current belief or revise its belief as a result of receiving a new message by performing a local computation. If it revises its belief, it communicates its revised state of belief to other concerned agents; else it does not revise its solution and remains silent.
(7) Finally, the negotiation terminates, or fails. If there is no consensus within a set time-out limit, the negotiation fails.

FUTURE TRENDS

Agent based systems will provide new approaches to modeling and simulation of complex business systems and also provide for new software engineering approaches (Lucena et al., 2004; Sichman et al., 2003). We will illustrate this by a simple example on modeling e-market and e-auction.

MODELING E-MARKET AND E-AUCTION

The agent negotiation system can model the e-market with many traders (agents), popularly known as buyers and sellers. These agents negotiate over the Internet to sell or buy shares or stocks in a stock market. In an e-market situation, it is possible that the negotiation ultimately leads to self-organization and criticality, causing crashes. That is, individual agents that correspond to a microscopic system can emerge as a self-organizing macroscopic system corresponding to a "percolation model" (Paul & Baschnagel, 1999).

In e-market situation (see Figure 1), to start with, the domain data D, rules P and organizational knowledge O can be based on three factors:

(1) the experience and economics knowledge of an agent deployed by a trader based totally on individualistic idiosyncratic criteria.
(2) the trader's acquired knowledge through communication with other selected agents; such a trader is called a fundamentalist.
(3) the trader's acquired knowledge by observing the trends on market from a collective opinion of other traders; such a trader is called a trend chaser.

The previous three factors play an important role in deciding the number of possible states that each agent will be in and his or her inclination to buy or sell or wait in an e-marketing decision.

In e-market, at every time instant a trader can adopt three possible states of action: buy, sell or wait, respectively represented by three states 1, -1 and 0. Each agent representing a trader can communicate with one another and this creates an imaginary bond or connectivity relationship among them, modifying the organizational knowledge O. This bond is created with a certain probability determined by a single parameter that characterizes the willingness of an agent to comply with others.

The three states of behavior are obviously a very complicated function of the behavioral property and personality of an individual and whether he or she uses elementary, derived or inferential beliefs. It is interesting to note that all the beliefs are again a function of the speed with which information is available to an agent, financial status, ability to reason and susceptibility to pressure from neighbors. Thus in a share market or auction situation we need to work out how the agents are linked in order to obtain information through communication, the personality factors such as age, financial status and the market trend. Using data mining techniques one can derive detailed information about the mechanism of bond formation among the agents. Based on this information, we can assume that any two agents are randomly connected with a certain probability. This will divide the agents into clusters of different sizes whose members are linked either directly or indirectly via a chain of intermediate agents. These groups are coalitions of market participants who share the same opinion about their activity. The decision of each group is independent of its size and the decision taken by other clusters. In this situation, using the random cluster model (Bak, 1996; Paul & Baschnagel, 1999) we can show that when every trader is on average connected to another, more and more traders join the spanning cluster, and the cluster begins to dominate the overall behavior of the system. This can give rise to "speculation bubble" (if the members all decide to buy), a crash (if the members all decide to sell) or a stagnation (if the members all decide to wait). These are cooperative phenomenon and depend upon trading rules, exchange of information- the speed and volume, and the connectivity relationship. For the three-state agents the critical probability $p(c) = 0.63$. That is, if an agent is even showing about 63% preference to the information from his or her neighbors, a crash or bubble is bound to happen. Crash is a highly cooperative phenomenon and depends upon trading rules, exchange of information- the speed and volume, and the connectivity relationship. Accordingly, an analogy exists between stock-market crashes and critical phenomena or phase transitions in physics. Thus a distributed agent system can eventually enter into a self-organized critical state or an emergent state. However, such a phenomenon is similar to an earthquake or epidemic, or fatigue in materials resulting in cracks and is hard to predict, since the cluster formation depends upon the various factors already mentioned.

CONCLUSION

This overview explained the use of multi-agent based planning, reasoning and negotiation in e-marketing. We explained how to use the techniques of AI planning to devise the multi-agent based negotiation protocol. We also explained how agents can reach a self-organized critical state that can result in crash, speculation bubble or stagnation in e-share market. The study of agent systems and their applications is a vastly expanding area for research; for more details, the following references may be consulted: (Horlait et al., 2003; Marik et al., 2003; Schillo et al., 2003; Woolridge, 2000).

REFERENCES

Bak, B. (1996). *How nature works: The science of self-organized criticality*. New York: Springer.

Chen, Q., & Dayal, U. (2000). Multi agent cooperative transactions for e-commerce. *Lecture Notes in Computer Science*, *1901*, 311-322. New York: Springer Verlag.

Dignum, F., & Sierra, C. (Eds.). (2001). Agent mediated e-commerce. *Lecture Notes in Artificial Intelligence, 1991, 2003*. New York: Springer Verlag.

Fisher, M. (1995). Representing and executing agent-based systems. *Lecture Notes in Computer Science, 890*, 307-323. New York: Springer-Verlag.

Genesereth, M.R., & Nilsson, N.J. (1987). *Logical foundations of artificial intelligence*. New York: Morgan Kaufmann.

Horlait, E., Magedanz, T., & Glitho, R.H. (Eds.). (2003). Mobile agents for telecommunication applications. *Lecture Notes In Artificial Intelligence*, *2691*. New York: Springer Verlag.

Ishida, T. (1994). Parallel, distributed and multiagent production systems. *Lecture Notes in Computer Science, 878*. New York: Springer Verlag.

Lucena, C. et al. (2004). Software engineering for multi-agent systems II . *Lecture Notes in Computer Science, 2940*. New York: Springer Verlag.

Marik, V., Muller, J., & Pechoucek, M. (Eds.). (2003). Multi-agent systems and applications. *Lecture Notes In Artificial Intelligence, 2691*. New York: Springer Verlag.

Murthy, V.K. (2002). Designing agent-based negotiation for e-marketing. In N. Shi & V.K. Murthy (Eds.), *Architectural issues of Web-enabled electronic business* (pp. 290-302). Hershey, PA: Idea Group Publishing.

Paul, W., & Baschnagel, J. (1999). *Stochastic processes*. New York: Springer Verlag.

Schillo, M., Klusch, M., Muller, J., & Tianfield, H. (Eds). (2003). Multi-agent system technologies. *Lecture Notes in Artificial Intelligence, 2831*. New York: Springer Verlag.

Sichman, J.S. et al. (2003). Multi-agent based simulation II. *Lecture Notes in Artificial Intelligence, 2581*. New York: Springer Verlag.

Woolridge, M. (2000). *Reasoning about rational agents*. Cambridge, MA: MIT Press.

Woolridge, M. (2002). *An introduction to multi-agent systems*. New York: John Wiley.

KEY TERMS

Agent: A system that is capable of perceiving events in its environment, or representing information about the current state of affairs and of acting in its environment guided by perceptions and stored information (current definition by AOIS, agent oriented information system community).

Coordinator: An that acts as a coordinator among several agents to carry out the negotiation protocol.

E-Auction: This is a centralized protocol for redistributing resources among agents. Each agent attaches a value to each resource. The seller asks a price for a resource and buyer offers a price and they negotiate over the Internet to achieve a desired outcome satisfying to both; else the negotiation fails.

E-Market: An Internet based market with many traders (agents) popularly known as buyers and sellers. These agents negotiate over the Internet to sell or buy products in any market (e.g., shares or stocks in a stock market).

Negotiation: This is an interactive process among a number of agents that results in varying degrees of cooperation, competition and ultimately to commitment that leads to a total agreement, consensus governed by a voting policy, or a disagreement.

Negotiation Protocol: A negotiation protocol is viewed as a set of public rules that dictate the conduct of an agent with other agents to achieve a desired final outcome.

Protocols: A set of rules that dictate the behavior of objects for communication and interaction.

Self-Organized Critical State or Emergence: A group of interconnected communicating, interacting and negotiating agents reach a state of self-organization resulting in an unpredictable decision—such as a share-market crash, a stagnation or a speculation bubble.

Agents and Payment Systems in E-Commerce

Sheng-Uei Guan
National University of Singapore, Singapore

INTRODUCTION

An emerging outcome of the popularization of the Internet are electronic commerce and payment systems, which present great opportunities for businesses, reduce transaction costs, and provide faster transaction times. More research has been conducted with new technologies like mobile Internet used by business models (Baek & Hong, 2003). However, before using the Internet, it is essential to provide security in transferring monetary value over the Internet. A number of protocols have been proposed for these secure payment systems, including NetBill, NetCheque, Open Market, iKP, Millicent, SET (Sherift, 1998), E-Cash (Brands, 1995), NetCash, CAFÉ (Mjolsnes, 1997), EMV cards (Khu-Smith & Mitchell, 2002), etc. These systems are designed to meet diverse requirements, each with particular attributes.

Automation and intelligence is another issue that poses challenges in the development of e-commerce. Agent technology has been incorporated into the area of e-commerce to provide automation and intelligence for the e-trade process. An agent is a software program capable of accomplishing tasks autonomously on behalf of its user. Agents must provide trustworthy consistency and fault tolerance to avoid eavesdropping and fraud. Also, agents should have roaming capability so as to extend their capability well beyond the limitations of owners' computers. To meet these requirements, this chapter will discuss some related components under the SAFER (Secure Agent Fabrication, Evolution, and Roaming) architecture (Zhu & Guan, 2000) and propose an agent-based payment scheme for SAFER.

Different types of electronic payment systems have been developed to meet its diverse requirements, which generally include integrity, authorization, confidentiality, availability, and reliability for security requirements (Asokan, 1997). Payment systems can be classified in a variety of ways according to their characteristics (Dahab & Ferreira, 1998), such as the exchange model (cash-like, check-like, or hybrid), central authority contact (online or offline), hardware requirements (specific or general), payment amounts (micropayment), etc.

Among the available payment schemes in the market, E-Cash is one of the best in terms of security, flexibility, and full anonymity. E-Cash is a cash-like online system that uses electronic coins as tokens. E-Cash has unique advantages, such as flexibility, integrity, and full anonymity that cannot be found in electronic check and credit card based systems. It uses cryptographic techniques to provide full anonymity. The agent-based payment scheme for SAFER adopts some similar principles and concepts of E-Cash.

MAIN THRUST OF THE ARTICLE

This chapter presents a brief overview of agents and payment system attributes used in e-commerce. An agent-based e-payment scheme built for the SAFER e-commerce architecture is proposed, which is aimed at providing a flexible and secure financial infrastructure for Internet commerce.

Software Agents in Electronic Commerce

Attributes of Agent-Based Systems for Electronic Commerce

Agents are bits of software performing routine tasks, typically in the background, on behalf of the user. Gathering, filtering, and presenting information are some of the small and well-defined tasks given to simple agents. An agent distinguishes itself from any other software by its intelligence. Intelligent agents are capable of "thinking" and producing intelligent feedback (Guan & Yang, 1999). Agents are increasing in the degree and sophistication of automation, on both the buyer's and seller's sides, commerce becomes much more dynamic, personalized, and context sensitive. These changes can be beneficial to both buyers and sellers (He, Jennings, & Leung, 2003).

The requirement for continuity and autonomy derives from our desire that an agent be able to carry out activities in a manner that is responsive to changes in the environment without requiring constant human guidance or intervention. According to Bradshaw (1997), agents have the following attributes, as shown in Table 1.

There are several software agent prototypes under development that will be capable of doing even more on behalf of buyers and sellers. One is Kasbah, wherein agents would proactively seek potential sellers and nego-

Table 1. Attributes of software agents

Attribute	Description
Reactivity	The ability to selectively sense an act
Autonomy	Goal-directness, proactive and self-starting behavior
Collaborative behavior	Can work in concert with other agents to achieve a common goal
Communication ability	The ability to communicate with persons and other agents
Personality	The capability of manifesting the attributes of a believable character, such as emotion
Temporal continuity	Persistence of identity and state over long periods of time
Adaptivity	Being able to learn and improve with experience
Mobility	Being able to migrate in a self-directed way from one host platform to another

tiate with them on the buyer's behalf, making the best possible deal, based on a set of constraints specified by the buyer, including the highest acceptable price and a transaction completion date (Chavz, 1996). A disadvantage of this software agent is that it always accepts the first offer that can meet its asking price, when even better offers might exist. This disadvantage is resolved by AuctionBot, a general-purpose Internet auction server. *AGENTics* is another agent prototype that develops what is referred to as "*online catalog integration for e-commerce*." AGENTics products shield the user from the technicalities of "where" and "how" the information was gathered, while it synthesizes many information pieces into a coherent whole (Mougayar, 1997).

Some agents can select desired items based on preferences, search databases to look for selected pieces of information, and conduct transactions. An example of such an adaptive agent is the SAFER architecture for e-commerce.

SAFER (Secure Agent Fabrication, Evolution, and Roaming) is a Web-based distributed infrastructure to serve agents to query, buy, and sell goods in e-commerce.

Figure 1. Cooperating agents for the SAFER payment scheme

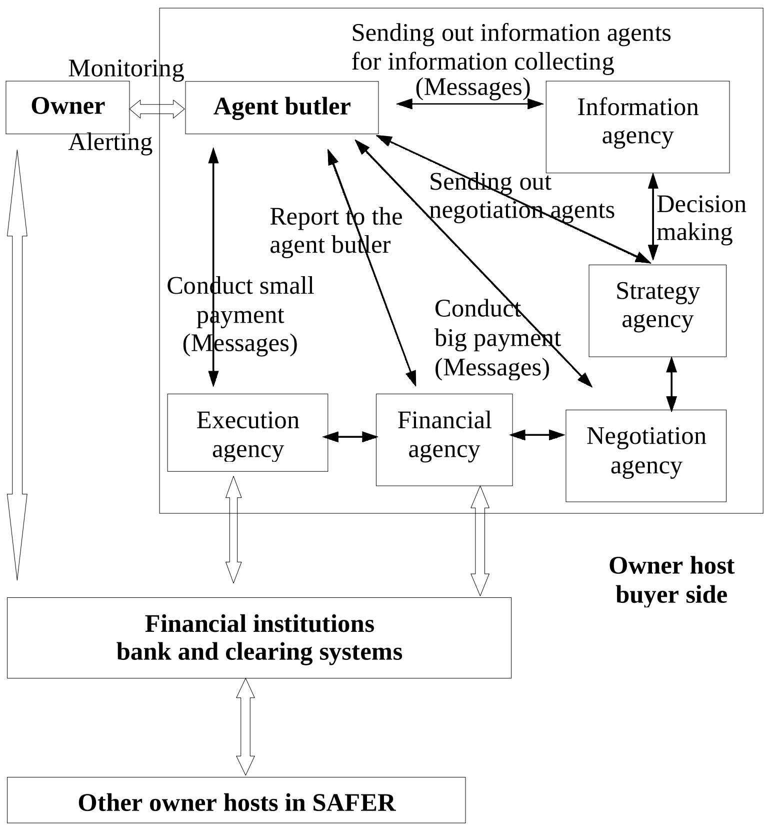
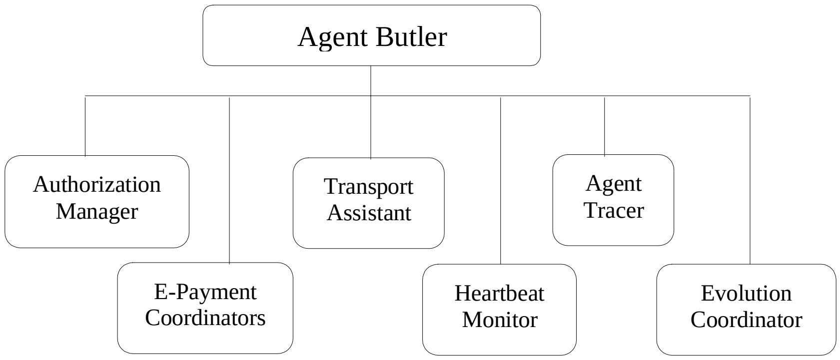


Figure 2. Prototype of agent butler



It establishes necessary mechanisms to transport, manufacture, and evolve all different types of agents. The goal of SAFER is to construct open, dynamic, and evolutionary agent systems for e-commerce (Zhu & Guan, 2000). There will be SAFER-compliant and noncompliant communities coexisting in the e-commerce network. Each SAFER community consists of several mandatory components: owner, butler, agent, agent factory, community administration center, agent charger, agent immigration, clearinghouse, and bank. Agent community is the basic unit in SAFER e-commerce, which offers virtual regions and vehicles to host and administrate mobile agents during roaming, transaction, and evolution. An owner is in charge of all his agents and of making respective authorizations to mobile agents and his agent butler, which is a 24-hour online watcher that would handle most of the tasks on behalf of the owner. When agents are sent roaming in the network, the butler has the responsibility of keeping track of agents' activities and locations by sending messages to agents and receiving messages from agents. At least one financial institution, usually a bank, that can link all value representations to real money, must also be involved. The payment scheme designed for SAFER is expected to fulfill flexibility and interoperability, which means that diverse representations of value will have the possibility of emerging in one framework for users' convenience. Given that, it is important that funds represented by one mechanism be easily converted into funds represented by others (Neuman, 1995).

An Agent-Based E-Payment Scheme for SAFER

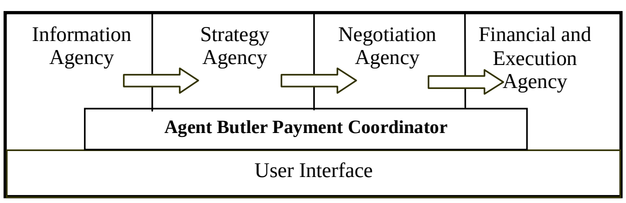
The payment module in the agent-mediated SAFER e-commerce architecture must contain several essential components: the marketplace, agents (including mobile agents, static agents, and agent butlers), financial institutions, and users. In SAFER, a community will offer virtual regions, factories, administration tools, vehicles to manipulate and administrate mobile agents during any activity, and provide security so that users can trust it.

Different types of agents fabricated by an agent factory of SAFER are running under the payment scheme for respective functions and tasks. They are briefly described in Figure 1.

In this scheme, a subsystem called agency is mentioned. Similar to the definition given by Dr. Larry Kerschberg in his Defense Personnel Support Center (DPSC) project, an agency can be thought of as a multi-layered agent group or a federation of agents with specific goals and functional roles in the architecture. It is also like a collection of cooperating intelligent agents with particular expertise (Kerschberg, 1997).

If the owner is interested in some items, he will assign tasks to his butler and agents. The agent butler will then send out information agents from the agency, taking note of the items of interest, and set parameters such as due date (by which the item should be purchased), desired

Figure 3. Payment coordinator



price, and highest acceptable price. The information agents used to sift, filter, and process information will roam in SAFER or even non-SAFER communities under a certain transport protocol, which is explained in the literature (Guan & Yang, 1999). It can help with dozens of purchasing decisions, thus lowering the cost and gaining efficiency. While roaming, agents are well tracked by the agent butler, by sending messages to report their activities and locations, which is described in detail in Zhu and Guan (2000). After gathering enough information, the information agent forwards all to the strategy agency, which will analyze the new data and settle on a decision for the user. All the recommendations will be reported to the agent butler first. Once a recommendation has been reported, the agent butler activates the negotiation agency that will send out negotiation agents to the shortlist merchant hosts. Negotiation is defined as follows in Green (1997), "negotiation is the communication process of a group of agents in order to reach a mutually accepted agreement on some matter" (21). If the negotiation agent and the receptionist agent reach an agreement, the result will be reported to the butler. The butler will inform the financial agency to initiate the contract for certain goods and make a transaction decision according to the amount of money involved, the distance from host to the destination vendor, etc. Financial agents will take charge of the goods reception and payment transaction under the authorization of the butler. They communicate with the merchant host, autonomously make a payment request, and sign a contract order against the right good.

Implementation

The implementation of SAFER is under way. The overall architecture consists of several closely related but separate modules: roaming, evolution, fabrication, negotiation, and electronic payment.

The implementation of the payment module began with the development of the agent butler, which is defined as a combination of several separate functions, as shown in Figure 2. They are authorization manager, e-payment coordinator, transport assistant, heartbeat monitor, agent tracer, and evolution coordinator.

In the e-payment coordinator module, communication channels are set up between agent butler and all agencies of diverse functionalities, each of which is running in a separate thread. User interfaces are designed so that the user can assign tasks, define needs and requirements, check records, and read alerting messages reported by his or her agent butler.

Making all types of agents and merchant hosts available to fit in the same framework will be difficult in the current research stage, because the attributes that agents require to communicate may differ. Given that, we have chosen to implement a limited number of typical agents to test the system functionality and will consider how the work could be generalized to e-commerce in the future.

FUTURE TRENDS

The foremost important feature of e-commerce is transaction security. If the system is not trustworthy, there will be no incentive for participants to cooperate. A prerequisite is the prevention of double spending of electronic cash or coins. Ensuring this is the crux of any system, and it will incur significant overhead. Electronic currency in the context of mobile-agent-based systems has one particular caveat: the credits that an agent may carry are essentially just data, to which the host potentially has access. To ensure system reliability and reduce overhead, there are still many open issues for further consideration and discussion in future work:

- Privacy issues in collecting data and negotiation, e.g., to prevent agents from divulging private information to other hosts and to protect agents from malicious hosts
- The extent that users will let agents make decisions for them (based on their preference)
- The agent's ability of negotiation (This should be protected to prevent a host from having access to the agent's negotiation function and then affect the agent's buying power.)
- Agent traceability (An agent butler keeps in contact with his mobile agents periodically.)
- Fault tolerance and credits recovery in case of sudden crash of remote systems or unexpected attack on mobile agents
- Protection of agent's public and secret key during agent roaming

CONCLUSION

The agent-based SAFER e-commerce payment scheme incorporated agent technologies and took advantage of some well-known secure payment transaction protocols. It aims to simulate and even enhance physical cash and is designed to support a multitude of currency types. By incorporating the concepts of agent, the system is expected to provide security, efficiency, flexibility, autonomy, and intelligence. It is designed to provide anonymity against other parties and audit ability (traceability) for the owner (or agent butler). At last, a number of potential improvements, practical aspects, and some open issues have been identified for future work.

REFERENCES

Asokan, N., & Janson, P. A. (1997). *The state of the art in electronic payment systems. Computer, 30*(9), 28–35.

Baek, J. M., & Hong, I.-S. (2003). *A study on mobile payment system with United Mileage using USIM* (pp. 394–403). HSI 2003, LNCS 2713.

Bradshaw, J. (1997). *Software agent*. Cambridge, MA: AAAI Press/The MIT Press.

Brands, S. (1995). *Electronic cash on the Internet.* In *Proceedings of the Symposium on Network and Distributed System Security, San Diego, California, 1995* (pp. 64–84).

Chavz, A., & Maes, P. (1996). MIT Media Lab. *Kashbah: An Agent Marketplace for Buying and Selling Goods*. Proceedings of the First International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology (Crabtree, B. and Jennings, N., Eds.), 75-90. The Practical Application Company Ltd, Blackpool.

Dahab, R., & Ferreira, L. C. (1998). *A scheme for analyzing electronic payment systems.* 14th Annual Computer Security Applications Conference, Phoenix, AZ.

Green, S. (1997). *Software agents*. A review. IAG Technical Report, Trinity College Available at *http://www.cs.tcd.ie/research_groups/aig/iag/toplevel2.html*

Guan, S. U., & Yang, Y. (1999). *SAFE: Secure-roaming agent for e-commerce*, 26th International Conference on Computers and Industrial Engineering, Australia, Vol 42 Issue 2-4, 481-493.

He, M., Jennings, N. R., & Leung, H. F. (2003). *On agent-mediated electronic commerce. IEEE Transactions on Knowledge and Data Engineering*, *15*(4), 985-1003.

Kerschberg, L., & Banerjee, S. (1997). *The DPSC electronic marketspace: The impact of intelligent agents, the Internet and the Web on electronic commerce and logistics*. Available at *http://cise.krl.gmu.edu/KRG/DPSCAgentHTML_folder/DPSCAgent.html*

Khu-Smith, V., & Mitchell, C. J. (2002). *Using EMV-cards to protect e-commerce transactions* (pp. 388–399). EC-Web 2002, LNCS 2455.

Lucas, F. & Dahab, R. (1998). A scheme for analyzing electronic payment systems. In 14th ACSAC-Annual Computer Security Applications Conference (ACSAC '98), Scottsdale, Arizona.

Mjolsnes, S. F., & Michelsen, R. (1997). *CAFÉ. Open transnational system for digital currency payment.* In Proceedings of the 30th Hawaii International Conference on System Sciences, Advanced Technology Track, IEEE Computer Society, Washington, D.C. (Vol. 5, pp. 198–207).

Mougayar, W. (1997). *The future of agent-based commerce on the Web.* CYBERManagement Inc. Retrieved from *http://www.cyberm.com/cyber/art2.htm*

Neuman, B. C., & Medvinsky, G. (1995). *Requirements for network payment: The NetCheque™ perspective*. Proceedings of IEEE Compcon'95, San Francisco, IEEE Computer Society, Washington, DC, 32.

Sherift, M. H., & Serhrouchni, A. (1998). SET and SSL: Electronic payments on the Internet. In *Proceedings of the Third IEEE Symposium on Computers and Communications, 1998. ISCC'98* (pp. 353–358), Athens, Greece.

Zhu, F. M., Guan, S. U., Yang, Y., & Ko, C. C. (2000). SAFER e-commerce: Secure agent fabrication, evolution, and roaming for e-commerce. In M. R. Syed & R. J. Bignall (Eds.), Internet *Commerce and Software Agents: Cases, Technologies and Opportunities*. Hershey, PA: Idea Group Publishing, 190-207. *http://www.webopedia.com/TERM/c/cryptography.html*

KEY TERMS

Adaptability: The ease with which software satisfies differing system constraints and user needs (Evans, 1987).

Agents: A piece of software that acts to accomplish tasks on behalf of its user.

Anonymity: The degree to which a software system or component allows for or supports anonymous transactions.

Confidentiality: The nonoccurrence of the unauthorized disclosure of information (Barbacci, 1995).

Cryptography: The art of protecting information by transforming it (encrypting it) into an unreadable format, called cipher text. Only those who possess a secret key can decipher (or decrypt) the message into plain text.

Flexibility: The ease with which a system or component can be modified for use in applications or environments other than those for which it was specifically designed (IEEE, 1990).

Alignment of Business and Knowledge Management Strategies

El-Sayed Abou-Zeid
Concordia University, Canada

INTRODUCTION

The role of knowledge as a crucial asset for an enterprise's survival and advancement has been recognized by several researchers (e.g., von Krogh, Ichijo & Nonaka, 2000). Moreover, by having knowledge (intellectual resources), an organization can understand how to exploit and develop its traditional resources better than its competitors can, even if some or all of those traditional resources are not unique (Zack, 1999).

However, realizing the importance of organizational knowledge and its management in creating value and in gaining competitive advantage is only the first and the easiest step in any knowledge management (KM) initiative. The second and almost as important step is to answer how and where to begin questioning (Earl, 2001). In fact, "many executives are struggling to articulate the relationship between their organization's competitive strategy and its intellectual resources and capabilities (knowledge)" (Zack, 1999). As Zack (1999) argued, they need pragmatic yet theoretically sound model. It has been highly accepted that a pragmatic and theoretically sound model should meet at least two criteria. First, it should explicitly include the external domains (opportunities/threat) and internal domains (capabilities/arrangements) of both business (B-) and knowledge (K-) strategies and the relationships between them. Second, it should provide alternative strategic choices.

In order address this issue a KM strategic alignment model (KMSAM) is presented. It stems from the premise that the realization of business value gained from KM investment requires alignment between the business (B-) and knowledge (K-) strategies of the firm and is based on the Henderson-Venkatraman SAM for IT (Henderson & Venkatraman, 1993).

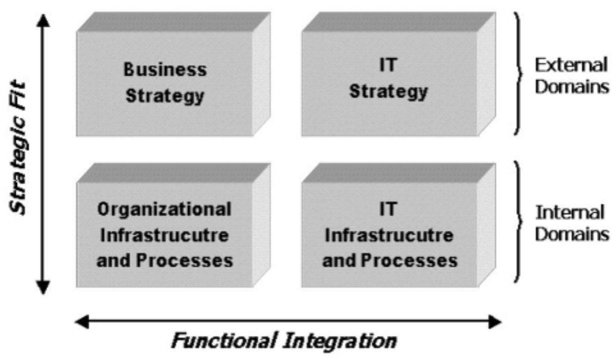
Overview of the Henderson-Venkatraman Strategic Alignment Model

The KM strategic alignment model is based on the theoretical construct developed by Henderson and Venkatraman (1993). In their model business success is viewed as the result of the synergy between four domains. The first two, the external domains, are business strategy and information technology (IT) strategy. The strategy domains are described in terms of (business/technology) scope, (distinctive business/IT systemic) competencies and (business/IT) governance. The second two, the internal domains, are organizational infrastructure and processes and IT infrastructure and processes. Both internal domains are described in terms of (administrative/IT) infrastructure, (business/IT) processes and (business/IT) skills. This synergy is achieved through two types of relationship:

- **Strategic fit** emphasizes the need for consistency between strategy (external domain) and its implementation (internal domain).
- **Functional integration,** which has two modes, extends the strategic fit across functional domains. The first mode, *strategic integration,* deals with the capability of IT functionality both to shape and to support business strategy. The second mode, *operation integration,* focuses on the criticality of ensuring internal coherence between organizational infrastructure and processes and IT infrastructure and processes.

Figure 1 shows the elements of the IT strategic alignment model (ITSAM).

Figure 1. IT strategic alignment model (Henderson & Venkatraman, 1993)



A

KM Strategic Alignment Model (KMSAM)

The premise of the original ITSAM is that ”the effective and efficient utilization of IT requires the alignment of IT strategies with business strategies” (Henderson & Venkatraman, 1993). In parallel way, the premise of KMSAM, in which knowledge strategy replaces IT strategy, is that “the effective and efficient use of organizational knowledge requires the alignment of knowledge strategies with business strategies”. Since strategy, whether business (B)-strategy or knowledge (K)-strategy, can be seen as a balancing act between the *external domain* (opportunities/threats) and the *internal domain* (capabilities/arrangements) of the firm (strengths and weaknesses) (Henderson & Venkatraman, 1993; Zack, 1999), the external and internal domains of K strategy have first to be defined.

K-Strategy External Domain

In the case of K-strategy, the *external domain* involves three dimensions: *K-scope* (what the firm must know), *K-systemic competencies* (what are the critical characteristics of the required knowledge) and *K-governance* (how to obtain the required K-competencies). The first dimension, K-scope, deals with the specific domains of knowledge that are critical to the firm’s survival and advancement strategies. Survival strategies aim at securing current enterprise profitability, while advancement strategies aim for future profitability (von Krogh et al., 2000).

Determining the K-scope can be achieved by constructing a business (B-) domain/ Knowledge (K-) thing matrix that documents the current and required state of organizational knowledge concerning some or all business domains. The first group of elements that constitute this matrix includes the list of B-domains (B_i). The second group of elements includes the K-things (K_j) that describe the current state of knowledge associated with each of the relevant B-domains. To relate this knowledge to enterprise business strategies, K-things are further classified according to the roles they play in such strategies. Von Krogh et al. (2000) have suggested that there are two types of strategies: survival and advancement. Survival strategies aim at securing current enterprise profitability, while advancement strategies aim for future profitability. Therefore, organizational knowledge, and consequently K-things, is classified into two categories: survival (K_S) and advancement (K_A). Figure (2) shows the generic form of this matrix.

The second dimension of the K-strategy external domain is K-systemic competencies. The focus of this dimension is the set of utilization-oriented characteristics of knowledge that could contribute positively to the creation of new business strategy or better support of existing business strategy. This set includes characteristics such as:

Figure 2. Generic form of B-things/K-things matrix (Abou-Zeid, 2002)

	Survival Knowledge			**Advancement Knowledge**		
B_1	K_{S11} (Current/Required States)	….	K_{S1n} (Current/Required States)	K_{A11} (Current/Required States)	….	K_{A1m} (Current/Required States)
B_2	K_{S21} (Current/Required States)	….	K_{S2k} (Current/Required States)	K_{A21} (Current/Required States)	….	K_{A2l} (Current/Required States)
….	….	….	….	….	….	….
B_N	K_{SN1} (Current/Required States)	….	K_{SNk} (Current/Required States)	K_{AN1} (Current/Required States)	….	K_{ANl} (Current/Required States)

- *Accessibility*, the extent to which organizational knowledge is made available to its members regardless of time or location (Buckman, 1998);
- *Transferability*, the extent to which the newly acquired knowledge can be applied in other contexts, for example organizational, cultural (Grant, 1996);
- *Appropriability*, the extent to which knowledge can be imitated. Things are said to have "strong" appropriability if they are difficult to reproduce by another organization. The converse is "weak" appropriability. A related concept is that of "sticky/slippery"; that is, sticky knowledge is such an integral part of a regime that it cannot be extracted in a meaningful whole (Grant, 1996; Narasimha, 2000);
- *Depth and breadth* (Narasimha, 2000);
- *Compositionality*, the amenability of knowledge to be synthesized from existing knowledge; and
- *Integrateability*, the extent to which the newly acquired knowledge can be integrated with existing knowledge.

Finally, K-governance dimension deals with the selection and use of mechanisms for obtaining the required K-competencies. The following are examples of some "acquisition mechanisms" (Probst, Raub & Romhardt, 2000):

- Bringing experts to the firm by recruiting specialists as full-time or temporary staff. Temporary hiring is becoming an increasingly interesting alternative.
- Tapping knowledge held by other firms through different inter-organizational co-operation forms such as joint ventures or strategic alliances.
- Utilizing the knowledge of stakeholders, for example, customers, suppliers, employees and owners. For example, involving customers early in the product-development process could generate valuable information about their needs.
- Acquiring knowledge products such as software, patents, and CD-ROMs.

K-Strategy Internal Domain

In the case of K-strategy, the internal domain involves three dimensions: *knowledge (K)- processes, knowledge (K)-infrastructures,* and *knowledge (K)-skills.*

Knowledge (K)-processes, the first dimension of the K-strategy internal domain, can be classified into two main categories: K-manipulating processes and K-enabling processes. The first category, K-manipulating processes, includes all the organizational processes needed to change the state of organizational knowledge such as K-generation, K-mobilization and K-application (Abou-Zeid, 2003). The second category, K-enabling processes, includes organizational processes that support K-manipulating processes such as managing conversation, mobilizing knowledge activists, creating the right context, and globalizing local knowledge (von Krogh et al., 2000).

Organizational knowledge processes are socially interaction-intensive. They involve social interactions and direct communication and contact among individuals and among members of "communities of practice". Therefore, they require the presence of social capital. Social capital is "the sum of actual and potential resources embedded within, available through, and derived from the network of relationships possessed by a social unit" (Nahapier & Ghoshal, 1998). Recognizing the importance of social capital, Gold et al. (2001) have identified three key K-infrastructures, the second dimension of the K-strategy internal domain, that is, technical, structural and cultural, that enable social capital. The *K-technical infrastructure* includes IT-enabled technologies that support KM activities such as business intelligence, collaboration and distributed learning, K-discovery, K-mapping, opportunity generation and security. The *K-structural infrastructure* refers to the presence of enabling formal organization structures and the organization's system of rewards and incentives. Finally, the *K-cultural infrastructure* involves elements such as corporate vision and the organization's system of values (Gold et al., 2001).

The last dimension of the K-strategy internal domain is K-skills. KM processes are by their very nature multifaceted. They involve many dimensions such as technical, organizational and human. This characteristic of KM processes reflects on the nature of skills required to perform them. For example, Malhotra (1997) defines a senior knowledge executive, such as a chief knowledge officer (CKO) or an organizational knowledge architect, as the person who should have the combined capabilities of a business strategist, technology analyst, and a human resource professional. The ability to facilitate the ongoing process of knowledge sharing and knowledge renewal, the ability to develop the human and cultural infrastructure that facilitates information sharing, and the ability to utilize the available technologies for serving the creation, sharing and documentation of knowledge are some examples of the required skills.

The Dynamics of KM Strategic Alignment Model (KMSAM)

Effecting a change in any single domain may require the use of three out of the four domains to assure that both strategic fit and functional integration are properly addressed. Therefore, applying KMSAM requires the identification of three domains: pivot, anchor and impacted

Figure 3. The dynamics of the strategic alignment process

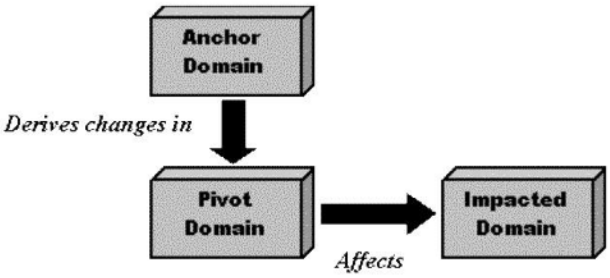
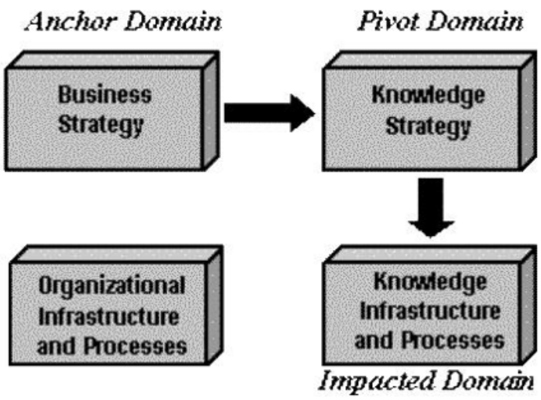


Figure 4. Knowledge potential perspective

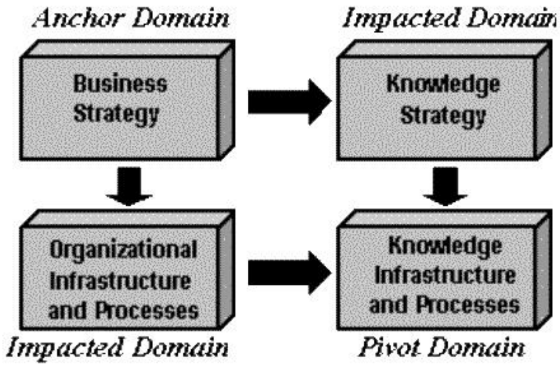


(Luftman, 1996). The pivot domain is the weakest and offers the greatest opportunity for improvement. The anchor domain is the strongest and will be the driver of change. Finally, the impacted domain is the area affected by a change to the pivot domain. Figure 3 shows the dynamics of the strategic alignment process.

Based on this distinction, different perspectives of strategic alignment can be identified. Each perspective represents a pattern of linkages between at least three elements of the four elements of KMSAM, that is, the two external domains (business strategy and knowledge strategy) and the two internal domains (organizational infrastructure and processes and knowledge infrastructure and processes). By identifying the strongest (anchor) domain and the *adjacent* weakest (pivot) domain, it becomes possible to identify the area that will be affected by the changes (the impacted domain). The direction the perspective flows is based on which domain is the strongest and which is the weakest.

For example, Figure 4 shows knowledge potential perspective in which business strategy, the strongest domain, derives changes to the adjacent weakest domain, knowledge strategy, and these changes will impact knowledge infrastructure and processes. In general, each alignment perspective has to include two types of relationships. The first is between external and internal domains of its business and knowledge components, that is, strategic fit. The second is the functional integration between business and knowledge domains. Eight single-path alignment perspectives can be then identified, namely: from anchor domain to adjacent pivot domain to impacted domain.

Figure 5. K-infrastructure fusion perspective



When the pivot and the anchor domains are not adjacent to one another, but rather across from each other on the diagonal, there will be two possible "paths" from the anchor domain to the pivot domain. This yields four fusion perspectives that result from fusing two of the eight single-path perspectives (Luftman, 1996). For example, Figure 5 shows K-infrastructure fusion perspective in which business strategy derives changes to the K-infrastructure and processes domain through organizational infrastructure and processes, and K- strategy domains.

Table 1 summarizes the 12 alignment perspectives.

CONCLUSION

Based on the premise that the realization of business value from KM investments requires alignment between the business and knowledge strategies and on the IT strategic alignment model (SAM) developed by Henderson and Venkatraman (1993), a KM strategic alignment model (KMSAM) is developed. Moreover, it provides executives with a logical framework for analyzing and assessing alternative strategic choices with regard to aligning K-strategy and B-strategy.

Extension of this work would move in two directions. The first would be to use KMSAM in cross-sectional

Table 1. KM strategic alignment perspectives

	Strategic Perspective \ Domain	Anchor Domain	Pivot Domain	Impacted Domain
1	**Strategy Execution**	Business Strategy	Organizational Infrastructure and processes	K-Infrastructure and processes
2	**Knowledge Potential**	Business Strategy	K-Strategy	K-Infrastructure and processes
3	**Competitive Potential**	K-Strategy	Business Strategy	Organizational Infrastructure and processes
4	**Service Level**	K-Strategy	K-Infrastructure and processes	Organizational Infrastructure and processes
5	**K-/Organizational Infrastructure**	K-Infrastructure and processes	Organizational Infrastructure and processes	Business Strategy
6	**K-Infrastructure/ K-Strategy**	K-Infrastructure and processes	K-Strategy	Business Strategy
7	**Organizational/ K-Infrastructure**	Organizational Infrastructure and processes	K-Infrastructure	K-Strategy
8	**Organizational Infrastructure/ Business Strategy**	Organizational Infrastructure and processes	Business Strategy	K-Strategy
9	**K-Infrastructure Fusion** (Perspectives 4 + 7)	Business Strategy	K-Infrastructure and processes	▪ Organizational Infrastructure and processes ▪ K-Strategy
10	**Organizational Infrastructure Fusion** (Perspectives 1 + 5)	K-Strategy	Organizational Infrastructure and processes	▪ Business Strategy ▪ K-Infrastructure and processes
11	**Business Strategy Fusion** (Perspectives 3 + 8)	K-Infrastructure and processes	Business Strategy	▪ Organizational Infrastructure ▪ K-Strategy
12	**K-Strategy Fusion** (Perspectives 2 + 6)	Organizational Infrastructure and processes	K-Strategy	▪ Business Strategy ▪ K-Infrastructure and processes

study of KM initiatives in order to identify the dominant patterns of K-strategy and B-strategy alignment. As "strategic alignment is not an event but a process of continuous adaptation and change" (Henderson & Venkatraman, 1993), the second direction would be a longitudinal study of each enterprise cycle around the alignment perspectives and how the adopted perspective is related to the degree of maturity of the KM initiative.

REFERENCES

Abou-Zeid, E. (2002). A knowledge management reference model. *Journal of Knowledge Management, 6*(5), 486-499.

Abou-Zeid, E. (2003). Developing business aligned knowledge management strategy. In E. Coakes (Ed.), *Knowledge management: Current issues and challenges* (pp. 156-172). IRM Press.

Buckman, R. (1998). *Lions, tigers and bears: Following the road from command and control to knowledge sharing*. White Paper. http://www.knowledge-nurture.com/

Earl, M. (2001). Knowledge management strategies: Toward a taxonomies. *Journal of Management Information Systems, 18*(1), 215-233.

Gold, A., Malhotra, A., & Segars, A. (2001). Knowledge management: An organizational capabilities perspective. *Journal of Management Information Systems, 18*(1), 185-214.

Grant, R. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal*, *17*(Winter Special Issue), 109-112.

Henderson, J., & Venkatraman, N. (1993). Strategic alignment: Leveraging information technology for transforming organization. *IBM Systems Journal*, *32*(1), 4-16.

Luftman, J. (1996). Applying the strategic alignment model. In J. Luftman (Ed.), *Competing in the information age* (pp. 43-69). Oxford University Press.

Malhotra, Y. (1997). Profile of the ideal knowledge manager/architect. *http://www.brint.com/wwwboard/messages/273.html*

Nahapier, J., & Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage. *Academy of Management Review*, *23*, 242-266.

Narasimha, S. (2000). Organizational knowledge, human resource management and sustained competitive advantage: Toward a framework. *CR*, *10*(1), 123-135.

Probst, G., Raub, S., & Romhardt, K. (2000). *Managing knowledge: Building block for success.* John Wiley.

von Krogh, G., Ichijo, K., & Nonaka, I. (2000). *Enabling knowledge creation: How to unlock the mystery of tacit knowledge and release the power of innovation.* Oxford University Press.

Zack, M.H. (1999). Developing knowledge strategy. *California Management Review*, *41*(3), 125-145.

KEY TERMS

Anchor Domain: The area that provides (drives, catalyzes, or enables) the change forces applied to the pivot domain.

Impacted Domain: The area affected by a change to the pivot domain.

K-Systemic Competencies: The set of utilization-oriented characteristics of knowledge that could contribute positively to the creation of new business strategy or better support of existing business strategy.

Pivot Domain: The problem or opportunity being addressed.

Alignment of Information Technology and Human Resources Strategies

K.F. Dery
University of Sydney, Australia

D.A. Samson
University of Melbourne, Australia

INTRODUCTION

Increased productivity within modern organizations over the past two decades has largely been generated by more effective use of information technology. However, the increase in technological capability has necessitated a wide range of organisational changes. These changes to process and work structure have implications for the recruitment, remuneration, reward, skill levels and general management of people. Therefore some of the greatest challenges have been associated with integrating and aligning information systems with the appropriate human resources management strategies.

While it is widely accepted that new information systems will have an impact on the management of human resources there is a significant lack of conceptual frameworks and practical guidelines to integrate all aspects of human resources policy with those of information systems strategy. HR and IT strategies should ideally be aligned in their competitive goals in order to maximise technology investment effectiveness.

This article provides a framework to:

(1) identify and analyse the degree of alignment between HR and IT and
(2) identify the enablers and inhibitors to the alignment process.

BACKGROUND

There is recognition in both academic and business practitioner literature that Human Resource (HR) management issues play a major part in the realization of improved business performance from Information Technology (IT) investment (Zuboff, 1988; Peters, 1988; Keen, 1991). While there is significant evidence of intent to align HR and IT activities, firms experience difficulties in the implementation process. There is therefore a need to better understand the relationship between the strategic planning and implementation of IT and the associated management of HR, to ensure that the firm has the capacity to maximize the opportunities enabled by technology innovations.

As contingency views of the role and activities of the HR and IT functions gained momentum, many management theories recognized the importance of aligning IT and HR with organization strategy to change work practices and increase productivity (Becker, Huselid, & Ulrich, 2001). New theories of management and organization process re-designs suggested by Business Process Redesign (Hammer & Champy, 1993; Davenport & Short, 1990), Knowledge Management (Nonaka, 1991; Ambrosio, 2000; Pfeffer & Sutton, 2000), and Virtual organizations (Ashkenas, Ulrich, Jick, & Kerr, 1995), were technology driven but required significant organizational change to implement successfully. People management was frequently cited as the reason for failures in the implementation of these management theories (Ambrosio, 2000), yet to date little is known about the issues that surround gaining alignment between the organizational demands of the technology and the strategic management of people.

CASE RESULTS AND ANALYSIS

A series of four Australian companies were examined in detail using case studies that described an IT project. While the unit of analysis was the firm for comparative purposes, the IT project was used as a lens to gain more specific insight into the minds and actions of executives' as the project was instigated, planned and implemented.

Each of these projects had significant implications for the HR strategy and therefore in each case we might expect to see some evidence of alignment in the planning activities and implementation programs between the IT and HR executives. Table 1 below summarizes the cases, the IT projects and implications for HR strategies.

A

Table 1. IT project selection for each company

	Mbank	PFS	WxPress	Snax Ltd
IT Project	On-line mortgage application system	On-line superannuation fund management system enhancement	On-line retail ordering system	R&D system to co-ordinate recipe development, ingredient sourcing and labeling
Implications for HR	Processes previously dominated by data entry and validating, now restructured with a greater focus on analysis and assessment skills. Additional opportunities to restructure jobs with part-time /job sharing rosters.	Organization built on corporate relationships moving to longer term relationships with individual policy holders. Increased complexity of consumer demands. Current bias favouring administrative skills moving to financial advisory skill requirements.	Customer contact staff initially employed for ability to communicate by telephone with customers, now required to move to a written medium and provide the same degree of advice and information.	Specialist operators detailing, proofing and checking compliance on labeling and recipe development will no longer be required in their current form. Potential for significant job redesign based on more co-operative, process driven systems.

(i) Identification and Analysis of the Degree of HR/IT Alignment

A series of unstructured interviews with executives in each company from IT, HR, user groups and senior executives were conducted. These were analyzed in conjunction with written evidence to create indicators that we might expect to observe in an organisation where HR and IT was aligned. The indicators included factors that suggested that (a) the organization intended to align the activities of HR and IT and (b) the behaviours that contributed to an organizational state of realized alignment. These two views of alignment are consistent with Reich and Benbasat's (1996, 2000) work that examined the intellectual (intended) and social (realized) perspectives of alignment, and will be further discussed in this paper.

Indicators of alignment were identified as:

a. Evidence of intellectual (intended) alignment between HR and IT strategies as evidenced by:

- written plans,
- internal consulting processes,
- reporting structures,
- regular meeting arrangements,
- other regular activities that suggest that the company's strategic requirements were being considered in the activities of the HR and IT functions.

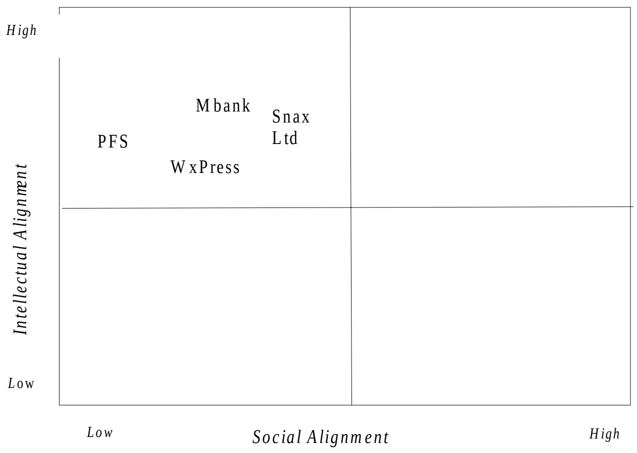
b. Evidence of social alignment (executive understanding and actions) as evidenced in the following factors:

- Regular communication between IT and HR executives.
- Company requirements that a representative from HR be included in all IT project teams.
- Written and verbal evidence that the CIO understood the importance of HR strategies to effectively implement the IT strategy.
- Understanding by the HR manager of the IT strategy and the implications of that strategy for the management of people in the organization.
- Recognition by the senior executives that having HR strategies built around the requirements generated by changes in IT was critical to realizing the rewards from the investment in the IT strategy.

Each case was assessed according to the alignment indicators identified above using both self assessments and documentary evidence. The findings were used to position each company on a spectrum of intellectual and social alignment (Figure 1).

Despite some indicators in written documents and discussion that alignment was intended between HR and IT, there was no evidence in the cases of systemic processes that ensured HR input into IT strategies and projects. Any social alignment observed was based on

Figure 1. Intellectual and social alignment ratings



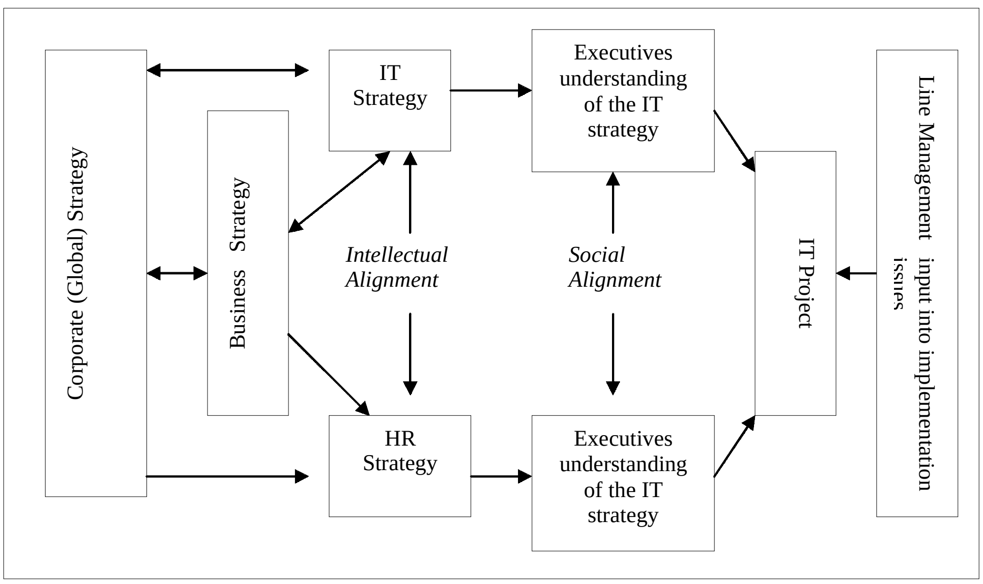
personal views of the HR managers' competencies and ability to add value. In three out of the four cases the HR manager was replaced (either during the research period or soon after) with a more experienced executive who not only had more strategic HR experience, but also a greater working knowledge of both the business and the industry. In Wxpress and Snax Ltd the stimulus for the new HR manager came from the acknowledged need for a well managed change management program in association with pending ERP implementations. Mbank's decision was a reactionary move to attempt to rescue their failing e-mortgage project.

While moves were in place in three of the cases to increase connectivity and effective communication between IT and HR, the smallest of the companies was progressing more slowly. It was recognized by the CEO of PFS that a lack of alignment was impeding their ability to maximize the returns from the IT investment, however, the size of the company made it difficult to employ an HR manager with the experience needed to have effective strategic input.

People management issues were typically devolved to line management (Figure 2) where the overwhelming response to the introduction of new technology was a training schedule. While training processes assisted in the short term implementation, there was little evidence of longer term strategic HR planning based on the expected impact of the IT strategy. In three of the companies the IT systems had not realised their expected returns and the fourth, Snax Ltd, dramatically changed its approach replacing both IT and HR managers as they embarked on a major strategic change program facilitated by new IT opportunities. As the CIO of Snax Ltd commented:

"It is no good assigning the responsibility for the people management issues and change process to the line manager because more often than not, it is the activities and responsibilities of that very person that have to change the most. The hardest thing for line management to let go of, is the power that comes with having more experience and knowledge than their staff. When everyone has access to the knowledge and the system takes over much of the processing, the line

Figure 2. Input and alignment indicators



manager is left searching for a new role and more often than not, they will hold on to the existing one."

(ii) Enablers and Inhibitors to the Social Alignment between HR and IT

Once a picture existed that indicated the nature of alignment between IT and HR, the Reich and Benbasat (1996, 2000) framework of social alignment, originally produced to examine the linkage between IT and business strategy, provided a valuable lens through which to examine the factors that influenced the understanding of executives' in the strategic alignment process. The framework was adapted to examine the inter-functional alignment between IT and HR and extended to reflect the research results.

The notion of social alignment was designed to focus on executives' understanding of each other's objectives and plans, as distinct from the intellectual intent of written documents (Figure 2). Given the nature of functional strategies, which are not necessarily deliberate and expressed in detailed, written plans (Lederer & Mendelow, 1989) but may emerge out of the collective actions of executives' (Mintzberg, 1987), and also may not necessarily be performed by a designated department that is readily observable (Ohmae,1982), the study of beliefs and behavior was important .

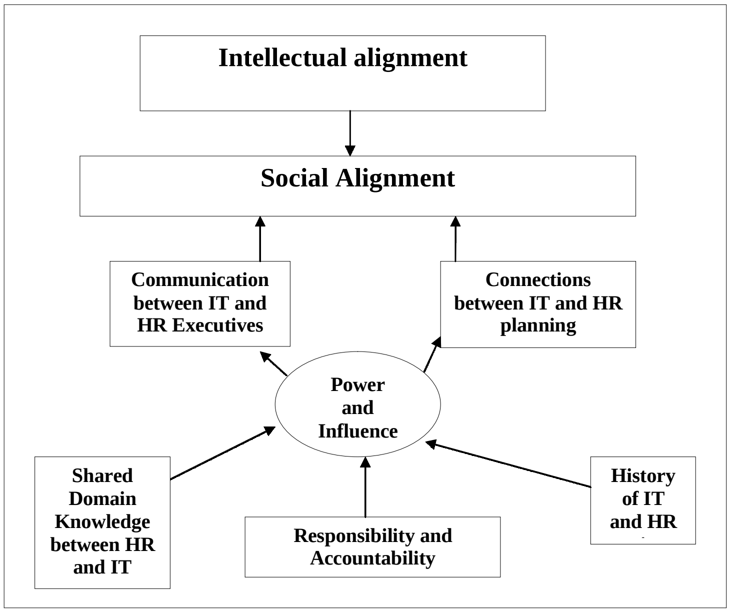
Reich and Benbasat (1996) defined the construct of alignment on two dimensions (adapted to examine the relationship between IT and HR):

- The intellectual dimension exists when the content of information technology and human resource strategies are consistent and externally valid.
- The social dimension exists when IT and HR executives understand and are committed to each others plans.

In order to examine the notion of social alignment in a meaningful and relevant manner a series of enabling factors were explored. These were all found to be significant in creating executives' understanding of each others strategies and therefore the necessary actions required in order to align their activities. Communication and planning connections supported by shared domain knowledge and a history of implementation success were identified as critical factors determining the degree of social alignment.

The revised map in Figure 3 (Dery, 2003) bears a close resemblance to that proposed by Reich and Benbasat (2000) but provides insight into factors that make the roads impassable at times and provide directions to help HR and IT work more closely together to reach their common goal.

Figure 3. Proposed framework of analysis to create a greater understanding of the alignment process between HR and IT (Dery, 2003)



Shared knowledge was the most powerful factor that had the potential to clear road blocks and help executives to visualise a way forward together. The HR executives initially involved in the IT projects examined had a limited knowledge of the business and their knowledge of IT was restricted to personal computing at best. New HR management in Mbank, Snax Ltd and Wxpress were significantly more experienced in the business processes and, with the exception of Wxpress, could contribute knowledgably to IT discussions. These new HR recruits were significantly more active in IT projects than their predecessors.

Both shared knowledge and past history are recognised bases for influence in the literature. Bacharach and Lawler (1980) identified these bases as expertise and personal characteristics. Specialist knowledge, or expertise, is a source of powerful influence as others turn to the specialist for knowledge that they do not possess. Thus specialist knowledge generates power based on dependence. Personal characteristics enhance the credibility of the executive's input and enables perspectives to be swayed in the desirable direction. The characteristics connected with reputation (Bacharach & Lawler, 1980) based on past experiences are recognized as important bases for influence.

Evidence suggested that the balance of power was generally tipped in favor of IT in the cases analyzed. The HR managers in all cases entered the "power and influence round- about" low on IT and business knowledge which reduced their ability to apply their specialist HR knowledge in an influential way. Past experiences of HR in IT departments had not been positive in all but one case (PFS). These experiences had a negative impact on the HR executives' personal credibility with the CIO and further reduced their ability to influence discussions. The paths to further communication and participation in planning activity were thus difficult to negotiate. In the three cases where HR management had been more recently replaced the evidence suggests that their increased knowledge had enabled them to exert more specialist power and thus smooth the road to improved communication and planning input. The HR managers in Mbank and Snax Ltd had moved into their new positions with proven track records with the CEO's of those companies and reputations as experienced and business-savvy executives.

When line management were charged with the task of assessing and managing the people issues, the power bases appear to adjust in their favor as they leveraged off specialist operations knowledge and, in most cases, had more IT knowledge than the HR manager. The CIO in all cases, except Snax Ltd, felt comfortable with referring people issues to line management who were generally more integrally involved with the IT project and had significantly more influence in project discussions than the HR Manager. The danger as documented in Mbank, and recognized by the CIO at Snax Ltd, is that old practices may simply be automated and anticipated benefits from the IT investment are inhibited by a lack of strategic HR thinking.

"It is very hard for line managers to envisage how the new order might look when they themselves are likely to be facing the most significant change. Most line managers source their power from their specialist operations knowledge so when all their employees can access the same information they tend to feel very threatened. CIO Snax Ltd."

When HR managers abrogated full responsibility to the line or project managers they offered little specialist support other than advice on compliance or redundancy programs. There was no link evident that suggested that the line or project manager was managing the people issues under the umbrella of the HR strategy. In fact, it was apparent that the only strategic input from HR, when a line or project manager took full responsibility for an IT project, was by default from the general managerial planning processes. This level of input was at best unreliable and at worst non-existent.

Responsibility and accountability were also recognized as important contributing factors. There was no evidence in the cases that indicated that CIO's were held accountable for an IT project beyond the development, technical implementation and support for the system itself. IT had no responsibility or accountability for the people management factors that were critical to implementation. Nor was there any evidence to suggest that HR were held accountable beyond the delivery of compliance advice. Line management were ultimately responsible and accountable for effective implementation but this was a short term response to ensuring that users accepted and could use the system effectively. Expectations were regularly revised and there was little attention paid to longer term strategic opportunities.

FUTURE TRENDS

The focus for IT research has moved to large, integrating systems such as ERP's which have significant implications for organisational change. People management challenges are increasingly creating obstacles to successful implementations and organisations are searching for new ways to enable greater alignment between the IT strategies and the organisational change process. The findings of this research add value to these discussions and provide a practical framework to more closely integrate the strategic planning and activities between HR and IT.

CONCLUSION

The evidence from these cases suggests a non-systemic approach to aligning IT and HR leaves the relationship vulnerable and often abrogated to line management. There is significant evidence to suggest that HR management is frequently unwilling to take a management role in IT projects and IT management seldom sees their role extending beyond system development. Organizations that invest in educating both the HR manager and the CIO and support them to maximize opportunities to align their activities are more likely to gain competitive leverage from their IT investment than those who assume that alignment will occur as a result of the general planning process and who abrogate strategic HR input for IT projects to the line.

REFERENCES

Ambrosio, J. (2000, July 3). Knowledge management mistakes. *Computerworld*, (1), 44.

Ashkenas, R., Ulrich, D., Jick, T., & Kerr, S. (1995). *The boundaryless organisation – Breaking the chains of organisational structure*. San Francisco: Jossey Bass.

Bacharach, S.B., & Lawler, E. (1981). *Power and politics in organisations*. London: Jossey Bass.

Becker, B. E., & Huselid, M.A. Ulrich, D. (2001). *The HR scorecard: Linking people, strategy, and performance*. Boston, MA: Harvard Business School Press.

Davenport, T.H. & Short, J.E. (1990). The new industrial engineering: IT and business process redesign. *Sloan Management Review*, *31*(4), 11-28.

Delaney, J.T. & Huselid, M.A. (1996). The impact of human resource management practices on perceptions of organisational performance. *Academy of Management Journal*, 39(4), 949-969.

Dery, K.F. (2003). *How do organisations align human resource management with information technology? An exploratory study of four Australian firms.* PhD Thesis (unpublished).

Dery, K.F. (2003, July). Alignment of human resources and information technology: A case study of a major Australian bank. *Pacific Asia Conference on Information Systems*, Melbourne: Victoria University.

Hammer, M., & Champy, J. (1993). *Re-engineering the corporation: A manifesto for business revolution*. London: Nicholas Brealey Publishing.

Henderson, J.C., & Venkatramen, N. (1993). Strategic alignment: The emerging concept for leveraging information technology for transforming organisations [Special Issue]. *IBM Systems Journal*.

Lederer, A.L. & Mendelow, A.L. (1989, Fall). Coordination of information systems plans with business plans. *Journal of Management Information Systems*, *6*(2).

Mintzberg, H. (1987). Crafting strategy. *Harvard Business Review*, *65*(4), 66-75.

Nonaka, I. (1991, November/December). The knowledge-creating company. *Harvard Business Review*, 96-104.

Ohmae, K. (1982). *The mind of the strategist: The art of Japanese business.* McGraw Hill Book Company.

Peters, T. (1988). *Thriving on chaos: A handbook for a management revolution*. New York: Alfred A. Knopf.

Pfeffer, J., & Sutton, R. (2000). *The knowing doing game: How smart companies turn knowing into action*. Boston, MA: Harvard Business School Press.

Reich, B.H., & Benbasat, I. (1996, March). Measuring the linkage between business and information technology objectives. *MIS Quarterly*, 55-81

Reich, B.H., & Benbasat, I. (2000, March). Factors that influence the social dimension of alignment between business and information technology objectives. *MIS Quarterly*, *24*(1), 81-113.

Zuboff, S. (1988). *In the age of the smart machine: The future of work and power*. New York: Basic Books.

KEY TERMS

Alignment: An orientation based on a series of ongoing processes based on the will to achieve a common goal. For the purposes of this research alignment is distinguished from the more static concept of fit. Alignment is conceptualized as an outcome with determinant factors that affect the degree to which one set of objectives and mutually consistent and supportive of the other.

Connections in Planning: Refers to the degree of contact between HR and IT executives in their respective planning processes.

History of Implementation Success: Describes the track record of IT projects in HR and the effectiveness of HR into IT people management issues.

Human Resources (HR): All activities of planning, staffing, appraisal and remunerating, improving employees and the working environment and establishing and maintaining working relationships.

Information Technology (IT): All hardware, software, communications, telephones, facsimiles, all personnel and resources dedicated to the delivery of information and processes via computerised mediums.

Intellectual Alignment: Describes a state where the content of IT and HR business plans are consistent and externally valid.

Shared Domain Knowledge: Refers to IT and HR executives' understanding of the business and of the knowledge in HR of the IT strategies and vice versa.

Social Alignment: Describes a state where IT and HR business executives understand and are committed to each others plans.

An ERP Life-Cycle Cost Model

A

José Esteves
Instituto de Empresa, Spain

Joan Pastor
Universidad Politécnica de Catalunya, Spain

INTRODUCTION

Nowadays, the enterprise resource planning (ERP) products and services industry is one of the most promising. Through the usage of ERP systems such as SAP, BAAN, Peoplesoft and Oracle, organizations try to integrate the information flow of the different business areas and, at the same time, improve efficiency and reduce costs. Theoretically, these integrated systems provide large functionality in terms of problem solving associated with data flow when they are integrated with different software systems.

Some critiques of ERP systems focus essentially on the high costs of ERP projects, the high failure rates and their complexity, which makes it difficult for users to perceive the benefits and the opportunities of these systems. ERP systems demand a large investment at the economic, human resource and organizational levels. This investment is made not only in the initial phase but also throughout their life-cycle. Thus, this study is intended to analyze the necessary investment to integrate an ERP system during its life.

BACKGROUND

One of the issues in the ERP systems area is that of determining if the investment made will be compensated in the future. A survey of Meta Group Inc. (Craig, 1999) shows that in financial terms, ERP projects cost more than the expected benefits. This is close to academic studies on this object. But there are some suggestions that financial terms and other strategic benefits should be considered. This issue is included in broader discussions about investment in information systems (IS) and their performance (Bender, 1986; Brynjolfsson & Hitt, 1994; Harris & Katz, 1988). Some authors found no or little relationship between them (Strassmann, 1990, 1999; Turner, 1985), while others concluded that the investment in IS has been detrimental to organizations (Roach, 1988).

A study conducted by Meta Group (2000) showed that only 20% of companies that have implemented ERP solutions actually know the total cost of ownership (TCO) of their deployments. The other 80% do not fully understand the magnitude of on-going support and infrastructure-related costs. Our study analyzes the necessary investments for the integration of an ERP system during its existence, along with an analysis of the costs associated with each phase.

ERP SYSTEMS LIFE-CYCLE

To define the ERP life-cycle model we use a simplified version of the model proposed by Esteves and Pastor (1999). Other authors have proposed models for ERP systems but focusing on the ERP implementation phase. However, there is a misunderstanding in relation to the concept of implementation. Our model is structured in phases and dimensions. Here, we only make reference to the phases as the different stages of the life-cycle of an ERP system in an organization. Next, we describe each phase, that is, adoption, acquisition, implementation, usage and maintenance, evolution and retirement.

- **Adoption decision phase:** During this phase managers examine the need for a new ERP system while selecting the general IS approach that will best address the critical business challenges and improve the organizational strategy. The decision phase includes the definition of system requirements, its goals and benefits, and an analysis of the impact of adoption at a business and organizational level.
- **Acquisition phase:** This phase consists of the product selection that best fits the requirements of the organization, thus minimizing the need for customization. A consulting company is also selected to help in the next phases of the ERP lifecycle, especially in the implementation phase. Factors such as price, training and maintenance services are analyzed and the contractual agreement is defined. In this phase, it is also important to make an analysis of the return on investment of the selected product.
- **Implementation phase:** This phase consists of the customization or parameterization and adaptation of the ERP package acquired according to the needs

of the organization. Usually this task is done with the help of consultants who provide implementation methodologies, know-how and training.

- **Use and maintenance phase:** This phase consists of the use of the product in a way that returns expected benefits and minimizes disruption. During this phase, one must be aware of the aspects related to functionality, usability and adequacy to the organizational and business processes. Once a system is implemented, it must be maintained, because malfunctions have to be corrected, special optimization requests have to be met, and general systems improvements have to be made.
- **Evolution phase:** This phase corresponds to the integration of more capabilities into the ERP system, providing new benefits, such as advanced planning and scheduling, supply-chain management, customer relationship management, workflow, and expanding the frontiers to external collaboration with other partners.
- **Retirement phase:** This phase corresponds to the stage when with the appearance of new technologies or the inadequacy of the ERP system or approach to the business needs, managers decide if they will replace the ERP software with other information system approaches more adequate to the organizational needs of the moment.

COSTS ALONG THE ERP LIFE-CYCLE

The literature on ERP research shows a lack of studies on ERP system costs and investments (Esteves & Pastor, 2001). Based on published case studies and literature review related to the cost analysis of IS, we developed a structure of costs along the ERP life-cycle. Next, we conduct an exploratory case study to validate this structure of costs (for more details see Esteves et al., 2001).

Table 1 summarizes the cost items, where costs are classified as tangible or intangible. Thus, tangible costs are the costs that can be measured in a direct way, taking always into account that, sometimes, tangible costs cannot be measured in monetary terms. Intangible costs are those costs that are difficult to be measured in a direct way, since they refer to vague concepts, as illustrated in Table 1. Next, we describe in detail each phase and related costs.

ADOPTION PHASE

Tangible Costs

- **Decision-making costs:** This phase is perhaps the phase that has the least associated costs, because it represents only the decision to adopt or not an ERP system. The associated cost is essentially the time spent by managers in the decision-making task. In the context of the decision-making process, the concept of avoidable costs is used to define the costs that can be eliminated when we opt for a specific choice or solution. The unavoidable costs refer to the costs that we cannot eliminate.
- **Consultancy costs:** Consultants help with knowledge and experience in the evaluation of the existent IS architecture and the definition of the new IS strategy.

Table 1. Cost items along the ERP life-cycle

Phase	Tangible Costs	Intangible Costs
Adoption	Consultancy	Decision-making costs
Acquisition	Consultancy Hardware Software licenses	Decision-making costs Opportunity costs
Implementation	Consultancy Training Human resources System specification	Customization, conversion and data analysis Time dedicated by staff Business process reengineering
Usage and Maintenance	System reconfiguration System adaptation Cost of system failure	Indirect costs of system failure Lost of competitiveness
Evolution	Cost of new applications Consultancy	
Retirement		Opportunity costs Decision=making costs

ACQUISITION PHASE

Tangible Costs

- **Consultancy costs:** Again, consultants help with knowledge and experience in the selection of the ERP system most adequate to the organizational needs, and they help in the implementation phase. They also act as mediators between the ERP provider and the organization. Analysts are unanimous in stating that the costs related to consultancy are the highest for an ERP system.
- **Hardware costs:** Acquisition of an ERP system implies changes in the existing hardware infrastructure. These changes can vary from a change and/or actualization of the actual hardware infrastructure to the complete installation of a new hardware infrastructure. Hardware needs must not be forgotten due to the fact that the new hardware infrastructure must have the maximal performance to provide access in real time to databases and the ERP system modules. Aspects such as network-communication capacity, servers and processing-speed rates are important.
- **Software licenses:** After the selection of ERP software, there is the need to make an agreement contract. The cost of this contract can be calculated in several ways. Some contracts consider the number of licenses and the number of users, while others include factors such as organization benefits and number of employees. Usually, providers offer price reductions depending on the number of modules acquired and the acquisition of extended applications and the purchase of maintenance services and upgrades of software.

Intangible Costs

- **Decision-making costs:** In the acquisition phase, decision-making costs must be analyzed again. The main decision consists of the selection of the ERP system that best addresses the organization needs.
- **Opportunity costs:** These costs measure the opportunity that is lost or is sacrificed when an option is abandoned. The analysis of the several ERP systems that exist in the market should take into account these costs.

Analysts recommend that at this point a return on investment (ROI) study should be done. The ROI has two important elements: the “how much” and the “when”. The “how much” element represents the benefits created by the investment, and the “when” represents the period of investment return.

IMPLEMENTATION PHASE

Tangible Costs

- **Consultancy costs:** Most organizations use consultants to implement the ERP system, due to the fact that they do not have the technical knowledge in-house to complete this process. Consultants incorporate knowledge, experience, implementation methodologies, and training programs for users. They also help re-engineering the organization and its business processes.
- **Training costs:** Here, we include the training of a project team and the end users. These costs are usually high because users need to learn a set of new processes and not just the usage of a new software application. To minimize these training costs, there presently exist Internet-based tools or videoconference tools that reduce the number of trainers and have a larger scope of training.
- **Human-resources costs:** The costs associated with the project team have a high weight. They are expert professionals and their price per hour is high and must be quantified. Sometimes there is the need to recruit these human resources from outside.
- **System-specification costs:** These costs refer not only to the number of human resource hours spent on this task but also the acquisition of tools and instruments (denominated enterprise modeling tools) that help to make specific the necessary business vision.

Intangible Costs

- **Customization, conversion and data analysis:** The effort made in the software customization and adaptation to organization needs is usually not measured, except in terms of time. The data to be inserted in the new system constitute a heavy burden to the project costs. The data converted from other systems and the new ones have to be verified due to inconsistencies, because usually they are not adequate to the new formats. Usually, there is the need to insert new data in the new system.
- **Time of dedicated staff:** The project team and staff have to keep many times their work and make the tasks related with the project. The costs associated with this effort and the respective loss of efficiency

are not usually measured. The option of keeping them totally dedicated to the project could result in the recruitment of personnel, whose costs are measurable.

- **Business-process re-engineering cost:** Forrester Research Inc. (Koch, 1997) estimated in 1997 that in a typical SAP installation, more than 50% of the implementation budget is spent in business process re-engineering. The implementation of an ERP system means in many cases a radical change in the organization business processes, forcing changes in organizational flows and in organizational structures.

USAGE AND MAINTENANCE PHASE

Tangible Costs

- **System reconfiguration:** Usually, after ERP system implementation, some inefficiency appears in its operation, which requires an adjustment to the ERP configuration. Reporting functions are particularly sensitive to reconfiguration needs.
- **System adaptations:** Organizations change over time, so there is always therefore the need to adapt the systems to whatever changes have been made. The introduction of new information flows and new business processes influence the ERP system usage and performance.
- **System-failure costs:** Failure costs refer to the costs arising from ERP system breakdowns. In big organizations the direct costs are very high, because orders cannot be satisfied, stocks cannot be updated, and in certain cases these failures cause a total disruption of the organization's business.

Intangible Costs

- **Indirect costs of system failure:** System failures cause many indirect costs, such as loss of access to vital information, loss of customer loyalty, or the ability to survive in the business (Uram, 1999).
- **Loss of competitiveness.** A common criticism of ERP systems is the standardization of business models through the imposition of ERP models. Most organizations adopt these business models to facilitate the software implementation and they may thereby lose competitive advantages over other organizations. These changes should be measured in economic terms.

EVOLUTION PHASE

Tangible Costs

- **Cost of new applications:** In this phase the main costs are related to the purchase and implementation of new applications, such as advanced planning and scheduling, customer relationship and management, workflow, e-commerce, and so forth.
- **Consultancy costs:** Most organizations use consultants to implement these new applications, due to the fact that they do not have the technical knowledge in-house to complete this process.

RETIREMENT PHASE

Here, the opportunity and decision-making costs repeat, along with all the tangible costs related with software. When this phase is reached, it is expected that the original investment would be recovered. In cases of abandonment (in implementation phase, for instance), the associated costs can constitute high losses, as happened to the FoxMeyer company (Scott, 1999). The company imputed its bankruptcy to the investment made in an ERP system.

CONCLUSION

This study attempts to analyze the costs associated with ERP systems along the typical ERP life-cycle phases. The article presents a list of costs that demonstrate that this type of system has a strong organizational perspective, and therefore the costs associated with this perspective should be taken into account and measured. The list does not intend to be exhaustive but attempts to give a general view of the whole ERP life-cycle. As for future research, we plan to validate this model with case studies and surveys to different organizations. We attempt to analyze the differences among industry sectors, size of companies, and countries. Another issue is the return on investment (ROI). With regard to this issue we attempt to study the reasons to measure ROI, how companies measure it, and the ROI results.

REFERENCES

Bender, D. (1986). Financial impact of information processing. *Journal of Management Information Systems, 3*(2), 232-238.

Brynjolfsson, E., & Hitt, L. (1994). *Paradox lost? Firm-level evidence of high returns to information systems spending*. Working Paper n. 162. The Center for Coordination Science, MIT Sloan School.

Caruso, D. (1998). *ERP vendors- AMR Research's top 20 roundup. The report on enterprise applications* (1st ed.). AMR Research Inc.

Craig Stedman. (1999). Survey: ERP costs more than measurable ROI. *Computerworld. http://www2.computerworld.com/home/print.nsf/all/9904059CFE*

Davenport, T. (1998, July/August). Putting the enterprise into the enterprise system. *Harvard Business Review*, 121-131.

Esteves, J., & Pastor, J. (1999). An ERP life-cycle-based research agenda. *1st International Workshop on Enterprise Management Resource and Planning Systems (EMRPS)*, Venice, Italy (pp. 359-371).

Esteves, J., & Pastor, J. (2001, August). Enterprise resource planning systems research: An annotated bibliography. *Communications of the Association for Information Systems (CAIS), 7*(8).

Esteves, J., Carvalho, J., & Santos, A. (2001). Towards an ERP life-cycle costs model. *Information Resources Management Association (IRMA) international conference*, Toronto, Canada (pp. 431-435).

Hagendorf, J. (1999, December). ERP partnerships remain elusive for distributors. *Computer Reseller News, 872.*

Harris, S., & Katz, J. (1988). Profitability and information technology capital intensity in the insurance industry. *21st Annual International Conference on System Sciences* (pp. 124-130).

Kale, V. (2000). *Implementing SAP R/3: The guide for business and technology managers*. SAMS Publications.

Koch, C. (1997, April). Lump it and like it. *CIO Magazine*.

Marion, L. (1998, May). The 15% solution. *Datamation*.

Meta Group. (2000). *ERP platform-related analysis total cost of ownership study*. White paper. Meta Group Consulting.

Roach, S. (1988). Technology and the service sector: The hidden competitive challenge. *Technological Forecasting and Social Change, 34*(4), 387-403.

Ross, J., & Vitale, M. (1998). *The ERP revolution: Surviving versus thriving*. Research paper. Center for Information Systems Research, Sloan School of Management, MIT.

SAS Institute. (1998). *Increasing ROI from ERP systems*. SAS Institute Inc.

Scott, J. (1999). The FoxMeyer Drug's bankruptcy: Was it a failure of ERP? *Americas Conference on Information Systems (AMCIS)*, Milwaukee.

Slater, D. (1999, February). An ERP package for you...and you...and you...and even you. *CIO Magazine*.

Strassmann, P. (1990). *The business value of computers*. New Canaan, CT: Information Economics Press.

Strassmann, P. (1999). The search for productivity. *Computerworld, 33*(32), 52.

KEY TERMS

Business Process Reengineering (BPR): This is related with the alignment between business processes and the ERP business model and related best practices. This process will allow the improvement of the software functionality according to the current and future organization needs. Managers have to decide if they do business process reengineering before, during or after ERP implementation.

Consultancy: The process of helping organizations to better understand complex processes through their knowledge and experience and provide solutions to achieve the objectives. Consultants may help organizations in the whole ERP life-cycle. The usage of external consultants will depend on the internal know-how that the organization has at the moment.

Data Conversion: This process involves moving the existing data from the legacy systems, performing an integrity check and then transporting them to the ERP system that has been implemented. There are diverse tools available to reduce the effort on conversion process. In some occasions, there is the need to introduce manually the data into the system. Due to this effort, the conversion process must be planned at the beginning of the ERP project in order to avoid delays.

Enterprise Resource Planning (ERP): Typically, ERP systems are software packages composed of several modules, such as human resources, sales, finance and production, and providing cross-organization integration of transaction-based data throughout imbedded business processes. These software packages can be customized to the specific needs of each organization.

ERP Life-Cycle: Consists of the several stages that an ERP system goes through during its whole life within the hosting organization.

Intangible Costs: Those costs that are difficult to be measured in a direct way, since they refer to vague concepts.

Return on Investment (ROI): A technique for measuring the return on investment. It is calculated by the earnings produced by the assets compared to the money invested in the asset. This technique is used in ERP contexts to justify ERP adoptions and also to measure how well an ERP was implemented.

Tangible Costs: Those costs that can be measured in a direct way, taking always into account that sometimes tangible costs cannot be measured in monetary terms.

Total Cost of Ownership (TCO): Accounts for all of the costs associated with procuring, deploying and operating an existing asset like an ERP system at a given point in time. Microsoft defines TCO as: "refers to the administrative costs associated with computer hardware and software purchases, deployment and configuration, hardware and software updates, training, maintenance, and technical support" (Microsoft glossary: www.microsoft.com/windows2000/techinfo/howitworks/management/glossary.asp).

A

Analyzing the Quality of Virtual Teams

Robert M. Verburg
Delft University of Technology, The Netherlands

J.H. Erik Andriessen
Delft University of Technology, The Netherlands

Joris P.G. de Rooij
Delft University of Technology, The Netherlands

INTRODUCTION

Global market developments and the large-scale use of diverse applications in the area of information and communication technology (ICT) have been key factors in the emergence of distributed teams. Such teams are often referred to as virtual teams. Virtual teams enable collaboration between people across traditional boundaries and offer tremendous opportunities for various achievements. Businesses are no longer tied to a single time zone and are, for example, able to develop software around the 24-hour clock. The Internet—as the almost universal medium for interaction across boundaries—has created an infrastructure that enables many organizations to launch virtual teams. Hardly any technical obstacle for communication and collaboration across geographic boundaries remains, as these processes are supported by high-tech collaboration solutions such as groupware and other collaborative applications (e.g., videoconferencing, electronic blackboards). Virtual teams have a number of opportunities that are not found with co-located teams, such as involving rare expertise.

For example, a group of eight scientists from different organizations rapidly developed a revolutionary rocket engine design by working under geographically dispersed conditions and without prior work relationships (Majchrzak, Rice, Malhotra, King & Ba , 2000). The complex and innovative design could not have been developed without the expertise of the eight highly specialized scientists. However, the design was not only a result of a careful combination of expertise, but required a number of interdependent iterative 'virtual' brainstorming sessions among the team of rocket scientists. All these activities were performed through a collaboration tool called "the Internet notebook," whereby the specialists spend no more than 15% of their time on the project.

As the above example illustrates, virtual teams have the advantage of bringing people together without the obvious constraints with regard to travel time, workspace, and socialization. Virtual teams perform a variety of tasks and are also defined in various ways. The term generally implies groups of geographically and culturally dispersed co-workers using a combination of communication and information technologies to accomplish an organizational task (Townsend, DeMarie & Hendrickson, 1998; Jarvenpaa, Knoll & Leidner, 1998). Hutchinson (1999) distinguishes three types of virtual teams: intra-organizational teams, inter-organizational teams, and inter-organizational distributed teams. In this overview we will have all three types in mind when discussing our approach.

BACKGROUND

Being 'virtual' is a matter of degree and refers, according to various authors, to dimensions such as spatial distance, time, cultural diversity, temporality, organizational contract, and mode of interaction (Mowshowitz, 1997; Jarvenpaa & Leidner, 1998; DeSanctis, Staudenmayer & Wong, 1999; Vartiainen, 2003). Mode of interaction is an important dimension. Some teams meet regularly face to face, but may also have some e-mail-based interaction, while other teams interact intensively and almost exclusively via various media and sophisticated groupware tools. Geographical distance and different timeframes may obviously be important reasons for groups to communicate electronically.

'Virtuality' refers to the extent to which a group is geographically distributed, is organizationally and culturally diverse, has different timeframes for work, communicates electronically ('mode of interaction'), and whose members are freelance or have fixed contracts with an organization. The degree of reliance on ICT, its availability, and the proficiency of the users are very important for virtual teams (Dubé & Paré, 2004). The more of the above, the more a team is considered to be a virtual group. 'Virtuality' is the highest in globally dispersed teams of culturally diverse members of different organizations (or freelancers) that interact temporarily and communicate exclusively via electronic means.

A useful definition of a team (or work group) is a collection of individuals who see themselves and who are seen by others as a social entity, who are interdependent because of the tasks they perform as members of a group, who are embedded in one or more larger social systems (e.g., community, organization), and who perform tasks that affect others (Guzzo & Dickson, 1996). Although often not defined, a number of implicit characteristics of conventional teams seems to include that members are often permanent employees of one organization, are often co-located, and the main form of interaction consists of face-to-face contact.

Virtual teams may not seem to be crucially different from co-located teams. There are comparable levels of responsibility for adequately performing basic processes of groups, such as information sharing, cooperation, coordination, and team building. Virtual teams also have to mobilize the necessary resources, and need to develop a cohesive team with clear goals. However virtual teams have to care for these processes under conditions of less-than-optimal communicative channels, of higher distance in time, space, and culture than face-to-face teams. Inadequate ICT tools or infrastructures and the incompatibility of technology will result in barriers for cooperation. But with sufficient attention to team building and adequate ICT tools, these problems may be overcome. The process of team building can be difficult in the virtual context, specifically when the 'life cycle' of a team is short, the stability of membership is limited, and face-to-face meetings are scarce.

Global virtual teams have to deal with the additional issues of communicating across different time zones, languages, and cultures (e.g., Dubé & Paré, 2001; Montoya-Weiss, 2001).

Other problems may include missing non-verbal cues in communication and a lack of unplanned social encounters, resulting in problems with 'awareness' of availability and state of others, of progress of the work, or of the setting in which others work (see, e.g., Steinfield, 2002). These barriers may result in a lack of trust and cohesion, which often may lead to lower performance levels. Jarvenpaa and Leidner (1998) confirmed that global virtual teams might start with a form of 'swift' trust (Meyerson, Weick & Kramer, 1996), but that such trust appears to be fragile and temporal. Cramton (1997) illustrates, for instance, the multiple interpretations members of virtual teams may give to the meaning of silence of their distant team members. Additionally, virtual team membership can be highly fluid, demanding continuous adaptation processes between the existing team and new members, who bring their own beliefs and frame of reference. It is this system of shared views and beliefs people hold that is often considered very important for team functioning. The specific situation of a virtual team hinders the development of such a system.

TEAM PERFORMANCE

A crucial difference between co-located and virtual teams is the fact that virtual teams have the opportunity to combine and integrate both co-located and distributed interaction. Virtual teams may combine the better of two worlds and may therefore have an advantage over conventional teams. Virtual teams require certain tools in the area of information and communication technology to support interaction. Some modern tools have sophisticated functionalities that provide such teams with opportunities that conventional teams do not have. One of the major effects of the introduction of collaboration technology has been that certain types of meetings can now be held with a large number of participants. Moreover, some tools allow for easy storage and retrieval of information and for collaborative editing of documents.

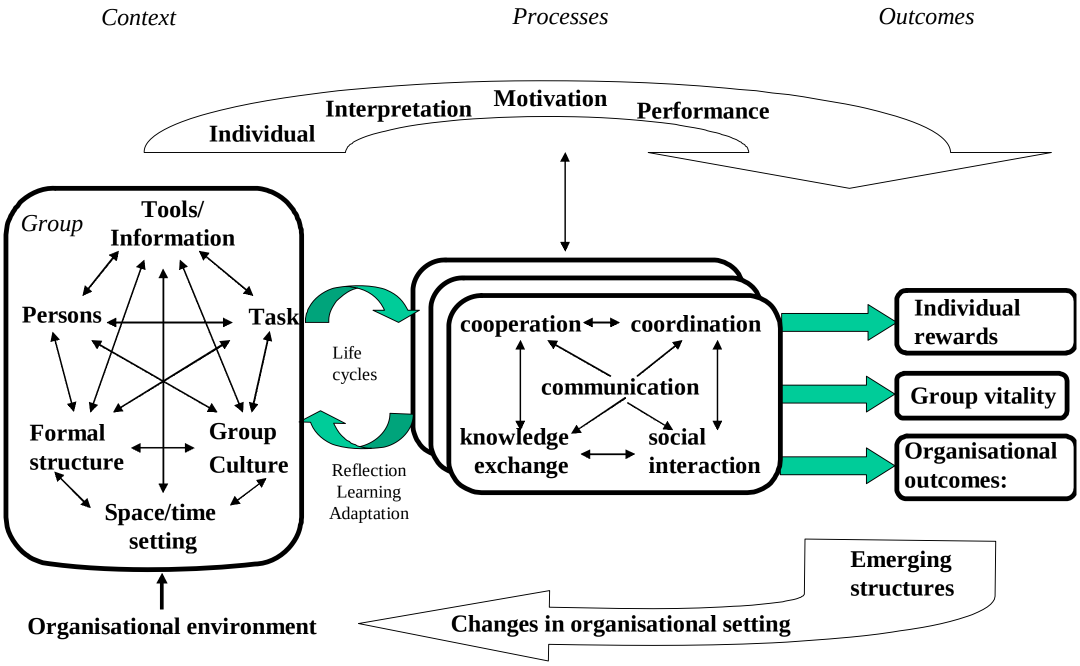
So far, the development of virtual teams has mostly been technology driven, almost neglecting other aspects of work, such as knowledge sharing, combining expertise, and dividing tasks. As a consequence, the performance of many virtual teams is far below their potential, thus producing poor business results (e.g., Jarvenpaa & Leidner, 1998).

In order to reach this optimal level of functioning, these new types of collaboration require new ways of organizing and managing. Major challenges for both managers and employees are the consequences of dealing with virtual teams. Systematic insight in the design and performance of effective (global) virtual teams is therefore an important prerequisite. It is clear that virtual teams may face substantial barriers for effective cooperation and that the probability of failure is ever present. The next section presents a model for analyzing the reasons for failure and can support the design of virtual groups.

ANALYZING VIRTUAL TEAMS: A MODEL

The model is based on a general model of group functioning, called the Dynamic Group Interaction model (DGIn-model), which is applied in several case studies (Andriessen, 2002; Andriessen & Verburg, 2004). The purpose of this model is not to limit the analysis of collaborative activities to specific aspects, but to structure the analysis by providing ideas and insights that have proven their value in other contexts.

Figure 1. Adapted from the Dynamic Group Interaction model (DGIn-model) (Andriessen, 2002)



In this model, elements of several theories are brought together. Three levels of behavior are taken into account: individual goal-directed behavior and cognitive processes (based on Action Theory, Activity Theory, Media Richness Theory); interpersonal and group processes (Activity Theory, Adaptive Structuration Theory, Social Information Theory, Coordination Theory); and a macro-social perspective (Structuration Theory). The various notions are brought together in a heuristic model concerning group processes, related to traditional 'input-process-output' schemas (see, e.g., McGrath, 1984; Hackman, 1987; Kraemer & Pinsonneault, 1990; McGrath & Hollingshead, 1994). However, they are enriched with interpretative and structurational notions and feedback cycles (see Figure 1).

The DGIn-model has the following four basic principles that can be applied to virtual teams:

1) Effectiveness—Some virtual groups cooperate only once, so in those cases vitality and continuity of the group as outcomes may not be that interesting for the members. In case of real (project) teams, however, it is not enough to come up with clear results and with rewards for their members. They also need to cater for group vitality and continuity in order to be effective. Virtual teams do not differ from conventional teams in this respect. However, developing vitality is more difficult than in co-located groups. The specific problems are related to the other principles of the model.

2) The quality of group processes—Six basic group processes were distinguished: communication, and the five other processes that can only exist on the basis of communication (cooperation, coordination, learning, reflection, and team building). These processes need to be aligned.

The type of *communication*—mediated to a smaller or larger extent—constitutes the core dimension for the concept of 'virtuality'. In case of virtual groups, the model implies that collaboration, coordination, knowledge exchange, social interaction, and reflection need to be adjusted to the degree of mediation of communication. This is reflected, among other things, in the fact that both remote *cooperation* and *social interaction* in mediated meetings need to be much more explicitly structured than face-to-face meetings in order to be effective. The already mentioned lack of non-verbal cues in communication, resulting in problems with 'awareness' of availability and state of others, makes it difficult to interact. Overall, face-to-face meetings allow for more flexibility during meetings and do not need to be as structured as mediated meetings. It is important to provide minutes of virtual meetings, as these help to assure that all members understand the same conclusions. In case of virtual student teams, Cramton (1997) showed that team members have difficulty in extracting information about the context in which their distant partners operate, while members themselves often fail to provide important information about their own context.

Globally distributed teams should give sufficient time and attention to group members who are less assertive than most members from a number of Western countries. *Leadership and coordination* of virtual teams therefore play a critical role in facilitating the work and organization of virtual teams (Bell & Kozlowski, 2002). In general the activities of virtual teams appear to need much more preparation and explicit coordination than co-located teams. The role of coordinators is, therefore, vital for the performance of virtual teams.

3) The quality and match of the 'context' characteristics—The quality of group processes depends on characteristics of the 'context'. Six groups of characteristics are distinguished: the task of the team, tools, member characteristics (knowledge, skills, attitudes), team structure (such as role division and meeting type), culture (norms, trust, cohesion, cognitive distance), and time-space setting (e.g., geographical distribution). The context characteristics need to match each other in order to optimally support the group processes.

ICT support. The technical tools and their usage should be adjusted to the virtuality of the group. The following suggestions can be made:

- Virtual groups require information storage and exchange tools.
- Virtual groups may benefit from a database with information on background and expertise of the group members (*'yellow pages'*).
- Virtual groups with intensive and non-routine interaction may benefit from tools for synchronous communication: chat features, where possible video links.
- Virtual groups with complex and time-sensitive tasks require workflow management tools for providing information regarding the progress of the project and activities of group members.
- The tools have to be easy to use and equally accessible to all members.
- Group members should be sufficiently trained in remote interaction and in using the technology.
- Global virtual teams should be careful in choosing and using the right tools. Research suggests that people from some cultures prefer direct expression of ideas, whereas others may be more sensitive to non-verbal cues and group relations (see Trompenaars, 1993, for examples). The first group of people would probably prefer a collaboration tool that enables synchronous communication, such as telephone, video, and chat. People from the other group would be happy with an asynchronous communication tool enabling them to express themselves more carefully. The choice for a suitable collaboration tool to facilitate both groups is, therefore, complicated.

Storage of information. Special attention should be given to information (document) exchange and storage. Effective virtual teams rely heavily on information exchange. Systems and procedures that allow for swift information exchange are therefore a prerequisite. Such systems need to be usable and accepted by all members of the team. In multicultural teams, such systems are not always easy to obtain. Differences in preferred practices of communication and storing information will limit the choice of an equally useable tool.

Cultural diversity may be large in virtual teams. In order to avoid conflicts and facilitate a smooth work process, group members should be trained to understand the potentially disturbing effect of diversity in national, organizational, and professional cultures (Dubé & Paré, 2004). The next step is to learn about each other's background so that differences in solving problems and ways of working will not form a source of major misunderstanding. As soon as members respect and trust distant team members, virtual teams will be able to benefit from the diversity of their members.

4) Development and adaptation: Team building—Groups develop and tools are adopted and adapted, through interpretation, interaction processes, and feedback. One of the processes through which this development and adaptation can be explicitly structured is team building. Team building proves to be a critical aspect of team performance and acts as the foundation for the development of necessary levels of trust, cohesion, and cognitive closeness among team members. In many cases, team building in virtual teams can benefit strongly from a face-to-face kick-off meeting (see Maznevski & Chudoba, 2001, for an overview). Coordinators should be alert to organize such meetings whenever needed or possible.

FUTURE TRENDS

As more and more organizations will explore the opportunities of working across boundaries, the number of virtual teams will increase in the coming years. These teams will experience the complexity of coordination of cross-border work activities. Future research in the area of virtual team analysis will highlight the relationship between virtuality and team effectiveness more closely. Especially, research with regard to the coordination and management of virtual teams will get more attention. So far, research into coordination of virtual teams has primarily focused on the role of leaders of virtual teams (e.g., Cascio & Shurygailo, 2003). Other possible ways of coordination in virtual contexts, such as the successful use of groupware tools and the role of substitutes for leadership, did not receive much attention yet, but will be trends for research in coming years.

CONCLUSION

Virtual teams offer great opportunities for collaboration across boundaries, which have encouraged many companies to form such teams. However, virtual teams also face challenges, particularly in the areas of communication and coordination. We have presented the DGIn-model for team analysis. On the basis of our analysis, we recommend that virtual teams should more explicitly pay attention to issues of team building, awareness, preparation, and information storage in order to work and collaborate effectively. Virtual teams should also benefit from the use of specialized groupware tools if applied properly.

REFERENCES

Andriessen, J.H.E. (2002). *Group work and groupware: Understanding and evaluating computer-supported interaction*. London: Springer-Verlag.

Andriessen, J.H.E. & Verburg, R.M. (2004). A model for the analysis of virtual teams. In A. Godar & S.P. Ferris (Eds.), *Virtual and collaborative teams: Process, technologies and practice*. Hershey, PA: Idea Group Publishing.

Bell, B.S. & Kozlowski, S.W.J. (2002). A typology of virtual teams: Implications for effective leadership. *Group & Organization Management, 27*(1), 14-49.

Cascio, W.F. & Shurygailo, S. (2003). E-leadership and virtual teams. *Organizational Dynamics, 31*(4), 362-376.

Cramton, C.D. (1997). Information problems in dispersed teams. *Academy of Management Best Paper Proceedings 1997* (pp. 298-302).

DeSanctis, G., Staudenmayer, N. & Wong, S.-S. (1999). Interdependence in virtual organizations. In C. Cooper & D. Rousseau (Eds.), *Trends in organizational behavior*. New York: John Wiley & Sons.

Dubé, L. & Paré, G. (2001). Global virtual teams. *Communications of the ACM, 44*(12), 71-73.

Dubé, L. & Paré, G. (2004). The multifaceted nature of virtual teams. In D.J. Pauleen (Ed.), *Virtual teams: Projects, protocols and processes* (pp. 1-40). Hershey, PA: Idea Group Publishing.

Guzzo, R.A. & Dickson, M.W. (1996). Teams in organizations: Recent research on performance effectiveness. *Annual Review of Psychology, 47,* 341-370.

Hackman, J.R. (1987). The design of work teams. In J.W. Lorsch (Ed.), *Handbook of organizational behavior* (pp. 315-342). Englewood Cliffs, NJ: Prentice-Hall.

Hutchinson, C. (1999). Virtual teams. In R. Stewart (Ed.), *Handbook of team working*. Aldershot, Hampshire, UK: Gower.

Jarvenpaa, S., Knoll, K. & Leidner, D. (1998). Is anybody out there? Antecedents of trust in global virtual teams. *Journal of Management Information Systems, 14*(4), 29-64.

Jarvenpaa, S.L. & Leidner, D.E. (1998). Communication and trust in global virtual teams**.** *Journal of Computer-Mediated Communication, 3*(4).

Kraemer, K.L. & Pinsonneault, A. (1990). Technology and groups: Assessment of the empirical research. In J. Galegher, R.E. Kraut & C. Egido (Eds.), *Intellectual teamwork: Social and technological foundations of cooperative work*. Hillsdale, NJ: Lawrence Erlbaum.

Majchrzak, A., Rice, R.E., Malhotra, A., King, N. & Ba, S. (2000). Technology adaptation: The case of a computer-supported inter-organizational virtual team. *MIS Quarterly, 24*(4), 569-600.

Maznevski, M.L. & Chudoba, K.M. (2001). Bridging space over time: Global virtual team dynamics and effectiveness. *Organization Science, 11*(5), 473-492.

McGrath, J.E. (1984). *Groups: Interaction and performance*. Englewood Cliffs, NJ: Prentice-Hall.

McGrath, J.E. & Hollingshead, A.B. (1994). *Groups interacting with technology: Ideas, evidence, issues and an agenda*. London: Sage Publications.

Meyerson, D., Weick, K.E. & Kramer, R.M. (1996). Swift trust and temporary groups. In R.M. Kramer & T.R. Tyler (Eds.), *Trust in organizations: Frontiers of theory and research* (pp. 166-195). Thousand Oaks, CA: Sage Publications.

Montoya-Weiss, M.M. (2001). Getting it together: Temporal coordination and conflict management in global virtual teams. *Academy of Management Journal*, *44*(6), 1251-1263.

Mowshowitz, A. (1997). Virtual organization. *Communications of the ACM*, *40*(9), 30-37.

Steinfield, C. (2002). Realizing the benefits of virtual teams. *IEEE Computer*, *35*(3), 104-106.

Townsend, A., DeMarie, S. & Hendrickson, A. (1998). Virtual teams: Technology and the workplace of the future. *Academy of Management Executive, 12*(3), 17-29.

Trompenaars, F. (1993). *Riding the waves of culture: Understanding cultural diversity in business*. London: Nicholas Brealey.

Vartiainen, M. (2003). *The functionality of virtual organizations.* Unpublished manuscript, Helsinki University of Technology, Finland.

KEY TERMS

Action Theory: Perspective on action facilitation that makes a distinction between acts, actions, and operations in performing a task. A basic principle of the theory is that the tools used should provide sufficient feedback to allow for adaptation of task execution.

Group Dynamics: Field of inquiry dedicated to advancing knowledge about the nature of groups.

Groupware: ICT applications that support communication, coordination, cooperation, learning, and/or social encounters through facilities such as information exchange, shared repositories, discussion forums, and messaging.

Dynamic Group Interaction (DGIn) Model: In this model elements of several theories with regard to group performance are brought together. Three levels of behavior are taken into account: individual goal-directed behavior, group processes, and a macro-social perspective. The various notions are brought together in a heuristic model concerning group processes. They are related to traditional input-process-output schemas.

Media Richness Theory: Theory on mediated communication that highlights the extent to which a medium is capable of sending rich information (i.e., text, smell, pictures, noise, etc.), as well as the proposition that media use is most adequate if the medium is matched with the complexity of the task at hand.

Structuration Theory: A theory of societal processes on a high abstraction level. Adaptive Structuration Theory (AST) focuses on the analysis of the way existing technologies are taken up by groups and evolve in their role during the appropriation process (i.e., the process of adaptation to new technical tools, which changes the original situation).

Team: A collection of individuals who see themselves and who are seen by others as a social entity, who are interdependent because of the tasks they perform as members of a group, who are embedded in one or more larger social systems (e.g., community, organization), and who perform tasks that affect others.

'Virtuality': The extent to which a group is geographically distributed, is organizationally and culturally diverse, has different timeframes for work, communicates electronically, and whose members are freelance or have fixed contracts with an organization.

A

Animated Characters within the MPEG-4 Standard

Françoise Prêteux
Institut National des Télécommunications/Groupe des Ecoles de Télécommunications, France

Marius Preda
Institut National des Télécommunications/Groupe des Ecoles de Télécommunications, France

INTRODUCTION

The first 3D virtual human model was designed and animated by means of the computer in the late ’70s. Since then, virtual character models have become more and more popular, making a growing population able to impact the everyday real world. Starting from simple and easy-to-control models used in commercial games, to more complex virtual assistants for commercial[1] or informational[2] Web sites, to the new stars of virtual cinema,[3] television,[4] and advertising,[5] the 3D character model industry is currently booming.

Moreover, the steady improvements within the distributed network area and advanced communication protocols have promoted the emergence of 3D communities[6] and immersion experiences (Thalmann, 2000) in distributed 3D virtual environments.

BACKGROUND

Animated Characters and 3D Standards

Creating, animating, and most of all, sharing virtual characters over Internet or mobile networks require unified data formats. If some animation industry leaders try—and sometimes succeed[7,8]—to impose their own formats in the computer world mainly by making available powerful authoring platforms, the alternative of an open standard is the only valid solution ensuring interoperability requirements, specifically when hardware products are to be built.

A dream of any content producer can be simply formulated as “creating once and reuse for ever and everywhere, in any circumstances.” Nowadays, content should be carried by heterogeneous networks (broadcast, IP,[9] mobile), available anywhere and for a large scale of devices (PCs, set-top boxes, PDAs,[10] mobile phones), and profiled with respect to user preferences. All these requirements make that the chain where content is processed more and more complex, and a lot of different actors must interfere: designers, service providers, network providers, device manufacturers, IPR[11] holders, end-users, and so on. For each one, consistent interfaces should be created on a stable and standardized basis.

Current work to provide 3D applications within a unified and interoperable framework is materialized by 3D graphics interchange standards such as VRML[12] ISO/IEC 14772-1:1997 and multimedia 2D/3D standards such as MPEG-4–ISO/IEC 14496. Each one addresses, more or less in a coordinated way, the virtual character animation issue. Moreover, some research groups proposed dedicated languages for modeling virtual faces and bodies, such as Face Markup Language (FML) and Virtual Human Markup Language (VHML). In the VRML community, the H-Anim[13] group released three versions of their specifications (1.0, 1.1 and 2001), as did the SNHC[14] sub-group of MPEG: MPEG-4 Version 1 supports face animation, MPEG-4 Version 2 supports body animation, and MPEG-4 Part 16 addresses the animation of generic and articulated virtual objects (including human-like). In MPEG-4, the specifications dealing with the definition and animation of human avatars are grouped under the name FBA—Face and Body Animation—while those referring to generic models are called BBA—Bone-Based Animation. The next section analyses the main similarities and differences of these two standardization frameworks: VRML and MPEG-4.

The VRML standard provides a textual description of 3D objects and scenes. It focuses on the spatial representation of such objects, while the time behavior is less supported. The major mechanism for supporting animation consists of defining it as an interpolation between key-frames.

The MPEG-4 standard, unlike the previous MPEG standards, does not only cope with highly efficient audio and video compression schemes, but also introduces the fundamental concept of media objects such as audio, visual, 2D/3D natural, and synthetic objects to make up a multimedia scene. As established in July 1994, the MPEG-4 objectives are focused on supporting new ways (notably content-based) of communicating, accessing, and manipulating digital audiovisual data (Pereira, 2002). Thus, temporal and/or spatial behavior can be associated with an object. The main functionalities proposed by the standard address the compression of each type of media objects, hybrid encoding of the natural and synthetic objects, universal content accessibility over various networks, and interactivity for the end-user. In order to specify the spatial and temporal localization of an object in the scene, MPEG-4 defines a dedicated language called BIFS—Binary Format for Scenes. BIFS inherits from VRML the representation of the scene, described as a hierarchical graph, and some dedicated tools such as animation procedures based on interpolators, events routed to the nodes, or sensor-based interactivity. In addition, BIFS introduces some new and advanced mechanisms such as compression schemes to encode the scene, streamed animations, integration of 2D objects, new 3D objects, and advanced time control.

In terms of functionalities related to virtual characters, both VRML and MPEG-4 standards define a set of nodes in the scene graph to allow for a representation of an avatar. However, only the MPEG-4 SNHC specifications deal with streamed avatar animations. A major difference is that an MPEG-4-compliant avatar can coexist in a hybrid environment, and its animation can be frame-based synchronized with other types of media objects, while the H-Anim avatar is defined in a VRML world and must be animated by VRML generic, usually non-compressed, animation tools.

The question that arises now is how to find a good compromise between the need for freedom in content creation and the need for interoperability? What exactly should be standardized, fixed, invariant, and in the mean time, ideally impose no constraints on the designer creativity? The long-term experience that the MPEG community has makes it possible to formulate a straight and solid resolution: in the complex chain of content, producing-transmitting-consuming the interoperability is ensured by only standardizing the data representation format at the decoder side. Pushing this concept to its extreme, an MPEG ideal tool is that one for which two requirements are satisfied: the designer can use any production tool he/she possesses to create the content, and it can be possible to build a full conversion/mapping tool between this content and an MPEG-compliant one. The same principle was followed when MPEG released the specifications concerning the definition and the animation of the virtual characters: there are no “limits” on the complexity of the virtual character with respect to its geometry, appearance, or skeleton and no constraints on the motion capabilities.

MAIN THRUST OF THE ARTICLE

The animation method of a synthetic object is strongly related to its representation model. A simple approach often used in cartoons is to consider the virtual character as a hierarchical collection of rigid geometric objects called segments, and to obtain the animation by transforming these objects with respect to their direct parents.

The second method consists of considering the geometry of the virtual character as a unique mesh and to animate it by continuously deforming its shape. While the former offers low animation complexity with the price of the seams at the joints between the segments, the latter ensures a higher realism of the representation, but requires more computation when synthesizing the animation. Both methods are supported by the MPEG-4 standard and will be detailed in the next sections.

Segmented Character: MPEG-4 FBA Framework

First efforts to standardize the animation of a human-like character (an avatar) within MPEG-4 were finalized at the beginning of 1999 and published as FBA. The approach used here consists of considering the avatar as a hierarchical structure of 3D rigid objects, each one corresponding to an anatomical segment (arm, forearm, etc.). There are no constraints on the geometry of a segment, but the hierarchical structure is predefined. The animation is performed by updating the geometric transformation between the segments. A separate, so-called FBA stream contains the compressed version of those transformations, expressed as extrinsic properties of an anatomical segment, that is, its 3D pose with respect to a reference frame attached to the parent segment.

The orientation of any anatomical segment is expressed as the composition of elementary rotations, namely twisting, abduction, and flexion. For some segments only one or two elementary rotations are supported. Hence, 296 angular joint values are enough to describe any 3D posture of a virtual human-like character. The angular values are specified with respect to the local 3D coordinate system of the anatomical segment. The origin of the local coordinate system is defined as the gravity center of the joint contour common to the considered anatomical segment and its parent. The rotation planes are specified, and/or anatomical segment rotation axes are standardized.

Due to the rigid nature of the anatomical segments, when performing the animation, seams can occur at joints between segments. This effect can be corrected by using local deformation tables that give hints on how to locally update the position of certain vertices on the anatomical segment with respect to the animation parameters (Capin, 1999; Preda, 1999, 2002).

To animate the face of the avatar, FBA uses a surface-based model driven by control points: a standardized number of key points (84) corresponding to the human features (e.g., middle point of upper lip) is defined on the face surface. Each control point can be locally translated on one, two, or three directions with respect to the face-specific animation parameters also defined in the FBA stream. It is possible to define a deformation behavior expressing how the translation of a key point is extended in its neighborhood. The complete animation is then performed by deforming the mesh in the vicinity of the key points (Doenges, 1997; Escher, 1998; Lavagetto, 1999).

Seamless Character: MPEG-4 BBA Framework

The purpose of this section is to introduce the animation framework for generic virtual objects as specified in Part 16 of the MPEG-4 standard. This framework is founded on a generic deformation model (Preda, 2002b), which relies on a deformation controller defined by means of a geometric support, an influence volume around this support, and a measure of affectedness within the influence volume. A best compromise between the efficient control of the geometrical support and appropriate volume specification is ensured by the 1D controller. For such a reason, MPEG-4 standardized two instances of the 1D deformation model, (1) bone controller and (2) muscle controller. The Skeleton, Muscle, and Skin (SMS) framework is built around these concepts—bone and muscle. An articulated character is now defined as a seamless mesh representing the skin, and a hierarchical structure of bones and muscles is attached to the skin. The geometric transformations of the bones and muscles lead to translations to the skin's vertex level. For one bone, the geometrical transformation is obtained by composing elementary transformations as rotations and, optionally, translations and scale factors. For one muscle, the main mechanism for deforming it consists of updating the curve-control points. Updating the degree of influence of the control points is also allowed. The efficient representation of the animation parameters is addressed by: (1) enriching the animation capabilities with temporal interpolation and inverse kinematics support; and (2) supporting the two data compression techniques, one using prediction, the second a DCT transform. The compressed representation of the animation parameters are grouped together in a so-called BBA stream, this stream being considered as an elementary MPEG-4 stream. It can be multiplexed with video, audio, and scene streams, allowing to build enriched multimedia applications.

In addition to the two controllers, bone and muscle, within the first amendment of AFX specifications, a morph space can be used to further control the animation quality. The tool standardized here consists of creating a base and some target shapes, and animating the shape as a result of weighting between target shapes. The origin of this technique is the 2D image-based face animation.

The main differences between the two animation frameworks developed by MPEG-4, FBA and BBA, consist of the geometry representation approach and skeleton structure: FBA builds the virtual character as a collection of 3D objects, and BBA considers the entire skin as a unique object; for FBA the skeleton structure is fixed (established by the standard), while for BBA the designer is free to build any kind of structure able to represent and animate a large category of objects (plants, animals, humans). Both approaches allow very low bit-rate animation (2-50kbps) by adopting compression techniques for the animation parameters.

FUTURE TRENDS

In recent years, major improvements have been reported in the field of virtual character animation, ranging from the creation of realistic models used in cinema movies, the development of animation production tools (e.g., motion capture systems), to the production of online animation in television shows and content streaming in distributed environments. However, research in this field is still in an initial stage and presents challenging issues. Despite large ongoing efforts, computer vision technologies for tracking human motion have not yet reached a level of maturity satisfactory for commercial use. In order to decrease the production cost of 3D content, retargeting motion from a motion capture data set to different avatars is still a hot topic of research. Another important research activity is oriented towards behavioral models for avatar or/and crowds, building autonomous agents able to own intelligence, making virtual characters "live" with the appropriate emotional content, and so forth. As a content representation standard, MPEG-4 did not explicitly analyze these issues closely, but provides a low-level framework able to represent/compress/transmit and playback the content. This approach, together with the extensibility offered by an open standard, allows content creators and researchers to imagine up-to-date and high-level tools and algorithms to provide content, map the results into a low-level representation, and then use the MPEG-4 chain for transmission and playback of this content.

CONCLUSION

This article is devoted to the standardization of virtual character animation. In particular, the MPEG-4 Face and Body, as well as the Skeleton, Muscle, and Skin animation frameworks have been presented. A generic deformation model and its implementation in the MPEG-4 standard through the bone and muscle controllers have been introduced.

REFERENCES

Capin, T.K. & Thalmann, D. (1999). Controlling and efficient coding of MPEG-4-compliant avatars. *Proceedings of IWSNHC3DI'99*, Santorini, Greece.

Doenges, P., Capin, T., Lavagetto, F., Ostermann, J., Pandzic, I. & Petajan, E. (1997). MPEG-4: Audio/video and synthetic graphics/audio for real-time, interactive media delivery. *Image Communications Journal*, *5*(4), 433-463.

Escher, M., Pandzic, I. & Magnenat-Thalmann, N. (1998). Facial animation and deformation for MPEG-4. *Proceedings of Computer Animation'98*.

Lavagetto, F. & Pockaj, R. (1999). The facial animation engine: Toward a high-level interface for the design of MPEG-4-compliant animated faces. *IEEE Transactions on Circuits System Video Technology*, *9*(2), 277-289.

Preda, M. (2002b, December). *Advanced virtual character animation within the MPEG-4 framework*. PhD Dissertation, Université Paris V-René Descartes, Paris.

Preda, M. & Prêteux, F. (2002a). Insights into low-level animation and MPEG-4 standardization. *Signal Processing: Image Communication*, *17*(9), 717-741.

Preda, M., Prêteux, F. & Zaharia, T. (1999). 3D body animation and coding within a MPEG-4-compliant framework. *Proceedings of IWSNHC3DI'99*, Santorini, Greece.

KEY TERMS

Avatars: Synthetic representation of a human body able to be animated. Avatars are often used in games for representing players or in virtual environments when the presence of the user must be visible.

BBA (Bone-Based Animation): A part of MPEG-4 specifications dealing with the definition and the animation at very low bit rate of a generic articulated model based on a seamless representation of the skin and a hierarchical structure of bones and muscles.

FBA (Face and Body Animation): A part of the MPEG-4 specifications dealing with the definition and the animation at very low bit rate of an avatar represented as a segmented object.

H-Anim (Humanoid Animation): Part of VRML specifications consisting of node prototypes allowing the definition of an avatar.

MPEG (Motion Picture Expert Group): Marketing name of the "ISOIEC SC29 WG11" standardization committee, affiliated with the ISO (International Standardization Office) and creator of the multimedia standards: MPEG-1, MPEG-2, MPEG-4, MPEG-7, and MPEG-21.

SMS (Skeleton, Muscle, and Skin): An animation framework based on BBA.

SNHC (Synthetic and Natural Hybrid Coding): An MPEG working group dealing with specifications of 3D graphics tools, and integration of synthetic and natural media in hybrid scenes.

Virtual Character: Synthetic representation of an entity (humans, animals, plants) able to be animated.

VRML (Virtual Reality Markup Language): Officially called ISOIEC 14772, this standard is the first attempt to provide a unified representation for 3D objects and 3D scenes.

ENDNOTES

1. Living actor technology, *www.living-actor.com*.
2. Scotland government Web page, *www.scotland.gov.uk/pages/news/junior/introducing_seonaid.aspx*.
3. Walt Disney Pictures & Pixar. "Geri's Game", "Toy Story" (1995), "A Bug's Life" (1998), "Toy Story 2" (1999), and "Monsters, Inc." (2001).
4. Vandrea news presenter, Channel 5, British Broadcasting Television.
5. Eve Solal, Attitude Studio, *www.evesolal.com*.
6. blaxxun Community, VRML-3D-Avatars-Multi-User Interaction, *www.blaxxun.com/vrml/home/ccpro.htm*.
7. 3D Studio Max™, Discreet, *www.discreet.com/index-nf.html*.
8. Maya™, Alias/Wavefront, *www.aliaswavefront.com/en/news/home.shtml*.
9. Internet Protocol.
10. Personal Digital Assistant.
11. Intellectual Property Rights.
12. The Virtual Reality Modeling Language, International Standard ISO/IEC 14772-1:1997, *www.vrml.org*.
13. H-Anim—Humanoid AnimationWorking Group, *www.h-anim.org*.
14. SNHC—Synthetic and Natural Hybrid Coding, *www.sait.samsung.co.kr/snhc*.

Antecedents of Trust in Online Communities

A

Catherine M. Ridings
Lehigh University, USA

David Gefen
Drexel University, USA

INTRODUCTION

Online virtual communities have existed on the Internet since the early 1980s as Usenet newsgroups. With the advent of the World Wide Web and emphasis on Web site interactivity, these communities and accompanying research have grown rapidly (Horrigan, Rainie, & Fox, 2001; Lee, Vogel, & Limayem, 2003; Petersen, 1999). Virtual communities arise as a natural consequence of people coming together to discuss a common hobby, medical affliction, or other similar interest, such as coin collecting, a devotion to a rock group, or living with a disease such as lupus. Virtual communities can be defined as groups of people with common interests and practices that communicate regularly and for some duration in an organized way over the Internet through a common location or site (Ridings, Gefen, & Arinze, 2002). The location is the "place" where the community meets, and it can be supported technologically by e-mail listservs, newsgroups, bulletin boards, or chat rooms, for example. The technology helps to organize the community's conversation, which is the essence of the community. For example, messages in a community supported by a listserv are organized in e-mails, sometimes even grouping together several messages into an e-mail digest. In bulletin board communities, the conversation is organized into message threads consisting of questions or comments posted by members and associated replies to the messages.

Virtual community members form personal relationships with strong norms and expectations (Sproull & Faraj, 1997; Sproull & Kiesler, 1991), sometimes developing deep attachments to the communities (Hiltz, 1984; Hiltz & Wellman, 1997). These developments are interesting, because the members of virtual communities are typically strangers to one another and may never meet face to face. Additionally, the nature of computer-mediated communication is such that nonverbal cues that aid in the interpretation of communication, such as inflections in the voice, gestures, dress, tone, physical personal attributes, and posture, are missing (Sproull & Kiesler, 1991), making the communication open to multiple interpretations (Korenman & Wyatt, 1996). Yet, despite these limitations, many virtual communities flourish by exchanging messages and building their conversation base. A key ingredient in sustaining the conversation in the community is the existence of trust between the members. Trust has a downstream effect on the members' intentions to give and get information through the virtual community (Ridings et al., 2002).

This chapter examines emergent virtual communities, that is, those arising without direction or mandate from an organization, government, or other entity for an expressed economic or academic purpose. For example, a discussion board for a strategic partnership work group between two companies or a chat room for a class taking a college course would not be considered emergent virtual communities. However, an online forum established by the Breast Cancer Young Survivors Coalition so that women could discuss their battles with the disease would be considered an emergent virtual community.

BACKGROUND

Trust is an essential ingredient in social relationships (Blau, 1964; Luhmann, 1979), and understanding and defining trust are dependent upon the situation in which they are considered. In communities, in general, trust is an integral part of interpersonal relations among members and defines an individual's expectations and behavior (Luhmann, 1979; Rotter, 1971). Trust has many definitions. It has been defined as a willingness to take a risk associated with the behavior of others (Mayer, Davis, & Schoorman, 1995) and, more generally, as a method of reducing social uncertainty (Gefen, Karahanna, & Straub, 2003; Luhmann, 1979). In this sense, trust is used in the virtual community to reduce social complexity associated with the behavior of other members, and as a way of reducing the fear that the trusted party will take advantage by engaging in opportunistic behavior (Gefen et al., 2003), much as it does in communities in general (Fukuyama, 1995).

Participating in a virtual community entails exposure to risk. Opportunistic behaviors could include selling personal information that was confidentially provided, adopting a fictitious persona, deliberately and stealthily marketing products and services when this is prohibited, flaming or spamming, making unfair practical jokes at members, providing false information, and, in general, behaving in a dysfunctional manner that ruins the community. Such behavior also applies to other types of communities, except that in the case of an online community, the anonymity provided by the Internet makes such behavior much easier to accomplish by the perpetrator and much harder to notice by the victim.

Scholarly research on trust has shown that trust is a multidimensional concept consisting of beliefs in ability, benevolence, and integrity (Blau, 1964; Butler, 1991; Giffin, 1967; Mayer et al., 1995; McKnight, Choudhury, & Kacmar, 2002). Ability deals with beliefs about the skills or expertise that another (i.e., trusted parties) has in a certain area. Ability relates to the belief that the other person knows what he or she is talking about. Because virtual communities are almost always focused on a specific topic, concerns about the abilities of others with respect to this topic are important. Benevolence is the expectation that others will have a positive orientation or a desire to do good to the trustee, typically by reciprocating with appropriate advice, help, discussion, and so on, such as contributing to the ongoing discussion with the intent to help, support, and care for others. Benevolence is important in virtual communities, because without positive reciprocation, the community would not exist. Integrity is the expectation that another will act in accordance with socially accepted standards of honesty or a set of principles, such as not telling a lie and providing reasonably verified information. Integrity applies in the virtual community context, because it is the existence of norms of reciprocity, closely linked with benevolence, that allow the community to properly function.

Research based upon surveying members of virtual communities has found that integrity and benevolence are united in this context, because the expected mode of behavior in many of the virtual communities is one of benevolence (Ridings et al., 2002). Hence, adhering to this expected mode of conduct, integrity, should overlap with actually behaving so, namely, with benevolence. Conformance to socially acceptable behavior or standards (integrity) and a desire to do “good” to others (benevolent intentions) are synonymous in the virtual community environment.

THE ANTECENDENTS OF TRUST

Trust in a virtual community is built through several mechanisms that are germane to the online context. As in personal contacts where successful interpersonal interaction builds trust (Blau, 1964; Gefen, 2000a; Luhmann, 1979), the responsiveness of other community members is necessary for trust to develop (Ridings et al., 2002). This can be shown through adherence to the social norms of the community (benevolence and integrity) and competency in the topic (ability). Members who post messages most often expect responses, and when these responses are absent, late, or lacking in number, there is no successful interpersonal interaction, and that hinders the development of trust. Responsiveness is also evident by members indicating gratitude for timely help. Trust is also built by reading what others post. If others post personal information about themselves, they appear less as strangers and more as acquaintances or friends. Divulging gender, age, name, e-mail address, or a personal problem may also add to the credibility of the member (ability) as well as make it easier for other members to shape beliefs regarding adherence to the community’s standards and principles (integrity and benevolence). Personal information can also be provided in site profiles. Thus, the confiding of personal information also builds trust in other members of a virtual community (Ridings et al., 2002). Finally, humans have some degree of a general willingness to depend on others, known as disposition to trust (McKnight, Cummings, & Chervany, 1998), and this has been found to be stable across situations (Mayer et al., 1995). In the virtual community where people are unfamiliar with one another, disposition to trust, at least initially before extensive interactions take place, is also an important factor leading to the development of trust in others. Disposition to trust has been empirically found to be directly related to trust in virtual settings (Gefen, 2000a; Jarvenpaa, Knoll, & Leidner, 1998) and in virtual communities, in particular (Ridings et al., 2002).

Because virtual communities lack an enforceable legal system to ensure appropriate behavior online, the actual membership in the community and the feeling of being part of a community, even if a virtual one, may provide a possible way to enforce honest behavior. Virtual communities enhance honest behavior through creating what Ba (Ba, 2001; Ba, Whinston, & Zhang, 2003) called a trusted third party (TTP) certification mechanism. Considering the problems with the three current trust-building mechanisms in online markets (feedback, insurance or guarantee, and escrow), as pointed out theoretically by Ba and with some empirical support by Pavlou and Gefen (2004), extralegal mechanisms might be especially useful in virtual communities. Extralegal mechanisms, such as gossip, reproach, and community appreciation, and the praise and sanctions they bring, may serve to create trust just as they do in regular community settings.

Another way virtual communities may be applied to build trust, according to Ba, is through the sense of community, that, as we know from economics, is crucial when there is a separation in time between the quid and the pro (Ba, 2001). Moreover, if the members of the community are held responsible for the actions of an offender, there will be more social pressure to adhere to the rules. This might only work with online groups with a strong sense of community, but many virtual communities are precisely that.

Trust also has implications with regard to user privacy. Many virtual communities center on personal topics, such as medical conditions, legal issues, or occupations. Participants may care to be anonymous when communicating in such communities. However, the economic viability of virtual communities may depend on the sale of advertising space or products to users. To accommodate this action, reliance is placed on the provision of user demographics, e-mail addresses, and traffic statistics, information somewhat at odds with the protection of user privacy. It may be possible for virtual communities to incorporate concepts of procedural fairness, where the provision and collection of personal information is perceived as conducted fairly. Procedural fairness has been found to address privacy concerns of customers. Culnan and Armstrong found that procedural fairness builds trust that customers have for an organization using personal information for marketing purposes (Culnan & Armstrong, 1999). Such procedural fairness policies could also be applied in the virtual community context.

FUTURE TRENDS

There are many directions to be investigated from the basic understanding of trust in virtual communities. It may be that trust develops differently in different kinds of communities. For example, trust in medical-based communities may be based more heavily on certain attributes than trust in communities organized for fans of a particular sports team. The level of trust for an individual may be related to the use of the community. Demographic variables such as gender, race, and culture may also influence trust and its development (Gefen, 2000b). Longitudinal studies of virtual communities may yield more information about the development of trust over time.

CONCLUSION

Virtual communities are a key resource on the Internet for individuals looking to exchange information with others as well as Web site sponsors desiring to provide interactivity to their sites. For virtual communities to survive, they must have conversation, and for the conversation to grow and flourish, there must exist trust between the virtual community members. Trust in the other members' abilities and benevolence and integrity is necessary, and this trust has been found to be built by the responsiveness of others, the confiding of personal information, and the member's general disposition to trust.

REFERENCES

Ba, S. (2001). Establishing online trust through a community responsibility system. *Decision Support Systems, 31*, 323–336.

Ba, S., Whinston, A. B., & Zhang, H. (2003). Building trust in online auction markets through an economic incentive mechanism. *Decision Support Systems, 35*, 273–286.

Blau, P. M. (1964). *Exchange and power in social life*. New York: John Wiley & Sons.

Butler, J. K. (1991). Toward understanding and measuring conditions of trust: Evolution of a condition of trust inventory. *Journal of Management, 17*(3), 643–663.

Culnan, M. J., & Armstrong, P. K. (1999). Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation. *Organization Science, 10*(1), 104–115.

Fukuyama, F. (1995). *Trust: The social virtues & the creation of prosperity*. New York: The Free Press.

Gefen, D. (2000a). E-commerce: The role of familiarity and trust. *Omega, 28*(6), 725–737.

Gefen, D. (2000b). Gender differences in the perception and adoption of e-mail and computer-mediated communication media: A sociolinguistics approach. In A. Kent (Ed.), *The encyclopedia of library and information science*. New York: Marcel Dekker.

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly, 27*(1), 51–90.

Giffin, K. (1967). The contribution of studies of source credibility to a theory of interpersonal trust in the communication process. *Psychological Bulletin, 68*(2), 104–120.

Hiltz, S. R. (1984). *Online communities: A case study of the office of the future*. Norwood, NJ: Ablex Publishing Corporation.

Hiltz, S. R., & Wellman, B. (1997). Asynchronous learning networks as a virtual classroom. *Communications of the ACM, 40*(9), 44–49.

Horrigan, J. B., Rainie, L., & Fox, S. (2001). Online communities: Networks that nurture long-distance relationships and local ties. Retrieved from *http://www.pewinternet.org/reports/toc.asp?Report=47*

Jarvenpaa, S. L., Knoll, K., & Leidner, D. E. (1998). Is anybody out there? Antecedents of trust in global virtual teams. *Journal of Management Information Systems, 14*(4), 29–64.

Korenman, J., & Wyatt, N. (1996). Group dynamics in an e-mail forum. In S. C. Herring (Ed.), *Computer-mediated communication: Linguistic, social and cross-cultural perspectives* (pp. 225–242). Philadelphia: John Benjamins.

Lee, F. S. L., Vogel, D., & Limayem, M. (2003). Virtual community informatics: A review and research agenda. *Journal of Information Technology Theory and Application, 5*(1), 47–61.

Luhmann, N. (1979). *Trust and power* (H. Davis, J. Raffan, & K. Rooney, Trans.). UK: John Wiley and Sons.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20*(3), 709–734.

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research, 13*(3), 334–359.

McKnight, D. H., Cummings, L. L., & Chervany, N. L. (1998). Initial trust formation in new organizational relationships. *Academy of Management Review, 23*(3), 473–490.

Pavlou, P. A., & Gefen, D. (2004). Building effective online marketplaces with institution-based trust. *Information Systems Research*, *15*(1), 37–59.

Petersen, A. (1999, January 6). Some places to go when you want to feel right at home: Communities focus on people who need people. *The Wall Street Journal,* p. B6.

Ridings, C., Gefen, D., & Arinze, B. (2002). Some antecedents and effects of trust in virtual communities. *Journal of Strategic Information Systems, 11*(3–4), 271–295.

Rotter, J. B. (1971). Generalized expectancies for interpersonal trust. *American Psychologist, 26*, 443–450.

Sproull, L., & Faraj, S. (1997). Atheism, sex and databases: The Net as a social technology. In S. Kiesler (Ed.), *Culture of the Internet* (pp. 35–51). Mahwah, NJ: Lawrence Erlbaum Associates.

Sproull, L., & Kiesler, S. (1991). *Connections: New ways of working in the networked organization*. Cambridge, MA: The MIT Press.

KEY TERMS

Disposition to Trust: A tendency to believe in the goodness of others based on lifelong socialization.

Reciprocity: Returning favors, which is a major way of building trust.

Trust: A willingness to take for granted that another person will behave as expected in a socially constructive manner. Trust generally reduces the perceived risk that another person will behave in an opportunistic manner.

Trust in Ability: The belief that a person has subject matter expertise in a certain area.

Trust in Benevolence: The belief that a person has a positive orientation or a desire to do good to others.

Trust in Integrity: The belief that a person will act in accordance with socially accepted standards of honesty or a set of principles.

Virtual Community: A group of people with common interests and practices that communicates regularly and for some duration in an organized way over the Internet through a common location or site.

"Anytime, Anywhere" in the Context of Mobile Work

Mikael Wiberg
Umea University, Sweden

INTRODUCTION

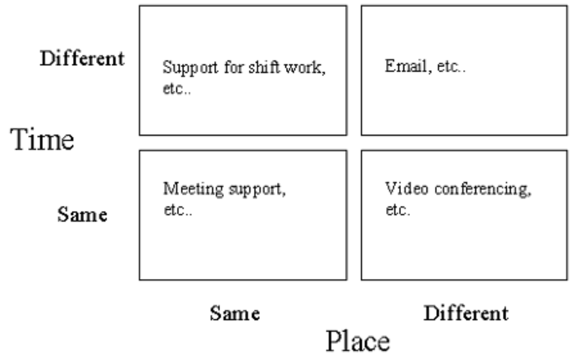
Kleinrock (1996, 1998) claims that advanced wireless technologies, the Internet, Global Positioning Systems, portable and distributed computing, and so forth will realize the vision of "anytime, anywhere" computing. We can today see the first signs of this vision. For example, telework is now possible, remote organizations can be engaged in close cooperation, and people can form communities on the Internet. The world has become a "global village," some claim (Castells, 1996; Preece, 1994), where you can interact with anybody independent of time and space.

The vision of anytime and anywhere describes a situation where people can do tasks wherever they want and without any consideration for time. Related to the vision is the 2x2 matrix often used in the field of CSCW (computer-supported cooperative work) to denote different kinds of computer-supported collaboration (e.g., Baecker et al., 1993; Johansen, 1988). This model has the dimensions of time and place, where each can be same or different. The model is shown in Figure 1 below.

The vision of anytime and anywhere is tasks that can be done independent of time and place, that is, in any of the four scenarios. This does not say anything about where or when the tasks should be done, only that these dimensions should not restrict them.

It is interesting to notice that the model does not take into consideration *mobility*. It assumes that people are either in the same place or in a different place, and whether or not they are mobile does not seem to make a difference.

Figure 1. Model showing different scenarios for groupware (Ellis et al., 1991)



BACKGROUND

In the past, people traveled because they had no choice. If you wanted to do business or talk to remote friends, you had to meet them face to face. However, transportation costs prohibited certain meetings and activities. A long series of technological developments (including the pony express, railroads, automobiles, and the telephone) have aimed at lowering the costs associated with transaction and conversation. Computer-mediated communications are the most recent development in that progression. Even so, people still travel and still meet in person.

To summarize, the adoption of Internet technologies, mobile phones, and so forth have increased and, in a sense, made the world smaller. Compared to 10 years ago, it is today much easier to communicate with remote sites, and the frequency of communication in many organizations has increased accordingly. Some people have even talked about the global village (Preece, 1994). A parallel trend is that people travel more than they used to do. According to predictions, this trend will sustain and even increase. For example, the national road agency of Sweden reports the number of flights will increase by a factor of four in the next 10 years. How can it be that the global village is so mobile? If people can interact and work independent of time and space, why then do they spend more and more time traveling? Is that not a paradox?

Reviewing the literature on the topic, we find no research that has explored this apparent paradox. Authors are either concerned with remote interaction (e.g., Brave, Ishii, & Dahley, 1998; Ellis et al., 1991; Kuzuoka, 1992; McDaniel, 1996; Tang & Minneman, 1991), mobility (e.g., Bejerano & Cidon, 1998; Luff & Heath, 1998; Porta, Sabnani, & Gitlin, 1996), or mobility as anytime, anywhere work (e.g., Dix, Rodden, Davies, Trevor, Friday, & Palfreyman, 2000; Perry, O'hara, Sellen, Brown, & Harper, 2001). Furthermore, research on mobility has mainly dealt with technology issues, for example, limited battery life, unreliable network connections, varying channel coding and characteristics, volatile access points, risk of data loss, portability, and location discovery (e.g., Bhagwat, Satish, & Tripathi, 1994; Dearle, 1998; Francis, 1997; Varshney, 1999). Accordingly, no research so far has explored the relation between, on one hand, the global village, with its idea that distance plays no role, and on the other hand, the trend of increased mobility. How do the two trends hang together?

EXPLORING THE "ANYTIME, ANYWHERE" MOBILITY PARADOX

In order to investigate this seeming paradox, we conducted an empirical study of mobile telecommunication engineers in a Swedish company (Wiberg & Ljungberg, 2000). Using qualitative research methods, we studied to what extent the work tasks they do are dependent on time and place. We analyzed the data using a 2x2 matrix, with the two axes, "time" and "space," which both have the categories "dependent" and "independent." One of the four situations is "anytime, anywhere," while the other three are dependent on time, place, or both (see Figure 2).

We found instances of work in all four categories. Some traveling seems very difficult to escape, simply because there are places that staff need to visit physically to do their job. For example, to repair a telephone pole, you need to go to it. We also found there are time frames that staff cannot escape. For example, rebooting parts of the telephone network has to be done at night. Lastly, there are work tasks that seem pretty much independent of time and space, for example, scheduling and rescheduling of activities.

As observed during this empirical study, there were just tiny parts of service work possible to perform anytime and anywhere. Most of the work is dependent on spatial factors such as the location of a breakdown in the telephone network system, the location of the client, and so forth, or has time-related dependencies such as fixing problems within 24 hours or coordinating schedules to cooperate around larger problems. For a more thorough description of the empirical material, see Wiberg and Ljungberg (2000). Overall, we found there are the following:

- traveling that seems difficult to remove because of places that people have to visit physically, for example, telephone poles, customers' houses since not all customers are mobile, network routers, locations where new cables need to be drawn, and so forth,
- time frames that seem very difficult for staff to not do certain tasks within, for example, customer service within 24 hours, rebooting parts of the telephone network has to be done at night, and so forth, and
- tasks that do not seem to be restricted by time and place, for example, scheduling and rescheduling of the activities over the day, coordination of activities between the technicians, experience and knowledge sharing among the technicians, and so forth. These tasks, however, are important for them since they are alone in their cars most of the day.

Accordingly, the vision of anytime and anywhere is not easy to realize in the case of the mobile workers we studied.

FUTURE TRENDS

Both work and leisure activities are becoming increasingly mobile. To describe the mobile worker, new concepts have been coined. Some examples are "road warriors" and "nomads" (Dahlbom, 1998) that distinguish mobile workers as moving from terms like distributed work, telework, and colocated work. One reason for this

Figure 2. Theoretical framework of the study

		Place Independent	**Place** Dependent
Time	**Independent**	**1. Anytime, anywhere**: tasks that can be done independent of time and place; they can be done anytime, anywhere	**2. Anytime, particular place**: tasks that need to be done in a particular place but can be done anytime
	Dependent	**3. Particular time, any place**: Tasks that can be done independent of place but at a certain time or in a certain order	**4. Particular time, particular place**: Tasks that must be done in a particular place within a particular time

increased mobility is the emergence of service work as the dominating profession in the postindustrial society. Service work very often takes place at the client site, and therefore it is often mobile. Another reason is the increased importance of cooperation in and between organizations. Some cooperation can take place remotely, but people also need to meet physically. A third important reason for increased mobility is the extensive adoption of mobile phones. Mobile phones enable people to be mobile and yet accessible. As people have become accessible independent of place, new ways of working have emerged in many organizations. So, for future development within this prominent area of mobile IT and mobility, it is important to keep in mind this “anytime, anywhere” paradox of the mobility vision.

CONCLUSION

This paper has shown some limitations to the vision of “anytime, anywhere” in the context of mobile work. Time and place are indeed very old ways for understanding context, and it seams like they are useful even for bringing light on the phenomena of the two parallel trends of the global village and mobility.

The paper has shown that work has moments, for example, time frames, which are not negotiable, so the work is dependent upon those. The paper has also shown that work has places of nonnegotiable importance; for example, you cannot reframe the earth by putting away distance nor go backwards in time, although computers are often described as being able to bridge those gaps in time and space. As seen above, there is not much service work possible to perform anytime and anywhere since service work is not only about moving around (i.e., mobility), but also about taking actions at various sites at specific times.

Kleinrock (1998) has argued the vision of mobility as being able to work anytime, anywhere. However, from the analysis of the empirical study presented above, we argue that there are several limits to that vision. In fact, this paper has argued that the concept of anytime and anywhere belongs to another trend, that is, the trend toward a global village, which is something altogether different from the trend of mobility. However, as the analysis has shown above, those two trends come together in practice.

So, finally, we conclude this paper arguing that the practical limitations of “anytime, anywhere” make it impossible for mobile service engineers to conduct work anytime, anywhere.

REFERENCES

Baecker, R. M. (Ed.). (1993). *Readings in groupware and computer supported cooperative work: Assisting human to human collaboration.* San Mateo: Morgan Kaufmann Publisher Inc.

Bejerano, Y., & Cidon, I. (1998). An efficient mobility management strategy for personal communication systems. *The Fourth Annual ACM/IEEE International Conference on Mobile Computing and Networking.*

Bhagwat, P., & Tripathi, S. K. (1994). Mobile computing. *Proceedings of Networks ’94*, 3-12.

Brave, S., Ishii, H., & Dahley, A. (1998). Tangible interfaces for remote collaboration and communication. *Proceedings of the ACM 1998 Conference on Computer Supported Cooperative Work.*

Castells, M. (1996). The information age: Economy, society and culture. In *The rise of the network society.* Oxford, UK: Blackwell Publishers Ltd.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology, 94*, 295-120.

Dahlbom. (1998). From infrastructure to networking. In N. J. Buch et al. (Eds.), *Proceedings of IRIS 21*. Department of Computer Science, Aalborg University.

Dearle, A. (1998). Towards ubiquitous environments for mobile users. *IEEE Internet Computing, 2*(1), 22-32.

Dix & Beale. (1996). *Remote cooperation: CSCW issues for mobile and teleworkers.* New York: Springer.

Dix, A., Rodden, T., Davies, N., Trevor, J., Friday, A., & Palfreyman, K. (2000). Exploiting space and location as a design framework for interactive mobile systems. *ACM Transactions on Computer-Human Interaction (TOCHI), 7*(3).

Ellis et al. (1991). Groupware: Some issues and experiences. *Communications of the ACM, 34*(1), 39-58.

Francis, L. (1997). Mobile computing: A fact in your future. *Proceedings of SIGDOC ’97*, 63-67.

Hammersley & Atkinson. (1995). *Ethnography: Principles in practice.* London: Routledge.

Hughes, J., Randall, D., & Shapiro, D. (1993). From ethnographic record to system design: Some experiences from the field. *Computer Supported Cooperative Work, An International Journal, 1*(3), 123-141.

Johansen, R. (1988). *Groupware: Computer support for business teams.* New York: The Free Press.

Kleinrock, L. (1996). Nomadicity: Anytime, anywhere in a disconnected world. *Mobile Networks and Applications, 1*(4), 351-357.

Kleinrock, L. (1998). Nomadic computing: Information network and data communication. *IFIP/ICCC International Conference on Information Network and Data Communication*, 223-233.

Kuzuoka, H., et al. (1994). GestureCam: A video communication system for sympathetic remote collaboration. *Proceedings of the Conference on Computer Supported Cooperative Work.*

Lindgren, R., & Wiberg, M. (2000). Knowledge management and mobility in a semi-virtual organization: Lessons learned from the case of Telia Nära. *Proceedings of Hicss33.*

Luff, P., & Heath, C. (1998). Mobility in collaboration. *Proceedings of the ACM Conference on Computer Supported Cooperative Work.*

Mason, R. O. (1989). MIS experiments: A pragmatic perspective. In I. Benbasat (Ed.), *The information systems research challenge: Experimental research methods* (Vol. 2, pp. 3-20). Boston: Harvard Business School Press.

McDaniel, S. (1996). Providing awareness information to support transitions in remote computer-mediated collaboration. *Proceedings of the CHI '96 Conference Companion on Human Factors in Computing Systems: Common Ground*.

Perry, M., O'hara, K., Sellen, A., Brown, B., & Harper, R. (2001, December). Dealing with mobility: Understanding access anytime, anywhere. *ACM Transactions on Computer-Human Interaction (TOCHI), 8*(4).

Porta, T., Sabnani, K., & Gitlin, R. (1996). Challenges for nomadic computing, mobility management and wireless communications. *Mobile Networking Applications*.

Preece, J. (Ed.). (1994). *Human-computer interaction.* New York: Addison & Wesley.

Tang, J., & Minneman, S. (1991). VideoWhiteboard video shadows to support remote collaboration. *Human Factors in Computing Systems: Conference Proceedings on Reaching through Technology*.

Varshney, U. (1999). Networking support for mobile computing. *Communications of the Association for Information Systems, 1*.

Wiberg, M., & Ljungberg, F. (2000). Exploring the vision of anytime, anywhere in the context of mobile work. In *Knowledge management and virtual organizations: Theories, practices, technologies and methods.* The Biztech Network, Brint Press.

KEY TERMS

"Anytime, Anywhere" Work: Describes a situation where people can do tasks wherever they want and without any consideration for time; that is, their work can be done anytime, anywhere.

Colocated Work: Collaborative work carried out by several persons at the same geographical location.

Distributed Work: Collaborative work carried out by several persons at different geographical locations.

Global Village: As computers all over the world become interconnected via the Internet, and the frequency of communication in and between organizations, countries, cultures, societies, and so forth has increased accordingly via these networks, we can now on a daily basis and quite easily maintain contact with anybody independent of time and space; that is, we are able to interact anytime, anywhere.

Mobile Work: The ability to carry out work while geographically moving around.

"Particular Time, Particular Place" Work: Tasks that must be done in a particular place within a particular time.

Remote Interaction: Information-technology-mediated human-to-human communication over a distance.

Telework: Ability to carry out work from a distance, for example, sitting at home and doing office work. Telework does not imply that the worker is mobile (i.e., in motion) in any sense even though the concept of telework is sometimes used as a synonym for mobile work.

Application of Fuzzy Logic to Fraud Detection

A

Mary Jane Lenard
University of North Carolina – Greensboro, USA

Pervaiz Alam
Kent State University, USA

INTRODUCTION

In light of recent reporting of the failures of some of the major publicly-held companies in the U.S. (e.g., Enron & WorldCom), it has become increasingly important that management, auditors, analysts, and regulators be able to assess and identify fraudulent financial reporting. The Enron and WorldCom failures illustrate that financial reporting fraud could have disastrous consequences both for stockholders and employees. These recent failures have not only adversely affected the U.S. accounting profession but have also raised serious questions about the credibility of financial statements. KPMG (2003) reports seven broad categories of fraud experienced by U.S. businesses and governments: employee fraud (60%), consumer fraud (32%), third-party fraud (25%), computer crime (18%), misconduct (15%), medical/insurance fraud (12%), and financial reporting fraud (7%). Even though it occurred with least frequency, the average cost of financial reporting fraud was the highest, at $257 million, followed by the cost of medical/insurance fraud (average cost of $33.7 million).

Statistical methods, expert reasoning, and data mining may be used to achieve the objective of identifying financial reporting fraud. One way that a company can justify its financial health is by developing a database of financial and non-financial variables to evaluate the risk of fraud. These variables may help determine if the company has reached a stress level susceptible to fraud, or the variables may identify fraud indicators. There are a number of methods of analysis that may be used in fraud determination. Fuzzy logic is one method of analyzing financial and non-financial statement data. When applied to fraud detection, a fuzzy logic program clusters the information into various fraud risk categories. The clusters identify variables that are used as input in a statistical model. Expert reasoning is then applied to interpret the responses to questions about financial and non-financial conditions that may indicate fraud. The responses provide information for variables that can be developed continuously over the life of the company. This article summarizes the specifics of fraud detection modeling and presents the features and critical issues of fuzzy logic when applied for that purpose.

BACKGROUND

Fraud Detection

The problem of fraudulent financial reporting is not limited to the U.S. In 2002, the Dutch retailer, Ahold, disclosed losses of $500 million related to accounting at its U.S. subsidiary (Arnold, 2003). Recently, Parmalat, an Italian firm, declared insolvency as a result of fraudulent financial reporting. The CEO of Parmalat has been accused of mishandling $10 billion and of hiding losses in offshore funds and bank accounts. The scandal at Parmalat could also have serious consequences for the company's auditor (Gallani & Trofimov, 2004).

The auditor's responsibility for fraud detection in the U.S. has been defined in Statement on Auditing Standards No. 99, *Fraud Detection in a GAAS Audit* (AICPA, 2002). This statement has four key provisions (Lanza, 2002): (1) increased emphasis on professional skepticism, (2) frequent discussion among audit team personnel regarding the risk of misstatement due to fraud, (3) random audit testing of locations, accounts, and balances, and (4) procedures to test for management override of controls. Auditors are discouraged from placing too much reliance on client representation and are required to maintain a skeptical attitude throughout the audit. The standard encourages auditors to engage in frequent discussion among engagement personnel regarding the risk of material misstatement due to fraud. SAS 99 also requires auditors to inquire of management and others not directly involved with fraud, perform analytical procedures, and conduct necessary tests to assess management override of controls. Finally, auditors are advised to evaluate the risk of fraud and steps taken by the client to mitigate the risk of fraud.

The U.S. Congress in 2002 passed the Sarbanes-Oxley Act, which spells out a number of steps firms must take to minimize fraudulent financial reporting. This legislation requires the principal executive officer and the principal financial officer of publicly traded companies to certify the appropriateness of the financial statements and disclosures in each quarterly and annual report that their company issues. These officers are also responsible for establishing and maintaining internal controls within the company. Further, they must disclose to auditors and the audit committee of the board of directors any fraud, whether or not material, involving management or employees who have a significant role in defining or implementing internal controls. As this law goes into effect, evaluation and reporting of a company's internal controls and financial statements in order to detect fraud becomes even more critical, and must be on-going.

Prior research shows that various kinds of decision aids may be used to assist the auditor in detecting financial reporting fraud. Bell, Szykowny, and Willingham (1993) used bivariate and cascaded logit to assess the likelihood of management fraud. Their model achieved within-sample correct classification of 97% on the fraud observations and 75% on the non-fraud observations. Hansen, McDonald, Messier, and Bell (1996) used a generalized qualitative response model to predict management fraud. They reported 89.3% predictive accuracy over 20 trials. Bell and Carcello (2000) developed a logistic regression model as a decision aid to assist in the auditor's fraud decision. Auditors may also use an expert system as a decision aid to assist in fraud determination. Eining, Jones, and Loebbecke (1997) examined the effect that the use of an expert system has on auditor decision-making ability in detecting fraud. Their research showed that in allowing the interaction between the auditor and the system, the expert systems that have been used to assist auditors in complex decision processes often give results that are more accurate and consistent. Similarly, Whitecotton and Butler (1998) found that allowing decision makers to select information for the decision aid increases decision aid reliance. Fuzzy clustering may also be used as a decision aid for an auditor to detect fraudulent financial reporting (Lenard & Alam, 2004).

Fuzzy Clustering

When available data does not suggest a clear answer, decision makers often look for patterns or groups in the underlying data to make a decision (Alam, Booth, Lee, & Thordarson, 2000). While discriminant analysis and logistic regression assign observations to groups that were defined in advance, cluster analysis is the art of finding groups in data (Kaufman & Rousseeuw, 1990). Fuzzy set theory, introduced by Zadeh (1965), attempts to classify subjective reasoning (e.g., a human description of "good", "very good", or "not so good") and assigns degrees of possibility in reaching conclusions (Lenard, Alam, & Booth, 2000). As opposed to hard clustering, where there is a clear-cut decision for each object, fuzzy clustering allows for ambiguity in the data by showing where a solution is not clearly represented in any one category or cluster. Fuzzy clustering shows the degree to which (in terms of a percentage) an item "belongs" to a cluster of data. In other words, a data item may belong "partially" in each of several categories. The strength of fuzzy analysis is this ability to model partial categorizations.

Lau, Wong, and Pun (1999) used neural networks and fuzzy modeling to control a plastic injection-molding machine. They suggested that the neural network and fuzzy technology complement each other and offset the pitfalls of computationally intelligent technologies. Alam et al. (2000) used a combination of fuzzy clustering and self-organizing neural networks, and were successful in identifying potentially failing banks. Ahn, Cho, and Kim (2000) reported results using these technologies to predict business failure, and stressed the importance of these predictions as useful in aiding decision makers. Lenard et al. (2000) used fuzzy clustering to identify two different categories of bankruptcy. Companies placed in the second bankrupt category exhibited more extreme values in terms of the financial ratios used in the study. Companies either showed much better results (such as a high current ratio) than would be expected for a company facing bankruptcy, or the companies showed very poor results, such as a much higher debt ratio than any of the other bankrupt companies in the data sample. Lenard and Alam (2004) operationalized a fuzzy logic model for fraud detection in an Excel spreadsheet. By using the fuzzy logic model to develop clusters for different statements representing red flags in the detection of fraud, non-financial data was included with financial statement variables for the analysis. The overall prediction accuracy for the model was 86.7%.

Nolan (1998) used expert fuzzy classification and found that fuzzy technology enables one to perform approximate reasoning, as when a student assignment is graded as "very good", or "not so good", and improves performance in three ways. First, performance is improved through efficient numerical representation of vague terms, because the fuzzy technology can numerically show representation of a data item in a particular category. The second way performance is enhanced is through increased range of operation in ill-defined environments, which is the way that fuzzy methodology can show partial membership of data elements in one or more categories that may not be clearly defined in traditional analysis. Finally, performance is increased because the fuzzy technology has decreased sensitivity to "noisy" data, or outliers. Ammar, Wright, and Selden (2000) used a multilevel fuzzy rule-based system to rank state financial management. The authors used fuzzy set theory to represent imprecision in evaluated information and judgments. Pathak, Viyarthi, and Summers (2003) developed a fuzzy logic based system for auditors to identify fraud in settled claimed insurance. They believe that their system was able to cut costs by detecting fraudulent filings.

CRITICAL ISSUES OF FUZZY LOGIC

The fuzzy clustering procedure used by Lenard et al. (2000) and Lenard and Alam (2004) is called FANNY (Kaufman & Rousseeuw, 1990). The program FANNY uses "fuzziness" to partition objects by avoiding "hard" decisions, or clustering into fixed, definite categories. For each item in the dataset, the algorithm provides k+1 pieces of information, where k is the number of clusters that are used in the clustering algorithm. The k+1 pieces of information are: U_{iv}, the membership coefficient of item i in cluster v, v = 1 ... k, and S_i , the silhouette coefficient of item i. A higher value of U_{iv} indicates a stronger association of item i and cluster v. The silhouette coefficients satisfy the constraints $-1 \leq S_i \leq 1$ and indicate how a well-clustered object uses average distances from its own cluster to the closest neighboring clusters. The closer S_i is to 1 the better the clustering of an individual item. A value of S_i close to -1 indicates that an item may be assigned to more than one cluster (Alam et al., 2000). The Euclidean distance measure is used to compute distances between objects and to quantify the degree of similarity for each object. The degree of similarity for each objects *i* and *j* is computed as follows (Kaufman & Rousseeuw, 1990):

$$d(i,j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \ldots + (x_{ip} - x_{jp})^2} \tag{1}$$

where the p^{th} measurement of the i^{th} object is given by x_{ip} and *d(i,j)* is the actual distance between objects *i* and *j*.

Several authors have expanded upon the fuzzy clustering algorithms. Van den Bergh and van den Berg (2000) developed a competitive learning algorithm using fuzzy frequency distributions. They emphasized that sometimes the discovery of exceptions is more important than the main rules. For example, profit opportunities often seem to appear randomly and infrequently, so the agent should concentrate on detecting the unusual, abnormal states, or exceptions, rather than the average normal states (van den Bergh & van den Berg, 2000). Their algorithm seeks to find a mapping from an *M*-dimensional input space X into an *N*-dimensional output space *Y*, given a representative data set *S*. The set *S* contains *P* data pairs $(x_p; y_p)$. The final formula for the mapping is depicted as:

$$y_b = \sum_{c=1}^{C_y} \overline{y_c} \quad \mathrm{X} \quad u_c^{-Y} \mid u_b^{-X} \tag{2}$$

Thus, the sum of output cluster centroids y_c is weighted by the "local" membership values $u_c^{-Y} \mid u_b^{-X}$.

Fuzzy clustering algorithms have also been extended by the work of Kaymak and Setnes (2000), who proposed fuzzy clustering algorithms with volume prototypes and similarity based cluster merging. These extensions reduce sensitivity of the clustering algorithms to bias in data distribution, and help determine the number of clusters automatically.

Finally, Mashor (2001) proposed a clustering algorithm called adaptive fuzzy *c*-means clustering. In this method, each data sample is assigned a membership grade to indicate the degree of belonging to each center rather than assigning the data sample to one center as in "hard" clustering algorithms like the *k*-means clustering. In addition, the clustering program is not as sensitive to initial centers and gives better performance. The algorithm only requires the data to be presented once, instead of requiring multiple presentations, and as such reduces the computational load.

Identifying the number of clusters in fuzzy clustering is a challenging task. The optimal clustering should consider both fuzzy compactness and separation. The current state of the art does not provide a theoretical basis for an optimal choice of clusters. The objective function based fuzzy clustering algorithms are often used to divide the data into a predetermined number of clusters. The fuzzy *c*-means algorithm is one of the most popular objective functions used for fuzzy clustering. It uses the similarity between objects to measure the distances between clusters. The validity of the clusters is often assessed after the clusters have been formed. Validity Measures typically address the issues of the compactness of the clusters and the distances between them (e.g., Pal & Bezdek, 1995; Xie & Beni, 1991). Gath and Geva (1989) proposed a validity measure, which is a ratio between fuzzy compactness and separation. Bensaid et al. (1996) argued that combining the validity guided clustering algorithm with the fuzzy *c*-means considerably improves the partitions generated by fuzzy *c*-means alone. Various other studies have addressed the issue of the number of clusters, but there is no generally accepted approach of a priori selecting the appropriate number of clusters. Cluster validity tests are the only means available to decide whether the number of clusters used captures the underlying characteristics of the data. Investigating and resolving these issues is crucial in the analysis of financial statement data. The assignment to a particular cluster would determine whether or not the data being analyzed is a high "red flag" indicator of fraud.

FUTURE TRENDS

In addition to fraud determination, there are also other decisions that accounting and financial personnel make that affect financial reporting. Specifically, in the field of auditing, there is the auditor's decision reflected in the audit report about whether the company can continue as a going concern. The auditor uses financial statement and non-financial statement data, and a framework of questions to make the going concern judgment. The auditor also applies judgment in the consideration of materiality. Materiality judgment is closely linked to the analysis of fraud because the auditor must decide the extent to which a discrepancy affects the credibility of financial statements. These decisions may be enhanced by the use of statistical models, expert reasoning, data mining tools, and now fuzzy logic, to provide support for the auditor's judgment.

CONCLUSION

Financial statement information is used by management, employees, outside analysts, investors and creditors to assess the health of publicly traded companies. Just as this information is now readily available through the Internet and online financial services, so should tools that help in the analysis of that information be readily available or easily obtained. As the different methods of fuzzy analysis become more prevalent, there will be additional opportunities for using fuzzy logic in various other applications.

REFERENCES

Ahn, B.S., Cho, S.S., & Kim, C.Y. (2000). The integrated methodology of rough set theory and artificial neural network for business failure prediction. *Expert Systems with Applications, 18*, 65-74.

Alam, P., Booth, D., Lee, K., & Thordarson, T. (2000). The use of fuzzy clustering and self-organizing neural networks for identifying potentially failing banks: An experimental study. *Expert Systems with Applications, 18*, 185-99.

American Institute of Certified Public Accountants (AICPA) (2002). *Consideration of fraud in a financial statement audit. Statement on Auditing Standards No. 99.* New York: AICPA.

Ammar, S., Wright, R., & Selden, S. (2000). Ranking state financial management: A multilevel fuzzy rule-based system. *Decision Sciences, 31*(2), 449-481.

Arnold, J. (2003). Worries mount for Ahold. BBC News. *http://news.bbc.co.uk/1/hi/business/2797097.stm*. February 26.

Bell, T.B. & Carcello, J.V. (2000). A decision aid for assessing the likelihood of fraudulent financial reporting. *Auditing: A Journal of Practice & Theory, 19*(1), 169-184.

Bell, T.B., Szykowny, S., & Willingham, J.J. (1993). *Assessing the likelihood of fraudulent financial reporting: A cascaded logit approach*. Working Paper, KPMG Peat Marwick, Montvale, NJ.

Bensaid, A.M., Hall, L.O., Bezdek, J.C., Clarke, L.P., Sibliger, M.L. Arrington, J.A., & Murtagh, R.F. (1996). Validity-guided (Re) clustering with applications to image segmentation. *IEEE Transactions on Fuzzy Systems, 4*(2), 112-123.

Eining, M., Jones, D.R., & Loebbecke, J.K. (1997). Reliance on decision aids: An examination of auditors' assessment of management fraud. *Auditing: A Journal of Practice & Theory, 16*(2), 1-19.

Gallani, A., & Trofimov, Y. (2004). Behind Parmalat chief's rise: Ties to Italian power structure. *Wall Street Journal*, March 8, A1.

Gath, J., & Geva, A.B. (1989). Unsupervised optimal fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 11*(7), 32-57.

Hansen, J.V., McDonald, J.B., Messier, Jr., W.F., & Bell, T.B. (1996). A generalized qualitative-response model and the analysis of management fraud. *Management Science, 42*(7), 1022-1032.

Kaufman, L., & Rousseeuw, P.T. (1990). *Finding groups in data: An introduction to cluster analysis*. New York: John Wiley.

Kaymak, U., & Setnes, M. (2000). Extended fuzzy clustering algorithms. *ERIM Report Series in Management, 51*, 1-24.

KPMG. (2003). *Fraud survey 2003*. Montvale, NJ.

Lanza, R.B. (2002). New audit standard approved-SAS 99 "consideration of fraud in financial statement audit." *http://www/aicpa.org/pubs/tpcpa/nov2002/anti fraud.htm*

Lau, H.C.W., Wong, T.T., & Pun, K.F. (1999). Neural-fuzzy modeling of plastic injection molding machine for intelligent control. *Expert Systems with Applications, 17*, 33-43.

Lenard, M.J., & Alam, P. (2004). The use of fuzzy logic and expert reasoning for knowledge management and discovery of financial reporting fraud. In H.R. Nemati & C.D. Barko (Eds.), *Organizational data mining: Leveraging* *enterprise data resources for optimal performance* (pp. 230-262). Hershey, PA: Idea Group Publishing.

Lenard, M.J., Alam, P., & Booth, D. (2000). An analysis of fuzzy clustering and a hybrid model for the auditor's going concern assessment. *Decision Sciences, 31*(4), 861-864.

Mashor, M.Y. (2001). Adaptive fuzzy *c*-means clustering algorithm for a radial basis function network. *International Journal of Systems Science, 32*(1), 53-63.

Nolan, J.R. (1998). An expert fuzzy classification system for supporting the grading of student writing samples. *Expert Systems with Applications*, *15*, 59-68.

Pal, N.R., & Bezdek, J.C. (1995). On cluster validity for the fuzzy c-means model. *IEEE Transactions on Fuzzy Systems, 3*, 370-379.

Pathak, J., Viyarthi, N., & Summers, S.L. (2003). *A fuzzy-based algorithm for auditors to detect element of fraud in settled insurance claims.* Working paper, University of Windsor.

Van den Bergh, W-M., & van den Berg, J. (2000). Competitive exception learning using fuzzy frequency distributions. *ERIM Report Series Research in Management, 6*, 1-12.

Whitecotton, S.M., & Butler, S.A. (1998). Influencing decision aid reliance through involvement in information choice. *Behavioral Research in Accounting, 10*(Supplement), 182-201.

Xie, X.L., & Beni, G. (1991). A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *13*(8), 841-847.

Zadeh, L.A. (1965). Fuzzy sets. *Information and Control, 8*, 338-353.

KEY TERMS

Cluster Analysis: Defining groups based on the "degree" to which an item belongs in a category. The degree may be determined by indicating a percentage amount.

Data Mining: Using powerful data collection methods to analyze a company's database or data stores and select information that supports a specific objective.

Expert Reasoning: Implementing rules or procedures, often programmed to occur automatically, in order to make a decision. Background and heuristics that identify how to reach a conclusion are based on the knowledge of human experts in that field.

Fraudulent Financial Reporting: Intentional or reckless conduct, whether by act or omission, that results in materially misleading financial statements.

Fuzzy Logic: A mathematical technique that classifies subjective reasoning and assigns data to a particular group, or cluster, based on the degree of possibility the data has of being in that group.

Internal Controls: Procedures applied by a business organization that ensure information is safeguarded, that it is accurate and reliable, and that it is processed efficiently and in accordance with management's prescribed policies.

Management Fraud: A situation in which management misrepresents the financial condition of their firm. They may do so for personal financial gain or to disguise the financial results of their company.

Application Service Provision for Intelligent Enterprises

Matthew W. Guah
Warwick University, UK

Wendy L. Currie
Warwick University, UK

ROAD TO ASP

Several historical shifts in information systems (IS) involved strategies from a mainframe to a client server, and now to application service provision (ASP) for intelligent enterprises. Just as the steam, electric, and gasoline engines became the driving forces behind the industrial revolution of the early 1900s, so the Internet and high-speed telecommunications infrastructure are making ASP a reality today. The current problem with the ASP model involves redefining success in the business environment of the 21st century. Central to this discussion is the idea of adding value at each stage of the IS life cycle. The challenge for business professionals is to find ways to improve business processes by using Web services.

It took mainframe computers a decade or two to become central to most firms. When IBM marketed its first mainframe computer, it estimated that 20 of these machines would fulfil the world's need for computation! Minicomputers moved into companies and schools a little faster than mainframes, but at considerably less costs. When the first computers were applied to business problems in the 1950s, there were so few users that they had almost total influence over their systems. That situation changed during the 1960s and 1970s as the number of users grew. During the 1980s the situation became even tighter when a new player entered the picture—the enterprise (McLeord, 1993). In the 21st century, information systems are developed in an enterprise environment (see Diagram 1).

Beniger (1986) puts forth a seemingly influential argument that the origin of the information society may be found in the advancing industrialisation of the late nineteenth century. The Internet is simply a global network of networks that has become a necessity in the way people in enterprises access information, communicate with others, and do business in the 21st century. The initial stage of e-commerce ensured that all large enterprises have computer-to-computer connections with their suppliers via electronic data interchange (EDI), thereby facilitating orders completed by the click of a mouse. Unfortunately, most small companies still cannot afford such direct connections. ASPs ensure access to this service costing little, and usually having a standard PC is sufficient to enter this marketplace.

The emergence of the ASP model suggested an answer to prevailing question: Why should small businesses and non-IT organisations spend substantial resources on continuously upgrading their IT? Many scholars believed that outsourcing might be the solution to information needs for 21st century enterprises (Hagel, 2002; Kern, Lacity & Willcocks, 2002; Kakabadse & Kakabadse, 2002). In particular, the emergence of the ASP model provided a viable strategy to surmount the economic obstacles and facilitate various EPR systems adoption (Guah & Currie, 2004). Application service provision—or application service provider—represents a business model of supplying and consuming software-based services over computer networks. An ASP assumes responsibility of buying, hosting, and maintaining a software application on its own facilities; publishes its user interfaces over the networks; and provides its clients with shared access to the published interfaces. The customer only has to subscribe and receive the application services through an Internet or dedicated intranet connection as an alternative to hosting the same application in-house (Guah & Currie, 2004). ASP is an IT-enabled change, a different and recent form of organisational change, evidenced by the specific information systems area (Orlikowski & Tyre, 1994). ASP has its foundations in the organisational behaviour and analysis area (Kern et al., 2002).

The initial attempt—by the ASP industry to take over the business world—was fuelled by the belief that utility computing offered a new business model to customers, similar to electricity, gas, and water. The commercialization of the Internet meant that, as network traffic increased in a firm's data centre, IT architecture would trigger other resources into action, including idle servers, applications, or pools of network storage. The firm would pay only for the amount of time it used the services. Thus, the concept of 'software-as-a-service' was created

(Kakabadse & Kakabadse, 2002). Accessing IT resources in this way would result in reduced up-front investment and expenditure, enabling firms to buy services on a variable-price basis (Dewire, 2000). This fuelled opportunities in the late 1990s for service providers to offer software applications and IT infrastructure on a rental, pay-as-you-go pricing model (Bennet & Timbrell, 2000). An ASP could be a commercial entity, providing a paid service to customers (Dussauge, Hart & Ramanantsoa, 1994) or, conversely, a not-for-profit organisation supporting end users (Currie, Desai & Khan, 2003).

ASP AREAS OF CONCERN

As evidence relating to the reality and basic features of the ASP market continues to grow, there begins to be less concern about confirming that any structural economic shift has continued historically, and more concern about understanding how the ASP industry is performing, and its impacts on productivity, investment, corporate capital formation, labour force composition, and competition.

The ASP business model is premised on the formation of strategic alliances and partnerships with technology and service providers (Ferergul, 2002). Telecommunications firms entering the ASP market with large IT infrastructures needed to partner with ISVs and hardware manufacturers. One of the significant strategic alliances was between Cable & Wireless (IT infrastructure), Compaq (hardware manufacturer), and Microsoft (ISV). Pure-play ASPs without a large investment in IT infrastructure needed to form strategic alliances with data centre and colocator firms (telcos) and ISVs. Some of the major reasons for businesses to implement an ASP business model are list in Table 1.

The ASP model was highly volatile, dynamic, and immature. A recent review of the ASP industry concluded that technological factors like scalability, the managerial aspects of speed and focus, and the behavioural aspects of price and flexibility were the key drivers of the model. The inhibitors of the model were poor connectivity, lack of trust in the model, reluctance to be locked into long-term contracts with suppliers, lack of customisation, poor choice and suitability of software applications from ASPs, and few opportunities to integrate disparate applications across technology platforms and business environments. These factors and others led Hagel (2002, p. 43) to conclude:

"ASPs in many respects represented a false start in the efforts to break out of the enterprise straitjacket. In particular, few of them adopted Web services architectures as their technology platform. Instead, they attempted to build businesses on the Internet using traditional technology architectures…this proved to be a significant flaw in the early ASP model and explains many difficulties these businesses experienced."

The business environment for intelligent enterprises (see Diagram 1) includes the enterprise itself and everything else that affects its success, such as competitors; suppliers; customers; regulatory agencies; and demographic, social, and economic conditions (Guah & Currie, 2004). As a strategic resource, ASP helps the flow of various resources from the elements to the enterprise, and through the enterprise and back to the elements.

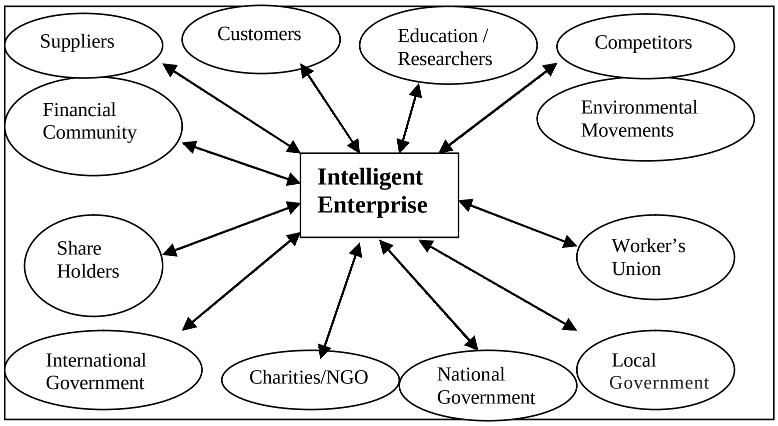
THE FUTURE OF THE ASP MODEL

According to Forester Research, the proportion of ASP business in the outsourcing market peaked at about $800 million in 2000 and was projecting for $25 billion by 2005. However, it actually declined by the year 2002 (due partly to the effect of stock market collapse) and currently is being projected at $15 billion by 2006. The overall business interests in the ASP model will continue to rise, with proportionally higher rates of investment by vendors

Table 1. A list of motivational factors to implement an ASP strategy

- To take maximise the capabilities of the Internet's latest technology
- To increase sales of products and services
- To reach a highly desirable demographic market
- To stay on top of competition
- To make changing information available quickly
- To test new products and services on the market
- To boast morale among staff and customers
- To experiment with an Internet model to business IT outsourcing

Diagram 1. A tool for controlling influences in a complex environment



versus traditional outsourcing. We attribute this optimistic forecast to four trends:

- continuing improvements in capabilities and cost-performance characteristics of Remote Support Services by vendors,
- improvements in capabilities and cost-performance characteristics of the technology at the system or application level,
- continual development of the telecommunications infrastructure to support ASP performance, and
- gradual reduction of institutional and social barriers to the introduction of the ASP model as a viable business strategy.

There are numerous papers warning that such accelerated Web service evolution increases the difficulty that other competitors have in adapting to ASP (Gottchalk, Graham, Kreger & Snell, 2002; Hondo, Nagaratnam & Nadalin, 2002; Stencil Group, 2002). By modifying the nature and the relative importance of the key factors for success in the ASP industry, Web service technological changes that are introduced by one or more of the vendors can lead to favourable changes in the competitive environment. In an industry built upon high volume, new technologies that are introduced by some of the competitors that nullify or minimise the impact of scale can significantly alter the nature of the competitive environment by making size a drawback rather than an advantage.

Diagram 2 shows that the holistic approach to technology always seems to work better than the piece-meal approach to information systems solution. Early stages of Web services are represented by a two-legged table. The current version of Web services being practiced by vendors after the dot.com crash is represented by the three-legged table in Diagram 2. But an even more successful model of Web services would be a properly architecture four-legged table, represented above. The analogy here is that a two-legged table is less stable than a three-legged table, while a four-legged table is even firmer.

CONCLUSION

We can safely conclude that policy makers in all fields, not just in IS, are forced into ill-considered conclusions and recommendations because they still view their management strategies in pre-Internet terms. Moreover, they are still constrained by statistical calculations based on outmoded and obsolete classification approaches, as well as on invalid assumptions about the fundamental sources of profit and capital formation—without full consideration for business environment.

Rethinking your business in terms of Internet economy, formulating new strategies for gaining competitive advantage, and raising the level of awareness of people throughout your enterprise to the notion that information itself can and should be looked upon as a strategic corporate asset¾these are great steps, but only the first steps for success in the 21st century. In addition, both structural and procedural changes must take place for an intelligent enterprise to put its convictions into operation. Could ASP provide you with such a necessary tool thereby directing your focus into the reality of a 21st century intelligent organisation?

Table 3.Summary of areas affecting the growth of the ASP market

Widespread model ignorance and perceptions among small and medium businesses. Lack of adequate understanding of the ASP business model and its understanding.

IS infrastructure raises a broad range of economic, social, and technical issues. Who should pay for infrastructure? Who should have access to/control over them and at what cost? Which technology should it include? Where ASP is involved, the economic question often puts telephone companies against cable companies, both of whom can provide similar capabilities for major parts of the telecommunications system.

Telecommunications facilitates ASP emancipation. Telecommunications has become virtually inseparable from computer with a paired value that is vital for integrating enterprises. As an e-commerce phenomenon, a few of the essentials of an ASP infrastructure are Common Carriers, Value-Added Networks, Private Line, and Private Networks.

Issues of security. As a program executed upon accessing a Web page, ASP carries a security risk because users end up running programs they don't know and/or trust. The latest solution is encryption, but does any mathematical encryption guarantee absolute security? No. Just as a physical lock cannot provide absolute safety, encryption cannot guarantee privacy—if a third party uses enough computers and enough time, they will be able to break the code and read the message. The encryption system only guarantees that the time required to break the code is so long that the security provided is sufficient. When someone asserts that an encryption scheme guarantees security, what they actually mean is that although the code can be broken, the effort and time required is great. Thus, an encryption scheme that requires a longer time to break than another scheme is said to be 'more secure'.

Overcoming organisational obstacles to a commercial future. The issue of organisation is based on the way ASP has evolved. The ASP industry lacks the type of clear organisation that would make it easy to use as a reliable and profitable business model.

ASP as competitive investment. Not so much return on investment for ASP, rather what is the cost of not investing? Will an enterprise lose customers and market share because it does not have a particular technology in place? Can you enter a new line of business without investing in the technology that competitors have adopted? What kinds of services do customers expect? These are competitive investment issues raised by ASP.

Change is also an opportunity. One stimulus for ASP solution implementation is its intention to transform the enterprise. In this sense, the investment in ASP is part of a larger change programme that is meant to enable intelligent enterprises' virtual and multiple-team structures. The resulting contributions can be described as part of the outcome of a general change effort.

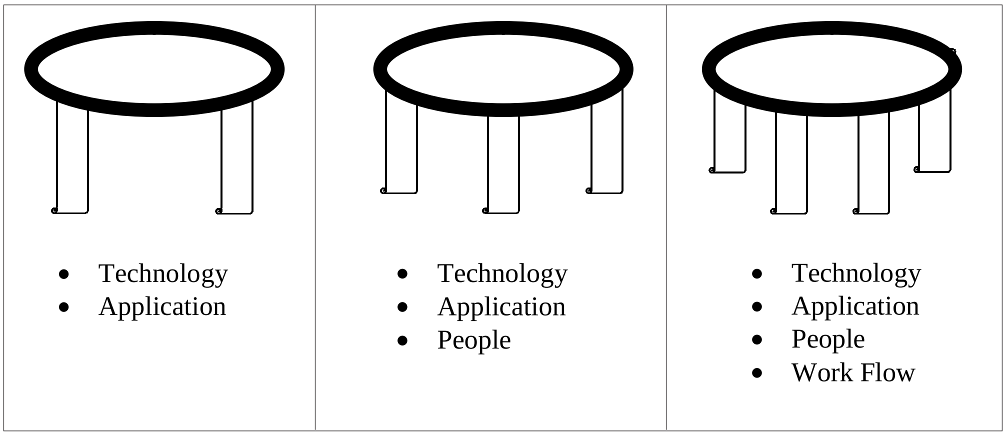
Social-technical issues. The social objectives refer to the expectations of major stakeholders (i.e., employees). An ASP business model that provides information and tools for employees increases involvement because they reinforce the employee's authority and responsibility for work. Those that provide information to managers or quality inspectors but do not support employees can reduce involvement by reinforcing the suspicion that the employee is not really responsible.

Tangible and intangible benefits. The tangible benefits of an ASP solution can be measured directly to evaluate system performance. Examples include reduction in the time for completion of transaction, improvement in response time, reduction in the cost of assets, and reduction in the error rate. Intangible benefits affect performance but are difficult to measure because they refer to comparatively vague concepts. A few intangible benefits of a solution are: better coordination; better supervision; better morale; better information for decision making; ability to evaluate more alternatives; ability to respond quickly to unexpected situations; and organisational learning. Although hard to quantify, intangible benefits are important and shouldn't be ignored, as many IS benefits to organisations are intangible.

The role of government. Modernisation of the machinery and process of government will accommodate many Internet strategies like ASP. This will involve reform of intellectual property law to accommodate access to and exploitation of works via the Internet. It will include administration of Internet domain names on an international basis. Finally, the facilitation of e-commerce development will include national and international initiatives and measures to protect both suppliers and consumers operating within this global electronic marketplace.

Blurring in-house IT and ASP services. As the industry evolves and becomes more complex, the need for new services and specialisation in the division of labour continues to increase. In-house migrates into strategic management and monitoring of IT standard, while ASP migrates into value-added services so that 'business IT becomes a service in a package form'. As the boundaries between in-house and ASP become more blurred through the use of improved communications technologies, the opportunities for entrepreneurs continue to increase.

Diagram 2. Evolution of Web services



REFERENCES

Beniger, J.R. (1986). *The control revolution: Technological and economic origins of the information society*. Boston: Harvard University Press.

Bennett, C. & Timbrell, G.T. (2000). Application service providers: Will they succeed? *Information Systems Frontiers*, *2*(2), 195-211.

Currie, W., Desai, B., Khan, N., Wang, X. & Weerakkody, V. (2003, January). Vendor strategies for business process and applications outsourcing: Recent findings from field research. *Proceedings of the Hawaii International Conference on Systems Sciences*, Hawaii.

Dewire, D.T. (2000). Application service providers. *Information Systems Management*, *17*(4), 14-19.

Dussauge, P., Hart, S. & Ramanantsoa, B. (1994). *Strategic technology management: Integrating product technology into global business strategies for the 1990s*. Chichester: John Wiley & Sons.

Ferergul, C. (2002). Best practices in Web hosting service level agreements. Stamford, CT: Meta Group. Retrieved May 2, 2002, from *techupdate.zdnet.com/techupdate/stories/main/*

Gottschalk, K., Graham, S., Kreger, H. & Snell, J. (2002). Introduction to Web services architecture. *IBM Systems Journal*, *41*(2).

Guah, M.W. & Currie, W.L. (2004). Application service provision: A technology and working tool for healthcare organizations in the knowledge age. *International Journal of Healthcare Technology and Management*, *6*(1/2), 84-98.

Hagel III, J. (2002). *Out of the box: Strategies for achieving profits today and growth tomorrow through Web services*. Boston: Harvard Business School Press.

Hondo, M., Nagaratnam, N. & Nadalin, A. (2002). Securing Web services. *IBM Systems Journal*, *41*(2).

Kakabadse, N. & Kakabadse, A. (2002). Software as a service via application service providers (ASPs) model of sourcing: An exploratory study. *Journal of Information Technology Cases and Applications*, *4*(2), 26-44.

Kern, T., Lacity, M. & Willcocks, L. (2002). *Netsourcing: Renting business applications and services over a network*. New York: Prentice-Hall.

McLeord Jr., R. (1993). *Management information systems: A study of computer-based information systems* (5th edition). New York: Macmillan.

Orlikowski, W.J. & Tyre, M.J. (1994). Windows of opportunity: Temporal patterns of technological adaptation in organizations. *Organization Science*, (May), 98-118.

Stencil Group. (2002). Understanding Web services management: An analysis memo. Retrieved May 2002 from *www.stencilgroup.com*

KEY TERMS

ASP: A third-party service firm that deploys, manages, and remotely hosts software applications through centrally located services in a rental or lease agreement

(ASP Consortium, 2000). Such application deliveries are done to multiple entities from data centres across a wide area network (WAN) as a service rather than a product, priced according to a license fee and maintenance contract set by the vendor. An ASP is considered by many to be the new form of IT outsourcing, usually referred to as application outsourcing.

ASP Aggregator: The ASP aggregator model is based on the premise that the rapid proliferation of firms offering ASP services has created an overly complex market for medium-sized enterprises to deal with when investigating application outsourcing options.

Agent-Based Approach to ASP: This approach to ASP is well equipped to address the challenges of multi-market package to e-procurement. Service agents within the ASP model are the system's gateway to external sources of goods and services. These agents are usually aware of the source's market model and of the protocols it uses (Zhang et al., 2000). Service agents are not only able to determine which requests it can service, but also proactively read these requests and try to find an acceptable solution.

Common Carriers: Companies that are licensed, usually by a national government, to provide telecommunications services to the public, facilitating the transmission of voice and data messages.

Infrastructure: An emerging class of companies have opted to approach the ASP market by providing infrastructure management and outsourcing services to ASPs, freeing up their resources to focus more directly on application management issues (telco, data centre, networking).

Internet Service Providers (ISP): Provides access to the Internet via different communications channels such as traditional telephone lines or a high-speed fibre optics channel.

Pure-Play ASP: Those with non-specific industry-required product or service, except Internet/Web-enabled software applications (e-mail/security/disaster recovery). Firms offering this service suffer from the sales of unprofitable commodity applications and therefore have greater reliance on venture capitalist funding as they operate in a rather unstable, volatile, and dynamic market.

Vertical Service Provider (VSP): Vertically focused ASPs offering industry-specific applications are also emerging. Their basic premise is that each industry (health, finance, transportation) has its own unique set of characteristics that can best be served by companies that focus exclusively on the given industry.

Virus: A malicious code added to an e-mail program or other downloadable file that is loaded onto a computer without the user's knowledge and which runs often without the user's consent. Computer viruses can often copy themselves and spread themselves to a user's e-mail address book or other computers on a network.

Web Services: Web Services technology is one of the most important foundations for ASP new-game strategies. Thus, by accelerating the pace of Web services in the industry, a competitor with good capability in the technology reinforces its own competitive position.

Applying a Metadata Framework to Improve Data Quality

Victoria Youngohc Yoon
University of Maryland, Baltimore, USA

Peter Aiken
Defense Information Systems Agency, USA

Tor Guimaraes
Tennessee Technological University, USA

INTRODUCTION

The importance of a company-wide framework for managing data resources has been recognized (Gunter, 2001; Lee, 2003, 2004; Madnick, Wang & Xian, 2003, 2004; Sawhney, 2001; Shankaranarayan, Ziad & Wang, 2003). It is considered a major component of information resources management (Guimaraes, 1988). Many organizations are discovering that imperfect data in information systems negatively affect their business operations and can be extremely costly (Brown, 2001; Keizer, 2004). The expanded data life cycle model proposed here enables us to identify links between cycle phases and data quality engineering dimensions. Expanding the data life cycle model and the dimensions of data quality will enable organizations to more effectively implement the inter- as well as intra-system use of their data resources, as well as better coordinate the development and application of their data quality engineering methods.

BACKGROUND

The proposed model has a number of inputs/outputs distributed throughout eight phases: metadata creation, metadata structuring, metadata refinement, data creation, data utilization, data assessment, data refinement, and data manipulation. Each of these phases is described next in more detail.

Two possible cycle "starting points" are shown bolded in Figure 1. The first starting point is applicable to new systems where there exists no data to be migrated, and/or converted from existing system(s). In these instances, the model cycle begins with metadata creation and proceeds counter- clockwise around the cycle. However, according to a survey of CIOs by Deloitte & Touche (1996), an average of more than 90% of organizational legacy systems were scheduled to be replaced by 2001. Legacy systems continue to impede business effectiveness for many. Only 6% of insurers plan to replace their systems in the next two years, while 29% plan to replace their legacy systems within four years (Chordas, 2004). In a survey of 115 business-technology professionals conducted by Optimize Research, 42% called their legacy IT a barrier to innovation and flexibility, while 8% said that it bogs down business and is expensive to maintain (Whiting, 2003). Thus, it is more likely that an organization's legacy data will become the major data asset to be managed. In these cases where data already exist, structural data quality re-engineering becomes necessary, and the cycle begins with data assessment. Next, each cycle phase is described in more detail.

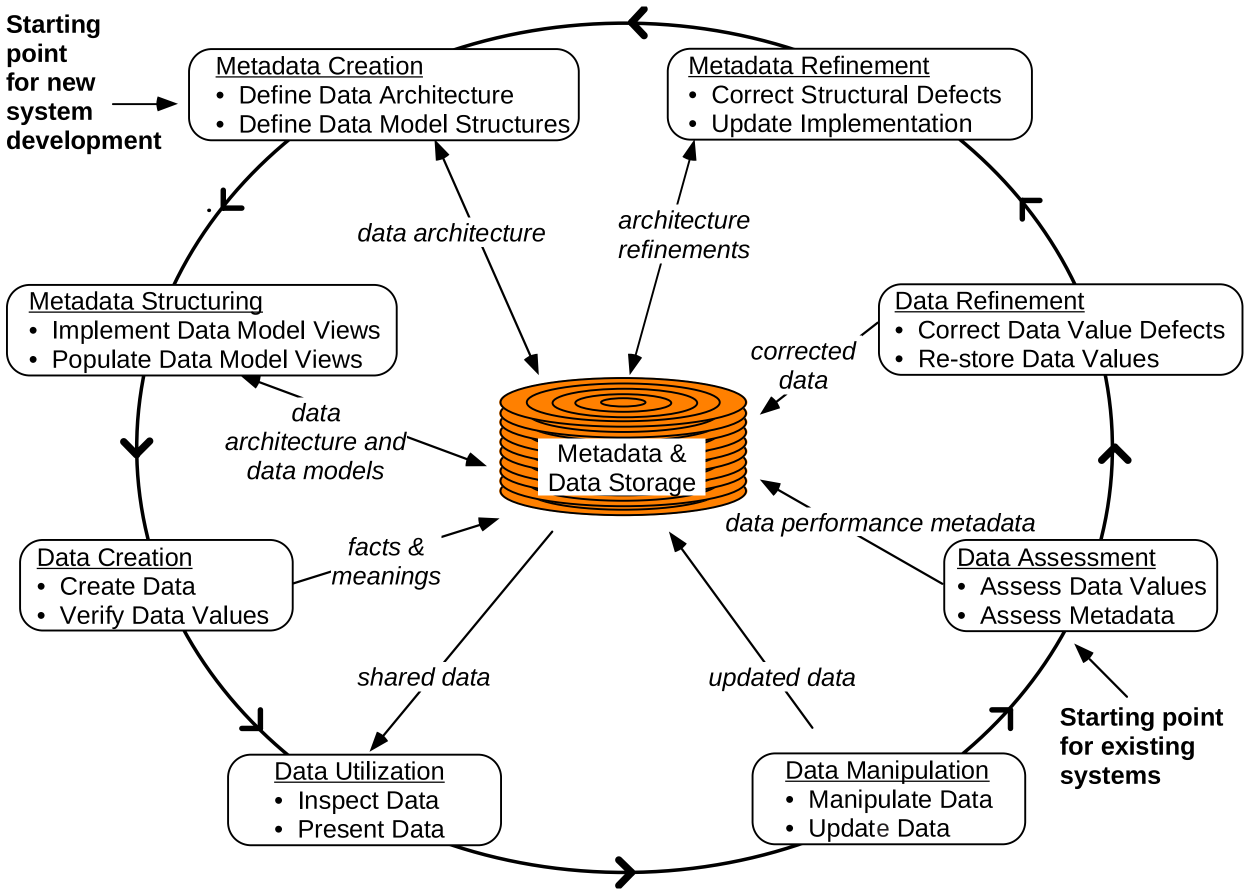
Metadata Creation

When the requirements dictate that users interact with multiple systems across functional area boundaries, a formal organizational data architecture is required to coordinate data quality engineering efforts. While all organizations have data architectures, only formally specified architectures can be formally managed. This phase typically corresponds to increasing awareness of data as an organizational asset. The architectural metadata created and evolved consist of the organizational data architecture structure definitions and specific associations among individual system data models.

Metadata Structuring

This phase focuses on developing a framework guiding the organizational data architecture implementation as it populates data models in the next phase. Metadata creation is followed by the development of a data model structure. Data models must also be evolved. The term

Figure 1. Newly proposed eight phases of extended data life cycle model with metadata sources and uses



"structuring" indicates the iterative development process that occurs as the organizational data architecture structure developed during the previous phase is populated with metadata. Defining data model structures permits organizations to understand the categories of data that comprise its data models. The process consists of populating the data architecture with data models describing the various specific systems. Each data model corresponds to one physical occurrence. In addition, when physically implemented, logical model components can be physically implemented by multiple systems, accessing common DESs. The process of defining data models as components extends the organizational data architecture comprehensiveness. Metadata structuring is complete when all entities can be associated with specific model components. Perfect model metadata occurs when a correct data model exists for each physical system, and each physical system component is associated with one and only one common organizational data architecture component.

Metadata Refinement

At various points, portions of some metadata can be determined imperfect. Architecture refinement implements an iterative approach to refining the existing metadata-based concepts, correcting factual errors, and evolving the structure to a more perfect state. This usually occurs in response to data assessment activities.

Data Creation

Data creation occurs when data values are captured from some external source and stored in systems. Data sources can range from a point of sale terminal, to EDI, to floppy disk exchange. Data creation is the most popular focus of data quality engineering efforts. These are commonly implemented as edit masking, range checking, or other forms of validation. Data value quality efforts are aimed at perfecting data values as they are captured and before they are stored or re-stored in the database.

Data Utilization

Data utilization occurs as the data are provided as information in response to a request from a user or a process. The focus of data quality engineering efforts for this phase is on appropriate data representation; that is, taking data from a storage location and properly presenting it to a user or a process as requested.

Data Assessment

Often occurs in response to complaints of imperfect data. It is assessed formally or informally to determine data suitability for current or future use. If data are judged inadequate, the assessment also determines if the problem causes are practice caused or structurally caused. Practice-caused problems are corrected through the data refinement phase, while structural problems are amended through the metadata refinement, creation, structuring phases. Structural changes must be applied at an organizational architecture level.

Data Refinement

If the cause of imperfect data is determined to be practice-oriented, the data values are corrected using a data refinement procedure. Data refinement refers to the process of altering data within the existing data structures. This continues to be a popular focus of data value quality engineering efforts.

Data Manipulation

Often times data are accessed in order to be altered, deleted, or otherwise manipulated. Data manipulation is the process of altering data forms or data values. Any change can introduce error, and the data quality engineering focus is similar to that described previously.

Developers have typically been task oriented when developing specific systems based on data modeling. Most data quality methods are also usually system focused, limiting their usefulness beyond the specific system boundaries. From an organizational perspective, absence of data definitions shared across functional areas makes data interchange among systems considerably more difficult. Thus, quality considerations based on these three dimensions can be insufficient to insure overall data quality. The increasingly widespread requirement that users interact with multiple systems, and the need for developers to build more highly integrated systems, demand an additional data quality dimension. Therefore, we propose a fourth dimension to data quality, which addresses organizational data architecture, to coordinate data management activities in cross-functional system development and operations. The chief characteristic of this dimension is an architecture providing an effective linkage between the organization-wide user data requirements and the specific systems implemented to satisfy such requirements.

Quality Engineering Dimension Attributes

Based on our literature survey, quality attributes for data values, data representations, and data models are recapped in Figure 2. Accompanying each attribute is its source(s) of reference(s) in the literature and a brief description.

Absent from the presently available literature are the attributes corresponding to data architecture quality. A list of nine organizational data architecture attributes is proposed as a collection of desirable characteristics for such architectures. This list has been developed based on the experience within the Department of Defense's (DOD) Center for Information Management's Information Engineering Directorate, where the second author worked as Program Manager and participated in the development of a DOD-wide data architecture development. We propose these attributes to increase reader awareness and further discussion in this critical area. The proposed attributes are:

Architectural Completeness

The architecture is comprehensive enough to be used by any functional area of the organization wishing to utilize it.

Architectural Correctness

The information describing the architecture is correctly represented with the appropriate methodology. That is, the organization can use the methodology to maintain uniform data definitions throughout the organization.

Management Utility

The data architecture is widely used by the organization in strategic planning and systems development as an indication of its utility. In practice, architectures too often wind up as shelf-ware.

Data Management Quality

The organization as a whole is data driven. Data models are developed and managed from an organization-wide perspective, guided by the organizational data architecture. Data are managed with distributed control from a centralized unit.

Data Sharing Ability

The data architecture serves as the basis for negotiating and implementing intra-organizational data exchange agreements by anticipating, defining and managing data sharing requirements within the organization and among its business partners using organization-wide standard metadata definitions.

Functional Data Quality

Data are engineered in support of business functional area requirements where data elements for individual systems are derived from organizational metadata requirements and implemented using organizational systems designed to support information representation.

Data Operation Quality

Data quality engineering is established as a functional area actively and consistently applying data quality engineering methods to data elements.

Evolvability

The organizational data architecture is maintained in a flexible, evolving fashion to enable the fulfillment of future user requirements.

Organizational Self-Awareness

Organization ability to investigate architecture use, and determine the types of value that it provides to end users. Feedback helps data architects refine the architecture to make it more useful organizationally.

FUTURE TRENDS

It is widely recognized that a company's data resources are vitally important for effective organizational operations. However, many organizations have experienced substantial financial losses and other negative impacts caused by imperfect data. A major reason for the negative impact is the lack of effective guidelines for data quality management. The problem is exacerbated by increased company dependence on computer technology for business operations and by strategic requirements necessary to improve company global competitiveness. To properly manage any data resource, companies must understand the roles it plays, its properties, the opportunities offered and the steps to take to explore those opportunities. While data present great possibilities there are also special challenges that must be addressed (Levitin & Redman, 1998).

Further adding to the problem is the trend toward systems integration within and among companies that has forced IS departments to change focus from developing and operating relatively isolated applications to new ones capable of supporting cross-functional and interorganizational business processes. Faced with such demands, data resources managers have received little guidance from researchers addressing the complete range of data quality dimensions they have to cope with. In order to obtain the required level of data quality, systems development efforts must consider all dimensions of data quality, including: data value; representation; data model; and data architecture quality. Based on previous research, a more comprehensive data quality model has been proposed, as well as the attributes applicable to each quality dimension. The scope of data quality engineering should be changed from multiple-triples within systems to organization-wide triple management. Data quality engineering will be more effective when also focused on rooting out structure-oriented causes of imperfect data by ensuring the quality of shared data across system boundaries, with common data definitions.

SUMMARY AND CONCLUSION

There are challenges and opportunities inherent in managing data. Development and use of the more comprehensive data quality metadata model will enable organizations to more effectively manage those dimensions, enhancing data quality. Data resources managers are encouraged to consider the expanded data life cycle model proposed here, and the attributes necessary to implement an organization-wide data architecture, for benchmarking purposes against their companies' presently used methodologies and data management activities.

REFERENCES

Brown, S. (2001). Making decisions in a flood of data. *Fortune*, *144*(3), 148.

Figure 2. Recap of research contributions to the attributes of data quality

	Described By:	Larry P. English	K. Laudon	A. Levitin and T. Redman 93	C. Fox, A. Levitin, & T. Redman 94	T. Redman 92	E. Tufte	Definition
Dimension	**Attribute**							
Data	Timeliness	X					X	Data should be promptly presented to the user at the time when it is needed.
Representation	Conciseness						X	Data presented to the users match user breadth/depth requirements without data loss.
Quality	Clarity						X	Data are presented in a form that is easiest for the user to understand given the request circumstances
	Consistency					X		Data presented to the users lacks nothing with respect to the user's information requirements.
	Detail					X		Data are presented in the level of detail most appropriate for the user's need.
	Accessibility						X	Data presented to the users is free from retrieval fault, data displayed unaltered from what was stored.
	Order							Data are presented in a sequence fitting the users need and their cognitive style.
	Flexibility					X		Data are able to be easily transformed between systems, formats, media to best match user needs.
	Portability					X		Data are able to be migrated from application to application without data loss.
	Presentation appropriateness	X					X	Data are presented in a format facilitating user comprehension.
	Media					X		Data are presented using media most effective for user comprehension.
	Unambiguousness/interpretability					X	X	Data presented to the users requires no interpretation to comprehend the correct value.
Data	Completeness	X	X		X	X		Attributes of entities requiring values have them.
Value	Correctness/accuracy	X	X		X	X		Data values maintained are free from fault, recording defects, or damage.
Quality	Currency				X	X		Data values maintained are the most up-to-date and match user expectations.
	Time Period					X		Data values maintained cover the time period required by the users.
	Clarity		X					Data values maintained match the breadth and depth of the user request parameters.
	Precision				X			Data values are maintained with the amount of precision or detail required by the user.
	Reliability				X			Data values stored can be depended upon by the user under stated conditions.
	Consistency					X		Data values continue to be maintained in a steady, dependable manner.
	Timeliness				X			Data values are updated as often as the user requires.
	Relevance					X		Data values stored are directly responsive to the specific user needs.
Data Model Quality	Completeness	X						The model is comprehensive enough to be used for a reference - containing complete enough subject areas to be of use.
	Definition clarity/unambiguity	X		X		X		The model is developed and maintained according to generally accepted modeling principles indicating the modelers consistently and correctly applied the techniques.
	Relevance			X		X		The model contents represent the facts of interest to the user.
	Value obtainability			X		X		The data model is structured so that users can obtain the facts they require.
	Comprehensiveness			X		X		This quality attribute addresses the issue "Did the modelers include all of the information they desired to in the model? Is this model populated with sufficient data to be useful?"
	Essentialness			X		X		The model contains only those elements fundamentally required to describe the subject.
	Attribute granularity			X		X		The model is structured so that it manipulates the level of detail desired by the users.
	Domain precision	X		X		X		The model maintains the factual precision desired by users.
	Naturalness			X		X		The model 'fits' with the way users assimilate facts into their work processes.
	Occurrence identifiability			X		X		The model maintains sufficient access means to uniquely identify facts required by users.
	Robustness			X		X		Both the model component definitions and the relationships between the entities are free from interpretation-based faults.
	Flexibility			X		X		The model maintained in a fashion where it is able to be useful in multiple applications
	Minimally redundant					X		The model is implemented using minimal factual duplication to meet user needs.

Chordas, L. (2004). Making ITs 'to-do' list. *Best's Review, 104*(11), 97-98.

Deloitte & Touche. (1996). Leading trends in information services. *Eighth Annual Survey of North American Chief Information Executives,* Wilton, CT.

English, L.P. (1996) Help for data quality problems – A number of automated tools can ease data cleansing and help improve data quality. *InformationWeek, 600,* 53.

Fox, C., Levitin, A., & Redman, T. (1994). The notion of data and its quality dimensions. *Information Processing and Management, 30*(1), 9-19.

Guimaraes, T. (1988). Information resources management: Improving the focus. *Information Resource Management Journal, 1*(1), 10-21.

Gunter, H. (2001). Taking a technology leap. *Computer Aided Engineering, 20*(5), 32-37.

Keizer, G., (2004). Gartner: Poor data quality dooms many IT projects. *TechWeb News, InformationWeek.* Retrieved May 25, 2004, from *http://informationweek.com/shared/printableArticle.jhtml?articleID=20301106*

Laudon, K. (1986). Data quality and due process in large interorganizational record systems. *Communications of the ACM, 29*(19), 4-11.

Lee, Yang W. (2003-2004, Winter). Crafting rules: Context-reflective data quality problem solving. *Journal of Management Information Systems, 20*(3), 93-119.

Levitin, A., & Redman, T. (1993). Models of data (life) cycle with applications to quality. *Information and Software Technology, 35*(3), 216-223.

Levitin, A., & Redman, T. (1998). Data as a resource: Properties, implications and prescriptions. *Sloan Management Review, 40*(1), 89-102.

Madnick, S., Wang, R., & Xiang X. (2003-2004). The design and implementation of a corporate householding knowledge processor to improve data quality. *Journal of Management Information Systems, 20*(3), 41-69.

Redman, T. (1992). *Data quality management and technology.* New York: Bantam Books.

Sawhney, M. (2001). Don't homogenize, synchronize. *Harvard Business Review, 79*(7), 86-95.

Shankaranarayanan, G., Ziad, M., & Wang, R. (2003). Managing data quality in dynamic decision environments: An information product approach. *Journal of Database Management, 14*(3), 14-32.

Tufte, E. (1990). *Chartjunk and the Challenger.* Cheshire, CT: Envisioning Graphics Press.

Whiting, R. (2003). Behind the numbers: Legacy-system updates get a second look. *InformationWeek, 962,* 70.

KEY TERMS

Data: Combination of facts and meanings that are processed into information.

Data Architecture: The underlying set of rules and descriptions of relationships that govern how the major kinds of data support the business processes defined in the business architecture.

Data Quality: A dimension or measurement of data in reference to its accuracy, completeness, consistency, timeliness, uniqueness and validity. Data are considered to be of high quality if they have all of the above attributes.

Data Resource Management: The analysis, classification and maintenance of an organization's data and data relationships.

Legacy Systems: Older computing technology used by an organization over time. Legacy systems perform business-critical functions and can be integrated with newer systems.

Metadata: Information that refers to one or more other pieces of information that can exist as separate physical forms. In short, data about data. In the information technology world the term is often used to indicate data that refers to digital resources available across a network.

Metadata Framework: A series of guidelines describing potential components required to define, describe, publish and maintain metadata.

Resource: Asset or means that fulfills a requirement.

Archival Issues Related to Digital Creations

Mark Kieler
Carnegie Mellon University, USA

Michael J. West
Carnegie Mellon University, USA

INTRODUCTION

The authors define "intellectual creations" as human expressions embodied in text, music, or other forms of art. Increasingly, we encode these creations in digital formats that have extremely short life cycles. Eventually, backward compatibility is lost. Thus, after very little time, a digital encoding format becomes obsolete, and intellectual works encoded in the format may become irretrievable. In contrast, the cultural worth of an intellectual creation may not be realized for generations. Additionally, future generations must access artifacts, including intellectual creations, to understand a culture in historical context. We contend that technology – intensive storage and manipulation of data may result in an inability to gain this access. Technology creators have some responsibility to facilitate future retrieval through careful documentation, and by selective maintenance of hardware that may be required to access archival media.

BACKGROUND

Cultural artifacts nearly always outlive the technologies that made them possible, which is particularly obvious with digital technology. Imagine the discovery hundreds of years from now of a floppy diskette containing a document written in WordStar®. Once a code has been lost, have all texts written in that code been lost as well?

At some point, supporting a technology is no longer economically viable if it is abandoned by enough people. It ceases to exist except in archival form. Currently, home videocassettes are being replaced by the DVD (Digital Video Disc), which subsequently will also be superceded.

Linguists stress the organic nature of language, which it is in constant evolution. Grammars are codifications constructed of "snapshot images" of spoken and written language usage. Despite the fact that both are considered examples of English, speakers of the vernacular of Shakespeare and the vernacular of rap music would find each other incomprehensible. With technology, as with language and culture, timing is everything, and context plays an important role in shaping how we interact with each other and with technology.

HOW DO WE KNOW ABOUT PAST CULTURES?

In engineering, Shannon (1948) described elements of a communications system: An information source generate data. A transmitter encodes information to travel over a channel through distance and time. At the other end, a receiver decodes the signal for the destination, which is the person or thing for which the message is intended. Since we are worried about humans, we consider them to be the source and destination. While parts of the transmission system have changed greatly over time, one could certainly argue that technology has caused the other portions of the system to change more rapidly than the human elements. Linguists, beginning with Ferdinand de Saussure, established semiotics as a science of signs, which likewise focuses on the sender and the receiver of a message. Semiotics also posited the arbitrary nature of the sign and looked at the ways human languages encode and transmit messages. (Saussure, 1974).

Physical artifacts that survive from the distant past reveal much about a culture, depending on their purpose and the quality of materials from which they are made. Our record from ancient civilizations is far from perfect, but archaeologists can construct some details about them from these clues.

Significant events have often been recorded in the living embodiment of a storyteller, and oral traditions still form an important part of many cultures. The historical record has often been related by language, as well as by performance (e.g., a ritual dance). Some of these oral histories survived long enough to be recorded in other more permanent media. However, not only have many oral traditions died with the last generation of storyteller, others have assumed inaccuracies and exaggerations as a result of being passed serially through generations.

As languages evolved and became standardized, it became possible to encode events in written form. Because writing has traditionally been the province of the learned few, written documents were recorded on long-lived media, and special care was accorded to their storage. Fortunately, many ancient documents have survived, albeit, with significant degradation. Given the constant change in a living language, when a culture dies, the language often dies with it. Language experts attempt to reconstruct meaning by looking for patterns that may establish types of words and contexts and similarities to more modern languages.

In addition to printed text, human expression is also accomplished through artistic means such as music, painting, and dance. While we only have been able to preserve musical and dance performances for a relatively recent portion of human history, for centuries we have had a written procedure for recreating music, using notes, measures, and time definitions. The methods for recording dance instructions are much less standardized and rely on interpretations and tradition.

Art works degrade over time, depending on the types of inks, paints, or dyes used, the media on which they are deposited, and overall interaction with the environment. Even sculpture is subject to environmental degradation, such as damage from acid rain.

THE INTRODUCTION OF TECHNOLOGY

The invention of the printing press made wide distribution of printed information possible, and wood pulp-based paper made it affordable for the general public. However, unlike expensive fiber-based paper, pulp-based paper has usually been manufactured through an acid-based process, and residual acid in the paper eventually destroys it. Thus, paradoxically, fiber-based books from the 19th century are often more legible than their 20th century wood pulp counterparts.

Text

Text was the first means of expression to be converted into electrical form. In fact, text went "direct to digital." Morse code is a duration-encoded digital signal, unrecognizable to anyone who does not understand it. Initially, storage was primitive, used mainly to "buffer" the information until a human could reconvert it to text. Thus, long-term storage and archiving of Morse code traffic was not an issue.

The first modern bit encoding of text occurred in 1874. Emile Baudot, a French telegraph engineer, devised a 5-bit code for each letter of the alphabet. Unlike Morse code, each symbol had a fixed length representation, dependent only on the presence or absence of electrical current. The Baudot code was durable, used by news service teletypewriters throughout the 1970s (Freed, 1995). Crude paper tape punches were often utilized for storage. The digital code could be read, albeit slowly, merely by holding the tape up to a light.

ASCII (American Standard Code for Information Interchange) uses 7 bits. The added bits allowed upper and lower case letters, as well as numbers, punctuation, and other special characters. It endures as the "plain text" standard.

The rise of WYSIWYG (what you see is what you get) computer interfaces, and the availability of sophisticated word processing programs, made it possible to digitally encode additional expressions to text. Different art styles of text (fonts) could be used, and these could embody visual variations such as italics, bold, superscripts and subscripts, and underlines. Word processing evolved to encode these variations. This was accomplished by adding bits to the original text data, or by software commands that controlled a section of ASCII text. These techniques represent a deviation from international standards into conventions of the word processing software. As the level of sophistication increases, we become increasingly unable to understand an encoded section of text without the software used to create it. In fact, the actual text becomes only a tiny portion of the code. If the page is graphically encoded, as occurs in programs such as Adobe Acrobat®, then plain text representations are lost all together (Kieler et al., 2004).

Audio and Visual Technology

Storage of audio and visual data began in the 19th century and progressed in sophistication throughout most of the 20th century. While not strictly correct in each instance, the general paths of progress could be viewed as follows:

- Visual Still and Audio-Only, to Visual Motion, to Visual Motion with Audio
- Mechanical or Chemical-Based Recording, and Storage to Electronic Recording and Storage
- Mechanical Reproduction to Electronic Reproduction
- Analog Encoding to Digital Encoding

The most significant impediments to easy retrieval of an intellectual creation involve electronic encoding and storage, and digital conversion.

Electronic encoding of visual information marks the historical point where the original signal cannot be recovered merely by converting stored data directly into light or motion. The visual image is stored as a series of lines, and voltage references signal new lines or new collections of lines (*frames*). Thus, one must be able to properly decode the various electronic levels and pulses unique to the method, and understand how the lines are ordered and encoded in order to correctly reconstruct the scanned image. Incompatible formats include PAL (Phase Alternating Line), NTSC (National Television Standards Committee) and SECAM (Systéme Electronique Couleur Avec Mémorie) (Abbott, 1994).

With the advent of magnetic storage, one could no longer directly "see" the information on the storage medium without a great deal of technological aid. As an example, by the late 1920s, light modulation of audio onto motion picture film was possible (Hochheiser, 1992), and both audio and video information are clearly visible on the film. In contrast, the advent of practical magnetic video recording in 1956 (Shima, 1984) resulted in a signal which can only be observed using microscope techniques (Chinn, 1957).

Finally, digital data types in streams of 1s and 0s are indistinguishable without the reverse algorithm to retrieve the original information. The significance is that in addition to the physical details of how the data was encoded, we also need mathematical processing information on how the data was changed after its conversion. Data protection techniques, including public and private key cryptography, hashing algorithms, and other methods (Wayner, 1997), added on top of the conversion technique, have deliberate obscuration of data as their intent. Thus, the data is not readable, even when the encoding techniques are known, unless the proper decrypting information is available.

Lossy Encoding

An additional dilemma is that digital conversions result in massive amounts of data. Compact disc audio, for instance, necessitates a bit rate of 4.32 million bits per second (Carrasso et al., 1982). Efficient storage and transmission virtually require data reduction. We must actually throw away some of the original information (hence the term *lossy*) to make any real progress in storage efficiency. We discard information based on redundancies in the retained data, or upon deeming it insignificant through models of human perceptual limitations. For instance, Moving Picture Experts Group Layer 3 audio (MPEG 3 or MP3) achieves a better than 10 to 1 savings (Jack, 1996). However, the encoded audio bit stream now represents a set of mathematical manipulations on top of the interleaving, control, data, and error correction algorithms of the storage format. It has become impossible to reconstruct the data without complete knowledge of each of these transformations.

IMPACT ON LONG-TERM DATA AVAILABILITY

If intellectual creations are stored in formats that require an exact knowledge of the encoding transforms, the issue for society is whether any of these techniques will be available, or even remembered, in the distant future.

In the short term, there is no problem, as new formats usually incorporate backward compatibility and conversion. Eventually, however, the overhead needed to retain compatibility hampers the effectiveness of a system. As demand collapses, the resources that support an obsolete format must be utilized for other purposes.

The digital revolution has compressed the time scale. Tremendous gains in hardware and software capability have resulted in unprecedented rates of obsolescence. For example, in both popular personal computing formats, operating systems have migrated to 32-bit code, abandoning the previous 16-bit systems. In doing so, they have shed much backward compatibility with the previous systems (White, 2002.) This transition has occurred in a mere 20 years. Thus, the ability to retrieve data created with a 16-bit program may well be lost once the last 16-bit compatible system is "upgraded."

New generations of artists and creative personalities use electronic and especially digital formats as freely as their predecessors used physical media. While most popular culture intellectual creations have little lasting value, a few will be significant enough to be studied and preserved far into the future. The problem is that human society does not always immediately recognize an intellectual work's significance, and it may fall into obscurity for decades or even centuries. A cursory encyclopedia search (Compton's, 1994) reveals examples from many fields. For example, painter Sandro Botticelli's works remained obscure from the late 1400s until the 19th century.

While not individually significant, ordinary popular culture creations collectively can reveal a great deal about a period in society. However objective evaluation often requires the passage of at least several decades. Even then, critiques and judgments may go through many modifications and revisions before any consensus is reached.

MINDSETS: CREATORS OF TECHNOLOGY VERSUS CREATORS OF IDEAS

Technology creators derive value from innovation. Profits and employment fall once a particular technology generation "matures" and a "shakeout" era ensues, where only the most efficient producers prevail. Software and digital data can be infinitely replicated with no degradation. Thus, barring legal impediments, there is virtually no value in creating additional copies. Even with legal protections against duplication, others may create software that operates differently but accomplishes the same goal. These characteristics shape the mindset of technology creators into a "cult of the new." Obsolescence is the fuel for continued employment and profitability.

With intellectual creations, there is also a progress-oriented motivation. Artists do not merely rest on past achievements. However, since these creations embody a combination of knowledge, ideas, and interpretation, their value is not necessarily superceded by later works. While only a few will have lasting value to society, there is no way to know when this value will finally be appreciated. The finest intellectual works must prove their value repeatedly through decades and centuries of cultural change.

The issue is that these works are increasingly created through a technological process that need only prove itself until the next "better" innovation comes along. The potentially tremendous but unpredictable future value of key intellectual works may never be realized if they are embodied in a technical process that has zero value five years later. While future societal worth is of little present economic value to the creation of intellectual work, the encoding technology creator derives nothing beyond the initial sale of the software.

Certainly, the same could be said of those who produce physical creative media such as paint or canvas. However, the medium usually survives regardless of the intentions of its original manufacturer. Deterioration is directly observable, and can often be repaired to foster preservation. With modern technology, however, the "medium" is a combination of the physical storage object and the encoding software/hardware systems. Even if the storage object can last centuries, the ability to retrieve it is governed by entities that have, if anything, a *dis*incentive to ensure continued long-term accessibility. This leaves society with no way to rediscover and realize the benefits of important intellectual creations in the distant future.

FUTURE TRENDS

While it is unfair to lay all responsibility for long-term accessibility on technology creators, it is clear that the nature of digital innovation fosters the problem. Among engineering ethicists, there is general agreement that those who create technology have a unique insight into potential side effects. Harris, Pritchard, and Rabins (1992) discuss the need for "preventive ethics" in engineering. Shinzinger and Martin (2000) describe engineering as social experimentation. This implies a responsibility for technology creators to envision possible social consequences and, at the very least, take steps to inform others of their existence and nature. This is especially true where the complexity of the technology is transparent to the end-user.

Thus, while digital technology creators may not solely be responsible for ensuring future accessibility of works expressed through their technology, one could argue that they have a responsibility to enter into dialogues about preserving accessibility, and to design their encoding technology to foster preservation.

Currently, preservation awareness is coming, not from the technology developers, but rather from the software archivists. Pennavaria (2003) details a number of concerns and efforts of groups such as the American Library Association and the American Film Institute. Although technologists are involved in these efforts, they tend to deal only with existing analog and digital formats. Also, they tend to concentrate largely on the deterioration mechanisms of the physical media, as opposed to preserving the availability of playback hardware and decoding software.

Some hardware, such as silicon chips, could last for centuries. Conversely, other pieces of machinery clearly have a limited shelf life due to physical deterioration and wear. It is the "soft" items, such as encoding and decoding software and the underlying algorithms and manipulations, that have the greatest chance to be lost, owing to their short lifetimes. Only the creators of these artifacts can take steps to ensure their survival. This implies a responsibility that the technologists cannot ignore. The broader culture must also realize the problem and work with the technologists to foster preservation. Otherwise, much of our current culture will disappear in less than a century.

CONCLUSION

The attractiveness and availability of digital processing has resulted in its use by creators of intellectual content. The rapid turnover of digital applications may make access to these creations impossible within a very short period, compared to the time it takes for society to determine their ultimate worth. Technology creators must work with society leaders to ensure that intellectual creations are not lost forever.

REFERENCES

Abbott, P. (1994, November-December). Video formats and resolution. *Nuclear Plant Journal*, 39-46.

Carasso, M., Peek, J., & Sinjou, J. (1982). The compact disc digital audio system. *Philips Technical Review, 40*(6), 151-156.

Chinn, H. (1957, August). Splicing video tape. *Industries & Tele-Tech*, 79.

Compton's interactive encyclopedia 1995. (1995) . *[CD-ROM*. d.). SoftKey International, Inc.

Freed, L. (1995). *The history of computers*. Emeryville, CA: Ziff-Davis Press.

Harris, Jr., C., Pritchard, M., & Rabins, M. (2000). *Engineering ethics: Concepts and cases* (2nd Ed.). Belmont, CA: Wadsworth.

Hochheiser, S. (1992). What makes the picture talk: AT&T and the development of sound motion picture technology. *IEEE Transactions on Education, 35*(4), 278-285.

Jack, K. (1996). *Video demystified: A handbook for the digital engineer* (2nd ed.). San Diego, CA: HighText Publications.

Kieler, M., & West, M. (2004). Digital orphans: Technology's wayward children. In Linda L. Brennan & Victoria E. Johnson (Eds.), *Social, ethical and policy implications of information technology* pp. 234-250). Hershey, PA: Idea Group Publishing.

Pennavaria, K. (2003). Nonprint media preservation: A guide to resources on the Web. *College Resources and Library News, 64*(8).

Saussure, F. (1974). *Course in general linguistics*. London: Fontana.

Schinzinger, R., & Martin, M. (2000). *Introduction to engineering ethics*. New York: McGraw-Hill.

Shannon, C.E. (1948, July, October). A mathematical theory of communications. *Bell Systems Technical Journal*, 379-423 (July) and 623-656 (October).

Shima, S. (1984). The evolution of consumer vtr's – Technological milestones. *IEEE Transactions on Consumer Electronics, CE-30*(2), 66-69.

Wayner, P. (1997). *Digital copyright protection*. Chestnut Hill, MA: Academic Press.

White, R. (2002). *How computers work* (6th Ed.). Indianapolis, Indiana: Que.

KEY TERMS

Analog: Encoding a physical phenomenon by a direct, perceptually continuous variation of a physical property such as electromagnetic intensity (recording tape), mechanical displacement (Vinyl disk), or opaqueness (photographic film).

ASCII: American Standard Code for Information Interchange. A standard method of encoding upper and lower case text and other symbols with a 7-bit code.

Baudot: A 5-bit standard encoding method for uppercase letters, Invented by Emilie Baudot in 1874.

Intellectual Creation: Any work of creation, such as authorship, visual arts, performing arts, or music.

Lossy Encoding: Removal of data that represents redundant information, or differences presumed imperceptible to humans, in order to reduce stored or transmitted quantities of digital data.

Masking: In psychoacoustic models, the ability of a loud tone to block the perception of sounds or noise occurring in nearby frequencies.

Psychoacoustic Models: Models of human aural perception, especially with regard to the ability or inability to perceive signals that are masked by other signals.

Semiotics: A theory of signs and the use of signs in languages. Semiotics posits the arbitrary nature of the sign and looks at how human languages encode and transmit messages.

Wood Pulp: An inexpensive paper stock. Acid, used as part of the production process, frequently remains in the paper and causes its destruction. This destruction is rapid compared to paper made from fiber.

WYSIWYG: "What you See is What you Get." A user display that shows text and graphical information exactly as it will appear in print or other subsequent distribution. This includes expressional and artistic variations of text fonts, such as italics, bold, and underlined text.

Artificial Intelligence Techniques in Medicine and Health Care

Rezaul Begg
Victoria University, Melbourne, Australia

INTRODUCTION

Now-a-days, researchers are increasingly looking into new and innovative techniques with the help of information technology to overcome the rapid surge in health care costs facing the community. Research undertaken in the past has shown that artificial intelligence (AI) tools and techniques can aid in the diagnosis of disease states and assessment of treatment outcomes. This has been demonstrated in a number of areas, including: help with medical decision support system, classification of heart disease from electrocardiogram (ECG) waveforms, identification of epileptic seizure from electroencephalogram (EEG) signals, ophthalmology to detect glaucoma disease, abnormality in movement pattern (gait) recognition for rehabilitation and potential falls risk minimization, assisting functional electrical stimulation (FES) control in rehabilitation setting of spinal cord injured patients, and clustering of medical images (Begg et al., 2003; Garrett et al., 2003; Masulli et al., 1998; Papadourokis et al., 1998; Silva & Silva, 1998).

Recent developments in information technology and AI tools, particularly in neural networks, fuzzy logic and support vector machines, have provided the necessary support to develop highly efficient automated diagnostic systems. Despite plenty of future challenges, these new advances in AI tools hold much promise for future developments in AI-based approaches in solving medical and health-related problems. This article is organized as follows: Following an overview of major AI techniques, a brief review of some of the applications of AI in health care is provided. Future challenges and directions in automated diagnostics are discussed in the summary and conclusion sections.

ARTIFICIAL INTELLIGENCE TECHNIQUES

There have been a number of artificial intelligence (AI) tools developed over the past decade or so (cf., Haykin, 1999; Keckman, 2002). Many of these have found their applications in medical and health-related areas. Commonly applied AI techniques can be listed as:

- Neural Networks
- Fuzzy Logic
- Support Vector Machines
- Genetic Algorithms
- Hybrid Systems

In the following, we give a brief overview of neural networks, fuzzy logic and the relatively new support vector machines.

Neural Networks

Artificial neural networks work much like the human brain and have the ability to learn from training data and store knowledge in the network. In the learning phase, it maps relation between inputs and the corresponding expected outputs. During the learning phase, knowledge is acquired and stored in the network in the form of synaptic weights and biases. This knowledge is used to make future predictions in response to new data or inputs during the testing phase. Usually, the network has one input and one output layer, and one or more hidden layers depending on the complexity of the problem. Learning can be supervised; that is, the network is provided with both the inputs and their desired outputs during the leaning process, or it can be unsupervised or self-organizing learning. There are a number of learning algorithms available (Haykin, 1999), and among them back-propagation learning algorithm is the most widely used. In this method, an error signal based on the difference between network-generated output (g_i) and desired output (d_i) is propagated in the backward direction to adjust the synaptic weights according to the error signal. During the learning process, the aim is to minimize an objective function such as the mean-squared error (E),

$$E = \frac{1}{n}\sum_{i=1}^{n}(d_i - g_i)^2$$

Neural networks are frequently used as diagnostics, and therefore it is important to have good generalization ability, that is, good performance in predicting results in response to unseen data. One limitation of neural networks is the possibility of being stuck in local minima during training rather than converging to the global minimum. To overcome this the network is usually trained several times with random initial weights to avoid converging to the local minima. Neural networks have found the majority of their applications in pattern recognition, time-series prediction, signal processing and financial forecasting.

Fuzzy Logic

Fuzzy sets were introduced by Zadeh (1965), and they deal with imprecise and uncertain information or data. Naturally, this has been found suitable for many medical and health-related problems, as it relates to the way humans think. Since the early work of Zadeh, there has been an exponential rise in the number of scientific papers applying fuzzy sets in biology, medicine and psychology areas (Teodorescu et al., 1998).

Support Vector Machines

Support vector machines are a relatively new machine learning tool and have emerged as a powerful technique for learning from data and solving classification and regression problems. This has been particularly effective for binary classification applications. SVMs originate from Vapnik's statistical learning theory (Vapnik, 1995). SVMs perform by nonlinearly mapping the input data into a high dimensional feature space (by means of a kernel function) and then constructing a linear optimal separating hyperplane by maximizing the margin between the two classes in the feature space.

For m training data with input-output pairs $(y_1,\boldsymbol{x}_1),\ldots,(y_m,\boldsymbol{x}_m)$ where each input data $\boldsymbol{x}_i \in \Re^N$ belongs to a class $y_i \in \{-1,+1\}_{i=1,\ldots,m}$, the decision function for a new data $(\boldsymbol{x}_i)$ can be given by the sign of the following function (Gunn, 1998):

$$\mathrm{f}(x) = sign\left(\sum_{i \in SVs} \alpha_i y_i \mathrm{K}(\chi_i, \chi) + \mathrm{b}\right)$$

where, a_i is a nonnegative Lagrange multiplier corresponding to $\boldsymbol{x}_i$, $K(.)$ is a kernel function and b is the bias. The Lagrange multipliers are obtained as the solution of a convex quadratic programming problem. The data points $\mathbf{x}_i$s corresponding to $a_i>0$ are called support vectors. Such $\boldsymbol{x}_i$s are the only data points in the training set relevant to classification since the decision surface is expressed in terms of these points alone (support vectors, SV). For linearly separable problems, the number of SVs and the hyperplane are determined by a subset of the training set only. For nonlinearly separable problems, a_i in SVs are constrained by an upper bound C, which is regarded as a regularization parameter. This parameter makes a trade-off between margin maximization and minimization of classification errors in the training data set (Gunn, 1998).

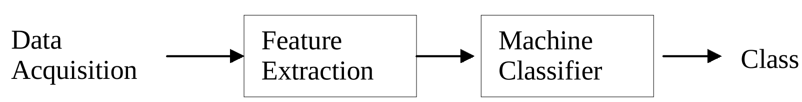
Hybrid Systems

Recently, researchers have started looking into ways of combining various AI tools in order to maximize performance of the AI system. The main idea behind this is to offset limitation of one system by cascading with another AI tool. As a result, hybrid systems like Neuro-Fuzzy (neural networks and fuzzy logic), Neuro–SVM (neural networks and support vector machines) and Fuzzy-SVM (fuzzy logic and support vector machines) systems have evolved. Hybrid systems have been applied in many applications, including some biomedical areas (Teodorescu et al., 1998).

APPLICATIONS IN HEALTH CARE AND MEDICINE

In addition to applications in medical diagnostic systems, AI techniques have been applied in many biomedical signal-processing tasks, including analysis of ECG, EEG and human movement data (Nazeran & Behbehani, 2001). Neural network models have played a dominant role in a majority of these AI-related applications in health and medicine. Many of these applications are for pattern recognition or classification. A typical classification application usually has a number of steps or procedures as shown by the flow diagram (see Figure 1). This involves feature extraction from the input data before feeding these features to the classifier for designing and developing

Figure 1. Stages of a typical pattern recognition task



automated classification models, and finally testing the models for generalization.

Medical Decision Support Systems

Medical decision support systems (MDSS) are designed to construct a knowledge database by way of receiving a list of symptoms as input features and their corresponding disease type(s) as the output. Such a developed symptom-to-disease mapping system then facilitates the diagnostic process by generating new responses due to a new set of symptoms. Neural networks have been used to aid MDSS. Silva and Silva (1998) developed such a neural network-based MDSS system for a relatively small set of 20 randomly selected diseases and reported encouraging results. Disease symptoms in this study were represented as sinusoidal damped waveforms. Hybridization has been shown to improve diagnostic accuracy. For example, Dumitrache et al. (1998) reported an improvement in accuracy (by 28%) for medical diagnosis in cardiac disease using a hybrid decision-making system compared to a classical expert system.

Cardiology

Several studies have applied neural networks in the diagnosis of cardiovascular disease, primarily in the detection and classification of at-risk people from their ECG waveforms (Nazeran & Behbehani, 2001). Celler and Chazal (1998) have applied neural networks to classify normal and abnormal (pathological) ECG waveforms: 500 ECG recordings (155 normal and 345 abnormal) were used to extract features from the QRS complex for training and testing the classifier. The abnormal ECG recordings had six different disease conditions. The classifier was able to recognize these waveforms with 70.9% accuracy.

Electroencephalography

AI tools, including neural networks, fuzzy clustering and SVMs, have been shown to be useful for analyzing electrical activity of the brain, the electroencephalogram (EEG) signals. Features extracted from EEG recordings of the brain have been used with AI tools for improving communication between humans and computers and also for effective diagnosis of brain states and epileptic seizures (Garrett et al., 2003; Geva & Kerem, 1998; Nazeran & Behbehani, 2001).

Opthalmology

Neural networks have been shown to be an effective diagnostic tool to identify glaucoma disease. Glaucoma is more prevalent in older age and can cause loss of vision. Papadourokis et al. (1998) applied backpropagation neural network to classify normal patients and patients with glaucomatic optic nerve damage from perimeter examination. Several neural network models were tested using 715 cases, including 518 glaucoma cases, and they reported 90% recognition accuracy with two hidden layer networks and training with 80% of the input data. In an effort to compare effectiveness of different AI techniques in recognizing glaucoma diagnosis, Chan et al. (2002) used standard automated perimetry data to compare classification performance of several classifiers including multiplayer perceptron and support vector machines (SVM). In-depth analysis of performance of these classifiers was carried out using areas under the receiver operating characteristic (ROC) curves and also sensitivity (true positive rates) and specificity (1 - false positive rates) measures. Machine classifiers were found to perform superiorly in the classification tasks, whereas SVM showed significantly improved performance compared to a multiplayer perceptron. A self-organizing fuzzy structure has also been developed and applied to predict the onset of hypoglycemia for diabetic patients (Hastings et al., 1998).

Gait Analysis and Rehabilitation

Gait is the systematic analysis of human walking. Various instrumentations are available to analyze different aspects of gait (cf. Begg et al., 1989). Among its many applications, gait analysis is being increasingly used to diagnose abnormality in lower limb functions, and also to assess the progress of improvement as a result of treatments and interventions. Recently, neural networks and fuzzy logic techniques have been applied for gait pattern recognition and clustering gait types. Barton and Lees (1997) classified gait patterns based on hip-knee angle diagrams and Holzreiter and Kohle (1993) used neural network models to identify normal and pathological gait patterns from measurements of foot-to-ground reaction forces using force platforms. Wu et al. (1998) applied a back propagation neural network to classify gait patterns of patients with ankle arthrodesis and normal subjects, and reported a superior classification by the neural networks compared to statistical technique (linear discriminant analysis) (98.7% vs. 91.5%). Gait analysis is being increasingly used in rehabilitation settings, and also combining with AI techniques to improve gait control and functionality. Tong and Grant (1998) applied neural networks to optimize sensor sets for control of FES system in people with spinal cord injury and showed improvements in accuracy as a result of neural network aided control. Fuzzy logic has also been recently applied with great success in: clustering children gait with and without neurological disorder (O'Malley, 1997) and also detection of gait events such as the foot contact and take-off during walking in the analysis of paraplegic gait (Skelly & Chizeck, 2001).

Support vector machine (SVM) has recently been applied to classify young and elderly gait patterns (Begg et al., 2003). Gait changes with aging, with potential risks of loss of balance and falls. Recognizing gait patterns with potential falls risks would help to detect at-risk people so that rehabilitation programs could be undertaken to minimize the risk of falls. AI techniques such as neural networks and SVMs have demonstrated their potentials for detecting gait degeneration due to aging and appear to have potential applications in falls prevention in the elderly.

SUMMARY AND FUTURE TRENDS

There are plenty of future challenges for AI to be routinely used in medicine and health. The use of automated medical decision support system in routine use in the clinic would make a significant impact on our health care system.

One possibility is that health care in the future will be built on knowledge networks (Erlandson & Holmer, 2003). Applications of telemedicine and informatics in health care can help to provide support to patients in remote areas, and to share expert knowledge or limited resources. Furthermore, effective networking between informatics and biomedical engineering can also help to complement each other's knowledge and to fight in partnership the various challenges faced by the medical and health care systems.

One major aim of a classification task is to improve its recognition accuracy or generalization performance. Selecting features that offer the most discriminatory information between the classes or medical groups could help to improve classification accuracy. At the same time, removing redundant features might also improve the medical recognition task. It has been demonstrated in a number of studies that selecting a small number of good features in fact improves the recognition accuracy (Begg et al., 2003; Chan et al., 2002; Yom-Tov & Inbar, 2002), in addition to the added advantages of classifier simplicity and fast processing time. Therefore, a pattern recognition task may be improved by decomposing it into the following stages (see Figure 2):

Recently hybrid systems are emerging, combining various AI tools with improved performance in medical diagnosis and rehabilitation. Application of hybrid systems in medical diagnosis has already provided increased efficiency in diagnosis of cardiac disease by as much as 28% compared to classical expert system. This and other applications provide much hope and encouragement for more future applications based on hybridization of AI systems in helping to solve problems in medicine and health care.

CONCLUSION

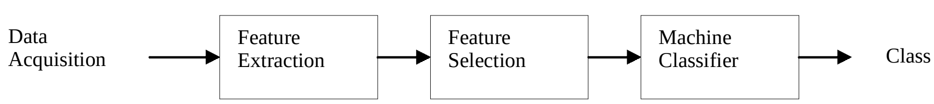
Artificial intelligence, particularly neural networks, fuzzy logic and the recently introduced support vector machines, played a key role over the years for many important developments in medicine and health care. Despite such developments there are many future challenges and currently only a few AI-based systems are routinely used in the clinical setting. Continued developments in AI fields are providing much impetus that is needed to tackle the many problems of the health care system.

REFERENCES

Barton, J.G., & Lees, A. (1997). An application of neural networks for distinguishing gait patterns on the basis of hip-knee joint angle diagrams. *Gait & Posture, 5*, 28-33.

Begg, R.K., Kamruzzaman, J., & Zayegh, A. (2003). An application of support vector machines for recognizing young-old gait patterns. *Proceedings of the World Congress on Medical Physics and Biomedical Engineering (WC2003*), Sydney, Australia.

Figure 2. Stages of a pattern recognition task involving "feature selection" sub-task



Begg, R.K., Wytch, R., & Major, R.E. (1989). Instrumentation used in clinical gait analyses: A review. *Journal of Medical Engineering & Technology, 5*, 115-121.

Celler, B.G., & Chazal, P. (1998). Low computational cost classifiers for ECG diagnosis using neural networks. *Proceedings of the International Conference of Engineering in Medicine & Biology Society (EMBC 1998)* (pp. 1337-1340).

Chan, K., Lee, T.W., Sample, P.A., Goldbaum, M.H., Weinreb, R.N., & Sejnowski, T.J. (2002). Comparison of machine learning and traditional classifiers in glaucoma diagnosis. *IEEE Transaction on Biomedical Engineering, 49*(9), 963-974.

Dumitrache, I., Mihu, I.R., & Rosu, K. (1998). Hybrid decision making system for medical diagnosis. In E.C. Ifeachor, A. Sperduti & A. Starita (Eds.), *Neural networks and expert systems in medicine and health care* (pp. 279-285). World Scientific Publishing.

Erlandson, B.E., & Holmer, N.G. (2003). Collaborative multimedia medical technologies in the management of health care. *World Congress on Medical Physics and Biomedical Engineering (WC2003*), Sydney, Australia.

Garrett, D., Peterson, D.A., Anderson, C.W., & Thaur, M.H. (2003). Comparison of linear, non-linear and feature selection methods for EEG signal classification. *IEEE Transactions on Neural Systems & Rehabilitation Engineering, 11*, 141-147.

Geva & Kerem. (1998). Brain state identification and forecasting of acute pathology using unsupervised fuzzy clustering of EEG temporal patterns. In T. Teodorrescu, A. Kandel & L.C. Jain (Eds.), *Fuzzy and neuro-fuzzy systems in medicine* (pp. 57-93). CRC Press.

Gunn, S. (1998). *Support vector machines for classification and regression.* ISIS Technical Report. University of Southampton, UK.

Hastings, G, Ghevondian, N., & Nguyen, H. (1998). A self-organizing fuzzy estimator for hypoglycaemia monitoring in diabetic patients. *Proceedings of the International Conference of Engineering in Medicine & Biology Society (EMBC 1998)* (pp. 1371-1374).

Haykin, S. (1999). *Neural networks: A comprehensive foundation* (2nd ed). Englewood Cliffs, NJ: Prentice Hall.

Holzreiter, S.H., & Kohle, M.E. (1993). Assessment of gait pattern using neural networks. *Journal of Biomechanics, 26*, 645-651.

Kecman, V. (2002). *Learning and soft computing: support vector machines, neural networks and fuzzy logic models.* NJ: IEEE MIT Press.

Masulli, F., Schenone, A., & Massone, A.M. (1998). Application of the probabilistic clustering to the segmentation of multimodal medical images. In E.C. Ifeachor, A. Sperduti & A. Starita (Eds.), *Neural networks and expert systems in medicine and health care* (pp. 241-249). World Scientific Publishing.

Nazeran, H., & Behbehani, K. (2001). Neural networks in processing and analysis of biomedical signals. In M. Akay (Ed.), *Nonlinear biomedical signal processing: Fuzzy logic, neural networks and new algorithms* (pp. 69-97).

O'Malley, M.J., Abel, M.F., Damiano, D.L., & Vaughan, C.L. (1997). Fuzzy clustering of children with cerebral palsy based on temporal-distance gait parameters. *IEEE Transactions on Rehabilitation Engineering, 5*, 300-309.

Papadourakis, G.M., Gross, H.G., & Alexakis, I. (1998). Detection of glaucomatic nerve damage using artificial neural networks. In E.C. Ifeachor, A. Sperduti & A. Starita (Eds.), *Neural networks and expert systems in medicine and health care* (pp. 140-150). World Scientific Publishing.

Silva, R., & Silva, A.C.R. (1998). Medical diagnosis as a neural networks pattern classification problem. In E.C. Ifeachor, A. Sperduti & A. Starita (Eds.), *Neural networks and expert systems in medicine and health care* (pp. 25-33). World Scientific Publishing.

Skelly, M.M., & Chizeck, H.J. (2001). Real-time gait event detection for paraplegic FES walking. *IEEE Transactions on Neural Systems & Rehabilitation Engineering, 9*, 59-68.

Teodorrescu, T., Kandel, A., & Jain, L.C. (1998). *Fuzzy and neuro-fuzzy systems in medicine.* In Teodorrescu, Kandel & Jain (Eds.). CRC Press.

Vapnik, V.N. (1995). *The nature of statistical learning theory*. New York: Springer.

Wu, W.L., Su, F.C., & Chou, C.K. (1998). Potential of the back propagation neural networks in the assessment of gait patterns in ankle arthrodesis. In E.C. Ifeachor, A. Sperduti & A. Starita (Eds.), *Neural networks and expert systems in medicine and health care* (pp. 92-100). World Scientific Publishing.

Yom-Tov, E., & Inbar, G.F. (2002). Feature selection for the classification of movements from single movement-related potentials. *IEEE Transactions on Neural Systems & Rehabilitation Engineering, 10,* 170-177.

Zadeh, L. (1965). Fuzzy sets. *Information and Control, 8,* 338-353.

KEY TERMS

ECG: The electrical activity of the heart recorded from the body surface using electrodes as electrical potential is known as electrocardiogram or ECG.

EEG: The electrical activity of the brain recorded from the scalp as electrical potential is known as electroencephalogram or EEG.

Fuzzy Logic: The concept of fuzzy logic is that many classes in the natural environment are fuzzy rather than crisp. It deals with imprecise and uncertain data.

Gait Analysis: Analysis of human walking patterns. It is used to analyze abnormality in lower limb problems and assess treatment or intervention outcomes.

Hybrid Systems: Integration of two or more artificial intelligence tools to improve efficiency or system performance.

Neural Networks: Neural networks resemble the human brain and able to store knowledge during training and use this for decision making during testing phase.

Support Vector Machines: Introduced by Vapnik and capable of learning from data for solving classification and regression problems.

A

Artificial Neural Networks in Financial Trading

Bruce Vanstone
Bond University, Australia

Clarence Tan
Bond University, Australia

INTRODUCTION

Soft computing represents that area of computing adapted from the physical sciences. Artificial intelligence (AI) techniques within this realm attempt to solve problems by applying physical laws and processes. This style of computing is particularly tolerant of imprecision and uncertainty, making the approach attractive to those researching within "noisy" realms, where the signal-to-noise ratio is low. Soft computing is normally accepted to include the three key areas of fuzzy logic, artificial neural networks, and probabilistic reasoning (which includes genetic algorithms, chaos theory, etc.).

The arena of investment trading is one such field where there is an abundance of noisy data. It is in this area that traditional computing typically gives way to soft computing, as the rigid conditions applied by traditional computing cannot be met. This is particularly evident where the same sets of input conditions may appear to invoke different outcomes, or there is an abundance of missing or poor-quality data.

Artificial neural networks (ANNs) are a particularly promising branch on the tree of soft computing, as they possess the ability to determine nonlinear relationships and are particularly adept at dealing with noisy data sets.

From an investment point of view, ANNs are particularly attractive, as they offer the possibility of achieving higher investment returns for two distinct reasons. First, with the advent of cheaper computing power, many mathematical techniques have come to be in common use, effectively minimizing any advantage they had introduced (Samuel & Malakkal, 1990). Second, in order to attempt to address the first issue, many techniques have become more complex. There is a real risk that the signal-to-noise ratio associated with such techniques may be becoming lower, particularly in the area of pattern recognition (Blakey, 2002).

Investment and financial trading is normally divided into two major disciplines: fundamental analysis and technical analysis. Papers concerned with applying ANNs to these two disciplines are reviewed.

BACKGROUND

There are a number of approaches within the literature that deal with applying ANN techniques to investment and trading. Although there appears to be no formal segmentation of these different approaches, this review classifies the literature into the topics proposed by Tan (2001), and augments these classifications with one more category, namely, hybrid. These categories of ANN, then, are as follows:

- Time series—Forecast future data points using historical data sets. Research reviewed in this area generally attempts to predict the future values of some time series. Possible time series include base time series data (e.g., closing prices) or time series derived from base data (e.g., indicators, which are frequently used in technical analysis).
- Pattern recognition and classification—These are attempts to classify observations into categories, generally by learning patterns in data. Research reviewed in this area involved the detection of patterns and segregation of base data into "winner" and "loser" categories as well as in financial distress and bankruptcy prediction.
- Optimization—This involves solving problems where patterns in the data are not known, often nonpolynomial (NP)-complete problems. Research reviewed in this area covered the optimal selection of parameters and determining the optimal point at which to enter transactions.
- Hybrid—This category was used to distinguish research that attempted to exploit the synergy effect by combining more than one of the above styles.

There appears to be a wide acceptance of the benefit of the synergy effect, whereby the whole is seen as being greater than the sum of the individual parts. This can be easily seen by inspecting the following reserach, which clearly shows that Hybrid techniques account for about one third of the research reviewed.

Further, the bias in this style of research toward technical analysis techniques is also evident from the research, with one third of the research pursuing the area of pattern recognition and classification. Technical analysis particularly lends itself to this style of research, as a large focus of technical analysis concerns the detection of patterns in data and the examination of the behavior of market participants when these patterns are manifest.

MAIN THRUST OF THE CHAPTER

This section briefly considers the characteristics of each of the four main categories described above. The selected papers were chosen as they are representative of current research directions, represent an important change in direction for this style of research, or represent a novel approach.

Research into Time Series Prediction

The area of time series predictions is normally focused on attempting to predict the future values of a time series in one of two primary ways:

- Predicting future values of a series from the past values of that same series
- Predicting future values of a series using data from different series

Typically, current research in this area focuses on predicting returns, or some variable thought to correlate with returns (e.g., earnings). Some researchers focus on attempting to predict future direction of a series (e.g., increasing from last known value, decreasing from last known value, no change). Research of this nature is essentially a classification problem and is discussed in the relevant section.

The following papers were selected and reviewed as they are representative of the current research in time series prediction. The authors of those papers are Chan and Foo (1995), Quah and Srinivasan (2000), Yao and Poh (1995), Hobbs and Bourbakis (1995), Austin et al. (1997), and Falas et al. (1994). The papers reviewed consider both fundamental and technical data. For example, Falas et al. (1994) used ANNs to attempt to predict future earnings based on reported accounting variables. They found no significant benefit using ANNs compared to the logit model and concluded that the accounting variables chosen were not appropriate earnings predictors. This conclusion represents one of the major problems encountered when working with ANNs, namely, their nonexistent explanatory capability. It is common to find conclusions of this type when reviewing ANN research, with noncorrelation often being reported as wrongly chosen input variables. Quah and Srinivasan (2000) used mainly accounting variables to predict excess returns (with limited success). Chan and Foo (1995) used ANNs to predict future time series values of stock prices and used these "future" values to compute a variety of technical indicators. The ANN produced showed particularly promising results. The authors concluded that the networks' ability to predict allows the trader to enter a trade a day or two before it is signalled by regular technical indicators, and this accounts for the substantially increased profit potential of the network.

In many ways, these two primary prediction methodologies relate to technical analysis strategies. For example, the use (and projection) of a moving average over a series of stock prices could be regarded as predicting future values of a series (the moving average) from past values of the same series. Indicators in technical analysis are often composed of a number of constituent data items, like price, volume, open-interest, etc. These indicators are commonly used to give indications of the future direction of price.

Research into Pattern Recognition and Classification

Pattern recognition techniques and classification techniques have been grouped together, as their goal is normally not to predict future values of a time series, but to predict future direction of a time series. For example, the primary goal of a chartist (a style of technical analyst) is to attempt to predict trend turning points by studying chart price action, looking for certain patterns. Chartists have noticed that these patterns tend to recur and are reasonably reliable indicators of the future direction of price trends. There are a number of these chart patterns, and different analysts attach different weightings to the predictive power of any given pattern. Also, these patterns normally need to be confirmed by values from another time series (such as volume) to be considered reliable. For more detail on this area, the reader is encouraged to refer to Pring (1999). Nonpattern matching techniques that also attempt to predict the future direction of a time series are also classification problems. Often, in addition to predicting the future direction of a time series, classification research attempts to classify stocks into two main groups, namely, "winners" and "losers," as in bankruptcy and financial distress predictions.

The following papers, listed by author, were selected and reviewed as they are representative of the current research in pattern recognition and classification: Tan (1996), Tan and Dihardjo (2001), Mizuno et al. (1998), Baba et al. (2001), Skabar and Cloete (2001), Suh and LaBarre (1995), Fu et al. (2001), Kamijo and Tanigawa (1990), Baba and Handa (1995), and Baek and Cho (2000). As described above, the research can generally be classified as winner and loser detection, or pattern matching. The works of Tan (1996), and later, Tan and Dihardjo (2001), use the concept of winner and loser classification, as do Longo, Suh and LaBarre (1995) and Skabar and Cloete (2001). Specifically, Skabar and Cloete do not predict winners and losers but predict two disparate categories, namely, "up" and "down" (direction of returns). The work of Kamijo and Tanigawa (1990) provides an excellent example of pattern matching, with the authors building ANNs to identify "triangle" patterns in stock market data (the triangle is a specific pattern used in technical analysis).

Classification involving pattern matching could also be validly discussed under the section above on time series prediction, due to the fact that pattern constructs must occur in specific time order, and the majority of patterns are not time invariant. This leads to the desire of researchers to identify time invariant patterns or attempt to determine a fixed period of time in which a pattern should occur. The work of Fu et al. (2001) provides examples of using genetic algorithms to "fix" the length of patterns, making them suitable for study using ANNs.

Research into Optimization

The focus of optimization is directed toward research that uses soft computing specifically to attempt to optimize an otherwise accepted achievable result. Typical of this style of research paper, an already accepted result is discussed and then considered for optimization. The optimization is characteristically proven by excess returns compared to the unoptimized case.

For an example of this style of optimization using ANNs, an index arbitrage timing has been proposed by Chen et al. (2001). Their model attempts to optimise the correct entry point timing for index arbitrage positions. Current arbitrage models propose establishing an arbitrage position immediately as an opportunity arises; the neural network approach is to attempt to locate the timing when there will be a maximum basis spread for the arbitrage, thereby increasing profit potential. Their research concludes that the neural model significantly outperforms the traditional approach.

Research into Ensemble Approaches

Research is classified as an ensemble approach if it combines work from more than one of the areas described below, effectively attempting to leverage the synergy effect by achieving an end result greater than that expected from each of the individual constituents. Among soft-computing research, there is a growing trend toward using the ensemble approach to analysis.

The following papers (listed by author and date of publication) were selected and reviewed, as they are representative of the current research in ensembles: Liu and Lee (1997), Leigh et al. (2002), Baba et al. (2002), Jang et al. (1991), Wong and Lee (1993), Chenoweth et al. (1995), and Abdullah and Ganapathy (2000). The majority of the ensembles draws their components from a variety of soft-computing methods. The use of ANNs and genetic algorithms (GAs) together is popular, and is used by Leigh et al. to combine pattern recognition techniques with price forecasting. Another approach combining ANNs and GAs is provided by Baba et al., using ANNs for their predictive ability and GAs to determine the best way to react to that information. Some ensembles combine multiple ANNs, for example, Jang et al. combine two ANNs: one that takes a short-term view of market movement, and one that takes a longer-term view. They then build a model that reacts to a weighted output sum of the outputs of both models. Liu and Lee (1997) and Abdullah and Ganapathy (2000) also used ensembles of ANNs and concluded that the predictive ability of the ensemble approaches exceeded that of the individual ANNs. Other research reviewed combined ANNs with fuzzy logic and expert systems.

FUTURE TRENDS

Essentially, the field of financial trading is in a state of transition between traditional pricing models, the efficient market hypothesis, and ideas about behavioral finance. The challenge that presents itself is how best to unify financial trading pricing models. There is much debate about the validity of the efficient market hypothesis, which effectively contends that prices cannot be predicted using methods such as technical analysis. There is a large body of evidence that appears to contradict the efficient market hypothesis, and there seems little chance of academically moving forward en masse unless an alternative valid pricing model exists. This offers substantial opportunity for soft-computing research techniques, particularly neural models. These models are capable of acting as universal approximators and determining complex nonlinear relationships. The goal with these methods is to attempt to mine deep relationships that can shed new light about the behaviour of markets and prices. These new relationships would inherently provide more scope for developing feasible and effective pricing models.

CONCLUSION

This chapter has surveyed recent and key literature in the domain of applying ANNs to investment and financial trading. Within the context of investment discipline, this survey shows that the majority of this type of research is being conducted in the field of technical analysis. As discussed earlier, soft computing is particularly data intensive, and it is suggested that this observation goes some way to explaining this obvious bias in research.

Within the area of soft-computing styles, the survey finds that the majority of research is within the area of both hybrid systems and pattern recognition and classification. It is suggested that the reason for this is that the technical analysis approach lends itself to the pattern recognition and classification areas. Also, many hybrid systems include pattern recognition and classification as one of their constituents.

REFERENCES

Abdullah, M. H. L. B., & Ganapathy, V. (2000). Neural network ensemble for financial trend prediction. *Tencon 2000: Proceedings: Intelligent Systems and Technologies for the New Millennium*, Kuala Lumpur, Malaysia, 157-161.

Austin, M., Looney, C. et al. (1997). Security market timing using neural network models. *New Review of Applied Expert Systems, 3*, 3-14.

Baba, N., & Handa, H. (1995). *Utilization of neural network for constructing a user friendly decision support system to deal stocks*. IEEE International Conference on Neural Networks, Australia, 818-823.

Baba, N., Inoue, N. et al. (2002). Utilization of soft computing techniques for contructing reliable decision support systems for dealing stocks. In *IJCNN'02: Proceedings of the 2002 International Joint Conference on Neural Networks*, Honolulu, Hawaii, 2150-2155.

Baba, N., Yanjun, Y. et al. (2001). Knowledge-based decision support systems for dealing Nikkei-225 by soft computing techniques. In *Knowledge-Based Intelligent Information Engineering Systems and Allied Technologies KES 2001*. The Netherlands: IOS Press, 728-732.

Baek, J. & Cho, S. (2000). "Left shoulder" detection in Korea Composite Stock Price Index using an auto-associative neural network. Intelligent Data Engineering and Automated Learning—IDEAL 2000: Data Mining, Financial Engineering and Intelligent Agents, Hong Kong, 286-291.

Blakey, P. (2002). Pattern recognition techniques. *IEEE Microwave Magazine, 3*, 28-33.

Chan, K. C. C., & Foo, K. T. (1995). *Enhancing technical analysis in the Forex market using neural networks*. IEEE International Conference on Neural Networks, Australia, 1023-1027.

Chen, A., Chianglin, C. et al. (2001). Establishing an index arbitrage model by applying neural networks method—A case study of Nikkei 225 Index. *International Journal of Neural Systems, 11*(5), 489-496.

Chenoweth, T., Obradovic, Z. et al. (1995). Technical trading rules as a prior knowledge to a neural networks prediction system for the S&P 500 Index. *IEEE Technical Applications Conference and Workshops*, Portland, Oregon, 111-115.

Falas, T., Charitou, A. et al. (1994). *The application of artificial neural networks in the prediction of earnings*. Orlando, FL: IEEE.

Fu, T. -C., Chung, F. -L. et al. (2001). Evolutionary segmentation of financial time series into subsequences. *Proceedings of the 2001 Congress on Evolutionary Computation*, Seoul, Korea, 426-430.

Hobbs, A., & Bourbakis, N. G. (1995). A neurofuzzy arbitrage simulator for stock investing. *Computational Intelligence for Financial Engineering (CIFER)*, New York, 160-177.

Jang, G., Lai, F. et al. (1991). An intelligent trend prediction and reversal recognition system using dual-module neural networks. *First International Conference on Artificial Intelligence Applications on Wall Street*, New York, 42-51.

Kamijo, K., & Tanigawa, T. (1990). Stock price pattern recognition: A recurrent neural network approach. In *Proceedings of the International Joint Conference on Neural Networks*, San Diego, 215-221.

Leigh, W., Purvis, R. et al. (2002). Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: A case study in romantic decision support. *Decision Support Systems, 32*(4), 361-377.

Liu, N. K., & Lee, K. K. (1997). An intelligent business advisor system for stock investment. *Expert Systems, 14*(3), 129-139.

Mizuno, H., Kosaka, M. et al. (1998). Application of neural network to technical analysis of stock market prediction. *Studies in Informatics and Control, 7*(2), 111-120.

Pring, M. J. (1999). *Martin Pring's introduction to technical analysis*. Singapore: McGraw-Hill.

Quah, T. -S., & Srinivasan, B. (2000). Utilizing neural networks in stock pickings. In *Proceedings of the International Conference on Artificial Intelligence,* 941-946.

Samuel, C., & Malakkal, I. (1990). Leading-edge investors downplay debate on fundamental vs. technical analysis. *Wall Street Computer Review, 7*, 22-28, 53.

Skabar, A., & Cloete, I. (2001). Discovery of financial trading rules. In *Proceedings of the IASTED International Conference on Artificial Intelligence and Applications*(AIA 2001), Marbella, Spain, 121-125.

Suh, Y. -H., & LaBarre, J. E. (1995). An application of artificial neural network models to portfolio selection. *Journal of Computer Information Systems, 36*(1), 65-73.

Tan, C. N. W. (2001). *Artificial neural networks: Applications in financial distress prediction and foreign exchange trading*. Gold Coast, QLD: Wilberto.

Wong, F., & Lee, D. (1993). Hybrid neural network for stock selection. *The Second Annual International Conference on Artificial Intelligence Applications on Wall Street,* New York, 294-301.

Yao, J., & Poh, H. -L. (1995). Forecasting the KLSE index using neural networks. *IEEE International Conference on Neural Networks (ICNN '95)*, Perth, Western Australia, 1012-1017.

KEY TERMS

Continuation Pattern: A pattern in technical analysis that suggests, on the balance of probabilities, that price trend will continue in its current direction.

Fundamental Analysis: The use of company-reported financial data to determine an intrinsic (or fair value) for a security. Used to identify cases where companies are undervalued, with a view to profiting from the future price movements. This style of analysis is generally long term.

Noisy Data: This term is generally used to describe data and data sets in which there is a low signal-to-noise ratio. Any algorithm attempting to filter out the signal has to be capable of identifying and dealing appropriately with noise. In this sense, noise is that element of the data that obscures the true signal.

Reversal Pattern: A pattern in technical analysis that suggests, on the balance of probabilities, that price trend will change direction.

Technical Analysis: The study of the behaviour of market participants, as reflected in the technical data. Used to identify early stages in trend developments, with a view to profiting from price movements. This style of analysis is generally short term.

Technical Data: "Technical data" is the term used to describe the components of price history for a security. These components are open price, low price, high price, close price, volume traded, and open interest.

Technical Indicators: Technical indicators are produced as results of various computations on technical data. They are primarily used to confirm price action.

Triangle Pattern: A triangle is a particular pattern observed using technical analysis. There are a variety of circumstances under which a triangle can occur, and dependent on the circumstances, the triangle can be either a reversal or continuation pattern.

Artificial Neural Networks Used in Automobile Insurance Underwriting

Fred L. Kitchens
Ball State University, USA

INTRODUCTION

As the heart of the insurance business, the underwriting function has remained mostly unchanged for nearly 400 years when Lloyd's of London was a place where ship owners would seek out men of wealth. The two would contractually agree to share the financial risk, in the unlucky event that the ship would be lost at sea (Gibb, 1972; Golding & King-Page, 1952).

Today, insurance underwriters perform a similar function on behalf of their respective insurance companies. Underwriters gathering pertinent information and analyze their potential clients to determine whether or not they should underwrite the risk; and if so, what premium they would require for the insurance policy. Insurance companies employ actuaries to help the underwriter in this process by studying past insurance losses and making predictive models for future risks. Using traditional statistical methods, insurance actuaries look for loss-contributing characteristics within the risk (Webb, Harrison et al., 1992). When the actuaries find positive relationships between the policy characteristics and subsequent losses, they create "underwriting guidelines" for the underwriters to follow, when analyzing potential clients (Malecki & Underwriters, 1986).

For hundreds of years, actuaries used pencil and paper to perform their statistical analysis; it was a long time before they had the help of a mechanical adding machine, still longer before they had computers. As recently as 1981, computers were not considered important to the underwriting process. Leading experts in insurance underwriting believed that the human-judgment factor involved in the insurance underwriting process was too complex for any computer to handle as effectively as a human underwriter (Holtom, 1981).

Recent research in the application of technology to the underwriting process has shown that Holtom's statement may no longer hold true (Kitchens, 2000; Lemaire, 1985; Rose, 1986). The time for computers to take-on an important role in the insurance underwriting judgment process may be here. The author intends to illustrate the applicability of artificial neural networks to the insurance underwriting process.

BACKGROUND

The American Institute for Chartered Property Casualty Underwriters reports that the most common considerations found in automobile underwriting guidelines are: age of operators, age and type of automobile, use of the automobile, driving record, territory, gender, marital status, occupation, personal characteristics of the operator, and physical condition of the vehicle. Traditionally, these comprise the basic variables used in determining the acceptability, classifying, and rating of private passenger automobile insurance policies (Malecki & Underwriters, 1986).

Private passenger automobile insurance is well suited for artificial intelligence applications applied to the underwriting function. There are three primary reasons for this: there is a fixed set of finite data used to make the underwriting decision; policies are highly standardized; and deviations from the standard insurance contract are rare.

In recent years, researchers have considered the application of computers to the process of automobile insurance underwriting. Two studies attempted to predict the acceptability of a given policy from a broad underwriting standpoint (Gaunt, 1972; Rose, 1986). Two other studies considered the possibility of predicting a loss on an individual-policy basis (Lemaire, 1985; Retzlaff-Roberts & Puelz, 1966). Another study focused the relationship between premium and customer retention from year-to-year. One study was designed to predict losses on individual policies using artificial neural networks (Kitchens, 2000).

The recent use of artificial neural networks represents what may result in the most accurate application of computers in the underwriting process. Originally developed in the 1940's, artificial neural networks were designed to replicate and study the thought process of the human brain (Cowan & Sharp, 1988). Early research showed that all processes that can be described with a finite number of symbolic expressions could be represented with a finite number of interconnected neurons (Whitley, Starkweather et al., 1990). Thus, artificial neural networks also provide a means of economic problem solving.

The author believes that for a number of reasons discussed in the following section, artificial neural networks can be successfully applied to the insurance underwriting process in order to reduce the ratio of insurance losses to insurance premiums.

NEURAL NETWORKS FOR INSURANCE UNDERWRITING

Artificial neural networks were first developed in the 1940's as a mathematical model used to study the human thought process (Cowan & Sharp, 1988). McCulloch and Pitts in 1943 proved that all processes which can be described with a finite number of symbolic expressions can be represented in a network of interconnected neurons (Whitley, Starkweather, & Bogart, 1990). This makes the artificial neural network a mathematical modeling tool in addition to a representation of the human brain.

Using a data set consisting of dependent and independent variables, an artificial neural network can be trained until it converges on an optimal solution for the dependent variable(s). If properly developed, the resulting model will be at least as accurate as traditional statistical models (White, 1989).

The insurance business, as practiced in the United States, has certain characteristics that produce less than optimal financial results. There are five basic reasons that the unique abilities of artificial neural networks can improve the underwriting process:

First, an artificial neural network model will be successful because the inequity of the current rate classification system will allow neural networks the opportunity to more accurately assess the risk level of each and every individual policy holder, rather than a class of policy holders (Wood, Lilly et al., 1984).

Second, an artificial neural network model will produce improved results because current actuarial methods of study will benefit from the broad range of available tools " such as more recent developments in the field of artificial intelligence (Cummins & Derrig, 1993; Kitchens, 2000).

Third, an artificial neural network model will improve the current state of actuarial research. Traditionally, the primary method of research in this field has been to predict the *pure premium* (the amount of premium required to pay all of the losses in a given class of insured accounts, a.k.a. "relative rates"). In comparison, *actual premiums* include the pure premium along with other important factors such as profit margin and operating expenses. The traditionally used pure premium models follow an actuarial approach, but not necessarily an underwriting approach. While it is intended to reduce corporate loss ratios, current actuarial research does not take an underwriting approach to the process. A fresh perspective on the problem could produce improved results.

Fourth, an artificial neural network will produce improved results because historically, statistical models used in predicting insurance losses have been able to produce only marginal incremental improvements. Given the current state of technology, the time has come for new insurance actuarial models to take advantage of the available speed and flexibility of artificial neural networks to solve what is clearly a complex problem, which will require extensive training and is likely to involve a complex architecture (Kitchens, 2000).

Fifth, even if the actuarial models are "perfect" (which the author contends they are not), the neural network should be capable of at least matching the current statistical results, if not improving upon them. This is because artificial neural networks comprise a class of nonlinear statistical models whose processing methods are designed to simulate the functioning of the human brain (Hawley, Johnson et al., 1990). The advantage of neural network models over other modeling methods grows with the complexity of the relationship between input and output variables; however, greater complexity of the underlying relationships between variables requires a more complex design (Lee, White et al., 1993). Provided the appropriate network architecture, a neural network output function can accurately approximate any mathematical function (White, 1989). Further, a model can achieve any degree of desired accuracy if the neural network is properly designed (Funahashi, 1989).

NEURAL NETWORK MODELS: DESIGN ISSUES

Automobile accidents occur with a certain degree of randomness, and it is expected that they will be very difficult to predict on an individual-policy basis. Previous research has shown that an underwriter's ability to predict the actual value of a paid claim is exceedingly difficult, if possible at all (Kitchens, Johnson et al., 2001). However, a successful system needs only to predict the incident (occurrence) of a loss, not the dollar value. In addition, a successful model would not have to predict each and every accident, as long as the predictions that the model makes are accurate. In fact, a new model needs only to outperform any current models, in order to prove its self worthwhile. As an industry rule-of-thumb, the average loss-to-gross-premium ratio is approximately 60%. The rest of the collected premium is used to pay operating expenses and a small profit of approximately 3%. Thus, if a new model could reduce losses by 1%, it would represent a 33% increase in operating profit. If a corresponding decrease in operating expenses such as loss-adjustment expenses is incurred, the operating profit could be increased by as much as 53%. This in itself is not a justification for using artificial neural networks, but it is enough incentive to try nontraditional techniques.

While it is theoretically possible for a computer program to handle the underwriting function, a reliable model has not yet been developed. Aside from the practical development, the human-side must also be considered. It will take time for society, insurance regulators, and the insurance industry to understand and accept a computer-based underwriting model. The development of a computer-based model that might aid the underwriter in the decision making process will be a good first step. An artificial neural network as an underwriter's tool could be used in several ways " to confirm an underwriter's decision; to provide a suggested course of action; or to handle routine policies, allowing the underwriter to spend more time on more complex policies. As an underwriter's tool, a model should be useful, reliable, and convenient while providing information of value. While there may be many methods of designing such a tool, the author believes that artificial neural networks hold the greatest likelihood of success.

In the development of an artificial neural network model as an underwriter's tool, several things must be determined: the output required, the input variables, the type of artificial neural network, the architecture of the artificial neural network, and the interpretability of the output.

The underwriter's decision-making process boils down to two basic decisions. First, he must decide whether to accept or reject the risk. Second, if accepted, a decision as to the premium that will be charged must be made.

Depending on the purpose of the model, the required output may place stringent requirements on the required input. One reason all previous models have had limitations on their applicability has been due to the lack of quantity or quality in the available data sets used to generate the model.

For purposes of insurance loss-prediction modeling, the Genetic Adaptive Neural Network (GANNT) algorithm is an appropriate choice. The GANNT algorithm is designed to overcome difficulties associated with the popular gradient and backpropagation techniques (Dorsey & Mayer, 1994).

The genetic algorithm was first proposed in 1975 (Holland, 1975; Nygard, Ficek et al., 1992). It was shown that the biological evolutionary process could be applied to an artificial mathematical modeling system (Konza, 1992). The concept is based on the theory that an optimization problem can be encoded as a list of concatenated parameters (nodal weights), which are used in the artificial neural network (Whitley, Starkweather et al., 1990). The genetic algorithm works through a process of modeling founded on the biological process by which DNA replicates, reproduces, crosses over, and mutates (Crane, 1950). These procedures are then modeled in a computer-based algorithm to solve complex problems (Nygard, Ficek et al., 1992). The actual operations of the genetic adaptive neural network training algorithm are explained in detail by Dorsey, Johnson, and Mayer (1991).

RESULTS

Recent research has shown that the automobile insurance underwriting process practiced in the United States is lacking in precision. Underwriters are not efficiently utilizing all of the information available to them. The current human-based underwriting process uses only a portion of the available information. In some cases, so much valuable information is overlooked that the remaining unutilized information can be used to make a more accurate accept/reject decision than the initial underwriter made (Kitchens, 2000). With the benefit of the flexibility and adaptability of an artificial neural network, the unutilized information may be used in the future to make more accurate and more precise underwriting decisions.

FUTURE TRENDS

Future research should be focused on two primary areas. First, to be properly trained, an artificial neural network requires data from both the "accepted" and the "unaccepted" policies. Thus, some effort needs to be focused on obtaining information about the policies that are currently being rejected by the underwriter. This is difficult information to obtain because insurance companies have no reason to track losses on policies they previously rejected. But, this data will be valuable in the development of a more complete underwriting model.

Second, future research should go beyond the accept-or-reject decision, and investigate the premium-setting decision. A model capable of second-guessing an underwriter's accept-or-reject decision might be capable of reducing an insurance company's losses. But, a model that can both accept-or-reject, and set the premium, might be capable of reducing the cost of underwriting, streamline the business process and produce policies that are more appropriately-priced.

CONCLUSION

Since insurance underwriting began 400 years ago, until recently, the biological neural network (human brain) has been the fastest, most efficient, and most readily available information processing tool available. It has naturally been the tool of choice for underwriters when making accept-or-reject and pricing decisions on insurance policies.

In 1981, it was a common belief that computers could not be used to replace insurance underwriters (Holtom, 1981). In the past 20 years or so, computers and technology have made tremendous advancements. During the same time period, sophisticated mathematical models and the algorithms used to generate them, including the Genetic Adaptive Neural Network Training Algorithm, have taken advantage of increased computing power and availability. Artificial neural networks and the technology used to run them have been shown to out perform the traditional human-based mental decision making practices, in both speed and accuracy; if only for limited domains and applications, such as insurance underwriting.

REFERENCES

Cowan, J.D., & Sharp, D.H. (1988). Neural nets. *Quarterly Reviews of Biophysics, 21*, 305-427.

Crane, H.R. (1950). Principles and problems of biological growth. *The Scientific Monthly, LXX* (6), 376-386.

Cummins, J.D., & Derrig, R.A. (1993). Fuzzy trends in property-liability insurance claim costs. *Journal of Risk and Insurance, 60*(3), 429-466.

Dorsey, R.E., & Mayer, K.J. (1994). *Optimizing using genetic algorithms*. Greenwich, CT: JAI Press.

Dorsey, R.E., Johnson, J.D., et al. (1991). The genetic adaptive neural network training (GANNT) algorithm for genetic feedforward artificial neural systems. Working Paper, The University of Mississippi.

Funahashi, K. (1989). On the approximate realization of continuous mappings by neural networks. *Neural Networks, 2*, 183-192.

Gaunt, L.D. (1972). Decision-making in underwriting: Policyholder selection in private passenger automobile insurance, Georgia State University “College of Business Administration: 00310.

Gibb, D.E.W. (1972). *Lloyd's of London: A study in individualism*. London: Lloyd's.

Golding, C.E., & King-Page, D. (1952). *Lloyd's*. New York: McGraw-Hill.

Hawley, D.D., Johnson, J.D., et al. (1990). Artificial neural systems: A new tool for financial decision-making. *Financial Analysts Journal, 46*(November/December), 63-72.

Holland, J. (1975). *Adaptation in natural and artificial systems*. University of Michigan Press.

Holtom, R.B. (1981). *Underwriting: Principles and practices*. Cincinnati, OH: The National Underwriter Company.

Kitchens (2000). Using artificial neural networks to predict losses in automobile insurance. Graduate School. Oxford, The University of Mississippi: 150.

Kitchens, F., Johnson, J.D., et al. (2001). Predicting severity in automobile insurance losses using artificial neural networks. *Production and Operations Management Society International Conference*, Sao Paulo, Brazil.

Konza, J.R. (1992). *Genetic programming: On the programming of computers by means of natural selection*. Cambridge, MA: MIT Press.

Lee, T.H., White, H., et al. (1993). Testing for neglected nonlinearity in time series model: A comparison of neural network methods and alternative tests. *Journal of Econometrics, 56*(3), 269-290.

Lemaire, J. (1985). *Automobile insurance: Actuarial models*. Boston, MA, U.S.A., Kluwer-Nijhoff, Distributors for North America Kluwer Academic Publishers.

Malecki, D.S., & A. I. F. P. A. L. Underwriters (1986). *Commercial liability risk management and insurance*. Malvern, PA: American Institute for Property and Liability Underwriters.

Nygard, K.E., Ficek, R.K., et al. (1992). Genetic algorithms: Biologically inspired search method borrows mechanisms of inheritance to find solutions. *OR/MS Today* (August), 28-34.

Retzlaff-Roberts, C., & Puelz, R. (1966). Classification in automobile insurance using a DEA and discriminant analysis hybrid. *Journal of Productivity Analysis, 7*(4), 417-27.

Rose, J.C. (1986). *An expert system model of commercial automobile insurance underwriting (knowledge base)*, The Ohio State University: 00280.

Webb, B.L., Harrison, C.M., et al. (1992). *Insurance operations*. Malvern, PA: American Institute for Chartered Property Casualty Underwriters.

White, H. (1989). Neural networks and statistics. *AI Expert, 49*(December).

Whitley, D., Starkweather, T., et al. (1990). Genetic algorithms and neural networks: Optimizing connections and connectivity. *Parallel Computing, 14*, 347-361.

Wood, G.L., Lilly, C.C., et al. (1984). Personal risk management and insurance. USA: American Institute for Property and Liability Underwriters.

KEY TERMS

Actuary: A statistician who practices the collection and interpretation of numerical data; especially someone who uses statistics to calculate insurance premiums.

Artificial Neural Network: (commonly referred to as "neural network" or "neural net") A computer architecture, implemented in either hardware or software, modeled after biological neural networks. Nodes are connected in a manner suggestive of connections between the biological neurons they represent. The resulting network "learns" through directed trial and error. Most neural networks have some sort of "training" algorithm to adjust the weights of connections between nodes on the basis of patterns found in sample or historical data.

Backpropagation: A learning algorithm for modifying a feed-forward neural network which minimizes a continuous "error function" or "objective function". Back-propagation is a "gradient descent" method of training in that it uses gradient information to modify the network weights to decrease the value of the error function on subsequent tests of the inputs. Other gradient-based methods from numerical analysis can be used to train networks more efficiently.

Biological Neural Network: A network of neurons that function together to perform some function in the body such as thought, decision-making, reflex, sensation, reaction, interpretation, behavior, and so forth.

Genetic Algorithm (GA): A class of algorithms commonly used for training neural networks. The process is modeled after the methods by which biological DNA are combined or mutated to breed new individuals. The crossover technique, whereby DNA reproduces itself by joining portions of each parent's DNA, is used to simulate a form of genetic-like breeding of alternative solutions. Representing the biological chromosomes found in DNA, genetic algorithms use arrays of data, representing various model solutions. Genetic algorithms are useful for multidimensional optimization problems in which the chromosome can encode the values for connections found in the artificial neural network.

Insurance: Protection against future loss. In exchange for a dollar value (premium), insurance is a promise of reimbursement in the case of loss. Contractual arrangement of insurance may be voluntarily or by government mandate (such as minimum requirements for automobile insurance for licensed drivers).

Nodal Connections: Connections between nodes in an artificial neural network. They are communication channels that carry numeric data. They simulate the axons and dendrites used to carry electrical impulses between neurons in a biological neural network.

Node: A mathematical representation of a biological neuron. Multiple layers of nodes are used in artificial neural networks to form models of biological neural networks.

Risk: An individual or organization's exposure to a chance of loss or damage.

Underwriter (Insurance Underwriter): An employee of an insurance company whose job duties include analyzing an application for insurance and making a decision whether to accept or reject the application. If accepted, the underwriter further determines the premium to be charged. Underwriters also review existing insurance contracts for renewal.

A

Assessing the Value of Information Systems Investments

Charlene A. Dykman
University of St. Thomas, USA

INTRODUCTION

Accurate assessment of the potential or realized value and impact of an information systems investment is becoming increasingly important. While hardware and software are, to a large extent, becoming commoditized, the overall cost of information systems projects continues to increase. Labor and other organizational costs continue to escalate. The result is that IS professionals must begin to view their systems projects with a business perspective. Such a business perspective includes the application of various quantitative approaches to assessing the value of information systems.

BACKGROUND

Strategies for success in an increasingly information and technology intensive business climate often include major investments in information systems, including hardware, software, networks, and personnel. Investments in information systems continue to rise in both percentage of operating budget and total dollars. And yet, significant early research showed no correlation between expenditures for information technology and any known measure of productivity (Brynjolfsson, 1993; Brynjolfsson & Hitt, 1996; Landauer, 1995; Strassman, 1990, 1997a; Weill, 1992). Although a major expenditure, most companies do not calculate the return on investment (ROI) of the information system investment like they do with other major expenditures. Nor is there much consideration of the social/subsystem cost and benefits of major information systems investments (Ryan, Harrison & Schkade, 2002). Why not?

In the technology and business climate of today, executives, in many cases, seem to have bought into the idea that they have no choice. We see small organizations making major investments in monolithic ERP systems such as SAP and PeopleSoft. Executives seem to have been convinced, by their IS staff, by the vendors, and by the media hype about leading edge technologies, that they must implement the latest fad or methodology in order to remain competitive. We seem to have forgotten that these types of investments are "business" and "organizational" decisions rather than simply "technology" decisions. There is clearly a need to use standard business practices in evaluating the potential costs and benefits of our investments in information systems and the associated technology (Davis, 2002; Talon & Gurbaxani, 2000).

Just as expenditures rise, many information systems projects are judged as failures. We hear repeatedly that information systems fail to deliver on the promises made and the expectations generated. These failures result in a costly mistake for the organization and the project manager. When projects fail, the dollars invested have been wasted and there is a clear impact on the bottom line of the organization. Additionally, several researchers have found that the intangible costs can also be very significant. This is known as the social subsystem cost and many researchers opine that these costs need to be more seriously considered (Ryan, Harrison & Schkade, 2002).

But what if the project is a success? The chairman of the board has just asked you how much of the investment has delivered quantifiable benefits, since your company has just spent eight months and millions of dollars implementing a new business application. Trying to justify expenditures is very common practice in business. It is relatively easy to quantify and justify expenditures for a piece of manufacturing equipment, additional personnel, or payroll processing systems. Trying to quantify and justify an integrated information system (IS) is not quite as easy. Many of the benefits realized from an IS are intangible and managers are not quick to recognize such benefits, just as they often undervalue such benefits. Some recent researchers have begun to show a payoff for IT investments (Brynjolffson & Hitt, 1995; Dewan & Min, 1997; Hitt & Brynjolfsson, 1996; Stratopoulos & Dehning, 2000). However, it is not clear when and where these payoffs will be found. Many business executives balk at measuring the value and performance of something so intimidating, complex and, well, so technical. However, a bottom line focus is really required, especially in organizations that have a very public image to present.

The information systems arena is facing continued scrutiny, due to the escalating costs that seem to spiral out of control. Stakeholders cannot afford to ignore assessment of the value of information systems, given the huge amount of expenditures invested in these systems. As the technology sector on Wall Street has come under increasing scrutiny, organizations have become extremely cost conscious. Some researchers urge that stock market impact must be evaluated, especially as these investments relate to e-commerce and/or outsourcing (Hayes, 2000; Subramani & Walden, 2001). Not only have IS expenditures skyrocketed over the past few years, but also internal corporate departments have become much more aware of the charges incurred through chargeback algorithms over which they have little control. This directly impacts the bottom line of the firm. If we are serious about being good stewards of the organization's resources, changing the approach to information technology investment decisions can decrease expenses and increase revenue opportunities.

ASSESSING THE BENEFITS OF AN INFORMATION SYSTEM INVESTMENT

Determination of the benefits and quantification of their projected value to the organization is a multifaceted task that is still more art than science. In fact, Ryan et al. (2002) found that there is a tendency for executives to view the post-implementation benefits as being more important than the implementation costs. There is renewed interest in establishing formal processes for managing IT investments due to the history of large capital investments with poor track records for success and perceptions of low return in value to the organization (Kim & Sanders, 2002). The ultimate value of IT is how it impacts business processes in line with the strategy of the organization. A good business case for an investment will show the appropriate linkage to this strategy. There are a variety of methods used to assist in this valuation:

- No justification: This includes a summary of operating and capital costs for the project and ongoing use of the system. There is no benefits analysis. This is used, and rightly so, for "cost of doing business" projects. Unfortunately, the real problem surfaces when every information systems investment decision is lumped into the "cost of doing business" group. This is an easy way to avoid scrutiny of the real expected cost or payoff from this investment. Senior general management in the firm must limit projects that fall into this "loophole" in order to accurately assess the costs and benefits of their IT investments.
- Total cost of ownership: This method is often used by consulting firms and includes the summation of all costs (purchase, operation, maintenance, and disposal of technology) to compare costs within a product line. Notice again that there is only a focus on costs, not on the organizational benefits to be derived from this investment.
- Financial metrics: These methods focus on costs and benefits in financial terms, including interest rate information and the time value of money. Several researchers have stressed the importance of accounting techniques in valuing IT investments (Dehning & Richardson, 2002). Several key financial indicators, also known as accounting performance measures, should be used in financial analysis. These include the net present value (NPV), return on investment (ROI) and the internal rate of return (IROR) calculations. Real option evaluation includes the notion of uncertainty and risk (Li & Johnson, 2002). The question that is asked with real option evaluation is whether making the investment today has enough net present value to make up for losing the option to delay the investment (Carlsson & Fuller, 2000). Microsoft's Rapid Economic Justification Model, through a five-step process of discernment, attempts to align IT investments with the success of the business by defining critical success factors, assigning probabilities to risks, and so forth. This approach is quite comprehensive, but hardly rapid.
  While appealing, the big risk with financial metrics is "spurious exactitude". We tend to believe, because there is a number associated with something, that we are really able to measure the phenomenon quantitatively. While it is imperative that we take a hard dollar approach to these hard to define and difficult to quantify variables, we must not lose sight of the threat to the validity of these results. Just because we cannot do it perfectly, does not mean that we should not do it. The process of such evaluation has a value in and of itself, in requiring us to focus on the business, its goals and strategies and to break down the components of a project and discern the relationships between the project and the business goals.
- Information economics: This is a scoring method that addresses the value of the information that results from the use of the system. This is difficult to measure since information itself is an intangible. Additionally, information itself has no inherent value. Value can be derived only when information is applied to specific organizational processes. If we are, for example, trying to evaluate the benefit of governmental spending on major information systems to track terrorists, then we can easily see the value of the information that might be provided. This is often compounded by that fact that the

information already exists in the organization, just not in as convenient a format, or in a way that it can easily be manipulated. At that point we are really looking at the "value-added" component of this investment.

- Balanced score card: This is a sophisticated management framework that translates strategy and vision into operational tactics. Measures and metrics for aligning processes with vision and the initiatives needed to meet the objectives of the business are identified. This complex method requires total organizational support to be successful. However, there is food for thought in just understanding the dimensions that are addressed by this method. We see once again the importance of the overall strategy of the organization in directing the investments made in the information systems arena.

FUTURE TRENDS

Investments in information systems represent major crossroads for 21st century organizations. These decisions are all unique decisions, at a point in time in the history of an organization and its competitive environment. Additionally, this point in time perspective is true of the technology itself. How many projects have we all seen that took so long to get to completion that the technology was obsolete when it was finally implemented? We must move quickly and decisively in the business of information systems. This brings risks that must be minimized through:

- Alignment of the technological environment with the organization's strategic plan. Of course, this presupposes that the organization has a strategic plan that can be translated into clear operational goals and measurements with well-defined operational processes and systems solutions. The successful chief information officer will be working at a level, with corporate leadership, where he or she can influence this process.
- The framing of all information systems investments around business requirements rather than technology requirements. This is not about technology looking for organizational problems to solve. Planning does not begin with preconceived hardware and software recommendations.
- The establishment of performance improvement goals and holding managers accountable for meeting them. Even when the benefits of information systems investments are intangible, managers must be measured against the accuracy of their tangible financial projections and the impacts on the social subsystem. These projects must be incorporated into the operating budget and long-term financial plan for both the IS function and the overall organization. Every investment must be justified with concrete revenue (benefit) and expense commitments.

CONCLUSION

Though the costs associated with implementing a new information system may seem exorbitant, the costs associated with delaying or not installing a new system (lost opportunity costs) may be even more devastating to an organization. As the use of information systems technology increases in most industries, so do the capital expenditures. Even though the cost of computer technology is becoming more affordable, the business process needs that can be addressed through information systems continue to increase. This contributes to the escalating aggregate costs of information systems.

It really is time for the academics to take a look at the various approaches to valuation of information systems projects and find out whether any of these are really working. We need to begin to understand the contingencies that are involved in deciding which approach to use for a specific project. In some ways, valuation of IT investments begins to look like economic impact assessments trying to determine the financial impact of, for example, a city hosting the Super Bowl or Olympics. Does anyone ever really see if these projections are accurate? Until we are able to do so in the realm of IT valuation, there will always be doubts about the numbers given in support of a project.

Determination of the benefits to be gained from a particular IS investment can be problematic, because information itself is an intangible. The various issues and approaches discussed in this article give an understanding of the questions that need to be asked and the information that needs to be gathered in order to begin the process of evaluating the potential benefits of an information systems investment. Remember that information systems investments are *business* decisions, not *technology* decisions. We must learn to make a business case in support of our recommendations and finally to compare the impact of the system to the expected benefits that were used to justify the investment in the first place. Finally, we must evaluate the accuracy of these projections after the information systems project is finished.

REFERENCES

Brynjolfsson, E. (1993). The productivity paradox of information technology: Review and assessment. *Communications of the ACM*, *36*(12), 67-77.

Brynjolfsson, E., & Hitt, L. (1995, May). Computers as a factor of production: The role of organizational differences among firms. *Journal of Economic Innovation and New Technologies, 3,* 183-199.

Brynjolfsson, E., & Hitt, L. (1996). Paradox lost? Firm-level evidence on the returns to information systems spending. *Management Science, 42*(4), 541-560.

Carlsson, C., & Fuller, R. (2000). Real option evaluation in fuzzy environments. *Proceedings of International Symposium of Hungarian Researchers on Computational Intelligence* (pp. 69-77).

Davis, C. (2002). *Technologies and methodologies for evaluating information technology in business.* Hershey, PA: Idea Group Publishing.

Dehning, B., & Richardson, V. (2002). Return on investment in information technology: A research synthesis. *Journal of Information Systems, 16*(1), 7-30.

Dewan, S., & Min, C. (1997). The substitution of information technology for other factors of production: A firm level analysis. *Management Science, 43*(12), 1660-1675.

Hayes, D., Hunton, J., & Reck, J. (2000). Information systems outsourcing announcements: Investigating the impact on the market value of contract-granting firms. *Journal of Information Systems, 14*(2), 109-125.

Hitt, L., & Brynjolfsson, E. (1996). Productivity, business profitability, and consumer surplus: Three different measures of information technology value. *Management Information Systems Quarterly, 20*(2), 121-141.

Kim, Y., & Sanders, G. (2002). Strategic actions in information technology investments based on real option theory. *Decision Support Systems, 33*(1), 1-12.

Landauer, T. (1995). *The trouble with computers.* Cambridge, MA: MIT Press.

Li, X., & Johnson, J.D. (2002). Evaluate information technology investment opportunities using real options theory. *Information Resources Management Journal, 15*(4), 32-47.

Ryan, S. Harrison, D., & Schkade, L. (2002). Information technology investment decisions: When do costs and benefits in the social subsystem matter? *Journal of Management Information Systems, 19*(2), 85-127.

Strassman, P. (1990). *The business value of computers: An executive's guide*. New Canaan, CT: The Information Economics Press.

Strassman, P. (1997a). *The squandered computer*. New Canaan, CT: The Information Economics Press.

Stratopoulos, T., & Dehning, B. (2000). Does successful investment in information technology solve the productivity paradox? *Information and Management, 38*(2), 103-117.

Subramani, M., & Walden, E. (2001). The impact of e-commerce announcements on the market value of the firm. *Information Systems Research, 12*(2), 135-154.

Tallon, P., Kraemer, K., & Gurbaxani, V. (2000). Executives' perceptions of the business value of information technology: A process-oriented approach. *Journal of Management Information Systems, 16*(4), 145-173.

Weill, P. (1992). The relationship between investment in information technology and firm performance: A study of the valve-manufacturing sector. *Information Systems Research, 3*(4), 307-333.

KEY TERMS

Accounting Performance Measures: Evaluation of the impact of information systems investments including typical accounting rations such as return on assets and return on equity.

Balanced Score Card: A management framework for translating strategy into tactics through identification of metrics for aligning processes with visions. Initiatives needed to meet the objectives of the business are identified.

Direct Impact: Information technology investments that can be evaluated as causally related to reduced cost or increased profits. For example, more effective management of inventory leads to reduced inventory carrying costs.

Indirect Impact: Information technology that is not clearly related to measurable impact. For example, a new system helps executives make better decisions through providing access to better or more accurate information.

Information Economics: A scoring method that assesses the value of the information that results from the use of an information system.

Market Performance Measures: Measures of impact that focus on event studies, market valuation of common equity, and Tobin's q as used to assess impact of investment in information systems technology.

Productivity Paradox: Major information technology investments with negative or zero documented returns. "Absence of positive relationship between spending on information technology and productivity or profitability" (Dehning, 2002).

Rapid Economic Justification Model: A Microsoft model that attempts to align information technology investments with the success of the underlying business through defining corporate strategy, critical success factors, project risks and impacts. Based loosely of the total cost of ownership approach to information technology investment evaluation.

Real Option Method: An evaluation approach that focuses on the notion of uncertainty and probability of risk. The theory is that a firm should invest today only if the net present value of the project is high enough to compensate for giving up the value of the option to delay the project.

Social Subsystem: The internal climate and culture of an organization, usually seen as several continua focusing on dimensions such as structure, randomness, stress, empowerment of individuals, knowledge-sharing, and so forth.

Strategic Relevance: A measure of the degree to which information systems directly facilitate the attainment of the organization's strategic goals.

Total Cost of Ownership: An evaluation method with a broad perspective on the costs of system ownership, including training, lost productivity during learning cycles, maintenance, and so forth. TCO methodologies bring into focus a softer side of the cost equation.

Audience Response Systems and Face-to-Face Learning

David A. Banks
University of South Australia, Adelaide, Australia

INTRODUCTION

Audience Response Systems (ARS) are gradually being introduced into educational settings, having previously proved their value in business. Hand-held keypads allow students to enter data in response to questions or statements displayed on a public screen. The captured data is displayed on the public screen and enables both academic and students to immediately see how the whole group has responded. The anonymity afforded by an ARS encourages individuals to fully participate without fear of ridicule or loss of face.

The technology is simple to use by both students and academics, can be used with large or small groups and has applications in all topics of study and at all levels of study. ARS are highly portable, require very little set-up time and are easy to use by anyone who has had some experience with software such as PowerPoint.

BACKGROUND

ARS developed within the general area of Computer Supported Collaborative Work (CSCW) and have been used in the business community since the late 1980's. Early work in the use of computer-based systems was concerned with providing managers with decision support tools that would provide support for, and improve effectiveness of, decision-making especially when dealing with semi-structured or unstructured situations. The systems were developed to be under the control of decision makers and support the process rather than attempt to automate it. Keen & Scott Morton (1978) initially described as Decision Support Systems they have also been described as Group Decision Support Systems (GDSS). Mallach (1994, p.7) defines a GDSS as "an information system whose primary purpose is to provide knowledge workers with information on which to base informed decisions". Young (1989, p.8) describes the key feature of such systems as being that "… they are intended to interact with and enhance the special mental capabilities of the user, thereby facilitating learning, creativity…". A key word in the name of these systems is "support", that is, they enhance the process they are supporting by providing rapid data capture, effective processing tools and immediate feedback to the participants as part of a broader human process.

The initial systems were based on personal computers and often appeared in networked laboratory-type settings but smaller hand-held systems using keypads were developed. From the early 1990's onwards these systems have been adopted by a growing number of businesses and are now most commonly described as ARS. The relatively low cost and ease of use meant that they attracted the attention of champions in organizations who often had a specific use for the system, often in the areas

Table 1. Typical uses for ARS in business

Quality (Self assessed, Baldrige, EQA…)	Control risk self assessment (CRSA)
Evaluating new product ideas	Budget and capital resource allocation
Corporate governance	Succession planning
Appointment interviews	Corporate ethics
Team building	Idea sharing, testing, approval
Conflict management	Customer value, customer satisfaction
Employee and customer focus groups	Stress management
Counselling related	Time and Project Management
Readiness for change	Benchmarking
Tapping employee wisdom	Product development

of quality self-assessment or team building. The initial emphasis on decision-making has now been broadened by recognition that these systems can provide support for a wide range of business processes. The feature of anonymity provided by these systems offers an opportunity for use in environments where there may be holders of widely differing levels of organisational power and authority engaged in discussion. The ease of use, portability, and flexibility in application provides an opportunity for groups to engage in a variety of reflective practices, with groups ranging from eight through to several thousand participants.

Typical business uses for these systems are shown in Table 1. Many of them offer great potential for adoption in higher education learning environments where they can be used as support tools in a variety of subject areas and teaching and learning philosophies.

PRACTICAL SYSTEMS

In practice these systems comprise hand-held input devices that transmit data to a receiving device connected to a personal computer. Software processes the data and presents it in a variety of formats to the participants for discussion. Key components of the system are:

- Hand-held input devices. A variety of sizes and designs exist, the credit-card size keypad (Figure 1) being typical of the most recent development.
- Receiver. Utilizes infrared or other wireless communication media to collect data from the keypads.
- Software. Manages collection and processing of data and supports display of the data in a variety of presentational formats. The software may be embedded in other container software such as PowerPoint. The output from the system is usually displayed on a public screen via a data projector (Figure 2).

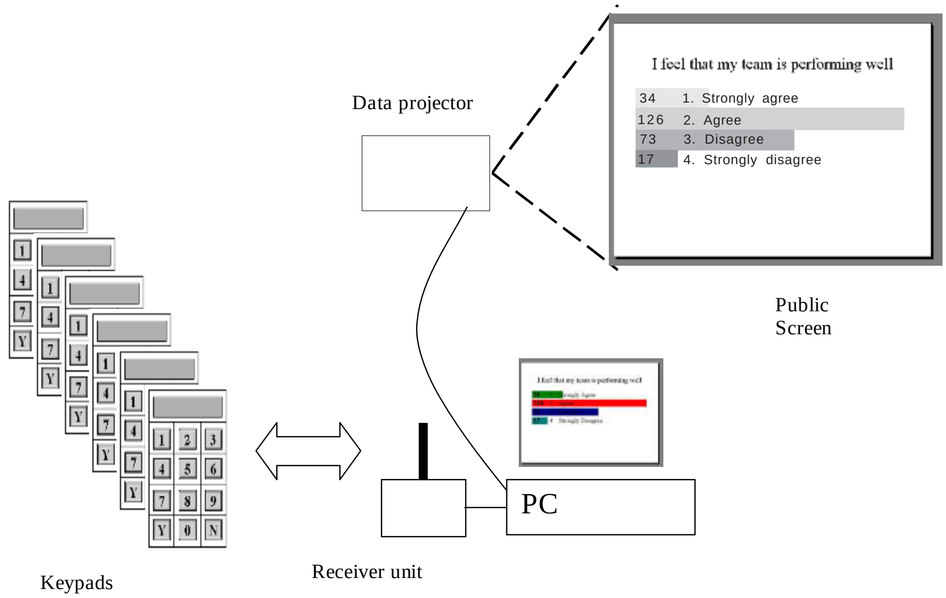
ARS IN HIGHER EDUCATION

The purpose of an ARS is to provide a system and set of tools that can support groups engaged in a face-to-face process that involves examining propositions, exploring alternatives and obtaining timely feedback from the participants engaged in the process. Draper and Brown (2004, p20) suggest that "The dream of personal teaching is really about adaptive teaching; where what is done depends on the learner's current state of understanding". ARS can provide timely feedback to support this adaptive

Figure 1. Credit-card size keypad (image courtesy of KEEpad Pty Ltd)

teaching goal, but Draper and Brown make the point that this can only be achieved through appropriate pedagogic design and action and not through the technology alone. In one-to-one or small group settings the learning facilitator may have a sense of the current state of the learner if the learner feels sufficiently comfortable in revealing it. With large groups in more formal settings the availability of cues to the learning facilitator can be more limited. The immediate feedback that an ARS offers can publicly identify differences or similarities of opinion within groups and provide a trigger for further discussion or analysis of data and re-adjustment of pacing or content. ARS can be used with both large (hundreds of participants) and small groups to support lectures, workshops, seminars, and to explore a wide range of subjects. They can be used at undergraduate and postgraduate levels, and within traditional and post-modern paradigms. Subject areas that value discussion, debate, multiple interpretations and direct challenges to accepted wisdom can benefit from this technology, but equally an ARS can be used in subject areas where demonstration of understanding of a fixed body of knowledge is vital. ARS can be used for formative and summative assessment, in the gauging of preliminary level and subsequent stages of understanding of a subject and in the exploration of the concepts that underpin critical issues.

Figure 2. Component parts of an ARS



Mitchell (2001) suggests that ARS can be used for mechanistic purposes such as monitoring class attendance via individual handsets, providing instant marking and feedback and for gathering data that can be used to support research activities related to classroom processes. McCabe, Heal and White (2001) used an ARS to support Computer Assisted Assessment (CAA) approaches with mathematics students and consider that it not only reinforced existing CAA activities but also served as a valuable tool for motivating higher levels of student learning. Hunt, Irving, Read and Knight (2003) used an ARS in a first-year information systems unit, in a decision-making subject in a third-year psychology course and also with second-year BSc Pharmacy students. In the pharmacy course questions were posed via the ARS and the resulting answers were displayed and discussed by the whole group. A key issue here is that what is being sought is not necessarily a "correct" answer but instead an examination and exploration of all possible answers and the reasons that individuals give for selecting a specific answer. The students expressed enthusiasm for the system, particularly in its ease of use, the ability to discuss answers immediately after making their choice and in the way it helped students identify where further reading was required. Importantly they also found it to be both easy and fun to use.

Post graduate HRM and MBA students using case-based approaches supported by an ARS indicated that the level of participation and number of ideas generated and explored was greater than usual and that the influence of individual personalities was greatly reduced (Jones, Gear, Connolly & Read, 2001). The students also observed that the technology was simple to use and to some extent became "invisible" to the users. Williams (2003) notes that students on an MBA course using an ARS were strongly in favour of the technology and had negative views about passive learning approaches that simply involved reading or listening. Uhari, Renko and Soini (2003) report that 80% of students studying a pediatric's course felt that an electronic voting system helped improve their learning and enhanced questioning during lectures. Witt (2003) found that 87% of students studying in a statistics for psychologists course saw more benefits than disadvantages in the use of keypads.

In a session run by the author in 1995 an ARS was used to explore students' evaluation of an engineering course (Banks, 2001). Students scored each item of the course feedback instrument using individual keypads instead of the normal paper and pencil approach. Once all items were completed they were worked through again, the students being shown the distribution of their scores and invited to comment. It quickly became evident that the criteria used to generate the scores differed widely from student to student and some discussion took place about what common and appropriate criteria might be. After some discussion the items were scored again and it was noted that greater consensus was evident in the second data set. This approach provided an opportunity for students to develop a common understanding of the criteria used in course evaluation and thus generated a more consistent interpretation of the evaluation instrument. The students found the opportunity to discuss their individual and collective views of the course to be a valuable experience.

The author has more recently used ARS to support peer review session in a Collaborative Information Systems Masters course (Banks, 2003). In this course students work in small groups collaboratively producing conference-style papers and peer review is an integral part of the learning process. Peer review was originally summative and was carried out by means of a paper-based instrument. In response to a student suggestion, a “diagnostic” electronic peer review session was introduced early in the course. This offers an opportunity for students to become aware of the perceptions of their peers of their performance against a number of agreed group performance indicators. The use of this diagnostic session allows problems to be surfaced and identified, solutions discussed and action plans developed before any major tensions affect the groups. The author has also used an ARS to support a “lost in the desert” scenario (Banks and Bateman, 2004) that allows groups to see how individuals and groups perform and to act as a trigger for discussion of issues of communication, trust, negotiation, and teamwork in general.

ISSUES IN THE USE OF ARS IN HIGHER EDUCATION

Benefits of Anonymity

In groups where a number of different cultures are represented there is a danger that in traditional face-to-face settings some students will not offer an answer for fear of “loss of face”. Some shy students may also feel inhibited in large groups and thus not feel able to contribute to the discussion. The anonymity afforded by an ARS provides an opportunity for these potentially disadvantaged students to fully participate in the learning process. The author has noticed that in small group peer review sessions, there is a tendency for overseas students to talk quite comfortably to the data on the screen rather than feel personally challenged. The system appears to offer a decontextualisation of the student data and permits a more open discussion than has been experienced with traditional face-to-face peer review sessions. On the other hand, anonymity can mean that students enter data mischievously, safe in the knowledge that they cannot be identified.

More Than Just the Numbers

For in-class summative tests the ability to use an ARS to collect numeric data from the individual keypads is an obvious asset. However, in more discursive learning environments the distribution patterns of the data can form a trigger for critical discussion. The issue is not what the average or mean score is, but in trying to determine why a particular pattern of data is captured. Differences in scoring patterns by students suggest differing worldviews or interpretations of the data and this provides an opportunity for reflection, sharing of views, critical thinking and deep learning.

Educational Paradigm

This technology can be used with any subject, at any level of education and within any educational paradigm. The paradigm that appears to be common to most current users of ARS is that of constructivism.

Constructivist approaches are based on the premise that learners should engage in active construction of their own knowledge rather than passively receive the views of others. This process requires that learners engage in critical and reflective behaviour and this takes place in response to interactions with other individuals as part of a social sense-making process. The emphasis in terms of teaching moves away from a “sage on the stage” model with lectures as information delivery mechanisms, towards a “guide on the side” model where the focus is upon challenge or justification of prevailing views. The approach exposes students to new experiences and modes of thinking and enables them to examine their own ideas and to determine the extent to which the new experiences make sense in the light of these ideas. They are also encouraged to consider the number of possible alternative explanations for what they have experienced and to evaluate the usefulness of a number of different perspectives (Dufresne, Gerace, Leonard, Mestre & Wenk, 1996). This critical process can take place within a structured learning environment through “justificatory arguments” which are “brought forward in support of assertions, in the structures they may be expected to have, the merits they claim and the ways in which we set about grading, assessing and criticising them” (Toulmin, 1958, p.12). ARS can provide a valuable support mechanism for such approaches.

Question, Then Question Again

By asking a question a number of times and critically evaluating the distribution of the ARS responses it becomes possible to explore the reasons for the differing or changing responses. Gauging the responses of students through the use of an ARS allows for quick and multiple loops around the material if learning appears to be problematic. d'Inverno, Davis and White (2003) report that the use of an ARS suggests that typically around 40% fail to identify the correct answer to simple questions, and that if the same question is asked again around 20% still provide the wrong answer. (They do, however, suggest that there may be some deliberate entry of incorrect answers as not all students feel that the technology offers them benefit.)

Over-Use of the Technology

Over-use or inappropriate use may trivialise the technology and lead to student dissatisfaction. The software for ARS typically interfaces with PowerPoint, and the risk of building a mechanized performance rather than genuinely engaging the learner must be borne in mind. As with any technology it should only be used when it offers a tangible benefit to the process.

Accessibility/Equity for All Students

One potential problem with the introduction of an ARS is that of the provision of the keypads for students. Even though the cost of keypads is already low, and falling, the provision of one keypad per student clearly represents a large investment and technology management issue for an educational institution. Ownership of keypads by the institution also raises the problem of issuing and collecting the keypads before and after learning sessions, maintenance and so on. On way to overcome this problem is to make the students responsible for the keypads. This can be achieved through the currently developing approach of providing students with shrink-wrapped packages of textbooks and keypad, the cost of the keypads being built in to the package purchase price.

FUTURE TRENDS

ARS have shown a dramatic growth in recent years. The increased availability of such technology combined with falling prices and innovative distribution models such as shrink-wrapping with textbooks should encourage more educators to adopt ARS. The simplicity of use and natural integration with all subject areas and teaching styles should ensure the painless adoption of the technology across institutions. Such widespread adoption will mean that students feel that the purchase of personal keypad systems represents a good investment.

In the near future, systems that utilise personal digital assistant (PDA) technology and mobile phone technology will add text input, further increasing the versatility and power of these systems. There are many research questions to be explored as this technology is introduced. These will include the potential for the use of the technology to promote deep learning, the long-term reactions of students to the technology versus the initial novelty factor, the effectiveness at different levels of study and in different subjects, and many other areas of concern will offer many opportunities for investigation. The use of frequent in-course evaluations rather than a single exit evaluation will allow student concerns to be addressed in a more timely way and will also allow discussion to take place between teacher and learner about both the evaluation instrument and the meaning of the captured data.

The ARS itself will provide a vehicle to support the gathering of research data in actual teaching and practice. The availability of data immediately after a session has been conducted should allow timely adjustments to be made in pacing and content.

CONCLUSION

Considerable time and effort is being invested in distance learning via the web, but it is equally important that the benefits of technology are applied to support and enhance more traditional face-to-face learning environments.

ARS provide educators with an opportunity support their existing teaching and learning strategies in a way that provides them with improved, immediate and dynamic insight to the progress of learners. Students have an opportunity to engage with learning in an active way that helps them see how they, and their peers, are performing on a moment-to-moment basis. The opportunities for learning to be a shared experience are improved and there is the potential for engagement with subject material at a deep level.

REFERENCES

Banks, D. A. (2001). A critical reflection on the value of course evaluation questionnaires: Using group support tools to explore student evaluations. *Proceedings of the 2001 Information Resources Management Association Conference*, Toronto, Canada, 292-295.

Banks, D. A. (2003). Using keypad-based group process support systems to facilitate student reflection. *Proceedings of the 20th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE)*, Adelaide, 37-46.

Banks, D. A., & Bateman, S. (2004). Using an audience response system to support a "lost in the desert" learning scenario. *Proceedings of the International Conference on Computers in Education (ICCE 2004)*, Melbourne.

d'Inverno, R. A., Davis, H. C., & White, S.A. (2003). Student feedback … A lesson for the teacher. *Teaching Mathematics and Its Applications*, 22, 163-169.

Draper, S., Cargill, J., & Cutts, Q. (2001). Electronically enhanced classroom interaction. *Proceedings of the 18th Annual Conference of the Australasian Society for Computers in Tertiary Education*, 161-167.

Draper, S.W., & Brown, M. I. (2004). Increasing interactivity in lectures using an electronic voting system. *Journal of Computer Assisted Learning*, 20, 81-94.

Dufresne, R. J., Gerace, W. J., Leonard, W. J., Mestre, J. P., & Wenk, L., Classtalk. (1996). A classroom communication system for active learning. *Journal of Computing in Higher Education*, 7, 3-47.

Hunt, A., Irving, A., Read, M. & Knight S. Supporting learning with a group decision support system (GDSS) [Online]. Retrieved May 16, 2003, from *http://www.pbs.port.ac.uk/~readm/GDSS/GDSS.htm*

Jones, C., Gear, A., Connolly, M., & Read, M. (2001). Developing the professional skills of postgraduate students through interactive technology. *Proceedings of the 2001 Information Resources Management Association Conference*, Toronto, Canada, 759-761.

Keen, P.G.W., & Scott Morton, M.S. (1978). *Decision support systems: An organizational perspective*. Reading, MA: Addison-Wesley.

KEEpad, Retrieved May 14, 2003, from *http://www.keepad.com*

Mallach, E. G. (1994). *Understanding decision support systems and expert systems*. Sydney, Australia: Irwin.

McCabe, M., Heal, A. & White, A. (2001). Computer assisted assessment (CAA) of proof = Proof of CAA: New approaches to computer assessment for higher level learning. *Proceedings of the 5th International Conference on Technology in Mathematics Teaching.*

Mitchell, C. (2004) PRS support system – A summer project [Online]. Retrieved May 16, 2004, from *http://grumps.dcs.gla.ac.uk/papers/PRS_Support_System3.doc*

Toulmin, S. E. (1958). *The uses of argument.* UK: Cambridge University Press.

Uhari, M., Renko, M., & Soini, H. Experiences of using an interactive audience response system in lectures. *BMC Medical Education*, 3:12. Retrieved February 23, 2004, from *http://www.biomedical.com/1472-6920/3/12*

Williams, J. B. (2003). "Learning by remote control": Exploring the use of an audience response system as a vehicle for content delivery. *Proceedings of the 20th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE)*, Adelaide, 739-742.

Witt, E. (2003). Who wants to be … The use of a personal response system in statistics teaching. *MSOR Connections*, 3(2).

KEY TERMS

Anonymity: A feature of an ARS that can protect the identity of a participant.

Audience Response System (ARS): An electronic system designed to support and enhance face-to-face group interaction by means of individual hand-held communication devices.

Computer Supported Collaborative Work (CSCW): Use of computer-based technology, including ARS, to support group processes.

Constructivism: An educational approach that takes the view that knowledge must be constructed within the cognitive structure of each individual.

Group Decision Support System (GDSS): A collection of hardware and software used to support decision-makers.

Keypad: A hand-held device that allows a participant to communicate data to an Audience Response System.

Audience-Driven Web Site Design

Olga De Troyer
WISE Research Group, Vrije Universiteit Brussel, Belgium

INTRODUCTION

In the last few years, Web sites have evolved from a simple collection of hypertext pages towards applications supporting complex business processes. Although it is still easy to publish a couple of pages, more and more it is recognized that appropriate design methods are needed to develop more complex Web sites. In the past, Web sites were created opportunistically without prior planning or analysis, and without any regard for methodology, resulting in the classical maintenance problems and development backlog. At the same time, a new problem unknown in classical information systems emerged: competition for the visitor's attention. Especially for commercial Web sites, it is important to hold the interest of the visitors and to keep them coming back. As stated by usability expert Jakob Nielsen: "all the competitors in the world are but a mouse click away" (Nielsen, 2000). Much more than in "classical" software systems, the *usability* of Web sites is a primary factor for their success.

BACKGROUND

One way to deal with the usability of a Web site is by assessing the usability of the Web site and improving the Web site afterwards. There are different ways to assess the usability of a Web site. The techniques that can be used are mainly the same as those used in usability testing of classical user interfaces, for example heuristic evaluation, expert-based evaluation, experimental evaluation, interviews, questionnaires… (Nielsen & Mack, 1994). Another approach to enhance usability (and complementary to the first approach) is to use a Web site design method that ensures a higher usability. The first methods proposed for Web site design are HDM (Garzotto, Paolini & Schwabe, 1993) and its successors HDM2 (Garzotto, Paolini & Mainetti, 1993) and OOHDM (Schwabe & Rossi, 1995), and RMM (Isakowitz et al., 1995) were originally designed for hypertext applications or came from the database research community. These methods used database design methods like E-R (Chen, 1976) or OMT (Rumbaugh et al., 1991), and focused on the organization of the data to be presented on the Web site. These methods could solve to some extent maintenance problems, but they did not address usability. Essential for good usability in Web sites is meeting the needs of the (different) visitors. WSDM was one of the first Web site design methods to recognize this. This method was presented at the WWW7 conference (1998) as a "user-centered" design method for Web sites (De Troyer & Leune, 1998). The starting point in the approach is the set of potential visitors (audiences) of the Web site. The method recognizes that different types of visitors have different needs and that this should drive the design of the Web site rather than the organization of the available data. Later on (De Troyer, 2001), the authors renamed their approach from "user-centered" to "audience-driven" to avoid confusion with the term "user-centered" from the HCI (human computer interaction) field. In HCI, a user-centered approach refers to a design process in which users are actively involved (by interviews, scenario analysis, prototyping, evaluation…). This explicit involvement is not necessary in WSDM. Often, most of the Web site users will be unknown; they cannot be interviewed in advance and they cannot be involved in the development process. In the audience-driven approach as defined by WSDM, the users play a central role but it is not necessary to involve them actively in the design process.

APPROACHES TO STRUCTURE WEB SITES

When designing a Web site, there are two important questions to be answered:

1. What information and services should be provided?
2. How should all this information and services be structured?

There exist different approaches to answer these questions. One of them is the audience-driven approach. Two other possible approaches are the *data-driven* approach and the *organization-driven* approach.

In a data-driven approach, the data (and services) available in the organization (in databases, brochures, internal documents…) are the design's starting point. Following this approach, the structure of the Web site will reflect the way the data are structured and maintained in the organization and the content will parallel the internal data. The same applies for services or functionality. Forms available in the organization will be converted into e-forms, and the current way of working will be reflected on the Web site. The advantage is that structuring the Web site is easy and that management of the Web content can be done in parallel with the internal data. However, the disadvantages are: (1) the data are presented and organized the way they are used in the organization. This is not necessarily how people external to the organization need them; (2) some information may be missing because it was not available in the form of a specific document and the designers were not aware of the fact that users may need it; (3) all information and all services are offered to all users. As a consequence, visitors may be drowned in information.

In an organization-driven approach, the internal structure of the organization is the starting point: the structure of the Web site reflects the structure of the organization. This approach is often used for large organizations with a lot of divisions, for example a university Web site that reflects its internal structure into faculties, departments and research institutes. As for the data-driven approach, it is easy to structure the Web site and the development and maintenance of the different parts can be assigned to the different divisions of the organization. The disadvantage is that it may be very difficult for visitors not familiar with the internal structure of the organization to know where to look for information.

In the audience-driven approach, the information and services needed in the Web site are determined by the needs and requirements of the target audiences (users). Also the main structure of the Web site will be based on the different types of audiences. This last point differentiates the audience-driven approach from many so-called user-centered approaches. We illustrate this with a university Web site. Following the audience-driven approach, the university Web site would (at least) contain a part with general information interesting to all visitors; a part with information specific for students and lecturers; and a part containing information for researchers and third parties interested in research. This approach gives consideration to the fact that Web sites usually have different types of visitors that may have different needs. Clearly, such Web sites may have a higher usability than the one structured using a data-driven or organization-driven approach. However, the downside of the medal is that the effort needed to design the Web site is higher and that the task of maintaining may be spread over the organization (usually, there will be no one-to-one mapping from the structure of the Web site onto the structure of the organization).

AUDIENCE-DRIVEN APPROACH

To arrive at an audience-driven organization of the Web site, the different types of audiences and their needs are identified already in an early stage of the design process. This is done by looking at the activities of the organization and the roles people play in these activities. These people are the potential users of the Web site. Next they are classified into *audience classes* by collecting their requirements (information, as well as functional, and usability requirements). Users with the same information and functional requirements become members of the same audience class. Users with additional requirements form audience subclasses. In this way a hierarchy of audience classes can be constructed, which will be the basis for the main structure of the Web site. For each audience class, a separated *audience track* will be created. Such an audience track can be considered as a sub-site that will provide all the information and services needed by the members of this audience class. To fill in the detailed navigation and content of such a track, the requirements of the corresponding audience class are translated into task- and object models. A task model is defined for each information- and functional requirement of each audience class. Each task defined in the task model is decomposed into elementary tasks, and temporal relationships among the tasks are expressed. For each elementary task an object model (called "object chunk") is created, which models the information and/or functionality needed to fulfill the requirement of that elementary task. In the next step, the task models of an audience class are translated and combined into an audience track. All audience tracks together form the navigational model. The navigational model defines the conceptual structure of the Web site and describes how the members of the different audience classes will be able to navigate through the site and perform the tasks. Based on this conceptual structure, a page structure will be defined as well as the "look and feel" of the Web site. The aim is to create a consistent, pleasing and efficient look and feel for the conceptual design made by taking into consideration the usability requirements and characteristics of the audience classes. If needed, a logical data design (database schema, XML schema, etc.) is made by integrating the different object chunks. Finally, the Web site can be realized using the chosen implementation environment, for example, HTML or XML.

FUTURE TRENDS

In the last years, many different researchers have recognized the need to take the users into consideration during the development process and adopted either the user-centered approach from the HCI field (as in Cato, 2001; Lazar, 2001; McCracken & Wolfe, 2004) or an approach similar to WSDM's audience-driven approach with respect to the fact that it is necessary to give due attention to the users but not necessary to actively involve them in the design process (as in Bomsdorf & Szwillus, 2003; Brinck, Gergle & Wood 2002; Lengels, 2002). More and more, we see that methods and approaches for Web site design incorporate all the goodies known from user interface design and usability research. It is expected that both fields will continue to combine in the future (a good example of this is the book by McCracken & Wolfe, 2004) and will benefit from each other.

In addition, we think that it is useful to combine different approaches to structure Web sites to accommodate the different ways users are working. For example some people prefer to use a search engine to locate the information; others may be more comfortable with the data-driven approach, while in some situations an organization-driven approach may be more appropriate. We expect that it must be possible to combine all of them with the audience-driven approach in a single Web site.

CONCLUSION

There exist different approaches to elicit and structure the information and services in a Web site. In the audience-driven approach, this is done by taking the different audiences of the Web site as the starting point. This results in Web sites where the information and the services are organized according to the needs of the different audience classes. This may result in higher usability, which is a primary factor for the success of a Web site.

REFERENCES

Bomsdorf, B., & Szwillus, G. (2003). User-centered modeling of interactive Websites. In I. King & T. Maray (Eds.), *CDROM Proceedings WWW2003 Conference.*

Brinck, T., Gergle, D., & Wood, S.D. (2002). *Usability for the Web: Designing Web sites that work.* Morgan Kaufmann Publishers.

Cato, J. (2001). *User-centered Web design.* Addison-Wesley Pearson Education.

Chen, P.P. (1976). The entity-relationship model: Towards a unified view of data. *ACM Transactions on Database Systems, 1*(1), 471-522.

De Troyer, O. (2001). Audience-driven Web design. In M. Rossi & K. Siau (Eds.), *Information modeling in the new millennium*. Hershey, PA: Idea Group Publishing.

De Troyer, O., & Leune, C. (1998). WSDM: A user-centered design method for Web sites. *Computer Networks and ISDN systems*, *Proceedings of the 7th International World Wide Web Conference* (pp. 85 – 94). Elsevier.

Garzotto, F., Paolini, P., & Mainetti, L. (1993). Navigation patterns in hypermedia databases. *Proceedings of the 26th Hawaii International Conference on System Science* (pp. 370-379). IEEE Computer Society Press.

Garzotto, F., Paolini, P., & Schwabe, D. (1993). HDM - A model-based approach to hypertext application design. *ACM Transactions on Information Systems, 11*(1), 1-26.

Isakowitz, T., Stohr, E.A., & Balasubramanian, P. (1995). RMM: A methodology for structured hypermedia design. *Communications of the ACM, 38*(8), 34-43.

Lazar, J. (2001). *User-centered Web development.* Jones and Bartlett Publishers, Inc.

Lengels, J.G. (2002). *The Web wizard's guide to Web design.* Addison-Wesley Pearson Education.

McCracken, D.D., & Wolfe, R.J. (2004). *User-centered Website development: A human-computer interaction approach.* Pearson Prentice Hall.

Nielsen, J. (2000). *Designing Web usability: The practice of simplicity*. New Riders Publishing.

Nielsen, J., & Mack, R.L. (Eds.). (1994). Usability inspection methods. John Wiley.

Rumbaugh, J., Blaha, M., Premerlani W., Eddy F., & Lorensen, W. (1991). *Object oriented modeling and design.* Prentice Hall Inc.

Schwabe, D., & Rossi, G. (1995). The object-oriented hypermedia design model. *Communications of the ACM, 38*(8).

KEY TERMS

Audience Class: Group of target visitors of a Web site with the same requirements.

Audience-Driven Web Design: The different audiences and their requirements are taken as the starting point for the design of the Web site. The information and services in the Web site are organized around the different audience classes.

Audience Track: Part of the Web site that provides information and services specifically tailored to an audience class.

Data-Driven Web Design: The data available in the organization are taken as the starting point for the design of the Web site.

Organization-Driven Web Design: The structure of the organization is taken as the starting point for the design of the Web site. The structure of the organization is reflected in the Web site.

Usability: The extent to which a product or system can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.

User-Centered Web Design: The requirements of the users of a Web site play a central role in the design process.

Audio Analysis Applications for Music

Simon Dixon
Austrian Research Institute for Artificial Intelligence, Austria

INTRODUCTION

The last decade has seen a revolution in the use of digital audio: The CD, which one decade earlier had taken over the home audio market, is starting to be replaced by electronic media which are distributed over the Internet and stored on computers or portable devices in compressed formats. The need has arisen for software to manage and manipulate the gigabytes of data in these music collections, and with the continual increase in computer speed, memory and disk storage capacity, the development of many previously infeasible applications has become possible.

This article provides a brief review of automatic analysis of digital audio recordings with musical content, a rapidly expanding research area which finds numerous applications. One application area is the field of music information retrieval, where content-based indexing, classification and retrieval of audio data are needed in order to manage multimedia databases and libraries, as well as being useful in music retailing and commercial information services. Another application area is music software for the home and studio, where automatic beat tracking and transcription of music are much desired goals. In systematic musicology, audio analysis algorithms are being used in the study of expressive interpretation of music. Other emerging applications which make use of audio analysis are music recommender systems, playlist generators, visualisation systems, and software for automatic synchronisation of audio with other media and/or devices.

We illustrate recent developments with three case studies of systems which analyse specific aspects of music (Dixon, 2004). The first system is BeatRoot (Dixon, 2001a, 2001c), a beat tracking system that finds the temporal location of musical beats in an audio recording, analogous to the way that people tap their feet in time to music. The second system is JTranscriber, an interactive automatic transcription system based on (Dixon, 2000a, 2000b), which recognizes musical notes and converts them into MIDI format, displaying the audio data as a spectrogram with the MIDI data overlaid in piano roll notation, and allowing interactive monitoring and correction of the extracted MIDI data. The third system is the Performance Worm (Dixon, Goebl, & Widmer, 2002), a real-time system for visualisation of musical expression, which presents in real time a two dimensional animation of variations in tempo and loudness (Langner & Goebl, 2002, 2003).

Space does not permit the description of the many other music content analysis applications, such as: audio fingerprinting, where recordings can be uniquely identified with a high degree of accuracy, even with poor sound quality and in noisy environments (Wang, 2003); music summarisation, where important parts of songs such as choruses are identified automatically; instrument identification, using machine learning techniques to classify sounds by their source instruments; and melody and bass line extraction, essential components of query-by-example systems, where music databases can be searched by singing or whistling a small part of the desired piece. At the end of the article, we discuss emerging and future trends and research opportunities in audio content analysis.

BACKGROUND

Early research in musical audio analysis is reviewed by Roads (1996). The problems that received the most attention were pitch detection, spectral analysis and rhythm recognition, areas which correspond respectively to the three most important aspects of music: melody, harmony and rhythm.

Pitch detection is the estimation of the fundamental frequency of a signal, usually assuming it to be monophonic. Methods include: time domain algorithms such as counting of zero-crossings and autocorrelation; frequency domain methods such as Fourier analysis and the phase vocoder; and auditory models which combine time and frequency domain information based on an understanding of human auditory processing. Recent work extends these methods to find the predominant pitch (e.g., the melody note) in a polyphonic mixture (Gómez, Klapuri, & Meudic, 2003; Goto & Hayamizu, 1999).

Spectral analysis is a well-understood research area with many algorithms available for analysing various classes of signals, such as the short time Fourier transform, wavelets and other more signal-specific time-frequency distributions. Building upon these methods, the specific application of automatic music transcription has a long research history (Chafe, Jaffe, Kashima, Mont-

Reynaud, & Smith, 1985; Dixon, 2000a, 2000b; Kashino, Nakadai, Kinoshita, & Tanaka, 1995; Klapuri, 1998, 2003; Klapuri, Virtanen, & Holm, 2000; Marolt, 1997, 1998, 2001; Martin, 1996; Mont-Reynaud, 1985; Moorer, 1975; Piszczalski & Galler, 1977; Sterian, 1999; Watson, 1985). Certain features are common to many of these systems: producing a time-frequency representation of the signal, finding peaks in the frequency dimension, tracking these peaks over the time dimension to produce a set of partials, and combining the partials to produce a set of notes. The differences between systems are usually related to the assumptions made about the input signal (e.g., the number of simultaneous notes, types of instruments, fastest notes, or musical style), and the means of decision making (e.g., using heuristics, neural nets or probabilistic reasoning).

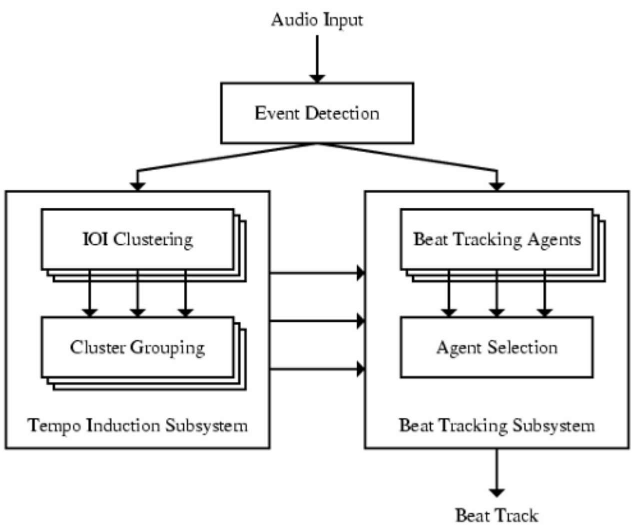
The problem of extracting rhythmic content from a musical performance, and in particular finding the rate and temporal location of musical beats, has also attracted considerable interest in recent times (Allen & Dannenberg, 1990; Cemgil, Kappen, Desain, & Honing, 2000; Desain, 1993; Desain & Honing, 1989; Dixon, 2001a; Eck, 2000; Goto & Muraoka, 1995, 1999; Large & Kolen, 1994; Longuet-Higgins, 1987; Rosenthal, 1992; Scheirer, 1998; Schloss, 1985). Previous work had concentrated on rhythmic parsing of musical scores, lacking the tempo and timing variations that are characteristic of performed music, but recent tempo and beat tracking systems work quite successfully on a wide range of performed music.

Music performance research is only starting to take advantage of the possibility of audio analysis software, following work such as Scheirer (1995) and Dixon (2000a). Previously, general purpose signal visualisation tools combined with human judgement had been used to extract performance parameters from audio data. The main problem in music signal analysis is the development of algorithms to extract sufficiently high level content, since it requires the type of musical knowledge possessed by a musically literate human listener. Such "musical intelligence" is difficult to encapsulate in rules or algorithms that can be incorporated into computer programs. In the following sections, three systems are presented which take the approach of encoding as much as possible of this intelligence in the software and then presenting the results in an intuitive format which can be edited via a graphical user interface, so that the systems can be used in practical settings even when not 100% correct. This approach has proved to be very successful in performance research (Dixon et al., 2002; Goebl & Dixon, 2001; Widmer, 2002; Widmer, Dixon, Goebl, Pampalk, & Tobudic, 2003).

BEATROOT

Compared with complex cognitive tasks such as playing chess, beat tracking (identifying the basic rhythmic pulse of a piece of music) does not appear to be particularly difficult, as it is performed by people with little or no musical training, who tap their feet, clap their hands or dance in time with music. However, while chess programs

Figure 1. System architecture of BeatRoot



compete with world champions, no computer program has been developed which approaches the beat tracking ability of an average musician, although recent systems are approaching this target. In this section, we describe BeatRoot, a system which estimates the rate and times of musical beats in expressively performed music (for a full description, see Dixon, 2001a, 2001c).

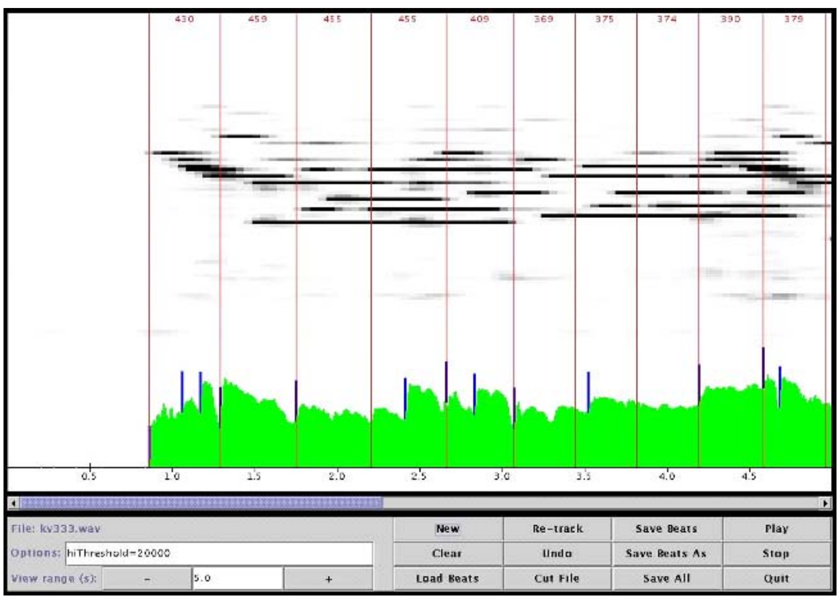
BeatRoot models the perception of beat by two interacting processes (see Figure 1): The first finds the rate of the beats (tempo induction), and the second synchronises a pulse sequence with the music (beat tracking). At any time, there may exist multiple hypotheses regarding each of these processes; these are modelled by a multiple agent architecture in which agents representing each hypothesis compete and cooperate in order to find the best solution. The user interface presents a graphical representation of the music and the extracted beats, and allows the user to edit and recalculate results based on the editing. Input to BeatRoot is either digital audio or symbolic music data such as MIDI. This data is processed off-line to detect salient rhythmic events, using an onset detection algorithm which finds peaks in the slope of the amplitude envelope of the signal (or a set of frequency bands of the signal). The timing of these events is then analysed to generate hypotheses of the tempo at various metrical levels.

First, inter-onset intervals (IOIs), the time differences between pairs of onsets, are calculated, and then a clustering algorithm is used to find groups of similar IOIs which represent the various musical units (e.g., half notes, quarter notes). Information about the clusters is combined by identifying near integer relationships between clusters, in order to produce a ranked list of tempo hypotheses, which is then passed to the beat tracking subsystem.

The beat tracking subsystem uses a multiple agent architecture to find sequences of events which match the various tempo hypotheses, and rates each sequence to determine the most likely sequence of beat times. Each agent represents a specific hypothesis about the rate and the timing of the beats, which is updated as the agent matches the detected onsets to predicted beat times. The agent also evaluates its beat tracking, based on how evenly the beat times are spaced, how many predicted beats correspond to actual events, and the salience of the matched events, which is calculated from the signal amplitude at the time of the onset. At the end of processing, the agent with the highest score outputs its sequence of beats as the solution to the beat tracking problem.

BeatRoot is written in Linux/C++, and comprises about 10,000 lines of code, with a graphical user interface consisting of 1,000 lines of Java. The user interface allows playback of the music with the beat times marked by clicks, and provides a graphical display of the signal and the beats with editing functions for correction of errors or selection of alternate metrical levels (Figure 2).

Figure 2. Screen shot of BeatRoot processing the first five seconds of a Mozart piano sonata, showing the inter-beat intervals in ms (top), calculated beat times (long vertical lines), spectrogram (centre), waveform (below) marked with detected onsets (short vertical lines) and the control panel (bottom)



BeatRoot is open source software (under the GNU Public License), and is available from:

http://www.oefai.at/~simon/beatroot

The lack of a standard corpus for testing beat tracking creates a difficulty for making an objective evaluation of the system. The automatic beat tracking algorithm has been tested on several sets of data: a set of 13 complete piano sonatas, a large collection of solo piano performances of two Beatles songs and a small set of pop songs. In each case, the system found an average of over 90% of the beats (Dixon, 2001a), and compared favourably to another state of the art tempo tracker (Dixon, 2001b). Tempo induction results were almost always correct, so the errors were usually related to the phase of the beat, such as choosing as beats onsets half way between the correct beat times. Interested readers are referred to the sound examples at:

http://www.oefai.at/~simon

Presently, BeatRoot is being used in a large scale study of interpretation in piano performance (Widmer, 2002; Widmer et al., 2003), to extract symbolic data from audio CDs for automatic analysis.

JTRANSCRIBER

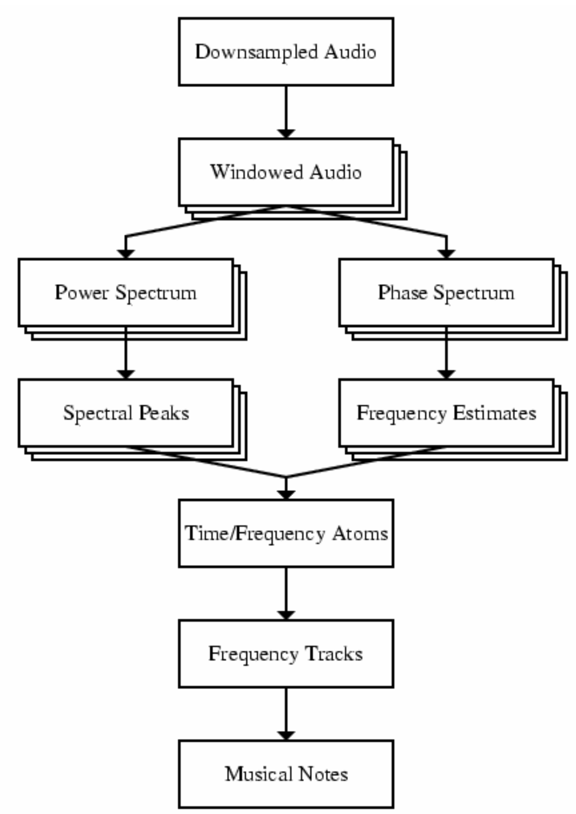
The goal of an automatic music transcription system is to create, from an audio recording, some form of symbolic notation (usually common music notation) representing the piece that was played. For classical music, this should be the same as the score from which the performer played the piece. There are several reasons why this goal can never be fully reached, for example, that there is no one-to-one correspondence between scores and performances, and that masking makes it impossible to measure everything that occurs in a musical performance. Recent attempts at transcription report note detection rates around 90% for solo piano music (Dixon, 2000a ; Klapuri, 1998; Marolt, 2001), which is sufficient to be somewhat useful to musicians.

A full transcription system is normally conceptualised in two stages: the signal processing stage, in which the pitch and timing of all notes is detected, producing a symbolic representation (often in MIDI format), and the notation stage, in which the symbolic data is interpreted in musical terms and presented as a score. This second stage involves tasks such as finding the key signature and time signature, following tempo changes, quantising the onset and offset times of the notes, choosing suitable enharmonic spellings for notes, assigning notes to voices in polyphonic passages, and finally laying out the musical symbols on the page. Here, we address only the first stage of the problem, detecting the pitch and timing of all notes, or in more concrete terms, converting audio data to MIDI.

The data is processed according to Figure 3: The audio data is averaged to a single channel and downsampled to increase processing speed. A short time Fourier transform (STFT) is used to create a time-frequency image of the signal, with the user selecting the type, size and spacing of the windows. Using a technique developed for the phase vocoder (Flanagan & Golden, 1966), a more accurate estimate of the sinusoidal energy in each frequency bin can be calculated from the rate of change of phase in each bin.

The next step is to calculate the peaks in the magnitude spectrum, and to combine the frequency estimates to give a set of time-frequency atoms, which represent packets of energy localised in time and frequency. These are then combined with the atoms from neighbouring frames (time slices), to create a set of frequency tracks, representing the partials of musical notes. Frequency tracks are assigned to musical notes by estimating the most likely set of fundamental frequencies that would give rise to the

Figure 3. Data processing steps in JTranscriber



observed tracks, and the pitch, onset time, duration and amplitude of each note are estimated from its constituent partials.

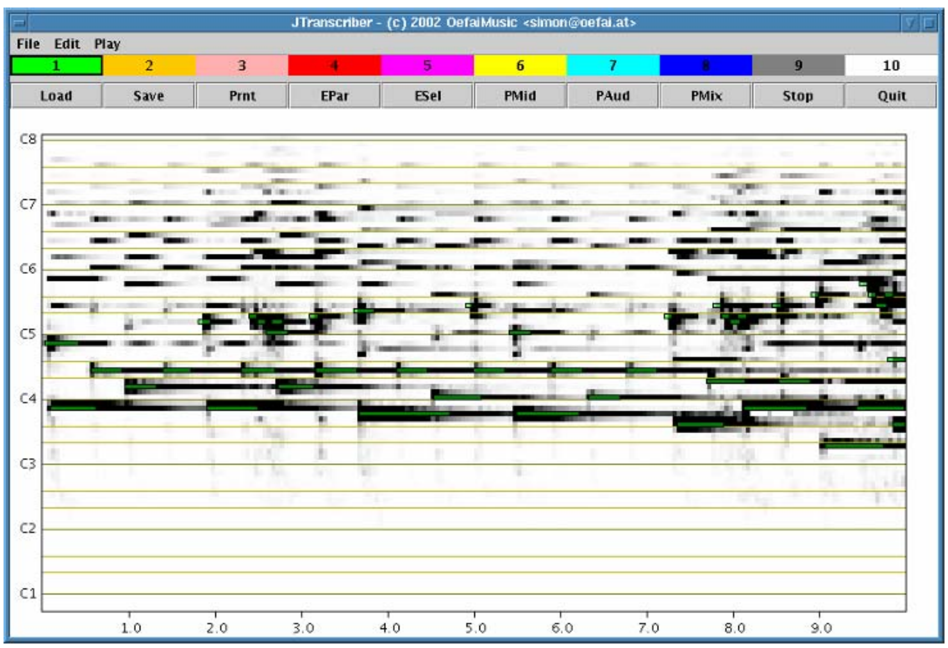
An example of the output is displayed in Figure 4, showing a spectrogram representation of the signal using a logarithmic frequency scale, labelled with the corresponding musical note names, and the transcribed notes superimposed over the spectrogram in piano roll notation. (The piano roll notation is colour and partially transparent, whereas the spectrogram is black and white, which makes the data easily distinguishable on the screen. In the grey-scale diagram, the coloured notes are difficult to see; here they are surrounded by a solid frame to help identify them.) An interactive editing system allows the user to correct any errors made by the automatic transcription system, and also to assign notes to different voices (different colours) and insert high level musical structure information. It is also possible to listen to the original and reconstructed signals (separately or simultaneously) for comparison.

An earlier version of the transcription system was written in C++, however the current version is implemented entirely in Java. The system was tested on a large database of solo piano music consisting of professional performances of 13 Mozart piano sonatas, or around 100,000 notes (Dixon, 2000a), with the results that approximately 10-15% of the notes were missed, and a similar number of the reported notes were false. The most typical errors made by the system are thresholding errors (discarding played notes because they are below the threshold set by the user, or including spurious notes which are above the given threshold) and octave errors (or more generally, where a harmonic of one tone is taken to be the fundamental of another, and vice versa).

THE PERFORMANCE WORM

Skilled musicians communicate high-level information such as musical structure and emotion when they shape the music by the continuous modulation of aspects such as tempo and loudness. That is, artists go beyond what is prescribed in the score, and express their interpretation of the music and their individuality by varying certain musical parameters within acceptable limits. This is referred to as expressive music performance, and is an important part of Western art music, particularly classical music. The Performance Worm (Dixon et al., 2002) is a real-time system for tracking and visualising the tempo and dynamics of a performance in an appealing graphical format which provides insight into the expressive patterns applied by skilled artists. This representation also forms the basis for automatic recognition of performers' style (Widmer, 2002; Widmer et al., 2003).

Figure 4. Transcription of the opening 10s of the second movement of Mozart's Piano Sonata K332. The transcribed notes are superimposed over the spectrogram of the audio signal (see text). It is not possible to distinguish fundamental frequencies from harmonics of notes merely by viewing the spectrogram.

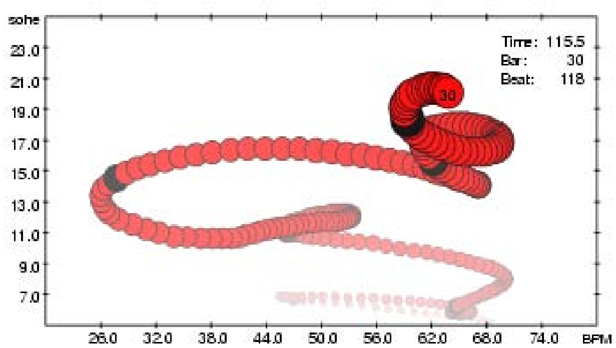


The system takes input from the sound card (or from a file), and measures the dynamics and tempo, displaying them as a trajectory in a 2-dimensional performance space (Langner & Goebl, 2002, 2003). The measurement of dynamics is straightforward: It can be calculated directly as the RMS energy expressed in decibels, or, by applying a standard psychoacoustic calculation (Zwicker & Fastl, 1999), the perceived loudness can be computed and expressed in sones. The difficulty lies in creating a tempo tracking system which is robust to timing perturbations yet responsive to changes in tempo. This is performed by an adaptation of the tempo induction subsystem of BeatRoot, modified to work in real time. The major difference is the online IOI clustering algorithm, which continuously outputs a tempo estimate based only on the musical data up to the time of processing. The clustering algorithm finds groups of IOIs of similar duration in the most recent eight seconds of music, and calculates a weighted average IOI representing the tempo for each cluster. The tempo estimates are adjusted to accommodate information from musically-related clusters, and then smoothed over time by matching each cluster with previous tempo hypotheses. Figure 5 shows the development over time of the highest ranked tempo hypothesis with the corresponding dynamics.

The Performance Worm is implemented in about 4,000 lines of Java, and runs in real time on standard desktop computers. The graphical user interface provides buttons for scaling and translating the axes, selecting the metrical level, setting parameters, loading and saving files, and playing, pausing and stopping the animation.

Apart from the real-time visualisation of performance data, the Worm can also load data from other programs, such as the more accurate beat tracking data produced by BeatRoot. This function enables the accurate comparison of different performers playing the same piece, in order to characterise the individual interpretive style of the performer. Current investigations include the use of AI pattern matching algorithms to learn to recognize performers by the typical trajectories that their playing produces.

Figure 5. Screen shot of the performance worm showing the trajectory to bar 30 of Rachmaninov's Prelude op.23 no.6 played by Vladimir Ashkenazy. The horizontal axis shows tempo in beats per minute, and the vertical axis shows loudness in sones. The most recent points are largest and darkest; the points shrink and fade into the background as the animation proceeds.



FUTURE TRENDS

Research in music content analysis is progressing rapidly, making it difficult to summarise the various branches of investigation. One major initiative addresses the possibility of interacting with music at the semantic level, which involves the automatic generation of metadata, using machine learning and data mining techniques to discover relationships between low-level features and high-level concepts. Another important trend is the automatic computation of musical similarity for organising and navigating large music collections. For other developments in this area, interested readers are referred to the web site at:

http://www.semanticaudio.org

CONCLUSION

The three systems discussed are research prototypes, whose performance could be improved in several ways, for example, by specialisation to suit music of a particular style or limited complexity, or by the incorporation of high-level knowledge of the piece being analysed.

This is particularly relevant to performance research, where the musical score is usually known. By supplying a beat tracking or performance analysis system with the score, most ambiguities are resolved, giving the possibility of a fully automatic and accurate analysis.

Both dynamic programming and Bayesian approaches have proved successful in score following (e.g., for automatic accompaniment, Raphael, 2001), and it is likely that one of these approaches would also be adequate for our purposes. A more complex alternative would be a learning system which automatically extracts the high-level knowledge required for the system to fine-tune itself to the input data (Dixon, 1996). In any case, the continuing rapid growth in computing power and processing techniques ensures that content-based analysis of music will play an increasingly important role in many areas of human interaction with music.

ACKNOWLEDGEMENTS

The Austrian Research Institute for Artificial Intelligence ackowledges the financial support of the Austrian Federal Ministry for Education, Science and Culture (BMBWK) and the Austrian Federal Ministry for Transport, Innovation and Technology (BMVIT). This work was supported by the START programme (project Y99-INF) of the BMBWK. Special thanks to the Bösendorfer Company, Vienna, for performance data used in this work.

REFERENCES

Allen, P., & Dannenberg, R. (1990). Tracking musical beats in real time. In *Proceedings of the International Computer Music Conference* (pp. 140-143), San Francisco CA. International Computer Music Association.

Cemgil, A., Kappen, B., Desain, P., & Honing, H. (2000). On tempo tracking:Tempogram representation and Kalman filtering. In *Proceedings of the 2000 International Computer Music Conference* (pp. 352-355), San Francisco CA. International Computer Music Association.

Chafe, C., Jaffe, D., Kashima, K., Mont-Reynaud, B., & Smith, J. (1985). Techniques for note identification in polyphonic music. In *Proceedings of the International Computer Music Conference* (pp. 399-405), San Francisco CA. International Computer Music Association.

Desain, P. (1993). A connectionist and a traditional AI quantizer: Symbolic versus sub-symbolic models of rhythm perception. *Contemporary Music Review, 9*, 239-254.

Desain, P., & Honing, H. (1989). Quantization of musical time: A connectionist approach. *Computer Music Journal, 13*(3), 56-66.

Dixon, S. (1996). A dynamic modelling approach to music recognition. In *Proceedings of the International Computer Music Conference* (pp. 83-86), San Francisco CA. International Computer Music Association.

Dixon, S. (2000a). Extraction of musical performance parameters from audio data. In *Proceedings of the First IEEE Pacific-Rim Conference on Multimedia* (pp. 42-45), Sydney. University of Sydney.

Dixon, S. (2000b). On the computer recognition of solo piano music. *Mikropolyphonie*, 6. *http://www.mikropol.net/volume6*

Dixon, S. (2001a). Automatic extraction of tempo and beat from expressive performances. *Journal of New Music Research, 30*(1), 39-58.

Dixon, S. (2001b). An empirical comparison of tempo trackers. In *Proceedings of the 8th Brazilian Symposium on Computer Music* (pp. 832-840). Brazilian Computing Society.

Dixon, S. (2001c). An interactive beat tracking and visualisation system. In *Proceedings of the International Computer Music Conference* (pp. 215-218), San Francisco CA. International Computer Music Association.

Dixon, S. (2004). Analysis of musical content in digital audio. In J. DiMarco (Ed.), *Computer graphics and multimedia: Applications, problems and solutions* (pp. 214-235). Hershey, PA: Idea Group Publishing.

Dixon, S., Goebl, W., & Widmer, G. (2002). Real time tracking and visualisation of musical expression. In *Music and Artificial Intelligence: Second International Conference, ICMAI2002* (pp. 58-68), Edinburgh, Scotland. Springer.

Eck, D. (2000). Meter through synchrony: Processing rhythmical patterns with relaxation oscillators. PhD thesis, Indiana University, Department of Computer Science.

Flanagan, J., & Golden, R. (1966). Phase vocoder. *Bell System Technical Journal, 45*, 1493-1509.

Goebl, W., & Dixon, S. (2001). Analysis of tempo classes in performances of Mozart sonatas. In *Proceedings of VII International Symposium on Systematic and Comparative Musicology and III International Conference on Cognitive Musicology* (pp. 65-76), Jyväskylä, Finland. University of Jyväskylä.

Gómez, E., Klapuri, A., & Meudic, B. (2003). Melody description and extraction in the context of music content processing. *Journal of New Music Research, 32*(1), 23-41.

Goto, M., & Hayamizu, S. (1999). A real-time music scene description system: Detecting melody and bass lines in audio signals. In Working Notes of the IJCAI-99 Workshop on Computational Auditory Scene Analysis (pp. 31-40). *International Joint Conference on Artificial Intelligence*.

Goto, M., & Muraoka, Y. (1995). A real-time beat tracking system for audio signals. In *Proceedings of the International Computer Music Conference* (pp. 171-174), San Francisco, CA. International Computer Music Association.

Goto, M., & Muraoka, Y. (1999). Real-time beat tracking for drumless audio signals. *Speech Communication, 27*(3-4), 311-335.

Kashino, K., Nakadai, K., Kinoshita, T., & Tanaka, H. (1995). Organization of hierarchical perceptual sounds:

Music scene analysis with autonomous processing modules and a quantitative information integration mechanism. In C.S. Mellish (Ed.), *Proceedings of the International Joint Conference on Artificial Intelligence* (pp. 158-164), Montréal, Canada: Morgan Kaufmann.

Klapuri, A. (1998). *Automatic transcription of music*. Master's Thesis, Tampere University of Technology, Department of Information Technology.

Klapuri, A. (2003). Automatic transcription of music. In R. Bresin (Ed.), *Proceedings of the Stockholm Music Acoustics Conference* (pp. 587-590).

Klapuri, A., Virtanen, T., & Holm, J.-M. (2000). Robust multipitch estimation for the analysis and manipulation of polyphonic musical signals. In *Proceedings of the COST-G6 Conference on Digital Audio Effects*, Verona, Italy.

Langner, J., & Goebl, W. (2002). Representing expressive performance in tempo-loudness space. In *Proceedings of the ESCOM 10th Anniversary Conference on Musical Creativity*, Liége, Belgium.

Langner, J., & Goebl, W. (2003). Visualizing expressive performance in tempo-loudness space. *Computer Music Journal, 27*(4), 69-83.

Large, E., & Kolen, J. (1994). Resonance and the perception of musical meter. *Connection Science, 6*,177-208.

Longuet-Higgins, H. (1987). *Mental processes*. Cambridge, MA: MIT Press.

Marolt, M. (1997). A music transcription system based on multiple-agents architecture. In *Proceedings of Multimedia and Hypermedia Systems Conference MIPRO'97*, Opatija, Croatia.

Marolt, M. (1998). Feedforward neural networks for piano music transcription. In *Proceedings of the XIIth Colloquium on Musical Informatics* (pp. 240-243), Gorizia, Italy. Associazione di Informatica Musicale Italiana.

Marolt, M. (2001). SONIC: Transcription of polyphonic piano music with neural networks. In *Proceedings of the Workshop on Current Directions in Computer Music Research* (pp. 217-224), Barcelona, Spain. Audiovisual Institute, Pompeu Fabra University.

Martin, K. (1996). *A blackboard system for automatic transcription of simple polyphonic music*. Technical Report 385, Massachussets Institute of Technology Media Laboratory, Perceptual Computing Section.

Mont-Reynaud, B. (1985). Problem-solving strategies in a music transcription system. In *Proceedings of the International Joint Conference on Artificial Intelligence* (pp. 916-919), Los Angeles, CA: Morgan Kaufmann.

Moorer, J. (1975). *On the segmentation and analysis of continuous musical sound by digital computer*. PhD Thesis, Stanford University, CCRMA.

Piszczalski, M., & Galler, B. (1977). Automatic music transcription. *Computer Music Journal, 1*(4), 24-31.

Raphael, C. (2001). Synthesizing musical accompaniments with Bayesian belief networks. *Journal of New Music Research, 30*(1), 59-67.

Roads, C. (1996). *The computer music tutorial*. Cambridge, MA: MIT Press.

Rosenthal, D. (1992). Emulation of human rhythm perception. *Computer Music Journal, 16*(1), 64-76.

Scheirer, E. (1995). *Extracting expressive performance information from recorded music*. Master's Thesis, Massachusetts Institute of Technology, Media Laboratory.

Scheirer, E. (1998). Tempo and beat analysis of acoustic musical signals. *Journal of the Acoustical Society of America, 103*(1), 588-601.

Schloss, W. (1985). *On the automatic transcription of percussive music: From acoustic signal to high level analysis*. PhD Thesis, Stanford University, CCRMA.

Sterian, A. (1999). *Model-based segmentation of time-frequency images for musical transcription*. PhD Thesis, University of Michigan, Department of Electrical Engineering.

Wang, A. (2003). An industrial strength audio search algorithm. In *4th International Conference on Music Information Retrieval (ISMIR 2003)* (pp. 7-13).

Watson, C. (1985). The computer analysis of polyphonic music. PhD thesis, University of Sydney, Basser Department of Computer Science.

Widmer, G. (2002). In search of the Horowitz factor: Interim report on a musical discovery project. In *Proceedings of the 5th International Conference on Discovery Science* (pp. 13-32), Berlin: Springer.

Widmer, G., Dixon, S., Goebl, W., Pampalk, E., & Tobudic, A. (2003). In search of the Horowitz factor. *AI Magazine, 24*(3), 111-130.

Zwicker, E., & Fastl, H. (1999). *Psychoacoustics: Facts and models* (2nd ed.). Berlin: Springer.

KEY TERMS

Automatic Transcription: The process of extracting the musical content from an audio signal and representing it in standard music notation.

Beat Tracking: The process of finding the times of musical beats in an audio signal, including following tempo changes, similar to the way that people tap their feet in time to music.

Clustering Algorithm: An algorithm which sorts data into groups of similar items, where the category boundaries are not known in advance.

Frequency Domain: The representation of a signal as a function of frequency, for example as the sum of sinusoidal waves of different amplitudes and frequencies.

Music Content Analysis: The analysis of an audio signal in terms of higher-level (cognitive) properties such as melody, harmony and rhythm, or in terms of a description of the signal's component sounds and the sound sources which generated them.

Music Information Retrieval: The research field concerning the automation of access to music information through the use of digital computers.

Onset Detection: The process of finding the start times of notes in an audio signal.

Time Domain: The representation of a signal, such as the amplitude or pressure of a sound wave, as a function of time.

Automation of American Criminal Justice

J. William Holland
Georgia Bureau of Investigation, USA

INTRODUCTION

Criminal Justice has been one of the public sectors in the forefront of the move toward automation and digital government. The effect of computerization on American criminal justice has been profound and it has transformed the criminal justice process in many fundamental ways. Starting with President Lyndon Johnson's government commission, *The Challenge of Crime in a Free Society: A Report by the President's Commission on Law Enforcement and the Administration of Justice*, public and private experts in criminal justice and technology laid out the information needs of the criminal justice system and the computer systems to meet those demands. At a time when computerization was minimal throughout the criminal justice system, these task force members developed the blueprint for today's multilayered automated criminal justice environment (Dallek, 1998, pp. 405-407, 409-411; *Challenge of crime in a free society*, 1967, pp. 268-271).

Among the major recommendations of the commission were the creation of a national directory of offenders' criminal records, what came to be known as Computerized Criminal History (CCH) and the development of similar directories at the state level. The commission also called for federal coordination of standards for criminal justice information and sharing. Finally, the report urged that a study of fingerprint classification techniques be undertaken with a view to automating much of the fingerprint search and identification effort and that work be intensified to create a national linkage of files on wanted persons and stolen vehicles under the name of the National Crime Information Center (NCIC) (*Challenge of crime in a free society*, 1967, pp. 255, 268-271; *Task force report: Science and technology*, 1967, p. 69).

BACKGROUND

One of the earliest responses to this report was the creation of the Law Enforcement Assistance Administration (LEAA) within the United States Department of Justice (DOJ). In 1969, LEAA funded Project SEARCH to create a nationwide computerized criminal history system. From this initial effort, SEARCH quickly evolved into an independent consortium of states with the mission of demonstrating a computerized system for the electronic exchange of criminal history information. On the national level, the United States Attorney General assigned management responsibility for the interstate and national portion of this system to the Federal Bureau of Investigation. The states also formed the National Law Enforcement Telecommunications System (NLETS) electronically linking the states as well as the FBI and the Royal Canadian Mounted Police. By 1976, 26 states had used LEAA funding to create state level central repositories for computerized criminal history information (U.S. Department of Justice, 2001c, p. 26).

It became apparent during the last half of the 1970s, however, that greater decentralization of the nation's criminal history systems was urgently needed. To respond to these issues and concerns, the various states, FBI and SEARCH created the Interstate Identification Index or Triple I (III) concept in 1980 (U.S. Department of Justice, 2001c, pp. 26-27, 76-82, 88). Designed to replace a centralized national criminal history file, III was an index of criminal offenders that pointed to the state or states where detailed criminal history information could be found. There was widespread acceptance of III for criminal justice purposes: By 2001, 43 states participated. Legal restrictions and concerns, however, limited use of III for non-criminal justice use and weakened any effort to achieve a truly decentralized criminal history system. Consequently, the FBI continued to maintain criminal histories on individuals to meet interstate non-criminal justice needs (U.S. Department of Justice, 2001c, pp. 76-82).

Another factor that prevented the decentralization of criminal history information was the vast effort required in the time-consuming fingerprint identification process. A new system called the NCIC classification was implemented in the 1970s. It did little, however, to speed up the overall identification process (*Challenge of crime in a free society*, 1967, p. 255; *Task force report*, 1967, p. 16; Ms. Shirley Andrews, personal communication, September 9, 2002).

During the mid 1980s, new technological solutions for fingerprint identification emerged on the market. These systems, called automated fingerprint identification systems (AFIS), significantly reduced the manual tasks needed to search a fingerprint and made true searching of latent crime scene fingerprints possible. By the close of the 1980s, many states and a few local agencies had purchased these systems. Most were stand alone systems dedicated to the fingerprint input, search, and presentation of potential candidates for human comparison. A few states, however, attempted to expand the capabilities of these systems and link them to other criminal history processes. When combined with the proven effectiveness of the AFIS latent search capability, the new technology contained the potential to transform criminal justice systems (U.S. Department of Justice, 2001b, pp. 43-44; U.S. Department of Justice, 2001c, pp. 61-63).

In the early 1990s, efforts were made through the National Institute of Standards and Technology (NIST) to devise a national fingerprint transmission standard; an effort spearheaded by the FBI. By 1993, a national standard for the electronic interchange of fingerprint information was approved by NIST and became the basis for the electronic linkage of local jurisdictions to state criminal history bureaus and the FBI. It formed the basis for the emerging national network of real-time identification and criminal history systems (See *Data format for the interchange of fingerprint, facial, and SMT information*, originally issued in 1993, amended in 1997 and further amended in 2000; U.S. Department of Justice, 2001c, pp. 61-63.)

CURRENT AND FUTURE TRENDS IN CRIMINAL JUSTICE AUTOMATION

Building on these past activities in fingerprint and criminal history automation, emphasis within state and national criminal justice circles has shifted to the need to share information, what is known as integrated criminal justice. With the explosion of the Internet and simultaneous cost limitations on criminal justice system development, both federal and state funding entities require that new criminal justice system developments build in the concept of information sharing, realignment of processing functions, and greater involvement of all criminal justice parties in individual systems development. The goal of this new focus is to eliminate duplicate entry of the same information and increase the overall completeness and accuracy of criminal justice information. (U.S. Department of Justice, 2001c, pp. 63-65; Harris, 2000, pp. 7, 14, 18-20, 41; U.S. Department of Justice, 2001b, pp. 47-48, 50; *Planning the integration of justice information systems*, 2002, pp. 2-3.)

Integrated justice efforts, however, have also resurrected older worries about privacy of such information and merged them with new concerns about greater linkage of criminal justice and non-criminal justice information on individuals. Questions about release of integrated information are linked to serious questions about the accuracy of the information released. These fears are intensified as private companies demand access to criminal history information, gathered at public expense, to market to customers for profit. In many jurisdictions, the old line between public and private responsibilities and authority has faded as private companies have assumed many of the traditional criminal justice information systems functions. In addition, the heightened threat of terrorist attacks has led to efforts to gather large amounts of information on individuals into databases to search for terrorist patterns. These efforts have collided with fears about loss of privacy and misuse of such information by the government. Initiatives such as the Total Information Awareness effort and the MATRIX project to correlate private and public data on suspicious individuals have ground to a halt in the face of protest from citizens fearful of the loss of civil liberties. (Ideas that mattered in 2003:9. No future for terror market, 2003; MATRIX Updates, 2003; *Planning the integration of justice information systems*, 2002, p.5; Stanford, 2003; U.S. Department of Justice, 2001a, pp. 8, 12; U.S. Department of Justice, 2001b, pp. 2-3, 27-28, 50).

CONCLUSION

In 1967, a national commission developed *The Challenge of Crime in a Free Society*, the roadmap for today's highly automated but incomplete criminal justice system. This report served the nation well but it is time to move beyond its confining vistas, time to recognize that dramatic developments in computer technology and digital government demand new answers to old questions and the formulation of entirely new questions. The events of September 11, 2001 have raised anew questions about lack of information on potential threats to society and posed new questions on how we as a nation can weave together governmental and private computerized information to detect dangerous individuals intent on mass murder without compromising constitutional safeguards and individual liberties. It is time to convene a new national task force charged with the duty to assess the challenge of crime and terror in a free digital society. Only then can criminal justice automation and digital government move forward in a planned and comprehensive way.

REFERENCE LIST

(*References marked with an asterisk indicate reports included in the Commission report.)

Challenge of crime in a free society: A report by the President's Commission on Law Enforcement and Administration of Justice. (1967). Washington, DC: US Government Printing Office.

Dallek, R. (1998). *Flawed giant: Lyndon Johnson and his times, 1961-1973*. NewYork: Oxford University Press.

Data format for the interchange of fingerprint, facial, and SMT information. (2000).Washington, DC: US Government Printing Office.

Harris, K.J. (2000, September). *Integrated justice information systems: Governance structures, roles, and responsibilities.* Retrieved July 10, 2002 from SEARCH Group, Inc. Web site at *http://www.search.org/images/pdf/governance.pdf/*

Ideas that mattered most in 2003: 9. No future for terror market (2003, December 28). *Atlanta Journal Constitution, (December 28, 2003),* p. G3.

MATRIX Updates. Retrieved February 23, 2004 from ACLU Website *http://www.aclu.org/Privacy/Privacy.cfm?ID=14240&c=130*

Planning the integration of justice information systems: Developing the justice information exchange model. Retrieved July 10, 2002 from SEARCH Group, Inc. Web site: *http://search.org/integration/pdf/JIEM.pdf*

Stanford, D. D. (2003). ACLU attacks MATRIX on privacy. *Atlanta Journal Constitution*, (October 31), p. G3.

**Task force report: Science and technology*. (1967). Washington, DC: US Government Printing Office.

Toward improved criminal justice information sharing: An information integration planning model (2000, April). Available from the International Association of Chiefs of Police, 515 North Washington Street, Alexandria, VA. 22314-2357.

US Department of Justice. (2001a). *Public attitudes toward uses of criminal history information: A privacy, technology, and criminal justice information report* (NCJ187663). Washington, DC.

US Department of Justice. (2001b). *Report of the National Task Force on privacy, technology, and criminal justice information* (NCJ187669). Washington, DC.

US Department of Justice. (2001c). *Use and management of criminal history record information: A comprehensive report, 2001 update* (NCJ187670). Washington, DC.

KEY TERMS

Automated Fingerprint Identification System (AFIS): A system that provides computerized fingerprint identification of arrestees, applicants for licensing and employment, and crime scene fingerprints of potential suspects.

Computerized Criminal History (CCH): A system containing offenders and their individual arrests, final disposition of those arrests, and custodial information for those arrests.

Federal Bureau of Investigation (FBI): The federal government's investigative, forensic, and criminal justice information system entity. It is part of the U.S. Department of Justice.

Interstate Identification Index (III): a national index of offenders housed at the FBI that points an inquiring entity to those states that contain detailed criminal histories on the requested individual.

Multistate Anti-Terrorism Information Exchange (MATRIX): A consortium of states attempting to create a database of public and private information on individuals to allow for advanced searches to uncover suspicious criminal and/or terrorist activity.

National Crime Information Center (NCIC): A national center housed at the FBI that provides 24 hour a day, seven day a week, real time access to law enforcement for warrants, stolen vehicles and other articles.

National Institute of Standards and Technology (NIST): A federal government standards setting body housed in the U.S. Department of Commerce.

National Law Enforcement Telecommunications System (NLETS): A state managed national network that provides state to state communications as well as links to the FBI and other large scale criminal justice entities.

SEARCH Group, Inc: A state managed consortium that represents the states as a body on criminal justice information systems issues at the national level.

Total Information Awareness: A discontinued effort by the federal government to create a vast database containing public and private information on individuals to allow for advanced search techniques to uncover suspicious activity or indications of possible terrorist links.

Autopoietic Approach for Information System Development

El-Sayed Abou-Zeid
Concordia University, Canada

INTRODUCTION

In the last decade a new generation of information systems (IS), such as enterprise resource planning, Web-based information systems and knowledge management support systems, have emerged in response to ever-changing organizational needs. As these systems are socio-technical phenomena in which social and technical factors interweave the ways in which people work, the issue of "how to integrate the work activity and social context of users into the IS which is being designed" becomes one of the principal problems of IS development (Bai et al., 1999). Therefore, the need for new information system design theories is recognized. According to Walls et al. (1992), an "IS design theory" must have two aspects—one dealing with the description of the system and one dealing with the prescription, that is, the process of developing of the system. The prescription aspect includes a description of procedures and guidelines for system development. In addition, these two aspects have to be grounded on theories from natural or social sciences, that is, kernel theories. Therefore, the development of new IS design theories requires a closer look at the system theories that go beyond the traditional system theory that is based, among other things, on Cartesian dualism, that is, mind/body or cognition/action, and on a model of cognition as the processing of representational information (Mingers, 2001). One of the candidate theories is the theory of autopoiesis, which can be best viewed as a system-grounded way of thinking with biological foundations, together with its extension into social domain.

THEORY OF AUTOPOIESIS

In order to conceive of living systems in terms of the processes that realized them, rather in terms of their relationships with an environment, Maturana and Varela (1980) coined the word *autopoiesis* (autos = self, poienin = creation, production) to denote the central feature of their organization, which is "autonomy". The meaning of this word coveys the very nature of living systems as systems that maintain their *identity* through their own operations of continuous self-renewal. Moreover, these systems could only be characterized with *reference to themselves* and whatever takes place in them, takes place as necessarily and constitutively determined in relation to themselves, that is, *self-referentiality*.

One of the key concepts of autopoiesis is the distinction between organization and structure. On one hand, organization is the capability of a system to re-produce its identity by referring constantly to itself, through the alternate re-production of its components together with the component-producing processes, that is, the capability of a recursive self-reproduction. On the other hand, *structure* is the realization of a system's organization through the presence and interplay of its components in a specific realization space. While *organization* is necessary to establish system unity and identity, *structure* is necessary because different spaces of its actualization impose different constraints on systems' components (Maturana & Varela, 1980). By rough analogy, an algorithm for solving certain problem can be viewed as a description of the system's organization, whereas the corresponding computer program can be viewed as the realization of this organization (structure) in a certain space (programming language).

Autopoietic Systems

An autopoietic system is defined by Maturana and Varela as

"a network of processes of production, transformation and destruction of components. These components constitute the system as a distinct unity in the space of its actualization and they continuously regenerate and realize, through their interactions and transformations, the network of processes that produce them." (Maturana & Varela, 1980, p. 135)

Among the distinct characteristics of the autopoietic systems, the most relevant ones are:

- The simultaneous openness and closure. Autopoietic systems are *open* with respect to structural interaction with the environment, that is, *structural openness*, which is unavoidable consequence of the fact that system elements must satisfy the particular requirements of the physical domain in which they occur, while they are *closed* with respect to their own organization, that is, *organizational closure*. The recognition of *the simultaneous openness and closure* of autopoietic systems is in opposition to the tradition for which a system is one or the other but not both. This interpretation is possible only because of the clear distinction between organization and structure (Bednarz, 1988).

- Structural determination. The state transition a system undergoes in response to environmental perturbations is entirely determined by its structure at that time. Moreover, a system specifies which environmental perturbations may trigger which structural changes. In other words, the environmental perturbations could trigger the system's structural changes but can never determine or direct these changes. Moreover, a system specifies which environmental perturbations may trigger which structural changes. Over time, through ongoing interactions with the environment, an autopoietic system will experience what Maturana and Varela (1992) describe as a *structural drift*, or a gradual change to their structure. The nature of this change is determined by previous system's history of structural changes, that is, its *ontogeny*.

Higher-Order Autopoietic Systems

Two (or more) lower-order autopoietic systems can be "structurally coupled" to form higher-order autopoietic system. Structural coupling is the ongoing process of the congruent structural changes between two (or more) systems that results from recurrent interactions between (among) them. Therefore, structural coupling has connotations of coordination and co-evolution. Moreover, following structural determination principle, two structurally coupled systems means that each of them selects from its possible structural changes those which are compatible with those in the other system and, at the same time, are suitable for the maintenance of its identity.

Social systems, such as enterprises, are constituted through the process of third-order structural coupling, or social coupling, the one that occurs between (or among) two (or more) second-order autopoietic systems. However, the unique feature of any human social system, such as an enterprise, is that the social coupling among its constituents occurs through "language in the network of conservations which language generates and which, through their closure, constitute the unity of a particular human society" (Maturana & Varela, 1992, p. 196). From this perspective, language is viewed as an example of social structural coupling that generates the self and creates meaning through interactions with others. Moreover, language represents what Maturana and Varela would describe as a consensual domain, which is the domain of arbitrary and contextual interlocking behaviors (Mingers, 1995a, p. 78). Within a consensual domain, two autopoietic systems would be able to observe the attribution of meaning to common events and undertake coordinated actions.

Autopoiesis and Cognition

Cognition is the term conventionally used to denote the process by which a system discriminates among differences in its environment and potential states of that environment. The evidence for this cognition is effectiveness of system behavior in response to the environmental perturbations. Today's dominant perspective on cognition, and consequently IS, is the idea that effective action is explainable in terms of manipulating formal and static representations of the objective reality (Mingers, 2001).

According to theory of autopoiesis, perception is neither objectivist nor purely constructivist (Varela, 1992, p. 254). Rather, it is co-determined by the linking of the structure of the perceiver and the local situations in which it has to act to maintain its identity. This is the basis of *enactive (embodied) cognition*, which implies that the autopoietic system's activities condition *that can be perceived* in an environment, and these perceptions, in turn, condition future actions. In this view, "A cognitive system is a system whose organization defines a domain of interactions in which it can act with relevance to the maintenance of itself, and the process of cognition is the actual (inductive) acting or behaving in this domain" (Maturana & Varela, 1980, p. 13). In addition, cognitive domain of an autopoietic system is defined as the domain of all the interactions in which it can enter without loss of identity (Maturana & Varela, 1980, p. 119).

APPLICATIONS OF THE CONCEPTS OF AUTOPOIESIS IN IS DEVELOPMENT RESEARCH

The use theory of autopoiesis in IS research can be classified into two main categories: metaphoric and theory-oriented approaches (Beeson, 2001).

Metaphoric Approaches

Kay and Cecez-Kecmanovic (2002) used the concepts of *social coupling* and *consensual domain* to explain processes underlying the IS-organization relationship and how it impacts on the competitive advantage of an organization. They showed how processes of recurrent interactions between members of different groups, that is, analysts, the IS team, and external clients, within the organizations work environment gave rise to commonalities in understanding, which in turn enabled continual IS organization co-emergence. In the same vein, Maula (2000) used the concepts of structural openness and organizational closure to identify two major knowledge management (KM) functions in four consulting firms. The first KM function, sensory function, is the realization of the structural openness of the firm and its structural coupling with its environment. The second KM function, memory function, is the realization of the concepts organizational closure and self-referentiality that enable the firm's effective functioning and continuous renewal. Finally, Carlsen and Gjersvik (1997) used autopoiesis metaphor to analyze possible organizational uses of workflow technology. They argued against "generic" business processes except as starting points for organizational adaptation. In addition, they indicated that the concept of autopoiesis implies that process models should include references to richer descriptions of the organizational environment and the environment the work process is situated in.

Theory-Oriented Approaches

Bai and Lindberg (1999) used first and second order cybernetics, together with Luhmann's social autopoiesis theory (Luhmann, 1995) and Engeström's activity theory (Engeström, 1987) to develop a framework for studying the relationship between IS design activity, use activity,

Table 1. Autopoietic implications for IS development

Characteristics of Autopoietic Systems	**Implications for IS Development**
Organizational Closure and Self-Referentiality	**Insider frame of reference.** The organizational closure and self-referentiality of an enterprise suggest it is best understood from *inside*. Therefore, an interpretive or hermeneutic approach could more reliably and intelligibly account for the experiences, intentions and interpretations of its members. Moreover, the main role of system developer is the role of "*catalyst* and/or *emancipator*" (Hirschheim et al., 1989), who helps an enterprise's members to develop the necessary inquiring, collaborative and communicative patterns needed to continuously explicate their information requirements. **Historicity.** As an enterprise is continuously reproducing itself, it must do so with constant reference to itself, its past practices, values, decisions, contracts, and commitments (Truex et al., 1999). Therefore, explicating an enterprise's history is an essential element in developing new knowledge and in introducing a new IS (von Krogh et al., 1994)
Structural Determination and Structural Coupling	**Context-dependency of IS development methodology.** Viewing ISD methodology as the means for realizing structural coupling between an enterprise and its new IS implies that it cannot be separated from an enterprise's context. In other words, autopoietic metaphor of an enterprise and its IS suggests "strong" approaches to systems development instead of the commonly used "weak" approaches (see key terms).
Embodied Cognition	**Minimal set of initial requirements.** The autopoietic view of cognition implies that requirements are always in motion, unfrozen, and negotiable (Truex et al., 1999). Therefore, information system development can be viewed as open-ended bootstrapping process that starts with a minimal set of requirements. Moreover, formal representation must be subordinated to the fostering of mutual understanding and coordinated action in the development team and between the team's members and the stakeholders (Beeson, 2001; Kay & Cecez-Kecmanovic, 2002).

and the embedded social context. This framework sheds light on the complex social context within which IS development takes place, and provides an epistemological understanding of the relationship among the elements involved in IS development. Moreover, it can be used to develop methodologies for the practice of IS development, and guide various research activities, such as the socio-technical approach.

IMPLICATIONS FOR IS DEVELOPMENT AND RESEARCH

The autopoietic metaphor provides ways of thinking about the mechanisms underpinning the development and the introduction of IS in an enterprise. Here third-order autopoietic system is used as a metaphor to explore referential correspondence between the characteristics of autopoietic systems and an enterprise and its IS. For example, organizational closure of autopoietic system implies that it is homeostatic and its own organization is the fundamental variable to be maintained constant. This concept may be used to explain why the behavior of IS developers and users seems to be stuck sometimes in persistent patterns or repertoires (Beeson, 2001). Moreover, the difficulties of system integration may be better understood from a perspective of structural coupling than from one of rational design or negotiation (Beeson, 2001). Structural coupling concept can also be used to explore the way common understandings between an enterprise's members emerge as a function of interactions in the work environment (Kay & Cecez-Kecmanovic, 2002).

From autopoietic view introducing a new IS in an enterprise can be conceptualized as a kind of perturbation that provokes or triggers an enterprise's structural-determined responses. Therefore, IS development process can be viewed as the means for realizing structural coupling between an enterprise and its new IS and becomes an integrated aspect of the recurrent interactions between developers and users in the work environment. Table 1 summarizes the implications of theory of autopoiesis for IS development process.

REFERENCES

Bai, G., & Lindberg, L.-A. (1999). A sociocybernetic approach to information systems development. *Kybernetes,* 28(6/7), 792-809.

Bednarz, J. (1988). Autopoiesis: The organizational closure of social systems. *System Research*, *5*(1), 57-64.

Beeson, I. (2001). Implications of the theory of autopoiesis for the discipline and practice of information systems. *Proceedings of the IFIP WG 8.2 Working Conference on Realigning Research and Practice in Information Systems Development: The Social and Organizational Perspective,* Boise, ID (pp. 317-332).

Carlsen, S., & Gjersvik, R. (1997). Organizational metaphors as lenses for analyzing workflow technology. *Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work: The Integration Challenge,* Phoenix, AZ (pp. 261-270).

Engeström, Y. (1987). *Learning by expanding. An activity-theoretical approach to developmental research.* Helsinki: Orienta-Konsultit.

Hirschheim, R., & Klein, H. (1989). Four paradigms of information systems development. *Communications of the ACM, 32*(10), 1199-1216.

Kay, R., & Cecez-Kecmanovic, D. (2002). Toward an autopoietic perspective on information systems crganization. *Proceedings of the Twenty-Third Annual International Conference on Information Systems,* Barcelona, Spain (pp. 383-390).

Luhmann, N. (1995). *Social systems.* Stanford, CA: Stanford University Press.

Maturana, H., & Varela, F. (1980). *Autopoiesis and cognition*. Dordrecht: Reidel.

Maturana, H., & Varela, F. (1992). *The tree of knowledge: The biological roots of human understanding* (rev. ed.). Boston: Shambhala.

Maula, M. (2000). The senses and memory of a firm - Implications of autopoiesis theory for knowledge management. *Journal of Knowledge Management, 4*(2), 157-161.

Mingers, J. (1995a). Information and meaning: Foundations for an intersubjective account. *Information Systems Journal,* 5, 285-306.

Mingers, J. (1995b). *Self-producing systems: Implications and applications of autopoiesis.* New York: Plenum Publishing.

Mingers, J. (2001). Embodying information systems: The contribution of phenomenology. *Information and Organization, 11*(2), 103–128.

Truex, D., Baskerville, R., & Klein, H. (1999). Growing systems e organizations. *Communications of the ACM, 42*(8), 117-123.

Varela, F. (1992). Whence perceptual meaning? A cartography of current ideas. In F. Varela & J. Dupuy (Eds.), *Understanding origins: Contemporary views on the origin of life, mind and society* (pp. 235-263). Dordrecht: Kluwer Academic.

Vessey, I., & Glass, R. (1998). Strong vs. weak approaches systems development. *Communications of the ACM, 41*(4), 99-102.

von Krogh, G., Ross, J., & Slocum, K. (1994). An essay on corporate epistemology. *Strategic Management Journal, 15,* 53-71.

Walls, J., Widmeyer, G., & El-Sawy, O. (1992). Building an information system design theory for vigilant EIS. *Information Systems Research, 3*(1), 36-59.

KEY TERMS

Cognitive Domain: The domain of all the interactions in which it can enter without loss of identity (Maturana & Varela, 1980, p. 119).

Consensual Domain: "The domain of interlocked conducts that results from ontogenetic structural coupling between structurally plastic organisms" (Mingers, 1995b).

Ontogeny: The history of structural changes that a system experiences without losing its identity.

Organization: The configuration of relationships among a system's components that define a system as a unity with distinctive identity, and determine the dynamics of interaction and transformations that it may undergo as such a unity.

Strong vs. Weak Methods: "Strong methods are those designed to address a specific type of problem, while weak methods are general approaches that may be applied to many types of problems" (Vessey et al., 1998).

Structural Coupling: The ongoing mutual co-adaptation between a system and its environment.

Structural Determination: The principle that the actual course of change in an autopoietic system is controlled by its structure rather than direct influence of its environment.

Structure: The physical embodiment of a system's organization in a certain physical domain.

B

Balancing Risks and Rewards of ERP

Joseph Bradley
Central Washington University, USA

INTRODUCTION

Enterprise Resource Planning (ERP) systems claim to meet the information needs of organizations. These off-the-shelf software packages replace hard to maintain solutions created by IS departments or older off-the-shelf packages that often provided only piecemeal solutions to an organization's information needs. ERP systems evolved from material requirements planning systems (MRP) and manufacturing resources planning systems (MRP II). ERP serves the entire enterprise, not just manufacturing and inventory control as with its predecessors. ERP integrates information for the entire organization in a single database. But ERP implementations are often complex and experience serious problems. Failures, abandoned projects and general dissatisfaction have been well publicized in the business press. ERP systems are "expensive and difficult to implement, often imposing their own logic on a company's strategy and existing culture" (Pozzebon, 2000).

BACKGROUND

Three characteristics distinguish ERP implementations from other IT projects (Somers, Ragowsky, Nelson, & Stern, 2001).

- ERP systems are "profoundly complex pieces of software, and installing them requires large investments of money, time and expertise" (Davenport, 1998).
- The packages may require changes in business processes and procedure, may induce customization, and leave the implementing firm dependent on a vendor for support and updates (Lucas, Walton, & Ginsberg, 1988).
- The adopting firm is usually required to reengineer its business processes. As a result, the project must be managed as a broad program of organizational change rather than a software implementation (Markus & Tanis, 2000; Somers et al., 2001).

Despite these risks, global firms were spending $10 billion on ERP software and another $10 billion on consultants to implement the systems in the late 1990s (Davenport, 1998). An AMR study expected firms to spend $47 billion on ERP packages in 2001(Cotteleer, 2002).

WHY DO FIRMS ADOPT ERP?

Firms adopt ERP for technical and business reasons. The technical reasons include: reducing systems operating costs, solving specific problems, such as Y2K, accommodating increased system capacity, and solving maintenance problems with legacy systems. Business reasons may include: presenting a single face to the customer, quoting realistic delivery times, accommodating business growth, improvement of business processes, standardization of data, reduction of inventory carrying costs, and elimination of delays in filling orders (Markus & Tanis, 2000).

Watson and Schneider (1999) attribute the rapid growth of the commercial market for ERP to the following factors:

- Use of the popular client/server platform.
- Can be used as an enabler for reengineering projects
- Y2K compliant.
- Marketed to CEO's and CFO's as "strategic solutions" rather than as transaction processing software.
- A way to outsource a significant part of the IS function. (Watson & Schneider, 1999).

Advantages of ERP systems include:

- Reliable information access by using a single database.
- Avoiding multiple data entries, reducing cost and improving accuracy.
- Delivery and cycle time reduction minimizing delays in reporting.
- Cost reduction including time saving and improved controls.
- Easy adaptability with business process options based on best practices easy to adapt.
- Improved scalability.
- Improved maintenance with long-term vendor contracts.
- Global outreach with extensions to modules such as CRM and SCM.

- E-commerce and e-business capabilities (Rashid, Hossain, & Patrick, 2002).

An example of a decision to adopt an ERP system is provided by Geneva Pharmaceuticals, a manufacturer of generic drugs. Faced with eroding margins and continuing price pressure, the existing systems were proving inadequate. Data shared across business units had to be re-keyed resulting in frequent errors. Data was locked in "functional silos" and did not support new processes. Geneva adopted ERP in order to:

- Implement best practices in business processes.
- Integrate data across business units (hence reduce Re-keying and maintenance costs).
- Enforce data standardization (to reduce software maintenance costs),
- Integrate well with new technologies or systems of acquired companies
- Provide scalability with growing product and customer base, and be Y2K (year 2000) compliant" (Bhattacherjee, 2000).

With the identification of the prospective benefits of ERP why have some firms not adopted ERP?

WHY DO FIRMS NOT ADOPT ERP?

Markus and Tanis (2000) identified three very broad categories of reasons why firms that otherwise have all or some of the reasons to adopt ERP systems, do not adopt it or only adopt ERP in part. These firms may adopt only certain modules and rely on legacy systems or new custom systems for their needs. Other firms may begin an implementation only to discontinue it for a variety of reasons. The reason for this non-adoption or partial adoption can be categorized as follows:

1. Lack of feature-function fit.
2. Company growth, strategic flexibility and decentralized decision-making.
3. Availability of alternatives to increase systems integration.

Lack of feature-function fit may be due to the design of most ERP for discrete manufacturing. Many companies have specialized processes common to their industry, which may not be solved by the best practices embedded in ERP systems. The various modules may not fully support process manufacturing industries, such as food processing and paper manufacturing, project industries, such as aerospace, or industries that manufacture products with dimensionality, such as clothing or footwear (Markus & Tanis, 2000). Although as the ERP market becomes saturated, vendors are designing packages for industries that were previously viewed as too complex.

Companies concerned with maintaining rapid growth rates, those needing strategic flexibility and those without a top down decision making style may be non-adopters or partial adopters of ERP systems. Dell Computer Corp. planned full implementation of SAP R/3 but discontinued the implementation after installing the human resource module. Dell's CIO expressed concern with the software's ability to keep pace with Dell's extraordinary growth rate. Visio, a software company subsequently acquired by Microsoft, expressed concern with the ability of SAP to handle the frequent changes it required to its sales analysis and commission requirements (Markus & Tanis, 2000).

The experiences of Dell and Visio focus on the need for efficiency and flexibility in dealing with the external environment and internal processes. In a stable environment, mechanistic structures are appropriate consisting of "high degrees of standardization, formalization, specialization and hierarchy" (Newell, Huang, Galliers, & Pan, 2003). In a dynamic environment, organic structures are needed to enable organizations to be flexible to change products, processes and structures. In these organizations low levels of standardization, formalization, specialization and hierarchy are most appropriate. ERP may maximize organizational efficiency at the cost of flexibility (Newell et al., 2003). The result may be an inability to respond quickly to changes in the environment, reducing the firm's competitiveness.

Organizational culture may also be a factor in non-adoption or partial adoption of ERP systems. Kraft Foods Inc. was highly decentralized but slowly moving to a one-company philosophy. ERP was regarded as culturally inappropriate with this strategy (Markus & Tanis, 2000).

Lean enterprises succeed "as a growth strategy for increasing sales by trimming the company's product delivery system into a competitive weapon" (Bradford & Mayfield, 2001). Lean enterprises have difficulty using ERP systems due to the lack of flexibility. "ERP creates many nonvalue-added transactions by making companies track every activity and material price in the factory. This is counter to Lean philosophy, which aims at speeding up and smoothing production" (Bradford & Mayfield, 2001).

Alternatives to ERP systems include data warehousing technologies that integrate data from source systems for query and analysis. These systems, sometimes described as "poor man's ERP," are limited by the quality of the underlying source systems (Markus & Tanis, 2000). In 1993 Great Atlantic & Pacific Tea Company, Inc. completed a supply chain and business process infrastructure based on a "robust data warehousing capacity for category management and other grocery-specific functionality" (Retek, 2003).

Other problems identified with implementation of ERP include time, expense, vendor dependence, complexity,

RISKS ASSOCIATED WITH ERP IMPLEMENTATION

Implementing ERP can be a risky proposition for firms. Brown and Vessey (2003) observe "although failures to deliver projects on time and within budget were an old IT story, enterprise systems held even higher risks - they could be a 'bet-our-company' type of failure" (Brown & Vessey, 2003).

Markus (2000) proposes 10 categories of IT related risks, all of which would apply to ERP systems:

Financial risk	Non-use, underuse, misuse risk
Technical risk	Internal abuse
Project risk	External risk
Political risk	Competitive risk
Contingency risk	Reputational risks

"IT-related risk includes anything related to IT that could have significant negative effects on the business or its environment from the perspective of an executive investing in IT" (Markus, 2000).

Some firms may be averse to the risks an ERP implementation can create. Scott (2003) discusses some of the risks identified by Markus (2000). He describes project risks, information systems risks, organizational risks, and external risks in ERP implementations.

Project risks stem from the customization of purchased packages and the difficulty of interfacing with legacy systems. When firms believe their business process are unique, they may customize ERP software instead of adopting best practices imbedded in a standard implementation. Data conversion can also be a problem when firms do not clean up their data before embarking on a project. After implementing SAP, Halliburton reported that inventory accuracy stood at less than 80% (Anderson, 2003). Project leadership, limiting project scope, avoiding customization, and a phased implementation (rollout) can minimize this risk (Scott, 2003).

Information systems risks arise from system performance problems ERP systems may be poorly configured or the hardware may need upgrading. Another risk arises when the use of multiple vendors creates the need for multiple interfaces. Multiple vendors contributed to the problems in the Hershey Food Corporation implementation. Information systems risks can be minimized by avoiding customization, use of data warehousing for reports and queries and avoiding multivendor implementations (Scott, 2003).

Organizational risks of a bad ERP implementation can impact the firm's operating profits. Customer deliveries can be delayed putting customer relationships at risk. Impacts can be with customers, financial performance, or internal business objectives. Organizational risks can be minimized with training and strong leadership, which assures that sufficient resources are allocated to the project and inspires employees who may resist the implementation (Scott, 2003).

External risks center on litigation associated with the implementation. Firms with implementation problems may sue consultants and/or ERP vendors. Overbilling by consultants and use of incompetent trainees have been sources of litigation (Scott, 2003). Gore-Tex claims its consultant promised expert staff and delivered incompetent trainees. Managing consultants, by specifying goals and individual competence of consultants, can minimize this risk (MacDonald, 1999).

Political risk occurs "if a dominant coalition attempts to use the ERP package as a means by which to impose its views on other functional areas" (O'Gorman, 2004). A case study at an international oil supply company where the ERP implementation was dominated by the financial management of the business left the supply chain management function without the tools they believed they needed (Bradley, 2004).

Another reason for non-adoption may be that a standard software package available to all potential purchasers may reduce a firm's competitive advantage. A resource-based view of the firm assumes that the individual firm's unique collection of resources and capabilities are a potential source of competitive advantage. Capabilities leading to competitive advantage may be embedded in current business processes.

To create competitive advantage such capabilities must be valuable, rare, costly to imitate and non-substitutable. "Valuable and rare organizational resource can only be sources of *sustained* competitive advantage if firms that do not possess these resources cannot readily obtain them" (Barney, 1991). An off-the-shelf package may be costly, but would not be rare or costly to imitate. Adoption of ERP packages based on "best practices" may cause the loss of the unique and valuable advantage imbedded in current business processes. A case study showed that:

…the introduction of SAP-specific business routines can threaten established core, enabling and supplemental capabilities and related knowledge sets. The integration of SAP's embedded business routines and reporting functionality contributed to the creation of (a) highly rigid reporting structures; (b) inflexible

managerial decision-making routines; and (c) reduced autonomy on the factory floor….(Butler & Pyke, 2004)

WELL PUBLICIZED FAILURES AND PROBLEMS

Numerous descriptions of ERP failures have appeared in the business press. The experience of serious problems at many well run, well-financed firms may be enough to discourage some firms from beginning an ERP implementation.

Hershey Foods embarked on an ERP investment in mid-1996 to solve its Y2K problems and improve its ability to perform just-in-time store deliveries to its customers (Severance & Passino, 2002). After spending $112 million on an ERP project, Hershey Foods Corporation was unable to fill Halloween candy orders in October 1999, resulting in a 19% drop in third quarter profits (Stedman, 1999). As a result of Hershey's problems its stock price fell by a third and the firm lost market share to Mars and Nestle (Severance & Passino, 2002). Hershey estimates it suffered a 3% permanent decrease in market share from this experience (Sutton, 2003).

A study by the PA Consulting Group found that "92% of companies are dissatisfied with results achieved to date from their ERP implementation and only 8% achieved a positive improvement in their performance" ("ERP Implementation Disappoints Companies," 2000).

Davenport (1998) identifies several unsuccessful implementation efforts:

- Fox-Meyer Drug claims that an ERP system led to its bankruptcy.
- Mobil Europe spent hundreds of millions on ERP, but abandoned the project when a merger partner objected.
- Dell found that its ERP system did not support its new decentralized management style.
- Applied Materials gave up on its ERP implementation when it became overwhelmed with the organizational changes it required.

ERP success stories receive much less publicity.

FUTURE TRENDS

Recently, the Gartner Group coined the term "ERP II" to describe a shift in ERP from an enterprise information base to moving information across the supply chain ("Taking the Pulse of ERP," 2001). New risks and challenges will be faced by organizations opening up information systems to supply chain partners. Supply chain partners could be potential competitors or pass information to existing competitors. Resolving trust issues will be key to advancement in ERP II. Gartner does not expect to see ERP II systems fully deployed before 2005 ("Taking the Pulse of ERP," 2001).

Small and mid-sized enterprises (SME) will become leading adopters of ERP system as the large company market becomes saturated. Vendors have been developing less expensive versions of their software to appeal to the SME market.

CONCLUSION

ERP implementation projects continue to present risks to adopting organizations. Continuing ERP spending demonstrates that most organizations have concluded that the benefits resulting from such implementations outweigh the substantial risks and cost of ERP systems. Perhaps the risks of not adopting ERP are determined to be greater than the risks faced by adopters. The prospect of extending ERP beyond organizational boundaries to supply chain partners makes ERP even more attractive and possibly more risky. Organizations will continue to adopt ERP as a strategic necessity to remain competitive in their industry, but few, if any, will gain any sustainable competitive advantage by adopting ERP.

REFERENCES

Anderson, A. (2003). *When Closeness Counts*. Retrieved Dec. 23, 2003, from *http://www.datasweep.com/ds/2003/article_2003.asp?page_id=newsln_002print*

Barney, J. B. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, *17*(1), 99-120.

Bhattacherjee, A. (2000). Beginning SAP R/3 implementation at Geneva Pharmaceuticals. *Communications of the Association for Information Systems, 4*(2).

Bradford, M., & Mayfield, T. (2001). Does ERP fit in a LEAN world? *Strategic Finance, 82*(11), 28-34.

Bradley, J. (2004). *Enterprise resource planning systems success: A management theory approach to critical success factors.* Unpublished PhD, Claremont Graduate University, Claremont, CA.

Brown, C. V., & Vessey, I. (2003). Managing the next wave of enterprise systems: Leveraging lessons from ERP. *MIS Quarterly, 2*(1), 65-77.

Butler, T., & Pyke, A. (2004). Examining the influence of ERP systems on firm-specific knowledge assets and capabilities. In F. Adam & D. Sammon (Eds.), *The enterprise resource planning decade: Lessons learned and issues for the future* (pp. 167-206). Hershey, PA: Idea Group Publishing.

Cotteleer, M. J. (2002). *An empirical study of operational performance convergence following enterprise-IT implementation* (Working Paper No. 03-011). Harvard Business School.

Davenport, T. H. (1998). Putting the enterprise into the enterprise system. *Harvard Business Review, 76*(4, July-August), 121-131.

ERP implementation disappoints companies. (2000, August 31). *Australian Banking & Finance, 9,* 8.

Lucas, H. C., Jr., Walton, E. J., & Ginsberg, M. J. (1988). Implementing packaged software. *MIS Quarterly* (December 1988), 537-549.

MacDonald, E. (1999, November 2). W. L. Gore alleges PeopleSoft, Deloitte botched a costly software installation. *The Wall Street Journal,* p. 14.

Markus, M. L. (2000). Toward an integrative theory of risk control. In R. Baskerville, J. Stage & J. I. DeGross (Eds.), *Organizational and social perspectives on information technology* (pp. 167-178). Boston, MA: Kluywer Academic Publishers.

Markus, M. L., & Tanis, C. (2000). The enterprise experience - From adoption to success. In R. W. Zmud (Ed.), *Framing the domains of IT research: Projecting the future through the past*. Cincinnati, OH: Pinnaflex Educational Resources, Inc.

Newell, S., Huang, J. C., Galliers, R. D., & Pan, S. L. (2003). Implementing enterprise resource planning and knowledge management systems in tandem: Fostering efficiency and innovation complementarity. *Information and Organization, 13*, 25-52.

O'Gorman, B. (2004). The road to ERP - Has industry learned or revolved back to the start? In F. Adam & D. Sammon (Eds.), *The enterprise resource planning decade: Lessons learned and issues for the future* (pp. 22-46). Hershey, PA: Idea Group Publishing.

Pozzebon, M. (2000). *Combining a structuration approach with a behavioral-based model to investigate ERP usage*. Paper presented at the AMCIS 2000, Long Beach, CA.

Rashid, M. A., Hossain, L., & Patrick, J. D. (2002). The evolution of ERP systems: A historical perspective. In F. F.-H. Nah (Ed.), *Enterprise resource planning solutions & management* (pp. 35-50). Hershey, PA: IRM Press.

Retek. (2003). *A&P completes supply chain/business process initiative: IBM and Retek deliver enterprise merchandising solutions*. Retrieved March 13, 2004, from *http://www.retek.com/press/press.asp?id/id=507*

Scott, J. (2003). What risks does an organization face from an ERP implementation? In D. R. Laube & R. F. Zammuto (Eds.), *Business driven information technology: Answers to 100 critical questions for every manager* (pp. 274-278). Stanford: Stanford Business Books.

Severance, D. G., & Passino, J. (2002). *Making I/T work*. San Francisco: Jossey-Bass.

Somers, T. M., Ragowsky, A. A., Nelson, K. G., & Stern, M. (2001). *Exploring critical success factors across the enterprise systems experience cycle: An empirical study* (Working Paper). Detroit, Michigan: Wayne State University.

Stedman, C. (1999, November 1). Failed ERP gamble haunts hershey: Candy maker bites off more than it can chew and "Kisses" big Halloween sales goodbye. *Computer World,* 1.

Sutton, S. (2003). Keynote Address, AIS Educator Annual Meeting, Copper Mountain, CO.

Taking the pulse of ERP. (2001, February) *Modern Materials Handling*, 44-51.

Watson, E. E., & Schneider, H. (1999). Using ERP in education. *Communications of the Association for Information Systems, 1,* Article 9.

KEY TERMS

Enterprise Resource Planning Systems (ERP): An off-the-shelf accounting-oriented information system that meets the information needs of most organizations. A complex and expensive information tool to meet the needs of an organization to procure, process and deliver customer goods or services in a timely, predictable manner.

ERP II: Opening up ERP systems beyond the enterprise level to exchange information with supply chain partners. ERP II extends beyond the four-walls of the business to trading partners.

IT-Related Risk: This risk includes "anything related to IT that could have significant negative effects on the business or its environment from the perspective of an executive investing in IT" (Markus, 2000).

Legacy Systems: Transaction processing systems designed to perform specific tasks. Systems that have become outdated as business needs change and the hardware and software available in the market place have improved.

Material Requirements Planning Systems (MRP): Processes that use bills of materials, inventory data and a master productions schedule to time phase material requirement, releasing inventory purchases in a manner that reduces inventory investment yet meets customer requirements.

Manufacturing Resources Planning (MRPII): Extends MRP by addressing all resources in addition to inventory. MRPII links material requirements planning with capacity requirements planning avoiding over and under shop loading typical with MRP.

B

Basic Notions on Multidimensional Aggregate Data

Maurizio Rafanelli
Italian National Research Council, Italy

INTRODUCTION

The term *multidimensional aggregate data* (MAD; see Rafanelli, 2003) generally refers to data in which a given *fact* is quantified by a set of *measures* obtained applying one more or less complex *aggregative function* (count, sum, average, percent, etc.) to row data, measures that are characterized by a set of variables, called *dimensions*. MAD can be modeled by different representations, depending on the application field which uses them. For example, some years ago this term referred essentially to statistical data, that is, data whose use is essentially of socio-economic analysis. Recently, the metaphor of the data cube was taken up again and used for new applications, such as On-Line Analytical Processing (OLAP), which refer to aggregate and non aggregate data for business analysis.

BACKGROUND

Generally there are two broad classes of multidimensional (statistical) data: microdata and macrodata. The former refers to SDBs containing elementary or raw data, that is, records of individual entities or events. The latter refers to databases containing multidimensional aggregate data, often shown as statistical tables, that result from the application of aggregate functions on raw data. Microdata are generally stored as relations in a relational database, and when an aggregation function is applied to them, the result is a complex data, called *macrodata*, which consists of a descriptive part and a summary part. The latter is called *summary attribute* or measure, and is characterized by the descriptive part mentioned above called *metadata*. Its simplest definition is "data describing data."

In order to model aggregate data, we define these data from both a conceptual and a logical point of view. The main difference between them is that in the case of conceptual data, which we will call *multidimensional aggregate data* (MAD), we do not consider their physical storage, while in the case of logical data, which we will call the *multidimensional aggregate data structure* (MADS), we refer explicitly to their physical storage. It is useful to remember that the *multidimensionality* concept was introduced by Shoshani and Wong (1985) describing a MAD as a mapping from the domains of the category attributes (independent variable) to the (numerical) domains of the summary attributes (dependent variable). Each category attribute often represents a level of a hierarchy present in that dimension of that MAD and ranges over an n-dimensional space (the space of the n-tuples of category attribute instances), from which derives the concept of *multidimensionality*.

We give now some definitions useful in describing the aggregation process, that is, the process that allows one to obtain the multidimensional *aggregate* database from a relational *disaggregate* database.

BASIC NOTIONS

Let Θ be the database universe, that is, the set of all the relations that form the very large relational database in which raw data (microdata) are stored. Let $\mathcal{R}$ be the subset of Θ relative to all the relations used in the definition of the multidimensional aggregate (macro) database and which, therefore, refers to all the phenomena studied. Note that each phenomenon consists of one or more *facts* that are the physical objects stored in the database. Let $\mathcal{R}_x = \{\mathcal{R}_i\}_{i=1,\ldots,h}$ be the set of all the relations $\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_h$ (each of them with attributes different in number and names), which refer to the x-th phenomenon. Let $A^1_1, A^1_2, \ldots, A^1_{k_1}$ be the set of attributes of the relation $\mathcal{R}_1$, where the superscript refers to the index which characterizes the considered relation, k_1, is the number of attributes of this relation (i.e., its cardinality), each of which has a definition domain $\Delta^i_1, \Delta^i_2, \ldots, \Delta^i_{k_1}$, and likewise for the other relations. To clarify how the subsets of $\mathcal{R}$ to be aggregated are characterized, let us analyze the concept of the *category attribute*. A category attribute is the result of an abstraction on one or more attributes of the microdata; analogously its *instances* are the result of an abstraction on the (numerical, Boolean, string, etc.) values actually associated with the single microdata.

Definition 1. Let $\mathcal{R}$ be the set of all the relations used in the definition of a multidimensional aggregate database, let Ω be the set of all the attributes which appear in $\mathcal{R}$, let

$A_x \in \Omega$ be a generic attribute of this database, and let a_{xy} be one of its instances (with y = 1,…, k, where k is the cardinality of the definition domain of A_x). The logical predicate ($A_x = a_{xy}$), defined on the microdata of $\mathcal{R}$, is called *base predicate*.

Definition 2. The *base set* of the base predicate ($A_x = a_{xy}$) is the subset of Θ consisting of all microdata which satisfy the base predicate. In the following such a subset will be denoted by $B_{A_x} = a_{xy}$.

Let A be the subset of all the attributes of Ω that will become descriptive (or category) attributes or measures of all the MADs that will form the multidimensional aggregate database at the end of the aggregation process. Then $\mathcal{A}$ is the set of all and only the attributes that describe all the facts that appear in the multidimensional aggregate database. Many of these attributes appear in different relations of $\mathcal{R}$. Different attributes can contribute to form one *hierarchy*. Different hierarchies can belong to the same *dimension*, on the condition that pairs of hierarchies have at least one attribute in common. Note that parallel hierarchies, called *specialization hierarchies*, can exist. Moreover, other attributes, which do not appear in $\mathcal{A}$, can complete the hierarchies mentioned above (on the condition that the relationship between them and the other attributes of the same hierarchy is defined). $\mathcal{A}^*$ is the set of these last attributes plus the attributes of $\mathcal{A}$. We call these hierarchies *primitive hierarchies* because all the hierarchies that refer to one of them are included in it. Analogously, we call *primitive dimension* the dimension which includes all its primitive hierarchies.

Let $\mathcal{H}$ be the set of all the hierarchies (including the specialized hierarchies) defined in $\mathcal{A}^*$. Let $\mathcal{D}$ be the set of all the dimensions defined in $\mathcal{A}^*$ (which can consist of different hierarchies). Note that the users often give the name of a dimension to descriptive variables of a MAD which are, in reality, levels of a hierarchy relative to this dimension. Let Δ be the set of all the definition domains (i.e., of all the instances) of the attributes of $\mathcal{A}$, and let Δ* be the set of all the definition domains of the attributes of $\mathcal{A}^*$ which also include all the possible instances that each attribute can assume (therefore, also including the instances not present in the relations of Θ). We call *primitive domains* these definition domains. This means that all the attributes (and all the relative instances) which appear in the multidimensional aggregate database are part of $\mathcal{A}^*$ and Δ* respectively.

Category attributes are not the only metadata of multidimensional aggregate data: several other properties may provide a semantic description of the summary data. Among them we consider, in particular, the following:

- The *aggregation type*, which is the function type applied to microdata (e.g., count, sum, average, etc.) to obtain the macrodata (i.e., a MAD, see Rafanelli & Ricci, 1993), and which defines the *summary type* of the measure. This property must always be specified.
- The *data type*, which is the type of summary attribute (e.g., real, integer, non-negative real, non-negative integer).
- The *fact* $\mathcal{F}_j$ described by the multidimensional aggregate table considered (e.g., production, population, income, life-expectancy).
- Other properties may be missing, for example "data source" (which may be unknown), "unit of measure," and "unit of count," as defined in the following.

Let Γ be the set of the *functional dependencies* which are possibly present in the multidimensional aggregate database and which, therefore, exist among groups of attributes. Given a phenomenon x and given the set of relations $\mathcal{R}_x \subset \mathcal{R}$, we consider the subset of $\mathcal{R}_x$ formed only by the relations involved in the building of the fact $\mathcal{F}_j$. We call this subset an *aggregation relation*, and denote it by $\mathcal{R}_j^x$, where $\mathcal{R}_j^x = \{\mathcal{R}_{j,1}, \ldots, \mathcal{R}_{j,s}\}^x$. Every fact $\mathcal{F}_j$ has its own descriptive space formed by s category attributes (where s is the cardinality of the j-th fact), which are a subset of all the attributes in the relations $\mathcal{R}_j^x$. We denote the set of the above-mentioned s category attributes by $A_j^x = \{A_{j,ks}\}^x = \{A_{j1}, \ldots, A_{js}\}^x$. We call the relation β_j^x, formed by these attributes, a *base relation* of the fact $\mathcal{F}_j$.

The measure values are the result of the aggregation process, i.e., of the application of the aggregation function to the base relation of the fact. The fact obtained by this aggregation process is called *base fact*, because its representation cannot even be disaggregated (i.e., only more aggregate views can be obtained). Each fact consists of a set of materialized views, obtained by applying different operators of aggregation (roll-up, group-by), or of reduction of the definition domains of its category attributes (dice). This set of materialized views defines the *lattice* of this fact. The source of this lattice is formed by the total of all the summary category instances of the base fact, and the sink formed by all the summary category instances at the lowest level of disaggregation.

Let $\mathcal{F} = \{\mathcal{F}_j\}$ be the set of all the *fact names* described by the multidimensional aggregate database. Let $\mathcal{S} = \{\mathcal{S}_j\}$ be the set of all the subjects described in the facts, in other words, the "what is" of the summary attributes (Cars, People, Fruit, Workers, Dollars, etc.). Let $\mathcal{R}_j^x = \{\mathcal{R}_{j,1}, \ldots, \mathcal{R}_{j,s}\}^x$ be the subset of the relations in the microdatabase which are involved in the x-th fact. Let $A_j^x = \{A_{j,ks}\}^x = \{A_{j1},$

..., $A_{js}\}^x$ be the set of attributes of $\mathcal{R}_j^x$ which are the only ones considered in the building of this MAD.

Definition 3 Let $\{\mathcal{R}_1, \ldots, \mathcal{R}_s\}^j_x$ be the set of all the relations involved in the building of the generic j-th fact $\mathcal{F}_j$ of the x-th phenomenon π_x. These relations are included in $\mathcal{R}_x$ and have *all* the category attributes $A_j^x = \{A_{j,ks}\}^x$ of the fact $\mathcal{F}_j$ simultaneously present in it (and possibly other attributes which are not used in this aggregation process). "j" characterizes the different category attributes of $\mathcal{F}_j$. We call the relation (in non-normal form) formed by all the tuples of the previous set $\{\mathcal{R}_1, \ldots, \mathcal{R}_s\}^j_x$ *aggregation relation* $\mathcal{R}_j^x$ of the j-th MAD which describes the fact $\mathcal{F}_j$.

Definition 4. We call *aggregation schema* $A_i = \{A_{ij}\}$ of the i-th MAD, with the aggregation relation $\mathcal{R}_i$ which describes the fact $\mathcal{F}_i$, the set of all the names of the category attributes A_{ij} of the fact $\mathcal{F}_i$. This set defines the dimensions of the MAD. The number of these category attributes defines the *cardinality* of the MAD.

Definition 5. The *base relation* $\mathcal{R}_{Bi}$ of the i-th MAD (with its predefined descriptive space), which describes fact $\mathcal{F}_i$, is the subset of the aggregation relation $\mathcal{R}_i^x$ of this fact which has all and only the descriptive attributes A_{Bij} of the fact $\mathcal{F}_i$.

Definition 6. The *base schema* A_{ij} of the i-th MAD is the tuple of the base relation attribute names of the MAD.

Now we can describe the aggregation process, starting from a relational database of raw data Θ.

THE AGGREGATION PROCESS

The *first step* consists of defining the set $\mathcal{R}$ of all possible microdata relations to which the aggregation process must be applied, and, within it, to choose the set $\mathcal{R}_x$ of relations that refer to a given phenomenon πx.

Defined the set Ω of all the attributes which appear in $\mathcal{R}$, the *second step* consists of choosing attributes that will become descriptive (or category) attributes in the multidimensional aggregate database, that is, set $\mathcal{A}$, and of defining the possible functional dependencies, as well as the attributes that possibly complete the eventual existing hierarchies.

The *third step* consists of recognizing all the attributes that belong to the same dimension, and the hierarchies (and their specializations) within this dimension, i.e., sets $\mathcal{H}$ and $\mathcal{D}$. All the attributes which do not appear explicitly but which complete a given hierarchy have to be inserted into it. For example, the hierarchy of the dimension "space," i.e., Region → Province → City, is connected to the specialization hierarchies Region (of residence) → Province (of residence) → City (of residence).

At this point, the part of the multidimensional aggregate database which refers to the described phenomenon can be built repeatedly by performing the following steps:

Fourth step: Selection of the subset $\mathcal{R}_i$ of relations that are involved in the i-th fact, and of the set A^i_{kn} of attributes of $\mathcal{R}_i$ which are the only ones considered in the building of the MAD. At the end of the aggregation process, all the sets A_j of attributes will form the descriptive space of the multidimensional aggregate database. In this step we also define the subsets in $\mathcal{D}$ and in $\mathcal{H}$ which characterize respectively the dimensions and the possible hierarchies in each dimension of the fact. Therefore, a hierarchy represents the same dimension at different granularity levels.

An innovative feature of multidimensional aggregate data has been the introduction of a third type of attribute, namely the *implicit category attribute*, which can considerably enlarge and enhance the manipulation capabilities of an aggregate database. In fact, if one or more attributes which define the dimensions of a MAD have a definition domain formed by only one instance, we transform each of them into part of the "fact name" and call it "implicit attribute" or *implicit dimension*. Note that also implicit attributes contribute to MAD cardinality.

Fifth step: Application of the aggregation function to the attributes of the relations in A_j. The result of this operation is the numeric data which represent the measure carried out on microdata, called simply *measure* (or summary attribute.) Depending on the type of function applied, the parameters that characterize each fact have to be defined, including:

- the *summary type* which is defined by the aggregate function type applied to microdata (a count, a sum, etc.);
- the *count unit*—suppose that the result instances are 100, 213, 47, and so forth, and suppose also that the subject is "fruit." The count unit defines if, for example, 100 really means 100 (100 *x 1*), or 100,000 (100 *x 1,000*), or 100,000,000 (100 *x 100,000*), and so on;
- the *measure unit*—in the previous example, the number could be "number of pieces of fruit," or "kilograms," or "tons," and so forth; and
- the *data source* (this information is not always available).

In general, the aggregation function applied is initially a *count* or a *sum*. Subsequent aggregations can be obtained by applying algebraic operations and statistical-mathematic operations to these aggregate data.

Sixth step: Definition of the fact name (name of the MAD). At the end of the aggregation process all the names of the *facts* defined will form the set $\mathcal{F}$.

Seventh step: Definition of the subject described in the fact, by choosing it from among the attributes of the relations in $\mathcal{A}$*. At the end of the aggregation process, all the subjects described in the facts will form set $\mathcal{S}$.

Eighth step: Definition of possible *Notes* which characterize possible anomalies of the MAD (this step can be lacking).

When we apply the aggregation process, all the MADs produced are conceptual structures, that is, for each of them we have not yet defined how to store them in the database, or any order among the attributes of a fact or among the domain instances of each attribute. To store them we have to define these orders.

The result of this further step, the *ninth step*, is the definition for each MAD of a corresponding logical structure, called *multidimensional aggregate data structure* (MADS). In it each instance of the measured data is characterized exactly by a tuple, whose elements are defined by one instance for each dimension (category attribute) of the MAD.

Definition 7. An *aggregation process* is formally described by the six-tuple

$$< \mathcal{P}, \mathcal{R}_i, A_i, \mathcal{S}_i, \mathcal{N}, f_{agg} >,$$

where:

$\mathcal{P}$ is a fact name which identifies a fact universe $\mathcal{R}$ through a mapping φ; this mapping is defined from the fact set $\mathcal{F}$ to the powerset of Θ.

$\mathcal{R}_i$ is a set of relations involved in the building of the i-th MAD.

A_i is a set of the descriptive (category) attributes $\{A_{i,1}, A_{i,2}, \dots, A_{i,k}\}$ of $\mathcal{R}_i$ (with its own instance domains $\Delta(A_{i,j})$, with $j = 1,\dots, k$) on which to apply an aggregation function.

$\mathcal{S}_i$ is the *subject* which characterizes the measure of the MAD.

$\mathcal{N}$ is a numerical domain.

$f_{agg}: \mathcal{R}_i \rightarrow \mathcal{S}_i$ is an aggregation function (initially, in general, a sum or a count).

A MAD is, therefore, a concise representation of the aggregation process result.

Definition 8. Let H and K be two sets whose elements are sets; we define the *cross-intersection* of H and K, and denote it by $H \otimes K$, as the set of all the possible intersections between each element in H and each element in K:

$$H \otimes K = \{ j : j = h \cap k, h \in H, k \in K \}$$

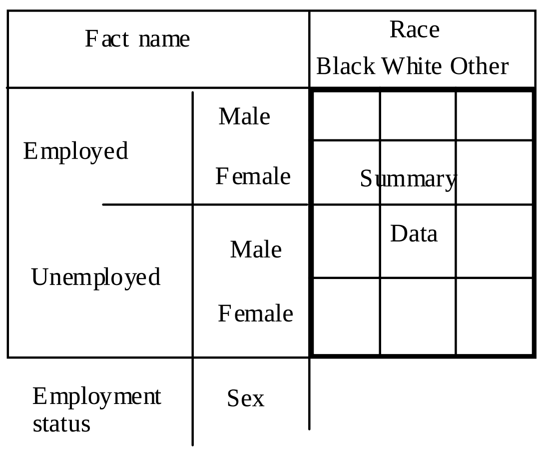
Now we describe the data structure of multidimensional aggregate data by a simple graphical representation.

THE DATA STRUCTURE

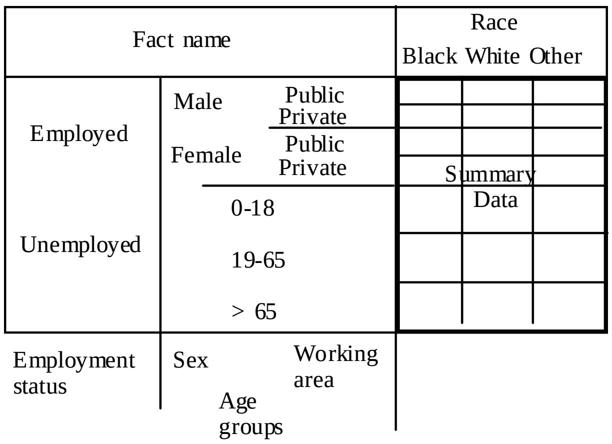
A measure of a MAD is described by a set of *dimensions*, that is, a set of category attributes; each of them can be a level of one of the possible hierarchies that form the primitive dimension. Each category attribute also has a *primitive category attribute*. The complexity of aggregate data is also due to the fact that the same operator may require different recomputation functions for the summary values, depending on the summary type of the multidimensional aggregate table, as illustrated in Rafanelli and Ricci (1993). Summary data are always fixed in time. It is always (implicitly or explicitly) present in every MAD, as well as the *space* dimension.

Most of the existing models for aggregate data represent a MAD as a mapping between category and summary attributes. The analysis of the aggregation process shows that a MAD represents a functional link between *aggregation sets* and *summary values*, rather than between tuples of category attribute instances and summary values. In Bezenchek, Rafanelli, and Tininini (1996), the formal definitions of a *simple, complex, and composite* MAD (and of the analogous MADS) are given. In the following figure we propose the tabular representation of a *simple, complex, and composite* MADS. A subtle distinction has to be made between the concepts of multidimensionality and of *polidimensionality*. The former refers to a data structure in which measured data is described by different parameters which define a multidi-

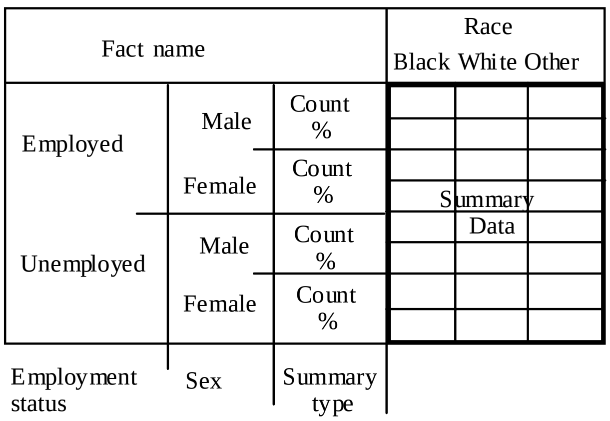
Simple MADS



Complex MADS



Composite MADS



mensional descriptive space. The latter refers to a descriptive space in which a category attribute has, for example, one part of its definition domain classified by one or more other category attributes, A, B, ... C, while the other part is classified by one or more different category attributes, E, F, ... H. Note that the number of the first and the second group of category attributes can be different. This situation exists in the complex MADS shown in the following. The difference between the simple/complex MADS and the composite MADS consists in their summary type, which is unique in the former and of at least two different types in the latter (in our case, Count and Percent).
Recently other proposals for modeling multidimensional databases have been presented, based on the cube metaphor, which represents the concept of multidimensionality by a three-dimensional representation (see Agrawal, Gupta & Sarawagi, 1997; Gyssens & Lakshmanan, 1997; Nguyen, Tjoa & Wagner, 2000; Pedersen & Jensen, 1999). Codd (1993) proposed the term On-Line Analytical Processing (OLAP) for rendering enterprise data in multidimensional perspectives, performing online analysis of data using mathematical formulas or more sophisticated statistical analyses, and consolidating and summarizing data according to multiple dimensions.

The new data models for OLAP typically categorize data as being "measurable business facts (measures) or dimensions." In OLAP research, most work has concentrated on performance issues. Techniques have been developed for deciding what subset of a data cube to precompute, for estimating the size of multidimensional aggregates and for indexing pre-computed summaries. Similarities and differences between OLAP and statistical databases are discussed in Shoshani (1997).

CONCLUSION

The MADS are data used all days by a large variety of users. The problems that arise in their manipulation are studied in different areas such as statistical data, geographical data, data warehouses, OLAP, and so forth. In this article, we gave some base concepts on their data structure and on the aggregation process that allows one to obtain them. Many other topics referring to multidimensional databases are treated in Rafanelli (2003).

REFERENCES

Bezenchek, A., Rafanelli, M., & Tininini, L. (1996). A data structure for representing aggregate data. *Proceedings of the 8th International Conference on Scientific and Statistical Database Management* (8th SSDBM) (pp. 22-31). Stockholm: IEEE Press.

Codd, E.F. (1993). *Providing OLAP (On-Line Analytical Processing) to user-analysts: An IT mandate.* Technical Report. E.F. Codd and Associates.

Gray, J., Chaudhuri, S., Bosworth, A., Layman, A., Reichart, D., Venkatrao, M., Pellow, F., & Pirahesh, H. (1997). Data cube: A relational aggregation operator generalizing group-by, cross-tab and sub-totals. *Journal of Data Mining and Knowledge Discovery*, *1*(1), 29-54.

Gyssens, M., & Lakshmanan, L.V.S. (1997, August 25-29). A foundation for multi-dimensional databases. *Proceedings of the 23rd International Conference on Very Large Data Bases* (VLDB'97) (pp. 106-115), Athens, Greece.

McCarthy, J.L. (1982). Metadata management for large statistical databases. *Proceedings of the 8th International Conference on Very Large Data Bases* (pp. 234-243), Mexico City, Mexico.

Nguyen, T.B., Tjoa, A.M., & Wagner, R. (2000). Conceptual multidimensional data model based on metacube. *Proceedings of the 1st International Conference on Advances in Information Systems* (ADVIS 2000) (pp. 24-33), Izmir, Turkey. Lecture Notes in Computer Science, 1909. Berlin: Springer-Verlag.

Pedersen, T.B., & Jensen, C.S. (1999). Multidimensional data modeling for complex data. *Proceedings of the 15th International Conference on Data Engineering* (ICDE'99) (pp. 336-345), Sydney, Australia. IEEE Computer Society Press.

Pedersen, T.B., Jensen, C.S., & Dyreson, C.E. (2001). A foundation for capturing and querying complex multidimensional data. *Journal of Information Systems*, *26*(5), 383-423.

Rafanelli, M. (2003). *Multidimensional databases: Problems and solutions*. Hershey, PA: Idea Group Publishing.

Rafanelli, M., & Ricci, F.L. (1993). Mefisto: A functional model for statistical entities. *IEEE Transactions on Knowledge and Data Engineering, 5*(4), 670-681.

Shoshani, A. (1997). OLAP and statistical databases: Similarities and differences. *Proceedings of the 16th ACM Symposium on Principles of Database Systems* (PODS'97) (pp. 185-196), Tucson, Arizona.

Shoshani, A. & Wong, H.K.T. (1985). Statistical and scientific database issues. *IEEE Transactions on Software Engineering, SE-11*(10), 1040-1047.

Tininini, L., Bezenchek, A., & Rafanelli, M. (1996). A system for the management of aggregate data. *Proceedings of the 7th International Conference on Database and Expert Systems Applications* (DEXA'96) (pp. 531-543), Zurich, Switzerland. Lecture Notes in Computer Science, 1134. Berlin: Springer-Verlag.

KEY TERMS

Aggregation Process: The process that allows one to obtain multidimensional *aggregate* data from *disaggregate* data.

Category Attribute: The variable that describes the summary data of an aggregate data structure.

Data Multidimensionality: The set of dimensions of a table or a data cube.

Dimension: Corresponds to a perspective under which facts can be fruitfully analyzed. It is a structural attribute of a fact, that is, a list of members (variables), all of which are of a similar type in the user's perception of the data.

Fact: An entity of an application that is the subject of decision-oriented analysis. It is usually represented graphically by means of a table or, using a metaphor, by a data cube.

Hierarchy: An ordered structure, where the order is established between individuals or between classes of individuals; the ordering function may be any function defining a partial order.

Metadata: Data which describe other data.

Multidimensional Aggregate Data: Data in which a given *fact* is quantified by a set of *measures*, obtained applying one more or less complex *aggregative function* (count, sum, average, percent, etc.) to row data.

Summary Attribute: The numerical result of the application of *aggregative functions* to row data. It is described by the set of category attributes that form the multidimensional aggregate data structure (table or data cube).

Basics of the Triune Continuum Paradigm

B

Andrey Naumenko
Triune Continuum Enterprise, Switzerland

INTRODUCTION

This article reviews the Triune Continuum Paradigm - a logically rigorous theoretical base for organization of modern conceptual frameworks that are used for system modeling in different contexts (e.g., in software development, in enterprise architecture, in the architecture of financial services, in jurisprudence, etc.). This paradigm is an important contribution to the system modeling domain, because currently none of the prevailing system modeling frameworks has a satisfactory formal theoretical foundation.

The absence of a theoretical foundation for modeling frameworks leads to the practical application experiences where modelers are constrained to be guided by chance and not by a founded reason. This often leads to the inadequate choices of modeling frameworks; that is, to the situations where a chosen modeling framework is not designed to deal with the targeted modeling problems. Possible consequences of such choices include: incorrect (e.g., inadequate with regard to the requirements) information systems specifications, contradictory data architectures, incomplete service specifications, and so forth – all of these being the decisive contributions to failures of many projects. The paradigm that we review in this article fixes this problem, providing missing theoretical foundations for frameworks positioned in the domain of general system modeling.

Many of the existing system modeling frameworks appeared as an integration of the best modeling practices. The reviewed paradigm does not repudiate the practical experience that was gathered by these different frameworks, but fixes its inconsistencies and complements it, supporting with logically rigorous theoretical foundations. Therefore the paradigm brings a significant constructive potential to the evolution of modern system modeling frameworks. This potential could be realized if people responsible for the design of modeling frameworks and tools would heed the proposed paradigm.

BACKGROUND

The Cambridge Dictionary of Philosophy (Audi, 1999) provides the following definition of the term “paradigm”: “Paradigm, as used by Thomas Kuhn (The Structure of Scientific Revolutions, 1962), a set of scientific and metaphysical beliefs that make up a theoretical framework within which scientific theories can be tested, evaluated and if necessary revised.”

In practice, a paradigm is usually defined for a collection of sciences. In this context a paradigm introduces and justifies a set of basic assumptions and principles on which any of the sciences from the collection can rely as on their foundations. Then, starting from the principles provided by a paradigm, different sciences build their specific frameworks of knowledge. And if some sciences share the same paradigm then they can bind and synchronize their specific frameworks of knowledge. By doing so they can mutually enrich each other with the knowledge obtained from the different (but consistent with regard to the basic principles) points of view.

The Triune Continuum Paradigm (Naumenko, 2002) is a paradigm for general system modeling. Thus the Triune Continuum Paradigm serves the sciences that have diverse interests in system modeling. As any paradigm, it introduces and justifies a set of principles that provide the sciences with the necessary starting points for building their diverse conceptual frameworks of scientific knowledge; in our case - the principles that are necessary for building modeling frameworks.

THREE PRINCIPLES OF THE TRIUNE CONTINUUM PARADIGM

Triune Continuum Paradigm is composed of three principles.

The first principle is the result of application of the Tarski’s Theory of Truth (Tarski, 1956) for the case of general system modeling. This application allows the definition of coherent semantics for the concepts of a modeling framework. This is done by constructing formal descriptions for the relations between the subjects that are interesting to be modeled on one side, and the concepts that have to represent these subjects on the other side. This principle is necessary to assure the *coherency* and *unambiguity* within modeling interpretations performed using a single system modeling framework.

An application of the first principle provided by the Triune Continuum Paradigm results in a system modeling framework that features modeling terms with a coherently defined semantics in the form of Tarski's declarative semantics. The justifications of importance of this principle for the information systems modeling were presented and analyzed in details (Naumenko, Wegmann & Atkinson, 2003). In particular, it was demonstrated that Tarski's declarative semantics are:

- formally *sufficient* for the definition of the application scope of a modeling language;
- formally *sufficient* for unambiguity in coherency of interpretations within modeling representations;
- formally *necessary and sufficient* for unambiguity in adequateness of modeling representations.

The second principle of the Triune Continuum Paradigm is the result of application of the Russell's Theory of Types (Russell, 1908) for the case of general system modeling. This application defines the way to categorize concepts of a modeling framework so that in applications of this framework the concepts make up *internally consistent* structures of propositions. Thus this principle is necessary to assure the consistency of descriptions and specifications, which are constructed with the aid of the modeling frameworks.

The importance of this principle is justified by the fact that Russell's Theory of Types was formulated to resolve Russell's paradox, "the most famous of the logical or set-theoretical paradoxes" (Irvine, 2003). Thus with an application of the second principle of the Triune Continuum Paradigm, the resulting modeling framework in its own applications will produce internally consistent system specifications (i.e., system specifications that are devoid of self-contradictions).

The name of Triune Continuum Paradigm originates from the third principle of the paradigm, which is the notion of Triune Continuum, defined in Naumenko (2002). This notion introduces and justifies a minimal set of modeling concepts that are *necessary and sufficient* to cover the representation scope of the general system modeling domain on the most abstract level. This principle is necessary for different system modeling frameworks to justify the existence of their basic modeling concepts.

The idea of Triune Continuum was defined through the introduction of three continuums representing the scope of general system modeling. The first two continuums are:

- *spatiotemporal continuum* (where subjective space-time metrics are defined to be used in the subjective representations);
- *constitution continuum* (where subjective constitutional metrics are defined to be used in the subjective representations, e.g., objects defined in relation with their environments).

These two continuums are introduced in relation to each other as complements within the universal general system modeling scope. In other words: everything in the scope that is not space-time is constitution; and everything in the scope that is not constitution is space-time.

The third continuum is:

- *information continuum* (which emerges from the mutual relations of the first two continuums and contains the corresponding information about these relations, e.g., information about objects and their environments being related to the spatiotemporal intervals or to the points in space-time).

Thus the three continuums are *triune*: none of them exist without the others; either the three exist altogether, or they do not exist at all. Indeed, as soon as the first (spatiotemporal) continuum is introduced, everything in the universal scope that does not belong to the first continuum immediately shapes the second (constitution) continuum; and the third (information) continuum immediately emerges as the information about the mutual relations of the first two continuums (e.g., as spatiotemporal information about the constitution).

The third principle of Triune Continuum Paradigm is important for various system modeling frameworks that are used in diversified domains of human activity (e.g., the frameworks used to analyze, design and develop coherent structures providing useful functionalities in domains spread from jurisprudence and health care to software engineering and machine-building industries). Using the notion of Triune Continuum it is possible to introduce and justify minimal sets of modeling concepts that are necessary and sufficient for those diversified frameworks to cover their respective representation scopes.

APPLICATIONS OF THE TRIUNE CONTINUUM PARADIGM

The Triune Continuum Paradigm can be applied in practice either to improve an existing system modeling framework or to design a new system modeling framework for a given purpose. Let us mention here three of the existing applications of the paradigm:

- case of the Unified Modeling Language (UML);
- case of the Reference Model of Open Distributed Processing (RM-ODP);

- case of the Systemic Enterprise Architecture Methodology (SEAM).

The first two of the three cases illustrate the paradigm applications targeting improvements of the existing system modeling frameworks. The third case illustrates the paradigm application contributing to the design of a new system modeling framework.

Case 1: Triune Continuum Paradigm Application for UML

"The Unified Modeling Language (UML) is a language for specifying, visualizing, constructing, and documenting the artifacts of software systems, as well as for business modeling and other non-software systems" (OMG, 2003, section 1.1). UML is a proposition of the Object Management Group (OMG) that emerged from the integration of different industrial practical experiences and became an influential phenomenon in the system modeling. As a matter of fact, due to the multiple efforts of different interested parties, UML has gained a relative domination over the other modeling techniques in the current industrial practices. This is why it was interesting to apply the Triune Continuum Paradigm for the case of UML conceptual framework. Results of this application were presented to the UML research community (Naumenko & Wegmann, 2002). With the aid of the Triune Continuum Paradigm it was shown that in its current state the metamodel of UML features a number of undesirable properties, in particular:

- absence of an explicit structural organization defined for the UML metamodel;
- absence of Tarski's declarative semantics in the UML metamodel;
- absence of theoretical justifications for the UML metamodel to represent the modeling scope that is targeted by UML.

The paradigm-based solutions were presented for each of the three identified problems (Naumenko & Wegmann, 2002), providing designers of UML with the benefits of the paradigm's logical rigor, of its formal presentation and of its solid theoretical foundations.

Case 2: Triune Continuum Paradigm Application for RM-ODP

The Reference Model of Open Distributed Processing (RM-ODP) is an ISO and ITU standard for system modeling, designed to model ODP-systems (ISO/ITU 1998). The result of Triune Continuum Paradigm application for the RM-ODP case is especially interesting because it allowed accomplishing a single consistent formalization of the RM-ODP conceptual framework, providing the denotational semantics for the basic modeling and specification concepts of RM-ODP. Such formalization was officially declared as a goal of the ISO and ITU activities in the scope of RM-ODP standardization (ISO/ITU 1998). But this goal was not achieved by the standard; and so far the paradigm-based formalization remains the only solution achieving the defined objective.

The formalization was expressed in a computer interpretable form using Alloy formal description technique (Jackson, 2002). Alloy was chosen because of the public availability of the corresponding software tool, "Alloy Constraint Analyzer," which allows simulation of the instances of conceptual structures formalized with Alloy and representation of these instances in a graphical form. However, due to the nature of denotational semantics (Naumenko, Wegmann & Atkinson, 2003), any choice of the formal description technique does not change semantic interrelations within the formalization. So, another formal description technique could also be used to express the paradigm-based formalization of the RM-ODP conceptual framework in a computer interpretable form.

The paradigm-based formalization of RM-ODP presents a concrete example of formal ontology for general system modeling. Thanks to the Triune Continuum Paradigm, the metamodel that is realized by the formal ontology is internally consistent, introduces logical coherency of interpretation of a subject of modeling, defines formal semantics for the modeling concepts, and its models are verifiable with the aid of computer tools. These results were presented to the RM-ODP research community (Naumenko & Wegmann, 2001; Naumenko, Wegmann, Genilloud & Frank, 2001), and they attracted interest of the ISO/ITU committee that is responsible for the RM-ODP standardization. This provides the Triune Continuum Paradigm with a chance to influence future evolution of the ISO/ITU standard.

Case 3: Triune Continuum Paradigm Application for SEAM

The Systemic Enterprise Architecture Methodology (SEAM) is a methodology proposed by LAMS-EPFL (Wegmann, 2003) for system modeling in the domain of enterprise architecture, which is the domain that considers integration of IT systems and business systems in the context of an enterprise.

Applying the Triune Continuum Paradigm, a logically rigorous framework of concepts covering the representation scope of SEAM was designed and implemented as a specialization of the RM-ODP standard conceptual framework (ISO/ITU, 1998). Thus in this case the paradigm application provided a formal ontology for SEAM. The corresponding research results were reported to the enterprise architecture community (Wegmann & Naumenko, 2001) and provided the necessary basis for ongoing evolution of SEAM.

FUTURE TRENDS

The Triune Continuum Paradigm provides a set of theoretical foundations for different frameworks of knowledge belonging to the general system modeling domain. In most of the cases, currently existing industrial and academic frameworks for system modeling do not feature such theoretical foundations, because these frameworks are developed using so-called "best practices" approach (when results of different practical experiences of system modeling within a given domain are integrated to build a modeling framework for the domain). And if the "best practices" approach is not accompanied by theoretical foundations then it is impossible to justify a number of important properties for the resulting system modeling frameworks (e.g., to guarantee the necessity and sufficiency of a framework for its domain representation, to assure internal consistency within different pieces of practical experience integrated in a single framework, etc.).

So, the Triune Continuum Paradigm provides an indispensable contribution to the general system modeling. And the future trends should assure practical realization of the significant constructive potential that the paradigm features for those numerous system modeling frameworks that currently do not have satisfactory theoretical foundations. This potential will be realized through the paradigm applications to these concrete system modeling frameworks. Some of the examples of such applications were presented in this article.

CONCLUSION

The Triune Continuum Paradigm provides system modelers (in particular, IS modelers) with a set of principles that are essential to build adequate system modeling frameworks. These principles are based on the solid theoretical foundations discovered in the last century: Tarski's Theory of Truth was presented in 1935, while Russell's Theory of Types was formulated in 1908. The authors of these two theories, Alfred Tarski and Bertrand Russell, are recognized to be among the greatest logicians throughout the history of humankind. Thus the Triune Continuum Paradigm through its applications used in the computer-aided environment promotes the use of fundamental logical theories to the practices of regular modelers, information system designers and architects. The paradigm-based theoretically founded approaches to the information systems development make a constructive difference in the IS development projects where the usual "best practices methodologies" do not perform well enough due to a number of reasons (e.g., lack of flexibility, lack of representation possibilities, lack of internal consistency, etc.).

REFERENCES

Audi, R. (Ed.). (1999). *The Cambridge dictionary of philosophy* (2nd ed.). Cambridge University Press.

Irvine, A.D. (2003, summer). Russell's paradox. In E.N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*.

ISO and ITU. (1998). Open distributed processing - reference model. *ISO/IEC 10746-1, 2, 3, 4 | ITU-T Recommendation X.901, X.902, X.903, X.904*.

Jackson, D. (2002, April). Alloy: A lightweight object modelling notation. *ACM Transactions on Software Engineering and Methodology, 11*(2), 256-290.

Kuhn, T.S. (1962). *The structure of scientific revolutions* (3rd ed.). University of Chicago Press.

Naumenko, A. (2002, June). *Triune Continuum Paradigm: A paradigm for general system modeling and its applications for UML and RM-ODP*. PhD thesis 2581. Swiss Federal Institute of Technology – Lausanne (EPFL).

Naumenko, A., & Wegmann, A. (2001, July). *A formal foundation of the RM-ODP conceptual framework*. Technical report No. DSC/2001/040. Swiss Federal Institute of Technology – Lausanne (EPFL).

Naumenko, A., & Wegmann, A. (2002, September/October). A metamodel for the Unified Modeling Language. In J.-M. Jézéquel, H. Hussmann & S. Cook (Eds.), *LNCS 2460: Proceedings of UML 2002* (pp. 2-17). Dresden, Germany: Springer.

Naumenko, A., Wegmann, A., & Atkinson, C. (2003, April). *The role of Tarski's declarative semantics in the design of modeling languages*. Technical report No. IC/2003/43. Swiss Federal Institute of Technology – Lausanne (EPFL).

Naumenko, A., Wegmann, A., Genilloud, G., & Frank, W.F. (2001, July). Proposal for a formal foundation of RM-ODP concepts. *Proceedings of ICEIS 2001, Workshop On Open Distributed Processing - WOODPECKER 2001*, Setúbal, Portugal (pp. 81-97).

OMG. (2003, March). *Unified Modeling Language Specification, version 1.5. http://www.omg.org/uml*

Russell, B. (1908). Mathematical logic as based on the theory of types. *American Journal of Mathematics, 30,* 222-262.

Tarski, A. (1956). *Logic, semantics, meta-mathematics*. Oxford University Press.

Wegmann, A. (2003, April). On the systemic enterprise architecture methodology (SEAM). *Proceedings of ICEIS 2003,* Anger, France.

Wegmann, A., & Naumenko, A. (2001, September). Conceptual modeling of complex systems using an RM-ODP based ontology. *Proceedings of the 5th IEEE Conference - EDOC 2001*, Seattle (pp. 200-211).

KEY TERMS

RM-ODP: Reference Model of Open Distributed Processing, an ISO and ITU standard for system modeling designed to model open distributed systems.

Russell's Theory of Types: A theory proposed by British logician Bertrand Russell to resolve Russell's paradox, which appears when the set of all sets that are not members of themselves is considered in naive set theory. The paradox is that such a set appears to be a member of itself if and only if it is not a member of itself.

SEAM: Systemic Enterprise Architecture Methodology, a methodology proposed by LAMS-EPFL for system modeling in the domain of enterprise architecture (the domain that considers integration of IT systems and business systems in the context of an enterprise).

Tarski's Theory of Truth: A theory proposed by Polish logician Alfred Tarski. The theory defines the criteria for a formal definition of a true sentence; the theory allows deriving the notion of Tarski's declarative semantics for a modeling language where the modeling language terms are put in the unambiguous correspondence with the subjects of modeling interest that they represent in applications of the language.

Triune Continuum: One of the principles of the Triune Continuum Paradigm. The principle introduces three continuums (spatiotemporal, constitution and information continuums) to justify a minimal set of modeling concepts that are necessary and sufficient to cover the representation scope of the general system modeling domain on the most abstract level.

Triune Continuum Paradigm: A paradigm for general system modeling. The paradigm introduces and justifies a set of principles that provide designers and architects of system modeling methodologies with the necessary theoretical support for building their modeling frameworks. The principles are derived from the Tarski's Theory of Truth, from the Russell's Theory of Types and use the notion of Triune Continuum.

UML: Unified Modeling Language proposed by the Object Management Group (OMG) for system modeling in the domains of software systems, of business systems and others.

Bayesian Analysis of Geographical Variation in Disease Risk

C.L. Vidal-Rodeiro
Norman J. Arnold School of Public Health, USA

M.I. Santiago-Perez
Public Health Department of Galicia, Spain

E. Vazquez-Fernandez
Public Health Department of Galicia, Spain

M.E. Lopez-Vizcaino
Galician Statistical Institute, Spain

X. Hervada-Vidal
Public Health Department of Galicia, Spain

INTRODUCTION

Disease mapping is a big focus of interest in the area of public health (Bailey, 2001; Moore & Carpenter, 1999), and the geographical distribution of a disease has an important role in understanding its origin, its causes or its evolution.

In recent years, there have been many efforts to map mortality or incidence from diseases (see, for example, Lopez-Abente, Pollán, Escolar, Errezola & Abraira, 2001). The most widely used indicator in geographical representation is the standardized mortality ratio (SMR); this offers the advantage of eliminating the confounding effect of the variable by which it is adjusted, usually age, but presents certain drawbacks when the population size varies over the map (Breslow & Day, 1975). In such a case, estimators of different accuracy are obtained in each area; areas having small populations—and thus fewer cases—tend to register very extreme estimations of risk, which then dominate the map and hinder epidemiological interpretation. This is a particular problem for rare diseases where thousands of individuals are needed before a single case is expected to occur. This makes the necessity to utilize information from neighbouring areas in order to produce better estimates.

Alternative risk measures may be obtained by applying other modelling techniques that take the sources of spatio-temporal variation into account (Lawson, 2001). A simple technique consists of adjusting a Poisson regression model that displays a log-linear relationship between risk and space-time variables. While successful in reducing the variability in risk, this method continues to pose a number of drawbacks. First, in geographical areas having few cases, this model yields unstable estimations due to extra-Poisson variation. Furthermore, if the hypothesis of spatial independence between risks does not hold, the model is not appropriate, as it takes no account of a possible correlation between areas.

A possible solution to these problems is the Bayesian extension of the model introduced by Clayton and Kaldor (1987) and further developed by Besag, York, and Mollié (1991). Basically, this approach provides a way to integrate, in the estimation of the unknown relative risk, local information consisting of the observed and expected number of cases in each area and prior information on the overall variability of the relative risk, their potential similarity in neighbouring areas, and their connection with geographically defined covariates.

In this article, we compare the behaviour of the mentioned techniques with the purpose of estimating and mapping mortality relative risks in small geographical areas; this is illustrated by the analysis of the geographical variation in men lung cancer mortality in Galicia (Spain).

THEORETICAL MODELS

Classical Approach

Let O_i denote the number of observed cases, E_i the number of expected cases, calculated by using the population broken down by age for each area or geographical unit, plus the specific mortality rates, and let ξ_i be the relative risk (*RR*).

B

The classical approach to disease mapping is based on the assumption that, conditional on the E_i's being known, the ξ_i's are mutually independent. Moreover, each O_i follows a Poisson distribution with mean $E_i\xi_i$:

$$[O_i \mid E_i, \xi_i] \sim Poisson\,(E_i\xi_i)$$

Under these assumptions, the maximum likelihood estimate of ξ_i, denoted by $\hat{\xi}_i$, is the *SMR*:

$$\hat{\xi}_i = SMR_i = \frac{O_i}{E_i}$$

Hierarchical Bayesian Model

Bayesian methods estimate the risk of an area by incorporating information from adjacent areas, so as to reduce the effect of random fluctuations unrelated to the risk. Furthermore, on taking account of spatial correlation between adjoining areas, the resulting smoothed maps prove more informative (Banerjee, Carlin & Gelfand, 2004; Richardson, 2003; Wakefield, Best & Waller, 2000).

In the Bayesian approximation, Poisson variation is modelled at a first level, and a model for the relative risks is specified at a second level, with area-specific random effects further decomposed into two components: a spatially structured component that takes into account the effects that vary in a structured manner in space (clustering) and a component that models the effects that vary in an unstructured manner between one area and the next (heterogeneity).

The spatio-temporal hierarchical model is formulated as follows (Bernardinelli, Clayton, Pascutto, Montomoli, Ghislandi & Songini, 1995):

Let O_{ij} be the number of observed cases, E_{ij} the number of expected cases, and ξ_{ij} the *RR* in area *i* and period *j*. A likelihood model is specified for the vector of observed cases, given the risk vector

$$[O_{ij} \mid \xi_{ij}] \sim \text{Poisson}(E_{ij}\xi_{ij})$$

and the *RR* is modelled as

$$\log \xi_{ij} = \alpha + \phi_i + \theta_i + (\beta+\delta_i)t_j + \Sigma\,\gamma_k X_{ki}$$

where α is the mean of the logarithm for *RRs* over all areas, ϕ_i the clustering component, θ_i the heterogeneity component, t_j the time, β the mean of the time trend across all areas, δ_i the space-time interaction effect, and X_k indicate the covariates, with coefficients γ_k.

Estimation of the risk across time in each area is given by $\exp(\beta+\delta_i)$. Following the Bernardinelli, Clayton, Pascutto, Montomoli, Ghislandi, and Songini (1995) notation, δ_i is named the differential trend for area *i*: a value of $\delta_i<0$ indicates that the trend in area *i* is below the mean, while a value of $\delta_i>0$ implies that the trend in area *i* is above the mean.

Bayesian modelling requires specification of prior distributions for random effects. Several prior distributions can be considered, and next, we will describe the specifications used in the practical application.

The distribution model for the heterogeneity component is

$$[\theta_i \mid \theta_j, i\neq j, \sigma_\theta^2] \sim \text{normal}\,(\bar{\theta}_{-i}, \sigma_\theta^2)$$

where

$$\bar{\theta}_{-i} = \frac{1}{I-1}\sum_{j\neq i}\theta_j\,, \quad I = \text{number of areas.}$$

By virtue of this prior distribution, it is assumed that variation in risk between areas is independent and, as a consequence, posterior estimations of the area effect will tend towards an overall mean.

For the clustering component, the estimations of the risk in any area depend on neighbouring areas; this was achieved by allocation of weights. Specifically, weights were taken equal to one in cases where the areas were adjacent (i.e., share a common boundary) and zero in cases where they were not. The conditional autoregressive (CAR) model proposed by Besag, York, and Mollié (1991) was used:

$$[\phi_i \mid \phi_j, i\neq j, \sigma_\phi^2] \sim \text{normal}(\bar{\phi}_i, \sigma_i^2)$$

where

$$\bar{\phi}_i = \frac{1}{\sum_j w_{ij}}\sum_j \phi_j w_{ij}$$

$$\sigma_i^2 = \frac{\sigma_\phi^2}{\sum_j w_{ij}}$$

$w_{ij} = 1$ if i,j are adjacent (or 0 if they are not).

Parameters σ_θ^2 and σ_ϕ^2 control the variability of θ and ϕ. Following the recommendations of Bernardinelli, Clayton, and Montomoli (1995), gamma distributions were considered for these parameters.

In order to carry out a full Bayesian analysis, hyperprior distributions for α, β and γ_k must be specified; uniform priors (improper) were assumed for all these parameters.

Estimation of Hierarchical Bayesian Models

For the estimation of the hierarchical model simulation techniques are required; the most commonly used is the Gibbs Sampler, which belongs to the Markov chain Monte Carlo (MCMC) methods and is implemented in the software WinBUGS (Spiegelhalter, Thomas, Best, & Lunn, 2002).

A convergence analysis must be carried out in order to assess whether the Markov chain converges to the stationary distribution (for a review of convergence diagnostics, see Cowles & Carlin, 1996).

Selection of Models

To select the best model from a collection of different Bayesian models, what is used is the Deviance Information Criterion (DIC) proposed by Spiegelhalter, Best, Carlin, and Van der Linde (2002) that enables simultaneous assessment of fit and model complexity:

$$\mathrm{DIC} = D + p_D$$

where D is the deviance based on posterior means of the parameters, and p_D is the effective number of parameters, which is obtained by subtracting a plug-in estimation of the deviance, calculated using posterior means of the parameters, from the posterior deviance.

It must be stressed that DIC is not intended for identification of the correct model, but rather as a method of comparing a collection of alternative formulations.

PRACTICAL APPLICATIONS

The methods described in the previous section have been applied to the study of geographical variation in men lung cancer mortality from 1978 to 1998 in Galicia, (Spain).

Data

Galicia is a region situated in the northwest of Spain and divided into 315 municipal areas. The study period, 1978-1998, was divided into three sub-periods of seven years each: 1978-1984, 1985-1991, and 1992-1998.

Age was categorized as under 15 years, 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, and over 74 years. Since very few deaths from lung cancer occur below age 44, we combine the first four categories.

Lung cancer mortality data in males, which consists of the number of lung cancer deaths in municipality i for time period j (i=1, …, 315 and j=1,2,3) were obtained from the Spanish National Statistics Institute for the period 1978-1986, and from the Galician Mortality Register for the period 1987-1998. Galician age-specific men population figures were obtained from the 1981 and 1991 Censuses and the 1986 and 1996 Population Registers.

Description of Covariates

The covariates we have considered are the total resident population per square kilometre (DEN) as a proxy for urban living, and a socio-economic score (SES), both at municipality level and referred to 1991. The SES, obtained using the methodology proposed in Silva-Ayçaguer (1997), is a summary of the following variables: infant mortality rate, population density, proportion of people that live in population centres with more than 1,000 inhabitants, proportion of men with high studies, proportion of active people in industrial activities, and proportion of people in households without running water or electricity.

Results

Before estimating and interpreting any results, the sensitivity of the conclusions to changes in Bayesian model specifications was investigated by using DIC. The best model in terms of DIC was the model that included the space-time interaction effects, without the heterogeneity term:

$$\log \xi_{ij} = \alpha + \phi_i + (\beta + \delta_i) t_j$$

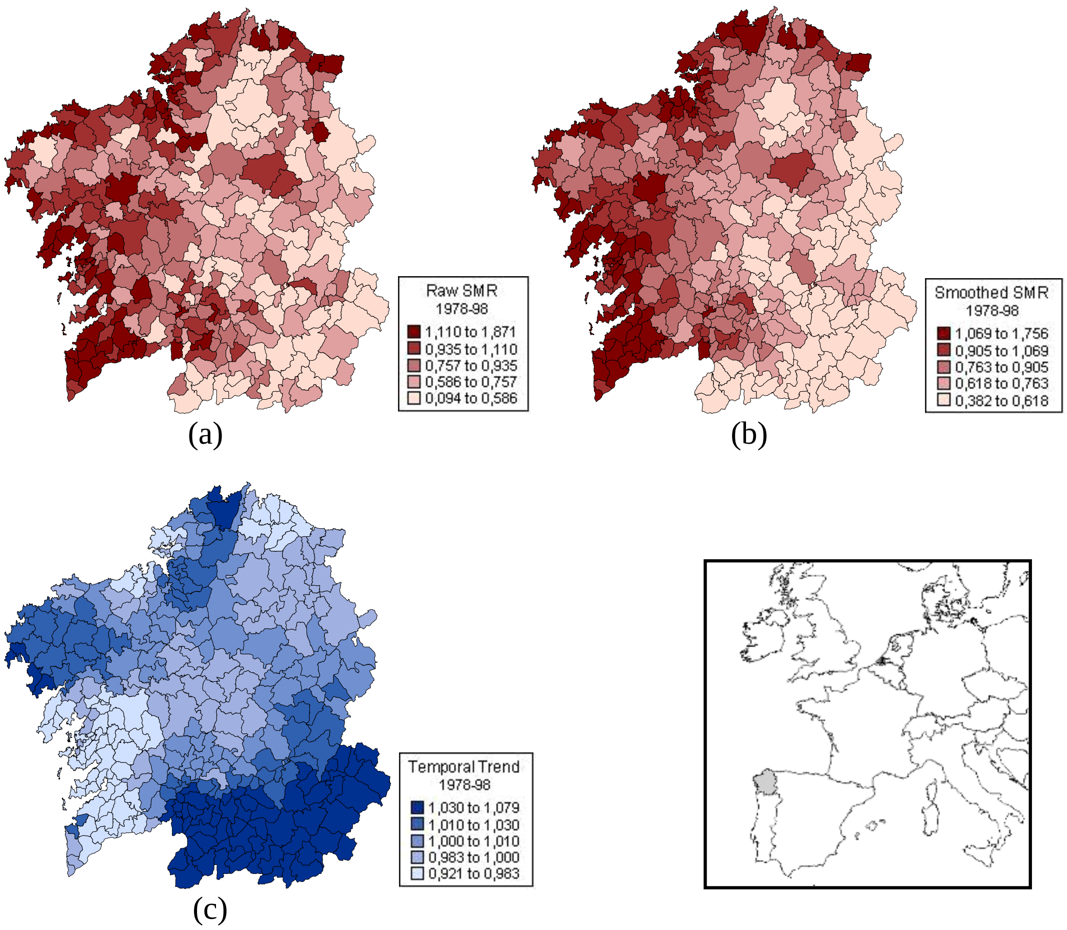
This model was, therefore, used for subsequent analyses, such as introduction of covariates.

Figure 1 shows the estimated *RR*s for lung cancer, obtained by classic approximation and Bayesian estimation.

The *RR*s are less variable in Figure 1(b) than in Figure 1(a), with the effect of smoothing being in evidence (*SMR*s ranges from 0,094 to 1,871; the Bayesian estimates of *RR* correct this dispersion, since the pulled estimates range from 0,382 to 1,756). The *SMR*s variability reflects random fluctuations due to different population size and corresponding small counts.

The maps highlight contrasts in the geographical distribution of risk; the northwest and the southwest

Figure 1. Lung cancer mortality in men in Galicia, 1978-1998: (a) Raw SMR; (b) Bayesian smoothed RR; (c) Temporal trend



parts of Galicia appear at higher risk. This pattern is stable in both maps, although, as expected, it emerges more clearly in the map obtained under the Bayesian model.

Figure 1(c) presents the temporal trend for lung cancer in the Galician municipalities and shows that, in areas with lower mortality, a rising trend has been in evidence over the course of the last few years.

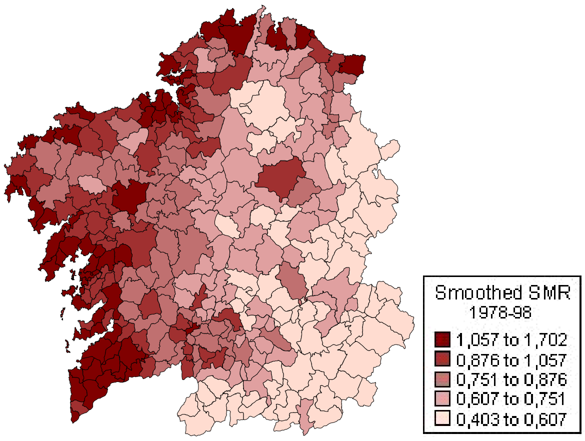
Figure 2 shows the estimations obtained by adjusting the model with both covariates. The *RRs* had shrunk further, that is, they have a smaller variation interval (*SMR*s ranges from 0.403 to 1.702). This result (the covariates providing information that helps to reduce the *RR*s variability), together with the fact that the DIC value for the model including the covariates is smaller than for the model without them, confirm the relationship between the covariates and the disease.

FUTURE TRENDS AND CONCLUSIONS

This study has addressed various issues in disease mapping: use of classical methods versus Bayesian models, choice of the prior distributions or incorporation of spatio-temporal structures. Its emphasis was more on comparing these methods with data than on making theoretical comparisons. Conclusions about comparison of methods have limited validity.

The use of Bayesian models for computing the *SMR* instead of the usual procedures is recommended. Maps of observed *SMR*s may be unstable and are prone to wrong interpretation, as attention is drawn to regions with high or low *SMR*s based on small populations. Smoothing overcomes the problem of overdispersion and enables us to highlight the differences between municipalities. The

Figure 2. Lung cancer mortality in men in Galicia, 1978-1998. Bayesian smoothed RR using the model with covariates.



estimation of the parameters of the hierarchical model can be carried out with algorithms belonging to the class of MCMC methods, which have improved in the last few years and allow us to compute summary statistics for the distribution of *RR*s.

The results obtained using Bayesian methods may be sensitive to the choice of priors. Nevertheless, Mollié (2000) discusses the insensitivity of these methods to the choice of priors, but in cases where data are sparse and there is a prevailing conviction of geographical variability, this independence ought to be explored with caution.

From this analysis, interesting results have arisen on the variability of the relative risks. There is indication that a main part of the variability is accounted for by a local clustering structure. Thus, there is some evidence that local risk factors, which might correspond to a complex combination of environmental and socio-demographic local characteristics, relatively stable over time, are influencing the spatial structure of lung cancer mortality.

By including covariate values measured on regions rather than individuals, this analysis is an example of "ecological modelling". As in any ecological study, and due to the "ecologic fallacy" (attributing effects measured in aggregate to individuals), one must interpret results with caution, even when pertinent covariates are included because there are many potential factors of bias which can create discrepancies between the results of a study that analyses data at an aggregate level and the results of those that do so at an individual level. Furthermore, cancer mortality rates rise with age, with the accompanying problem of worse diagnostic quality and ensuing death certification, which may lead to bias in estimation.

In this study, we assumed that the covariate measures are error-free. However, this is rather unrealistic. Indeed, Bernardinelli, Pascutto, Montomoli, and Komakec (1999) and Xia and Carlin (1998) incorporate covariate measures that are not error-free into the standard hierarchical Bayesian model, and demonstrate that when the underlying error is taken into account, estimations prove more accurate.

Subsequent investigations have led us to assess the extension of the model, with the introduction of the age factor and even its interaction with time because the mortality behaviour through the study years is different in the age-groups. This necessarily implies a greater number of observations, giving rise to problems in the production of sufficiently accurate posterior estimations.

Another improvement involves the choice of neighbours. A recent study (Best, Arnold, Thomas, Waller & Conlon, 1999) indicates that the option used in this paper yields better results, but it would nonetheless be interesting to experiment with other alternatives, such as taking all municipalities lying within a given radius of kilometres as neighbours, or taking *k* neighbours lying closest to a given municipality.

REFERENCES

Bailey, T.C. (2001). Spatial statistical analysis in health. *Cadernos de Saúde Pública, 17,* 1083-1098.

Banerjee, S., Carlin, B.P., & Gelfand A.E. (2004). *Hierarchical modeling and analysis for spatial data*. Boca Raton: Chapman&Hall/CRC.

Bernardinelli, L., Clayton, D., & Montomoli, C. (1995). Bayesian estimates of disease maps: How important are priors? *Statistics in Medicine, 14,* 2411-2431.

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., & Songini, M. (1995). Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine, 14,* 2433-2443.

Bernardinelli, L., Pascutto, C., Montomoli, C., & Komakec, J. (1999). Ecological regression with errors in covariates: An application. In A. Lawson, A. Biggeri, D. Böhning, E. Lesaffre, J.F. Viel & R. Bertollini (Eds.), *Disease mapping and risk assessment for public health* (pp. 329-348). Chichester, England: Wiley.

Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics, 43,* 1-59.

Best, N.G., Arnold, R.A., Thomas, A., Waller, L.A., & Conlon, E.R. (1999). Bayesian models for spatially correlated disease and exposure data. In J.M. Bernardo, J.O. Berger, A.P. Dawid & A.F. Smith (Eds), *Bayesian statistics 6*. Oxford: Clarendon Press.

Breslow, N.E., & Day, N.E. (1975). Indirect standardization and multiplicative models for rates, with reference to the age adjustment of cancer incidence and relative frequency data. *Journal of Chronic Diseases, 28,* 289-303.

Clayton, D., & Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics, 43,* 671-681.

Cowles, M.K., & Carlin, B.P. (1996). Markov Chain Monte Carlo convergence diagnosis: A comparative review. *Journal of the American Statistical Association, 91,* 883-904.

Lawson, A.B. (2001). *Statistical methods in spatial epidemiology*. Chichester: Wiley.

Lopez-Abente, G., Pollán, M., Escolar, A., Errezola, M., & Abraira, V. (2001). *Atlas of cancer mortality and other causes of death in Spain 1978-1992*. Madrid, Spain: Instituto de Salud Carlos III.

Mollié, A. (2000). Bayesian mapping of Hodgkin's disease in France. In P. Elliott, J. Wakefield, N.G. Best & D. Briggs (Eds.), *Spatial epidemiology. Methods and applications*. Oxford: Oxford University Press.

Moore, D.A., & Carpenter, T.E. (1999). Spatial analytical methods and geographic information systems: Use in health research and epidemiology. *Epidemiologic Reviews, 21,* 143-161.

Richardson, S. (2003). Spatial models in epidemiological applications. In P.J. Green, N.L. Hjort & S. Richardson (Eds.), *Highly structured stochastic systems*. Oxford: Oxford University Press.

Silva Ayçaguer, L.C. (1997). *Cultura estadística e investigación científica en el campo de la salud: Una mirada crítica*. Madrid, Spain: Diaz de Santos.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., & Van der Linde, A. (2002). Bayesian measure of model complexity and fit. *Journal of the Royal Statistical Society. Series B, 64,* 583-639.

Spiegelhalter, D.J., Thomas, A., Best, N.G., & Lunn, D. (2002). *WinBUGS 1.4.* [Computer Software]. Cambridge, England: MRC Biostatistics Unit.

Wakefield, J.C., Best, N.G., & Waller, L. (2000) Bayesian approaches to disease mapping. In P. Elliott, J. Wakefield, N.G. Best & D. Briggs (Eds.), *Spatial epidemiology. Methods and applications*. Oxford: Oxford University Press.

Xia, H., & Carlin, B.P. (1998). Spatio-temporal models with errors in covariates: mapping Ohio lung cancer mortality. *Statistics in Medicine, 17,* 2025-2043.

KEY TERMS

Clustering: Unusual grouping of events in space and/or time. It is also the tendency of observations to be situated closer to one another than they would be expected.

Deviance Information Criterion (DIC): Measure introduced by Spiegelhalter, Best, Carlin, and van der Linde (2002) as a tool for comparing and selecting complex hierarchical models. DIC combines a measure of fit, the deviance statistic, and a measure of complexity, the number of free parameters in the model. Since increasing complexity is accompanied by a better fit, the DIC trades off these two quantities.

Disease Mapping: It refers to the visual display of geographical patterns of disease in a map. The maps typically show standardized mortality or incidence ratios for geographic areas such as countries, counties, or districts.

Ecological Fallacy: The relationship between geographical variation in disease incidence or mortality and explanatory covariates (e.g., environmental agents or lifestyle characteristics) measured on groups is often interpreted as a proxy of the relationship between disease and exposure in individuals. However, the association observed at the group or ecological level will not necessarily represent the association between the corresponding variables at the individual level. This type of bias is known as ecological fallacy.

Heterogeneity: Effects that unobserved variables can introduce in disease rates and that vary in an unstructured manner in either space or time.

Hierarchical Bayesian Model: The enterprise of specifying a model over several levels is called hierarchical modelling, with each new distribution forming a new level in hierarchy. Suppose that we have a collection of observations $\boldsymbol{y}=(y_1, y_2, \ldots, y_k)$ that are assumed to come from a distribution $p(\boldsymbol{y}|\boldsymbol{\theta})$ where $\boldsymbol{\theta}=(\theta_1,\theta_2,\ldots,\theta_k)$. It can also be assumed that $\boldsymbol{\theta}$ is a random quantity draw from the distribution $p(\boldsymbol{\theta}|\boldsymbol{\lambda})$ where $\boldsymbol{\lambda}$ is a vector of parameters. In principle, the parameter $\boldsymbol{\lambda}$ can depend itself on a collection of other parameters. This sequence of parameters and distributions constitutes a hierarchical model. The hierarchy must stop at some point, with the remaining parameters known.

Prior Information: Existing knowledge about a phenomenon under investigation. This knowledge can be incorporated into a statistical analysis using probability distributions. For instance, in a hierarchical Bayesian model, prior information is the set of probability distributions of the parameters in the model.

Relative Risk: Measure that compares the incidence of a disease in a group of interest with the incidence in a control group.

Smoothing: Adjustment of mortality rates using a combination of the knowledge about the rate in each area and the knowledge about the rates in surrounding areas. The smoothing removes the noise in disease maps and allows to assess specific hypotheses concerning incidence or mortality.

Standardized Mortality Ratio (SMR): Ratio of the observed mortality in a group of interest to the expected mortality in that group, based on some standard population. It gives information about regions with unusual high or low numbers compared to the expected numbers.

B

Bayesian Machine Learning

Eitel J.M. Lauría
Marist College, USA

INTRODUCTION

Bayesian methods provide a probabilistic approach to machine learning. The Bayesian framework allows us to make inferences from data using probability models for values we observe and about which we want to draw some hypotheses. Bayes theorem provides the means of calculating the probability of a hypothesis (posterior probability) based on its prior probability, the probability of the observations and the likelihood that the observational data fit the hypothesis.

$$\mathrm{P}(H \mid D)=\frac{\mathrm{P}(D \mid H)\cdot \mathrm{P}(H)}{\mathrm{P}(D)} \tag{1}$$

$\mathrm{P}(H \mid D)$ is defined as the probability of a certain hypothesis based on a set of observational data given a certain context (posterior probability of hypothesis H); $\mathrm{P}(D \mid H)$ is the likelihood of the observations given a certain hypothesis; $\mathrm{P}(H)$ is the intrinsic probability of hypothesis H, before considering the evidence D (prior probability); and $\mathrm{P}(D)$ is the probability of the observations, independent of the hypothesis, that can be interpreted as a normalizing constant. Bayes rule can therefore be reformulated as shown in expression . This means that the probability of the hypothesis is being updated by the likelihood of the observed data.

$$\mathrm{P}(H \mid D) \propto \mathrm{P}(D \mid H)\cdot \mathrm{P}(H) \tag{2}$$

BACKGROUND

The practical application of Bayes rule to machine learning is rather straightforward. Given a set of hypotheses H and a set D of observational data we can estimate the most probable hypothesis H given D, by comparing different instances of the previous expression for each hypothesis H and choosing the one that holds the largest posterior probability (also called maximum a posteriori probability or MAP).

$$\text{Most probable } H \equiv H_{\mathrm{MAP}} = \arg\max \left[\mathrm{P}(D \mid H)\cdot \mathrm{P}(H)\right] \tag{3}$$

Suppose we have a classification problem where the class variable is denoted by C and can take values $c_1, c_2, \ldots, c_k$. Consider a data sample D represented by n attributes $A_1, A_2, \ldots, A_n$ of which the observations $(a_1, a_2, \ldots, a_m)$ have been taken for each instance of D. Suppose that each instance of the data sample D is classified as $c_1, c_2, \ldots, c_k$. The Bayesian approach to classifying a new instance would then be to assign the most probable target value (a class value of type c_i) by calculating the posterior probability for each class given the training data set, and from them choosing the one that holds the maximum a posteriori probability.

$$c_{\mathrm{MAP}} = \arg\max_{c_i \in C} \left[\mathrm{P}(D \mid c_i)\cdot \mathrm{P}(c_i)\right] \tag{4}$$

NAIVE BAYES CLASSIFICATION

Although the idea of applying full-blown Bayesian criteria to analyze a hypothesis space in search of the most feasible hypothesis is conceptually attractive, it usually fails to deliver in practical settings. Although we can successfully estimate $\mathrm{P}(c_i)$ from the training data, calculating the joint probability $\mathrm{P}(D \mid c_i)$ is usually not feasible: unless we have a very large training data set, we would end up with estimates that are representative of a small fraction of the instance space and are therefore unreliable. The naive Bayesian classifier attempts to solve this problem by making the following assumptions:

- Conditional independence among attributes of the data sample. This means that the posterior probability of D, given c_i is equal to the product of the posterior probability of each attribute.

$$\mathrm{P}(D \mid c_i) = \prod_{j=1}^{n} \mathrm{P}(A_j = a_j \mid c_i) \quad , c_i \in C \tag{5}$$

- The conditional probabilities of each individual attribute can be estimated from the frequency distributions of the sample data set D as N_{ij}/N_i, where N_{ij} is the number of training examples for which attribute $A_j = a_j$ and class value is c_i; and N_i is the number of training examples for which the class value is c_i. If the prior probabilities $P(c_i)$ are not known, they can also be estimated by drawing its probabilities from the sample data set of frequency distributions.
- To solve the cases in which there are very few or no instances in the data set for which $A_j = a_j$ given a certain class value c_i, which would in turn render poor estimates of $\mathrm{P}(A_j = a_j \mid c_i)$ or make it equal to zero, a common approach is to estimate $\mathrm{P}(A_j = a_j \mid c_i)$ as:

$$\mathrm{P}(A_j = a_j \mid c_i) = \frac{N_{ij} + \alpha_{ij}}{N_i + \alpha_i} \qquad (6)$$

where α_{ij} and α_i can be seen as fictitious counts coming out of our prior estimate of the probability we wish to determine. In rigor, this implies considering a conjugate prior probability given by a Dirichlet distribution (for more details see Ramoni & Sebastiani, 1999). A typical method for choosing α_{ij} and α_i in the absence of other information is to assume uniform distribution of the counts, which means that if an attribute has r possible values, $\alpha_{ij} = 1$ and $\alpha_i = r$. This results in:

$$\mathrm{P}(A_j = a_j \mid c_i) = \frac{N_{ij} + 1}{N_i + r} \qquad (7)$$

These assumptions have the effect of substantially reducing the number of distinct conditional probability terms that must be estimated from the training data. To illustrate the use of the naïve Bayes classifier, consider the example in Table 1 adapted from Mitchell (1997). We are dealing with records reporting on weather conditions for playing tennis. The task is to build a classifier that, by learning from previously collected data, is able to predict the chances of playing tennis based on new weather reports. We can estimate the class probabilities P(play=yes) and P(play=no) by calculating their frequency distributions as follows:

P(play=yes) = (# of instances were play=yes) / (total # of instances) = 9/14

P(play=no) = (# of instances were play=no) / (total # of instances) = 5/14

The conditional probabilities can be estimated by applying equation , as shown in Table 1(d). For a new weather report W={outlook=rain, temp=hot, Humidity=high, windy =false} the classifier would compute

Table 1. Weather data set

outlook	temperature	humidity	windy	play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	No

(a) List of Instances

Attribute Name	Values
outlook	sunny, overcast, rainy
temperature	hot, mild cool
humidity	high, normal
windy	true, false
play (class attribute)	yes, no

(b) List of Attributes

(play = yes)	9/14
(play = no)	5/14

(c) Prior Class Probabilities

	play				play	
Outlook	yes	no		Humidity	yes	no
sunny	3/12	4/8		high	4/11	5/7
overcast	5/12	1/8		normal	7/11	2/7
rain	4/12	3/8				
Temperature				Windy		
hot	3/12	3/8		false	4/11	4/7
mild	5/12	3/8		true	7/11	3/7
cold	4/12	2/8				

(d) Conditional Probabilities

$$P(\text{play=yes} \mid W) = P(W \mid \text{play=yes}) \cdot P(\text{play=yes})$$
$$= \frac{4}{12} \cdot \frac{3}{12} \cdot \frac{4}{11} \cdot \frac{4}{11} \cdot \frac{9}{14} = 0.007084$$

$$P(\text{play=no} \mid W) = P(W \mid \text{play=no}) \cdot P(\text{play=no})$$
$$= \frac{3}{8} \cdot \frac{3}{8} \cdot \frac{5}{7} \cdot \frac{4}{7} \cdot \frac{5}{14} = 0.020499$$

and therefore predict that no tennis matches will be played.

Although naive Bayes classification makes a major simplification (conditional independence is a very strong assumption), in practice it often competes well with more sophisticated algorithms. See Rish (2001) for an empirical analysis of the naive Bayes classifier.

BAYESIAN BELIEF NETWORKS

Bayesian belief networks (BBNs) follow a middle-of–the-road approach when compared to the computationally intensive optimal Bayesian classifier and the oversimplified naive Bayesian approach. A BBN describes the probability distribution of a set of attributes by specifying a set of conditional independence assumptions together with a set of causal relationships among attributes and their related joint probabilities. When used in this way, BBNs result in a powerful knowledge representation formalism, based in probability theory, that has the potential of providing much more information about a certain domain than visualizations based on correlations or distance measures.

Building a Graphical Model

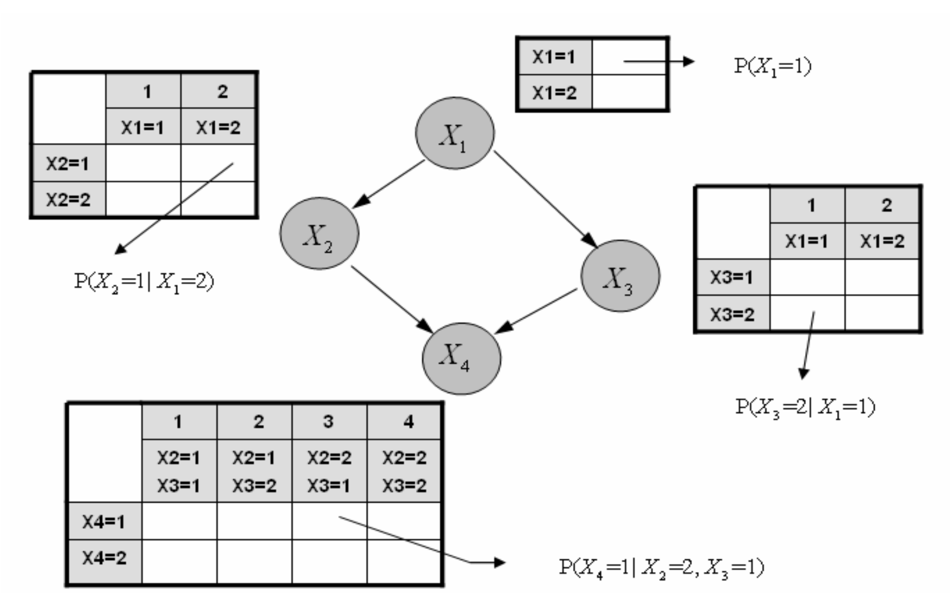
The set of relationships in a Bayesian network can be modeled by means of a directed acyclic graph (DAG), where each node represents a variable to which a group of conditional probabilities are attached. Given a set of variables $X_1, X_2, ..., X_n$, a directed acyclic graph (DAG) is constructed in which each node of the graph represents each of the variables and the arcs represent the relationship between the arcs (see Figure 1). Recalling the product probability rule, the joint probability distribution can be written as:

$$p(X_1 = x_1, X_2 = x_2, ..., X_n = x_n) = p(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} p(x_i \mid x_{\text{pa}(i)}) \tag{8}$$

We use the notation $p(x_1, x_2, ..., x_n)$ as an abbreviation for the probability of a conjunction of particular assignments $x_1, x_2, ..., x_n$ to the set of variables $X_1, X_2, ..., X_n$, and $x_{\text{pa}(i)}$ is a vectorized notation to identify a given configuration of the list of direct parents of X_i, linked to X_i through the causal arcs in the graph model.

In addition to the graph structure, it is necessary to specify the parameters of the model. This means that we

Figure 1. Structure and parameters of a Bayesian network



must specify the conditional probability distribution at each node. If the variables are discrete, this can be represented as a table that lists the probability the child node takes on each of its different values for each combination of values of its parents (see Figure 1).

Inference in Bayesian Belief Networks

Once the belief network is formulated it can be used for probabilistic inference, that is, to make probabilistic statements concerning the network attributes (nodes). Since a BBN uniquely defines a joint probability distribution, any probability can be computed from the network by specifying the joint probability distribution and applying basic rules of probability, such as conditioning and marginalization.

As indicated before, because a BBN determines a joint probability distribution for the set of attributes in the network, the Bayesian network can, in principle, be used to compute any probability of interest. For problems with many variables, though, this approach is not practical. Pearl's message passing algorithm (1988) provides an improved approach for singly connected DAGs (graphs where no pair of nodes has more than one path). The new evidence is propagated over the network by sending messages to neighbor nodes. Through the arcs, acting as communication channels, the nodes send messages providing information about the joint probability distribution that is defined by the network and the evidence obtained so far. The algorithm has been extended to multiple connected DAGs where messages are passed in a junction tree that maps the original BBN (Jensen et al., 1990; Lauritzen & Spieglelhalter, 1988; Pakzad & Anantharam, 2002). Also, several approximate algorithms have been devised, based on Markov Chain Monte Carlo (MCMC) sampling (Heckerman et al., 1994; Neal, 1993) or variational methods (Jordan et al., 1997).

Training Bayesian Belief Networks

There are a number of possible scenarios to consider when addressing the problem of training a BBN:

- When the structure of the BBN is known and all the attributes for all the instances are observable in the training examples, learning the conditional probabilities is quite straightforward. We simply estimate the conditional probabilities by maximizing the likelihood of the training data, as in the case of the naive Bayes classifier (estimating relative frequencies with zero frequency corrections, for example).
- When the structure of the BBN is known but not all of the variables are observable (partially or totally) in the training data, we come across a more complicated problem. In such a case we can resort to algorithms intended to deal with missing values, such as the Estimation Maximization (EM) algorithm (Dempster et al., 1977). Another approach is assimilating the problem to the case of estimating the weights of the hidden nodes in a neural network. In that case a gradient ascent approach can be used, where the algorithm searches through the space of hypotheses corresponding to the set of all possible entries in the conditional probability table.
- When the structure of the network is not known, we face a problem of model selection, typically a much more complicated problem than the two previously described cases. The goal is to find the network, or group of networks that best describe the probability distribution over the training data. This optimization process is implemented in practice by using heuristic search techniques to find the best model over the space of possible BBNs. A scoring system is commonly used for choosing among alternative networks. Several scores have been considered, including the posterior probability of the model given the data, and approximations based on the Bayesian Information Criteria (Heckermann, 1999; Schwartz, 1978). Lam and Bachus (1994) and Lauría (2003) use scores based on the Minimum Description Length Principle (Rissanen, 1986, 1996), an information theoretic perspective of Occam's Razor principle according to which simple, sparse models should be preferred to complex overfitted models. Several authors discuss the idea of doing structure discovery by approximating the full Bayesian model averaging. Madigan et al. (1996) and Giudici et al. (2000) discuss the use of a MCMC search over the set of structures. Friedman and Koller (2001) introduce the idea of heuristic searching over orderings.

FUTURE TRENDS

Bayesian machine learning is being applied in a wide variety of domains, and the breadth of applications will undoubtedly increase in the future. Variations of the naive Bayes classifier are used in data mining and text summarization tasks (Arnborg, 2003; Eyheramendy et al., 2003; Peng & Schuurmans, 2003; Wiering, 2002). Lauría and Tayi (2003) describe the application of naive Bayes classification in network intrusion detection. Stewart (2002) uses a naive Bayes classifier to predict software project metrics. Friedman et al. (2000) use Bayesian networks to analyze gene expression data. Bilmes (2000) and Nefian et al. (2002) apply dynamic Bayesian networks to speech recognition. Onisko (2001) describes Hepar II, a system for diagnosing liver disorders. Kennett et al. (2001) have developed a weather forecasting system that predicts sea breezes. Hudson et al. (2001) have applied Bayesian networks as probabilistic expert systems for terrorist risk assessment. Breese and Heckerman (1996), and Jensen et al. (2001) target the problem of troubleshooting, fault diagnosis and reliability analysis. Microsoft has developed a number of Bayesian network applications, including Microsoft Office's Answer Wizard that uses naive Bayes to select help topics based on queries (Haddaway, 1999).

A plethora of theoretical and applied research is being performed in Bayesian networks and its related domains, including such topics as causality and Bayesian networks (Pearl, 2000), dynamic networks, learning with incomplete data, and cyclic networks. Heckerman (2000) points out that Microsoft uses dependency networks (a formalism similar to Bayesian networks that admits cyclical graphs) in two of its recent products—SQL Server 2000 and Commerce Server 2000—for both collaborative filtering and data visualization.

CONCLUSION

In this article we have provided an overview on Bayesian methods as applied to the field of machine learning. In doing so we have deliberately focused on reviewing the most relevant concepts, techniques and practical issues. Over the last few years, Bayesian machine learning has emerged as a prominent modelling and data analysis approach from which both academicians and practitioners can benefit. We envisage that Bayesian machine learning will continue to expand in the future both from a theoretical as well as a practical application perspective

REFERENCES

Arnborg, S. (2003). A survey of Bayesian data mining. In J. Wang (Ed.), *Data mining: Opportunities and challenges* (pp. 1-26). Hershey, PA: Idea Group Publishing.

Bilmes, J. (2000, July). Dynamic Bayesian multi-networks. *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence,* Stanford.

Breese, J.S., & Heckerman, D. (1996). Decision-theoretic troubleshooting: A framework for repair and experiment. *Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence,* San Francisco, CA (pp. 124–132). Morgan Kaufmann Publishers.

Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, B 39,* 1-38

Eyheramendy, S., Lewis, D., & Madigan, D. (2003) On the naive Bayes model for text categorization. To appear in *Artificial Intelligence & Statistics.*

Friedman, N., & Koller, D. (2001). Being Bayesian about Bayesian network structure: A Bayesian approach to structure discovery in Bayesian networks. *Machine Learning*. Accepted for publication. Earlier version appeared in UAI 2000.

Friedman, N., Linial, M., Nachman, I., & Pe'er, D. (2000). Using Bayesian networks to analyse expression data. *Proceedings of the 4th Annual International Conference on Computational Molecular Biology*.

Giudici, P., Green, P., & Tarantola, C. (2000). Efficient model determination for discrete graphical models. *Quaderni di Dipartimento 116 (7-00).* Dipartimento di economia politica e metodi quantitativi.Universita' degli studi di Pavia.

Haddaway, P. (1999, Spring). An overview of some recent developments in Bayesian problem solving techniques. *AI Magazine*.

Heckermann, D. (1999). A tutorial on learning with Bayesian networks. In M. Jordan (Ed.), *Learning in graphical models*. Cambridge, MA: MIT Press.

Heckerman, D., Chickering, D., Meek, C., Rounthwaite, R., & Kadie, C. (2000). *Journal of Machine Learning Research, 1. http://www.ai.mit.edu/projects/jmlr//papers/volume1/heckerman00a /html/heckerman00a.html*

Heckerman, D., Geiger, D., & Chickering, D. (1994). *Learning Bayesian networks: The combination of knowledge and statistical data.* Technical Report MSR-TR-94-09. Microsoft Research.

Hudson, L., Ware, B., Mahoney, S., & Laskey, K. (2001). *An application of Bayesian networks to antiterrorism risk management for military planners.* Technical report. Department of Systems Engineering and Operations Research, George Mason University.

Jensen, F., Lauritzen, S., & Olesen, K. (1990). Bayesian updating in causal probabilistic networks by local computations. *Computational Statistics Quarterly, 4,* 269-282.

Jensen, F.V., Kjærul, U., Kristiansen, B., Langseth, H., Skaanning, C., Vomlel, J., & Vomlelov´a, M. (2001). The SACSO methodology for troubleshooting complex systems. *Artificial Intelligence for Engineering, Design, Analysis and Manufacturing, 15*(5), 321–333.

Jordan, M., Ghahramani, Z., Jaakkola, T., & Saul, L. (1997). An introduction to variational methods for graphical models. In M. Jordan (Ed.), *Learning in graphical models* (pp. 105-158). MIT Press.

Kennett, R., Korb, K., & Nicholson, A. (2001). Seabreeze prediction using Bayesian networks. In D. Cheung, G.J. Williams & Q. Li (Eds.), *Proceedings of the 5th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining - PAKDD 2001* (pp. 148-153). Springer-Verlag.

Lam, W., & Bachus, F. (1994). Learning Bayesian networks. An approach based on the MDL principle. *Computational Intelligence, 10*(3), 269-293.

Lauría, E. (2003). *Learning structure and parameters of Bayesian belief networks: An application and a methodology in information technology implementation*. Dissertation. University at Albany, State University of New York.

Lauría, E., & Tayi, G. (2003). A comparative study of data mining algorithms for network intrusion detection in the presence of poor quality data. *Proceedings of the 8th international Conference on Information Quality* (ICIQ-2003), Massachusetts Institute of Technology (pp. 190-201).

Lauritzen, S.L., & Spiegelhalter, D.J. (1988). Local computations with probabilities on graphical structures and their application to expert systems. Journal of the Royal Statistical Society Series B, *50*(2), 157-224.

Madigan, D., Andersson, S., Perlman, M., & Volinsky, C. (1996). Bayesian model averaging and model selection for Markov equivalence classes of acyclic graphs. *Communications in Statistics: Theory and Methods, 25,* 2493–2519.

Mitchell, T. (1997). *Machine learning*. McGraw-Hill

Neal, R.M. (1993). *Probabilistic inference using Markov Chain Monte Carlo methods*. Technical Report CRG-TR-93-1. Dept. of Computer Science, University of Toronto.

Nefian, A., Liang, L., Liu, X., Pi, X., & Murphy, K. (2002). Dynamic Bayesian networks for audio-visual speech recognition. *EURASIP, Journal of Applied Signal Processing, 2002*(11), 1274-1288.

Onisko, A., Lucas, P., & Druzdzel, M. (2001). Comparison of rule-based and Bayesian network approaches in medical diagnostic systems. *AIME 2001* (AI in Medicine in Europe), 283-292.

Pakzad, P., & Anantharam, V. (2002). A new look at the generalized distributive law. *2002 Conference on Information Sciences and Systems,* Princeton University.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Mateo, CA: Morgan Kaufman Publishers.

Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge University Press.

Peng, F., & Schuurmans, D. (2003). *Combining naive Bayes and n-gram language models for text classification*. Submitted to the 25th European Conference on Information Retrieval Research (ECIR).

Ramoni, M., & Sebastiani, P. (1999). Bayesian methods for intelligent data analysis. In M. Berthold & D.J. Hand (Eds.), *Intelligent data analysis: An introduction*. New York: Springer.

Rish, I. (2001). An empirical study of the naive Bayes classifier. IJCAI-01 *Workshop on Empirical Methods in AI.*

Rissanen, J. (1986). Stochastic complexity and modelling. *Annals of Statistics, 14*(3), 1080-1100.

Rissanen, J. (1996). Fisher information and stochastic complexity. *IEEE Transaction on Information Theory, 42,* 40-47.

Stewart, B. (2002). Predicting project delivery rates using the naive-Bayes classifier. *Journal of Software Maintenance: Research and Practice, 14*(3), 161-179.

Wiering, M.A. (2003). Hierarchical mixtures of naive Bayesian classifiers. In P. Larranaga, J. Lozano, J. Pena & I. Inza (Eds.), *European Conference on Machine Learning. Workshop on Probabilistic Graphical Models for Classification. ERCIM news* (pp. 93-104).

KEY TERMS

Bayes' Rule: Shows how probabilities can be updated in the light of evidence. Bayes' rule derives from a simple reordering of terms in the product rule and by subsequently applying the law of total probability.

Bayesian Network: A graphical model that encodes the probability distribution of a set of random variables by specifying a set of conditional independence assumptions together with a set of relationships among these variables and their related joint probabilities.

Classification: Mapping a data item into one of several pre-defined categories. Stored data are used to locate data in predetermined groups. For example, a retail store chain could use customer purchase data to determine when customers visit and what they typically buy.

Directed Acyclic Graph: A graph in which each edge can be followed from one vertex to the next, and where no path starts and ends at the same vertex.

Naive Bayes: An extremely simple classification algorithm based on the use of the posterior probability as the defining criterion to choose among a set of hypotheses (classes). The naive Bayes classifier simplifies calculations by assuming conditional independence among the dataset attributes, given the class.

Probability: A number between 0 and 1 attached to a logical proposition that represents the degree of truth in that proposition.

Statistical Machine Learning: The process of deriving the parameters that characterize a statistical model from a set of observations.

Bayesian Modelling for Machine Learning

Paul Rippon
The University of Newcastle, Australia

Kerrie Mengersen
The University of Newcastle, Australia

INTRODUCTION

Learning algorithms are central to pattern recognition, artificial intelligence, machine learning, data mining, and statistical learning. The term often implies analysis of large and complex data sets with minimal human intervention. Bayesian learning has been variously described as a method of updating opinion based on new experience, updating parameters of a process model based on data, modelling and analysis of complex phenomena using multiple sources of information, posterior probabilistic expectation, and so on. In all of these guises, it has exploded in popularity over recent years.

General texts on Bayesian statistics include Bernardo and Smith (1994), Gelman, Carlin, Stern, and Rubin (1995), and Lee (1997). Texts that derive more from the information science discipline, such as Mitchell (1997) and Sarker, Abbass, and Newton (2002), also include sections on Bayesian learning.

Given recent advances and the intuitive appeal of the methodology, Bayesian learning is poised to become one of the dominant platforms for modelling and analysis in the 21st century. This article provides an overview of Bayesian learning in this context.

BACKGROUND

Bayesian Modelling

Bayesian learning aims to provide information about unknown characteristics of a population (such as a mean and/or a variance) or about relationships between characteristics (for example, via a regression equation or a neural network). We often have a set of alternative models or hypotheses, H_1, H_2,…, H_m, that could describe these unknowns, such as possible values for the unknown mean or alternative neural network representations. The Bayesian approach allows prior beliefs about these models to be updated in the light of new data. The fundamental enabling mechanism is Bayes' rule:

$$p(H_i \mid D) = \frac{p(D \mid H_i)\, p(H_i)}{p(D)}, \qquad (1)$$

which states that the *posterior* probability $p(H_i|D)$ of a particular model, H_i, conditional on data, D, is proportional to the probability $p(D|H_i)$ of the data, given the model multiplied by the *prior* probability $p(H_i)$ of the model. The denominator $p(D)$, a normalizing constant designed to make the posterior probability sum or integrate to one, can be termed the probability of the data and is expressed as

$$p(D) = \sum_{i=1}^{m} p(H_i)\, p(D \mid H_i).$$

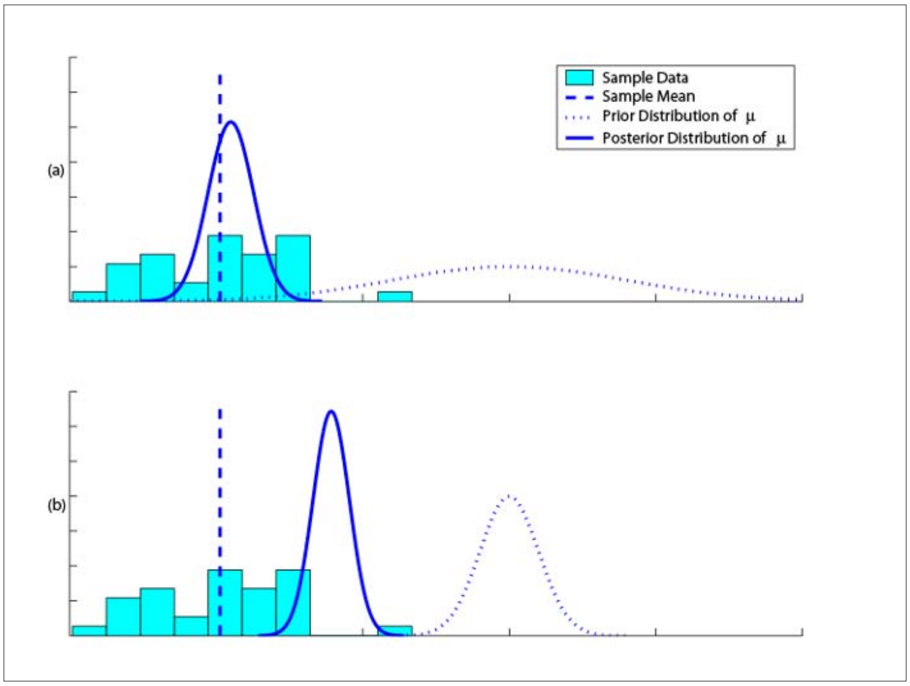
The number of plausible models might be infinite, for example, when the different models are represented by unknown values of a continuously distributed population mean. In this case, probability distributions become densities and the summation in $p(D)$ is replaced by an integral. In either case, it is this denominator, $p(D)$, that is often intractable. This motivates the development of numerical methods such as Markov chain Monte Carlo, described in the next section.

As a simple example, consider sampling n data points y_1, y_2,…, y_n from a population of normally distributed measurements in order to estimate an unknown mean, μ, and assume that the population variance, σ^2, is known. Thus, H is the set of all possible values that μ may take. The sample mean, $\bar{y}$, represents the information contained in the data so that $p(D|H)=p(\bar{y}|\mu)=N(\mu,\sigma^2/n)$.

In practice, we often have some prior knowledge about μ, such as, "μ is known from experience to be around a value μ_0." We might express this prior knowledge as $\mu \sim N(\mu_0, \tau_0^2)$, where τ_0^2 represents the uncertainty around the best guess, μ_0. Now, according to Bayes' rule:

$$p(H \mid D) = p(\mu \mid \bar{y}) \propto p(\bar{y} \mid \mu)\, p(\mu) = N(\mu_n, \tau_n^2) \qquad (2)$$

Figure 1.



with

$$\mu_n = \left(\frac{\mu_0}{\tau_0^2} + \frac{n\bar{y}}{\sigma^2}\right)\tau_n^2 \text{ and } \qquad \frac{1}{\tau_n^2} = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2}.$$

The posterior distribution can be considered as a merging of the prior opinion about μ and the data. Figure 1 illustrates this updating of opinion about μ for different priors.

In 1a, the prior knowledge about μ is fairly vague. Thus, the posterior distribution for μ is dominated by the data. In 1b, the prior knowledge about μ is more precise and has more influence on the posterior distribution.

Bayesian Computation

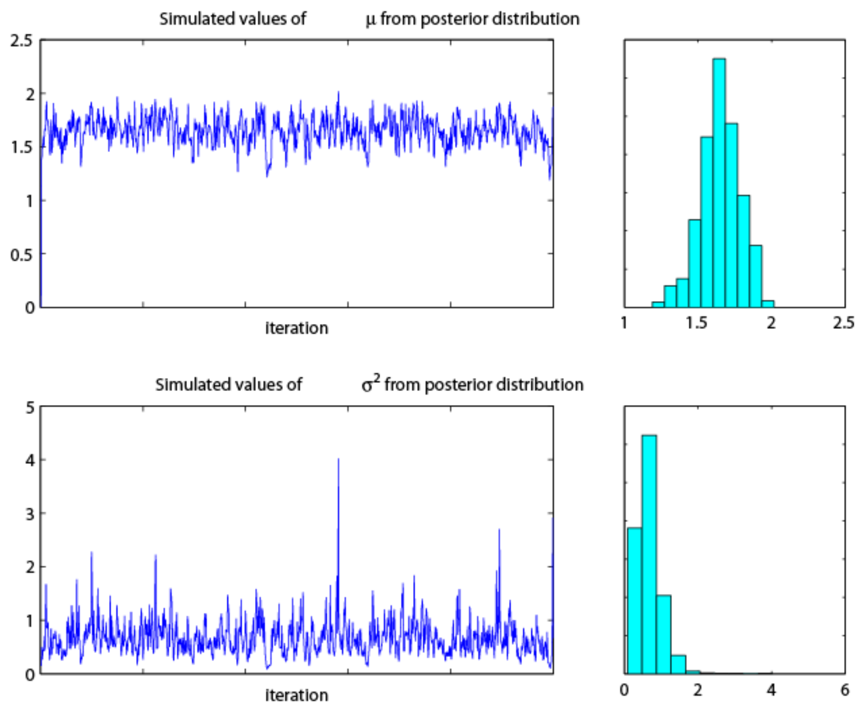
Continuing this example, suppose more realistically that σ^2 is also unknown. A distribution that reflects theoretical properties of a variance is the inverse Gamma distribution, so we might take $\sigma^2 \sim IG(a_0, b_0)$, where a_0 and b_0 are chosen to reflect our prior knowledge. Application of Bayes' rule results in a joint posterior distribution of μ and σ^2 that is nonstandard and multidimensional, making analytical solutions difficult.

A popular numerical solution is Markov chain Monte Carlo (MCMC). MCMC algorithms allow simulation from a Markov chain whose stationary distribution is $p(H|D)$. If it is not easy to simulate directly from $p(H|D)$, values can be proposed from some easily simulated distribution (such as uniform or normal) and accepted or rejected according to a rule that ensures that the final set of accepted values are from the target posterior distribution. If $p(H|D)$ is high dimensional, it can often be decomposed into a series of lower dimensional, conditional distributions, and (possibly different) MCMC algorithms can iterate around these, eventually forming a sample from the joint distribution (Besag, 1974).

For this example problem, a basic MCMC algorithm would be as follows.

- Choose initial values μ_1, σ_1^2
- Repeat for i=2:k for k large
  - Randomly draw σ_i from $\mu \mid \sigma_{i-1}^2, \bar{y}$ given in Equation (2).
  - Randomly draw σ_i^2 from the conditional posterior distribution $\sigma^2 \mid \mu_i, \bar{y} \sim IG(a_n, b_n)$, where $a_n = a_0 + \frac{1}{2}$ and $b_n = b_0 + \frac{n}{2}(\bar{y} - \mu_i)^2$
- Discard the first part of the above chain as "burn-in," in which the Markov chain is still approaching the target posterior distribution from the (perhaps unlikely) initial values.
- The remaining iterates, μ_i and σ_i^2, represent a large sample from the target posterior distribution. Graphs

Figure 2.



of μ_i and σ_i^2 or summaries such as $E(\mu)\approx\Sigma\mu_i/n$ can be used to present information about the posterior distribution. Figure 2 shows the results of using such an algorithm to simulate a sample from the posterior distribution.

Chen, Shao, and Ibrahim (2000) and Robert and Casella (1999) describe a wide range of methods for modelling and simulation of multiparameter data. While Metropolis-Hastings and Gibbs remain popular traditional MCMC algorithms, many variations have now been developed. The well-known variations include adaptive rejection sampling and adaptive rejection Metropolis sampling (Gilks, Roberts, & Sahu, 1998), slice sampling (Mira & Tierney, 2002; Neal, 1997), perfect simulation (Kendall, 1998; Propp and Wilson, 1996), delayed rejection sampling (Tierney & Mira, 1999), adaptive algorithms, black box algorithms (Chen & Schmeiser, 1998), and hybrid methods that employ combinations of MCMC algorithms in a single analysis (Tierney, 1994). For further description of Monte Carlo algorithms for Bayesian analysis, see Carlin and Louis (2000), Chen et al. (2000), Denison, Holmes, Mallick, and Smith (2002), Gustafson (2003), Ibrahim, Chen, and Sinha (2001), Robert (1998), Robert and Casella (1999), and West and Harrison (1997). Various software for MCMC is now available, including the freeware package BUGS (Spiegelhalter, Thomas, Best, & Gilks, 1995).

With all Monte Carlo methods, it is imperative to assess convergence, which includes ensuring that the length of burn-in is adequate (so that the effect of the initial values have worn off), the simulated values are indeed being drawn from the correct target distribution, the algorithm adequately explores the whole target space, and sufficient values are collected to allow adequate approximations of expectations of interest. Convergence assessment for MCMC includes theoretical evaluation exploiting the Markov chain features of these chains (Mengersen & Tweedie, 1996; Tierney, 1998; among many others) and empirical diagnostics (see, for example, Best, Cowles, & Vines, 1995, and stand-alone routines at http://lib.stat.cmu.edu).

Applications

Flexible, and if necessary quite complex, models can be constructed to reflect the linkages and uncertainties in any particular problem. The simple example described above could be extended further by placing priors on the hyperparameters μ_0, τ_0^2, a_0, and b_0 to create a hierarchical model. This extends to general hierarchical models (Congdon, 2001), neural networks (Neal, 1996), regression and classification trees and belief networks (Chipman, George, & McCulloch, 1998; Mello & Brown, 1999), and so on. For example, Bayesian belief networks, BBNs, describe associations between sets of variables. Prior knowledge informs the initial construction of the network and Bayes' rule is used to improve the network by learning both the parameters and the structure from the data. Similarly, a Bayesian neural network (BNN) applies a probabilistic interpretation to the traditional artificial neural network technique, using the training set of inputs and targets to calculate the predictive distribution of target values in a new test case, given the inputs for that case.

Bayesian learning has been used for pattern recognition (Serrano, Andreas, Savakis, & Luo, in press), mixture model identification (Ma, Wang, & Xu, 2004), artificial intelligence in medicine and neurocomputing (Cervera & del Pobil, 1997; Kononenko, 2001), operations research (Anderson & Vastag, 2004), and Internet evaluation and security (Cho & Cha, in press; Mladeni & Grobelnik, 2003). Bayesian networks have been applied to complex problems in areas as diverse as genetics (Hsu, 2004), agriculture (Moshou et al., 2001), marketing (Ahn & Ezawa, 1997), information retrieval (Campos et al., 2003), text classification (Klopotek, 2004), biomedicine (Lucas, van der Gaag, & Abu-Hanna, 2004), and forensics (Taroni, Biedermann, Garbolino, & Aitken, 2004). Bayesian genetic and evolutionary models, including neural networks, have also been described in an information science context (Pelikan, Goldberg, & Tsutsui, 2003) and applied to software failure time prediction (Tian & Noore, in press), drug discovery (Winkler & Burden, 2004), and environmental modelling (Danaher, Datta, Waddle, & Hackney, 2004).

FUTURE TRENDS

The theory of Bayesian modelling and analysis are comprehensively described by Bernardo and Smith (1994) and Robert (1994). These discussions include issues of point estimation, tests and confidence regions, admissibility and invariance, and priors. All of these areas are still actively researched. For example, priors may be objective so that they have no influence over the data (Bernardo & Smith) or subjective, reflecting expert opinion (Garthwaite & O'Hagan, 2000; Kadane and Wolfson, 1998). Gelman et al. (1995) discuss informative and noninformative prior distributions for single-parameter models. Berger and Berry (1988) maintain that no known method of statistical inference is completely objective, but that a Bayesian approach explicitly acknowledges and controls the subjective element through formulation of the prior distribution. Similarly, research continues into the ethics and practicalities of Bayesian hypothesis testing (Bernardo & Rueda, 2002) and empirical Bayes methods (Carlin & Louis, 2000).

The problem of learning about the model structure from the data, or model identification, remains an area of active research interest. For example, the alternative models, H_i, might represent a set of regression equations with different explanatory variables or mixture models with different numbers of components. While the traditional Bayesian approach to such model selection problems is through Bayes' factors (Bernardo & Smith, 1994), recent alternatives include methods based on building "reversible jumps" between models and then choosing the model as another step in the simulation (Green, 1995), and methods based on birth-and-death processes (Stephens, 2000). Instead of choosing a single model, the Bayesian paradigm also allows *model averaging*, in which parameter estimates, expectations, and predictions can be combined over some or all potential models. Effectively, a weighted sum of estimates over the set of models is taken, with the weights reflecting the posterior probability of each model given the data (Hoeting, Madigan, Raftery, & Volinsky, 1999).

Other models, such as generalized linear models, multivariate models, time-series, and spatial models, have also been cast in a Bayesian framework. See, for example, Carlin and Louis (2000), Congdon (2001), Gilks, Richardson, & Spiegelhalter (1996), and Gelman et al. (1995). These authors also describe methods for different study designs, nonstandard data, model checking, and sensitivity analysis.

Although the computational efficiency of Bayesian learning continues to improve, the computational cost can be high compared to other methods. However, Mitchell (1997) explains that even in those cases where a Bayesian algorithm is not practical as a final implementation, the Bayesian approach can often offer useful insight and/or provide a benchmark for quality decision making against which other methods can be assessed. Sometimes an alternative method can be viewed as a practical approximation to a more rigorous Bayesian method.

CONCLUSION

Bayesian learning differs from other algorithms in the way all relevant knowledge about model parameters and model structure and any associated uncertainty is expressed as probability distributions. In some applications, the Bayesian approach quite simply leads to efficient algorithms. In other applications, it is the Bayesian framework that is appealing. For example, the ability to explicitly combine prior knowledge with observed data is useful in any situation where expert opinion needs to be incorporated into the learning algorithm, in which information from different sources needs to be combined or where a model is regularly updated as more data arrives. The need to provide explicit prior distributions for all parameters is sometimes viewed as a disadvantage of Bayesian learning rather than a strength. Some critics object to the entire Bayesian approach, viewing the choice of the prior as arbitrary. Alternatively, the problem can be how to formulate the prior knowledge in a practical way including those situations where there essentially is no prior knowledge. Use of appropriate representational systems such as Bayesian networks can help.

Bayesian learning quantifies and highlights the inherent uncertainty in the model through the posterior distribution of the model parameters. This also makes it very easy to combine predictions from multiple models through a model-averaging approach rather than being restricted to a single most likely model.

Now established as an analytic tool, Bayesian learning still offers great opportunities for advancements in theory, methodology, computation, and applications. Importantly, the Bayesian revolution is interdisciplinary, bringing together probabilists, statisticians, computer scientists, information scientists, researchers, and practitioners.

REFERENCES

Ahn, J., & Ezawa, K. J. (1997). Decision support for real-time telemarketing operations through Bayesian network learning. *Decision Support Systems, 21*(1), 17-27.

Anderson, R. D., & Vastag, G. (2004). Causal modeling alternatives in operations research: Overview and application. *European Journal of Operational Research, 156*(1), 92-109.

Berger, J. O., & Berry, D. A. (1988). Statistical analysis and the illusion of objectivity. *American Scientist, 76*, 159-65.

Bernardo, J. M., & Rueda, R. (2002). Bayesian hypothesis testing: A reference approach. *International Statistical Review, 70*, 351-372.

Bernardo, J. M., & Smith, A. F. M. (1994). *Bayesian theory*. New York: Wiley.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society: Series B, Statistical Methodology, 36*, 192-236.

Best, N. G., Cowles, M. K., & Vines, S. K. (1995). *CODA: Convergence diagnostics and output analysis software for Gibbs sampling output, Version 0.3* (Tech. Rep.). Cambridge, UK: MRC Biostatistics Unit.

Carlin, B., & Louis, T. A. (2000). *Bayes and empirical Bayes methods for data analysis* (2nd ed.). London: Chapman and Hall.

Cervera, E., & del Pobil, A. P. (1997). Multiple self-organizing maps: A hybrid learning scheme. *Neurocomputing, 16*(4), 309-318.

Chen, M. H., & Schmeiser, B. W. (1998). Towards black-box sampling: A random-direction interior-point Markov chain approach. *Journal of Computational and Graphical Statistics, 7*, 1-22.

Chen, M. H., Shao, Q. M., & Ibrahim, J. G. (2000). *Monte Carlo methods in Bayesian computation*. New York: Springer-Verlag.

Chipman, H. A., George, E. I., & McCulloch, R. E. (1998). Bayesian CART model search. *Journal of the American Statistical Association, 93*, 935-948

Cho, S., & Cha, S. (in press). SAD: Web session anomaly detection based on parameter estimation. *Computers & Security*.

Congdon, P. (2001). *Bayesian statistical modelling*. London: Wiley.

Danaher, S., Datta, S., Waddle, I., & Hackney, P. (2004). Erosion modelling using Bayesian regulated artificial neural networks. *Wear, 256*(9-10), 879-888.

de Campos, L. M., Fernández-Luna, J. M., & Huete, J. F. (2003). The BNR model: Foundations and performance of a Bayesian network-based retrieval model. *International Journal of Approximate Reasoning, 34*(2-3), 265-285.

Denison, D. G. T., Holmes, C. C., Mallick, B. K., & Smith, A. F. M. (2002). *Bayesian methods for nonlinear classification and regression*. New York: Wiley.

Garthwaite, P. H., & O'Hagan, A. (2000). Quantifying expert opinion in the UK water industry: An experimental study. *The Statistician, 49*, 455-477

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. Chapman and Hall Texts in Statistical Science.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice*. London: Chapman and Hall.

Gilks, W. R., Roberts, G. O., & Sahu, S. K. (1998). Adaptive Markov chain Monte Carlo through regeneration. *Jour-*

nal of the American Statistical Association, 93, 1045-1054.

Green, P. J. (1995). Reversible jump MCMC computation and Bayesian model determination. *Biometrika, 82*(4), 711-732

Gustafson, P. (2003). *Measurement error and misclassification in statistics and epidemiology: Impacts and Bayesian adjustments*. Boca Raton, FL: Chapman and Hall/CRC Press.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning: Data mining, inference, and prediction*. Springer Series in Statistics.

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science, 14*, 382-401.

Hsu, W. H. (2004). Genetic wrappers for feature selection in decision tree induction and variable ordering in Bayesian network structure learning. *Information Sciences, 163*(1-3), 103-122.

Ibrahim, J. G., Chen, M. H., & Sinha, D. (2001). *Bayesian survival analysis*. Springer-Verlag.

Jensen, F. V. (2001). *Bayesian networks and decision graphs*. Springer.

Kadane, J. B., & Wolfson, L. J. (1998). Experiences in elicitation. *The Statistician, 47*, 3-19.

Kendall, W. (1998). Perfect simulation for the area-interaction point process. In C. C. Heyde & L. Accardi (Eds.), *Probability towards 2000* (pp. 218-234). New York: Springer-Verlag.

Klopotek, M. (in press). Very large Bayesian multinets for text classification. *Future Generation Computer Systems*.

Kononenko, I. (2001). Machine learning for medical diagnosis: History, state of the art and perspective. *Artificial Intelligence in Medicine, 23*(1), 89-109.

Lee, P. M. (1997). *Bayesian statistics: An introduction* (2nd ed.). London: Arnold.

Lucas, P. J. F., van der Gaag, L. C., & Abu-Hanna, A. (2004). Bayesian networks in biomedicine and healthcare. *Artificial Intelligence in Medicine, 30*(3), 201-214.

Ma, J., Wang, T., & Xu, L. (2004). A gradient BYY harmony learning rule on Gaussian mixture with automated model selection. *Neurocomputing, 56*, 481-487.

Macrossan, P., & Mengersen, K. (2002). Bayesian learning. In R. Sarker, H. A. Abbass, & C. S. Newton (Eds.), *Heuristic and optimization for knowledge discovery*. Hershey, PA: Idea Group Publishing.

Mello, K. L., & Brown, S. D. (1999). Novel "hybrid" classification method employing Bayesian networks. *Journal of Chemometrics, 13*, 579-590.

Mengersen, K. L., & Tweedie, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *The Annals of Statistics*, *24*, 101-121

Michie, D., Spiegelhalter, D. J., & Taylor, C. C. (Eds.). (1994). *Machine learning, neural and statistical classification*. Prentice-Hall Inc.

Mira, A., & Tierney, L. (2002). Efficiency and convergence properties of slice samplers. *Scandinavian Journal of Statistics, 29*, 1-12

Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.

Mladeni, D., & Grobelnik, M. (2003). Feature selection on hierarchy of Web documents. *Decision Support Systems, 35*(1), 45-87.

Moshou, D., Vrindts, E., De Ketelaere, B., De Baerdemaeker, J., & Ramon, H. (2001). A neural network based plant classifier. *Computers and Electronics in Agriculture, 31*(1), 5-16.

Neal, R. M. (1996). Bayesian learning for neural networks. In *Lecture notes in statistics*. New York: Springer-Verlag.

Neal, R. M. (1997). *Markov chain Monte Carlo methods based on "slicing" the density function* (Tech. Rep. No. 9722). University of Toronto.

Pelikan, M., Goldberg, D. E., & Tsutsui, S. (2003). Getting the best of both worlds: Discrete and continuous genetic and evolutionary algorithms in concert. *Information Sciences, 156*(3-4), 147-171.

Propp, J. G., & Wilson, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms, 9*, 223-252.

Robert, C. P. (1994). *The Bayesian choice*. New York: Springer-Verlag.

Robert, C. P. (Ed.). (1998). *Discretization and MCMC convergence assessment*. New York: Springer-Verlag.

Robert, C. P., & Casella, G. (1999). *Monte Carlo statistical methods*. New York: Springer-Verlag.

Sarker, R., Abbass, H. A., & Newton, C. S. (Eds.). (2002). *Heuristic and optimization for knowledge discovery*. Hershey, PA: Idea Group Publishing.

Serrano, N., Andreas, A. E., Savakis, E., & Luo, J. (in press). Improved scene classification using efficient low-level features and semantic cues. *Pattern Recognition*.

Spiegelhalter, D. J., Thomas, A., Best, N. G., & Gilks, W. R. (1995). *BUGS: Bayesian inference using Gibbs sampling*. Cambridge: MRC Biostatistics Unit.

Stephens, M. (2000). Bayesian analysis of mixture models with an unknown number of components: An alternative to reversible jump methods. *Annals of Statistics, 28*(1), 40-74.

Taroni, F., Biedermann, A., Garbolino, P., & Aitken, C. G. G. (2004). A general approach to Bayesian networks for the interpretation of evidence. *Forensic Science International, 139*(1), 5-16.

Tian, L., & Noore, A. (in press). Evolutionary neural network modeling for software cumulative failure time prediction. *Reliability Engineering & System Safety*.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics, 22*, 1701-1786.

Tierney, L. (1998). A note on Metropolis-Hastings kernels for general state spaces. *Annals of Applied Probability, 8*(1), 1-9.

Tierney, L., & Mira, A. (1999). Some adaptive Monte Carlo methods for Bayesian inference. *Statistics in Medicine, 18*, 2507-2515.

West, M., & Harrison, J. (1997). *Bayesian forecasting and dynamic models* (2nd ed.). New York: Springer-Verlag.

Winkler, D. A., & Burden, F. R. (2004). Bayesian neural nets for modeling in drug discovery. *Drug Discovery Today: BIOSILICO, 2*(3), 104-111.

KEY TERMS

Bayes' Rule: Mathematical equation relating prior and posterior probabilities.

Bayesian Learning: Learning algorithm based on Bayes' rule.

BBN: Bayesian belief network. A graphical model of joint probability distributions in which nodes represent random variables and arcs represent (causal) relationships. May also be called a causal net or influence diagram (see Jensen, 2001).

BNN: Bayesian neural network. A neural network where a Bayesian approach is used to calculate the posterior distribution for the weights. Rather than selecting the single most likely set of weights, model averaging is used to predict outputs.

Conditional Distribution: Probability distribution of a parameter, given the values of other parameters and/or the data.

MCMC: Markov chain Monte Carlo computational algorithms for estimating multidimensional integrals.

Posterior Probability: Probability distribution of an unknown characteristic after combining prior information and data.

Prior Probability: Probability distribution of a characteristic prior to observing data. Way of formalizing current knowledge about that characteristic.

Behavioral Factors in Strategic Alliances

Purnendu Mandal
Lamar University, USA

Dale H. Shao
Marshall University, USA

Chong W. Kim
Marshall University, USA

STRATEGIC ALLIANCE

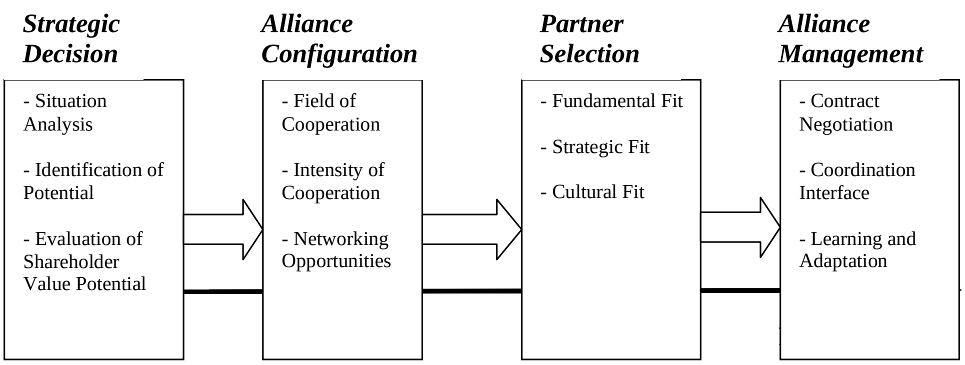
Recently, there has been a growing trend among information technology (IT) organizations to form strategic alliances to increase competitive advantages in the marketplace. For an organization to exploit the benefits of alliances, human factors and IT factors must be among the basic components of any strategic plan (Kemeny & Yanowitz, 2000). Despite the obvious need to consider human and IT factors when developing a long-term plan, many strategic plans developed in the past that led to alliances have failed to consider human aspects. Examples of failure in the implementation of IT systems due to the lack of consideration of human factors have come to light in recent years, but a comprehensive study of the consideration of human factors in the development of strategic alliances resulting in a major IT system alignment for a firm, is still rare in IT literature.

A successful alliance should not imply an imposition of one organization's culture over another. It is not a requirement that both organizations change the social structure, but the unique personalities of the two cultures should be considered when combining the resources of two organizations. The new organization should create a new working culture that brings together the best elements of each (Rule & Keown, 1998). Unfortunately, the creation of a new culture is rarely practiced, as alliances are often viewed solely from a financial perspective, leaving the human issues as something to be dealt with later, and many times with a minimal amount of effort. The creation of a new culture involves operations, sales, human resources management, technology, and structure, as well as many other internal and external entities and forces. It is undoubtedly an expensive and time-consuming endeavor to create a new working culture, but in the end, more value is created, and employees are more content and productive.

Strategic alliances are "co-operative relationships between two or more independent organizations, designed to achieve mutually beneficial business goals for as long as is economically viable" (Paris & Sasson, 2002). The main purpose of an alliance is to create one or more advantages such as product integration, product distribution, or product extension (Pearlson, 2001). In strategic alliances, information resources of different organizations require coordination over extended periods of time.

Bronder and Pritzl (1992) suggest that a strategic alliance exists when the value chains between at least two organizations (with compatible goals) are combined for the purpose of sustaining and/or achieving significantly competitive advantage. They derived four critical phases of a strategic alliance; namely, strategic decision for an

Figure 1. Strategic alliance phases (Bronder& Pritzel, 1992)



alliance, alliance configuration, partner selection, and alliance management, as shown in Figure 1. These four phases provide the basis for a continuous development and review of the strategic alliance, which increases the likelihood of the venture's success.

Typically, the first phase of a strategic alliance is the decision to go forward with the development of a strategic alliance (i.e., it asks this question: Is this strategic alliance justified?). Phase II (Configuration of a Strategic Alliance) focuses on setting-up the alliance's structure. Phase III (Partner Selection) is one of the most important success factors of the strategic alliance. This phase addresses whether the firms that are considering the partnership have characteristics that are conducive to a successful strategic alliance. Some of the concerns in this phase are fundamental fit (e.g., Do the company's activities and expertise complement each other in a way that increases value potential?), strategic fit (e.g., Do strategic goal structures match?), and cultural fit (e.g., Is there a readiness to accept the geographically and internally grown culture of the partners?). The final phase, Phase IV, is concerned with managing a strategic alliance (e.g., How do partners continually manage, evaluate, and negotiate within the alliance to increase the odds of continued success?). People-related issues are the major focus of this phase.

Before an organization commits to a strategic alliance, it should have a management plan developed to deal with the human behavior aspects of the newly created organization. Parise and Sasson (2002) discuss the knowledge management practices that organizations should follow when dealing with a strategic alliance. They break down the creation of a strategic alliance into three major phases.

- *Find:* making alliance strategy decisions and screening and selecting potential partners.
- *Design:* structuring and negotiating an agreement with the partners.
- *Manage:* developing an effective working environment with the partner to facilitate the completion of the actual work. This phase includes collecting data relating to performance and feedback from both partners on how they think the alliance is progressing. Managing relationships and maintaining trust are particularly critical during the Manage Phase.

The application of proper knowledge management techniques is especially important for a successful alliance (Parise & Sasson, 2002). There must be a systematic approach for capturing, codifying, and sharing information and knowledge; a focus on building social capital to enable collaboration among people and communities; an emphasis on learning and training; and a priority on leveraging knowledge and expertise in work practices. Parise and Sasson (2002) suggest a list of the building blocks of alliance management. Four of these building blocks relate specifically to human behavior factors.

- *Social Capital:* Building trust and communication channels that allow unambiguous discussions with the partner is a necessary ingredient for an effective relationship.
- *Communities:* Communities of practice allow for the sharing of personal experiences and tacit knowledge based on individuals with a common interest or practice. Communities can be realized by using electronic meeting rooms, forums, or more formal alliance group structures.
- *Training:* Companies that rely heavily on strategic alliances should provide formal training for managers and team members in their strategic plans. Providing staff with the skills necessary to exist in a new system (in this case, a strategic alliance) is often overlooked in the development of the new system.
- *Formal Processes and Programs:* Alliance know-how should be institutionalized. An example of this is Eli Lilly, a leading pharmaceutical firm, which created a dedicated organization called the Office of Alliance Management, which was responsible for alliance management.

The literature on strategic alliances shows that organizations that use alliance management techniques to provide for stress and knowledge management are more successful than those who do not. Leveraging knowledge management across a company's strategic alliance is a critical success factor for partnering companies. The greatest contributors to knowledge management in an organization are the information-literate knowledge workers—mainly the IT professionals.

CULTURAL ASPECTS IN ALLIANCES

Alliances among firms would naturally result in many organizational changes. Leavitt (1965) concluded that there are four types of interacting variables to consider when dealing with organizational change, especially in large organizations. These variables are task variables, structural variables, technological variables, and human variables. He proposed structural, technological, and people approaches to organizational changes, which derive from interactions among these four variables.

The four variables are highly interdependent so that a change in any one variable usually results in compensatory changes in other variables. The introduction of new technological tools (e.g., computers) may cause changes in structure (communication system), changes in people (their skills and attitudes), and changes in performance and tasks. Therefore, it is imperative to consider all areas that might be affected when a company plans to introduce change to an organization.

Pre-existing people-related problems at a target company often cause many alliances to fail to reach their full financial and strategic potential. Numerous case studies report failure of alliances due to a lack of consideration for the potential impact of behavioral and structural aspects (Numerof & Abrams, 2000). To build an effective alliance, institutions must pay particularly close attention to cultural, personality, and structural incompatibilities. Leaders from alliance institutions need to recognize the personality differences in their managers, as well as the demands required by the stage of the organizational life cycle stage that their organization is in (Segil, 2000). It has also been demonstrated that successful alliance partners share many strong similarities in performance and relationships (e.g., people skills) (Whipple & Frankel, 2000). Understanding potential incompatibilities gives institutions that are contemplating alliances a solid foundation on which to explore the feasibility of joint projects. It also increases the likelihood that the alliance will operate successfully (Whipple & Frankel, 2000).

Successful alliances are impeded when the culture of one or both associations highly differs in value. "High control value" is inconsistent with tolerance for ambiguity and the "willingness to compromise" often required for strategic alliances. Maron and VanBremen (1999) suggest the use of William Bridges' Organizational Character Index, which can be a useful tool for analyzing the cultural differences between two associations to determine how well they might work together. It promotes better understanding between two associations; it fosters an appreciation for what both partners could bring to an alliance; and it identifies underdeveloped qualities in both associations that could inhibit the success of an alliance.

IT ISSUES IN ALLIANCES

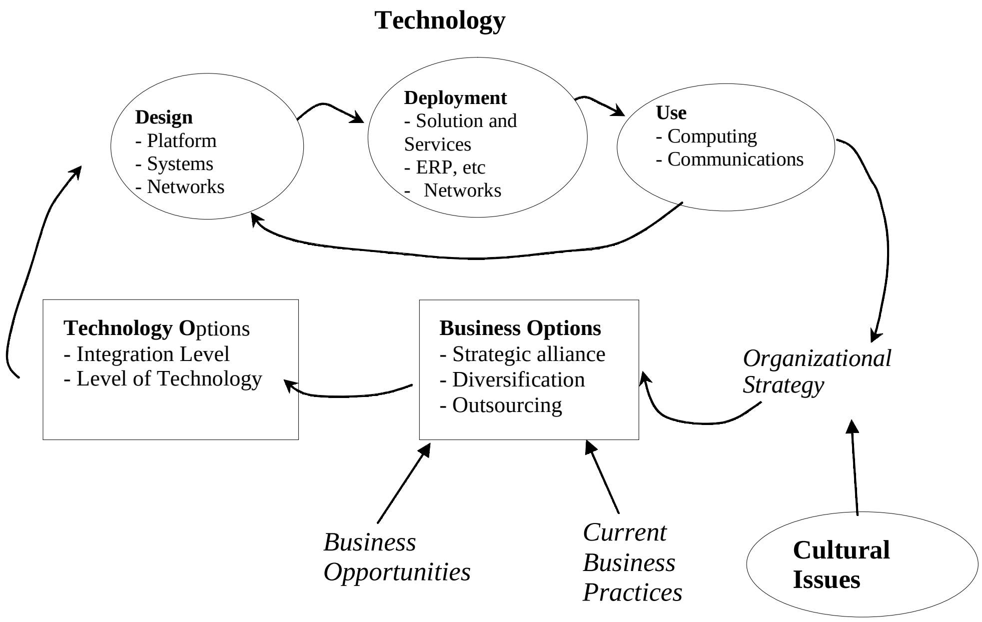
Long-term IT considerations, such as IT architecture, is another major consideration when developing a strategic alliance. A strategic consideration, such as new alliances, requires the visioning of a different IT architecture. Applegate, McFarlan, and McKenney (1999) view IT architecture as an overall picture of the range of technical options available to a firm, as well as standard business options. "Just as the blueprint of a building's architecture indicates not only the structure's design but how everything – from plumbing and heating systems, to the flow of traffic within the building – fits and works together, the blueprint of a firm's IT architecture defines the technical computing, information management and communications platform" (p. 209).

Figure 2 brings out the dynamic nature of the IT architecture development process. The technology component, shown by the dotted oval, is concerned with design, deployment, and how it is used. This part is the core of IT architecture, and a large proportion of IT professionals' time is devoted to these activities. Consideration of business options that feed to various technology options is a higher-level activity in the IT architecture development process. Business options such as strategic alliances, mergers and acquisitions, outsourcing, diversification, and so forth are influenced by major internal, as well as external, factors such as current business practices, business opportunities, and organizational strategy. There is a direct link between technology and organizational strategy. The technology (with its operational and technical settings) exerts a strong influence on the organization's future strategic direction. Thus, one can observe a close link between technical and other business factors, and, like ever changing business, the IT architecture is a dynamically evolving phenomenon (see Figure 2 through connecting lines).

An alliance can exist between any types of organizations. For example, a telecommunications organization could form an alliance for international joint ventures, or an alliance could be established between a banking organization and an IT supplier. The notion of developing a strategic alliance suggests that an organization's performance can be significantly improved through joint, mutually dependent action. For a strategic alliance to be successful, business partners must follow a structured approach to developing their alliances to ensure that all major items have been addressed, and should include, as part of this process, strategic planning, communication, efficient and effective decision-making, performance evaluation, relationship structure, and education and training.

Strategists have often suggested that organizations should consider entering into similar or somewhat related market sectors to broaden their product/service portfolios (Henderson & Clark, 1990; Markides & Williamson, 1997). Both of the dimensions of market (customer and product, Ansoff, 1986) in a related market can be identified easily and strategies formulated for deployment. The main advantage of adopting such a strategy is that an organization can easily use its competencies and strategic assets in generating a strategic competitive advantage (Markides & Williamson, 1997). Determining the design and the requirements of a new information system (IS) is a relatively simple task. In

Figure 2. Forces affecting overall IT architecture



contrast, diversification into a significantly different market for an IT/IS organization is a very challenging task that requires considerable evaluation of the IT infrastructure and human relations.

CONCLUSION

Strategic alliance is a complex business decision that involves careful consideration of business processes, IT architecture, human issues, and many other factors. Over emphasizing one factor or downplaying another may lead to a compromising situation that can have seriously negative consequences for both of the organizations involved. Behavioral issues in their broader sense impact all phases of a strategic alliance. For IT professionals, the understanding of behavioral or human issues is of critical importance in the analysis and design of the system that will support a strategic alliance. The new system must take into account not only traditional system design considerations, but knowledge management and its growing importance in system design, which increases organizational effectiveness and ensures the firm's long-term existence.

REFERENCES

Ansoff, I.H. (1986). *Corporate strategy*. London: Sidwick and Jackson.

Applegate, L.M., McFarlan, F.W., & McKenney, J.L. (1999). *Corporate information systems management: Text and cases*. Boston: Irwin McGraw-Hill.

Bronder, C., & Pritzl, R. (1992). Developing strategic alliances: A successful framework for cooperation. *European Management Journal*, *10*(4), 412-420.

Henderson, R., & Clark, K. (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative Science Quarterly*, *35*, 9-30.

Kemeny, J.M., & Yanowitz, J. (2000, March). Strategic alliances: Finding the hidden leverage for success. *Reflections, 1*(3), 62-71.

Leavitt, H.J. (1965). Applied organizational change in industry: Structural, technological and humanistic approaches. In James March (Ed.), *Handbook of organizations* (pp. 1144-1170). Randy, McNally & Company.

Markides, C.C., & Williamson, P.J. (1997). Related diversification, core competencies and corporate performance. In A. Cambell, & K. Sommer Luchs (Eds.), *Core competency-based strategy* (pp. 96-122). London: International Thomson Business Press.

Maron, R.M., & VanBremen, L. (1999). The influence of organizational culture on strategic alliances. *Association Management, 51*(4), 86-92.

Numerof, R.E., & Abrams, M.N. (2000). Subtle conflicts that wreck merger visions. *Mergers and Acquisitions, 35*(3), 28-30.

Parise, S., & Sasson, L. (2002, March/April). Leveraging knowledge management across strategic alliances. *Ivey Business Journal*, 41-47.

Pearlson, K.E. (2001). *Managing and using information systems*. New York: John Wiley & Sons.

Rule, E. & Keown, S. (1998, September/October). Competencies of high-performing strategic alliances. *Strategy & Leadership, 26*(4), 36-38.

Segil, L. (2000). Understanding life cycle differences. *Association Management, 52*(8), 32-33.

Whipple, J., & Frankel, R. (2000, Summer). Strategic alliance success factors. *The Journal of Supply Chain Management: A Global Review of Purchasing and Supply*, 21-28.

KEY TERMS

Alliance Management: Allows two different organizations to effectively work together and combine resources, which is expected to bring benefits to both organizations.

Culture: A societal manifestation influenced by traditions, religion, history, acceptable behavior, and many other factors.

IT Strategic Alliance: A broad agreement between business partners to operate cooperatively, usually facilitated by IT systems.

IT Architecture: A conceptual framework of IT in an organization that supports business processes. IT includes hardware, software, telecommunications, database management, and other information processing technologies used in computer-based information systems.

Information-Literate Knowledge Worker: A worker who knows what information is needed, knows how and where to obtain that information, understands the meaning of information, and can act based on the information to help the organization achieve its greatest advantage.

Strategic Planning: Corporate long-term planning that ensures the existence of an organization.

Behavioral Perspective of Groupware Adoption

Yining Chen
Ohio University, USA

Hao Lou
Ohio University, USA

INTRODUCTION

Over the past decade, groupware technologies, such as e-mail, electronic bulletin boards, and group support systems, have become an important part of the business-computing infrastructure in many organizations. This evolving software category has captured the attention and imagination of information technology professionals, line of business managers and end users, not to mention software suppliers. Organizations adopt groupware applications to enhance communication, collaboration, and coordination among group members and thus improve group performance ("Lotus Development," 1995). While some groupware applications, for example, e-mail, have been commonly accepted, many other applications, especially those that require significant collaboration and cooperation among users, are not widely used in organizations and their potential benefits are far from being fully realized (Orlikowski, 1993). Although many laboratory and field studies have consistently shown the relevance and positive impact of group support systems on group work, more research is needed in understanding how to increase the rate of diffusion and adoption of the technology (Nunamaker, 1997).

Behavioral-related elements (e.g., an individual's normative beliefs, attitude, and motivation), recognized by many, are the primary causes of users' resistance toward a newly implemented system or technology. Information technology (IT) research, however, tends to under-utilize existing knowledge in the behavioral science (Burton et al., 1993; DeSanctis, 1983; Melone, 1990; Turner, 1982). Expectancy theory has been recognized as one of the most promising conceptualizations of individual motivation (Melone, 1990; Snead & Harrell, 1995). Many researchers have proposed that expectancy theory can provide an appropriate theoretical framework for research that examines a user's acceptance of and intent to use a system (DeSanctis, 1983; Melone, 1990). This study uses expectancy theory as part of a student-based experiment to examine users' behavioral intention (motivation) to utilize a groupware application.

BACKGROUND

Groupware refers to a class of computer technologies designed to support communication, collaboration, and cooperation among a group of knowledge workers. It covers a variety of technologies, ranging from simple e-mail systems to complex workflow applications. Although the use of some groupware technologies, such as e-mail, has become ubiquitous, organizations have encountered many difficulties in adopting and utilizing more sophisticated groupware applications, such as group support systems and Lotus Notes (Nunamaker, 1997; Orlikowski, 1993).

Groupware applications are designed to support communication, cooperation, and collaboration among a group of users rather than to improve productivity of individuals. Therefore, usage and resulting benefits are only achieved if a majority of the users whose work is affected by a groupware application accept and use the system (Grudin, 1994). Otherwise, the application will not only fail to improve group performance but will also create additional communication and coordination barriers. While many factors (e.g., users' background and commitment, organizations' reward systems, work norms, and policies and procedures) can contribute to the success of a groupware application, achieving a "critical mass" of users has been recognized as one of the keys for successful groupware acceptance (Ehrlich, 1987; Grudin, 1994; Markus, 1990; Markus & Connolly, 1990). The intrinsic value of a groupware technology increases and becomes more apparent as more and more users accept the technology. Consequently, more and more functions are available to adopters, which in turn reinforce their opinion about the technology and reaffirm their acceptance decisions.

Expectancy theory is considered one of the most promising conceptualizations of individual motivation (DeSanctis, 1983; Melone, 1990). Expectancy models are cognitive explanations of human behavior that cast a person as an active, thinking, predicting creature in his or her environment. He or she continuously evaluates the outcomes of his or her behavior and subjectively assesses the likelihood that each of his or her possible actions will lead to various outcomes. The choice of the amount of effort he or she exerts is based on a systematic analysis of (1) the values of the rewards from these outcomes, (2) the likelihood that rewards will result from these outcomes, and (3) the likelihood of reaching these outcomes through his or her actions and efforts.

According to Vroom (1964), expectancy theory is comprised of two related models: the valence model and the force model. In our application of the theory, each user of a groupware application first uses the valence model to evaluate the application's outcomes (e.g., enhanced communication, increased ability to coordinate, better collaboration, and improved competence) and subjectively assesses the likelihood that these outcomes will occur. Next, by placing his or her intrinsic values (or weights) on the various outcomes, each user evaluates the overall attractiveness of the groupware application. Finally, the user uses the force model to determine the amount of effort he or she is willing to exert to use the application. This effort level is determined by the product of the attractiveness generated by the valence model and the likelihood that his or her effort will result in a successful use of the application. Based on this systematic analysis, the user will determine how much effort he or she would like to exert in using the groupware application.

The general research question examined by this study is "Can the valence and force models of expectancy theory explain the motivation of a user to use a groupware application?" Specifically, under the valence model, we investigate the impact of the potential outcomes of a groupware application upon users' motivation to use such an application. The four outcomes of groupware applications examined by this study are (1) enhancing communications among coworkers; (2) increasing ability to coordinate activities; (3) facilitating collaboration among coworkers; and (4) improving competence of job performance. Under the force model, we investigate the extent that the difficulty of using a groupware application will affect users' motivation to actually use the application.

RESEARCH METHOD AND RESULTS

The within-person or individual focus of expectancy theory suggests that appropriate tests of this theory should involve comparing measurements of the same individual's motivation under different circumstances (Harrell, Caldwell & Doty, 1985; Murray & Frazier, 1986). In response to this suggestion, this study adapts a well-established within-person methodology originally developed by Stahl and Harrell (1981) and later proven to be valid by other studies in various circumstances (e.g., Burton, Chen, Grover & Stewart, 1993; Geiger & Cooper, 1996; Snead & Harrell, 1995).This methodology uses a judgment modeling decision exercise that provides a set

Table 1.

If you use the groupware application (e.g., Domino Discussion or Lotus Notes) to the MAXIMUM extent in your job, the likelihood that:	
You will enhance your communications with your coworkers is	HIGH (90%)
You will improve your ability to coordinate job-related activities is	HIGH (90%)
You will achieve a better collaboration among your coworkers is	HIGH (90%)
You will increase your general level of competence in performing your job is ………	LOW (10%)

Table 2.

DECISION A: With the above outcomes and associated likelihood levels in mind, indicate the ***attractiveness*** to you of using the groupware application in your job.

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Very Unattractive										Very Attractive

FURTHER INFORMATION: If you exert a great deal of effort to use Lotus Notes in your job, the likelihood that you will be successful in doing so is .... LOW (10%)

DECISION B: Keeping in mind your attractiveness decision (DECISION A) and the FURTHER INFORMATION, indicate the level of ***effort*** you would exert to use the groupware application.

0	1	2	3	4	5	6	7	8	9	10
Zero Effort										Great Deal of Effort

of cues, which an individual uses in arriving at a particular judgment or decision. Multiple sets of these cues are presented, with each representing a unique combination of strengths or values associated with the cues. A separate judgment is required from the individual for each unique combination of cues presented.

The participants are 86 business undergraduate students who enrolled in courses that utilize groupware applications. In each of the 16 proposed cases (groupware applications), the participants are asked to make two decisions. The first decision, Decision A, represents the overall attractiveness of using the groupware application, given the likelihood (10% or 90%) that the four second-level outcomes would result from the use. The second decision, Decision B, reflects the strength of a participant's motivation to use the groupware application, using (1) the attractiveness of the application obtained from Decision A and (2) the expectancy (10% or 90%) that, if the user exerts a great deal of effort, he or she would be successful in using the application. As an example, one of the 16 cases is presented next.

If you use the groupware application (e.g., Domino Discussion or Lotus Notes) to the MAXIMUM extent in your job, the likelihood that:

We first examine whether the valence model of expectancy theory can explain a user's perception of the attractiveness of using a groupware application. Through the use of multiple regression analysis, we seek to explain each participant's perception of the attractiveness of using a groupware application. Decision A serves as the dependent variable, and the four second-level outcome instruments serve as the independent variables. The resulting standardized regression coefficients represent the relative importance (attractiveness) of each of the second-level outcomes to each user in arriving at Decision A. The mean adjusted-R^2 of the regressions and the mean standardized betas of each outcome are presented in Table 1.

As indicated in Table 1, the mean R^2 of the individual regression models is .6876. The mean R^2 represents the percentage of total variation in users' response that is explained by the multiple regression. Thus, the relatively high mean R^2 indicates that the valence model of expectancy theory explains much of the variation in users' perception of the attractiveness of using a groupware application. Among the 86 individual regression models, 79 are significant at the level of .05.

The standardized betas of V1, V2, V3, and V4 are significant, at the level of .05, for more than half of the individuals. This implies that all four of the secondary outcomes are important factors, to a majority of the individuals, in determining the attractiveness of a groupware application. Although all four factors are important, some factors are more important than others. It is the *mean* of these standardized betas that explains how users, on average, assess the attractiveness of potential outcomes resulting from a groupware application. The users, on average, place the highest valence on the outcome V4. The other valences, in descending order of their strength, are V1, V2, and V3. These results imply that

*Table 3. Valence and force model regression results**

	N	Mean	Standard Deviatio n	Range	Frequency of Significance at .05
Valence Model:					
Adjusted R^2	86	.6876	.2034	-.0267 to .9388	79/86
Standardized Beta Weight					
V1	86	.3748	.1745	-.4423 to .7646	62/86
V2	86	.3320	.1619	-.1506 to .6129	53/86
V3	86	.3190	.1830	-.5897 to .6803	51/86
V4	86	.5197	.2444	-.3965 to .9197	73/86
Force Model:					
Adjusted R^2	86	.7205	.2301	-.1141 to .9999	75/86
Standardized Beta Weight					
B1	86	.5997	.2530	-.1960 to 1.000	72/86
B2	86	.4976	.3110	-.2302 to .9763	64/86

* Results (i.e., mean, standard deviation, range, and frequency of significant at .05) of individual within-person regression models are reported in this table.
V1: valence of communication enhanced V2: valence of coordination ability increased
V3: valence of collaboration improvement V4: valence of competence improvement
B1: weight placed on attractiveness of the groupware application
B2: weight placed on the expectancy of successfully using the system

the users believe improving job competence (V4) is the most attractive outcome of a groupware application and improving collaboration among coworkers (V3) is the least attractive outcome. In the middle is the enhanced communication (V1) and increased coordination (V2).

We then examine whether the force model can explain a user's motivation to use a newly implemented groupware application. We again use multiple regression analysis to examine the force model (Decision B) in the experiment. The dependent variable is the individual's level of effort to use the groupware application. The two independent variables are (1) each user's perception about the attractiveness of the application from Decision A, and (2) the expectancy information (10% or 90%).

The mean R^2 (.7205) indicates that the force model sufficiently explains the users' motivation of using a groupware application. The mean standardized regression coefficient B1 (.5997) indicates the impact of the overall attractiveness of the groupware application, while B2 (.4976) indicates the impact of the expectation that a certain level of effort leads to successfully using the application. These results imply that both factors, the attractiveness of the groupware application (B1) and the likelihood that the user's efforts will lead to a successful use (B2), are of similar importance to the user's motivation.

FUTURE TRENDS

This study successfully applies a behavioral theory, expectancy theory, to a system implementation area. This application (1) helps close the gap between the capabilities of a groupware application and the extent to which it is used, and (2) responds to the claim of previous research that the gap can be better explained by behavioral elements rather than by technical attributes.

Future research should revalidate the application of expectancy theory in different contexts. Various factors such as social norms, one's job requirements, and organization's reward system can be examined for their impact on the valence and force models. Along with the direction of several recent studies (Lucas & Spitler, 1999; Szajna & Scamell, 1993), the relationship among attitude (i.e., perceived system quality, perceived usefulness, and perceived ease of use), intention, and actual use needs to be further validated. The ultimate goal of this line of research is to gain more rigorous and consistent insight into understanding the effectiveness of groupware applications and our ability to explain or predict user acceptance to a groupware application. To increase practical value and external validity, field studies that offer different conditions than lab environments using student subjects are necessary.

CONCLUSION

Groupware technologies have become an important part of the business computing and communicating infrastructure in many organizations. However, literature suggests that many groupware applications, especially those that require significant collaboration and cooperation among users, are still not adequately used (Nunamaker, 1997; Orlikowski, 1993). Their potential benefits are far from being fully realized due to the lack of user acceptance (Venkatesh et al., 2003). While there are studies that show the relevance and positive impact of group support systems on group work, very few have looked into users' perception of the groupware technologies and their motivation to participate (Nunamaker, 1997).

Expectancy theory is considered one of the most promising models of individual motivation. This study examines the use of expectancy theory in explaining the behavioral intention (motivation) to use a groupware application. Our results suggest that expectancy theory is appropriate for assessing and understanding users' motivation to use a groupware application and, subsequently, its acceptance and success. Since user acceptance is an essential antecedent of a successful groupware application, the results of this study should be considered thoughtfully when a groupware application is designed, implemented, and operated.

REFERENCES

Burton, G.F., Chen, Y., Grover V., & Stewart, K.A. (1993). An application of expectancy theory for assessing user motivation to utilize an expert system. *Journal of Management Information Systems, 9*(3), 183-198.

DeSanctis, G. (1983). Expectancy theory as explanation of voluntary use of a decision support system. *Psychological Reports, 52*(1), 247-260.

Ehrlich, S.F. (1987). Strategies for encouraging successful adoption of office communication systems. *ACM Transactions on Office Information Systems, 5*(4), 340-357.

Geiger, M.A., & Cooper, E.A. (1996). Using expectancy theory to assess student motivation. *Issues in Accounting Education, 11*(1), 113-129.

Grudin, J. (1994). Groupware and social dynamics: Eight challenges for developers. *Communications of the ACM, 37*(1), 93-75.

Harrell, A.M., Caldwell, C., & Doty, E. (1985). Within-person expectancy theory predictions of accounting students' motivation to achieve academic success. *Accounting Review, 60*(4), 724-735.

Lotus Development Corporation. (1995). *White Paper*. Boston, MA.

Lucas, H.C., Jr., & Spitler, V.K. (1999). Technology use and performance: A field study of broker workstations. *Decision Sciences, 30*(2), 291-311.

Markus, M.L. (1990). Toward a critical mass theory of interactive media. In J. Fulk & C.E. Steinfield (Eds.), *Organizations and communication technology* (pp. 194-218). Newbury Park, CA.

Markus, M.L., & Connolly, T. (1990). Why CSCW applications fail: Problems in the adoption of interdependent work tools. *Proceedings of CSCW '90*, Los Angeles, CA.

Melone, N.P. (1990). A theoretical assessment of the user-satisfaction construct in information systems research. *Management Science*, *36*(1), 76-89.

Murray, D., & Frazier, K.B. (1986). A within-subjects test of expectancy theory in a public accounting environment. *Journal of Accounting Research, 24*(2), 400-404.

Nunamaker, J.F., Jr. (1997). Future research in group support systems: Needs, some questions and possible directions. *International Journal of Human-Computer Studies*, *47*(3), 357-385.

Orlikowski, W.J. (1993). Learning from notes: Organizational issues in groupware implementation. *Information Society*, *9*(3), 237-250.

Snead, K.C., & Harrell, A.M. (1995). An application of expectancy theory to explain a manager's intention to use a decision support system. *Decision Sciences, 25*(4), 499-513.

Stahl, M.J., & Harrell, A.M. (1981). Modeling effort decisions with behavioral decision theory: Toward an individual differences model of expectancy theory. *Organizational Behavior and Human Performance, 27*(3), 303-325.

Szajna, B., & Scamell, R.W. (1993). The effects of information system user expectations on their performance and perceptions. *MIS Quarterly, 17*(4), 493-516.

Turner, J.A. (1982). Observations on the use of behavioral models in information systems research and practice. *Information and Management, 5*(6), 207-213.

Venkatesh, V., Morris, M.G., Davis, G.B., & Davis, F.D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425-478.

Vroom, V.C. (1964). *Work and motivation*. New York: John Wiley & Sons.

KEY TERMS

Critical Mass: A point at which a sufficient number of individuals have adopted an interactive communication technology to cause a rapid acceleration in its diffusion.

Expectancy Theory: Expectancy models are cognitive explanations of human behavior that cast a person as an active, thinking, predicting creature in his or her environment. He or she continuously evaluates the outcomes of his or her behavior and subjectively assesses the likelihood that each of his or her possible actions will lead to various outcomes.

Force Model: The force model shows that a user's motivation to exert effort into using an application is the summation of the products of the attractiveness of the application and the probability that a certain level of effort will result in successfully using the application.

Group Support Systems: Group support systems provide computer based support for group communication, decision making, and work activities of co-located (same time, same place) or dispersed (same time, different place; or different time, different place) members of a group.

Groupware Application: A class of computer technologies designed to support communication, collaboration, and cooperation among a group of knowledge workers.

User Acceptance: With respect to information technology in general, user acceptance has been conceptualized in several ways. First, it is equated with routinization, the point at which an innovation ceases to be perceived as something unusual and is considered to be a routine part of the process into which it was introduced. Second, user acceptance is equated with time since adoption, implying that an innovation that endures in an organization has been accepted. Finally, user acceptance is equated with usage. Usage is considered an objective measure of acceptance when certain conditions are met: (a) users are motivated to use the system; (b) users have convenient access to the system; and (c) users understand the system and what it can do.

Valence Model: The valence model shows that the overall attractiveness of an application to a user is the summation of the products of the attractiveness of those outcomes associated with the application and the probability that the application will produce those outcomes.

B

Benefits and Challenges of Blended Learning Environments

Charles R. Graham
Brigham Young University, USA

Stephanie Allen
Brigham Young University, USA

Donna Ure
Brigham Young University, USA

INTRODUCTION

The term "blended learning" has become a corporate buzzword in recent years (Lamb, 2001). Recently, the American Society for Training and Development identified blended learning as one of the top ten trends to emerge in the knowledge delivery industry in 2003 (Rooney, 2003). In higher education, the term blended learning is being used with increased frequency in academic conferences and publications. Issues related to the design and implementation of blended learning environments (BLE) are surfacing as technological advances continue to blur the lines between distributed learning and the traditional campus-based learning. Many universities are beginning to recognize the advantages of blending online and residential instruction. *The Chronicle of Higher Education* recently quoted the president of Pennsylvania State University as saying that the convergence between online and residential instruction was "the single-greatest unrecognized trend in higher education today" (Young, 2002). Along the same lines, the editor of *The Journal of Asynchronous Learning Networks* is predicting a dramatic increase in the number of hybrid (i.e., blended) courses to include as many as 80-90% of the range of courses (Young, 2002). The article provides an overview of blended learning environments (BLEs) and outlines the most common benefits and challenges identified in the research literature.

BACKGROUND

The use of the term "blended learning" is a relatively new phenomenon in higher education. Historically, academicians have referred to blended learning environments (BLEs) as hybrid environments. But with the explosion in the use of the term "blended learning" in corporate training environments, the academic literature has increasingly followed suit, and it is common to see the terms used interchangeably (Voos, 2003). In this section of the article, we address the following two questions:

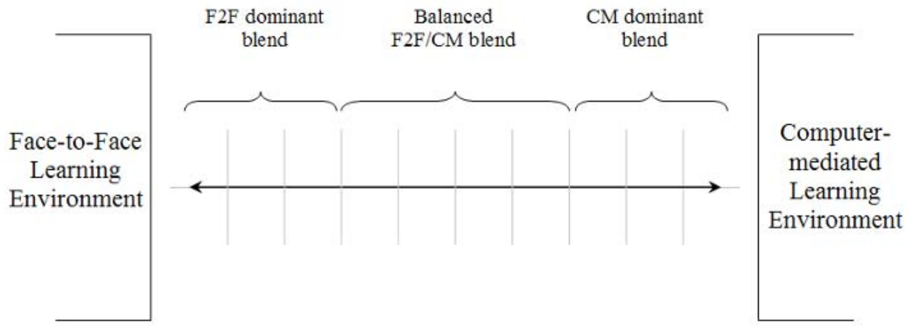
- What is being blended in a BLE?
- How much to blend in a BLE?

What Is Being Blended?

By nature, both the terms "hybrid" and "blended" imply a mixing or combining of *something*. It is that *something* that people do not always agree upon. Some understand blended learning to be a combination of different instructional methods (soft technologies) (Singh & Reed, 2001; Thomson, 2002), while others define blended learning as a combination of different modalities or delivery media (hard technologies) (Driscoll, 2002; Rossett, 2002). Blended learning is most commonly considered to be the combination of two archetypal "learning environments" using *both* the hard and soft technologies most common in each instructional environment. In short, *blended learning environments combine face-to-face (F2F) instruction with computer-mediated (CM) instruction*.

Blending occurs at the instructional (or course) level as opposed to the institutional level. A whole body of literature talks about dual-mode institutions that deliver both F2F and distributed instruction, but don't explicitly blend these environments at a course level (Rumble, 1992). Historically, the on-campus and distributed education branches of dual-mode universities served different populations of learners. However, increasingly the lines between traditional on-campus learners and distance learners are being blurred. This same phenomenon is happening between on-campus course offerings and distributed course offerings. This blurring of boundaries is often referred to as the "hybridization" of the university (Cookson, 2002).

Figure 1. Blended learning environments combine F2F and computer-mediated instruction



How Much to Blend?

As might be expected, no magic blend is optimal for all learning contexts. As Figure 1 suggests, a range of combinations can occur in a blended environment. Figure 1 divides this range into three general levels: blends that have a dominant F2F component, blends that have a dominant CM component, and blends that are fairly balanced in mixing the two environments. In higher education and corporate training, blends of all varieties exist. At the F2F end of the spectrum, many on-campus instructors and corporate trainers are enhancing their courses or training programs by using computer-based technologies. In these instances, the instructors and trainers may change what they do in the F2F environment because of the added CM portion, but they typically do not reduce the F2F contact time. At the computer-mediated end of the spectrum, an increasing number of higher education distributed education courses have a F2F component. These courses range from requiring F2F orientation activities and in-person testing (Martyn, 2003; Schrum & Benson, 2000) to allowing for optional participation in discussion or lecture sessions. In the corporate world, companies often add F2F sessions to e-learning training modules (Bielawski & Metcalf, 2002; Thorne, 2003) to give employees the chance to practice and apply skills and knowledge they've gained via the online instruction. In the middle of the spectrum, both university courses and corporate training modules reduce F2F class time by increasing the time the learners spend in online instructional activities.

Why Blend?

There are many reasons why an instructor or corporate trainer might choose to design a BLE over a non-blended environment. The most predominant benefits and challenges in the literature are presented in the following two sections.

Benefits to Blending

The phrase most commonly used by advocates of BLEs is that they allow one to have the "best of both worlds" (Morgan, 2002; Young, 2002). BLEs can also mix the least effective elements of both worlds if they are not designed well. Beyond this general statement, we identified three major themes that are often referred to as reasons for blending: (1) more effective pedagogy, (2) increased convenience and access, and (3) increased cost effectiveness.

More Effective Pedagogy

The opportunity to improve upon prevalent pedagogical practices is one of the most commonly cited possibilities that blending provides. For example, in the on-campus environment much of teaching and learning is still focused on the "transmission" model with the lecture used by 83% of higher education instructors as the predominant teaching strategy (U.S. Department of Education, 2001). Constraints such as class duration, size, and location can provide a formidable barrier to making changes to that strategy. Introducing online instructional components opens the range of instructional strategies that can be used. Proponents of BLEs have mentioned such benefits as:

- an increase in active learning strategies used (Collis, 2003; Morgan, 2002);
- a change from a more teacher-centered to learner-centered focus (Hartman, Dziuban, & Moskal, 1999; Morgan, 2002);
- a greater emphasis on peer-to-peer learning (Collis, 2003);
- a change in the way faculty allocate time, allowing for increased mentoring of individual students

(Bourne, 1998; Waddoups, Hatch, & Butterworth, 2003); and
- the possibility for interaction with remote experts or remote peer review of projects (Levine & Wake, 2000).

In distance education, there is often a similar problem of focusing on "transmissive" rather than "interactive" learning strategies (Waddoups & Howell, 2002). This typically results from making large quantities of information available on the Web for students to learn independently of anyone else. Many learners are not ready for this type of independent online learning and feel isolated in distance learning courses as a result. BLEs are seen as one way to balance independent learning with human interaction (Hartman et al., 1999; Morgan, 2002). Adding a F2F component to a distance course or online training program can improve social interaction and facilitate future online interaction (Willett, 2002). Occasional F2F interaction can also motivate students to discipline themselves in a predominantly online course (Leh, 2002).

Convenience and Access

Many learners want the convenience of a distributed learning environment, yet don't want to sacrifice the human touch they are used to in the F2F class environment. BLEs are viewed as a way to balance these two factors by increasing convenience, but still maintaining some F2F contact (Collis, 2003; Morgan, 2002). Reducing the amount of F2F seat time reduces time and place constraints for learners (Hartman et al., 1999; Leh, 2002), as well as commuting stress (Willett, 2002) for individuals who travel distances and then have to fight for parking on campuses with large commuter populations.

In corporate settings, increased convenience and access to instructional materials are also compelling reasons for adopting blended learning solutions. Access to training materials was significantly increased when many companies adopted e-learning materials in the form of self-paced computer-based training and Web-based training. However, interaction with fellow learners and instructors was severely limited, if not completely

Table 1. Four general categories describing technology use in Pew grant projects

Technology Use	Description
Technology Enhanced	• F2F contact time (seat time) is not reduced. • Technology may be used to engage learners in activities that were not possible or practical in a F2F classroom (e.g., simulations, online tutorials, individualized assessment and feedback, etc.). • Technology may change the nature of the F2F contact (e.g., information transmission via technology enabling discussion rather than lecture in class; change from lecture to tutoring in computer lab, etc.)
Reduced Seat Time	• F2F contact time is reduced to accommodate time spent in computer-mediated learning activities.
Entirely Distributed	• Courses that have no F2F contact with instructors or learners and are conducted entirely via the internet or other technological tool. • Courses may include both synchronous or synchronous distributed interaction. • Courses may be self paced independent study or more collaborative.
Distributed with optional F2F sessions	• Required components of courses are entirely distributed (see above) but optional F2F sessions can be attended (lectures, discussion sessions, one-on-one lab help, etc.)

annihilated with the computer-mediated delivery method on its own. As a result, completion rates of e-learning courses often sagged dramatically (Singh & Reed, 2001). Much of the slump has been attributed to the lack of a social, interactive, mentored learning environment. Blended learning approaches are popular because they provide the convenience and accessibility of e-learning environments, while simultaneously promoting personal interaction.

Cost Effectiveness

Cost effectiveness is a third major benefit that can result from using BLEs. On the academic side, The Center for Academic Transformation, with the support from the Pew Charitable Trusts, recently completed a three-year grant program designed to support universities in redesigning their instructional approaches using technology to achieve cost savings and quality enhancements. Grant awards were used in the redesign of large enrollment introductory courses offered at 30 institutions across the U.S. In order to achieve cost savings, many of the redesign efforts transformed traditional F2F lecture-based course offerings to courses with a blend of F2F and CM instruction. Table 1 details four general patterns of technology use identified in the PEW-sponsored courses.

Just as higher educational institutions look for cost savings from BLEs, most corporations that adopt blended learning solutions do so because they are a cost-effective means of training and distributing critical information to employees. Numerous studies indicate that travel costs and time associated with corporate training objectives could be reduced by as much as 85% using blended learning methods (Singh & Reed, 2001). For example, a report on an executive MBA program for mid-career doctors shows that blended learning programs can be completed in half the time and at less than half the cost of classroom based instruction (Jones International University, 2002). Another report on corporate blended learning indicates that blended learning solutions can be cost-effective for companies because they provide a means of reaching a large, globally-dispersed audience in a short period of time with both a consistent and a semi-personal content delivery (Bersin & Associates, 2003).

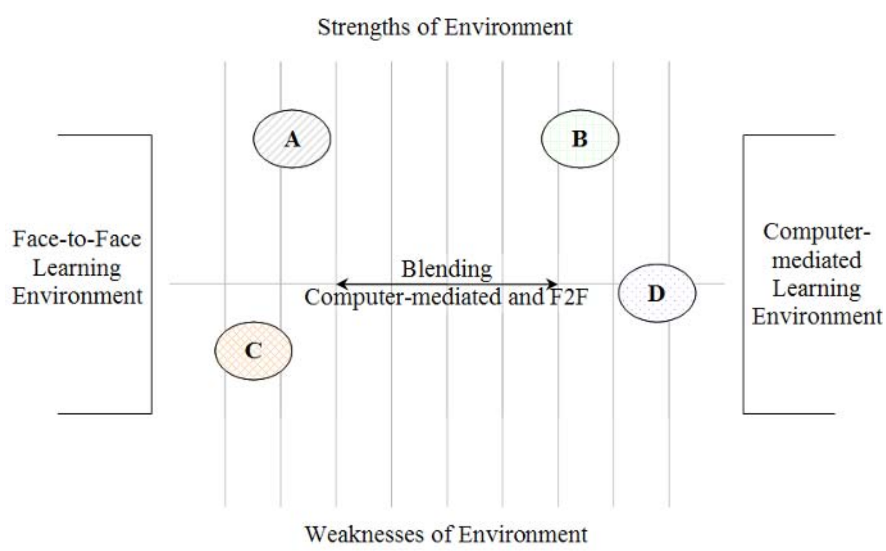
Challenges to Blending

The current literature tends to focus much more on the positives of BLEs than the challenges that institutions, instructors, and learners face from adopting BLEs. Three major categories of challenges addressed in the literature are: (1) finding the "right" blend, (2) the increased demand on time, and (3) overcoming barriers of institutional culture. These three main challenges will be addressed individually in the following sections.

Finding the Right Blend

The most significant challenge faced by people developing and delivering BLEs is identifying the instructional

Figure 2. Blending the strengths of F2F and CM learning environments (Adapted from Osguthorpe and Graham, 2003, p. 229)



strategies that match well with the conditions that are present in these two quite different environments. This challenge is complex because it relates to achieving the right blend, both from a learning and cost-effective standpoint.

Both F2F and CM learning environments have affordances that endow them with particular strengths and weaknesses. The affordances enable the effective use of particular instructional methods within the specific learning environment. By blending F2F with CM instruction, the range of possible instructional strategies that can be used is greatly increased. From a pedagogical standpoint, the goal of blending learning environments for a particular context and audience is to find an appropriate combination of the two environments that takes advantages of the strengths of each environment and avoids their weaknesses (Martyn, 2003; Osguthorpe & Graham, 2003). Figure 2 adds a dimension to Figure 1 and shows how blended solutions can either capitalize on the strengths of both learning environments or not. In Figure 2, for example, blends A and B both capitalize on the strengths of both environments while blends C and D use methods in an environment for which they are not well suited.

Increased Demand on Time

Both faculty instructors and corporate trainers are generally comfortable with creating and presenting instructional materials either in a F2F or in a CM learning environment, but not necessarily in both learning environments. When institutions decide to utilize both learning environments for a single course offering, the time demands of the instructor or trainer increase because now instructional materials must be develop for both CM and F2F environments. Additionally, instructors and trainers typically incur an increase in the time they spend interacting with learners in BLEs. When BLEs are instituted, instructors and trainers who have typically only interacted with students in either F2F or CM learning environments now are responsible for interacting with the students in both environments. As Hartman studied BLEs in higher educational institutions, he found that adding an online component to a F2F course put increased time demands and stress on the faculty developing and delivering the blended course (Hartman et al., 1999).

Overcoming Barriers of Institutional Culture

There are cultural barriers for both learners and instructors that must be overcome in order to use BLEs. The CM components of BLEs require a large amount of self discipline on the part of the learners because learning in that setting is largely independent (Collis, 2003). In current higher educational online learning environments, students tend to procrastinate when they have little required contact (Leh, 2002). In corporate settings, many employees don't finish online courses because they lack the discipline or the motivation to complete the courses. Currently, the culture in both higher educational institutions and corporations allows for student dropouts and don't necessarily require the learners to have the discipline to finish an online course. If BLEs are to be highly successful, the culture of these organizations, as it relates to persistence, must change.

Another cultural element that can work against the use of BLEs is organizational and management support. In higher educational institutions, faculty may hesitate to try blended approaches because they are not sure that they have departmental support or that it fits into the culture of the department or greater institution (Hartman et al., 1999). Similarly, management support for BLEs is essential if they are to succeed in corporations because executives have a large influence on company culture (Bersin & Associates, 2003).

FUTURE TRENDS

Blended learning environments promise to be an important part of the future of both higher education and corporate training. Over the past decade, with the increased availability of technology and network access, the use of BLEs has steadily grown. However, the amount of research done related to the design and use of BLEs is relatively small and additional research is needed. In particular, research is needed that will help instructors to understand the strengths and weaknesses of methods used in F2F and CM instructional environments and know how to appropriately combine both types of instruction. Cases across various levels and types of blends need to be documented and analyzed in order to better understand the tradeoffs that must be made when designing instruction in a blended learning environment.

CONCLUSION

During the past decade, online learning has made huge strides in popularity in both higher education and corporate sectors of society. The use of technology has increased access to educational resources and facilitated communication in a way that was not previously possible. Despite the strengths that online learning environments provide, there are different strengths inherent in traditional face-to-face learning environments. The current trend toward blending both online and F2F instruction is a positive direction that merits our further attention and study.

REFERENCES

Bersin & Associates. (2003). *Blended learning: What works?: An industry study of the strategy, implementation, and impact of blended learning*. Bersin & Associates.

Bielawski, L., & Metcalf, D. (2002). *Blended elearning: Integrating knowledge, performance, support, and online learning*. Amherst, MA: Human Resource Development Press Inc.

Bourne, J.R. (1998). Net-learning: Strategies for on-campus and off-campus network-enabled learning. *Journal of Asynchronous Learning Networks, 2*(2), 70-88.

Collis, B. (2003). Course redesign for blended learning: Modern optics for technical professionals. *International Journal of Continuing Engineering Education and Lifelong Learning, 13*(1/2), 22-38.

Cookson, P. (2002). The hybridization of higher education: Cross-national perspectives. *International Review of Research in Open and Distance Learning, 2*(2).

Driscoll, M. (2002, March 1). Blended learning: Let's get beyond the hype. *E-Learning,* 54.

Hartman, J.L., Dziuban, C., & Moskal, P. (1999, August 16-18). Faculty satisfaction in ALNs: A dependent or independent variable? Paper presented at the *Sloan Summer ALN Workshops: Learning Effectiveness and Faculty Satisfaction*, Urbana, IL.

Jones International University. (2002). *E-learning: Going the distance*. Jones International University. Retrieved July 9, 2003, from the World Wide Web at *http://www.international.edu/eprise/main/PressReleases/e-Learning_White_Paper.pdf*

Lamb, J. (2001, June 4). "Blended learning" is the new buzz phrase. *FT.com*.

Leh, A.S. (2002). Action research on hybrid courses and their online communities. *Educational Media International, 39*(1), 31-38.

Levine, S.L., & Wake, W.K. (2000, October 20). Hybrid teaching: Design studios in virtual space. Paper presented at the *National Conference on Liberal Arts and the Education of Artists*, SVA, New York.

Martyn, M. (2003). The hybrid online model: Good practice. *Educause Quarterly, 26*(1), 18-23.

Morgan, K.R. (2002). *Blended learning: A strategic action plan for a new campus*. Seminole, FL: University of Central Florida.

Osguthorpe, R.T., & Graham, C.R. (2003). Blended learning environments: Definitions and directions. *Quarterly Review of Distance Education, 4*(3), 227-233.

Rooney, J.E. (2003). Blending learning opportunities to enhance educational programming and meetings. *Association Managment, 55*(5), 26-32.

Rossett, A. (2002). *The ASTD e-learning handbook*. McGraw-Hill.

Rumble, G. (1992). Open learning. *Open Learning, 7*(2), 31-45.

Schrum, L., & Benson, A. (2000). Online professional education: A case study of an MBA program through its transition to an online model. *Journal of Asynchronous Learning Networks, 4*(1), 52-61.

Singh, H., & Reed, C. (2001). *A white paper: Achieving success with blended learning*: Centra Software.

Thomson, I. (2002). *Thomson job impact study: The next generation of corporate learning*. Thompson, Inc. Retrieved July 7, 2003, from the World Wide Web: *http://www.netg.com/DemosAndDownloads/Downloads/JobImpact.pdf*

Thorne, K. (2003). *Blended learning: How to integrate online & traditional learning*. Sterling, VA: Kogan Page Ltd.

U.S. Department of Education. (2001). *The condition of education 2001* (2001-072). Washington, DC: National Center for Edcuational Statistics.

Voos, R. (2003). Blended Learning - What is it and where might it take us? *Sloan-C View, 2*(1), 3-5.

Waddoups, G., & Howell, S. (2002). Bringing online learning to campus: The hybridization of teaching and learning at Brigham Young University. *International Review of Research in Open and Distance Learning, 2*(2).

Waddoups, G.L., Hatch, G.L., & Butterworth, S. (2003). Blended teaching and learning in a first-year composition course. *Quarterly Review of Distance Education, 4*(3), 271-278.

Willett, H.G. (2002). Not one or the other but both: Hybrid course delivery using WebCT. *The Electronic Library, 20*(5), 413-419.

Young, J.R. (2002, March 22). 'Hybrid' teaching seeks to end the divide between traditional and online instruction. *Chronicle of Higher Education*, pp. A33.

KEY TERMS

Affordances: Features of an environment or artifact that "afford" or permit certain behaviors.

Blended Learning Environment: A learning environment that combines face-to-face and computer-mediated instruction.

Distributed Learning Environment: A learning environment where participants are not co-located and use computer-based technologies to access instruction and communicate with others.

Dual-Mode Institution: Institutions that simultaneously offer on-campus courses and distance education courses.

Hard Technologies: Computer equipment, software, networks, and so forth.

Hybrid Course: Another name for a blended course. Typically, a course that replaces some F2F instructional time with computer-mediated activities.

Hybridization: The phenomenon occurring at universities where there is an increasing overlap of the boundaries of traditional on-campus courses and distance education courses.

Soft Technologies: Instructional innovations, methods, strategies, and so forth.

Best Practices for Effective Virtual Teams

D. Sandy Staples
Queen's University, Canada

Ian K. Wong
Queen's University, Canada

Ann Frances Cameron
Queen's University, Canada

INTRODUCTION

The use of teams as fundamental building blocks in organizations is growing (Furst, Blackburn & Rosen, 1999), as is the frequency of teams to be distributed geographically (which we call virtual teams). Virtual teams are now being used by many organizations to enhance the productivity of their employees and to reach a diversity of skills and resources (Majchrzak, Malhotra, Stamps & Lipnack, 2004). Virtual teams are groups of individuals who work on interdependent tasks, who share responsibility for outcomes, and who work together from different locations. While the use of virtual teams is more common in today's organization, the practices that make virtual teams most effective are not fully understood and challenges remain (Markus, 2004).

Virtual team best practices are identified below from three perspectives: organizational best practices, team leadership best practices, and team member best practices. Ideas for best practices were identified from three sources: six case studies of actual virtual teams (Staples, Wong & Cameron, 2004); the existing literature on virtual teams; and the existing literature on traditional (i.e., collocated) teams and telecommuting (i.e., research on virtual work at the individual level).

ORGANIZATIONAL BEST PRACTICES

There are six best practices that the organizations that have virtual teams should follow. Table 1 contains a list of these practices, each of which is explained next.

Carefully Select Team Members for Diversity

The importance of creating teams with the appropriate diverse mix of skills and individual traits was identified in the case studies and has been identified in the traditional team literature (e.g., Bettenhausen, 1991; Cohen, 1994). The distributed nature of virtual teams allows a diversity of backgrounds, experiences, ideas, thoughts, abilities, and perspectives to be assembled within a single team. Organizations forming virtual teams should take advantage of this, selecting team members with diverse backgrounds and skills. Working on a diverse team can also be more rewarding, interesting, and fun, as team members get the opportunity to learn about new cultures and interact with people beyond their own work location.

Supply Sufficient Resources and Support

Organizations have to supply virtual teams with sufficient resources, including such things as financial resources, time, facilities, hardware, software, communication channels, technical equipment, and proper training. The traditional team literature suggests that team building activities and training members how to work in teams is important because they ensure that employees develop the knowledge required to contribute to organizational performance (Cohen, 1994). In virtual teams where it is especially difficult for team members to get to know one another, organizations may need to provide extra resources for extensive team building exercises. Since virtual teams often need to communicate electronically, appropriate information technology (IT) equipment, training on how to use IT and communication systems, and constant technical support are also important to virtual teams (Duarte & Snyder, 2001; Fisher & Fisher, 2001; O'Hara-Devereaux & Johansen, 1994; Pinsonneault & Boisvert, 2001; Staples et al., 2004).

Develop Human Resource Policies that Stimulate High Virtual Team Performance

Policies must be designed in such a way that virtual team members are recognized, supported, and rewarded for their work (Duarte & Snyder, 2001). Providing team-based

Table 1. Organizational best practices for effective virtual teams

- Carefully select team members for diversity
- Supply sufficient resources and support to the team
- Develop human resource policies that reward team efforts and stimulate virtual team performance
- Provide appropriate autonomy to the team
- Use standard process and procedures
- Develop an organizational culture that stimulates the sharing of information

rewards to team members can increase team cohesiveness, motivation, and effectiveness (e.g., Cohen, Ledford & Spreitzer, 1996; Lawler, 1986, 1992). Since virtual team members are not seen every day in a central office, it is also possible that they may be overlooked for promotional opportunities (Duarte & Snyder, 2001). Therefore, special career development opportunities should be created for virtual team members so that this "out of sight, out of mind" phenomenon does not occur (Pinsonneault & Boisvert, 2001).

Provide Appropriate Autonomy to the Team

Consistent with traditional team research (Cohen & Bailey, 1997), virtual team members interviewed in the case studies reported that little involvement from senior management was usually preferred over hands-on management, as long as the organization still provided the funds and resources necessary. Worker autonomy is shown to have clear benefits such as enhanced worker attitudes and performance. Organizations should give team members the power to take action and make decisions while still providing the team with the information it needs to make sound business decisions (Cohen, 1994). Organizations should provide information on processes, quality, customer feedback, business results, competitor performance, and organizational changes.

Use Standard Processes and Procedures

The use of standard processes and procedures can reduce the time needed for team start-up and may eliminate the need for unnecessary reinvention of operating practices every time a new team is needed (Duarte & Snyder, 2001). For virtual teams that rarely meet face to face, standard communication procedures and policies are extremely important (Duarte & Snyder, 2001; Fisher & Fisher, 2001; Grenier & Metes, 1995). A good face-to-face start-up can allow team members to develop communication norms and agreements on how members are going to work together.

Develop an Organizational Culture that Stimulates the Sharing of Information

The organizational culture plays a large role in determining how well a virtual team functions within the organization since it influences how individuals in an organization behave. Therefore, organizations should work to build norms and values that promote communication and the sharing of information. The traditional team research also identified the importance of having a supportive culture. Therefore, organizations should create a cooperative work environment where norms are established that reinforce and support team behaviors such as sharing information, responding appropriately to team members, and cooperating (Bettenhausen, 1991). This has been found to be critical for effective team performance (Tjosvold, 1988).

TEAM LEADERSHIP BEST PRACTICES

There are six best practices relating to the leadership and management of the virtual team. Table 2 contains a list of all six team leadership practices, each of which is explained next.

Set Goals and Establish Direction

The virtual team literature strongly suggests that effective leaders understand the importance of defining a vision for the virtual team (Fisher & Fisher, 2001; Grenier & Metes, 1995; Lipnack & Stamps, 1997; O'Hara-Devereaux & Johansen, 1994). According to Lipnack and Stamps (1997), a predictor of virtual team success is the clarity of its purpose and vision. To succeed, teams must turn their purpose and vision into action by assigning roles and responsibilities (Fisher & Fisher, 2001; Grenier & Metes, 1995; O'Hara-Devereaux & Johansen, 1994; Pinsonneault & Boisvert, 2001) and setting goals (Lipnack & Stamps, 1997).

In the virtual team case studies, 64% of team members recognized this need to carefully set realistic, clear goals

Table 2. Management and team leader best practices for effective virtual teams

- Set goals and establish direction
- Provide feedback via coaching and modeling
- Build trust through open communication, honest behavior, and delivering on commitments
- Empower the team
- Motivate the team
- Use appropriate leadership styles at appropriate times

and timelines. To accomplish this, management and team leaders can first develop a 'roadmap' with realistic timelines that are compatible with the expectations of senior management. Next, the critical path through the project should be identified. Based on this path, major milestones should be set. Whether or not it affects them directly, all team members should be constantly reminded of the next milestone. Focusing on milestones and deliverable dates will help members keep the 'big picture' in mind when working on their individual tasks. Successful virtual teams are those that, with the help of a focused manager, are consistently able to meet these milestones within the allotted time.

Provide Feedback via Effective Coaching and Modeling

Team leaders need to provide members with timely feedback about their performance so team members know what they can do to continuously improve their performance (Duarte & Snyder, 2001). A manager's ability to provide remote employees with advice and help can increase the effectiveness of the remote employees (Staples, 2001). In virtual teams, this might require getting informal input from various people who interact with team members both within and outside of the organization. Virtual leaders also need to model appropriate virtual work habits (Staples, Hulland & Higgins, 1999). To accomplish this, the telecommuting literature suggests managers should keep remote employees well-informed of organizational activities, provide regular feedback on performance and progress, establish well-structured and constant communications, and be available at hours that fit with work routines of remote employees (Pinsonneault & Boisvert, 2001).

Build Trust Through Open Communication, Honest Behavior, and Delivering on Commitments

Without trust, productivity suffers as team members spend time playing politics instead of working on real business issues (Fisher & Fisher, 2001). To build trust, it is important for team leaders to communicate openly and frequently with team members. Perhaps the single most important variable that affects trust is honesty. Leaders who demonstrate openness about their actions will find that members respond with sincerity. How a leader listens and communicates with his or her team members is very much related to team effectiveness and trust (Cohen & Bailey, 1997). Listening to team members and visibly keeping commitments increases trust, whereas broken promises diminish it (Fisher & Fisher, 2001).

Empower the Team

This best practice is related to the organizational practice of more generally providing autonomy to teams. At a lower level, team leaders also have to provide the appropriate level of autonomy, setting overall team goals and establishing direction while allowing individual team members to decide how they carry out their own specific tasks (Cohen & Bailey, 1997; Fisher & Fisher, 2001). Leaders who trust team decisions can give the members a sense of ownership. This approach is particularly important in a virtual team environment where geographic separation makes micromanagement impractical.

Motivate the Team

In a virtual team environment where tasks may appear unconnected, the 'big picture' is not always easy to visualize and it may be difficult for employees to remain committed to the project. Thus, team leaders can play a key role in keeping virtual team members motivated. Motivation can be stimulated by making the importance of the team's task clear (such that passion for the team's cause is created) and by demonstrating how the project will result in significant outcomes for the individual team members (Fisher & Fisher, 2001; Staples et al., 2004). By linking team success to individual success and opportunities, team members will be highly motivated to succeed on the project.

Use Appropriate Leadership Style

Over a quarter of case study team members (Staples et al., 2004) reported that appropriate leadership at the *appro-*

priate time was one of the key elements of a successful virtual team. During the initial phases of the project, the appropriate leader is one who can "whip up enthusiasm" and motivate the team. During the later stages, the effective leader is someone who is "getting the right people together and keeping everybody on task and keeping everything going." Therefore, the style and activities of team leaders have to be appropriate for the stage the team is at and the needs at that particular time.

TEAM MEMBER BEST PRACTICES

Suggestions for what makes individual members of virtual teams effective include specific behaviors, as well as attitudes and beliefs. The five general characteristics of effective virtual team members are listed in Table 3 and described.

Communicate Effectively

Research in tradition teams and telecommuting recognizes that the ability to communicate effectively is a critical skill for telecommuters, remote workers, and managers of remote workers (Cohen, 1994; Pinsonneault & Boisvert, 2001; Staples, 2001). Eighty-four percent of team members interviewed in the case studies also recognized the importance of effective communication in building a successful team. Communication involves transferring ideas, sharing information, listening and internalizing the ideas of others, and notifying team members of any problems or issues. This can be challenging in a virtual team where face-to-face communication and impromptu meetings are infrequent, if not impossible. To solve this problem, virtual team members suggest working hard to keep lines of communication open, and developing or finding the right communications tools that make up for the loss of face-to-face time and provide for informal interactions. For example, team members can use e-mail or instant messaging as a "virtual coffee pot or water cooler" around which personal conversations can occur (O'Hara-Devereaux & Johansen, 1994). In addition, team members themselves have to be responsive, quickly returning telephone calls and responding to e-mails, even if it is just to say, "I don't have time right now, but I'll get back to you in two days with the answer." Recipients can also confirm that the message had been received and ensure that the major points in the message were understood. Setting communication norms such as these helps to avoid misunderstanding that can occur when communicating electronically.

Have the Necessary Skill Sets

The effectiveness of a team depends on the collective knowledge and skills of its members. In order to make good decisions, team members need the appropriate knowledge and skills (Cohen, 1994). Specific skills mentioned by the virtual team members in the case studies included: the ability to organize effectively, a strong competency in an individual's functional area of responsibility, adequate technical skills to use the information and technology tools available, and good time management skills. Specific social skills mentioned in the virtual team literature include learning how to negotiate creatively, mediating online disputes, and making new members of the team feel included (Grenier & Metes, 1995).

Be Highly Motivated

In some cases, a particular practice may exist on both the management and individual level. As described above, motivating is an activity that must be performed by team leaders. The self-managing nature of many virtual teams means that *self*-motivation and *self*-discipline are also essential. Virtual team members must be able to work independently and be motivated to make appropriate decisions. This is made possible by having clear goals and responsibilities, and having high personal commitment and motivation to the team, along with having the resources and information needed to do the job (Lipnack & Stamps, 1997). The ability to self-motivate is important since virtual team members are often working far apart from their other team members. Virtual workers should also have a low preference or need for social interaction or have the ability to fulfill this need outside of the work team (Pinsonneault & Boisvert, 2001).

Be Supportive of Other Team Members

The way in which team members interact with each other has an effect on team effectiveness (Cohen, 1994). Supporting team members involves working together with a sense of energy and team spirit, and sharing ideas and expertise to help others (Cohen, 1994). Several dimensions of a supportive team emerged during the case study interviews. First, team members felt that it was important to recognize when someone else did a good job, and to congratulate or thank them accordingly. Second, interviewees sought a respectful team environment where members were not afraid to openly discuss ideas. Third, interviewees reported that the ability to get along with other team members was an essential quality of an effective virtual team member.

Table 3. Team member best practices for effective virtual teams

• Communicate effectively
• Have the necessary skill sets
• Be highly motivated
• Be supportive of other team members
• Be action-oriented

Be Action-Oriented

Interviewees felt that individuals should have an action-oriented approach when participating in a virtual team. Phrases that case study members of one team used to describe a top performing virtual team member included: being a "doer"; being "proactive"; "uses an entrepreneurial approach"; and "looks for solutions." One virtual team member who was interviewed stated that successful virtual team members are those who "organize their thoughts into actions or proposals that get a good buy" or influence the rest of the group.

FUTURE TRENDS AND OPPORTUNITIES

As competition increases globally and trading barriers decline, the creation of teams comprised of members from many different locations is expected to become the norm. With this virtual team growth, knowing the best practices for team members, leaders, and organizations with virtual teams becomes even more important. While some useful best practices are explained above, some questions are left unanswered. These questions create many opportunities for future research in this area such as:

- Which of the best practices are most critical for team effectiveness?
- Does the impact of certain practices on effectiveness vary depending on the task and/or the organizational context? How?
- Does one set of practices (i.e., individual, managerial, or organizational) take precedence such that those practices have to be in place before the other practices have a positive effect?
- How does an organization ensure that best practices are followed?
- Can training programs be developed for managers and leaders, and for members of virtual teams? What should be in these training programs and how should they be delivered?
- Can policies be developed and norms established in organizations such that supportive practices, which research suggests leads to effective virtual work, are followed? How can this be done most effectively?

CONCLUSION

The ideas presented in this article should help organizations create and maintain more effective virtual teams. Individual virtual team members, virtual team leaders, and organizations can all follow best practices that contribute to virtual team effectiveness. Given the growing use of virtual teams in organizations today, there is a great need to more deeply understand what makes a virtual team effective. We hope we have made a contribution in that direction, and we look forward to other researchers and practitioners answering some of the questions posed previously.

REFERENCES

Bettenhausen, K.L. (1991). Five years of group research: What we have learned and what needs to be addressed. *Journal of Management, 17*(2), 345-381.

Cohen, S.G. (1994). Designing effective self-managing work teams. In M.M. Beyerlein, D.A. Johnson & S.T. Beyerlein (Eds.), *Advances in interdisciplinary studies of work teams, volume 1, series of self-managed work teams.* Greenwich, CT: JAI Press.

Cohen, S.G., & Bailey, D.E. (1997). What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management, 23*(3), 239-290.

Cohen, S.G., Ledford, G.E., & Spreitzer, G.M. (1996). A predictive model of self-managing work team effectiveness. *Human Relations, 49*(5), 643-676.

Duarte, D.L., & Snyder, N.T. (2001). *Mastering virtual teams: Strategies, tools, and techniques that succeed*. San Francisco: Jossey-Bass.

Fisher, K., & Fisher, M.D. (2001). *The distance manager: A hands-on guide to managing off-site employees and virtual teams*. New York: McGraw-Hill.

Furst, S., Blackburn, R., & Rosen, B. (1999, August). Virtual teams: A proposed research agenda. *Proceedings of the Academy of Management Conference*, Chicago, IL.

Grenier, R., & Metes, M. (1995). *Going virtual*. Upper Saddle River, NJ: Prentice-Hall.

Hofstede, G. (1997). *Cultures and organizations: Software of the mind*. New York: McGraw-Hill.

Lawler, E.E. (1986). *High-involvement management: Participative strategies for improving organizational performance*. San Francisco: Jossey-Bass.

Lawler, E.E. (1992). *The ultimate advantage: Creating the high involvement organization*. San Francisco: Jossey-Bass.

Lipnack, J., & Stamps, J. (1997). *Virtual teams: Reaching across space, time, and organizations with technology*. New York: John Wiley & Sons.

Majchrzak, A., Malhotra, A., Stamps, J., & Lipnack, J. (2004). Can absence make a team grow stronger? *Harvard Business Review, 82*(5), 131.

Markus, M.L. (2004). The reality of virtual teams. *Optimize*, (April), 69.

O'Hara-Devereaux, M., & Johansen, R. (1994). *Global work: Bridging distance, culture & time*. San Francisco: Jossey-Bass.

Pinsonneault, A., & Boisvert, M. (2001). The impacts of telecommuting on organizations and individuals: A review of the literature. In N.J. Johnson (Ed.), *Telecommuting and virtual offices: Issues & opportunities* (pp. 163-185). Hershey, PA: Idea Group Publishing.

Staples, D.S. (2001). Making remote workers effective. In N.J. Johnson (Ed.), *Telecommuting and virtual offices: Issues & opportunities* (ch. 11, pp. 186-212). Hershey, PA: Idea Group Publishing.

Staples, D.S., Hulland, J.S., & Higgins, C.A. (1999). A self-efficacy theory explanation for the management of remote workers in virtual organizations. *Organization Science, 10*(6), 758-776.

Staples, D.S., Wong, I.K., & Cameron, A.F. (2004). Best practices for virtual team effectiveness. In D. Pauleen (Ed.), *Virtual teams: Projects, protocols and processes* (ch. VII, pp. 160-185). Hershey, PA: Idea Group Publishing.

Tjosvold, D. (1988). *Working together to get things done: Managing for organizational productivity*. Lexington, MA: Lexington Books.

KEY TERMS

Communication Norms: In the context of virtual teams, communication norms are typical routines and expectations for communicating within a virtual team using the communication media that the team has available to them (e.g., electronic communication such as e-mail or instant messaging, telephone, etc.). For example, responding within a day to all e-mails, even if just "I'm busy but will get to this tomorrow," is an example of a virtual team communication norm.

Organizational Culture: "The collective programming of the mind which distinguishes the members of one organization from another" (Hofstede, 1997).

Remote Worker: An individual who works at a different location than his/her co-workers and/or manager. That person is remote, in terms of physical presence from his/her colleagues.

Team Diversity: The combined variety of skills, backgrounds, experiences, ideas, thoughts, abilities, and perspectives that individuals bring to their team.

Team Effectiveness: The ability of a team to perform its tasks on time, on budget, and with acceptable quality, as well as the satisfaction, motivation, and commitment of the team members.

Telecommuting: The practice of working from the employee's home instead of physically commuting to a company office location. The connection to the office is done via telecommunications, rather than physically commuting (i.e., traveling) to the office.

Traditional Teams/Collocated Teams: A group of individuals who work on interdependent tasks, who share responsibility for outcomes, and who work together at the same location (i.e., their office/work area is in the same general location).

Virtual Coffee Pot/Virtual Water Cooler: Using electronic communication (such as e-mail or instant messaging) to conduct informal interactions (personal or non-work conversations) that would normally be discussed around the office water cooler in a face-to-face work environment.

Virtual Team: A group of individuals who work on interdependent tasks, who share responsibility for outcomes, and who work together from different locations (i.e., they are geographically dispersed, in terms of their normal work location).

Better Executive Information with the Dashboard Approach

Frédéric Adam
University College Cork, Ireland

Jean-Charles Pomerol
Université Pierre et Marie Curie, France

INTRODUCTION

After more than 30 years of research on how the work of managers can be supported by computers, the observation that developing computer systems that are truly useful for top management is a highly complex and uncertain task is still as valid as ever. Information systems for executives raise specific problems, which have primarily to do with the nature of managerial work itself (Mintzberg, 1973), as they are intended to tackle the needs of users whose most important role is "to create a vision of the future of the company and to lead the company towards it" (King, 1985, p. xi).

BACKGROUND

The major difficulty in supporting managers with computer systems comes from the very nature of management work (Mintzberg, 1973, 1975, 1976), which is concerned with communication, coordination, and people's management for more than 80%. At the time of his research, Mintzberg (1973) had noted how little time is left for reflection and for "playing" with computer systems. This has been a significant difficulty from the origins of MIS systems because their primarily "operational" focus was not central to executives' concerns (Ackoff, 1967; Keen & Scott Morton, 1978). Twenty years later, this difficulty has also been largely responsible for the shift from decision support systems (DSSs) to executive information systems (EISs). EISs were intended to be very easy to use and to help users manipulate required data without the need for much training, which would be very attractive to top executives who want to have, at a glance, a very comprehensive view of their business. Specific descriptions of the differences between DSSs, EISs, and cooperative decision systems can be found in Pomerol and Brézillon (1998). Naturally, computer literacy among executives has increased to a great extent, notably thanks to the development of electronic mail and the World Wide Web. However, whatever designs were put forward over the years, it has remained true that managers are not inclined to spend countless hours browsing computer data, such is the time pressure under which they operate.

Beyond the time pressures under which executives must operate, there are issues of trust and of credibility of the information that can be found in a computer system, which mitigate against intensive executive reliance on information systems, especially in a long-term perspective. First of all, the lack of confidence of executives in their models has been noted by many researchers (e.g., Wallenius, 1975; Cats-Baril & Huber, 1987; Abualsamh, Carlin & McDaniel, 1990). The idea that decision makers need sophisticated models may actually be wrong. People in charge of the preparation of decisions would probably be able to understand and use smart models, but the high-level executives who most commonly make the final decisions are far too busy to train with and use involved systems. On the contrary, they appear to prefer simple systems that they trust and understand, and that display very timely simple information. More often, the data required to make the best decisions will already reside in some form or another in the database of the organization or can be captured with an online feed into a computer system, and what is really needed is a device to filter and display and to warn executives about the most important variances (Simon, 1977). As noted by Kleinmutz (1985): "the ability to select relevant variables seems to be more important than procedural sophistication in the processing of that information" (p. 696).

In EIS, the underlying models built into the system are normally very simple and easily understandable, which is a great help in increasing the acceptability of a computer system.

To conclude, the specificities of managerial decision making can be synthesized as follows:

- Most decisions are made very quickly under considerable time pressure (except some strategic decisions).
- Strategic decision making is often the result of collaborative processes.

- Most decisions are linked to individuals who have specific intentions and commitments to personal principles and ideas.

It is therefore very difficult to support managers, and despite many years of research, little is known about the way information systems could support such unstructured tasks.

A VEHICLE FOR INFORMATION REQUIREMENTS ANALYSIS: CRITICAL SUCCESS FACTORS

In pre-EIS days, Rockart (1979) put forward a methodology called critical success factors or CSF to guide information systems planning. The method had its advantages, though it failed to make a general impact on the planning process of organizations. Its potential in other areas, notably the development of information systems, has been explored by a number of researchers. It is argued in this article that it can be very useful as a guide for the development of executive systems, as both from an information content perspective as for the design of the interface of these systems.

CSF assumes that the performance of organizations can be improved by focusing on "the few key areas where things must go right for the business to flourish" (Rockart, 1979). In simple terms, the method seeks to isolate, using the expertise and gut feeling of managers, the factors which may make the difference between success and failure for the firm.

A number of key points about CSF make it a very attractive technique. First of all, while CSF is essentially a generic framework, it recognizes that all firms are different and operate in different markets. Thus, CSFs are different for different organizations. Secondly, the CSF theory takes into account that the needs of managers within the same organizations are also different based on their hierarchical level, but more importantly, based on their style and their specific areas of responsibility. In general, there are only a limited number of factors that each manager should monitor closely, and this guarantees that managers can concentrate their limited attention to factors that really matter and that are within their control. The attractive thing about this breakdown of responsibility is that the CSF sets controlled by the different managers add up to a complete organizational set that covers all the key areas of the business.

Van Bullen and Rockart (1986) identified a number of primary categories of CSF that are useful in guiding the analysis of the organizational CSF set. These generic sources of CSFs are: (1) the industry where the organization operates (these CSFs are shared by mainstream organizations in this industry), (2) the competitive position and strategy pursued by the organization (which are unique to its set of circumstances and objectives set by its top managers), (3) the environmental factors surrounding the organization (which it has no control over, but which it must monitor closely to compete), (4) temporal factors (which relate to specific events or change programs currently facing the organization, and require the temporary monitoring of additional factors), and finally, (5) CSFs that are specific to each manager and their role in the company. Other authors have added other potential sources such as CSFs related to the analysis of main competitors (especially industry leaders) and the evolution of their business (Leidecker & Bruno, 1984). These sources add up to a wealth of potential factors and measurements that are sufficient for effective monitoring of the business of most organizations.

Dashboards and Control Rooms

In the next stage of the development of executive systems, designers must create an interface for displaying the CSFs. The design of this interface is nearly as important as the selection of the indicators in shaping the perception of managers of the usefulness of their information systems and keeping their interest in the long run. One technique that has worked well in selecting and presenting indicators is the application of the dashboard concept to the management of organizations.

Fundamentally, the concept of dashboard reflects the application of the concept of control room to the management of the firm and echoes the call for a warning or exception reporting functionality in EIS-type systems. In engineering, the control room is a specially designed physical area of a plant where the proper operation of key equipment can be monitored. Control rooms have developed because of the need to monitor increasingly complex processes, such as petrol refining or the operation of nuclear power plants. The control room allows operators to control a process without looking at it with their own eyes, and with a degree of accuracy and completeness that could not be achieved with human perception alone.

This suggests that dashboards may be developed that considerably help managers in their day-to-day search for problems and matching solutions. Naturally, the nature of management itself is highly dynamic and diverse and involves consideration of infinite number of parameters in a way that is fundamentally different from the monitoring of a manufacturing process. Thus, management has a significant "human interaction" component that cannot easily be supported by computer systems. Simon (1977), Gorry and Scott Morton (1971), and others have commented comprehensively on the degree to which managerial decisions are programmable or not, however it remains that the implementation of many of the objectives of the firm, however elusive, can be monitored using a dashboard-type interface. Further, the CSF method can be a powerful vehicle in selecting the indicators to be shown on each manager's dashboard.

The difficulty with CSF-based dashboards resides in the operationalization of managers' key concerns and the identification of specific targets for CSF monitoring, the design of measurement logic for each indicator, and in the development of the interfaces that can be used by managers to easily and effectively review the performance of the firm in relation to each of the indicators.

TOWARDS A METHODOLOGY FOR DASHBOARD DEVELOPMENT

At the height of the EIS movement, King (1985) remarked:

"It is so easy to lose sight of reality—to believe that the computer model's numerical forecasts are real and that they describe future outcomes that will, in fact, come to pass…The computer model's forecasts are based solely on those predictions about the future that we are able to quantify. Those things that are not readily quantifiable are usually omitted, and in being omitted there is a danger that they may be ignored." (p. xi)

This illustrates the dangers inherent in approaching management based solely on numbers, however obtained. This also explains why observational plant tours are still regarded as one of the most reliable methods for collecting data in manufacturing environments (Jones, Saunders & McLeod, 1988). The basis of any dashboard approach to management must therefore take into account the following four key issues:

(1) *Limited Attention:* Given the limited attention of managers and the costs inherent in sourcing certain data, the indicators displayed on the dashboard must be carefully selected using the CSF.

(2) *Performance Measurement:* The measurements used to monitor indicators or CSFs are crucial. The usefulness and effectiveness of the dashboard is totally dependent on the accuracy of the data used and the realism of the calculations presented to managers.

(3) *Operator Training:* It is critical that managers understand the assumptions built into the dashboard and the algorithms used to reach the results presented to them. They must also be fully aware of how data are collected and what limitations applied to the accuracy of the measurements. For instance, drill down facilities can make the difference between "using information to manage more intelligently and more effectively and making the same old mistakes but with more speed" (Meall, 1990).

(4) *Dashboard Layout:* The layout of the dashboard has a direct impact on the understanding derived by managers. The interface of the dashboard must be consistent so that managers can visualize immediately where they should focus their attention as a matter of priority. Exception reporting and color coding (Watson, Rainer & Koh, 1991) can be used to achieve maximum visual impact.

It is also useful if the development of the dashboard can be operationalized as an evolutionary activity where managers can feed back their perception of the dashboards to developers so that the design can be improved over time and the indicators refined or replaced (as in the case of temporal CSFs). In this article, we propose a framework based on 11 questions to support this evolutionary process, and help managers and developers to establish a fruitful dialogue:

- **Question 1: Who will use the indicators?**
  The answer may not be simple when not one, but a number of individuals are interested in monitoring certain indicators. However, managers should concentrate on monitoring the parameters most closely associated with their own performance or that of the areas directly under their control.
- **Question 2: Can be mapped out to a specific objective at a higher level?**
  In the perspective of a top-down CSF exercise, indicators are mapped out to specific objectives pursued by top management. In a bottom-up scenario, it will help if indicators can be merged into higher level composite indicators presented to higher level managers. Developers can use the hierarchy of indicators as a blueprint for the drill-down facility to be built into the dashboard so that top managers can understand the underlying causes of poor or good performance.
- **Question 3: How frequently will managers need to monitor each indicator?**
  Managers' perception of how frequent significant or revelatory variations are likely to occur should be used as a guide for deciding how frequently indicators should be updated. The scope of the benefits that may arise as a result of the monitoring should also be considered if high costs are likely to be incurred.

- **Question 4: What calculation methods are available? What unit of measurement will be used?**
  The choice of calculation method can greatly influence the variation of an indicator and shift the burden of responsibility from one area to another. It can also influence the way the performance of operators or workshops is measured.
  The choice of the unit of measurement is normally straightforward for quantitative analysis, but can become far more complex for less tangible CSFs that involve the estimations of qualitative factors. Customer satisfaction, for instance, will require vision and creativity if it is to be measured properly. Some quantitative measures may be applicable such as the number of complaints received per time interval, but other measures may have to be found that can act as surrogates of customer satisfaction.
- **Question 5: What data source exists? What should be created?**
  Certain data may be missing from existing organizational information systems. Other data may reside in a proprietary system (e.g., a custom-built process control system) that does not integrate well with other systems. Significant investment in equipment and special devices (such as scanners and sensors) or in software such as OLAP and ROLAP (Relational OLAP) may have to be made.
- **Question 6: How detailed should the analysis presented in the dashboard be? How can the indicators be broken down to be more meaningful?**
  Many indicators are too broad to be suitably presented as one figure, and some disaggregating may be required. Typical organizations sell multiple products in multiple markets. Thus, sales figures need to be disaggregated to present £ figures, volumes, and variances for each product on each market while also presenting the aggregated data. Multi-dimensional modeling can be used to support the organization and retrieval of such data.
- **Question 7: What threshold values should be used to differentiate between adequate and inadequate performance? What comparisons can be made to assess the company's performance?**
  Absolute measurement figures presented by a dashboard may not be meaningful to managers unless they can be examined in light of other data. Most companies already have a tight budget system in place, and this can be used as a source of normative values.
- **Question 8: How can each indicator be represented for maximum visual impact?**
  Developers must seek to reduce information overload and use the latest graphical user interface (GUI) technology. Some software tool boxes are now available to help designers create displays and objects that mirror the type of controls normally found on dashboards. Gauges with specific color-coded threshold values can easily be created, and special charts can be made clickable to build intuitive drill down into the data. These speed up and facilitate the data reading of managers.
- **Question 9: What action must be taken when good or bad performance is measured? Is there scope for corrective action to be taken based on the indicator?**
  Whenever good or bad results are presented, managers should know what avenues can be pursued. Reporting mechanisms (e.g., electronic mail) can be built into the dashboard to facilitate and accelerate the dissemination of interesting results and their discussion. In the longer term, increased familiarity with indicators and what their evolution means should have practical decision-making implications for all managers and staff. Thus, users' reaction times to certain signals should be reduced and their responses should improve, especially in recurrent situations.
- **Question 10: How will indicators be monitored/archived in the long term?**
  A key element of our approach is the learning that can be achieved when CSFs are monitored over long periods of time. Staff and managers learn from regularly sampling their performance and that of their areas, and seeing it compared to other data, such as budgets and previous performance of industry standards. Greater learning will be derived if managers and staff set time aside to review and discuss indicators on a regular basis.
- **Question 11: Is there any potential bias inherent in the methods and data used for calculations? What incentives are being given to staff?**
  The development of new performance measurement systems, such as a dashboard of indicators, should always be guided by consideration of the incentives given to actors and the behavior likely to result from the implementation of the underlying indicators. There may also be a change management side to the project, as managers negotiate with staff the implementation of the system. Staff may object to a certain type of measurement (which they may perceive to be threatening or invasive) or the implementation of devices dedicated to monitoring their work.

CONCLUSION

The case for managing the firm solely based on numbers has already been argued and lost. At this point, it is well established that managing firms cannot and should not be compared to the administration of a power plant. This does not mean, however, that the concept of control room does not have potential when applied to the management of organizations. Faced with increasingly complex situations and responsibility for the administration of increasingly complex business processes, managers have less and less time to spend monitoring the key factors of the business. The development of a dashboard can speed up this process and help managers catch far more information than they normally would without assistance.

Following the steps highlighted in this article will also give organizations a much better idea of what parameters they should worry about and how to measure performance. Peter Swasey, one of the directors of the Bank of Boston, commented that "what you don't measure, you don't manage" (McGill, 1990). The preparatory analysis work on the CSFs of the firm will give much confidence to organizational actors that they understand their business and have a comprehensive hold upon its vital functions, and the dashboard ultimately developed will provide flexible and speedy access to vital information, thereby freeing time for other key activities such as business or staff development. As a by-product, managers may also be able to use the analysis carried out for their dashboard as a blueprint for the incentive systems of their company.

REFERENCES

Abualsamh, R., Carlin, B. & McDaniel Jr., R.R. (1990). Problem structuring heuristics in strategic decision making. *Organizational Behavior and Decision Process, 45,* 159-174.

Ackoff, R.L. (1967). Management MISinformation systems. *Management Science, 14*(4), 147-156.

Cats-Baril, W.L. & Huber, G. (1987). Decision support systems for ill-structured problems: An empirical study. *Decision Science, 18,* 350-372.

Gorry, A. & Scott Morton, M. (1971). A framework for management information systems. *Sloan Management Review,* (Fall), 55-70.

Jones, J., Saunders, C. & McLeod, R. (1988). Information media and source patterns across management levels: A pilot study. *Journal of Management Information Systems, 5*(3), 71-84.

Keen, P.G. & Scott Morton, M.S. (1978). Decision support systems: An organizational perspective. Reading, MA: Addison-Wesley.

King, W.R. (1985). Editor's comment: CEOs and their PCs. *Management Information Systems Quarterly, 9,* xi-xii.

Kleinmutz, D.N. (1985). Cognitive heuristics and feedback in a dynamic decision environment. *Management Science, 31*, 680-702.

Leidecker, J. & Bruno, A. (1984). Identifying and using critical success factors. *Long Range Planning, 17*(1), 23-32.

McGill, P. (1990). Executive support systems. *Business Quarterly*¸ (Summer).

Meall, L. (1990). EIS: Sharpening the executives' competitive edge. *Accountancy,* (September).

Mintzberg, H. (1973). *The nature of managerial work*. New York: Harper and Row.

Mintzberg, H. (1975). The manager's job: Folklore and fact. *Harvard Business Review*, (July/August), 49-61.

Mintzberg, H. (1976). Planning on the left side and managing on the right. *Harvard Business Review*, (July/August), 120-130.

Pomerol, J.-Ch. & Brézillon, P. (1998). From DSSs to cooperative systems: Some hard problems still remain. In R. Dolk (Ed.), *Proceedings of HICCS 31* (IEEE Publication, Volume 5, pp. 64-71).

Rockart, J. (1979). Chief executives define their own data needs. *Harvard Business Review, 57*(2), 81-93.

Simon H. (1977). *The new science of management decisions*. Englewood Cliffs, NJ: Prentice-Hall.

Van Bullen, C. & Rockart, J. (1986). A primer on critical success factors. In J. Rockart & C. Van Bullen (Eds.), *The rise of management computing*. Homewood, IL: Dow Jones Irwin.

Wallenius, J. (1975). Comparative evaluation of some interactive approaches to multi-criterion optimization. *Management Science, 21,* 1387-1396.

Watson, H.J, Rainer, K.R. Jr. & Koh, C.E. (1991). Executive information systems: A framework for development and a survey of current practices. *MIS Quarterly*, *15*(1), 13-50.

KEY TERMS

Control Room: A special location in a plant where operators can monitor a process in great detail without having to physically be looking at it. This is particularly useful in dangerous environments.

Critical Success Factors: A methodology for managing projects and firms that concentrates on the areas where things must go right if the endeavor is to flourish.

Dashboard: Specific display of information that presents key information about a process or device. A dashboard may or may not be computerized.

Evolutionary Design: System development methodology where an ongoing approach is taken to analyzing the requirements of the application.

Interface: Portion of a computer application that is used by the user to communicate with the application. It is particularly important for a dashboard, because it may impinge on the ability of users to properly interpret the variations in the indicators shown to them.

Managing by Numbers: A school of thought that sought to demonstrate that firms could be managed solely based on watching key (mostly financial) indicators. This is now largely discredited.

Model: A simplified representation of reality that concentrates on predicting how a factor or a series of related factors would evolve based on the variation of a set of parameters. Also, a simplified representation of reality.

Biblomining for Library Decision-Making

Scott Nicholson
Syracuse University, USA

Jeffrey Stanton
Syracuse University, USA

INTRODUCTION

Most people think of a library as the little brick building in the heart of their community or the big brick building in the center of a campus. These notions greatly oversimplify the world of libraries, however. Most large commercial organizations have dedicated in-house library operations, as do schools, non-governmental organizations, as well as local, state, and federal governments. With the increasing use of the Internet and the World Wide Web, digital libraries have burgeoned, and these serve a huge variety of different user audiences. With this expanded view of libraries, two key insights arise. First, libraries are typically embedded within larger institutions. Corporate libraries serve their corporations, academic libraries serve their universities, and public libraries serve taxpaying communities who elect overseeing representatives. Second, libraries play a pivotal role within their institutions as repositories and providers of information resources. In the provider role, libraries represent in microcosm the intellectual and learning activities of the people who comprise the institution. This fact provides the basis for the strategic importance of library data mining: By ascertaining what users are seeking, bibliomining can reveal insights that have meaning in the context of the library's host institution.

Use of data mining to examine library data might be aptly termed *bibliomining*. With widespread adoption of computerized catalogs and search facilities over the past quarter century, library and information scientists have often used bibliometric methods (e.g., the discovery of patterns in authorship and citation within a field) to explore patterns in bibliographic information. During the same period, various researchers have developed and tested data mining techniques—advanced statistical and visualization methods to locate non-trivial patterns in large data sets. Bibliomining refers to the use of these bibliometric and data mining techniques to explore the enormous quantities of data generated by the typical automated library.

BACKGROUND

Forward-thinking authors in the field of library science began to explore sophisticated uses of library data some years before the concept of data mining became popularized. Nutter (1987) explored library data sources to support decision-making, but lamented that "the ability to collect, organize, and manipulate data far outstrips the ability to interpret and to apply them"(p. 143). Johnston and Weckert (1990) developed a data-driven expert system to help select library materials and Vizine-Goetz, Weibel, and Oskins (1990) developed a system for automated cataloging based on book titles (also see Aluri & Riggs, 1990; Morris, 1991). A special section of *Library Administration and Management* ("Mining your automated system") included articles on extracting data to support system management decisions (Mancini, 1996), extracting frequencies to assist in collection decision-making (Atkins, 1996), and examining transaction logs to support collection management (Peters, 1996).

More recently, Banerjeree (1998) focused on describing how data mining works and ways of using it to provide better access to the collection. Guenther (2000) discussed data sources and bibliomining applications, but focused on the problems with heterogeneous data formats. Doszkocs (2000) discussed the potential for applying neural networks to library data to uncover possible associations between documents, indexing terms, classification codes, and queries. Liddy (2000) combined natural language processing with text mining to discover information in "digital library" collections. Lawrence, Giles, and Bollacker (1999) created a system to retrieve and index citations from works in digital libraries. Gutwin, Paynter, Witten, Nevill-Manning, and Frank (1999) used text mining to support resource discovery.

These projects all shared a common focus on improving and automating two of the core functions of a library—acquisitions and collection management. A few authors have recently begun to address the need to support management by focusing on understanding library users: Schulman (1998) discussed using data mining to examine changing trends in library user behavior; Sallis, Hill, Jance, Lovetter, and Masi (1999) created a neural network that clusters digital library users; and Chau (2000) discussed the application of Web mining to personalize services in electronic reference.

The December 2003 issue of *Information Technology and Libraries* was a special issue dedicated to the bibliomining process. Nicholson (2003) presented an overview of the process, including the importance of creating a data warehouse that protects the privacy of users. Zucca (2003) discussed an implementation of a data warehouse in an academic library. Wormell (2003), Suárez-Balseiro, Iribarren-Maestro, and Casado (2003), and Geyer-Schultz, Neumann, and Thede (2003) used bibliomining in different ways to understand use of academic library sources and to create appropriate library services.

We extend these efforts by taking a more global view of the data generated in libraries and the variety of decisions that those data can inform. Thus, the focus of this work is on describing ways in which library and information managers can use data mining to understand patterns of behavior among library users and staff and patterns of information resource use throughout the institution.

INTEGRATED LIBRARY SYSTEMS AND DATA WAREHOUSES

Most managers who wish to explore bibliomining will need to work with the technical staff of their integrated library system (ILS) vendors to gain access to the databases that underlie that system to create a data warehouse. The cleaning, pre-processing, and anonymizing of the data can absorb a significant amount of time and effort. Only by combining and linking different data sources, however, can managers uncover the hidden patterns that can help to understand library operations and users.

EXPLORATION OF DATA SOURCES

Available library data sources are divided in three groups for this discussion: data from the *creation* of the library, data from the *use of the collection*, and data from *external sources* not normally included in the ILS.

ILS Data Sources from the Creation of the Library System

Bibliographic Information

One source of data is the collection of bibliographic records and searching interfaces that represent materials in the library, commonly known as the Online Public Access Catalog (OPAC). In a digital library environment, the same type of information collected in a bibliographic library record can be collected as metadata. The concepts parallel those in a traditional library: take an agreed-upon standard for describing an object, apply it to every object, and make the resulting data searchable. Therefore, digital libraries use conceptually similar bibliographic data sources as traditional libraries.

Acquisitions Information

Another source of data for bibliomining comes from acquisitions, where items are ordered from suppliers and tracked until received and processed. Because digital libraries do not order physical goods, somewhat different acquisition methods and vendor relationships exist. Nonetheless, in both traditional and digital library environments, acquisition data have untapped potential for understanding, controlling, and forecasting information resource costs.

ILS Data Sources from Usage of the Library System

User Information

In order to verify the identity of users who wish to use library services, libraries maintain user databases. In libraries associated with institutions, the user database is closely aligned with the organizational database. Sophisticated public libraries link user records through zip codes with demographic information in order to learn more about their user population. Digital libraries may or may not have any information about their users, based upon the login procedure required. No matter what data is captured about the patron, it is important to ensure that the identification information about the patron is separated from the demographic information before storing this information in a data warehouse; this will protect the privacy of the individual.

Circulation and Usage Information

The richest sources of information about library user behavior are circulation and usage records. Legal and ethical issues limit the use of circulation data, however. This is where a data warehouse can be useful, in that basic demographic information and details about the circulation could be recorded without infringing upon the privacy of the individual.

Digital library services have a greater difficulty in defining circulation, as viewing a page does not carry the same meaning as checking a book out of the library, although requests to print or save a full text information resource might be similar in meaning. Some electronic full-text services already implement server-side capture of such requests from their user interfaces.

Searching and Navigation Information

The OPAC serves as the primary means of searching for works owned by the library. Additionally, because most OPACs use a Web browser interface, users may also access bibliographic databases, the World Wide Web, and other online resources during the same session; all of this information can be useful in library decision-making. Digital libraries typically capture logs from users searching their databases and can track, through "clickstream" analysis, the elements of Web-based services visited by users. In addition, the combination of a login procedure and cookies allow connecting user demographics to the services and searches they used in a session.

External Data Sources

Reference Desk Interactions

In the typical face-to-face or telephone interaction with a library user, the reference librarian records very little information about the interaction. Digital reference transactions, however, occur through an electronic format, and the transaction text can be captured for later analysis, which provide a much richer record than is available in traditional reference work. The utility of these data can be increased if identifying information about the user can be captured as well, but again, anonymization of these transactions is a significant challenge.

Item Use Information

Fussler and Simon (as cited in Nutter, 1987) estimated that 75-80% of the use of materials in academic libraries is in-house. Some types of materials never circulate, and therefore, tracking in-house use is also vital in discovering patterns of use. This task becomes much easier in a digital library, as Web logs can be analyzed to discover what sources users examined.

Interlibrary Loan and other Outsourcing Services

Many libraries using Interlibrary Loan and/or other outsourcing methods to get items on a "just-in-time" basis for users. The data produced by this class of transactions will vary by service, but can provide a window to areas of need in a library collection.

FUTURE TRENDS

Bibliomining can provide understanding of the individual sources listed earlier; however, much more information can be discovered when sources are combined through common fields in a data warehouse.

Bibliomining to Improve Library Services

Most libraries exist to serve the information needs of users, and therefore, understanding those needs of individuals or groups is crucial to a library's success. For many decades, librarians have suggested works; market basket analysis can provide the same function through usage data to aid users in locating useful works. Bibliomining can also be used to determine areas of deficiency and predict future user needs. Common areas of item requests and unsuccessful searches may point to areas of collection weakness. By looking for patterns in high-use items, librarians can better predict the demand for new items.

Virtual reference desk services can build a database of questions and expert-created answers, which can be used in a number of ways. Data mining could be used to discover patterns for tools that will automatically assign questions to experts based upon past assignments. In addition, by mining the question/answer pairs for patterns, an expert system could be created that can provide users an immediate answer and a pointer to an expert for more information.

Bibliomining for Organizational Decision-Making within the Library

Just as the user behavior is captured within the ILS, the behavior of library staff can also be discovered by connecting various databases to supplement existing performance review methods. While monitoring staff through their performance may be an uncomfortable concept, tighter budgets and demands for justification require thoughtful and careful tracking of performance. In addition, research has shown that incorporating clear, objective measures into performance evaluations can actually improve the fairness and effectiveness of those evaluations (Stanton, 2000).

Low use statistics for a work may indicate a problem in the selection or cataloging process. Looking at the associations between assigned subject headings, call numbers and keywords along with the responsible party for the catalog record may lead to a discovery of system inefficiencies. Vendor selection and price can be examined in a similar fashion to discover if a staff member consistently uses a more expensive vendor when cheaper alternatives are available. Most libraries acquire works both by individual orders and through automated ordering plans that are configured to fit the size and type of that library. While these automated plans do simplify the selection process, if some or many of the works they recommend go unused, then the plan might not be cost effective. Therefore, merging the acquisitions and circulation databases and seeking patterns that predict low use can aid in appropriate selection of vendors and plans.

Bibliomining for External Reporting and Justification

The library may often be able to offer insights to their parent organization or community about their user base through patterns detected with bibliomining. In addition, library managers are often called upon to justify the funding for their library when budgets are tight. Likewise, managers must sometimes defend their policies, particularly when faced with user complaints. Bibliomining can provide the data-based justification to back up the anecdotal evidence usually used for such arguments.

Bibliomining of circulation data can provide a number of insights about the groups who use the library. By clustering the users by materials circulated and tying demographic information into each cluster, the library can develop conceptual “user groups” that provide a model of the important constituencies of the institution’s user base which can fulfill some common organizational needs for understanding where common interests and expertise reside in the user community. This capability may be particularly valuable within large organizations where research and development efforts are dispersed over multiple locations.

In the future, organizations that fund digital libraries can look to text mining to greatly improve access to materials beyond the current cataloging / metadata solutions. The quality and speed of text mining continues to improve. Liddy (2000) has researched the extraction of information from digital texts, and implementing these technologies can allow a digital library to move from suggesting texts that might *contain the answer* to just *providing the answer*, extracted from the appropriate text or texts. The use of such tools risks taking textual material out of context and also provides a few hints about the quality of the material, but if these extractions were links directly into the texts, then context could emerge along with an answer. This could provide a substantial asset to organizations that maintain large bodies of technical texts because it would promote rapid, universal access to previously scattered and/or uncataloged materials.

CONCLUSION

Libraries have gathered data about their collections and users for years, but have not always used those data for better decision-making. By taking a more active approach based on applications of data mining, data visualization, and statistics, these information organizations can get a clearer picture of their information delivery and management needs. At the same time, libraries must continue to protect their users and employees from misuse of personally identifiable data records. Information discovered through the application of bibliomining techniques gives the library the potential to save money, provide more appropriate programs, meet more of the user’s information needs, become aware of gaps and strengths of their collection, and serve as a more effective information source for its users. Bibliomining can provide the data-based justifications for the difficult decisions and funding requests library managers must make.

References

Atkins, S. (1996). Mining automated systems for collection management. *Library Administration & Management*, *10*(1), 16-19.

Chau, M.Y. (2000). *Mediating off-site electronic reference services: Human-computer interactions between libraries and Web mining technology*. Fourth International Conference on Knowledge-based Intelligent Engineering Systems & Allied Technologies (vol. 2, pp.695-699). Piscataway, NJ: IEEE.

Chaudhry, A.S. (1993). Automation systems as tools of use studies and management information. *IFLA Journal*, *19*(4), 397-409.

Doszkocs, T.E. (2000). Neural networks in libraries: The potential of a new information technology. Retrieved October 24, 2001, from *http://web.simmons.edu/~chen/nit/NIT%2791/027~dos.htm*

Geyer-Schulz, A., Neumann, A., & Thede, A. (2003). An architecture for behavior-based library recommender systems. *Information Technology and Libraries*, *22*(4), 165-174.

Guenther, K. (2000). Applying data mining principles to library data collection. *Computers in Libraries, 20*(4), 60-63.

Gutwin, C., Paynter, G., Witten, I., Nevill-Manning, C., & Frank, E. (1999). Improving browsing in digital libraries with keyphrase indexes. *Decision Support Systems, 2I*, 81-104.

Johnston, M., & Weckert, J. (1990). Selection advisor: An expert system for collection development. *Information Technology and Libraries*, *9*(3), 219-225.

Lawrence, S., Giles, C.L., & Bollacker, K. (1999). Digital libraries and autonomous citation indexing. *IEEE Computer, 32*(6), 67-71.

Liddy, L. (2000, November/December). Text mining. *Bulletin of the American Society for Information Science*, 13-14.

Mancini, D.D. (1996). Mining your automated system for systemwide decision making. *Library Administration & Management*, *10*(1), 11-15.

Morris, A. (Ed.) (1991). *Application of expert systems in library and information centers*. London: Bowker-Saur.

Nicholson, S. (2003). The bibliomining process: Data warehousing and data mining for library decision-making. *Information Technology and Libraries*, *22*(4), 146-151.

Nutter, S.K. (1987). Online systems and the management of collections: Use and implications. *Advances in Library Automation Networking, 1*, 125-149.

Peters, T. (1996). Using transaction log analysis for library management information. *Library Administration & Management*, *10*(1), 20-25.

Sallis, P., Hill, L., Jance, G., Lovetter, K., & Masi, C. (1999). A methodology for profiling users of large interactive systems incorporating neural network data mining techniques. *Proceedings of the 1999 Information Resources Management Association International Conference* (pp. 994-998). Hershey, PA: Idea Group Publishing.

Schulman, S. (1998). Data mining: Life after report generators. *Information Today*, *15*(3), 52.

Stanton, J.M. (2000). Reactions to employee performance monitoring: Framework, review, and research directions. *Human Performance, 13*, 85-113.

Suárez-Balseiro, C.A., Iribarren-Maestro, I., & Casado, E.S. (2003). A study of the use of the Carlos III University of Madrid library's online database service in scientific endeavor. Information Technology and Libraries, 22(4), 179-182.

Wormell, I. (2003). Matching subject portals with the research environment. *Information Technology and Libraries*, *22*(4), 158-166.

Zucca, J. (2003). Traces in the clickstream: Early work on a management information repository at the University of Pennsylvania. *Information Technology and Libraries*, *22*(4), 175-178.

KEY TERMS

Bibliometrics: The study of regularities in citations, authorship, subjects, and other extractable facets from scientific communication using quantitative and visualization techniques. This allows researchers to understand patterns in the creation and documented use of scholarly publishing.

Bibliomining: The application of statistical and pattern-recognition tools to large amounts of data associated with library systems in order to aid decision-making or justify services. The term "bibliomining" comes from the combination of bibliometrics and data mining, which are the two main toolsets used for analysis.

Data Mining: The extraction of non-trivial and actionable patterns from large amounts of data using statistical and artificial intelligence techniques. Directed data mining starts with a question or area of interest, and patterns are sought that answer those needs. Undirected data mining is the use of these tools to explore a dataset for patterns without a guiding research question.

Data Warehousing: The gathering and cleaning of data from disparate sources into a single database, optimized for exploration and reporting. The data warehouse holds a cleaned version of the data from operational systems, and data mining requires the type of cleaned data that lives in a data warehouse.

Integrated Library System: The automation system for libraries, combining modules for cataloging, acquisition, circulation, end-user searching, database access, and other library functions through a common set of interfaces and databases.

B

Online Public Access Catalog (OPAC): The module of the integrated library system designed for use by the public to allow discovery of the library's holdings through the searching of bibliographic surrogates. As libraries acquire more digital materials, they are linking those materials to the OPAC entries.

ENDNOTE

[1] This work is adapted from: Nicholson, S. & Stanton, J. (2003). Gaining strategic advantage through bibliomining: Data mining for management decisions in corporate, special, digital, and traditional libraries. In H. Nemati, & C. Barko (Eds.), *Organizational data mining: Leveraging enterprise data resources for optimal performance (pp.247-262)*. Hershey, PA: Idea Group Publishing.

Bridging the Digital Divide in Scotland

Anna Malina
e-Society Research, UK

INTRODUCTION

Perceptions of the different meanings and issues surrounding the term *digital divide* have set the scene for policy development in various countries. In recent times, broader analysis of the meanings and problems have altered understanding, and a new range of initiatives to tackle perceived problems is being devised in the United Kingdom (UK) and its regions. In what follows, digital divide perspectives are outlined and action to close the divide in Scotland is discussed.

BACKGROUND

For some time now, the Information Society vision in many countries has been accompanied by knowledge of risk of exclusion and strategies to close the so-called "Digital Divide," often seen as a short-hand term to indicate significant inequalities in access across social groups, and in particular between those who have access to ICTs (i.e., the "haves" and the "have-nots" or those who do not have access) (Civille, 1995; Raab, 1996). EU directives (e.g., eEurope 2002 [2000] and eEurope 2005 [2002]) support the goal of cheaper and more widespread access. The 2002 goal is to achieve physical access to the Internet, and the next stage is to consider content to stimulate access. The hope is that the benefits that emerge as a result of access to modern information and communication technologies (ICTs) will be felt by regional and local economies and communities.

Extending access to ICTs will help ensure innovation, economic development, and the new economy (Tambini, 2000a). In discussing universal Internet access, Tambini (2000b) points to arguments that without widespread access, e-commerce would not be able to support greater innovation and entrepreneurship. In addition, the intense concentration on developing e-government could not be legitimised. Moreover, benefits said to be associated with the design of ICTs to improve efficiency, effectiveness, and transparency in public service delivery could not be realised. Norris (2001) and others note the views of pessimists who fear an escalation of existing inequalities, and optimists who hold that new ICTs have the potential to widen opportunities for more democratic participation. However, without universal Internet access, it is unlikely that wider forms of electronic participation and actions associated with e-governance and e-democracy could be supported. Additionally, distance learning and public education resources would not reach wider audiences or increase literacy levels.

BRIDGING THE DIGITAL DIVIDE IN SCOTLAND

The Scottish Household Survey shows that access to the Internet in Scotland is growing quickly. People who are excluded comprise the unemployed, those with low incomes, low levels of education, and poor literacy and numeracy levels. The Scottish Executive, the Scottish Parliament, the voluntary sector, and other organizations in Scotland have designed a range of initiatives to tackle problems associated with the digital divide. The Scottish framework is based on raising awareness, widening access, increasing skills, building support, developing content, and motivating and involving communities.

Scottish Executive and Scottish Parliament

Similar to the UK Government, the Scottish Executive (i.e., the devolved government of Scotland) is committed to achieving universal access to the Internet by 2005. The Scottish Executive initiative—Digital Scotland—set out to ensure that Scotland obtains and retains maximum economic and social advantage from the development of ICTs. Digital divide problems are associated with a lack of telecommunications infrastructure and with poverty, lack of awareness, and low skill levels (Digital Inclusion: Connecting Scotland's People, 2001). Emphasis has been placed on expanding Scotland's communication infrastructure to stimulate demand for broadband and to test innovative delivery technologies. A three-year Digital Champions programme[1] was set up to improve inclusive ICT provision in Social Inclusion Partnerships (SIPs) in deprived areas of Scotland. This project provides community professionals with local knowledge to engage local people in various initiatives in order to drive local ICT activities forward.

In 2002, a £3.2 million initiative was launched by the Social Justice Minister to create a network of 1,000 new Internet access points in areas where current public provision was still poor.[2] The Scottish Executive also promotes awareness of existing public access to the Web, and provides an online service to help people find the nearest access point.[3]

The Scottish Parliament's Web site[4] provides extensive information online. To help address problems of the digital divide, the Parliament worked with the public library sector to establish a network of 80 partner libraries throughout Scotland, many of which now provide public access to the Internet through freely available terminals.

With a key focus on the citizen, government portals across the UK are offering services that they suggest are relevant to life episodes. Closely linked to the UK Government Portal (www.ukonline.gov.uk) is the Scottish Government Portal that operates under the brand *Open Scotland*[5] to promote choice and public take-up of services.

Voluntary Sector Initiatives

In August 1999, com.com/holyrood, a public-private partnership between the Scottish Centre for Voluntary Organisations (SCVO) and British Telecom (BT), was given the task of installing 200 PCs into local halls and community centres throughout Scotland. However, access alone was not considered enough to address the digital divide, and SCVO also began to develop voluntary sector content. In June 2002, a Web portal[6] was launched to act as a single gateway to Scotland's voluntary sector. In addition, a lobby channel allowed voluntary organisations to conduct their own e-consultations. Moreover, online questions could be forwarded to Members of the Scottish Parliament (MSP).

Education and Learning Initiatives

The National Grid for Learning (NGfL) Scotland was set up by the Scottish Executive Education Department in September 1999 to connect all schools, colleges, universities, and libraries in Scotland to the Internet by 2002. A key objective of the NGfL Scotland Communities team[7] is to use ICT to improve opportunity, access, and quality of life for excluded groups, and to actively involve communities in worthwhile local projects. The communities channel of NGfL Scotland aims to supply information, advice, and assistance to all those providing support for ICTs in their community. In addition, NGfL Scotland's Connecting Communities Training Programme[8] promotes the effective use of ICT in community learning agencies across Scotland.

The Scottish University for Industry (SUfI) was set up to promote public/private partnership, commission research, draw on other analyses, and investigate the needs of market and client groups in Scotland. SUfI built on partnerships already existing in Scotland and worked closely with Highlands and Islands Enterprise and other partners to develop skills in using ICT. The subsequent development of a variety of IT centres in different locations of Scotland has provided Internet access and a learning environment for people to meet, to learn about new ICT, and to achieve new skills.

Scottish radio and television broadcasts were organised in late 1999 to promote learning directly and supporting the BBC's Webwise campaign, building on an earlier programme entitled Computers Don't Bite. Drawing from £200 million pounds allocated to support Community Access to Lifelong Learning (CALL) across the UK, the new Opportunities Fund (NOF) in Scotland was allocated £23 million pounds to support LearnDirect, an organisation providing public information about local learning centres.

Scotland is aiming to close the digital divide and to encourage people in deprived areas to acquire the key IT skills that are suitable to the demands of an Information Society.

Wired Communities in Scotland

During 2001, in an attempt to promote digital inclusion, a three million pound initiative outlined the intention to create two pilot digital communities in Scotland. The document, titled Digital Inclusion: Connecting Scotland's People (*Scottish Executive*, 2001, p. 22) outlines intention and funding for:

- the provision of entry level PCs, software, and Web access to up to 2,000 homes in each community;
- the development of links with school-based ICT and Web access initiative;
- the development of a community Web portal for each community with local content relevant to that community, including relevant online public and commercial services;
- the provision of training to increase the level of ICT and Web skills;
- the promotion to raise awareness of the benefits of the Web; and
- the creation of a network of local people to provide ongoing support.

The same report also outlines the aims of the digital communities Scotland project as follows:

- to use ICTs to help tackle social exclusion in these communities;

- to create a "critical mass" of Web users in each community;
- to contribute to achieving universal access to the Web;
- to increase the take-up of computers and the Web in disadvantaged households;
- to increase ICT skills in disadvantaged communities;
- to increase community involvement to develop local online content and a local support network; and
- to create partnerships with the private sector.

In late 2001, communities across Scotland were invited to submit bids to the Scottish Executive to receive funding and support to become digital communities. Towards the end of March 2002, the Scottish Executive selected two winning communities—one rural and one urban.

The Argyll Islands, the rural community selected, covers 13 of the 26 inhabited islands in the archipelago. A total of 2,145 households from a population of 4,290 was included. The digital communities submission for funding suggests that while the islands "are rich in biodiversity, culture, history, archaeology, renewable energy opportunities, and landscape, they suffer from many issues of deprivation and disadvantage, both caused and accentuated by their geographical isolation" . Physical and demographic characteristics of the islands combine to produce economic and social conditions constraining economic activity and growth. As such, the islands of Argyll are categorised as "fragile areas" by Highlands and Islands Enterprise and as "disadvantaged areas" by the Council's Economic Development Strategy (Argyll and Bute Council First Round Digital Communities Submission, 2002).

All local people in the Argyll Islands are expected to profit from new opportunities to use computers and access the Internet. The aim is to make it easier to communicate speedily and cheaply in regard to public services, to interact with peers, to telework, and to access health services and lifelong learning materials. Wherever possible, the intention is to include the business community of these islands as well as community groups. While some members of the community already have computer skills, the initiative is expected to extend training and support to any of those households that currently have computers but lack sufficient knowledge and training to use them to full advantage. In addition, the intention is to build e-services into a comprehensive community portal.

Bellsmyre, situated to the north of the town of Dumbarton, is the urban digital community selected for funding. Bellsmyre has 789 households and a population of 1,694 people. No telecommunications company had cabled the community, and at the time of the bid, Bellsmyre only had an analogue telephone infrastructure. The area is described in the First Round Proposal document as the most deprived within West Dunbartonshire.

At the time, people living in Bellsmyre had not yet responded to previous encouragement to train and use technology and, as a result, were considered in danger of being excluded from society even further. Technology is being taken to the people to promote the advantages of learning. Drawing from partnership arrangements, the key aim outlined in Bellsmyre's First Round Proposal is to help achieve the objectives set out in Scotland's digital inclusion strategy. It is envisaged that delivery of training and ICT services by Digital Champions and local community groups and agencies will help develop trust. It is hoped that attitudes toward ICT will improve, fear of technology will subside, confidence will improve, and the benefits of being connected to the Web will emerge. It is also hoped that attitudes toward learning will change and become more positive. A central aim is to ensure that levels of educational achievement will rise locally. Key aspirations are to reduce digital exclusion, create an ICT-skilled local population that will develop a routine taste for further learning and training and go on to employment or start up new businesses.

E-Democracy and the Digital Divide in Scotland

The UK government has established a Cabinet Committee to oversee e-democracy in the UK, and to develop a strategy for its development and rollout. An e-democracy charter divides the e-democracy domain into two clear fields: e-voting and e-participation. It is possible to design technology specifically to support local authorities by taking a creative, flexible, and democratic approach at local levels (Sisk, 2001). However, the ways in which technology might be developed and implemented to support better democratic participation in local areas have not been researched previously in Scotland. An added difficulty is that because each local authority is charged with representing multiple communities with diverse needs, the local authorities may need to use different e-democracy processes in different ways in different communities. This highlights a gap in knowledge about the numerous ways e-democracy systems could be tailored to suit the democratic needs of dissimilar communities. In addition, there is a matter of ensuring authentic participation of citizens in ways in which they are willing to be involved and that reflect their views and experiences about issues they feel are relevant in their own communities and beyond.

FUTURE TRENDS

A New Zealand report, "The Digital Divide – Examining the Main Characteristics that Influence Household Internet Connection in New Zealand" (2004), reflects international research in suggesting that the expansion of information and communication technologies is mainly utilised by households with higher incomes and by households whose members have formal educational qualifications. The report also found that "although the age of the youngest occupant, ethnicity, labour force status and geographic location played important roles in determining household Internet access, the most important variables identified as influencing household connectivity levels were household income, the level of educational qualification and household composition" (Chapter 3 Conclusion).

Research conducted by the Greater London Authority to outline a strategy to address the digital divide in London shows that while socioeconomic factors such as low income, low levels of education, low-skilled jobs, unemployment, and lack of technology skills are a barrier to the adoption and use of ICTs, sociopersonal aspects such as low levels of awareness, interest, understanding, and acceptance of ICTs are also very important (Foley et al., 2002).

A great deal of attention and policy and resource support over recent years has helped to ensure that excluded groups and communities living outside mainstream society are provided with different kinds of access to contemporary ICTs. However, Wilhelm (2000) questions whether or not the information under class can now be wholly defined in terms of access. Instead, he emphasizes broader contexts of information-seeking behaviour (i.e., media use patterns and cultural and environmental contexts). Drawing from Schon, Sanyal, and Mitchell (1999), he argues that these contexts will "provide a thicker description of the various shades of information and communications inequalities" (p. 70). Hague and Loader (1999) argue that experiences of using ICTs will vary enormously, since people do not generally share the same traits and since members of the public are very different in terms of gender, race, disability, class, location, and religion.

Others suggest that digital divide arguments should move away from issues of access alone. For example, Warschauer (2003) argues that increased emphasis should be put on action to ensure that people develop more meaningful use of technology within their own social arenas and daily practices.

CONCLUSION

Consideration of the Scottish context suggests that commitment to develop locally based ICT infrastructure and innovative new initiatives is fundamental in addressing problems associated with the digital divide. Critical evaluations of these initiatives would be invaluable, and it is crucial now for long-term research programmes to broaden understanding of meaningful routine use. Some researchers underline the importance of gathering empirical data relating to how and why people use ICTs in different settings (Malina & MacIntosh, 2003). Action research is suggested to assess the democratic requirements of those living in different kinds of communities, and to enable more meaningful design, development, and continuous assessment of ICT-based systems underpinning new democratic practices. It is suggested that this research approach could be applied initially to the two "digital communities" in Scotland—Argyle & Bute and Bellsmyre. In seeking to consult citizens, we must listen to their democratic needs and consider their perceptions. Research would help to improve the quality of citizenship and the level of democratic participation in local communities. The research work would also provide a framework to better appreciate the significance of technology in supporting e-democracy at local community levels and, in so doing, to contribute knowledge to strategy and planning policies and social and digital inclusion agendas to address problems of the digital divide in Scotland.

REFERENCES

Argyll and Bute Council first round digital communities submission (2002). Argyll & Bute Council.

Bellsmyre, West Dunbartonshire first round digital communities proposal (2002). West Dunbartonshire Council.

Civille, R. (1995). The Internet and the poor. In B. Kahlin and J. Keller (Eds.), *Public access to the Internet*. Cambridge MA: MIT Press.

The digital divide – Examining the main characteristics that influence household Internet connection in New Zealand (2004, March 5). Statistics New Zealand (Te Tari Tatau). Retrieved from *http://www.stats.govt.nz/domino/external/pasfull/pasfull.nsf/web/Reference+Reports+The+Digital+Divide+2004?open*

eEurope2002 (2000). *http://www.europa.eu.int/information_society/eeurope/action_plan/pdf/actionplan_en.pdf* eEurope2005: An information society for all (2002, June 28). Retrieved from *http://europa.eu.int/information_society/eeurope/news_library/eeurope2005/index_en.htm*

Foley, P., Alfonso, X., & Ghani, S. (2002). *The digital divide in a world city*. London: Greater London Authority.

Hague, B., & Loader, B. (Eds.) (1999). *Digital democracy: Discourse and decision making in the information age*. London: Routledge.

Malina, A., & Macintosh, A. (2003). Bridging the digital divide. Development in Scotland. In A. Anttiroiko, M. Mälkiä , & R. Savolainen (Eds.), *eTransformation in governance: New directions in government and politics*. Hershey, PA: Idea Group Publishing.

Norris, P. (2001). *Digital divide: Civic engagement information poverty, and the Internet worldwide*. Cambridge, MA: Cambridge University Press.

Raab, C., Bellamy, C., Taylor, J., Dutton, W.H., & Peltu, M. (1996). The information polity: Electronic democracy, privacy, and surveillance. In W. Dutton (Ed.), *Information and communication technologies: Vision and realities*. Oxford: Oxford University Press.

Schon, D., Sanjal, B., & Mitchell, W.J. (Eds.) (1999). *High technology and low-income communities: Prospects for the positive use of advanced information technology*. Cambridge, MA: MIT Press.

Scottish Executive (2001). Digital inclusion: Connecting Scotland's people. Retrieved from *http://www.scotland.gov.uk/library3/enterprise/dics-00.asp*

Scottish Executive (2002). Scottish household survey bulletin no. 7: Life cycles. Retrieved from *www.scotland.gov.uk/library3/housing/shs7-09.asp*

Sisk, T. (2001). *Democracy at the local level: The international IDEA handbook on participation, representation, conflict management, and governance*. Hershey, PA: Idea Group Publishing.

Tambini, D. (2000a). *Digital danger*. London: IPPR.

Tambini, D. (2000b). *Universal Internet access: A realistic view*. London: IPPR.

Warschauer, M. (2003). *Technology & social inclusion*. Cambridge, MA: MIT Press.

Wilhem, Anthony (2000). *Democracy in the digital age: Challenges to political life in cyberspace*. New York and London: Routledge.

KEY TERMS

Digital Divide: Refers to individuals or members of communities and groups whose social, cultural, political, economic, or personal circumstances constrain access to electronic communications or limit benefit to their lives from contemporary electronic technologies.

Digital Inclusion: Strategies and actions to assure more equal access to digital technologies and Web facilities and to strengthen effective, meaningful, and beneficial use for all members of the public in their day-to-day lives.

Distance Learning: Learners are connected with educational resources beyond the confines of a traditional classroom, and instructed via computer-mediated communication and different types of electronic technologies that can overcome the constraints of distance, time, physical presence, or location that separate instructors and students. Learning may be synchronous or asynchronous.

e-Commerce: The buying and selling of commercial goods; the conduct of financial transactions using digital communications and electronic networks such as the World Wide Web; and aspects of sharing of business information, maintenance of business relationships, and provision of information services.

e-Democracy: The use of electronic communications to support and increase democratic engagement and deepen and widen citizen participation.

e-Government : The ability of government to design and use ICTs to interact internally and externally with government bodies, citizens, and businesses in order to deliver integrated electronic public services.

e-Governance: Communication by electronic means to place power in the hands of citizens to determine what laws need to be made and how these laws should be written.

Information and Communication Technologies (ICTs): While often meaning different things in different timescales, places, and contexts, ICTs describe all media and a mix of converging technology tools involved in the dynamic transfer and storage of analogue and digital data. In addition to Internet-based technologies such as computers, telephones, and networks, ICTs in a broad sense include digital television, cable and satellite technologies, and music formats (e.g., MP3), DVDs, and CDs. ICTs may be used to facilitate remote human interaction for good and evil purposes. In the context of this article, ICTs are used to increase human communication; broaden education, literacy, and knowledge; and enhance social, cultural, political, and economic capacity. It is hoped that this will help address problems attributed to the so-called digital divide.

ENDNOTES

[1] http://www.scotland.gov.uk/library3/enterprise/dics-06.asp, access November 2002.
[2] http://www.scotland.gov.uk/pages/news/extras/00007100.aspx, accessed November 2002.
[3] http://www.scotland.gov.uk/digitalscotland/webaccess/default.asp, accessed November 2002.
[4] www.scottish.parliament.uk, accessed November 2002.
[5] http://www.openscotland.gov.uk/, accessed November 2003.
[6] www.workwithus.org, retrieved November 2002.
[7] http://www.ngflscotland.com/communities/, retrieved May 2002.
[8] http://www.ngflscotland.com/communities/training/connectingcommunities/, retrieved May 2002.

B

Bridging the Growing Digital Divide

Ioannis Tarnanas
Kozani University of Applied Science, Greece

Vassilios Kikis
Kozani University of Applied Science, Greece

INTRODUCTION

That portion of the Internet known as the World Wide Web has been riding an exponential growth curve since 1994 (Network Wizards, 1999; Rutkowski, 1998), coinciding with the introduction of NCSA's graphically-based software interface Mosaic for "browsing" the World Wide Web (Hoffman, Novak, & Chatterjee, 1995). Currently, over 43 million hosts are connected to the Internet worldwide (Network Wizards, 1999). In terms of individual users, somewhere between 40 to 80 million adults (eStats, 1999) inthe United States alone have access to around 800 million unique pages of content (Lawrence & Giles, 1999), globally distributed on arguably one of the most important communication innovations in history.

Yet even as the Internet races ambitiously toward critical mass, some social scientists have begun to examine carefully the policy implications of *current* demographic patterns of Internet access and usage (Hoffman & Novak, 1998; Hoffman, Kalsbeek, & Novak, 1996; Hoffman, Novak, & Venkatesh, 1997; Katz & Aspden, 1997; Wilhelm, 1998). Looming large is the concern that the Internet may not scale *economically* (Keller, 1996), leading to what Lloyd Morrisett, the former president of the Markle Foundation, has called a "digital divide" between the information "haves" and "have-nots." For example, although almost 70% of the schools in this country have at least one computer connected to the Internet, less than 15% of classrooms have Internet access (Harmo, 1997). Not surprisingly, access is not distributed randomly, but correlated strongly with income and education (Coley, Cradler, & Engel, 1997). A recent study of Internet use among college freshman (Sax, Astin, Korn, & Mahoney, 1998) found that nearly 83% of all new college students report using the Internet for school work, and almost two-thirds use email to communicate. Yet, closer examination suggests a disturbing disparity in access. While 90.2% of private college freshman use the Internet for research, only 77.6% of students entering public black colleges report doing so. Similarly, although 80.1% of private college freshman use email regularly, only 41.4% of students attending black public colleges do.

Further, although numerous studies (CyberAtlas, 1999; Maraganore & Morrisette, 1998) suggest that the gender gap in Internet use appears to be closing over time and that Internet users are increasingly coming from the ranks of those with lower education and income (Pew Research Center, 1998), the perception persists that the gap for race is not decreasing (Abrams, 1997).

We now raise a series of points for further discussion. We believe these issues represent the most pressing unanswered questions concerning access and the impact of the digital divide on the emerging digital economy. This chapter is intended to stimulate discussion among scholars and policy makers interested in how differences in Internet access and use among different segments in our society affects their ability to participate and reap the rewards of that participation in the emerging digital economy. In summary, we have reviewed the most recent research investigating the relationship of race to Internet access and usage over time. Our objective is twofold: 1) to stimulate an informed discussion among scholars and policy makers interested in the issue of diversity on the Internet, and 2) propose a research agenda that can address the many questions raised by this and related research.

BACKGROUND AND PERSPECTIVE

Laugksch (1999) pointed out that scientific literacy has become an internationally well-recognized educational slogan, buzzword, catchphrase, and contemporary educational goal. The same applies to the case of digital divide. Courtright and Robbin (2001) contend that "the metaphor of the digital divide" — has become part of the national discourse of the United States, an abstract symbol that condenses public concerns about social inequality and evokes hopes for solutions related to the use of information technology. In addition, "the digital divide is a potent resource whose symbolic properties and communicative power have activated a wide array of participants in the policy debates about how to create a more just society."

According to Hoffman (2001, cf.: Arquette, 2001), the term *digital divide* was first used by Lloyd Morrisett who vaguely conceived of a divide between the information-haves and have-nots. However, the divide herein mainly is a gap of PC penetration in the early days of the Apple II in 1980 (Arquette, 2001). The term then grasped public's attention with the issuance of the first National Telecommunications and Information Administration (NTIA) survey on Internet adoption and use in the US in 1994 with the catchy title: *Falling Through the Net.* Since then, numerous articles, either popular or academic, on this issue have been published. According to a convenient sample of newspapers, journal articles, newswires and similar mass media sources in the Lexis-Nexis database from January 1999 to December 2000 (Arquette, 2001), the increasing rate of digital divide related articles hits almost 3000%.

In developing countries, digital divide is receiving similar social saliency. A quick search with the key words "digital divide" in one of Greece leading news Web site *Daily Online* (www.in.gr) shows that at least 500 articles somehow related to this term are available. On July 2001, a high-level forum on public understanding of information technology with the special topic of *Pay Attention to the Digital Divide* was held in Greece. A wide range of representatives, including governmental officials, IT experts, educators, social scientists and media practitioners, presented their viewpoints and comments on this issue. Digital divide has been incorporated into daily conversational discourse.

Ironically, while the term *digital divide* has frequently appeared in varied contexts, including academic writings, both the connotative and denotative meanings of it are confusingly incoherent. The presence of other similarly prevalent terminologies, such as digital equality, information equality, e-development, network readiness, etc., adds additional confusion. People seem to debate on the issue without a shared understanding of what is meant by the digital divide. As Arquette (2001) contends, the entire researcher community is plagued by a lack of definitional clarity of the concepts such as digital divide: "each researcher assumes other researchers use the same definitional frameworks for these terms while in fact there is no such shared meaning in nomenclature" (p. 3).

While the comment of Arquette (2001) mainly refers to the phenomenon in the English speaking world, the use of its minority counterpart of the term *digital divide* is also in a similar situation. For example, among more than 30 articles collected by the book *Pay Attention to the Digital Divide in Developing Countries* (Leng, 2002), no consistent conceptual definition is available across the writings. While some are talking about the Internet penetration divide among different social groups categorized by age, occupation and educational level, others refer the concept to an uneven development of e-infrastructure among different areas or nations. So, whenever the term digital divide is confronted, the following question can always be raised: *in terms of what*?

This chapter intends to introduce a new approach of operationalizing digital divide from the perspective of Developing Countries. We first make a brief review of different definitional perspectives of the term *digital divide*. Then a detailed introduction of National Informatization Quotient is presented which will be employed as the operational definition of the informatization level of a region. Finally we will investigate the geographical digital divide in Developing Countries in terms of NIQ.

CONCEPTUAL REVIEW

Conceptual definition involves verbal descriptions of the essential properties that are to be included in the intended meaning of a concept. In research practice, it often involves specifying the essential dimensions of a concept (McLord & Pan, 2002, p. 62). On the other hand, o*perational definition* involves procedures by which a concept is to be observed, measured, or manipulated. It details the rules, specific steps, equipment, instruments, and scales involved in measuring a concept (p. 65). In this section, we will briefly review the multiple conceptions around digital divide.

Digital divide is a fresh term not unfamiliar to communication scholars (Zhu, 2002). As early as 1970, a theory called *knowledge gap* (Tichenor, Donohue & Olien, 1970) was developed which has been one of the most active inquiry fields hereafter in communication studies. The supposition of knowledge gap mainly concerns the different knowledge possession through mass media by social groups with varied social-economic-status. In 1980s, with the development of ICTs, especially with the wide application of PC in diverse contexts, a divide between the information-haves and have-nots was sensitively observed and warned (Compaine, 2001). Since early 1990s, digital divide has gradually become a convenient label, or more precisely, a metaphor (Courtright & Robbin, 2001), in describing the inequality of possessing and using ICTs, especially the Internet connectedness.

The first group of definitions varies on the concrete referents of what digital means. In a narrow sense of the definition, digital divide is particularly referred to the inequality of Internet access and use among different social groups or localities. US Department of Commerce's (1995, 2001) Falling through the net reports represent the most influential version of the stream. Zhu (2002) also takes Internet penetration as the sole indicator of what digital means in his construction of digital divide index (DDI) while taking age, sex, education and occupation collectively as the categorizing factors. In short, in this stream of definitions, digital divide is operationalized to

Internet access/penetration divide categorized by demographics and social status factors.

However, to many people, the term *digital* means a wide range of ICTs other than the Internet. Arquette (2001) labeled it as the concept fit disjuncture in the studies of digital divide, i.e., to measure global digital equality in terms of teledensity or Internet penetration. Employing the so-called Information Intelligence Quotient (IIQ) analytical framework, he uses ICT infrastructure rather than a single ICT such as Internet or telephony as the subject of the "digital".

A second clue of conceptualizing the digital divide basically focuses on the meaning of "divide". Many different analytical perspectives on this concern are available. Jackel (2001) exemplifies some of these:

- a macro-level-comparison of the so-called First and Third world or a comparison of rich and poor countries;
- a comparison of differences in modern societies according to the level of integration in the labor market;
- a comparison of differences in modern societies according to education groups, gender and age, more general a comparison of generations;
- a specification of differences in modern societies according to communication skills;
- a comparison of different diffusion curves as a result of differing demands.

As can be seen, the dimensions identified by these perspectives are noticeably diverse.

Synthesizing the prior research on digital divide, Arquette (2001) proposed an organizational framework based on three dimensions of digital divide: ICS infrastructure, access, and use. ICS infrastructure refers to the technophysical means by which voice, audio, video and/or data communication circulates. The operationalization of the dimension involves the specification of sixteen indicators, including telephony penetration (wire line and wireless), Internet hosts, and costs of calls, etc. The second dimension is ICS access which focuses on the ability of persons interested in using the infrastructure (regardless of that infrastructure quality or quantity) to gain access to the infrastructure. Nineteen indicators are developed to operationalize the dimension. The third dimension of digital divide that Arquette (2001) specifies is ICS use. Use-based conceptualizations of digital divide are conceived in terms of how people employ the technologies. Another nineteen indicators are developed to measure the situation of this dimension of digital divide. In summary, IIQ is an aggregate meta-analytic framework for assessing the state of digital divide among different nations or regions.

A notable point implied by the IIQ is its consideration of the dimension of ICT use. In fact, access is not enough is becoming a recognizable consensus (e.g., Blau, 2002; Jackel, 2001; Nagaraj, 2002). In other words, merely connecting people and computers will not bridge the digital divide (Blau, 2002), and there's digital discrimination among the information haves, too (Nagaraj, 2002). Jackel (2002) labels the divide among ICT haves as the second level of "divide".

NIQ is a composite index comprised of 20 indicators in six dimensions. It is the operational definition of the National Informatization Level (NIL). In the remaining part of the article, the digital divide is discussed in terms of this NIL which is operationally defined as NIQ. The six dimensions of NIQ are:

- The development and application of information resources (IR): The indicators under this umbrella term include Radio and TV broadcasting hour/per 1,000 people; Bandwidth per person; Telephone use frequency per person; Total capacity of Internet data base.
- Information network construction (IN). There are four components in this dimension, including Total length of long distance cable; Microwave channels; Total number of satellite stations; Number of telephone lines per 100 people.
- The application of information technologies (IT): the indicators for this dimension include Number of cable TV stations per 1,000 people; Number of Internet users per 1 million people; Number of computers per 1,000 people; Number of TV sets per 100 people; E-commerce trade volume; Proportion of investment in the information industry by enterprises to the total fixed investment.
- Information industry development (II): there are two indicators designed to reflect the situation of this dimension: Added value contributed by the information industry to the total GDP; Contributions made by the information industry to the total GDP increase.
- Human resources of informatization (HR): there are two indicators for this dimension: Proportion of university graduates per 1,000 people; and Information index which refers to the proportion of expenditure other than fundamental consumption to the total expenditure.
- The environment for informatization development (EI). Two indicators are designed to measure the situation of the dimension: proportion of expenses for research and development of the information industry to the country's total budget in R&D; Proportion of investment on the infrastructural

development of the information industry to the country's total investment in capital construction.

Compared to other index used to reflect the digital divide or ICT development in a country or region, NIQ has several characteristics:

1. It is a multi-dimensional composite index. Therefore, NIQ is a comprehensive reflection of the state informatization level rather than the development of some particular ITs.
2. As for its application in assessing digital divide, the divide it evaluates is a geographical divide rather than informatization divide among different social groups or divides defined by other factors.
3. The index covers a wide range of the aspects regarding the informatization development. Particularly, NIQ emphasizes the importance of information industry in its structure of dimensions. The proportion of indicators related to information industry is notably high which, reflects the fact that NIQ will be a guideline for the promotion and adoption of IT in developing countries.

DEVELOPING A RESEARCH AGENDA

We now raise a series of points for further discussion. We believe these issues represent the most pressing unanswered questions concerning access and the impact of the digital divide on the emerging digital economy.

Computers in the Home

While previous research has shown that inequalities in Internet access in schools persist (Educational Testing Service, 1997, Sax et al., 1998), the research reviewed here suggests that inequalities in Internet access at home may be even more problematic. The role of access to the Internet at home needs to be much more clearly understood (Abrams, 1997). Whites are more likely to have access to the Internet and to have ever used the Web than Minorities and these gaps appear to be *increasing* over time. Probing deeply, we have discovered that among recent Web users, who by definition have access, the gaps in Web use have been *decreasing* over time. Over time, there appear to no or only slight differences between whites and Minorities in how recently they had used the Web, how frequently, or in their length of time online. Gaps in general Web access and use between Different minorities and whites appear to be driven by whether or not there is a computer present in the home. Access to a personal computer, whether at home, work, school or somewhere else, is important because it is currently the dominant mechanism by which individuals can access the Internet. We believe that access translates into usage. Overall, individuals who own a home computer are much more likely than others to use the Web. This suggests that programs that encourage home computer ownership (see, for example, Roberts, 1997) and the adoption of inexpensive devices that enable Internet access over the television should be aggressively pursued, especially for Minorities.

Morrisette (1999) forecasts that by the year 2003, over half of all households in the United States will have access to the Internet, but that PC penetration could stall at 60% of households. Research is necessary to understand what motivates individual-level adoption of home computers and related technologies, as well Internet adoption, both within and outside the home. Additionally, research is required to understand the long-term impact of home computer ownership on Internet access and use. Katz and Aspden (1997) investigated the role of social and work networks in introducing people to the Internet. The dominant three ways people were originally introduced to the Internet were 1) taught by friends or family, 2) learned at work and 3) self taught. Formal coursework was the *least* often mentioned way people were introduced to the Internet. Long term Internet users were most likely to have learned at work; for recent Internet users, friends/family and self-taught were equally important. These results reinforce the importance of the presence of a computer at home, or the opportunity to access the Web from locations other than the home, in stimulating Web use.

Insight into the importance of reducing this gap in Web use between whites and African-Americans is provided by Anderson and Melchior's (1995) discussion of *information redlining*. Information redlining signifies the relegation of minorities into situations where satisfying their information needs is weighed against their economic and social worth. From the minority point of view, this is both an access issue and a form of discrimination. The new technologies of information are not simply tools of private communication as is a telephone, or tools of entertainment as is a television. They provide direct access to information sources that are essential in making social choices and keeping track of developments not only in the world at large, but also within their immediate neighborhoods. Unless the neighborhoods are properly served, there is no way out of information redlining for most of these disadvantaged groups. Research on this topic is warranted.

There are also interesting differences in media use between whites and Minorities that also deserve further probing. For example, although the rate of home PC ownership among Minorities is flat or even decreasing, the rates of cable and satellite dish penetration are increasing dramatically for Minorities. At a minimum, these results suggest that Minorities may make better immediate prospects than whites for Internet access through cable modems and satellite technology.

Web Use Outside of the Home

In addition to gaps in home computer ownership, the implications of differential Internet access at locations outside the home, including school, the workplace and other locations needs to be clearly understood. Research suggests that additional access points stimulate usage. Further research is necessary to understand the impact of multiple access points on Web use, particularly for individuals who have no access at home. Public-private initiatives such as Bell Atlantic's efforts in Union City and Bill Gates announcement of a $200 million gift to provide library access to the Internet are a step in the right direction (Abrams, 1997). It has also been noted that "community networks and public access terminals offer great potential for minority communities" (Sheppard, 1997). Further, the recent roll-out of E-rate funds (Schools and Libraries Corporation, 1998) provides a significant opportunity for researchers to understand the factors important in stimulating Web usage among those least likely to have access.

School Web Use

The role of Web access in the schools, compared to other locations, needs to be clearly understood. Students enjoy the highest levels of Internet access and Web use, especially when there are computers in their households. However, white students are still more likely than Minority students to have access and to use the Internet, and these gaps persist over time. Indeed, our findings closely parallel statistics comparing student Internet use at private universities and minority public colleges (Sax et al., 1998). As a recent report by the Educational Testing Service (1997) makes clear:

- There are major differences among schools in their access to different kinds of educational technology.
- Students attending poor and high-minority schools have less access to most types of technology than students attending other schools.
- It will cost about $15 billion, approximately $300 per student to make all our schools "technology rich."

This is five times what we currently spend on technology, but only 5% of total education spending. Anderson and Melchior (1995) cited lack of proper education as an important barrier to technology access and adoption. Access to technology does not make much sense unless people are properly educated in using the technologies. Our data do not speak to the quality of the hardware/network connections, or the quality of information technology education that is provided by school.

Differences in Search Behavior

Reasons for the gap between different minorities and whites in Web search behavior need to be clearly understood. Such differences could have important implications for the ultimate success of commercial efforts online. White Web users are more likely to report searching for product or service-related information than Minorities. One possibility is that despite sites such as NetNoir and Black Entertainment Television, general purpose search agents may not be perceived as an effective way to locate Web content that is compelling to minority users (*New Media Week* 1997). This suggests the development of search engines and portals targeted to the interests of racial/ethnic groups.

Shopping Behavior

There appear to be no differences between different minorities and whites in the incidence of Web shopping. Is this because race doesn't matter for "lead users?" who are most likely to shop, or is this because commercial Web content better targets racial and ethnic groups than does non-commercial Web content? Previous research (Novak, Hoffman, & Yung, 1999) suggests that more skill is required to shop online than to search. However, as noted above, whites are more likely to search for information online than are Minorities. More generally, consumer behavior in the commercial Web environment is complex and only weakly understood. Further research is needed to explore fully the differences in consumer behavior on the Web and their implications for commercialization.

Multicultural Content

Studies investigating the extent of multicultural content on the Web are needed. Another possibility for the gap between different minorities and whites in Web search behavior is that there is insufficient content of interest to different minorities. *Interactive Marketing News* (1997) claimed that "while there are about 10 million sites on the Web, there are fewer than 500 sites targeted" to different minorities. However, others have commented on the multicultural diversity of the Web. Skriloff (1997) reported, "there are thousands of Web sites with content to appeal to other ethnic groups. Many of these sites are ready-for-prime time with high quality content, graphics, and strategic purpose."

Community Building

Are there different cultural identities for different parts of cyberspace? Schement (1997) notes that by the year 2020, major U.S. cities such as Los Angeles, Chicago, and New York will have increasingly divergent ethnic profiles, and will take on distinctive cultural identities. An important question is whether there are divergent ethnic profiles for areas of cyberspace. While the questions in the three IDS do not allow us to directly address this issue, our analyses provide some preliminary evidence of divergent ethnic profiles for various Web usage situations. For example, minorities appear to be more likely to use the Web at school and at other locations, and in some cases, are more likely to use the Web at work. How much of this is driven by the lack of a PC in the home and how much by other factors we have yet to hypothesize and investigate?

In addition to facilitating community building at the global level, the Web also facilitates neighborhood-level community building. Schwartz (1996) discusses how the Internet can be used as a vehicle for empowering communities. Anderson and Melchior (1995) raise the issue of the ways in which telecommunications can be used to strengthen communities. Thus, we should expect to find neighborhood Web sites emerging as an important aspect of cyberspace, and that these Web sites will parallel the ethnic profiles of the corresponding physical communities.

Income and Education

Income matters, but only after a certain point. Household income explains race differences in Internet access, use, home computer ownership and PC access at work. In terms of overall access and use, higher household income positively affects access to a computer. But at lower incomes, gaps in access and use between whites and Minorities existed and were increasing. Research is necessary to determine the efforts most likely to be effective to ensure access for lower-income people, especially Minorities. The situation is different with education. As with income, increasing levels of education positively influences access, Web use, PC ownership and PC access at work. However, whites are still more likely than Minorities to have access to and use the Internet, and own a home computer, and these gaps persist even after controlling for educational differences.

The policy implication needs to be carefully considered: To ensure the participation of all People in the information revolution, it is critical to improve the educational opportunities for Minorities. How this might best be achieved is an open research question.

CONCLUDING REMARKS

The consequences to civil society of the digital divide in Internet use are expected to be severe (Beaupre & Brand-Williams 1997). Just as A.J. Liebling observed for the freedom of the press (Liebling 1960), the Internet may provide for equal economic opportunity and democratic communication, but only for those with access. The United World economy may also be at risk if a significant segment of our society, lacking equal access to the Internet, wants for the technological skills to keep National firms competitive.

The broad policy implications for these findings should not be overlooked. By identifying those who are truly in need, policymakers can prudently and efficiently target support to these information disadvantaged. Only when this point is reached can *all* those who desire to access the Online Services be possibly accommodated. However, connectivity to all such households will not occur instantaneously; rather, there is a pivotal role to be assumed in the new electronic age by the traditional providers of information access for the general public — the public schools and libraries. These and other "community access centers" can provide, at least during an interim period, a means for electronic access to all those who might not otherwise have such access. Policy prescriptions that include public "safety nets" would complement the long-term strategy of hooking up all those households who want to be connected to the Online Services.

REFERENCES

Abrams, A. (1997, June). Diversity and the Internet. *Journal of Commerce.*

Anderson, T. E .& A. Melchior. (1995). Assessing telecommunications technology as a tool for urban community building. *Journal of Urban Technology, 3*(1), 29- 44.

Arquette, T. J. (2001). *Assessing the digital divide: Empirical analysis of a meta-analytic framework for assessing the current state of information and communication system development.* Unpublished manuscript, Northwestern University, Department of Communication Studies.

Beaupre, B. & Brand-Williams, O. (1997, February 8). Sociologists predict chasm between black middle-class, poor will grow. *The Detroit News.*

Blau, A. (2002, June/July). Access isn't enough. *American Libraries*, 50-52.

Compaine, B. M. (Ed.). (2001). *The digital divide. Facing a crisis or creating a myth?* Cambridge, MA: MIT Press.

Courtright, C. & Robbin, A. (2001, November). Deconstructing the digital divide in the United States: An interpretive policy analytic perspective. Paper presented at *International Association of Media and Communication Research and International Communication Association Symposium on the Digital Divide*, Austin, Texas, USA.

Educational Testing Service. (1997). *Computers and classrooms: The status of technology in U.S. schools*, Policy Information Center. Retrieved from *http://www.ets.org/research/pic/compclass.html*

eStats (1999). Net Market Size and Growth: U.S. Net Users Today. May 10, *http://www.emarketer.com/estats/nmsg_ust.html*

Interactive Daily (1997). More different minorities plan to go online. February 18.

Jackel, M. (2001). *Inclusion, exclusion and the diversity of interests. Is digital divide an adequate perspective?* Paper prepared for International Association of Media and Communication Research and International Communication Association Symposium on the Digital Divide, November 15-17, 2001, Austin, Texas, USA.

Katz, J. & Aspden, P. (1997). *Motivations for and barriers to Internet usage: Results of a national public opinion survey*. Paper presented at the 24th Annual Telecommunications Policy Research Conference, Solomons, Maryland, October 6, 1996.

Laugsksch, R. C. (2000) Scientific literacy: A conceptual overview. *Science Education, 84(1)*, 71-94.

Liebling, A.J. (1960). *The New Yorker*, 36, 105, May 14.

McLeod, J. M. & Pan, Z. (n.d.). *Concept explication and theory construction.* School of Journalism and Mass Communication, University of Wisconsin-Madison. Obtained electronically on July 2002.

Morrisette, S. (1999). *Consumer's digital decade.* Forrester Report, January, Forrester Research, Inc. Online: *http://www.forrester.com/*

Nagaraj, N. (2002). The other divides. *Businessline*, April 24.

New Media Week (1997). BET, Microsoft sees potential in African-American audience. March 3.

Nielsen Media Research (1997a). *The Spring '97 CommerceNet/Nielsen Media InternetDemographic Survey, Full Report.* Interviews Conducted in December '96/Jan '97. Volumes I & II.

Novak, T.P., Hoffman, D.L. & Yung, Y.F. (1999). Modeling the flow construct in online environments: A structural modeling approach. Manuscript under review, *Marketing Science.*

Roberts, R.M. (1997). Program lowers costs of going online: Families can get break on equipment. *The Atlanta Journal and Constitution*, June 19.

Rutkowski, A.M. (1998). *Internet trends.* February. Center for Next GenerationInternet, Washington, D.C. Online: *http://www.ngi.org/trends.htm*

Sartori, G. (1984). Guidelines for concept analysis. In G. Sartoni (Ed.), *Social science concepts: A systematic analysis* (pp. 15-85). Beverly Hills, CA: Sage

Sax, L.J., Astin, A.W., Korn, W.S., & Mahoney, K.M. (1998). *The American Freshman: National Norms for Fall 1998.* Higher Education Research Institute, UCLA Graduate School of Education & Information Studies. *http://www.acenet.edu/news/press_release/1999/01January/freshman_survey.html*

Schement, J.R. (1997). *Thorough Americans: Minorities and the New Media.* Paper presented at the Aspen Institute Forum, October 1996.

Schools and Libraries Corporation (1998). *First Wave of E-Rate Funding Commitment Letters Sent.* November 23 news release.

Schwartz, E. (1996). *NetActivism: How Citizens Use the Internet.* Sebastopol, CA: O'Reilly & Associates, Inc.

Sheppard, N. (1997). Free-Nets Reach Out to Communities' Needs. *The EthnicNewsWatch*, April 30.

Skriloff, L. (1997). Out of the Box: A Diverse Netizenry. *Brandweek*, February 17.

Tichenor, P. J., Donohue, G. A., & Olien, C. N. (1970). Mass media and differential growth in knowledge. *Public Opinion Quarterly, 34*, 158-170.

U.S. Department of Commerce (1995). Falling through the net: A survey of the have nots in rural and urban American. Retrieved from *http://www.ntia.doc.gov/ntiahome/fallingthru.html*

U.S. Department of Commerce (2001). Falling through the net: Toward digital inclusion. Retrieved from *http://www.esa.doc.gov/fttn00.pdf*

KEY TERMS

DDI: Digital Divide Index

Digital Divide: a "**digital divide**" between the information "haves" and "have-nots."

IIQ: Information Intelligence Quotient

ICT: Information and Communication Technologies

IT: Information Technologies

Knowledge Gap: the different knowledge possession through mass media by social groups with varied social-economic-status.

NIQ: National Informatization Quotient

Bridging the Industry–University Gap through Action Research

Ned Kock
Texas A&M International University, USA

INTRODUCTION

Virtually every university in the US and overseas has seen a significant increase in demand for information technology (IT) courses and programs in the last 10 years (Greenspan, 1999; Monaghan, 1998; Ross, 1998). At the source of this demand is an ever-growing need for qualified IT professionals in most companies, whether the companies are in technology industries or not (Alexander, 1999; Andel, 1999; Trunk, 2000; Wilde, 1999).

Given the practical motivation above, one would expect university IT courses to be closely aligned with the industry's basic needs. Nevertheless, the gap between industry and academia in the field of IT (King, 1998; Kock et al., 2002; Richter, 1999) seems to be widening rather than contracting, which is evidenced by some symptoms: (a) students complaining about their lack of "real-world" IT experience when they graduate; (b) industry representatives pointing out that universities do not prepare students for the challenges and complexity of corporate IT management; and (c) faculty teaching topics that are related to their research yet far removed from the daily reality faced by IT professionals.

One way of addressing the problematic situation above is to establish industry-university partnerships. Such partnerships, particularly those involving research universities, have been commonplace for quite some time, and are arguably on the rise (Burnham, 1997; Wheaton, 1998). Irrespective of economic sector or industry, the vast majority of industry-university partnerships are of the *research partnership* type, which predominantly involves applied firm-specific research. In this type of partnership, funding from the industry partner is received in exchange for "intellectual horsepower" in the form of research services and technology transfer (Hollingsworth, 1998; Meyer-Krahmer, 1998).

A much less common type of industry-university partnership is what we refer here to as a *course partnership*, which gravitates around a regular university course (or set of courses) rather than a research project or program. In these types of partnerships, the industry partner agrees to sponsor one or more courses in which the students are expected to apply concepts and theory learned in class to the solution of some of the industry partner's key problems. Students benefit from the direct contact with the industry they are likely to join after they graduate as well as professional relationships they are able to establish during the course.

This article discusses a *course partnership* involving a large engineering and professional services company, and a public university, both headquartered in Philadelphia. An action research study of the course partnership is used as a basis.

Like typical action research studies (Checkland, 1991; Lau, 1997; Peters & Robinson, 1984; Winter, 1989; Wood-Harper, 1985), ours aimed at providing a service to the research clients (Jonsonn, 1991; Rapoport, 1970; Sommer, 1994) while at the same time performing an exploratory investigation of the effect of Web-based collaboration technologies on course partnerships. The research clients in question were the students and the industry partner. Also, in line with a subclass of action research, namely participatory action research (Elden & Chisholm, 1993; Greenwood et al., 1993; McTaggart, 1991; Whyte, 1991), one of the research clients, the industry partner, participated actively in the compilation and analysis of the exploratory research data, as well as in the interpretation of the findings.

BACKGROUND

Our study was centered on a different and arguably promising approach to implementing course partnerships that was recently proposed to address the problems outlined previously (Kock et al., 2000, 2002, 2003). The approach involves conducting certain courses, particularly senior undergraduate and graduate courses, in close partnership with companies. Such courses are designed so that the concepts and theory discussed in class are applied in team course projects geared at solving immediate problems at the company partner. Other fundamental characteristics of these course partnerships are:

- All team projects are conducted in one single organization. Letting student teams identify organizations they would want to work with, based on criteria defined by the instructor, usually leads to different student teams conducting projects in different organizations, and thus to significant discrepancies in project complexity, project scope, and organizational support across different student teams. These problems can have a negative impact on learning, and are considerably reduced when all team projects are conducted in one single organization.

- Potential projects are identified in advance. The identification of a potential project by student teams can take up to 5 weeks of a 14-week course. One may argue that this is acceptable, as long as concepts and theory are covered in the classroom during those initial 5 weeks. However, in addition to identifying a project, a student team also needs to learn about the organizational culture, key people, and specific business processes they will be dealing with. This can easily take up another 5 weeks, leaving little time for other key project activities (e.g., business process redesign and IT implementation). The solution to this problem is to identify potential projects in advance, prior to the formal start of the course, and distribute them among student teams in the first week of the course.
- Top management personally sponsors the course partnership. Often, when students are asked to come up with their own company-sponsored course projects, the individuals who sponsor the projects are not senior managers. As a result, a project sponsor may be reluctant or lack the authority to approve organizational changes or purchases of hardware and software necessary for a project to be effectively completed. These difficulties are mitigated when top management directly sponsors team projects.

It is important to note that course partnerships with these characteristics require a considerable amount of extra time and effort from the students and instructor, well beyond what is usually expected in traditional courses. In addition to applying the concepts and theory learned in class, students would also have to learn "on-the-fly" how to effectively deal with issues that are found in real-world projects (e.g., organizational culture and politics). The instructor, on the other hand, has to also take on project management, industry-university liaison, and inter-organizational team facilitation responsibilities, in addition to traditional course delivery and student mentoring duties.

IMPLEMENTING A COURSE PARTNERSHIP: TEMPLE UNIVERSITY AND DAY & ZIMMERMANN, INC.

The course partnership idea discussed previously has been implemented through a collaboration between Temple University, a large public university located approximately two miles from downtown Philadelphia, and Day & Zimmermann, Inc., a US$ 1.5 billion engineering and professional services company headquartered in downtown Philadelphia. The course was a pilot version of CIS650 - Process Design and Information Technology, a newly developed course in Temple's Computer and Information Science Department dealing with process analysis and redesign issues.

The course instructor (the author of this article) initiated the course partnership by sending a letter to one of the senior executives at Day & Zimmermann. In the letter, the course instructor inquired if the company would be interested in partnering with Temple University, providing details about the partnership. The partnership was approved after an initial meeting involving the course instructor and senior managers at the company.

The course project required students to analyze and redesign five of Day & Zimmermann's business processes using the concepts, theory and techniques taught in class. The course partnership and related project had direct support from Day & Zimmermann's Chief Information Officer (CIO) from the outset. A senior manager at Day & Zimmermann, who reported directly to the CIO, was assigned the responsibility of managing the project together with the course instructor. The project involved, directly and indirectly, over 30 Day & Zimmermann employees and 26 Temple students.

The students were split into five process redesign teams, which periodically met with key Day & Zimmermann employees at the company's headquarters in downtown Philadelphia. Each team analyzed and redesigned one process, generated three reports, and delivered an oral presentation to Day & Zimmermann management at the end of the course. The first report generated by each team contained a detailed description of the process targeted; the second a detailed description of the redesigned process and the rationale behind the redesign decisions; and the third a detailed analysis of IT solutions to enable the new (redesigned) process.

WEB SITE REMOVES OBSTACLES TO PARTICIPATION

Before the course partnership was started, two main obstacles had to be dealt with. First, Day & Zimmermann employees were expected to actively participate in the process redesign efforts. In order to do so, they had to understand the concepts and theory used by the students. Yet, most of the Day & Zimmermann employees likely to be involved in this project could not come to Temple to audit the course together with the students. Also, given that Temple students and Day & Zimmermann employees were not co-located, a great deal of their interaction would have to occur by means other than face-to-face meetings. The solution to overcome these two obstacles was the development of a password-protected Web site, which allowed Day & Zimmermann employees to access all the course material online. The Web site also supported interaction between them and Temple students through shared document areas, multimedia components, and discussion boards.

WAS THE COURSE PARTNERSHIP SUCCESSFUL?

The partnership was considered a success by Day & Zimmermann management and employees, as well as by Temple students. Managers emphasized the anytime/anyplace collaboration between Day & Zimmermann employees and Temple students enabled by the Web site as one of the key elements that made the course partnership a very successful collaborative effort.

Temple students emphasized the real-world experience as one of the most positive aspects of the course. Following is a representative comment by a student extracted from one of the anonymous course evaluation forms completed at the end of the course:

"The learning experience was very rich. The group project gave us hands on experience in applying the redesign techniques we learned in the course. It was a great experience to work with upper level IT management!"

Table 1 shows the average scores for several question/statements asked from the students in connection with the course. The question/statements were part of a standard course and faculty evaluation, and were answered anonymously.

Several students pointed out that the course required considerably more time and effort from them than most traditional university courses they had taken before. In spite of that, their anonymous evaluations of the course were very positive, as it can be seen in Table 1. The average answer to the question/statement "Overall, this is one of the best courses I have had at Temple" was 3.53, on a 0-to-4 scale. The average answer to the question/statement "Overall, I have learned a great deal from this course" was 3.41, also on a 0-to-4 scale.

An added benefit for Day & Zimmermann was the ability to identify young talent based on observation of business-relevant action (as opposed to the traditional analysis of resumes). Day & Zimmermann were able to observe a group of 26 students in action over a two-month period and identify several students whom they would like to consider hiring. This is not as easy to accomplish with other approaches for identifying new graduates for hiring, of which internships are perhaps the most popular. There are two key reasons for this. First, the number of interns that could be hired for a two-month period by an organization would typically be considerably smaller, thus significantly reducing the number of students that

Table 1. Course evaluation scores

Question/statement	Score
The objectives and requirements of the course were made clear.	3.82
The instructor clearly communicated the subject.	3.88
The instructor was open to questions and comments.	3.82
The instructor was accessible outside the classroom.	3.70
The instructor was impartial in evaluating my performance.	3.47
The instructor expected academic excellence from students.	3.47
Overall, the instructor did an excellent job teaching.	3.82
Overall, I have learned a great deal from this course.	3.41
Overall, this is one of the best courses I have had at Temple.	3.53

(Score = average score for the class regarding the question; Range: 0-4)

Day & Zimmermann managers would be able to observe in action during that period of time. Second, the tasks that the interns would be assigned to would not usually be nearly as complex and strategically relevant as those carried out in this course.

CONCLUSION AND LESSONS FOR FUTURE PARTNERSHIPS

The general perception at the end of the course partnership was that it had been an exciting and rewarding experience for all those involved. Students saw the course as a valuable experience that provided them with a unique view of IT management and which complemented the concepts, theory and techniques learned in the course and throughout their university program. Day & Zimmermann managers perceived the input provided by the students as very valuable and likely to lead to concrete business process improvements.

Also, a few lessons have been learned along the way that can be useful for universities and companies planning to implement similar course partnerships in the future. These lessons are summarized below.

- The course partnership should have two main co-project managers, one from academia and one from industry. As with most inter-organizational initiatives, the scope of management authority does not usually extend beyond organizational boundaries. Students will respond more quickly to requests by the instructor than to requests by a manager of the partner organization. Similarly, employees of the partner organization will respond more quickly to a request by someone who has formal authority within the organization than to a request by the instructor. Therefore, the instructor should share the responsibility of managing the project with a member of the partner organization who has enough formal authority to oversee all the team projects.
- The course partnership should include at least one purely social activity. Social activities allow for important information exchanges that would not normally occur in formal meetings. For example, prior to the final oral presentation by student teams at Day & Zimmermann, a pizza party (paid for by the company) was held in downtown Philadelphia. After the party, several people remarked that the personal knowledge they learned from informal conversations during the party was invaluable to them. The party was also seen as a "thank you" gesture by the partner organization to the students. Not only did this boost morale, but it also helped the students relax for the presentation the next day, as they got to know the people they would be presenting to at a more personal level.
- The business problems addressed through the course partnership should be "real" and of high relevance to the partner organization. Because students are involved, not professional consultants, the partner organization may be tempted to create "toy" problems to be solved through the course partnership, rather than address real and relevant business problems. The motivation for this may be the perceived risks linked to not accomplishing the goals of the project (e.g., wasted time, internal conflicts, and reluctance to engage in future organizational change efforts). The problem with this approach is that it is likely to relieve the students from any real responsibility and, at the same time, decrease the motivation for employees to get involved. A better alternative to reduce risk is to involve experienced consultants at critical points in the course partnership (the costs are relatively low since few consultant-hours are likely to be used).
- Partner organization employees should be asked to report on and make an oral presentation of their projects too. Course partnerships such as the one described here are, as the name implies, collaborative endeavors in which members of the two organizations involved should contribute evenly. Therefore, *key me*mbers of the partner organization should also be asked to make presentations about their projects. This is particularly important because, in some cases, despite full dedication from the students, a project may fail to accomplish its initial goals. And, a closer examination may indicate that the failure was not the students' fault, but that it was caused by lack of interest or commitment from the part of one or more employees. Employee reporting and presentations are useful in assessing whether this is the case, which in turn is important for appropriately grading the students' coursework. Moreover, the requirement to report and present their projects communicates to the partner organization employees that they are equally responsible for the outcomes of the project, which is likely to increase their level of commitment and dedication to the project.

The course partnership discussed in this article went well beyond the idea of having students conduct their course projects in real organizations, which is an approach adopted by many project-based courses around the world. It seems that the close cooperation between Temple University and Day & Zimmermann that characterized the course partnership presented here was the key reason for its success. This type of cooperation requires extra time and effort from everyone involved – students, instructor, company management and employees. Yet, the positive (tangible and intangible) outcomes of the partnership seem to easily outweigh its costs.

REFERENCES

Alexander, S. (1999). High demand for hot skills. *Computerworld, 33*(39), 4-6.

Andel, T. (1999). IT is your business. *Material Handling Engineering, 54*(7), 18.

Burnham, J.B. (1997). Evaluating industry-university research linkages. *Research Technology Management, 40*(1), 52-56.

Carnes, K.C., & Gierlasinski, N.J. (1999). Have you considered a faculty intern? *The National Public Accountant, 44*(3), 31-32.

Checkland, P. (1991). From framework through experience to learning: The essential nature of action research. In H. Nissen, H.K. Klein & R. Hirschheim (Eds.), *Information systems research: Contemporary approaches and emergent traditions* (pp. 397-403). North-Holland..

Elden, M., & Chisholm, R.F. (1993). Emerging varieties of action research. *Human Relations, 46*(2), 121-141.

Greenspan, A. (1999). The interaction of education and economic change. *The Region, 13*(1), 6-11.

Greenwood, D.J., Whyte, W.F., & Harkavy, I. (1993). Participatory action research as a process and as a goal. *Human Relations, 46*(2), 175-191.

Hollingsworth, P. (1998). Economic reality drives industry-university alliances. *Food Technology, 52*(7), 58-62.

Jonsonn, S. (1991). Action research. In H. Nissen, H.K. Klein & R. Hirschheim (Eds.), *Information systems research: Contemporary approaches and emergent traditions* (pp. 371-396). North-Holland.

King, J. (1998). Labor confab issues call for training. *Computerworld, 32*(2), 1, 16.

Kock, N., Auspitz, C., & King, B. (2000). Using the Web to enable industry-university collaboration: An action research sudy of a course partnership. *Informing Science* (special issue on Organizational Learning), *3*(3), 157-167.

Kock, N., Auspitz, C., & King, B. (2002). Bridging the industry-university gap: An action research study of a Web-enabled course partnership. In E. Cohen (Ed.), *Challenges of information technology education in the 21st century* (pp. 166-186). Hershey, PA: Idea Group Publishing.

Kock, N., Auspitz, C., & King, B. (2003). Web-supported course partnerships: Bringing industry and academia together. *Communications of the ACM, 46*(9), 179-183.

Kock, N., Gray, P., Hoving, R., Klein, H., Myers, M., & Rockart, J. (2002). IS research relevance revisited: Subtle accomplishment, unfulfilled promise, or serial hypocrisy? *Communications of the AIS, 8*(23), 330-346.

Lau, F. (1997). A review on the use of action research in information systems studies. In A.S. Lee, J. Liebenau & J.I. DeGross (Eds.), *Information systems and qualitative research* (pp. 31-68). London: Chapman & Hall.

McTaggart, R. (1991). Principles for participatory action research. *Adult Education Quarterly, 41*(3), 168-187.

Meyer-Krahmer, F. (1998). Science-based technologies: University-industry interactions in four fields. *Research Policy, 27*(8), 835-852.

Monaghan, P. (1998). Growing demand for computer animators spurs a new program at U. of Washington. *The Chronicle of Higher Education, 44*(48), 23-35.

Peters, M., & Robinson, V. (1984). The origins and status of action research. *The Journal of Applied Behavioral Science, 20*(2), 113-124.

Rapoport, R.N. (1970). Three dilemmas in action research. *Human Relations, 23*(6), 499-513.

Richter, A. (1999, November 7). Silicon Island, wired but underpopulated. *New York Times,* 14LI, 1.

Ross, P.E. (1998). Enjoy it while it lasts. *Forbes, 162*(2), 206.

Sommer, R. (1994). Serving two masters. *The Journal of Consumer Affairs, 28*(1), 170-187.

Trunk, C. (2000). Information technology in logistics: Material flow at Moen. *Material Handling Management, 55*(1), 8-10.

Wheaton, Q. (1998). Government-university-industry cooperation: Does it work? *Quality, 37*(5), 20-24.

Whyte, W.F. (Ed.). (1991). *Participatory action research.* Newbury Park, CA: Sage.

Wilde, C. (1999). Hiring in triplicate. *Computerworld, 33*(28), 77.

Winter, R. (1989). *Learning from experience: Principles and practice in action-research.* New York: The Falmer Press.Wood-Harper, A.T. (1985). Research methods in information systems: Using action research. In E. Mumford, R. Hirschheim, G. Fitzgerald & A.T. Wood-Harper (Eds.), *Research methods in information systems* (pp. 169-191). North-Holland, Amsterdam.

KEY TERMS

Action Research: Type of research approach in which the researcher attempts to improve the research client, which can be an organization, while at the same time generating relevant academic knowledge.

Co-Partnership Managers: The course instructor, on the university side, and a senior manager of the client organization, who jointly manage a course partnership project.

Course Partnership: Course-based industry-university partnerships, where a course is designed so that the concepts and theory discussed in class are applied in team course projects geared at solving immediate problems at the company partner.

Industry-University Gap: Disconnect between the knowledge and skill needs of industry practitioners and the knowledge and skills imparted on students by universities.

Process: Set of interrelated activities through which an organization transforms inputs into value-added outputs. Inputs and outputs can be tangible (e.g., materials, parts) or intangible (e.g., services, information) items.

Process Redesign: Approach to organizational improvement through transformation of business processes. The term refers to business process change approaches emphasizing gradual change (e.g., total quality management) or radical change (e.g., business process reengineering).

Research Partnership: Industry-university partnership involving collaboration in applied firm-specific research. In this type of partnership, funding from the industry partner is received in exchange for "intellectual horsepower" in the form of research services and technology transfer.

Web-Based Course Partnership: A course partnership in which a Web site is developed to serve as a central communication hub and document repository for the partnership.

Building and Management of Trust in Information Systems

István Mezgár
Budapest University of Technology and Economics, Hungary &
Hungarian Academy of Sciences, Budapest, Hungary

INTRODUCTION

Thanks to the rapidly developing information and communication technologies, the complexity of networked organizations becomes very high, so the representation of their structure, the description of their operation, and their control needs new technologies and approaches. The availability of individuals independent from location and time means mobility, and that is an important attribute of today's society. This mobility can be achieved by using different types of wireless networks such as wireless wide area networks (WWANs—GSM, GPRS, and UMTS), wireless local area networks (WLANs, such as WiFi 802.11b, 802.11a), and wireless personal area (or Pico) networks (WPAN—Bluetooth, IrDA2).

In spite of the application of high-tech approaches, tools, and methodologies, there is a common point in all of the organizations; human beings make the most of important decisions, and they operate and use systems. According to experience, improper application of this human factor can make operation very inefficient, even in the case of technically advanced systems. The lowest level of connection among systems is made through protocols; the highest contact level is among decision makers and users, namely, human beings. A very important element of this human contact is trust. In a networked organization, trust is the atmosphere, the medium in which actors are moving (Castelfranchi & Yao-Hua Tan, 2001). Only trust can bridge the cultural, geographical, and organizational distances of team members (and even of firms) and keep them from turning into unmanageable psychological distances. Trust is the base of cooperation, the normal behavior of the human being in society. The ability of enterprises to form networked systems depends on the existing level of trust in the society and on the capital of society (Fukuyama, 1995). As the rate of cooperation is increasing in all fields of life, the importance of trust is evolving even faster.

Lack of trustworthy security services is a major obstacle in the use of information systems in private, in business (B2B [business to business]), as well as in public services. Trust is intimately linked to consumers' rights, like security, identification, authentication, privacy, and confidentiality. Secure identification, authentication of the users, and communication security are main problems in networked systems.

Information management (IM) is a fuzzy term covering the various stages of information processing from production to storage and retrieval to dissemination toward the better working of an organization, where information can be from internal and external sources and in any format. The role of trust in these processes is definitive as human-to-human and human-to-system communications form the base of information management.

BACKGROUND

Definitions of Trust

The word trust is used by different disciplines, so there are many definitions of the term fulfilling the demands of the actual theory or application. In everyday life, without trust, one would be confronted with the extreme complexity of the world in every minute. No human being could stand this, so people have to have fixed points around them. One has to trust in family members, partners, trust in the institutions of a society and its members, and trust within and between organization partners. The diversity of approaches is one reason that trust has been called an "elusive concept to define" (Gambetta, 1988, p. 213).

Trust can be defined as a psychological condition comprising the truster's intention to accept vulnerability based upon positive expectations of the trustee's intentions or behavior (Rousseau, Sitkin, Burt, & Camerer, 1998). Those positive expectations are based upon the truster's cognitive and affective evaluations of the trustee and the system and world as well as of the disposition of the truster to trust. Trust is a psychological condition (interpreted in terms of expectation, attitude, willingness, perceived probability). Trust can cause or result from trusting behavior (e.g., cooperation, taking a risk) but is not behavior itself.

According to Luhman (1979), trust can be viewed as a cognitive and social device able to reduce complexity, enabling people to cope with the different levels of uncertainty and sometimes the risks that, at different degrees, permeate our life. Without trust, an individual would freeze in uncertainty and indecision when faced with the impossibility to calculate all possible outcomes of a situation. Engaging trust automatically can reduce the number of decision nodes that are being analyzed and facilitate the decision-making processes. From a social perspective, trust permits the necessary knowledge sharing of delegation and cooperative actions.

The following components are included in most definitions of trust (Harrison, McKnight, & Chervany, 1996):

- the willingness to be vulnerable and to rely,
- confidence and having positive expectations and attitudes toward others, and
- interdependence and taking risks as necessary conditions.

Trust has different forms such as the following.

1. **Intrapersonal trust:** trust in one's own abilities (self-confidence) and a basic trust in others
2. **Interpersonal trust:** expectation based on cognitive and affective evaluation of partners in primary relationships (e.g., family) and nonprimary relationships (e.g., business partners)
3. **System trust:** trust in depersonalised systems and a world that function independently (e.g., economic system, regulations, legal system, technology); requires voluntary abandonment of control and knowledge
4. **Object trust:** trust in nonsocial objects; trust in correct functioning (e.g., in an electronic device)

Trust is a Multifaceted Construct

There is compelling evidence originating from the organizational research community to support the idea that trust is a many sided, complex construct. McAllister (1995) has proposed two critical dimensions: emotional trust and cognitive trust. Emotional trust is the development of noncalculating and spontaneous emotional bonds, and affects two or more people. Emotional trust is demonstrated through confidence and openness in sharing ideas, feelings, and concerns. Cognitive trust refers both to judgments of competence (predictably professional behavior) and reliability (the congruence between words and actions) about the other members of a team.

Represented Forms of Trust

There are two basic modeling approaches in describing trust: the cognitive approach (Castelfranchi & Falcone, 1999) and the mathematical approach (Marsh, 1994). In the case of applying cognitive models, trust is made up of underlying beliefs, and trust is a function of the value of these beliefs. The mathematical modeling approach ignores the role of underlying beliefs and uses a trust metric based on variables like perceived competence, perceived risk, utility of a situation for the agent involved, importance of a situation, and so forth. These models incorporate some aspects of game theory and the evolution of cooperation models. Both modeling approaches see trust as a variable with a threshold for action. When the value of the variable crosses the threshold, the agent executes an action. In the Marsh model, the action is cooperation; in the Castelfranchi and Falcone model, the action is delegation. The action is Boolean in nature; the agent either delegates or not, or the agent either cooperates or not.

Classifying the Meanings of Trust

Harrison et al. (1996) made a very deep and thorough analysis of the word trust from many aspects in their working paper. The goal of the paper was to develop a classification system for the types of trust, and to develop trust definitions and types that can be accepted by most disciplines.

In the following, the main groups of the classification system for trust constructs are given to better understand the definitional problem. Impersonal and structural trust refers to those definitions of trust that differentiate it from being a property or state of a person or persons. Dispositional trust means that trust is based on the personality attributes of the trusting party. Personal and interpersonal trust means that one person trusts another person, persons, or thing(s) in the situation.

Guided by the classification system, six related types of trust have been defined in the working paper. The six constructs are as follows: trusting intention, trusting behavior, trusting beliefs, system trust, dispositional trust, and situational decision to trust. Both cognitive and affective components are included in trusting beliefs, trusting intention, and trusting behavior. The six constructs cover the more common of the dictionary definitions of trust. This multidimensional view of trust provides a parsimonious way to organize measurable trust types while clearly distinguishing one type from another.

BUILDING TRUST

Connection of Trust and Information Management

Information technology management deals with the management of the different steps of information processing, and trust has a role where human beings are involved in this process. Human beings basically have two types of connections in these processes: human-to-human relationships through networks and human-to-computer-system communication through interfaces. In the first case, trust management of virtual teams can be analyzed; in the second case, special effects of computer interfaces and the role of security technologies in trust building and maintenance can be studied (Mezgár & Kincses, 2002). Information management has to take into consideration the aspects of the trust-building process to develop, modify, and influence information handling into the direction that increases trust of human beings participating in these processes.

Trust is More Than a Simple Technology

In building trust there are two approaches: the information technology approach and human-centered approach based on culture and morality. The information technology approach means that security has to increase by different architectures, protocols, certifications, cryptography, authentication procedures, and standards, and this increased security generates the trust of users. This means access control (passwords, firewalls) protects the integrity and privacy of messages, databases (cryptography), and the identification of users. Stressing the effectiveness of these technologies for humans (users) can cause trust in the systems based on this convincing action. Based on the technological approach, 100% security can never be obtained (there will be always security holes somewhere in the systems), so full trust cannot be guaranteed based on these mechanisms.

The human side of trust is more complicated. There was different research (e.g., Hoffmann, Novak, & Peralta, 1999) focusing on this side of trust. From this aspect, the user interface has the main role, that is, the menu structure and the messages sent for the user by the system. In the case that the user feels that the interface is easy to use, it is transparent; he or she can control the system (even with low-level computer knowledge), that is, the system is "user friendly." Through this he or she can be convinced that this is a trustworthy system.

It would be a mistake to think that applying only the human-centered approach would generate trust; the technological part has to be added as well (e.g., biometrical identification), so mainly, the structured integration of the two approaches can result in the expected level of trust.

Technical Side of Trust Building: Application of Security Mechanisms

Security is a very complex term. There is computer, communication, information system, physical, and a lot of other types of security. As an addition, these terms are overlapping each other in a lot of cases. Approaching security from the side of trust, security is the set of different services, mechanisms, and software and hardware tools for generating trust with pure technology. More generally, security is a condition that results from the establishment and maintenance of protective measures that ensure a state of inviolability from hostile acts or influences (FED, 2003).

The totality of protection mechanisms within a computer system, including hardware, orgware, and software, is responsible for enforcing a security policy called trusted computing base (TCB). The ability of a trusted computing base to enforce correctly a unified security policy depends on the correctness of all types of mechanisms within the trusted computing base, the protection of those mechanisms to ensure their correctness, and the correct input of parameters related to the security policy.

In network management, the security management covers the set of functions (a) that protects telecommunications networks and systems from unauthorized access by persons, acts, or influences, and (b) that includes many subfunctions, such as creating, deleting, and controlling security services and mechanisms; distributing security-relevant information; reporting security-relevant events; controlling the distribution of cryptographic keying material; and authorizing subscriber access, rights, and privileges (FED, 2003; Tipton & Krause, 1998).

The building blocks, elements of security, are the security services and the security mechanisms. The security services are the following (Schneier, 1996).

- **Access control:** protects against unauthorized use
- **Authentication:** provides assurance of someone's identity
- **Confidentiality:** protects against disclosure to unauthorized identities
- **Integrity:** protects from unauthorized data alteration
- **Nonrepudiation:** protects against originator of communications later denying it

The means for achieving these properties depends on the collection of security mechanisms that supply security services, the correct implementation of these mechanisms, and how these mechanisms are used.

Security mechanisms with crypto functions are the following (three basic building blocks are used).

- Encryption is used to provide confidentiality, and also can provide authentication and integrity protection.
- Digital signatures are used to provide authentication, integrity protection, and nonrepudiation.
- Checksums or hash algorithms are used to provide integrity protection and can provide authentication.

Human Side of Trust Building: Feeling of Trust

The feeling of security experienced by a user of an interactive system does not depend on technical security measures alone. Other (psychological) factors can play a determining role; the user's feeling of control can be one of these factors.

Trust is a dynamic process and it alters based on experience. The trusting process begins when an individual perceives indications that suggest a person or organization may be worthy of trust. These indications can include behaviors such as manners, professionalism, and sensitivity, and these forms are designed to represent trustworthiness. These formal claims to trustworthiness become strengthened over time and are eventually transformed into "character traits," such as dependability, reliability, and honesty. The research study of Cheskin (1999) identifies six fundamental factors that communicate trustworthiness in case of e-commerce Web sites.

- **Brand:** the importance of the company's reputation in the choice to do business with them
- **Navigation:** the ease of finding what the user seeks
- **Fulfillment:** the process the user experiences from the time a purchase process is initiated until the purchase is received to the user's satisfaction
- **Presentation:** ways in which the appearance of the site, in and of itself, communicates meaningful information
- **Technology:** ways in which the site technically functions
- **Seals of approval:** symbols that represent the companies that specialize in assuring the safety of Web sites

It has to be analyzed why people feel safe and secure and what causes these feelings. The hypothesis of D'Hertefelt (2000, p. 1) was, "The feeling of security experienced by a user of an interactive system is determined by the user's feeling of control of the interactive system." The more a user feels in control of an interactive program, the more the user will trust the site or the program. An interactive system that generates the user to feel in control has to be (a) comprehensible (the client needs to know what and how he or she can accomplish, and needs confirmation that it actually has been accomplished), (b) predictable (user has to know what is going to happen when clicking on a button or menu item), and (c) flexible and adaptable (the user will feel in control if he or she can choose the way a task is executed instead of having to figure out how the system requires it to be done).

The process of building trust is slow; trust is formed gradually. It takes quite a lot of time and repeated positive experiences (Cheskin, 1999). Online trust can be described as a kind of human relationship. The initial stage is that of interest and distrust; there has to be a motivation, a need, to get interested in the service, or coworking. In subsequent phases the trust will evolve or, in case of negative experiences, the cooperation will terminate.

Trust depends on the time span of cooperation and the type of connection as well. It can be stated that there are differences in the trust building process in short-term and long-term relationships. In case of short-term relationships (e.g., in a virtual organization), trust must be achieved quickly and then be maintained with no or rare face-to-face interaction. The members of these teams must assume that other remote team members are trustworthy, and then later on modify their assumptions according their positive or negative experiences.

In long-term relationships there are four factors that influence trust building (Rocco, Finholt, Hofer, & Herbsleb, 2001).

- The expectation of future interaction may motivate greater investment in building trustworthy relationships.
- Long-term relationships offer more time to establish trustworthiness through routines and culture.
- People have more communication channels, which may affect trust to the extent that participants have additional ways to clarify misunderstandings or to correct failures.
- Participants are interested in successful task performance, and trust formation may assume a higher priority.

TRUST AND INFORMATION MANAGEMENT

Devices, Services and Technologies for Trust Building in Information Management

As security, one important base of trust involves many topics—not only information security technology, but also the structure of the firms; the management techniques also have to follow the ongoing changes of the IC technologies. In case the security challenges are put into a broader, management-oriented interpretation, they can be classified into four categories according to Dhillon (2001).

- Establishing good management practices in a geographically distributed environment and yet being able to control organizational operations
- Establishing security policies and procedures that adequately reflect the organizational context and new business processes
- Establishing correct structures of responsibility, given the complex structuring of organizations and information-processing activities
- Establishing appropriate information technology emergency recovery plans

A solution for addressing trust is to build a chain of trust where each link is strong but also connects to its neighbor by verifying its trustworthiness. In particular, beginning with a priori grounding in a physically trustworthy base, each link of the chain checks signature-based trust tokens of its neighbor before the chain is allowed to carry any weight. Such a chain begins with processors and extends to operating systems, to applications, to interconnected protocols, and ultimately, to end users.

Trust management unifies the notions of security policy, credentials, access control, and authorization. An application that uses a trust-management system can simply ask the compliance checker whether a requested action should be allowed. Trust-management policies are easy to distribute across networks, helping to avoid the need for application-specific distributed policy configuration mechanisms, access control lists, and certificate parsers and interpreters (Blaze, Feigenbaum, & Lacy, 1996).

Virtual Teamwork and Trust Management

Today the different types of networked organizations need new types of cooperation as the members of the working teams are geographically (physically) separated; they use shared documents and communicate through e-mail and high quality audio and video channels. These teams are called “virtual teams” as they never meet personally and have no face-to-face contact. The work of teams without face-to-face contact is less effective and reliable based on the observation stated by Handy (1995, p. 46): “trust needs touch.” According to case studies, it is evident that trust of virtual team members is significantly lower than trust in conventional teams (Rocco et al., 2001). In other experiments where interaction was primarily via e-mail, very similar results were gained, as in geographically distributed teams (Jarvenpaa & Leidner, 1999)

The latest research shows if people meet before using computer-mediated communication (CMC), they trust each other, as trust is being established through touch. In case participants do not meet formerly but initiate various getting-acquainted activities over a network, trust is much higher than if they do nothing before, nearly as good as a prior meeting. Using chat forums to get acquainted is nearly as good as meeting, and “even just seeing a picture is better than nothing” (Zheng, Veinott, Bos, Olson, Gary, & Olson, 2002, p. 141).

Information Technology Tools for Generating Trust

Identification of a user or customer of an information system is a complex task. In computer science, the identifier is a string of bits or characters that names an entity, such as a program, device, or system, so that other entities can call that entity. In the context of information systems, the purpose of identification is very concrete: It is used to link a stream of data with a certain person, so the following definition can be given, “Human identification is the association of data with a particular human being” (Clarke, 1994, p. 1).

Information systems have tended to use codes rather than names as the primary identification mechanism. The most reliable mode to identify a person is to apply biometric techniques (Jain, Bolle, & Pankanti, 1999). The applied techniques in biometry systems in IT includes a physiological (fingerprint, iris, facial features, etc.) element or factor as well as a behavioral one (e.g., vocal patterns, typing rhythm). Biometric identification is preferred over current methods involving passwords and PINs (personal identification numbers) as the person to be identified is required to be physically present at the point of identification, so the person of the user is identified, not the device as in the case of PIN and password. The layered biometric verification (LBV) technology entails layering different biometric technologies into a complex identification process.

The main factor of trust is confidentiality that can be achieved by technologies that convert or hide the data or text into a form that cannot be interpreted by unauthorized persons. There are two major techniques to fulfill this goal: encryption and steganography. Encryption is transforming the message into a ciphertext, such that an enemy who monitors the ciphertext cannot determine the message sent (Schneier, 1996). Steganography is the art of hiding a secret message within a larger one in such a way that the opponent cannot discern the presence or contents of the hidden message (Johnson, Neil, Duric, & Jajodia, 2000). For example, a message might be hidden within a picture by changing the low-order pixel bits to be the message bits.

Generating Trust by Human-Computer Interfaces

Generally speaking, the goal of an interface is to interconnect two or more entities at a common point or shared boundary. As a communication- or information-system term, an interface is the point of communication between two or more processes, persons, or other physical entities. Interfaces are the key points for gaining the trust of the user or customer. They are the first connection points between the user and the system; identification of the users takes place at these points (e.g., password input, fingerprint reader, smart-card reader), so they have to be designed very carefully. In the design of both types of interfaces, ergonomic and psychological aspects are taken into consideration besides the technical ones.

Trusted Time

Trusted time is important in global network communication. Trusted time is essential to ensuring that the network and operations infrastructure run accurately, securely, and reliably, and that all relying applications execute with traceability that provides auditability to resolve disputes, avoid repudiation, and ensure legal enforceability of electronic business transactions. The auditable sourcing and synchronization of time involves a chain of trust.

FUTURE TRENDS IN INFORMATION TECHNOLOGY MANAGEMENT

Nearly all types of systems in all fields of the economy have become distributed and virtual structures appeared. The result is big structural and cultural changes in enterprises. A new communication technology appeared as well: the wireless technology. The management of these new systems has brought new aspects into focus also in the field of information management. The main advantages of wireless and mobile communication are that anybody from anywhere at anytime can make contacts. According to market researcher Gartner, 45% of the American workforce is using mobile technology of some kind, including laptops, PDAs (personal digital assistants), and new sensor networks (Gartner Symposium, 2003). By 2007, more than 50% of enterprises with more than 1,000 employees will make use of at least five wireless networking technologies.

Mobile devices became far more popular than it was estimated before, and thanks to the extremely fast development of electronic industry, these devices have grown into multifunctional tools. Mobile Internet rewrites many rules. People everywhere, in different positions, are using new services that are relevant to their personal needs and preferences and are accessible at anytime, anywhere. New terms are developing in all sectors of the industry, in finance, in the government, and in the society as well.

Information architectures, structures, business processes, and business models of enterprises have to be modified according to new information communication technology. Mobile technologies add new value in being continuously connected. Short response times assure the validity of information. Productivity is no longer restricted by place or time. Best of all, it is possible to experience new ways of communicating and sharing information. In an information- and communication-centered world, security and trust are exceptionally important as the value and strategic role of reliable information is extremely high. Information technology management systems and the managers themselves have to adapt to this fast, nonstop, changing environment. Smart cards, personal trusted devices, and secure agent architectures (Wong & Sycara, 1999) are some of the main elements in information technology that will change in the future.

CONCLUSION

The chapter has shortly introduced the main characteristics of trust and its connections between security services and mechanisms, and its role in information technology management. The importance of trust is increasing very fast as the main characteristic of the information and knowledge society is cooperation through computer and communication networks. As it is pointed out by different analyses based on real-life statistics, when users do not trust a system or service, they do not use it. The organizations have to adapt themselves to this requirement even by changing their culture or organization structures as well.

Integrated mobile technologies are speeding up this tendency as they offer mobility and freedom for citizens. Distributed information systems with different sizes are playing a definite role, but originating from their openness and flexibility, the information systems will always be a security risk. As absolute (100%) security and trust cannot be reached with pure information technology approaches, they have to be integrated with the human-centered techniques based on psychological methods.

There is a need for complex, flexible security systems that are user friendly and platform independent at the same time. The developments of hardware and software elements of such systems are going on and the potential users have to get acquainted with them. The managers of information technology have to adapt these technologies, tools, and devices into their systems to provide a high security and trust level that can induce trust in all humans involved in the different phases of the life cycle of information systems.

REFERENCES

Blaze, M., Feigenbaum, J., & Lacy, J. (1996). Decentralized trust management. *Proceedings of the 17th Symposium on Security and Privacy*, 164-173.

Castelfranchi, C., & Falcone, R. (1999). The dynamics of trust: From beliefs to action. *Proceedings of the Autonomous Agents Workshop on Deception, Fraud and Trust in Agent Societies*.

Castelfranchi, C., & Yao-Hua Tan. (Eds.). (2001). *Trust and deception in virtual societies*. Kluwer Academic Publishers.

Cheskin. (1999). *eCommerce trust: A joint research study with Studio Archetype/Sapient and Cheskin*. Retrieved from *http://www.cheskin.com/p/ar.asp?mlid=7&arid=10&art=0*

Clarke, R. (1994). Human Identification in Information Systems: Management Challenges and Public Policy Issues, *http://www.anu.edu.au/people/Roger.Clarke/DV/HumanID*

D'Hertefelt, S. (2000). *Trust and the perception of security*. Retrieved from http://www.interactionarchitect.com/research/report20000103shd.htm

Dhillon, G. (2001). *Information security management: Global challenges in the new millennium*. Hershey, PA: Idea Group Publishing.

FED. (2003). *Telecommunications: Glossary of telecommunication terms*. Retrieved from *http://glossary.its.bldrdoc.gov/fs-1037/*

Fukuyama, F. (1995). *Trust: The social virtues and the creation of prosperity*. New York: The Free Press.

Gambetta, D. (1988). *Can we trust trust*? In D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations* (pp. 213-237). Oxford, UK: Basil Blackwell.

Gartner Symposium/Itxpo. (2003). Lake Buena Vista, Florida, USA, 19-24, October.

Handy, C. (1995). Trust and the virtual organization. *Harvard Business Review*, *73*(3), 40-50.

Harrison, D., McKnight, N., & Chervany, L. (1996). *The meanings of trust* [Working paper]. University of Minnesota, Management Information Systems Research Center (MISRC).

Hoffman, D. L., Novak, T. P., & Peralta, M. (1999). Building consumer trust online. *CACM, 42*(4), 80-85.

Jain, A., Bolle, R., & Pankanti, S. (1999). *Biometrics: Personal identification in networked society*. Kluwer Academic Publishers.

Jarvenpaa, S. L., & Leidner, D. E. (1999). Communication and trust in global virtual teams. *Organization Science*, *10*(6), 791-815.

Johnson, Neil, F., Duric, Z., & Jajodia, S. (2000). *Information hiding: Steganography and watermarking—Attacks and countermeasures*. Kluwer Academic Publishers.

Luhman, N. (1979). *Trust and power*. Chichester: Wiley.

Marsh, S. (1994). *Formalising trust as a computational concept*. PhD thesis, University of Stirling, Stirling, Scotland.

McAllister, D. J. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal, 38*(1), 24-59.

Mezgár, I., & Kincses, Z. (2002). The role of trust in information management. In A. Gunasekaran (Ed.), *Knowledge and information technology management in the 21st century organizations: Human and social perspectives* (pp. 283-304). Hershey, PA: Idea Group Publishing.

Rocco, E., Finholt, T. A., Hofer, E. C., & Herbsleb, J. D. (2001, April). *Out of sight, short of trust.* Presentation at the founding conference of the European Academy of Management, Barcelona, Spain.

Rousseau, D. M., Sitkin, S. B., Burt, R., & Camerer, C. (1998). Not so different after all: A cross-disciplinary view of trust. *Academy of Management Review, 23*, 1-12.

Schneier, B. (1996). *Applied cryptography*. John Wiley & Sons.

Tipton, H., & Krause, M. (Eds.). (1998). *Handbook of information security management.* CRC Press LLC.

Wong, H. C., & Sycara, K. (1999). Adding security and trust to multi-agent systems. *Proceedings of Autonomous Agents '99* (Workshop on Deception, Fraud and Trust in Agent Societies), 149-161.

Zheng, J., Veinott, E., Bos, N., Olson, J. S., Gary, Olson, G. M. (2002). Trust without touch: Jumpstarting long-distance trust with initial social activities. *Proceedings of the SIGCHI Conference on Human factors in Computing Systems: Changing our World, Changing Ourselves*, 141-146.

KEY TERMS

Biometry (synonym: biometrics): Generally, biometrics refers to the study of measurable biological characteristics. In computer security, biometric technologies are defined as automated methods of identifying or authenticating the identity of a living person based on his or her physiological (e.g., fingerprint, hand, ear, face, eye [iris or retina]) or behavioral (e.g., signature, voice, keystroke) characteristic. This method of identification is preferred over current methods involving passwords and PINs as the person to be identified is required to be physically present at the point of identification, so the person of the user is identified, not the device as in the case of PIN and password.

Encryption: Encryption is the transformation of plaintext into an apparently less readable form (called ciphertext) through a mathematical process. The ciphertext may be read by anyone who has the key that decrypts (undoes the encryption of) the ciphertext.

Layered Biometric System: Multilayer, layered, and multimodal biometric systems combine more than one physiological or behavioral characteristic for verification or identification. By promoting multilayered identification and authentication—that means parallel use of strong passwords, smart tokens, and biometrics—many significant security problems can be eliminated. The combination of multiple biometric methods such as voice and fingerprints, in conjunction with, for example, digital certificates and smart cards, offer the companies an ideal solution to provide a very high level of protection for their sensitive information.

Personal Trusted Device: People like smart, little equipment, tools that they can keep in their handsand bring permanently with themselves, so they can control them both physically and in time. According to the concept of the personal trusted device, they have to be personal, always carried by the user, small, cheap, battery powered, secure as a smart card, and they have a common user interface. Mobile phones can fulfill the role of a personal trusted device as mobile phones are well placed as identity tokens; they have dynamic authentication already proven in GSM, mass market, and secure communications. Mobile phones are (the only) mass-market smart card readers and they are highly personal. Users are usually very attached to their phones, and they can be made more personal by use of PIN or (later) biometrics.

Steganography: Steganography (literally meaning *covered writing*) dates back to ancient Greece. It includes a vast array of methods of secret communications that conceal the very existence of the message. In today's computer security, steganography is the science of hiding information by embedding messages within other seemingly harmless messages. Steganography works by replacing bits of useless or unused data in regular computer files (such as graphics, sound, text, HTML [hypertext markup language]) with bits of different, invisible information. This hidden information can be plaintext, ciphertext, or even images. Unlike encryption, steganography cannot be detected. Therefore, steganography is used to supplement encryption. An encrypted file may still hide information using steganography, so even if the encrypted file is deciphered, the hidden message is not seen.

Trust: Trust can be viewed as a cognitive and social device able to reduce complexity, enabling people to cope with the different levels of uncertainty and sometimes the risks that, at different degrees, permeate our life. Without trust, an individual would freeze in uncertainty and indecision when faced with the impossibility to calculate all possible outcomes of a situation. From a social perspective, trust permits the necessary knowledge sharing of delegation and cooperative actions (Luhman, 1979).

Trust Chain: A solution for addressing trust is to build a chain of trust where each link is strong but also connects to its neighbor by verifying its trustworthiness.

Trust Management: Trust management is a unified approach to specifying and interpreting security policies, credentials, and relationships; it allows direct authorization of security-critical actions. A trust-management system provides standard, general-purpose mechanisms for specifying application security policies and credentials. Trust-management credentials describe a specific delegation of trust and subsume the role of public key certificates; unlike traditional certificates, which bind keys to names, credentials can bind keys directly to the authorization to perform specific tasks.

Trusted Time: Trusted time is emerging as the industry's best practice. Trusted time is auditable, secure, available, warranted, accurate, and managed. Trusted time is essential to ensuring that the network and operations infrastructure run accurately, securely, and reliably, and ensure legal enforceability of electronic business transactions. The auditable sourcing, setting, and synchronization of time involve a "chain of trust" that is complex, requires time specialization, and involves multiple sources of risk.

Trustworthiness: Trustworthiness is the ability to attain and maintain a "trusted state," which is definable, measurable, validated, and demonstrable over time. Digital trustworthiness is a verifiable level of electronic process integrity, security, control, authenticity, and reliability that captures, preserves, retrieves, verifies, renders, and makes available in human readable form the essential transaction content, context, notice, intent, and consent to meet the electronic forensic-evidence requirements necessary for legal admissibility and regulatory compliance.

B

Building Educational Technology Partnerships through Participatory Design

John M. Carroll
The Pennsylvania State University, USA

INTRODUCTION

Educational technology provides many examples of how efficient software development and deployment is not enough. Teachers work in a complex and dynamic context in which measurable objectives and underlying values collide on a daily basis. Traditionally, teachers work in isolation from their peers; individual teachers have well-established personal practices and philosophies of education. Teachers have enormous discretion with respect to what goes on in their classrooms, yet are also routinely interrogated by supervisors, by parents and other community members, and by educational bureaucracies. This has led to an abiding tension in the culture of schools: Teachers' innovative practices are often not adequately acknowledged or valued, and at the same time, teachers often passively resist school reforms that are imposed top-down.

Technology is a particularly problematic element in the culture of schools. The isolation and discretion of the teacher's work environment requires that technology for classroom use be highly appropriate and reliable. Yet it is generally assumed that teachers are to be *trained* on new technologies, not asked to *define* what those technologies should be. From the teacher's standpoint, classroom technology often is itself the problem, not the solution. This culture of technologydevelopment in the schools has been singularly ineffective—film and radio in the 1920s, television in the 1950s, and computer-assisted instruction in the 1980s, among others, have been notable failures (Tyack & Cuban, 1995).

An alternative to merely efficient technology development is *participatory design*, the inclusion of users within a development team such that they actively help in setting design goals and planning prototypes. This approach was pioneered, and has been widely employed, in Europe since the 1970s, and now consists of a well-articulated and differentiated set of engineering methods in use worldwide (Carroll, 2000; Clement & Van den Besselaar, 1993; Muller, 2003; Muller, Haslwanter, & Dayton, 1997; Rosson & Carroll, 2002).

In 1994, a design collaboration was formed between Virginia Tech and the public schools of Montgomery County, Virginia. The objective was to develop and investigate a high-quality communications infrastructure to support collaborative science learning. Montgomery County is located in the rural Appalachian region of southwestern Virginia. In March 2000, one of its high schools was listed among the top 100 in the US by *Newsweek* magazine. However, in others, physics is only offered every other year and to classes of only three to five students. The initial vision was to give students in this diverse and dispersed school district access to peers through networked collaboration.

We felt it was critical for the teachers to contribute as collaborators in design analysis, implementation, deployment, testing, and refinement, and as leaders in the development of courseware and classroom activities that would exploit the software. For a classroom-technology partnership to succeed, the university researchers must eventually fade and leave the teachers to maintain and develop its achievements. In the end, the technology-development goals of this project were achieved, though this is not the topic of this paper (Isenhour, Carroll, Neale, Rosson, & Dunlap, 2000).

BACKGROUND

We analyzed our participatory engagement with the teachers as "developmental" in the sense of Piaget and Inhelder (1969) and Vygotsky (1978). We believe the teachers developed qualitatively different roles through the course of our collaboration. In some cases, these roles were suggested to them; in other cases, they defined and claimed new roles. But in all cases, these transitions exemplified the defining characteristics of *developmental change*: active resolution of manifest conflicts in one's activity, taking more responsibility, and assuming a greater scope of action. Each successive stage can be seen as a relatively stable organization of knowledge, skills, and attitudes that resolves the instigating conflict.

During the six years of this project, we distinguished four stages in our collaboration with the teachers. At first, the teachers were *practitioner-informants*; we observed their classroom practices and we interviewed them. Subsequently, the teachers became directly and actively involved in the requirements-development process as *analysts*. Later, the teachers assumed responsibility as *designers* for key aspects of the project. Finally, the teachers became *coaches* to their own colleagues within the public school system.

In a classic Piagetian example, a child in the preoperational stage perceives single dimensions of quantity. This produces conflicts: A given quantity of liquid poured from a short, wide container into a tall, thin container appears suddenly to be more, but of course cannot be more. These conflicts eventually precipitate a cognitive reorganization called the concrete operational stage, in which constant quantities are perceived as constant regardless of varying shapes and arrangements.

Developmental change in adults is of course more complex. The stages we describe are not singular competencies, but relatively complex ensembles of collaboration, social norms, tool manipulation, domain-specific goals and heuristics, problem solving, and reflection in action. They are social constructions achieved through enculturation, constituted by the appropriation of the artifacts and practices of a community (Vygotsky, 1978).

In the Piagetian notion of stages in child development, successive stages build upon the cognitive structures and enabled activity of prior stages, but ultimately replace those structures. A child who enters the concrete operational stage can no longer function at the preoperational stage. Adult growth, however, is not static achievement, but continual elaboration. The teachers are still practitioners whose classroom practices we regularly observe and whose classroom expertise we still interrogate; they seem to us and to themselves to be representative practitioner-informants. However, they are now *also* analysts and designers, and often coaches. Indeed, effective design coaches probably must be experienced designers, successful designers must be skilled analysts, and analysts must have attained significant domain knowledge (Carroll, Chin, Rosson, & Neale, 2000).

MAIN THRUST OF THE CHAPTER

Developmental theory explains transitions between stages as resolutions of conflict. Thus, the preoperational child's conflicting perceptions of quantity based on single dimensions, such as height and width, are resolved in the abstraction of quantity as an invariant in the concrete operational stage. For development to take place, the child must have attained the requisite competencies to experience the triggering conflict, and then be able to reconceptualize the situation in such a way that the conflict dissolves.

This analytical schema seems to fit the transitions between the stages of cooperation we identified. The general mechanism appears to be that successive increases in knowledge, skill, and confidence empowered the teachers to resolve conflicts by assuming successively greater scope of action and responsibility in the project. Early on, the teachers faced the conflict that their pedagogical concerns and perspectives would be adequately represented and fully considered by the group only if they themselves championed those concerns. This went beyond the practitioner-informant role they had played in the project up to then. But they were both motivated and competent to resolve this conflict by assuming the analyst role in the project.

Once the teachers were functioning as analysts in the project team, further conflicts and resolutions arose. The teachers experienced a conflict between their own analyses of system requirements and the current state of our project software and development plans. They resolved these conflicts by formulating their own design proposals, ultimately a radical reorientation of the project's vision of classroom activity. They became designers. Subsequently, the teachers recognized that they were the best qualified project members to train new teachers and to pursue specific curricular extensions of the project. They became coaches.

The teachers' behavior also reflects development *within* the four general stages we have described. For example, cognitive scaffolding (via examples, reflective prompts) was needed to engage the teachers in the novel and relatively abstract activity of design analysis. But as the project progressed, teachers spontaneously identified and presented design trade-offs to the group as a way to articulate personal positions. This is consonant with the general notion of learning as movement through a zone of proximal development (Vygotsky, 1978).

The designer stage also reflects several different levels of development. Initially, the teachers were able to collaborate with a research assistant in focused design sessions, cowriting scenarios of technology-mediated activities for their classroom. Later they banded together as a subgroup, pooling their goals and expertise to develop a scenario that specified a new vision of collaborative learning activities. Ultimately, each learned to function as an independent designer, envisioning and specifying activities optimized for their own teaching styles, objectives, and classroom environments. In their coach role, the teachers also worked first as a group, but subsequently recruited and mentored colleagues in a one-to-one fashion.

In sum, it appears that the transitions among stages were triggered by conflicts with respect to the teachers' role in the project. In each case, a series of scaffolded activities enabled them to attain the knowledge, skill, and confidence that led them to expand their role (Carroll et al., 2000).

FUTURE TRENDS

We originally committed to a long-term participatory-design method because we conjectured that such an approach would be crucial for success in this educational technology setting. We believe we could not have succeeded to the extent we have had we not made this commitment. Working from the national agenda for school reform, educational technology, and science education (AAAS, 1993; NTSA, 1992), and from our own initial vision of a "virtual school," we would have built the wrong system—we would not have had effective support from teachers. Little or nothing would have been sustained after the initial project funding ended.

Participatory design is fundamentally a process of mutual learning, and thus of personal development for participants. But it is often exemplified by rather singular and ephemeral learning interactions. Our study expands the scope of the design participants' personal development by examining a case of long-term cooperative design interaction, and by describing a developmental sequence of roles with constituent capacities and responsibilities.

Much research on participatory design has focused on relatively short-term collaborative relationships. This is especially true in North America, where many participatory-design techniques are directed at brief user-interface-design interactions of perhaps one hour (Muller et al., 1997). Such methods are both effective and democratic, but it seems unlikely that the experience of manipulating a user-interface mock-up during a brief participatory session can have a significant developmental effect on a person's knowledge, skills, self-confidence, or other professional capacities.

In our project, user-interface design per se was a secondary issue. We used brief participatory exercises, but this level of engagement was more a starting point than the objective of our work. More specifically, we wanted the teachers to have a significant voice in designing the functionality and the use of the virtual school, not merely its appearance. We needed to learn about pedagogical goals and practices, classroom management, school-system politics, the relationship of the community and the schools, and so forth.

Where participatory-design investigations *have* focused on longer term interactions, chiefly in Europe, these often involve extremely well-organized user groups with well-defined roles and prerogatives in the design process. In many cases, the users are represented by labor unions whose personnel provide legal representation of user interests in the design process. In these cases, there is sometimes a clear demarcation, even conflict, between the user (union) interests and management's technology strategy. Indeed, this is an important element of the context for many of these studies. Because the user role in many of these studies is both specified a priori and representative (versus individual), the personal development of user-designers is not a central issue. These case studies also typically involve situations in which the development and deployment of new information technology is a given, and the challenge is to define appropriate technology for the users and their activities (Bjerknes & Bratteteig, 1993; Bødker, Ehn, Kammersgaard, Kyng, & Sundblad, 1987; Merkel et al., 2004).

In the educational domain, the deployment of new information technology is far from given. Indeed, the introduction of new technology has historically almost always failed in school settings. One of the key questions for us was whether a concept for appropriate technological support could be developed at all.

The users in our domain are very loosely organized. As mentioned earlier, teachers traditionally work in isolation from peers; they manage their own work practices and environments (classrooms). The notion of a "user community" in this domain is almost ironic. Teachers unions in the US are also extremely weak and play no role in the introduction of classroom technology. Indeed, school administrations in the US rarely have technology strategies at all. Thus, unlike the European case studies, the issue is almost never one of recognized conflict, but rather finding a direction at all.

The teachers in our team do not represent other teachers; they are individuals who, as members of our team, have become teacher-designers. This is precisely why their personal development as designers is a central issue in our study. Of course, we do hope that they are representative teachers—allowing us to generalize our investigation to other teachers participating in similar development projects—but this is a separate issue. The point is that in our project, and unlike many long-term participatory-design efforts in Europe, the teachers act as individual professionals just as university researchers do.

CONCLUSION

The stages we have described here are specific to our project; they emerged through specific things that we did and are rooted in the specific goals of our project. At the same time, they suggest a schematic programme for developing cooperative engagement more generally. Most participatory-design work engages users at the practitioner-informant stage. This would seem to be an obvious and general starting point for any participatory-design collaboration. In our project, the teachers transitioned to the analyst stage through their inclusion in a requirements-analysis workshop and a significant process of iterative requirements development (Carroll, Rosson, Chin, & Koenemann, 1998). This is perhaps not typical of participatory-design practice, but it is a modest extension. Nevertheless, the teachers found it quite stimulating to be invited to objectify their own experience, to dissect it and not merely describe it.

ACKNOWLEDGEMENTS

This work was partially supported by the Hitachi Foundation and the National Science Foundation (REC-9554206 and DGE-9553458). This paper is a summary of Carroll, Chin, Rosson, Neale, Dunlap, and Isenhour (2002).

REFERENCES

American Association for the Advancement of Science (AAAS). (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

Bjerknes, G., & Bratteteig, T. (1995). User participation and democracy: A discussion of Scandinavian research on system development. *Scandinavian Journal of Information Systems, 7*(1), 73-98.

Bødker, S., Ehn, P., Kammersgaard, J., Kyng, M., & Sundblad, Y. (1987). A utopian experience. In G. Bjerknes, P. Ehn, & M. Kyng (Eds.), *Computers and democracy: A Scandinavian challenge* (pp. 251-278). Brookfield, VT: Avebury.

Carroll, J. M. (2000). *Making use: Scenario-based design of human-computer interactions.* Cambridge, MA: MIT Press.

Carroll, J. M., Chin, G., Rosson, M. B., & Neale, D. C. (2000). The development of cooperation: Five years of participatory design in the virtual school. In D. Boyarski & W. Kellogg (Eds.), *DIS'2000: Designing interactive systems* (pp. 239-251). New York: Association for Computing Machinery.

Carroll, J. M., Chin, G., Rosson, M. B., Neale, D. C., Dunlap, D. R., & Isenhour, P. L. (2002). Building educational technology partnerships through participatory design. In J. Lazar (Ed.), *Managing IT/community partnerships in the 21st century* (pp. 88-115). Hershey, PA: Idea Group Publishing.

Carroll, J. M., Rosson, M. B., Chin, G., & Koenemann, J. (1998). Requirements development in scenario-based design. *IEEE Transactions on Software Engineering, 24*(12), 1-15.

Clement, A., & Van den Besselaar, P. (1993). A retrospective look at PD projects. *Communications of the ACM, 36*(4), 29-37.

Isenhour, P. L., Carroll, J. M., Neale, D. C., Rosson, M. B., & Dunlap, D. R. (2000). The virtual school: An integrated collaborative environment for the classroom. *Educational Technology and Society, 3*(3), 74-86.

Merkel, C. B., Xiao, L., Farooq, U., Ganoe, C. H., Lee, R., Carroll, J. M., et al. (2004). Participatory design in community computing contexts: Tales from the field. In *Artful integration: Interweaving media, materials and practices-Proceedings of the 2004 participatory design conference.* Palo Alto, CA: Computer Professionals for Social Responsibility.

Muller, M. J. (2003). Participatory design: The third space in HCI. In J. Jacko & A. Sears (Eds.), *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications.* Mahwah, NJ: Erlbaum.

Muller, M. J., Haslwanter, J. H., & Dayton, T. (1997). Participatory practices in the software lifecycle. In M. Helander, T. K. Landauer, & P. Prabhu (Eds.), *Handbook of human-computer interaction* (2nd ed., pp. 255-297). Amsterdam: Elsevier.

National Science Teachers Association (NTSA). (1992). *Scope, sequence and coordination of secondary school science: Vol. 1, The content core*. Washington, DC.

Piaget, J., & Inhelder, B. (1969). *The psychology of the child*. New York: Basic Books.

Rosson, M. B., & Carroll, J. M. (2002). *Usability engineering: Scenario-based development of human-computer interaction.* San Francisco: Morgan Kaufmann.

Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform*. Cambridge, MA: Harvard University Press.

Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

KEY TERMS

Analysts: Users who collaborate with designers as domain experts analyzing constraints and trade-offs in existing and envisioned work practices are called *analysts*. This is the second stage in the developmental theory of participatory-design relationships between users and designers.

Coaches: Users who help other users participate in design work by coaching them are called *coaches*. This is the fourth stage in the developmental theory of participatory-design relationships between users and designers.

Designers: Users who collaborate with designers as domain experts envisioning new work practices and tools are called *designers*. This is the third stage in the developmental theory of participatory-design relationships between users and designers.

Developmental Theory: Theories of learning that involve growth and other qualitative changes in skills, knowledge, and capacities. Developmental theory is contrasted to accretive theories in which learning is conceived of as a matter of quantitative improvements—more knowledge or faster performance.

Educational Technology: Technology used in formal educational contexts, such as classrooms. Recent examples are television, personal computers, and the Internet.

Participatory Design: Design methods in which users (and other stakeholders) provide special expertise and play active and autonomous roles in design work.

Practitioner-Informants: Users who collaborate with designers as domain experts providing information about work practices are called *practitioner-informants*. This is the initial stage in the developmental theory of participatory-design relationships between users and designers.

Building Local Capacity via Scaleable Web-Based Services

Helen Thompson
University of Ballarat, Australia

INTRODUCTION

Information communications technology (ICT) has been identified as a key enabler in the achievement of regional and rural success, particularly in terms of economic and business development. The potential of achieving equity of service through improved communications infrastructure and enhanced access to government, health, education, and other services has been identified. ICT has also been linked to the aspiration of community empowerment, where dimensions include revitalizing a sense of community, building regional capacity, enhancing democracy, and increasing social capital.

In Australia, there has been a vision for online services to be used to open up regional communities to the rest of the world. Government support has been seen "as enhancing the competence levels of local economies and communities so they become strong enough to deal equitably in an increasingly open marketplace" (McGrath & More, 2002, p. 40). In a regional and rural context, the availability of practical assistance is often limited. Identification of the most appropriate online services for a particular community is sometimes difficult (Ashford, 1999; Papandrea & Wade, 2000; Pattulock & Albury Wodonga Area Consultative Committee, 2000). Calls, however, continue for regional communities to join the globalized, online world. These are supported by the view that success today is based less and less on natural resource wealth, labor costs, and relative exchange rates, and more and more on individual knowledge, skills, and innovation. But how can regional communities "grab their share of this wealth" and use it to strengthen local communities (Simpson 1999, p. 6)? Should communities be moving, as Porter (2001, p. 18) recommends (for business), away from the rhetoric about "Internet industries," "e-business strategies," and the "new economy," to see the Internet as "an enabling technology—a powerful set of tools that can be used, wisely or unwisely, in almost any industry and as part of almost any strategy?"

Recent Australian literature (particularly government literature) does indeed demonstrate somewhat of a shift in terms of the expectations of ICT and e-commerce (National Office for the Information Economy, 2001; Multimedia Victoria, 2002; National Office for the Information Economy, 2002). Consistent with reflections on international industry experience, there is now a greater emphasis on identifying locally appropriate initiatives, exploring opportunities for improving existing communication and service quality, and for using the Internet and ICT to support more efficient community processes and relationships (Hunter, 1999; Municipal Association of Victoria and ETC Electronic Trading Concepts Pty Ltd., 2000; National Office for the Information Economy, 2002).

The objective of this article is to explore whether well-developed and well-implemented online services can make a positive contribution to the future of regional and rural communities. This will be achieved by disseminating some of the learning from the implementation of the MainStreet Regional Portal project (www.mainstreet.net.au). To provide a context for this case study, the next section introduces some theory relevant to virtual communities and portals. The concept of *online communities* is introduced and then literature is reviewed to identify factors that have been acknowledged as important in the success of online community and portal initiatives.

BACKGROUND

In regional Australia, many Web-based initiatives have been premised on fear of external electronic commerce ventures adversely affecting local industry (McGrath & More, 2002, p. 50). Media and government reports have reinforced notions that those who ignore the adoption of electronic commerce will do so at their peril (Department of Communications Information Technology and the Arts, 2000). Recent research however identifies a movement beyond the "starry-eyed fascination with, and high expectations of, technology per se," with the focus now more pragmatically on how ICT can enable enhanced business and community processes and more effective organizational relationships (More & McGrath, 2003).

The term *online community* means different things to different people (Preece, 2000). In early definitions, the term described communication facilitated through bulletin boards (Rheingold, 1994, pp. 57-58). More recent definitions reflect the expansion of Web-based technologies and often link online communities with concepts of regional communities and local strengths (Keeble & Loader, 2001).

In Australia the terms *online community, regional portal, Web portal,* and *community portal* are often used more or less interchangeably. Web portals "provide focal points on the Web, places to start, places to go to find things" (Gronlund, 2001, p. 88). They have been identified as one strategy for encouraging regional participation in the information economy. For example, according to the Department of Communications Information Technology and the Arts (2001), a regional portal can achieve the online aggregation of potential and existing regional presence into a comprehensive portal, gateway, or regional Web site. In funding initiatives, preference has been given to projects that offer inclusive regional aggregation of business, government, and community services, and which provide interactive services to clients both in and external to the region.

Some definitions of online communities capture the concepts of both *communities of interest* and *communities of location,* and identify the role of encouraging communication and information sharing among members as important (McGrath & More, 2002). Australia's largest telecommunications provider describes online communities as providing a focal point for the provision of local regional information. In terms of functionality, these community portals generally incorporate local news services, local weather reports, a directory of community organizations, and features such as bulletin boards, discussion forums, a calendar of events, and transaction services (Telstra Country Wide, 2002).

To achieve optimum online collaboration, various issues require consideration. These include notions of community, trust and commitment, processes and structure, knowledge management, learning, and collaboration (More & McGrath, 2003). Some further factors more specific to the success of online community or portal initiatives are considered in the next section.

In forging and managing online collaboration, people issues rather than technological ones have been identified as the most challenging. "Certainly across a broad range of projects, many have come to realize that managing people, relationships, and business processes is harder than managing technology" (McGrath & More, 2002, p. 66). It is easy to underestimate the amount of planning and effort that is needed to build and sustain an online community; therefore care should be taken to avoid miscalculations. In particular, "overlooking the key role of the human facilitator is perhaps the greatest reason that online communities fail to meet the expectations of their designers" (Bernal, 2000, p. 4).

For many projects, collaboration is the key to survival, renewal, and growth, especially in regional areas "where the threat of global competitive dynamics often drove alliances" (McGrath & More, 2002, p. 67). Initiatives, however, with a broad geographical focus, can "encounter difficulties in establishing and maintaining cooperative relationships across multiple communities in their regions" (Simpson, 2002, p. 8).

"Many projects that have adopted a 'build it and they will come' approach have been doomed to early failure" (Simpson, 2002, p. 4). Developers need to work with community members to ensure that the goals of the site owner and the needs of community members are met (Preece, 2000). Good online services provide multiple levels of entry, many-to-many relationships, and rapid movement between the services and content of disparate providers (Local Government Association of Tasmania and Trinitas Pty Ltd., 2001).

Community members also need compelling reasons to use and return to an online community again and again. There will be a need to balance supply-side investment (access, technical platforms) and demand-side investment (content and services) (Local Government Association of Tasmania and Trinitas Pty Ltd., 2001).

"If you get this right—if you can identify and fill a need in the lives of your community members—you can go a long way on very little technology. If you miss this, no amount of technology is going to make you successful as an online community." (Kim, cited in Bernal, 2000, p. 3)

Engaging and relevant content are vital to increase uptake and sustained use of the Internet. Portal content management strategies should be *bottom-up* in their approach. This can be achieved by providing multiple opportunities for interaction and by providing permission-based access to software that allows members to produce content for their online community (Brumby, 2001; Telstra Country Wide, 2002).

Soft technologies are also essential in building user confidence and comfort with new technology. "Individualized awareness raising…training activities, and learner support are key elements in creating within the community the desire, motivation, and enthusiasm to trial and take up the technology" (Simpson, 2002, p. 7).

This review has highlighted a number of factors which can impact the success or otherwise of portal type initiatives. This background information provides a context for introducing the MainStreet case study in the next section.

MAIN THRUST OF ARTICLE

In May 1999 a collective of regional stakeholder organizations engaged the University of Ballarat to research the requirements and make recommendations on how the Central Highlands and Wimmera regions of Victoria could capture greater advantages from new information and communications technologies.

The research, documented in *Victoria's Golden West Portal Project Business Case* (Thompson, 1999), involved a number of different stages. These included confirming existing regional Web content, examining community portal developments, identifying portal tools, researching potential revenue streams, conducting focus group sessions, and other forms of stakeholder consultation.

The research report described how an environment could be established that would be conducive to the widespread adoption of electronic commerce. Specific recommendations included: establish a membership-based regional association with a specific focus on electronic commerce; establish infrastructure for a manageable and economically sustainable Internet presence in a way that would encourage the enhancement of community service and facilitate communities of interest and trading communities; and through a regional portal, achieve better Web content coordination, provide a valuable information source for residents, and also enhance efforts to promote all the attributes of the region.

The Chamber of Electronic Commerce Western Victoria Inc. (the Chamber) was established to facilitate the advancement of electronic commerce and implement the MainStreet portal project. Funding applications were prepared, and in November 1999 the MainStreet project secured funding of AUD 274,000 through Networking the Nation, with a further AUD 135,000 approved in May 2000. The University's Centre for Electronic Commerce and Communications (CECC) was then contracted to implement the project because it had the specialist skills necessary to develop the portal infrastructure and services.

Research had identified that many portal projects had produced 'static' or 'fixed' solutions. The MainStreet model, with the inclusion of a technical team as a critical element, was different, but the decision to have this team was significant in determining how the MainStreet project would evolve. The technical officer and part-time programmers would develop a portal framework based on the core services identified during the preliminary study. All tools would be selected or developed with non-technical end users in mind. The initial toolset would include event calendars; news publishing tools; online registration, payment, and product systems; and access to Web wizards and other Web publishing tools. This would be achieved by incorporating a range of in-house developments, with some integration of externally sourced product. The core services would create capacities to link regional Internet information and services, construct searchable directories, dynamically generate content like news and weather, distribute publishing and authoring rights, and promote community news and events.

The MainStreet project was actively promoted in the period leading up to its official launch in July 2000. This promotion was important as it helped to maintain interest in the project while technical developments proceeded behind the scenes.

During the early part of 2002, the MainStreet project attracted its first major client. Success in securing the Ararat Online project (www.ararat.asn.au) was attributed to involving regional stakeholders right from the project's beginning. Ararat's Economic Development Manager had participated in a range of activities, meetings, and focus group sessions. Through these activities he developed a strong understanding of how MainStreet offered Ararat something different that could be applied immediately to benefit his local community.

The Ararat Online project would include a range of elements with more than 80 businesses and community groups to benefit directly from an upgrade of their Web presence. They would also be given the opportunity to undertake training so that each organization would gain the skills to manage their own site. A further opportunity would be available for six businesses through an e-commerce mentoring program. Selected businesses would be assisted in the implementation of electronic commerce initiatives developed to match their particular business needs.

The value derived from the Ararat Online project was substantial. First, although the project did not represent a significant 'bottom-line' contribution in the context of the overall project budget, the investment of AUD 8,000 in a regional electronic commerce context represented a significant buy-in for the MainStreet product. Second, the Ararat Online project provided an opportunity to showcase the full product suite, its technical capabilities, the Web products, and the training and consulting services. Third, the project would help to address one of the early barriers: people in the target region had a very limited understanding of what a portal was. The Ararat Online project would provide a 'real' example, which it was hoped could be used to demonstrate the value and benefits that were associated with the efficient linking of Internet-based information and services in an easily searchable form. In other words, the Ararat Online project would establish the first 'before' and 'after' images. This proved to be a very powerful marketing mechanism for the project.

The project's technical team, however, had their task doubled—they were now expected to build not one, but two portals, and to deliver these within very short periods. They were successful in developing a way to replicate the MainStreet functionality through Ararat Online (www.ararat.asn.au) and later through projects with the

Birchip Cropping Group (www.bcg.org.au), Moorabool Shire (www.mconline.com.au), and Pyrenees Shire (www.pyreneesonline.com.au).

The original goal had been to establish MainStreet as the "point of first electronic contact for the region" (Thompson 1999, p. iv.). The vision was that people would find MainStreet, and from there be able to search and access information about a particular region or locate services of a particular type. What, however, was now understood was that 'communities' were much more motivated if the functionality of MainStreet could be delivered with local Web addresses and branding. Information could then be filtered up to the MainStreet umbrella so that client communities could be either accessed directly or through MainStreet. While this turned the original concept upside down, there was a strong indication that communities in the region were prepared to pay to both establish and maintain a service based on the 'replicable portal framework' developed through the MainStreet project.

FUTURE TRENDS

The MainStreet portal infrastructure and tools have since been replicated to suit a range of different clients, with this approach proving to be a very effective way of getting people actively engaged online. Appendix 1 contains a selection of URLs for clients including local governments, town-based communities, membership-based organizations, industry groups, and small and medium enterprises.

While a number of factors have been highlighted, the most successful and distinctive aspect has been the development of the replicable portal framework. It has been this capability that has been leveraged to cause increase in 'buy-in', participation, and ongoing investment in regional Web-based services. Members of 'geographic communities' and 'communities of interest' are able to work with CECC to design and implement sophisticated Web-based services, customized to meet their specific communication, promotional, and/or electronic commerce needs. Through this university/community partnership, initiatives are then sustained by putting community members in charge of the management of their online community. Local ownership and the sustainability of infrastructure and technical support services have been achieved by effectively aggregating regional demand for portal services.

The MainStreet project has made a significant contribution to the advancement of ICT and electronic commerce uptake in the Central Highlands and Wimmera regions of Victoria. Many individuals and communities have been assisted in advancing their uptake of electronic commerce as they update their community sites, publish event information and news items, or show others how to build simple Web sites. The level of functionality and services accessed is high and, because clients have strong ownership of their online activities, maintain their own Web-based information, and are committed to annually investing to maintain the portal infrastructure and services, the services can continue to be delivered after the initial seed funding period.

The MainStreet project has also supported and encouraged a staged uptake of electronic commerce, with a number of organizational clients becoming increasingly confident in both selecting and investing in electronic commerce solutions.

Services have also been customized to meet the needs of small groups such as Birchip Cropping Group, and also larger communities such as Moorabool, Ararat, and the Pyrenees Shire regions. This has overcome a barrier where under most models, the costs to establish (and sustain) a local portal have been substantial, and therefore prohibitive for small towns and community groups.

CONCLUSION

Through the MainStreet project, regional and rural communities have a greater ability to build on local strengths and capitalize on the opportunities that are provided by electronic commerce and ICT. Communities, however, just like businesses, require assistance in identifying the most appropriate online service for their particular circumstances. Policies that encourage communities to enter collaborative partnerships, and which leverage existing infrastructure, knowledge and learning, should thus be seen as preferable to the funding or establishment of discrete or stand-alone initiatives. Well-developed and well-implemented online services can make a positive contribution to the future of regional and rural communities. Case studies such as the one presented in this article are effective in illustrating the impacts, influences, and challenges that can be experienced in operationalizing and sustaining online communities in a regional and rural context.

ACKNOWLEDGEMENTS

The author acknowledges Brian West from the University of Ballarat who has been generous in the provision of advice and encouragement that greatly assisted in the preparation of this work.

REFERENCES

Ashford, M. (1999). Online WA: A trickle-up approach to using communications to enhance regional economic and social development. *Proceedings of the Regional Australia Summit,* Canberra, Australia.

Bernal, V. (2000, November). *Building online communities: Transforming assumptions into success.* Retrieved from *benton.org/Practice/Community/assumptions.html.*

Brumby, H.J. (2001). *Connecting communities: A framework for using technology to create and strengthen communities.* State Government of Victoria, Melbourne.

Department of Communications Information Technology and the Arts. (2000). *Taking the plunge: Sink or swim? Small business attitudes to electronic commerce.* Commonwealth of Australia, Canberra.

Department of Communications Information Technology and the Arts. (2001). *Funding priorities and principles, networking the nation, the commonwealth government's regional telecommunications infrastructure fund.* Commonwealth of Australia, Canberra.

Gronlund, A. (2001). Building an infrastructure to manage electronic services. In S. Dasgupta (Ed.), *Managing Internet and intranet technologies in organizations: Challenges and opportunities.* Hershey, PA: Idea Group Publishing.

Hunter, A. (1999). Opportunities through communications technology for regional Australia. *Proceedings of the Regional Australia Summit,* Canberra.

Keeble, L. & Loader, B.D. (2001). *Challenging the digital divide: A preliminary review of online community support.* CIRA, University of Teesside, UK.

Local Government Association of Tasmania and Trinitas Pty Ltd. (2001). *Online service delivery strategy paper—gaining the maximum benefit for our communities from the local government fund.* Local Government Association of Tasmania, Hobart.

McGrath, M. & More, E. (2002). *Forging and managing online collaboration: The ITOL experience.* National Office for the Information Economy and Macquarie University, Canberra, Australia.

More, E. & McGrath, M. (2003). Organizational collaboration in an e-commerce context: Australia's ITOL project. *The E-Business Review III*, 121-123.

Multimedia Victoria. (2002). *Connecting Victoria: A progress report 1999-2002.* State Government of Victoria, Melbourne.

Municipal Association of Victoria and ETC Electronic Trading Concepts Pty Ltd. (2000). *Local government—integrated online service delivery strategy and implementation plan, executive summary—final.* Municipal Association of Victoria, Melbourne.

National Office for the Information Economy. (2001). *B2B e-commerce: Capturing value online.* Commonwealth of Australia, Canberra.

National Office for the Information Economy. (2002). *The benefits of doing business electronically—e-business.* Commonwealth of Australia, Canberra.

National Office for the Information Economy. (2002). *Guide to successful e-business collaboration.* Commonwealth of Australia, Canberra.

Papandrea, F. & Wade, M. (2000). *E-commerce in rural areas—case studies.* Rural Industries Research and Development Corporation, Canberra.

Pattulock, E. & Albury Wodonga Area Consultative Committee. (2000). *Facilitation of e-commerce and Internet use by regional SMEs.* Albury Wodonga, La Trobe University, Australia.

Porter, M.E. (2001). Strategy after the Net. *BOSS*, (April), 17-23.

Preece, J. (2000). *Online communities: Designing usability, supporting sociability.* Chichester, UK: John Wiley & Sons.

Rheingold, H. (1994). A slice of life in my virtual community. In L.M. Harasim (Ed.), *Global networks: Computers and international communication* (pp. 57-80). Cambridge, MA: MIT Press.

Simpson, L. (2002). Big questions for community informatics initiatives: A social capital perspective. *Search Conference: Community and Information Technology The Big Questions,* Centre for Community Networking Research, Monash University, Melbourne, Australia.

Simpson, R. (1999). Brave new regions. *Proceedings of the Regional Australia Summit,* Canberra, Australia.

Telstra Country Wide. (2002). *Our community online.* Letter and brochure distributed to local government conference delegates, 31 October 2002, Telstra Corporation Limited.

Thompson, H. (1999). *Victoria's Golden West portal project business case.* Centre for Electronic Commerce and Communications, University of Ballarat, Australia.

KEY TERMS

'Bottom-Up' Approach: Development approach founded upon the principle that communities are better placed to coordinate and integrate efforts at the local level.

Case Study: The intensive examination of a single instance of a phenomenon or where one or just a few cases are intensively examined using a variety of data-gathering techniques.

Community Informatics: A multidisciplinary field for the investigation of the social and cultural factors shaping the development and diffusion of new ICT and its effects upon community development, regeneration, and sustainability.

Community Portal: Online initiative often developed through participative processes which aims to achieve better coordination of relevant Web-based information and provide communication services for community members.

Regional Development: The act, process, or result of actions to grow, expand, or bring a regional place to a more advanced or effective state.

Web Portal: Focal points on the Web which provide a place to start. Web portals facilitate the location of information by incorporating the strengths of search engines and additionally provide more efficient access to information by categorizing it into easily recognizable subcategories or channels.

APPENDIX 1

University of Ballarat URL	
University of Ballarat	www.ballarat.edu.au
CECC	www.cecc.com.au
MainStreet portal URL	
Mainstreet.net.au	www.mainstreet.net.au
Geographical portal URLs examples	
Ararat Online	www.ararat.asn.au
Moorabool Online	www.mconline.com.au
Pyrenees Online	www.pyreneesonline.com.au
Membership based communities URLs examples	
Birchip Cropping Group	www.bcg.org.au
Young Australian Rural Network	www.yarn.gov.au
Rural Regional Research Network	www.cecc.com.au/rrrn
Pyrenees Hay Processors	www.exporthay.com.au
Comprehensive Web site URLs examples	
Ballarat A Learning City	www.ballaratlearningcity.com.au
Central Highlands Area Consultative Committee	www.chacc.com.au
Pyrenees Shire	www.pyrenees.vic.gov.au
Regional Connectivity Project	www.regionalconnectivity.org

Building Police/Community Relations through Virtual Communities

Susan A. Baim
Miami University Middletown, USA

INTRODUCTION

Over the past two decades, police departments around the globe have been involved in a slow, but steady transition from call-based policing to community-oriented policing. The former approach, while effective at closing cases once a crime has occurred, does little to develop crime prevention partnerships between officers on the beat and the citizens of local communities. Community-oriented policing serves to increase awareness of issues and potential problems before they occur, thus assisting police departments to provide a more proactive approach to stopping crime within their communities.

One of the greatest difficulties in developing effective community-oriented policing programs is establishing solid, two-way communications links between police officers and the populations that they serve. Information flow to the police and suggestions back to the citizenry often fall victim to the same constraints—lack of time to interact effectively and lack of a ready-made mechanism to deliver the information in a timely manner. To reduce or eliminate these constraints, interactive police department Web sites and virtual communities (that involve both police officers and citizens) can provide actionable and measurable performance increases in the efficiencies and the effectiveness of community-oriented policing efforts. Although the IT hardware, software, and design expertise needed to create interactive Web sites and virtual communities are readily available, online efforts at community-oriented policing will remain more of a theoretical interest than a broad-scale application until police departments truly understand the needs and the wants of the citizens within their local communities.

This article explores a service-learning approach for use in a university classroom that combines IT applications with current research practices in the use of citizen satisfaction surveys conducted for local police departments. Examples are drawn from three primary-based research studies involving police departments that are turning away from call-based policing practices and proactively moving toward community-oriented policing practices.

BACKGROUND

Descriptions of community-oriented policing efforts may be found in the literature as early as the 1960s, although the majority of papers published date from the mid-1990s to the present day. Successful community-oriented policing programs began to emerge as departments returned to fundamental cop-on-the-beat policing that put officers back in close contact with citizens in their neighborhoods and in their places of business (Sissom, 1996). The knowledge gained from the early community-oriented policing efforts was used to improve departmental training efforts and also used to focus police officers more closely on crime prevention techniques. Community-oriented policing practices have continued to evolve in more recent studies where most authors focus on how the police can better identify specific issues that are divisive within their respective communities (Culbertson, 2000; Vincent, 1999; Woods, 1999; Rohe, Adams & Arcury, 2001).

Community-oriented policing programs rely heavily on current and ongoing issues of citizen concern received from both the police departments and the citizens of the communities served. The basic premise of community-oriented policing involves both sides becoming very familiar with each other's needs, wants, and expectations, and then forming a true community partnership to create a safe environment for citizens to live and work. Police officers are, in a sense, being asked to enroll in a police version of a basic marketing course in order to learn how to sell this new approach to the residents of their communities (Cummings, 2001). Residents, long accustomed to seeing police officers only when an emergency has been reported, can represent a tough sell for police officers in terms of forming proactive crime prevention citizen partnerships. Additionally, many police departments, themselves, may believe that the extra time and effort necessary to create community-oriented policing programs is not worth the increased costs, given the difficulties in measuring the perceived benefits of crime prevention programs. Crime, itself, is measurable and actionable in terms of police performance. For example, the widespread incorporation of computerized emergency call systems (e.g., 911 in the United States and similar systems in other nations) has given police departments ready access to tools capable of tracking performance measures such as call volume, time from call to officer arrival, clearance rate of calls, and so forth (Siegel, 1999). Particularly for police departments that score well on these measures and are rewarded appropriately by their city councils and/or their citizenry, the impetus to move toward more time-consuming and less easily quantified community-oriented policing objectives appears to be small. Like many governmental agencies, operational change in police departments tends to be extremely slow and very difficult to implement.

Regardless of these prevalent views, however, one finding that all parties seem to agree on is that the proper incorporation of new computer-based technologies will help police departments get closer to the citizens that they protect and serve. Computer-based technologies also help the individual officer solve crimes at the same time. An excellent example of such a technology is the mobile laptop computer now found in a high percentage of patrol cars (Greenemeier, 2002; Couret, 1999). Officers in the field now have access to virtually the same information as their office-based counterparts, and they can get at that information in real time without the translation losses associated with working through a radio dispatcher. Frequent reliance on the Internet and e-mail for sharing information and communicating between local police departments and other external police agencies also adds to the active network in use by the majority of police departments today. Given the significant improvements in computer technology, the timing is right for police departments to begin to implement community-based policing practices.

The design and development of efficient and effective "customized" community-oriented policing programs clearly places a burden on police departments to solicit, collect, analyze, and interpret data from their citizenries in order to make wise choices regarding the scope and size of any program that is set up. Obtaining high-quality data can be a formidable task, due to the diversity of the population to be sampled and due to the fact that not all respondents share the same expectations regarding active participation in crime prevention with their local police departments. It is with this background framework in mind that modern IT techniques, combined with current research practices, can significantly boost the ability of all parties to communicate and to share information that is critical in moving a police department from call-based policing practices to community-orientated police practices.

SURVEYS, INTERACTIVE WEB SITES, AND VIRTUAL COMMUNITIES

Police departments often lack not only the knowledge of what citizens might respond favorably to in terms of interactive Web sites or virtual communities, but also to the expertise that is needed to conduct unbiased research surveys among their constituencies to generate the required data input. Citizen satisfaction surveys are becoming highly efficient and effective tools for a variety of city government purposes with credible studies cited in the literature over the past several years (Kearney, Feldman & Scavo, 2000; Oleari, 2000). Often, such citizen surveys, conducted among random samples of the community's population, will be highly revealing of the type of information needed to set up initial Web site and/or virtual community structure(s).

To generate useful data that can drive the development of interactive Web sites and virtual communities, it may be necessary to enlist the services of professional researchers. Given large citizen populations, researchers may want to choose conducting either a mail or an online survey. If neither format is selected, they then need to develop a survey instrument, a process plan, and a timeline for conducting the survey, and at the conclusion of the survey, an unbiased set of concrete, actionable, and meaningful recommendations on how to put the data to use. Unfortunately, hiring a professional research firm can cost thousands of dollars that taxpayers cannot afford to spend given today's tight city budgets. A workable alternative, therefore, is to "hire" university students (under the direction of an instructor knowledgeable in current research techniques) who want to have a "hands-on" educational experience that benefits themselves, their university, and their local community in a service learning setting.

The three citizen satisfaction surveys described in this article were conducted following a traditional mail survey format. Table 1 briefly summarizes the city locations, numbers of surveys sent and returned by respondents (including response rate), and the dates that the research studies were conducted over the past few years. In each case, the study was run collaboratively with a police department in Southwestern Ohio (all looking to implement community-based policing practices in their communities) and an undergraduate Marketing course at Miami University. Students in each of the courses developed the database (in Excel or in Access), handled all of the data tabulation, and analyzed the results under the guidance of the author (as instructor of the courses).

Table 1. Citizen satisfaction studies (mail surveys) involving local police departments

City	No. of Surveys Sent	No. of Surveys Returned	Response Rate	Date Conducted
Middletown, Ohio	2,000	636	32%	Spring 2000
Oxford, Ohio	1,857	522	28%	Spring 2001
Trenton, Ohio	1,600	478	30%	Spring 2001

Since future citizen satisfaction surveys of this type may be conducted using Internet-based survey instruments in situations where there is reason to believe that a sufficient concentration of "online" respondents is available, citizen respondents were asked to provide data on their Internet usage and their previous access to city and/or police department-sponsored Web sites. Data were also collected on the desirability and projected usage level of advanced interactive Web services, in case these types of services should be offered by police departments at a later date. Selected data are summarized in Table 2.

A careful examination of the data generated across all three surveys reveals several important findings for police departments that may desire to set up interactive Web sites and/or virtual communities to enhance police/community relations. These findings were generated by examining the cross-tabulations of the questions noted in Table 2 with various demographic parameters tracked in the surveys and also the respondent verbatims related to Web site usage. Results are summarized qualitatively in Table 3.

- In all three communities surveyed, personal Internet connectivity varied with age group. Older residents were less likely to have Internet access at home, and they did not appear to make up for this lack by using Internet access at work or public Internet access at libraries or other similar locations.
- Among residents with Internet connections, most users felt comfortable using Internet sites as casual observers or to download information, but only a small percentage of individuals were comfortable sending personal information online. Subsequent questions revealed that this reluctance could be attributed to the desire to keep personal information secure—regardless of any printed claims regarding "secure" Web sites or other online privacy statements relayed by Internet providers.
- Even among residents who felt comfortable with two-way communications online or with using interactive Web sites, it is evident that traditional communication mechanisms would have to be maintained in order to meet all needs, especially as they relate to police departments. Most residents believe that traditional means of communication are a "safety net" that must be used when electronic communications go down.

Police departments seeking to increase citizen participation in community-oriented policing efforts through the incorporation of electronic communications mechanisms such as interactive Web sites and/or virtual communities may face a similar situation as outlined above. While it would be too simplistic to state that the general citizenry will not accept such efforts, it is imperative to acknowledge that a significant percentage of city residents may be reluctant to participate in online communications at the present time. As such, it is critically important for those involved in setting up interactive Web sites and/or virtual communities to bear in mind that it may take time for these initiatives to "catch on" and that a keen eye

Table 2. Citizen satisfaction studies—Internet usage

City	No. of Surveys Returned	Percent of Respondents Using the Internet	Percent of Internet Users Accessing City and/or Police Web Sites	Percent of Internet Users w/Interest in an Upgraded Police Web Site
Middletown, Ohio	636	40%	29%	13%
Oxford, Ohio	522	81%	28%	3%
Trenton, Ohio	478	66%	15%	48%

Table 3. Key findings regarding citizens' use/desire to use the Internet for police services

- In all three communities surveyed, personal Internet connectivity varied with age group. Older residents were less likely to have Internet access at home, and they did not appear to make up for this lack by using Internet access at work or public Internet access at libraries or other similar locations.
- Among residents with Internet connections, most users felt comfortable using Internet sites as casual observers or to download information, but only a small percentage of individuals were comfortable sending personal information online. Subsequent questions revealed that this reluctance could be attributed to the desire to keep personal information secure—regardless of any printed claims regarding "secure" Web sites or other online privacy statements relayed by Internet providers.
- Even among residents who felt comfortable with two-way communications online or with using interactive Web sites, it is evident that traditional communication mechanisms would have to be maintained in order to meet all needs, especially as they relate to police departments. Most residents believe that traditional means of communication are a "safety net" that must be used when electronic communications go down.

is essential to make the Web sites attractive (with frequent updates) for casual and/or new Internet users.

John Hagel and Arthur Armstrong contend that there is "nothing more uninviting to passing Web traffic than a community without members" (1997, p. 134). Interpreting this comment in the context of community-oriented policing is straightforward. To attract proactive, involved citizens, interactive Web sites and virtual communities must offer something of value—something to cause these citizens to check back frequently, to offer suggestions, and to take away information that is valuable and that cannot be obtained as easily through alternative means. In bringing interactive Web sites and/or virtual communities to fruition, it may be advantageous to encourage members of the police department to participate vigorously and frequently—especially as the sites are first introduced. Some citizens may take a "watch and see" attitude before beginning to participate in online community-based policing efforts. The presence of helpful and interesting information from the police department may help draw these citizens from the mode of casual observer to the mode of full-partner participant.

The technology required to offer interactive Web sites and virtual communities is well established, and it will not be addressed specifically in this article. Police departments have numerous choices in how to achieve these goals, ranging from employing the services of professional Web site and virtual community design/support organizations, to "bootlegging" efforts from members of the city workforce and/or police department who are proficient at computer operations. What is mandatory, however, is the list of criteria given in Table 4.

- The Web site and/or virtual community must be in operation 24/7. Long periods of inactivity or excessive "down" periods must be avoided.
- Police officers and/or a department spokesperson(s) must monitor the Web site(s) frequently and actively participate in the citizen/police partnership.
- Continuous solicitation of new members will keep the Web site(s) fresh and productive.
- Police departments must not become discouraged if it takes a relatively long period for the concept to "catch on" in their communities.

Many police departments have a preliminary foundation on which to build an interactive Web site or a virtual community since they already have a police department Web site in place. (Among the three Ohio cities surveyed, Middletown, Oxford, and Trenton, all have police department Web sites or Web pages in an overall directory of city services.) Even if these sites exist only to provide a brief overview of the departments' operations, they can serve as a starting point to build citizen involvement. The key for the developer charged with expanding a basic Web site into an offering capable of engaging the local citizenry in community-oriented policing is to recognize that marketing the partnership process and the end result of the project (in terms of actionable and measurable performance indicators) will be as critical as executing the technical details of the assignment. Considered to be one of the true Internet pioneers, Howard Rheingold stated in an interview that the three most critical things that a developer must determine before beginning a project of this type are: 1) how the site will be marketed, 2) what is expected in return for visits by potential members, and 3) what technologies will be needed to put the site into proper operation (Moore, 2001).

While interactive Web sites can provide a forum for police officers and citizens to exchange information, in general these exchanges are likely to be highly discreet in nature—involving only one citizen and one police officer

Table 4. Criteria for successful use of interactive Web sites and virtual communities

- The Web site and/or virtual community must be in operation 24/7. Long periods of inactivity or excessive "down" periods must be avoided.
- Police officers and/or a department spokesperson(s) must monitor the Web site(s) frequently and actively participate in the citizen/police partnership.
- Continuous solicitation of new members will keep the Web site(s) fresh and productive.
- Police departments must not become discouraged if it takes a relatively long period for the concept to "catch on" in their communities.

or department representative at a time. The real opportunity to magnify the positive benefits of community-oriented policing will occur as groups of citizens and officers begin to communicate on a frequent basis. If this is handled electronically, the communications may evolve into a highly functional virtual community. Police departments are well positioned for such an effort because police officers need to interact with the citizenry in order to keep a finger on the pulse of the community. Citizens, on the other hand, often have trouble interacting with the police unless an emergency has occurred. The mindset is that police officers are incredibly busy and that they do not have time to "just chat," when it is actually through such interactions that community-oriented policing makes its mark.

Police departments will need to go the extra mile in establishing virtual communities because individual citizens, or even groups of business people, will be unlikely to set up a virtual community infrastructure with sufficient credibility to attract widespread participation. Police departments may want to borrow guidelines on community participation from authors such as Rachel Gordon, who has published an extensive discussion on "hints for success" in virtual community interactions (Gordon, 2000). Privacy and confidentiality concerns must also be properly addressed. Even though an interactive Web site or a virtual community is considered to be a public forum, clear statements regarding the collection and use of information are warranted (Sheehan, 2002; Sheehan & Hoy, 2000; Milne & Culnan, 2002). Finally, police departments should also keep in mind that they are not setting up interactive Web sites and virtual communities in order to run for-profit businesses. As such, the goal is not to extract data from citizen participants in order to judge how to market revamped and/or new police services. Rather, the goal is to generate critical information and upgrade the ability of both sides to deal with the process of maintaining an environment in which all citizens feel safe to live and work (Brewer, 2000; Wood, 2000).

FUTURE TRENDS

Community-oriented policing is here to stay. For the three Southwestern Ohio communities studied, the benefits of the process far outweigh the anticipated costs and extra effort put forth by their police departments. Data consistently showed that community-oriented policing programs would be greatly appreciated by citizens and business owners alike. Nevertheless, time pressures and resource constraints still continue to plague police departments in the same manner as they do other for-profit businesses, non-profit organizations, and governmental agencies. The key to successful implantation of community-oriented policing efforts thus becomes finding efficient and effective ways to interact with the public at large. Fortunately, the rapid increase in Internet connectivity and the general upswing in personal e-mail and Web usage provide a suitable vehicle for police departments to use in establishing better communication links with the citizenry.

It is highly likely that police departments will continue to increase their presence online in a variety of formats. The relevant issues requiring future study are primarily application oriented and not rooted in the basic technologies of setting up Web sites and/or offering access to virtual communities. Fortunately, police departments can ride the coattails of IT developers who are already generating continuously more advanced Internet communications technologies for business and industrial clients. This is not to say, however, that engaging the citizenry of a community to communicate with the police online is an easy or well-understood proposition. The state of the art today is one of relatively infrequent successes. Solid, two-way communications take time and money to implement, and both issues are at a premium in most police departments—regardless of whether those communications take place through traditional channels or in an online format.

Citizen satisfaction surveys, conducted through a mail survey format as described in this article or conducted in a fully online research mode, are likely to grow in importance as police departments continue to focus more energy on meeting the specific needs and wants of their communities. In terms of service learning at the university level, surveys of this type are natural service learning projects for post-secondary students in business, sociology, criminal justice, and other similar majors where a quantitative and qualitative study of police/citizen interactions is of value. Educators are likely to find students highly willing to participate in such work, and police departments eager for the unbiased "third-party" perspective that a well-run research study can provide.

Future research should focus on identifying more precisely what information citizens would value receiving through interactive Web sites and/or virtual communities. Individuals charged with developing interactive Web sites and/or virtual communities for police and citizen use would also be well advised to consider surveying police officers to better elucidate their department's needs and wants before beginning the design process. Collectively, the issues surrounding the "true" needs and wants of citizens in a local community transcend the technical aspects and challenges of mainstream research in information technology to include additional aspects of public policy, criminal justice, and citizens' rights. Citizen satisfaction surveys are, nevertheless, one form of practical application toward the future direction of applied information technology solutions.

CONCLUSION

Interactive Web sites and virtual communities represent two of the most innovative ways to generate meaningful, two-way dialogs over the Internet. As the technologies for connecting multiple Internet users in these manners mature, information technology researchers and developers can turn significant attention toward solving the specific application problems posed by unusual clients. Connecting the citizenry of a community with their police department in an efficient and effective manner is just one of a growing number of novel applications made possible by advanced Web site design and virtual community hosting technologies. As community-oriented policing efforts continue to grow in the coming years, it is highly likely that the efforts of information technology professionals will play a critical role in their success—making all of us feel safer at home and at work.

REFERENCES

Brewer, C. (2000). Community is the fly paper of your site. *Computer User, 18*(12), 49.

Couret, C. (1999). Police and technology. *The American City & County, 114*(9), 31-32+.

Culbertson, H.M. (2000). A key step in police-community relations: Identifying the divisive issues. *Public Relations Quarterly, 45*(1), 13-17.

Cummings, B. (2001). NYPD meets marketing 101. *Sales and Marketing Management, 153*(4), 14.

Gordon, R.S. (2000). Online discussion forums. *Link-up, 17*(1), 12.

Greenemeier, L. (2002). Sacramento cops take e-tools on the beat. *Information Week, 886*, 60.

Hagel III, J. & Armstrong, A.G. (1997). *Net gain*. Boston: HBR Press.

Kearney, R.C., Feldman, B.M. & Scavo, C.P.F. (2000). Reinventing government: City manager attitudes and actions. *Public Administration Review, 60*(6), 535-547.

Milne, G. & Culnan, M. (2002). Using the content of online privacy notices to inform public policy: A longitudinal analysis of the 1998-2001 U.S. Web surveys. *The Information Society, 18*, 345-359.

Moore, R. (2001). Focus on virtual communities. *B to B, 86*(7), 14.

Oleari, K. (2000). Making your job easier: Using whole system approaches to involve the community in sustainable planning and development. *Public Management, 82*(12), 4-12.

Rohe, W.M., Adams, R.E. & Arcury, T.A. (2001). Community policing and planning. *Journal of the American Planning Association, 67*(1), 78-90.

Sheehan, K. (2002). Toward a typology of Internet users and online privacy concerns. *The Information Society, 18*, 21-32.

Sheehan, K. & Hoy, M. (2000). Dimensions of privacy concern among online consumers. *Journal of Public Policy & Marketing, 19*(1), 62-73.

Siegel, F. (1999). Two tales of policing. *Public Interest, 134*, 117-121.

Sissom, K. (1996). Community-oriented policing means business. *FBI Law Enforcement Bulletin, 65*(12), 10-14.

Vincent, E. (1999). How citizens' voices are heard in Jacksonville. *Sheriff Times, 1*(10). Retrieved October 10, 2003, from *www.communitypolicing.org/publications/shtimes/s10_fa99/s10vince.htm*

Wood, J.M. (2000). The virtues of our virtual community. *Instructor, 110*(1), 80-81.

Woods Jr., D.D. (1999). Supervising officers in the community policing age. *Sheriff Times, 1*(10). Retrieved October 10, 2003, from *www.communitypolicing.org/publications/shtimes/s10_fa99/s10woods.htm*

KEY TERMS

Call-Based Policing: Traditional policing approach whereby officers respond to emergency calls for assistance and address crimes and other situations after the fact.

Citizen Satisfaction: Term coined to describe the overall approval rating of services received by citizens within their communities. A 100% rating indicates total satisfaction with services received. Ratings may be taken "per service" or represent satisfaction with an entire government structure.

Community-Oriented Policing: Contemporary policing approach that builds relationships between police officers and the citizens of a community on an ongoing basis. Crime prevention is stressed as a partnership approach before an actual emergency situation(s) develops.

Interactive Web Site: A Web site or page configured so as to invite correspondence between the user and the originator/sponsor of the site. Such sites go beyond passively providing information to those who browse the site. Customarily, there are options to complete online surveys, send e-mail to the sponsor, request specialized or personalized response(s), and so forth.

Internet Privacy: Concerns expressed by Internet users regarding the security and confidentiality of information transmitted electronically. Government agencies and business/industry professional groups share responsibility to address Internet privacy concerns.

Internet-Based Survey: A contemporary survey technique through which researchers may obtain respondents' opinions via online survey processes. Respondents may either be asked to go to a Web site to complete a survey (Web-based) or the survey questionnaire may be e-mailed to the respondents (e-mail-based) for them to complete and return electronically.

IT Applications: Research and development work performed to create a situation-specific bridge between new or existing IT hardware and software technologies and the information needs/wants of a customer. The combination of proper hardware, software, and tailored application delivers a well-rounded IT solution for the customer's problem.

Mail Survey: A traditional survey technique in which a multi-part survey questionnaire is mailed to a randomized sample of individuals (within a larger population) who are asked to complete the questionnaire and return it to the survey researcher for tabulation and analysis.

Service Learning: Educational projects structured to take students out of the traditional lecture-based classroom and involve them in "real-world" problem-solving opportunities of importance to their communities. Students apply theories learned in their classroom studies in a practical manner that generates a completed work product for one or more clients, while helping to cement in students' minds the usefulness of the theoretical concepts under study.

Virtual Community: An online forum or discussion group through which members may interact either in real time or asynchronously. Most virtual communities use discussion groups and message boards that are accessible online to all members. Members correspond by posting messages back and forth within the forum. Membership in a virtual community indicates that the user shares one or more common interests with others in the same forum.

Business Model Application of UML Stereotypes

Daniel Brandon, Jr.
Christian Brothers University, USA

OVERVIEW

The UML (Unified Modeling language) has become a standard in design of object-oriented computer systems (Schach 2004). UML provides for the use of stereotypes to extend the utility of its base capabilities. In the design and construction of business systems, the use of stereotypes is particularly useful stereotypes, and this article defines and illustrates these.

UML STEREOTYPES

“Stereotypes are the core extension mechanism of UML. If you find that you need a modeling construct that isn’t in the UML but it is similar to something that is, you treat your construct as a stereotype” (Fowler & Kendall, 2000). The stereotype is a semantic added to an existing model element and diagrammatically it consists of the stereotype name inside of guillemots (a.k.a., chevrons) within the selected model element. The guillemot looks like a double angle bracket (<< … >>), but it is a single character in extended font libraries (Brown, 2002). The UML defines about 40 of these stereotypes such as “<<becomes>>”, “<<include>>”, and “<<signal>” (Scott, 2001). However, these 40 standard stereotypes are not particularly useful in business models and do not add the meaning necessary for automatic code generation in a UML CASE tool.

One common general use of the stereotype is for a metaclass. A metaclass is a class whose instances are classes, and these are typically used in systems in which one needs to declare classes at run time (Eriksson & Penker, 1998). A similar general use is for powertypes. A powertype is an object type (class) whose instances are subtypes of another object type. Figure 1 shows an example of the use of stereotypes for powertypes (Martin & Odell, 1998).

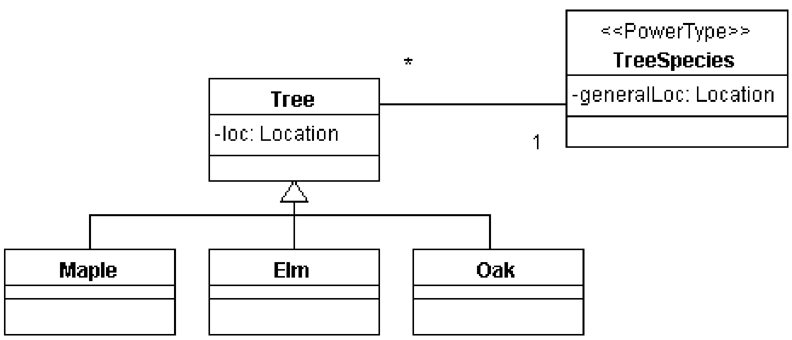
USER DEFINED STEREOTYPES FOR BUSINESS SYSTEMS

In the design of business systems we have found some stereotypes that were useful, and two stereotypes that are extremely useful. When defining stereotypes it is necessary to describe (Eriksson & Penker, 1998):

1. on which (UML) element the user-defined stereotype should be based;
2. the new semantics the stereotype adds or refines; and
3. pne or more examples of how to implement the user-defined stereotype.

A common use of stereotypes in business systems is for interfaces as found in Java or CORBA; this is shown in Figure 2. An interface typically has public functionality but not data (unless holding data for global constants). The class model element has been modified with the “<<interface>>” notation. This is commonly used for

Figure 1.



UML CASE products that do not have separate interface symbols or where these symbols do not allow data (i.e., global constants).

Still another common stereotype usage in business systems is to clarify or extend a relationship. Figure 3 shows a stereotype called “history” which implies a “many” cardinality for history purposes, that is, each person has zero or one current employers but may have many employers in terms of the employee’s history. It may imply some common functionality upon code generation such as (Fowler & Kendall, 2000):

```
Company Employee::getCompany(Date);
```

CODE WRITING AND GENERATION

Most modern UML CASE (Computer Aided Software Engineering) products can generate “skeleton” classes from the UML class diagrams and possibly other diagrams. For business systems design, we need to write the code for our classes (usually implemented in Java or C++) based on both the Structural Model (UML Class Diagram) and the Dynamic Model (UML Activity Diagram). This process is shown in Figure 4. It is very important that consistency between the two diagrams is achieved.

Many such CASE products allow the user to write their own “class generation scripts” in some proprietary scripting language or in a general scripting language (i.e., Python). With user defined stereotypes, the user can

Figure 2.

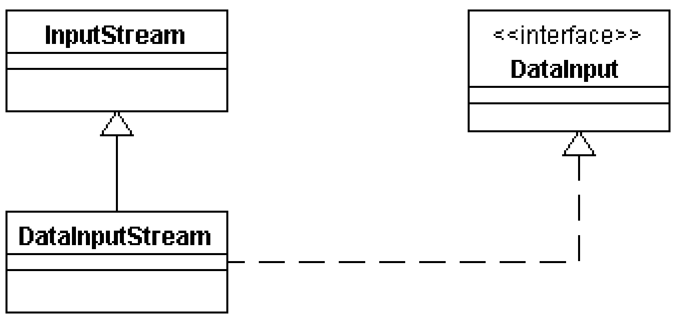


Figure 3.

Figure 4.

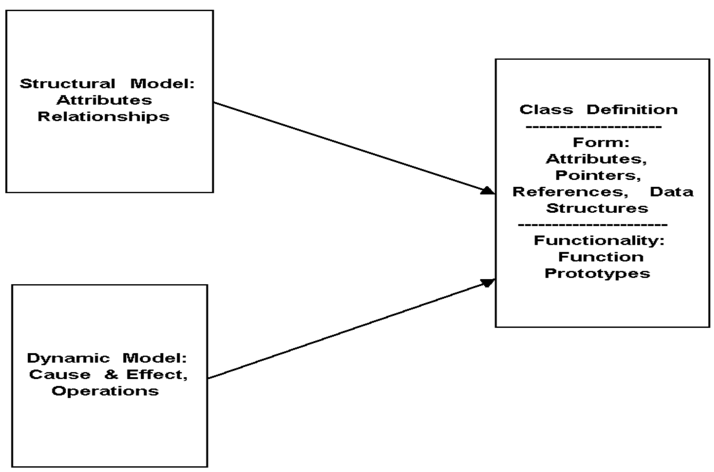


Figure 5.

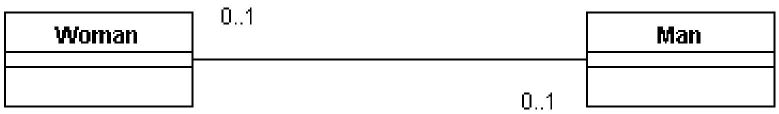
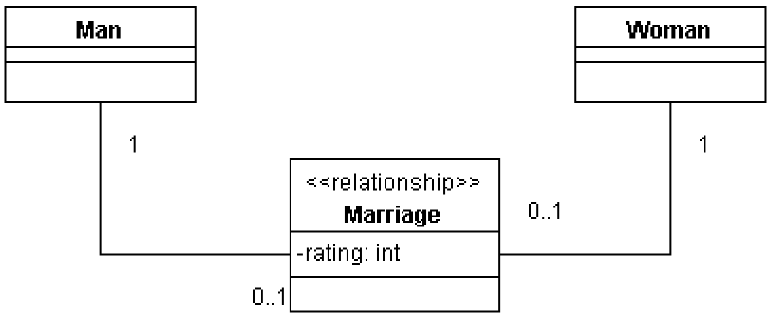


Figure 6.



modify the class generation script code to use their stereotypes as needed.

RELATIONSHIP OBJECT TYPES

As an example, consider a simple association between two object types. Often these simple relationships need to be modeled as object types because these relationships have data content and/or functionality. Figure 5 shows a simple association between two object types representing the relationship "current marriage". If we need to maintain an attribute on each marriage (such as rating), then we can more effectively represent the relationship as an object type as shown in Figure 6. Here we use the "relationship" stereotype to indicate that this object type is a relationship; and the code generation can use a more appropriate class representation. Others authors have suggested other notations for relationship object types such as "placeholders" (Martin & Odell, 1998), and UML suggests using the dotted line from a standard object type (class) to the relationship line. But implementing these other diagramming techniques in code generation is difficult and has ambiguity problems.

ACTIVITY DIAGRAMS

A UML Activity Diagram is a state diagram in which most of the states are action states, and most of the transitions are triggered by the completion of these action states. This is the case in most models of business systems. Activity diagrams identify action states, which we call operations (Martin & Odell, 1998), and the cause and effect between operations. Each operation needs to belong to an object type, at least for a C++ or Java implementation. Operations may be nested, and at some point in the design the operations need to be defined in terms of methods. The methods are the processing specifications for an operation and can be so specified in lower lever activity diagrams, pseudo code, or language specific code. Note that the term "methods" may cause some confusion here since in programming terminology, a method is a function defined within a class and it is invoked upon an object (unless it is a static method).

Drawing Methodology

Figure 7 shows a typical UML activity diagram for a simple ordering process. The operations are represented in the ovals and the arrows show the cause and effect scenario or the "triggers". In this diagram, there are two "fork/join" model elements, and the use of "conditional branch states" is also common. Each of the operations must be associated with a particular object type. The standard way to do that in this UML type diagram is to use "swimlanes", and these are the vertical lines shown in Figure 7.

There are two problems with the standard representation as shown in Figure 7. The first problem is that, as the system gets more complex (more object types and

Figure 7.

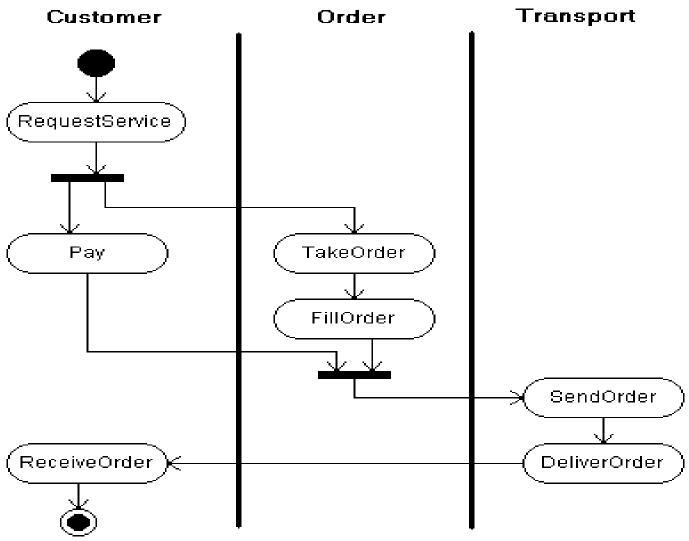
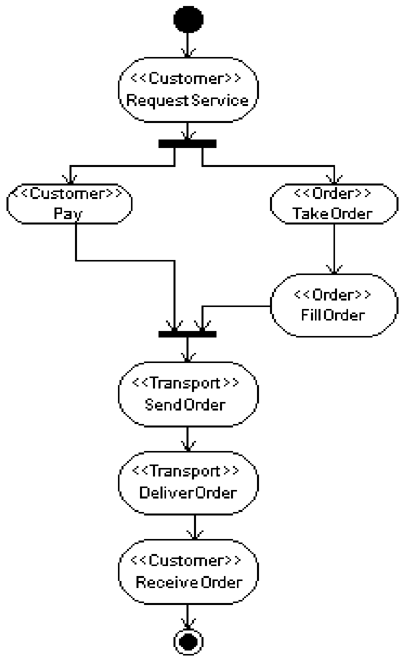


Figure 8.



operations), it is very difficult to draw in swimlanes. The second problem is that code generation is very difficult in UML CASE products since you have to scan the geometry of the drawing to find out which operations lay in which swimlanes. A solution to the above problems with standard UML activity diagrams is to use a stereotype for the operation element to indicate the object type (class) owning that operation. Figure 8 shows the same systems as Figure 7 drawn with the "operation owner" stereotype.

Model Consistency

The use of these UML stereotypes allows a greater degree of consistency checking of business models. A final business system design will involve several UML diagram types. For example business systems typically have static structural diagrams (UML Class Diagram) and dynamic diagrams (UML Activity Diagram). These diagrams must be consistent with one another, in particular:

1. The object types (shown with the operation stereotype notation) that contain the operations in activity diagrams must be included on the structural diagrams.
2. The operations shown in the activity diagrams (along with the object types identified with the stereotype notation) must be included as operations in the same object type on the structural diagrams.

For a UML business system example (including implementation in C++), the reader is referred to the full book chapter on this subject (Brandon, 2003).

FUTURE TRENDS

As UML becomes more accepted for general use in the design of business systems, we could expect to see more universal stereotypes being formulated. Eventually, libraries of these stereotypes should become generally available and UML tool vendors would include support for these libraries in their products.

CONCLUSION

UML stereotypes can be very useful in designing business systems. For example, the use of a "relationship" stereotype is helpful in static structural models (UML Class Diagrams) and the use of an "operation owner" stereotype is most helpful in dynamic models (UML Activity Diagrams). These stereotypes aid in both the design/drawing phase and in the implementation (coding) phase of the overall system construction.

REFERENCES

Brandon, D. (2003). Use of UML stereotypes in business models. Chapter in L. Faver (Ed.), *UML and the Unified Process*. Hershey, PA: IRM Press.

Brown, D. (2002). *An introduction to object-oriented analysis*. John Wiley & Sons.

Eriksson, H.-E. & Penker, M. (1998). *UML toolkit*. John Wiley & Sons.

Fowler, M., & Kendall, S. (2000). *UML distilled*. Addison-Wesley.

Martin, J., & Odell, J. (1998). *Object oriented methods – A foundation* (UML ed.). Prentice Hall.

Object Domain. (2001). Object Domain Systems Inc. Available at *www.objectdomain.com*

Schach, S. (2004). *Introduction to object oriented analysis and design*. Irwin McGraw Hill.

Scott, K. (2001). *UML explained*. Addison-Wesley.

KEY TERMS

Activity Diagram: An UML diagram showing operations and triggers between operations; a diagram which shows system dynamics via cause and effect relationships. An activity diagram is a state diagram in which most of the states are action states, and most of the transitions are triggered by the completion of these action states.

CASE: Computer Aided Software Engineering.

Class: A program construct representing a type of thing (abstract data type) which includes a definition of both form (information or data) and functionality (methods); the implementation of the design concept of "object type".

Composition: A new class in an objected programming language that is composed of other classes.

Dynamic Model: A UML model describing dynamic behavior such as state changes, triggers, and object type operations.

Encapsulation: The ability to insulate data in a class so that both data security and integrity is improved.

Framework: A software foundation which specifies how a software system is to be built. It includes standards at all levels both internal construction and external appearance and behavior.

Function: A programming construct where code that does a particular task is segregated from the main body of a program; the function may be sent arguments and may return arguments to the body of the program.

Implementation: The code placed inside of methods. For some languages, this code is pre-compiled or interpreted.

Include: Some code stored separately from the main body of a program, so that this code can be used in many programs (or multiple places in the same program).

Inheritance: A feature of object-oriented languages that allow a new class to be derived from another class (a more general class); derived classes (more specific classes) inherit the form and functionality of their base class.

Interface: The specification for a method ("what" a method does); how that function is called from another program. Interfaces are provided in source form as opposed to implementations which are secure. This allows one to use a method without regard for "how" that method is coded. It also allows multiple implementations of the same interface.

Libraries: A group of functions and/or classes stored separately from the main body of the main program; an "include" file consisting of functions and/or classes.

Metaclass: A class whose instances are classes.

Method: A function defined inside of a class, a processing specification for an operation.

Object Type: A specification of a type of entity, both structure (attributes) and operations (functions) are specified for object types; the notion of object type is a design notion being implemented as a "class".

Operation: A process related notion in UML; operations cause state changes in objects.

Packages: Similar to a library, but just containing classes.

Patterns: A software library for a common business scenario. A framework may be a design framework (possibly expressed in UML [Unified Modeling Language]) or an implementation framework (possibly in C++, Java, or PHP).

Polymorphism: The ability of object-oriented programs to have multiple implementations of the same method name in different classes in an inheritance tree. Derived classes can override the functionality defined in their base class.

Relationship: A connection concept between object types. There are several types of relationships including: aggregation, composition, association, and inheritance (generalization/specialization).

Reuse: Reuse (software) is a process where a technology asset (such as a function or class) is designed and developed following specific standards, and with the intent of being used again.

Separation: The separation of what a method does (interface) from how the method does it (implementation).

Stereotype: The core extension mechanism of UML.

Structural Model: An UML model describing static structure (relationships and properties).

Trigger: One operation invoking another operation; a call from one method to another method within the same or different object type (class).

UML: Unified Modeling Language

Business Model Innovation in the Digital Economy

Chung-Shing Lee
Pacific Lutheran University, USA

INTRODUCTION

Most Internet ventures failed because they did not have viable business models and sustainable long-term strategies. Their business models failed to satisfy the two fundamental questions associated with the characteristics of the Digital Economy (Lee & Vonortas, 2004):

- Does your organization's business model follow the fundamental economic principles of the Digital Economy? That is, what is the underlying economic logic that explains how your organization can deliver value to customers at an appropriate cost?
- Does your organization's business model capitalize on the "disruptive attributes" of the Digital Economy? That is, how can you organization capture the full benefits of the Internet innovation?

These two fundamental questions lead business executives to consider several strategic questions regarding the implementation of an innovative business model in the Digital Economy.

- What are the functions and components of a viable business model in the Digital Economy?
- What are the disruptive attributes of the Internet innovation and how does an organization capitalize on them for competitive advantage and profits?
- What are the differences between the traditional organizational transformation process and the value creation process in the Digital Economy?
- How do transaction costs and network effects in the Internet economy change a company's competitive position?
- How do the cost and revenue structures in the Digital Economy differ from in the traditional industrial economy?

BACKGROUND

A business model is the method of doing business by which a company can generate revenue to sustain itself (Rappa, 2003; Turban, King, Lee & Viehland, 2004). It describes the basic framework of a business. It also tells what market segment is being served (who), the service that is being provided (what), the means by which the service is produced (how) (Chaudhury & Kuilboer, 2002), and how it plans to make money long term using the Internet (Afuah & Tucci, 2003, p. 51).

A firm's business model should also spell out how the company is positioned in the value chain or within the business ecosystem. Weill and Vitale (2001) define an e-business model as a description of the roles and relationships among a firm's consumers, customers, allies, and suppliers that identifies the major flows of product, information, and money, and the major benefits to participants. Timmers (1998) defines business model as an architecture for the product, service, and information flows, including: a description of the various business actors and their roles, a description of the potential benefits for the various business actors, and a description of the sources of revenues.

A business model consists of multiple components and performs different functions. A "new economy" business model requires four choices on the part of senior management, argued by Rayport and Jaworski (2001). These include the specification of a value proposition or a value cluster for targeted customers; a scope of marketspace offering, which could be a product, service, information, or all three; a unique, defendable resource system, that is, the associated resource system to deliver the benefits; and a financial model, which includes a firm's revenue models, shareholder value models, and future growth models. In a similar effort, Chesbrough and Rosenbloom (2002) identify the functions of a business model as: 1) to articulate the value proposition; 2) to identify a market segment; 3) to define the structure of the firm's value chain; 4) to specify the revenue generation mechanisms(s) for the firm; 5) to describe the position of the firm within the value network; and 6) to formulate the competitive strategy to gain advantage over rivals. Other efforts to bring together the various lines of thought and to establish a common denominator for the business model discussion include Alt and Zimmermann (2001) and Dubosson-Torbay, Osterwalder, and Pigneur (2002).

Rappa (2003) identifies nine basic Internet business models: brokerage, advertising, infomediary (e.g., recommender system, registration model), merchant, manufacturer (direct marketing), affiliate (provide commission for online referrals), community (voluntary contributor model or knowledge networks), subscription, and utility (e.g., pay by the byte). In addition, Turban et al. (2004) also identify several types of Internet business models, including name your price, find the best price, dynamic brokering, affiliate marketing, group purchasing, electronic tendering systems, online auctions, customization and personalization, electronic marketplaces and exchanges, supply chain improvers, and collaborative commerce.

In order to sustain a successful business venture, a viable business model should address a number of issues and the dynamics of the respective elements which include: what value to offer customers (strategic goals and value proposition); which customers to provide the value to (scope of offerings); how to price the value (pricing); how much and who to charge for it (revenue models); quantity of resources required and the associated costs to provide the value; what strategies, structures, and processes need implementing to offer value; and the legal issues that may influence the general vision of the business model (Alt & Zimmermann, 2001). In addition, in order to prosper in e-commerce, a firm's Internet business model must capitalize on the "disruptive" attributes and characteristics of the Internet or Digital Economy to enable it to offer innovative solutions and value to customers.

BUSINESS MODEL INNOVATION

Although the changes made possible by the Internet are strategic and fundamental (Ghosh, 1998), the underlying technologies are not radically different from the existing technologies that support business operations. Computing and communication technologies, which are the technological foundations for the Internet, have both been improved incrementally over the past few decades. Bower and Christensen (1995) argue that the technological changes that damage established companies are usually not radically new or difficult from a technological point of view. However, the Internet is considered a disruptive innovation to many businesses. It is disruptive to the traditional way of doing business in that it is transforming the rules of competition and inventing new value propositions and business models. The successful implementation of a viable business model in the Digital Economy requires a paradigm shift. In moving toward e-commerce as an enabler, a business executive must be able to identify the disruptive nature of the innovation and then capture the benefits. Table 1 lists several disruptive attributes of the Internet and e-commerce identified by Lee (2001). Organizations in the Digital Economy must understand and capitalize on the disruptive attributes of the Internet and e-commerce to transform their business models for success.

Business model innovation in the Digital Economy is the use of new knowledge (both technological and market) that capitalizes on the disruptive attributes of the Internet to design and implement an innovative way of offering products or services that customers want. Examples of successful business model innovations that were able to capitalize on some of the disruptive attributes of the Internet include channel innovation (e.g., Dell's build-to-order virtual integration model), process innovation (e.g., Boeing virtual design and e-procurement PART page), customer experience innovation (e.g., Schwab), auction and reverse auction model (e.g., eBay and Priceline), online virtual community (e.g., iVillage), customer-relationship (e.g., Yahoo!), and affiliate network (e.g., Amazon).

VALUE CREATION IN THE DIGITAL ECONOMY

In a competitive market environment, businesses exist to create value for their customers. To understand how a business creates value, a simple "input-transformation-output" model can be utilized to describe the process of value creation. In the Industrial Economy, inputs to a value creation process are raw materials or all of the necessary physical inputs that are required to produce the finished products or services. Outputs are finished products or intermediate goods used as inputs to another transformation or value creation process. Information, such as design and engineering know-how as well as production methods and procedures, is applied to facilitate the "physical" transformation process, which involves one or more of the four value-adding activities: alter, transport, inspect, and store (Meredith & Schaffer, 1999). Management's focus is to make the transformation process more efficient by implementing techniques such as lean manufacturing, total quality management, and business process re-engineering. In contrast, input to the value creation process in the Digital Economy is information (e.g., customer profiles and preferences, i.e., the digital assets, as well as production and distribution status) that firms gather, organize, select, synthesize, and distribute (Rayport & Sviokla, 1995) in the transformation process to provide individual customers a bundle of customized solutions. In the Digital Economy, information is a source of value, and every business is an information business (Earl, 1999). Since physical and digital economies co-exist within a firm or supply chain, management should go beyond concentrating on improving the

B

Table 1. Summary of disruptive attributes of the Internet and e-commerce

Open Platform. Internet provides an open and nonproprietary platform for communication and collaboration. The open source movement in software development (e.g., Raymond, 1999) has contributed to Internet-enabled collaboration and free information sharing.

Network Effects. Network effects exist in the industrial economy (e.g., regular telephone service) but are much stronger in the Digital Economy. For knowledge-intensive products, such as software operating systems, characterized by high upfront costs, network effects, and customer groove-in (Arthur, 1996), achieving a critical mass of installed customer base is vital for success.

Connectivity and Interactivity. E-commerce enables close connections with customers and among supply chain or business ecosystem partners' information systems. The benefits include real-time pricing, flexible products and services versioning, gathering customer information, and a very low cost for the distribution of information goods.

Information Sharing and Exchange. In the Digital Economy, the traditional trade-off between richness and reach in information exchange no longer exists (Evans & Wurster, 1997). Information can reach many customers or business ecosystem partners through the Internet without sacrificing the richness of the contents.

Prosumption. *Prosumption* (Tapscott, 1996) is the term to describe the convergence of design with development process, and the production of goods and services by customers in the e-commerce environment. Internet-enabled collaborations can reduce both concept-to-design and design-to-production cycle times.

Digital Assets. Digital assets are information about customers (e.g., purchasing patterns and profiles). A firm that exploits the Internet should build and utilize its digital assets in order to provide customer value across many different and disparate markets. In the Digital Economy, information is a source of revenue, and every business is an information business (Earl, 1999). A firm should use information to create new businesses and/or reinvent customer relationships through the implementation of a virtual value chain (Rayport & Sviokla, 1995).

Cost Transparency. The vast amount of information about prices, competitors, and features that is readily available on the Internet helps buyers "see through" the costs of products and services (Sinha, 2000).

Industry Scope. Value generated in Internet-enabled business transcends traditional industrial sectors. A firm or business ecosystem (Gossain & Kandiah, 1998) or business Web (Tapscott, Ticoll & Lowy, 2000) must provide unique (and customized) "solutions" (as opposed to single product or service) to individual customers.

Speed and Frequency of Changes. Change is fast and frequent in the Digital Economy. Firms in every industry must learn to adapt quickly to changing business and economic environments. Arthur (1996) states that adaptation in a turbulent environment means watching for the next wave that is coming, figuring out what shape it will take, and positioning the company to take advantage of it.

Virtual Capacity. The advance in network and storage technologies gives customers the feeling that it has infinite virtual capacity to serve them (Afuah & Tucci, 2003). General online merchandise stores (e.g., Amazon.com) and all-purpose business-to-business mega exchanges can offer enormous variety without building huge physical display areas that rack up costs and alienate many shoppers. In addition, virtual communities have infinite capacity for members to participate in discussions and sharing of information anywhere and anytime for as long as they want.

transformation process itself, to focus on leveraging information assets and capitalize on the disruptive features of the Internet and e-business to create more value for the customers. Table 2 compares the transformation processes, and Table 3 presents organizational goals and value creation strategies to assist business executives in designing and implementing a viable business model in the Digital Economy.

COST AND REVENUE MODELS

Rather than searching for the single dominant effect or cost advantage that will provide a long-term sustainable competitive advantage for a company, companies wishing to take full advantage of the disruptive power of Internet commerce must understand the underlying economic logic of the Digital Economy (Lee & Vonortas, 2004). There are four cost factors: scale, scope, switching costs, and transaction costs from both the demand and supply sides that must be examined to determine the viability of the business model.

Internet commerce and virtual value chain have redefined the concepts of economies of scale, allowing small companies to achieve low unit costs in markets dominated by big companies. In addition, online mega store or exchange models are able to spread fixed costs over a larger customer base and offer a wide selection of goods. On the demand side, marketing strategies in markets in which strong network effects exist must be designed to influence customer expectations in order to achieve a critical mass of users.

In the Digital Economy, companies can not only identify and take advantage of the economies of scope in production and distribution, they can also redefine economies of scope by drawing on a single set of "digital assets" to provide value across many different and disparate markets (Rayport & Sviokla, 1995). The combination of demand-side scale and scope economies reinforces network effects in the Digital Economy. For example, Amazon and Yahoo! are able to expand their scope of offerings to provide customers a package of "solutions" across numerous industrial sectors due to the ability to exploit its large installed based of customers.

In the Digital Economy, strategies to increase trading partners' and customers' switching costs include prosumption (e.g., Dell's direct order model), virtual communities (e.g., fool.com), an extranet that allows partners access trade-specific information (Riggins & Rhee, 1998), and developing a strong trust relationship with the end-customers by participating in business ecosystems (Gossain & Kandiah, 1998) or b-Webs (Tapscott et al., 2000).

The Internet has dramatically reduced the costs of many kinds of market transactions. It could prove easier and cost effective to disaggregate many value-creating activities out to the open market so companies can concentrate on their core competences. Companies must also reduce customers' (transaction) costs of doing business with them by making it easy for customers to obtain and

Table 2. Comparison of organizational transformation and value creation processes

	Traditional Economy	**Digital Economy**
Input:	raw materials	Information in digital form
Transformation: (Value Creation)	inspect, alter, transport, and, store	Gather, organize, select, synthesize, and distribute
Output:	Intermediate or finished products/services	Knowledge-or solution-based services
Roles of Information:	A supporting element that facilitates the physical transformation process	A source of value that enable that enables firms to offer new value propositions
	It serves as the connection between the various value-adding stages within a supply chain	Real-time information enables supply chain integration, collaboration, and synchronization

Table 3. Organizational goals and value creation strategies in the Digital Economy

INPUT

Organizational Goals:

- Design innovative products and services to meet customer's latent needs
- Reach the critical mass by building an installed base of customer
- Take advantage of the network effects to build a successful e-business

Management Strategies:

- Increase collaboration throughout the product design process
- Synchronize product design requirements within a supply chain or business ecosystem in the early stages of the development process
- Leverage knowledge capital critical to the design process through external linkages (e.g., alliance partners, research labs and universities)
- Achieve demand-side economies of scale by increase installed customer base's collective switching costs
- Reduce customer's transaction costs (i.e., make it easy for customers to do business with you)

TRANSFORMATION VALUE ACTIVITIES

Organizational Goals:

- Improve efficiency and effectiveness of business or supply chain transformation process
- Match the performance of the physical activities to the digital world

Management Strategies:

- Sustaining innovation approach: Apply conventional management techniques (e.g., lean manufacturing and total quality management) to improve process efficiency
- Achieve production economies of scale and scope
- Disruptive innovation approach: Take advantage of the lower transaction costs in the digital economy to redesign organizational structures and to reconfigure value creation systems

OUTPUT

Organizational Goals:

- Create numerous innovative knowledge- or solution-based products and services
- Provide a package of solutions to satisfy or exceed customer's expectations

Management Strategies:

- Transform value proposition by taking advantage of the network effects and demand-side economies of scope, i.e., leverage on a single set of "digital assets" to provide many solutions to individual customers
- Increase user's (or buyer's) switching costs by offering value across many different and disparate markets
- Influence users' decision-making process through the use of extranet and Web-based collaborative planning tools
- Look beyond costs as the sole arbiter of value

compare product-related information, identify customer decision process, and be able to provide assistance.

Internet commerce provides companies with new sources of revenues via opportunities to offer new "customized" services or solutions in addition to, and sometimes independent of, the traditional products or services sales. In the Digital Economy, pricing can be done in real time. Product can even price below unit cost (in the long run) as long as other e-commerce revenue models, such as online advertising and referral fees, are sustainable. Internet business models also offer companies new opportunities to test prices, segment customers, and adjust to changes in supply and demand. They can create enormous value in the process because online pricing allows companies to make adjustments in a fraction of the time and to profit from even small fluctuations in market conditions, customer demand, and competitors' behavior (Baker, Marn & Zawada, 2001).

Finally, Tapscott et al. (2000) and Gossain and Kandiah (1998) argue that in the new business ecosystems or b-Webs, customers perceive greater value in one-stop shopping with a known and trusted company. Customers receive "a package of solutions" that satisfies their explicit and implicit needs, rather than purchasing individual products from different vendors. For example, Internet companies, such as Amazon.com, Edmunds.com, and Marshall Industries, are perceived as a single "trusted" source to offer complementary products and services provided by their ecosystem partners to end-customers.

MANAGERIAL IMPLICATIONS

This article identifies major elements and provides specific guidelines to assist organizations in designing an innovative business model in the Digital Economy. A viable business model in the Digital Economy must transform value propositions, and organizational structures and systems to enhance value creation (see Tables 2 and 3). It must be able to take advantage of the Internet network effects and other disruptive attributes (see Table 1) to achieve and sustain a critical mass of installed base of customer. A viable business model in the Digital Economy must also develop trust relationships with business ecosystem partners and customers through virtual communities to reduce their costs of doing business with the company, and to increase their costs of switching to other vendors. Companies must build and maintain a large set of digital assets and leverage them to provide value across many different and disparate markets. Finally, companies must identify customers' latent needs and transform their business models from a product- or component-based model to a knowledge- or solution-based model.

FUTURE TRENDS

Future research on the analysis, design, development, implementation, and controlling of business model innovation in the Digital Economy include, but are not limited to the following areas:

- Analytical and architectural frameworks for new business models
- Models and modeling techniques
- Specific details and guidelines for business model innovation
- Specific industry perspectives on business models (e.g., converging of technologies and industries, media industry, and communication industry)
- Sustainable business models for digital contents, mobile commerce, collaborative commerce, peer-to-peer architecture
- Trust and business model innovation in business ecosystems

CONCLUSION

A viable business model must follow the fundamental economic principles outline in this article (i.e., costs and revenue models). Most important, an innovative business model in the Digital Economy must capitalize on the disruptive attributes of Internet commerce. E-business initiatives can be implemented as a sustaining innovation to enhance the current way of doing business (i.e., focus primarily on improving the efficiency of organizational transformation). In such case, companies fail to identify and capitalize on many of the Internet's opportunities. Business model innovation in the Digital Economy requires business planners and strategists to recognize and capture the full benefits of the disruptive characteristics of the Internet and e-commerce.

REFERENCES

Afuah, A. & Tucci, C.L. (2003). *Internet business models and strategies: Text and cases* (2nd Edition). New York: McGraw-Hill/Irwin.

Alt, R. & Zimmermann, H. (2001). Preface: Introduction to special section—business models. *Electronic Markets, 11*(1), 3-9.

Arthur, W.B. (1996). Increasing returns and the new world of business. *Harvard Business Review*, (July-August), 100-109.

Baker, W., Marn, M. & Zawada, C. (2001). Price smarter on the Net. *Harvard Business Review*, (February), 122-127.

Bower, J.L. & Christensen, C.M. (1995). Disruptive technologies: Catching the wave. *Harvard Business Review*, (January-February), 43-53.

Chaudhury, A. & Kuilboer, J. (2002). *E-Business and e-commerce infrastructure: Technologies supporting the e-business initiative*. New York: McGraw-Hill/Irwin.

Chesbrough, H. & Rosenbloom, R. (2002). The role of the business model in capturing value from innovation. *Industrial and Corporate Change*, *11*(3), 529-556.

Dubosson-Torbay, M., Osterwalder, A. & Pigneur, Y. (2002). E-business model design, classification, and measurements. *Thunderbird International Business Review*, *44*(1), 5-23.

Earl, M.J. (1999). Strategy-making in the information age. In W.L. Currie and B. Galliers (Eds.), *Rethinking management information systems* (pp. 161-174). New York: Oxford University Press.

Evans, P.B. & Wurster, T.S. (1997). Strategy and the new economics of information. *Harvard Business Review*, (September-October), 70-82.

Ghosh, S. (1998). Making business sense of the Internet. *Harvard Business Review*, (March-April), 126-135.

Gossain, S. & Kandiah, G. (1998). Reinventing value: The new business ecosystem. *Strategy & Leadership*, *26*(5), 28-33.

Katz, M.L. & Shapiro, C. (1986) Technology adoption in the presence of network externalities. *Journal of Political Economy*, *94*(4), 822-841.

Lee, C.-S. (2001). An analytical framework for evaluating e-commerce business models and strategies. *Internet Research*, *11*(4), 349-359.

Lee, C.-S. & Vonortas, N.S. (2004). Business model innovation in the digital economy. In G. Doukidis, N. Mylonopoulos & N. Pouloudi (Eds.), *Social and economic transformation in the digital era* (pp. 164-181). Hershey, PA: Idea Group Publishing.

Meredith, J.R. & Schaffer, S.M. (1999). *Operations management for MBAs*. New York: John Wiley & Sons.

Miller, W.L. & Morris, L. (1999). *Fourth generation R&D*. New York: John Wiley & Sons.

Rappa, M. (2003). Business models on the Web. *Managing the digital enterprise*. Retrieved from *digitalenterprise.org/models/models.html*

Raymond, E.S. (1999). *The cathedral & the bazaar: Musings on Linux and Open Source by an accidental revolutionary*. Sebastopol, CA: O'Reilly & Associates.

Rayport, J.F. & Jaworski, B.J. (2001). *E-commerce*. New York: McGraw-Hill/Irwin.

Rayport, J.F. & Sviokla, J.J. (1995). Exploiting the virtual value chain. *Harvard Business Review*, (November-December), 75-85.

Riggins, F.J. & Rhee, H. (1998). Toward a unified view of electronic commerce. *Communications of the ACM*, *41*(10), 88-95.

Tapscott, D. (1996). *The Digital Economy: Promise and peril in the age of networked intelligence*. New York: McGraw-Hill.

Tapscott, D., Ticoll, D. & Lowy, A. (2000). *Digital capital: Harnessing the power of business webs*. Boston: Harvard Business School Press.

Timmers, P. (1998). Business models for electronic markets. *Electronic Markets*, *8*(2), 3-7.

Tirole, J. (1988). *The theory of industrial organization*. Cambridge, MA: MIT Press.

Turban, E., King, D., Lee, J. & Viehland, D. (2004). *Electronic commerce: A managerial perspective*. Upper Saddle River, NJ: Prentice-Hall.

Weill, P. & Vitale, M.R. (2001). *Place to space: Migrating to e-business models*. Boston: Harvard Business School Press.

KEY TERMS

Business Model: The method of doing business by which a company can generate revenue to sustain itself (Rappa, 2003).

Digital Assets: The information (in digital form) a company collected about its customers. Companies that create value with digital assets may be able to reharvest them through a potentially infinite number of transactions (Rayport & Sviokla, 1995).

Digital Economy: The economy for the age of networked intelligence (Tapscott, 1996). The Digital Economy is also a knowledge economy. Information, in all its forms becomes digital, is the input of organizational transformation or value creation process.

Disruptive Innovations: Innovations that typically present a different package of performance attributes—ones that, at least at the outset, are not valued by existing customers. The performance attributes that existing customers value improve at such a rapid rate that the new innovation can later invade those established markets (Bower & Christensen, 1995). Innovations are disruptive when the current organization lacks the necessary models of competitive architecture and organizational capabilities, and are therefore unable in critical ways to do what must be done (Miller & Morris, 1999).

Economies of Scale: *Supply-side economies of scale*—Reductions in unit costs resulting from increased size of operations. *Demand-side economies of scale*—The value of a technology or product increases exponentially as the number of users increase (network effects lead to demand-side economies of scale).

Economies of Scope: *Supply-side economies of scope*—Cost of the joint production of two or more products can be less than the cost of producing them separately. *Demand-side economies of scope*—A single set of digital assets can provide value for customers across many different and disparate markets.

Network Effects or Network Externalities: A technology or product exhibits network externalities when it becomes more valuable to users as more people take advantage of it (Katz & Shapiro, 1986).

Switching Costs: Investment in multiple complementary and durable assets specific to a particular technology or system (Tirole, 1988).

Transaction Costs: Costs associated with contractual relationships in a market economy, such as costs that the consumer or the producer pays to make the market transaction happen.

B

Business Modeling with Client–Oriented Requirements Strategy

María Carmen Leonardi
UNCPBA, Argentina

INTRODUCTION

Rational unified process (RUP) (Jacobson, Booch, & Rumbaugh, 1999) is an iterative, incremental and use case driven methodology. RUP starts the software development with the requirements capture stage, taking into account that *"the major challenge is that the customer, who we assume to be primarily a non-computer specialist, must be able to read and understand the result of requirements capture. To meet this challenge we must use the language of the customer to describe these results"* (Jacobson et al., 1999, p. 113). As requirements are originated from the system's context, RUP proposes the definition of it through a business model, more concretely, a *business use cases model* and a *business objects model*. There are several approaches to enhance this first stage of the RUP development. In this article, the author describes the most important proposals and briefly presents her strategy that defines a set of activities and heuristics to define a UML conceptual object model starting from stakeholder oriented requirements models. These models describe the overall context in which the software will be developed and operated, known as universe of discourse (Leite & Leonardi, 1998). The strategy enhances traceability (Pinheiro, 2000) between requirements and RUP models.

BACKGROUND: BUSINESS MODELING IN THE CONTEXT OF RUP

As mentioned earlier, RUP proposes a business model to define the organization without defining any explicit techniques or strategy to guide the construction of it. There are some works that present different proposals to enhance this stage, taking into account different starting points and perspectives. In this section, the author briefly describes the most important.

From the RUP/UML and business technology community, there are some proposals to model the organization. The work of Eriksson and Penker (2000) presents a combination of techniques to model the business with UML organization models representing processes, events, resources, goals, business rules and general mechanisms. The business architecture and its elements are represented from four different views: business view, business process, business structure and business behavior. They propose a set of UML business extensions to represent those concepts using the standard UML extension mechanisms. The authors propose three categories of business patterns to describe common modeling solutions during the business modeling: resource and rules patterns, process patterns and goals patterns. In Barros, Duddy, Lawley, Milosevic, Raymond, and Wood (2000), an interesting proposal is presented extending UML in order to define the organizational enterprise model for systems that will be implemented using distributed objects technology. This model is described in terms of processes, entities, lists and events of the business. Finally, in Marshall (1999), some key concepts of an enterprise and their components are modeling. Specifically, they define the purpose, processes, entities and organization of the enterprise with standard UML diagrams.

From the requirement community, one of the most important works is Santander and Castro (2002). This strategy allows the definition of use cases starting from the i* organizational model. This strategy captures organizational requirements to define how the system fulfills the organization goals, why it is necessary, what the possible alternatives are and what the implications to the involved parts are, all of them represented by the two i* models. The approach presents some guidelines to develop use cases from these models. This approach is a goal-based methodology useful for RUP software requirements definition phase since it generates use cases for the software system.

Although it is not directly related to RUP business modeling, one of the most important works in generating conceptual object-oriented specifications from natural language is the strategy presented in Juristo, Moreno, and López (2000). This strategy consists of analyzing natural language-based descriptions (expressing the problem to be solved) from a syntactic and semantic perspective in order to identify the key elements of the object-oriented conceptual model (modeling the problem in the computer). The proposal is a semi-formal model that lets organizations systematically produce object-oriented models. It is based on two components: a mathematical component that defines formal rules to identify the elements of the object-oriented conceptual models from the natural language structures; and a method that guides the analyst in the development of such models, by means of the definition of a set of defined steps.

A CONSTRUCTION PROCESS OF THE RUP BUSINESS MODEL BASED ON STAKEHOLDER-ORIENTED REQUIREMENTS MODELS

During the early stages of development, when the interaction with the stakeholders is crucial, the use of natural language-oriented requirements engineering techniques seems necessary in order to enhance communication. These techniques are very useful to obtain a first specification of the universe of discourse that will be easily validated with the stakeholders and will be the basis for a development. Therefore, the necessity of integrating stakeholder-oriented requirements models and strategies to enhance the construction process of the business model keeping the RUP philosophy of using the customer language is highlighted. Stakeholder-oriented models are based on natural language, therefore the communication between engineers and stakeholders are facilitated. The author presents a strategy based on heuristics that guide the construction of the RUP starting from the stakeholders-oriented requirements models. Due to space reason, the author does not present the full strategy, which may be found in Leonardi (2003). The section is organized in three subsections: one to present the stakeholder-based models, the second presents the construction process and finally, in the third, its use is discussed.

Stakeholder-Oriented Requirements Models

The models presented in this section are well-known, used and accepted by the requirements engineering community (Leite et al., 1997; Leite & Leonardi, 1998). In this proposal, they are used as the first models to obtain a RUP business object model. The models are: language extended lexicon model, scenario model and business rules models.

- **Language Extended Lexicon:** The language extended lexicon (LEL) (Leite et al., 1997) is a structure that allows the representation of significant terms of the universe of discourse. The purpose of the lexicon is to help understand the vocabulary and its semantics, leaving the comprehension of the problem for a next step. It unifies the language allowing communication with the stakeholder. LEL is composed of a set of symbols with the following structure: *symbol name* “ word or phrase and set of synonyms; *notions* defining the denotation of the symbol; and *behavioral responses* describing the symbol connotation. In the description of the symbols, two rules must be followed simultaneously: the “closure principle” that encourages the use of LEL symbols in other LEL symbols forming a graph, and the “minimum vocabulary principle” where the use of symbols external to the application language is minimized. LEL terms define objects, subjects, verbal phrase and states. Figure 1 shows the heuristics to define each type of symbol.
- **Scenario Model:** A scenario describes situations of the universe of discourse (Leite et al., 1997). A scenario uses natural language as its basic representation, and it is connected to LEL. Figure 2

Figure 1. Heuristics to represent LEL terms

Subject	Notions: who the subject is.
	Behavioral responses: register actions executed by the subject.
Object	Notions: define the object and identify other objects with which the former has a relationship.
	Behavioral responses: describe the actions that may be applied to this object.
Verb	Notions: describe who executes the action, when it happens, and procedures involved in the action.
	Behavioral responses: describe the constraints on the happening of an action, which are the actions triggered in the environment and new situations that appear as consequence.
State	Notions: what it means and the actions which triggered the state.
	Behavioral responses: describe situations and actions related to it.

describes the components. In Leite, Hadad, Doorn, and Kaplan (2000), the scenario construction process is described.

- **Business Rule Model:** We define business rules as statements about the enterprise's way of doing business, reflecting business policies (Leite & Leonardi, 1998). Organizations have policies in order to: satisfy the business objectives, satisfy customers, make good use of resources, and conform to laws or general business conventions. The business rule model distinguishes between functional rules and non-functional rules. *Functional rules* are general policies regarding organization functionality. Non-functional rules are divided in *macrosystem rules* that describe policies that constrain the behavior and structure of the organization and *quality rules* that are demands of the organization on the characteristics of its processes or products. Rules are built following some syntax patterns, in which the main components must be a LEL symbol. In Leonardi, Leite, and Rossi (1998), the authors describe the construction process.

THE CONSTRUCTION PROCESS OF THE RUP BUSINESS MODEL

Once the models are constructed and validated with the stakeholders, they have to be manipulated in order to be mapped to a RUP business model, more concretely a business use case model and a business object model. The strategy is organized in four stages, defining for each one a set of heuristics that guide the definition of the different models enhancing traceability. The stages are:

- **Business Use Cases Model Definition:** The business use case starting from the scenario model is described; each scenario will be a business use case. As there is a well-defined process for the scenario construction, the use cases obtained are a representative and validated group defining the behavior of the main actors of the organization.
- **Business Classes Identification and Definition:** The proposed heuristics in this stage guide the identification of the classes starting from the LEL symbols, taking into account the taxonomy of these symbols in order to define them as classes or responsibilities (see Figure 1). It is important to remark that when the universe of discourse is analyzed, active and passive entities can be distinguished. This distinction is reflected in the classification as actor or resource considered in the scenario model and as object or subject considered in the LEL. As business classes define the universe of discourse, they will have the same distinction. Therefore, primary and secondary classes can be considered. Although from the object-oriented paradigm there is no distinction between the kind of objects, heuristics, to define them, are different because inside the universe of discourse, they have different semantics; the primary classes are those that represent active entities, and the secondary classes are the ones that represent passive or resources entities. Therefore, we propose different heuristics for defining them: primary classes are defined from the subject LEL terms and some of the verbal phrase LEL terms; and secondary classes are derived principally from object LEL terms. Responsibilities of the classes are defined taking into account the behavioral responses of the corresponding LEL terms. As each business use case is a detailed description of a particular situation, it is used to complete the responsibilities and collaborations between the classes that appear in it as LEL terms. This is possible due to the traceability between LEL terms and classes. Finally, the functional business rules model is utilized to add more responsibilities to the defined classes or to create new ones if the relevance of a set

Figure 2. Components of scenario

Title: identifies a scenario.
Objective: describes the purpose of the scenario.
Context: defines geographical and temporal locations and preconditions.
Resources: identify passive entities with which actors work.
Actors: detail entities actively involved in the scenario.
Set of episodes: each episode represents an action performed by actors using resources.

Note: Notions of c*onstraint* and e*xception* are added to some of the components of a scenario: a *constraint* refers to non-functional requirements and may be applied to context, resources or episodes. An *exception*, applied to episodes, causes serious disruptions in the scenario, asking for a different set of actions which may be described separately as an exception scenario.

of related business rules justifies the creation of a new class to represent them.

- **Business Object Diagram Definition:** Once the classes and their responsibilities have been identified, structural aspects of the classes are modeled (methods, attributes, relationships) defining a UML object diagram. Attributes are defined taking into account the notions of the corresponding LEL terms. Methods are derived from the previously defined responsibilities; the heuristics consider the possibility of split responsibilities in several methods or group some of them into a single method, with the aim of enhancing the quality of the interface of the defined classes. LEL terms corresponding to the defined classes are analyzed in order to define inheritance, part-of and associations relationships between them. A set of heuristics is also defined to attach non-functional rules to the involved class. Rules are modeled as OCL expressions (Warmer & Kleppe, 2003) associated to methods, attributes or associations. The business object diagram defines the universe of discourse, and as such, it does not present software limits, leaving this task for the analysis stage where the actors of the system are defined and the use cases of the software system are identified from the business use cases.
- **Business Use Cases Model Realization:** The next step is to describe the *business use cases* based on the previously defined classes. Business use cases realizations may be defined by means of interaction diagrams (collaboration or sequence) or activity diagrams. The heuristics may be used as an informal validation mechanism, because by applying them, the business use cases are "executed", and the engineer may analyze if the defined classes and their responsibilities perform the behavior of them.

Analyzing the Use of the Strategy

The strategy was used in several case studies. From this experience, some advantages of it are mentioned:

- Natural language allows communication with stakeholders, enhancing models validation.
- LEL and business rules models allow engineer to define objects from a global perspective, enhancing the specific behavior defined in a use case.
- From the object-oriented perspective, this strategy improves the analysis phase, by the incorporation of stakeholder-oriented techniques.
- From the requirements perspective, this strategy extends the requirements capture phase to the definition of an object-oriented conceptual model that allows engineer to visualize the problem in terms of objects, the same concept that he will use in the rest of the development process. A conceptual object model is obtained that serves as a basis for RUP oriented development, reducing the gap between analysis and design phases allowing that transition between conceptual modeling and design phase results in a more natural process (Kotonya & Sommerville, 1998).
- This strategy identifies and maintains the trace relationships generated by the application of the heuristics, allowing traceability between the RUP business model and LEL, scenarios and business rules models. It also defines internal RUP models traces and internal requirements models traces. Each trace relationship links two or more model components, defining its semantics, that is, the origin component, the source, cardinality and the kind of relationship between the components.

The main difficulties found during the application of the strategy are:

- One of the most significant problems is the level of detail of the specifications. Depending on the level, more or less classes can be found. But, this problem is inherent to specifications, independent on a particular strategy.
- It is very important to deal with the redundancy between the LEL, business rules and scenarios to avoid translating it to the object model.
- As the strategy generates a lot of traceability information, it is necessary to define what trace relationships are managed, since if all of the possible links are generated, the maintaining will be tedious, time consuming and labor-intensive.

FUTURE TRENDS

Integration between business modeling, requirements engineering techniques and the rest of the software development steps are essential. It is necessary to define some strategy to guide in the translation of the requirements models to an architectural model, in order to obtain a first description of an adaptable architecture that reflects the non-functional and functional requirements. New proposals are emerging from different areas with the objective of working in the integration (*STRAW*, 2003). Moreover, research must be set in the context of the OMG model driven architecture (MDA, 2004), to define business models that can be used as the computer independent model (CIM) and requirements model that can be used as the platform independent model (PIM) integrated to a completely MDA software development.

CONCLUSION

In this work, the author presented a brief survey of proposals that enhance the first stage of RUP development where the RUP business model is defined through the business use cases and the business-objects models. The author also presented her strategy based on well-known and used stakeholder-oriented requirements models to define this model. This strategy is compatible with the aims of RUP of enhancing the interaction with the customer by the use of easy reading and understanding models (Jacobson, Booch, & Rumbaugh, 1999). As use cases model is the key model of RUP, our strategy has incorporated them by means of scenario representation, but it has also added complementary models to enhance the knowledge of the business organization. LEL model and the business rules model also increase the abstraction degree of object-oriented modeling, allowing engineers to define objects from a global perspective, enhancing the specific behavior defined in a use case. The strategy proposes a set of heuristics to guide in the manipulation of the great quantity of information generated by the requirements models to define the RUP business model. A trace mechanism is defined in order to enhance traceability between models.

REFERENCES

Barros, A., Duddy, K., Lawley, M., Milosevic, Z., Raymond, K., & Wood, A. (2000). Processes, roles and events: UML concepts for enterprise architecture. *Proceedings UML 2000: The Unified Modeling Language. Third International Conference* (pp. 62-77). York, UK: Springer.

Booch, G., Rumbaugh, J., & Jacobson, I.(1999). The unified language user guide. Reading, MA: Addison-Wesley.

Eriksson, H., & Penker, M. (2000). Business modeling with UML: Business patterns at work. New York: John Wiley & Sons.

Jacobson, I., Booch, G., & Rumbaugh, J. (1999). *The unified software development process*. Reading, MA: Addison-Wesley.

Juristo, N., Moreno, A., & López, M. (2000, May/June). How to use linguistic instruments for object-oriented analysis. *IEEE Software*, 80-89. Los Alamitos, CA: IEEE Computer Society Press.

Kotonya, G., & Sommerville, I. (1998). *Requirements engineering*. Chichester: John Wiley & Sons.

Leite, J.C.S.P., & Leonardi, C. (1998). Business rules as organizational policies. *Proceedings of IEEE Ninth International Workshop on Software Specification and Design* (pp. 68-76). Los Alamitos, CA: IEEE Computer Society Press.

Leite, J.C.S.P., Hadad, G., Doorn, J., & Kaplan, G. (2000). A scenario construction process. *Requirements Engineering Journal*, *5*(1), 38-61.

Leite, J.C.S.P., Rossi, G., Balaguer, F., Maironana, V., Kaplan, G., Hadad, G., & Oliveros, A. (1997). Enhancing a requirements baseline with scenarios. *Proceedings of IEEE Third International Requirements Engineering Symposium* (pp. 44-53). Los Alamitos, CA: IEEE Computer Society Press.

Leonardi, C. (2003). Enhancing RUP business model with client-oriented requirements models. In L. Favre (Ed.), *UML and the unified process* (pp. 80-115). Hershey, PA: IRM Press ISBN 1-931777-44-6, Chapter 6.

Leonardi, C., Leite, J.C.S.P., & Rossi, G. (1998). Estrategias para la identificación de Reglas de Negocio. *Proceeding of Simposio Brasilero de Engenharia de Software* (pp. 53-67). Sociedad Brasilera de Computacao, Maringa, Brasil.

Marshall, C. (1999). *Enterprise modeling with UML*. Reading, MA: Addison-Wesley.

MDA (2004). OMG model driven architecture. Retrieved September 17, 2004 from *http://www.omg.org/mda/*

Pinheiro, F.(2000). Formal and informal aspects of requirements tracing. *Anais III Workshop de Engenharia de Requisitos* (pp. 1-21). Rio de Janeiro.

Santander, V., & Castro, C. (2002). Deriving use cases from organizational modeling. *Proceedings of IEEE Joint International Requirements Engineering* (pp. 32-39). Los Alamitos, CA: IEEE Computer Society Press.

STRAW03 (2003). *Proceedings of Second International Workshop from Software Requirements to Architectures*. Retrieved September 17, 2004 from *http://se.uwaterloo.ca/~straw03.2003*

Warmer J., & Kleppe A. (2003). *The object constraint language: Getting your models ready for MDA* (2nd ed.). Reading, MA: Addison-Wesley.

KEY TERMS

Business Model: A model of a company that shows the company's function in the world, what it does, how and when. It is designed to serve the needs of one or more types of handlers, and it should contain the information these handlers need " no more and no less (Marshall, 1999).

Conceptual Model: A mental model that allows understanding and simplification of the problem. The purpose of domain modeling is to contribute to an understanding of the system's context, and thereby also to understanding of the system's requirements as they originate from this context. The system's internal way of solving this problem will be dealt with in the analysis, design and implementation workflows (Jacobson, Booch, & Rumbaugh, 1999).

Rational Unified Process: A use case driven, architecture-centric, iterative and incremental software development process that uses UML to describe the artifacts (Jacobson, Booch, & Rumbaugh, 1999).

Requirements: Defined during the early stages of a system development as a specification of what should be implemented. They are descriptions of how the system should behave, application domain information, constraints on the system's operation, or specifications of a system property or attribute. Sometimes they are constraints on the development process of the system (Kotonya & Sommerville, 1998).

Requirements Traceability: The ability to define, capture and follow the traces left by requirements on other elements of the software development environment and the traces left by those elements on requirements (Pinheiro, 2000).

Stakeholder: People or organizations who will be affected by the system and who have a direct or indirect influence on the system requirements. They include end-users of the system, managers and others involved in the organizational processes influenced by the system, engineers responsible for the system development and maintenance, customer of the organization who will use the system to provide some services, external bodies such as regulators or certification authorities (Kotonya & Sommerville, 1998).

Unified Modeling Language (UML): A standard graphical language for visualizing, specifying, constructing, and documenting the artifacts of a software-intensive system (Booch, Rumbaugh, & Jacobson, 1999).

B

Business Process and Workflow Modeling in Web Services

Vincent Yen
Wright State University, USA

INTRODUCTION

In large organizations, typical systems portfolios consist of a mix of legacy systems, proprietary applications, databases, off-the-shelf packages, and client-server systems. Software systems integration is always an important issue and yet a very complex and difficult area in practice. Consider the software integration between two organizations on a supply chain; the level of complexity and difficulty multiply quickly. How to make heterogeneous systems work with each other within an enterprise or across the Internet is of paramount interest to businesses and industry.

Web services technologies are being developed as the foundation of a new generation of business-to-business (B2B) and enterprise application integration (EAI) architectures, and important parts of components as grid (*www.grid.org*), wireless, and automatic computing (Kreger, 2003). Early technologies in achieving software application integration use standards such as the common object request broker architecture (CORBA) of the Object Management Group (*www.omg.org)*, the distributed component object model (DCOM) of Microsoft, and Java/RMI, the remote method invocation mechanism. CORBA and DCOM are tightly coupled technologies, while Web services are not. Thus, CORBA and DCOM are more difficult to learn and implement than Web services. It is not surprising that the success of these standards is marginal (Chung, Lin, & Mathieu, 2003).

The development and deployment of Web services requires no specific underlying technology platform. This is one of the attractive features of Web services. Other favorable views on the benefits of Web services include: a simple, low-cost EAI supporting the cross-platform sharing of functions and data; and an enabler of reducing integration complexity and time (Miller, 2003). To reach these benefits, however, Web services should meet many technology requirements and capabilities. Some of the requirements include (Zimmermann, Tomlinson & Peuser, 2003):

- *Automation Through Application Clients:* It is required that arbitrary software applications running in different organizations have to directly communicate with each other.
- *Connectivity for Heterogeneous Worlds:* Should be able to connect many different computing platforms.
- *Information and Process Sharing:* Should be able to export and share both data and business processes between companies or business units.
- *Reuse and Flexibility:* Existing application components can be easily integrated regardless of implementation details.
- *Dynamic Discovery of Services, Interfaces, and Implementations:* It should be possible to let application clients dynamically, i.e., at runtime, look for and download service address, service binding, and service interface information.
- *Business Process Orchestration Without Programming:* Allows orchestration of business activities into business processes, and executes such aggregated process automatically.

The first five requirements are technology oriented. A solution to these requirements is XML-based Web services, or simply Web services. It employs Web standards of HTTP, URLs, and XML as the lingua franca for information and data encoding for platform independence; therefore it is far more flexible and adaptable than earlier approaches.

The last requirement relates to the concept of business workflow and workflow management systems. In supply chain management for example, there is a purchase order process at the buyer's side and a product fulfillment process at the supplier's side. Each process represents a business workflow or a Web service if it is automated. These two Web services can be combined into one Web service that represents a new business process. The ability to compose new Web services from existing Web services is a powerful feature of Web services; however, it requires standards to support the composition process. This article will provide a simplified exposition of the underlying basic technologies, key standards, the role of business workflows and processes, and critical issues.

WHAT ARE "WEB SERVICES"?

The phrase "Web services" has been defined in many different ways (Castro-Leon, 2002; Ambrosio, 2002). In the working draft of Web Services Architecture (W3C, 2003), it is defined as:

"…a software system designed to support interoperable machine-to-machine interaction over a network. It has an interface described in a machine-processable format (specifically WSDL). Other systems interact with the Web service in a manner prescribed by its description using SOAP-messages, typically conveyed using HTTP with an XML serialization in conjunction with other Web-related standards."

A simplified Web service architecture based on this definition is conceptually depicted in Figure 1.

Main features of Web services are that services (Burner, 2003):

1. Expose programmable application logic.
2. Are accessed using standard Internet protocol.
3. Communicate by passing messages.
4. Package messages according to the SOAP specification.
5. Describe themselves using WSDL.
6. Support the discovery of Web services with UDDI.
7. Are loosely coupled.

WEB SERVICES TECHNOLOGIES

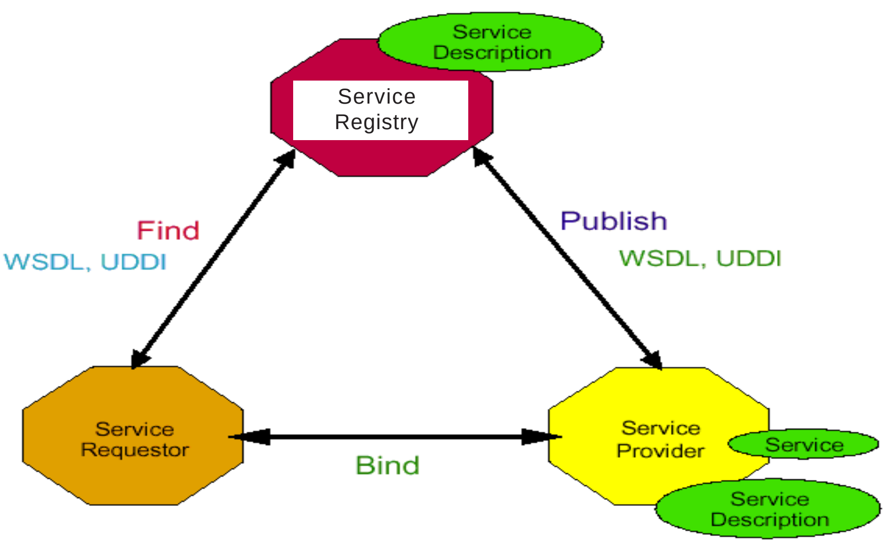
Three XML-based protocols—one for communication, one for service description, and one for service discovery—have become de facto standards (Curbera et al., 2002). They are:

- SOAP (the Simple Object Access Protocol) provides a message format for communication among Web services.
- WSDL (the Web Services Description Language) describes how to access Web services.
- UDDI (the Universal Description, Discovery, and Integration) provides a registry of Web services descriptions.

Another area of importance in Web services is the capability of constructing new composite Web services from existing Web services. Many standards in this area are being developed (Van der Aalst, 2003), for example, Business Process Execution Language for Web Services (BPEL4WS) by IBM and Microsoft (Fischer, 2002). It is not clear if there will be a common standard. However, regardless of the differences among vendor groups, the composition of Web services uses the concept of business processes and workflow management.

As noted earlier in this article, the development and deployment of Web services do not require a particular platform, nevertheless most Web services development

Figure 1. A simplified Web services architecture (W3C, 2003)



is being accomplished today using either Microsoft .NET or Sun Microsystems' J2EE specifications (Miller, 2003). It is not clear which of the two competing platforms is most suitable for the developers and their future directions (Miller, 2003; Williams, 2003).

THE ROLE OF BUSINESS PROCESSES AND WORKFLOW MODELING IN WEB SERVICES COMPOSITION

The need to integrate software components within and across companies is to economically re-use components for supporting business processes automation. A business process such as borrowing a book from a library may involve: 'check user identification', 'enter call numbers', and 'generate a receipt' activities. Each activity in the business process is a piece of work and is performed according to a sequence defined by business rules. That is, a business process contains a workflow (Allen, 2000).

The Workflow Management Coalition (Norin & Shapiro, 2002) defines the *business process* as:

"…a set of one or more linked procedures or activities which collectively realize a business objective or policy goal, normally with the context of an organizational structure defining functional roles and relationships."

Workflow management is further defined as:

"…the automation of a business process, in whole or part, during which documents, information, or tasks are passed from one participant to another for action, according to a set of procedural rules."

To understand complex business processes, workflow modeling techniques provide logical descriptions of the flow of activities that achieves the goal of the process. Companies compete to provide workflow-based tools for Web-service integration (Ganesarajah & Lupu, 2002). Such tools allow developers to use the concept of workflow to build complex business processes from Web services. Since a complex business process contains a large number of activities, managing the execution of these activities, (e.g., monitoring the workflow progress, sending activities to the right servers, and scheduling activities for execution) can be critical. The advent of the workflow management system or the workflow engine software is to serve this purpose.

Existing methods for creating business processes are not designed to work with cross-organizational components and Web services platforms (Peltz, 2003). This gives rise to multiple Web services standards for business process modeling and execution (Van der Aalst, 2003). However, regardless of standards, the composition of Web services requires the modeling (and construction) of a Web service from one party perspective and the interaction among each involved party. This is accomplished by using business process and workflow modeling in two aspects: 1) specifying internal details of the composition to offer an executable business process, and 2) planning the message sequences between parties and sources. The former is called *orchestration* and the latter is called *choreography* in creating business processes from composite Web services (Peltz, 2003).

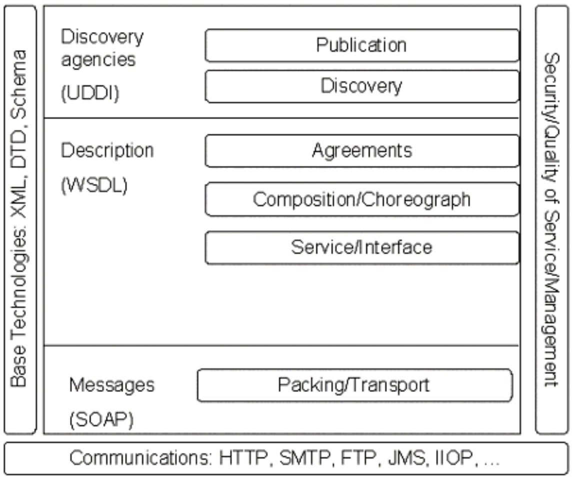
OUTSTANDING AND CHALLENGING ISSUES IN WEB SERVICES

Web service technology is still emerging, and researchers are still developing important functionalities for meeting technical and business requirements. Although some Web service standards such as SOAP, WSDL, and UDDI have been widely accepted, others such as Web services composition language are being developed by many organization and industry groups (Van der Aalst, 2003). These languages build upon the concept of workflow management systems to allow for orchestrating a group of required Web services in support of a business process. Due to real-world demands for more complex business problems integration, many companies and standards organizations have built extensions to the core specifications. Unfortunately, few of these extensions are interoperable. To resolve this problem and ensure application interoperability and extensibility, the W3C (2003) is developing a formal Web service architecture. There are different competing approaches (Kleijnen & Raju, 2003) to this problem, notably, the ebXML specifications suite, an independent technology for business-to-business integration (Patil & Newcomer, 2003). Additional challenges lie in XML and Web services security standards (Naedele, 2003). Obviously, the creation and support of Web service standards for mission-critical applications is a challenging, complex, and time-consuming process. Figure 2 adapted from W3C (2003) shows more completely the critical service areas of Web services.

A brief explanation of some Web service architectural components follows:

- *Packing/Transport:* This is the message layer that packs information into messages and transports it between parties.
- *Service/Interface:* Describes operational interfaces of a Web service.

Figure 2. A Web service architecture



- *Composition/Choreography:* Describes how services interact with each other in business processes.
- *Agreements:* The service level agreement defines the specific performance, usage, costs, and so forth, and the business level agreement defines a contractual agreement between the two business partners in business engagement using Web services.
- *Security Management and Quality of Service:* Describes reliability and security characteristics.

Each component has varying degrees of achievement by industries and standard setting organizations. However, standards are not enough to realize the great expectations of Web services. Langdon (2003) describes several key inhibitors of Web services adoption, including the lack of complementary service providers for metering, accounting, and billing services; the lack of ready-to-use Web services from either internal sources or third parties; and the lack of third-party support of how to decompose an automation problem and how to deliver it.

CONCLUSION

Web services involve many technologies and standards. Although a very young field, it has made tremendous progress in the last two years. Software vendors like Microsoft's .NET and Sun's J2EE have tools for Web services development. However, the field has many challenging issues, with standards being one of them. This article provides an overview of Web services in general and the role of workflow in developing a Web services application in particular. Since workflow modeling in the composition of Web services utilizes knowledge of business processes, that is where MIS professionals should be actively involved.

REFERENCES

Allen, R. (2000). Workflow: An introduction. In L. Fisher (Ed.), *The workflow handbook 2001*. Workflow Management Coalition.

Ambrosio, J. (2002). Web services: Report from the field. *Application Development Trends*, *9*(6).

Burner, M. (2003). The deliberate revolution. *ACM Queue*, *1*(1), 28-37.

Chung, J., Lin, K. & Mathieu, R. (2003). Web services computing: Advancing software interoperability. *Computer*, *36*(10).

Curbera, F., Duftler, M., Khalaf, R., Nagy, W., Mukhi, N. & Weerawarana, S. (2002). Unraveling the Web services web: An introduction to SOAP, WSDL, and UDDI. *IEEE Internet Computing*, *6*(2), 86-93.

Ferris, C. & Farrell, J. (2003). What are Web services? *Communications of the ACM*, *46*(6), 31.

Fischer, L. (2002). The WfMC heralds BPEL4WS standards for business process management industry. Retrieved from *xml.coverpages.org/WfMC-Heralds-BPEL4WS.html*.

Ganesarajah, D. & Lupu, E. (2002). Workflow-based composition of Web services: A business model or a programming paradigm? *Proceedings of the 6th International Enterprise Distributed Object Computing Conference*, Lausanne, Switzerland.

Kleijnen, S. & Raju, S. (2003). An open Web services architecture. *ACM Queue*, *1*(1).

Kreger, H. (2003). Fulfilling the Web services promise. *Communications of the ACM*, *46*(6), 29-34.

Langdon, C.S. (2003). The state of Web services. *Computer*, *36*(7), 93-94.

Marin, M., Norin, R. & Shapiro, R. (Eds.). (2002). *Workflow process definition interface—XML process definition language*. Document No. WFMC-TC-1025. Document Status: 1.0 Final Draft.

Miller, G. (2003). .NET vs. J2EE. *Communications of the ACM*, *46*(6), 64-67.

Naedele, M. (2003). Standards for XML and Web services security. *Computer*, *36*(4), 96-98.

Patil, S. & Newcomer, E. (2003). ebXML and Web services. *IEEE Internet Computing*, *7*(3), 74-82.

Peltz, C. (2003). Web services orchestration and choreography. *IEEE Computer*, *16*(10).

Van der Aalst, W.M.P. (2003). Don't go with the flow: Web services composition standards exposed. *IEEE Intelligent Systems*, (January/February).

W3C. (2003, August). *Web services architecture*. W3C Working Draft. Retrieved from *www.w3.org/TR/2003/WD-ws-arch-20030808/*

Williams, J. (2003). J2EE vs. .NET. *Communications of the ACM*, *46*(6), 59-63.

Zimmermann, O., Tomlinson, M. & Peuser, S. (2003). Perspectives on Web services. Berlin, Heidelberg, New York: Springer-Verlag.

KEY TERMS

EAI: Projects involving the plans, methods, and tools aimed at modernizing, consolidating, and coordinating the computer applications and data in an enterprise.

Grid Computing: A form of distributed computing that involves coordinating and sharing computing, application, data, storage, or network resources across dynamic and geographically dispersed organizations.

HTML (Hypertext Markup Language): A standard language for representing text, formatting specifications and hyperlinks.

HTTP (Hypertext Transfer Protocol): The standard for requesting and transmitting information between a browser and a Web server.

Java/RMI: A Java application programming interface known as remote method invocation.

Protocol: A set of rules and procedures governing transmission between two points in a network.

URL (Universal Resource Locator): A text string used as a reference to a Web resource. A URL consists of a protocol, a host name, and a document name.

XML (Extensible Markup Language): An extension of HTML that is being extensively used for transmitting data/information on the Internet.

Business Processes and Knowledge Management

John S. Edwards
Aston Business School, UK

INTRODUCTION

Knowledge has been a subject of interest and enquiry for thousands of years, since at least the time of the ancient Greeks, and no doubt even before that. "What is knowledge" continues to be an important topic of discussion in philosophy.

More recently, interest in *managing* knowledge has grown in step with the perception that increasingly we live in a knowledge-based economy. Drucker is usually credited as being the first to popularize the knowledge-based economy concept, for example, by linking the importance of knowledge with rapid technological change in Drucker (1969). Karl Wiig coined the term knowledge management (hereafter KM) for a NATO seminar in 1986, and its popularity took off following the publication of Nonaka and Takeuchi's book *"The Knowledge Creating Company"* (Nonaka & Takeuchi, 1995). Knowledge creation is in fact just one of many activities involved in KM. Others include sharing, retaining, refining, and using knowledge. There are many such lists of activities (Holsapple & Joshi, 2000; Probst, Raub, & Romhardt, 1999; Skyrme, 1999; Wiig, De Hoog, & Van der Spek, 1997). Both academic and practical interest in KM has continued to increase throughout the last decade.

In this article, first the different types of knowledge are outlined, then comes a discussion of various routes by which knowledge management can be implemented, advocating a process-based route. An explanation follows of how people, processes and technology need to fit together, and some examples of this route in use are given. Finally there is a look towards the future.

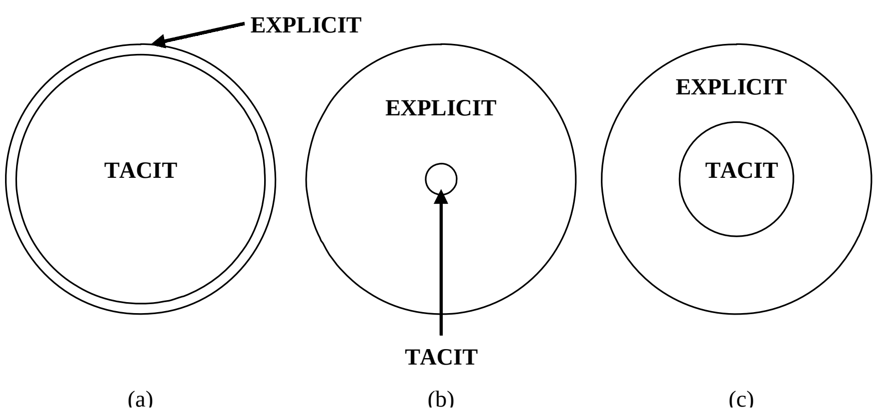
BACKGROUND

Types of Knowledge: Tacit and Explicit

Nonaka and Takeuchi's book (1995) popularized the concepts of tacit and explicit knowledge, as well as KM more generally. They based their thinking on that of Michael Polanyi (1966), expressed most memorably in his phrase "we know more than we can tell".

It is, however, most important to realize that tacit and explicit knowledge are not mutually exclusive concepts. Rather, any piece of knowledge has both tacit and explicit elements, as shown in Figure 1. The size of the inner circle represents the proportion of tacit knowledge: the "tacit core" at the heart of the knowledge that we "cannot tell". Figure 1(a) shows a case where the knowledge is almost entirely tacit, as in riding a bicycle. Figure 1(b) shows mainly explicit knowledge, where the tacit core is very small, for example, how to process a claim for travel expenses in an organization. Figure 1(c) shows an intermediate case, such as making a piece of furniture, where

Figure 1. The relationship between tacit and explicit knowledge



substantial amounts of both tacit and explicit knowledge are involved.

The Role of KM Systems

KM systems represent a deliberate, conscious attempt to manage knowledge, usually in an organization. Hansen, Nohria, and Tierney (1999) identified that there are two fundamental KM strategies, codification and personalization. Codification concentrates more on explicit knowledge (often relying very heavily on information technology), personalization more on tacit knowledge. Again, it is important to realize that these are not mutually exclusive, and that a strategy combining elements of both is likely to be the most successful.

ROUTES TO IMPLEMENTING KM

Many organizations have found it difficult to implement knowledge management systems successfully. Identifying "who", "what", and "why" – who is involved in knowledge management, what knowledge is being managed, and why is it being managed – can be problematic. The routes they have attempted to follow can be put into five generic categories, which will now be described.

Knowledge World Route

A substantial amount of the literature on knowledge management addresses knowledge at the level of the whole organization, or in a "world of knowledge" that is not specifically linked to the activities that a particular organization carries out. On an abstract level, such discussion of knowledge management can be extremely valuable. However, it has weaknesses in terms of practical implementation. For example, it is necessary not only to understand how individuals learn, but also how they learn in a given organization, and how the organizational systems may help or hinder the individual's learning process. The same issue applies even more forcefully to group learning, since the organization provides a crucial element of the group's context.

The practical focus in Nonaka and Takeuchi (1995) was very much on knowledge creation. As a result, organizations attempting to follow their principles for other aspects of KM, such as sharing or retaining knowledge, have sometimes found it difficult to make a specific connection from abstract ideas about knowledge to what the organization actually does, or could do, or should do. Often only the "why" is present, not the "who" or even the "what". Something more concrete is needed.

IT-Driven Route

This route assumes that the fundamental requirement is for the codification of as much knowledge as possible. Advocates of this approach sometimes refer to this as "extracting" the knowledge from the people who possess it; see for example Johannsen and Alty (1991). This is an inadvisable term to use, for two reasons. First, it is logically incorrect; their knowledge is being *shared*, not extracted. The people still have the knowledge after the "operation" has taken place. Second, it gives the people the wrong impression – that their knowledge is being taken away. This is not a recipe to encourage their cooperation. For an organization of any size, such a codification task evidently requires IT support, and the thrust of this route is that once the "correct" form of IT support for managing knowledge has been chosen, it is simply a matter of a great deal of hard work.

This technology-driven route only works well in a limited range of situations where the "what" questions are most important, for example, where the main KM task is managing the knowledge held by a company in the form of patents. In other circumstances, it may not achieve any improvement in knowledge management at all. One example of this from the author's experience is of a heavy manufacturing firm. Knowledge management in this organization was seen solely as an information systems issue; the KM group was part of the information systems department. The "solution" was seen in terms of the implementation of a knowledge sharing system based on Lotus Notes™. However, there was no real consideration as to who would share what knowledge or for what specific purpose. Consequently, the eventual use of the installed IT was poor; the only really successful use was by the knowledge management project team itself, where the "who" and "why" questions had been properly addressed, as well as the "what" questions.

Functional Route

An alternative route that has the potential to address the "who", "what" and "why" questions is to organize the implementation around the existing organizational structure. The most commonly found structural elements intended to facilitate learning and knowledge sharing in organizations are departmental groupings based on functions. These have clear advantages in terms of what might be termed professional development and allegiance. Davenport and Prusak (1998) report examples of successful knowledge transfer between groups of surgeons, and groups of tunneling engineers, among others. However, this functional route also has the disadvantage that it encourages the compartmentalization of knowledge. This problem can only worsen over time, as specialisations multiply and sub-divide. In addition, professional divisions can actively prevent sharing of knowledge. It has, for example, taken decades for hospital doctors in the UK National Health Service to allow other professionals such as pharmacists and physiotherapists to participate in decision-making about treatment of individual patients on an equal footing. On a wider scale, modern Western medical science has come to separate "diet" and "drugs", at least until the very recent past, in a way that Chinese medicine, for example, never has done. The problems of running an organization in this manner, and the "functional silos" mentality that tends to result, were recognized by authors such as Hammer (1990) as part of the business process reengineering movement, when KM was in its infancy.

Therefore, although the functional route to implementation will allow some improvement in KM, progress may be limited by the characteristics of the existing structure, and in the worst cases (for example, where transferring knowledge between functions is the greatest KM issue in the organization) this route may be counter-productive.

People-Centric Route

A people-centric route to KM is the essence of the Hansen et al. (1999) personalization strategy. By definition, such an approach, appropriately implemented, will answer all the "who" questions that might be involved in KM. Thus in organizations where there is general consensus on "what" knowledge is important and "why" it needs to be managed, such a route should prove effective.

However, as was mentioned in the previous sub-section, organizations have become increasingly diverse in their activities, and in the range of specialized knowledge that they need to access. This means that consensus even on what knowledge the organization has, never mind what is important, may be difficult to achieve. On the one hand, it may not be easy for a particular specialist to fully appreciate "what the organization does". Equally, even the most conscientious senior manager will find it literally impossible to understand all the expertise and knowledge possessed by the specialists in his or her organization. To repeat the quotation from Hewlett Packard CEO Lew Platt (Davenport & Prusak, 1998, p. xii), "If HP knew what HP knows, we would be three times as profitable."

Business Processes Route

The managers in an organization have to translate the goals of any strategic program or initiative – whether on knowledge management or something else – into practical, implementable reality. In other words, to connect with "what the organization does". Various management thinkers have presented models of this, for example:

- Porter's (1985) value chain,
- Earl's (1994) view of core processes, the ones that are done directly for external customers,
- Beer's (1985) "System Ones", the systems that make the organization what it is,
- Core competences/competencies as espoused by Hamel and Prahalad (1994).

Although there are some significant differences between them, their common theme is that the effectiveness – indeed, the competitive advantage – of organizations depends not on how they are structured, or on what resources they have, but on what they *do*. In the terminology of this article, this means their underlying business processes. Note that the term *business* processes is used throughout, but such processes exist equally in not-for-profit organizations.

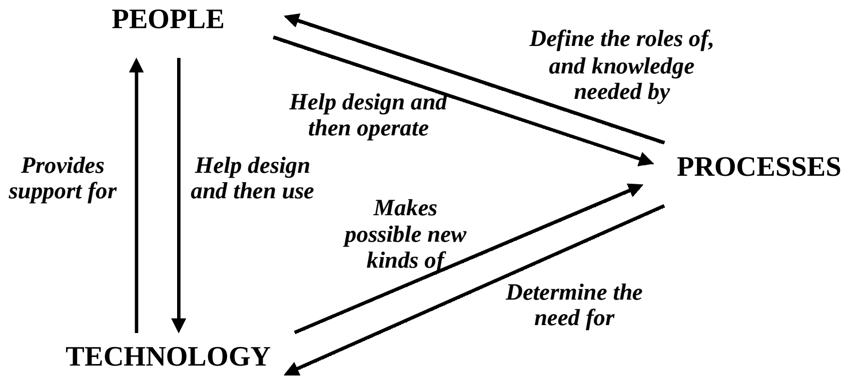
Business processes possess five characteristics that justify their use as a foundation for knowledge management in organizations.

1. Business processes have identifiable customers, whether internal or external. Knowledge is of little relevance unless put to use for a customer of some kind.
2. Business processes cut across organizational boundaries. Knowledge does not need to, and does not, obey the artificial boundaries within an organization.
3. Business processes consist of a structured set of activities. Choosing the appropriate way to structure activities is an important part of the knowledge.
4. Business processes need to be measured. Without some form of measurement as a comparison, knowledge cannot be validated.
5. While the parts of a business process are important, the overriding requirement is that the overall process works. Valid knowledge in an organizational context must take a holistic view.

An additional argument (Braganza, 2001) is that viewing knowledge management in terms of an organization's processes gives a much-needed *demand-side* view of

B

Figure 2. People, processes and technology in a KM system



knowledge. This is complementary to the supply-side view of knowledge that stems, for example, from considerations "of data leading to information leading to knowledge". Beer and Earl particularly concentrate on this demand-side perspective. Beer indeed goes even further, to include the informal processes and activities of the organization as well as the more formalized ones.

Completing this argument for a greater use of the business processes route, the knowledge that an organization requires must, logically, be related not just to what that organization does, but also to how it does it. Thus, people in organizations should think about this knowledge, and how to manage it, by reference to that organization's business processes.

PEOPLE, PROCESSES AND TECHNOLOGY

From the earlier discussion, it may be seen that, whichever route is chosen, effective KM requires the consideration of both tacit and explicit knowledge. The need is to coordinate people, processes and technology successfully. The interaction of these three elements is shown in Figure 2.

Not only does a knowledge management system consist of more than technology, it is important to realize that the technology used to support KM does not have to be "KM software". Recent studies have found that generic software such as e-mail or an Intranet may be at least as important as specific software (Edwards, Shaw, & Collier, 2004 (to appear); Zhou & Fink, 2003).

KM BY A BUSINESS PROCESSES ROUTE

As it has so far been less frequently attempted than the other routes, some examples of organizations that have implemented KM by a business processes route will now be given.

Unisys (Wizdo, 2001) have embarked upon a company-wide knowledge management initiative with an explicit process focus. Its objectives include:

- Accelerating the speed and scope of organisational learning,
- Decreasing the time it takes to reach critical mass in new markets,
- Uniting cross-boundary groups,
- Increasing innovation in product and process.

Wizdo identifies three increasingly ambitious categories of "transformation" in business: efficiency, innovation and re-invention. The Unisys knowledge management program regards a focus on processes as essential in achieving the two "higher" categories.

Objective Corporation (Fisher, 2001) changed most of their processes over a period of some five years. They found that such an emphasis not only improved knowledge management within the business, but also had a significant impact on the performance of the business itself. In this case, it was most likely the effect of a coherent training programme, with an emphasis on understanding, increasing the overall organisational performance through the people involved operating more effectively.

Both of these examples involved substantial use of information technology. However, that does not have to be the case. The author's group has been working with a component manufacturer in the aerospace industry, whose KM initiative also has an explicit process focus. Typically, their manufacturing processes use a machine operated by one person. The operators' choice of the best way to retain and share knowledge does not use IT at all (except for a word processor). The agreed best operating procedure, with illustrations, is put on a laminated sheet of paper mounted near the machine, which includes the names of the people who had contributed to designing the procedure. A suitable pen is provided to annotate the laminated sheet. At regular intervals, office staff come round to produce a revised version of any of the "Standard Operating Sheets" that have been annotated.

FUTURE TRENDS

A further justification for the use of business processes as the foundation for implementing knowledge management is that they are now becoming part of the mainstream of management thought. For example, the latest version of the ISO9000 family of standards for Quality Management Systems, including ISO9001: 2000, is constructed on the basis of a "process approach". The ISO9000 term *realisation process* is equivalent to Earl's core process or Beer's primary activity as discussed earlier. Significantly, the latest editions of strategic management textbooks (Johnson & Scholes, 2001) typically discuss the business process view of organizations, whereas earlier ones did not.

It seems clear, therefore, that the business processes route to implementing knowledge management is likely to become more common in future, and that this will encourage the development of ever more appropriate information technology for supporting it. Equally, new information technologies will enable new types of process. Intelligent agents, smart cards and "picture phones", for example, all offer different possibilities for supporting KM which have only just begun to be considered.

CONCLUSION

This article has considered the implementation of knowledge management systems by looking at five different generic routes towards achieving it: knowledge world, IT-driven, functional, people-centric and business processes. While each of these routes has some merits, it has been argued that the business processes route offers potential for the greatest integration between knowledge management and "what the organization does". It is, thus, likely to be increasingly common in the future.

REFERENCES

Beer, S. (1985). *Diagnosing the system for organisations*. Chichester, UK: Wiley.

Braganza, A. (2001). Knowledge (mis)management…and how to avoid it. *Information Management 2001*. Olympia, London.

Davenport, T.H. (1993). *Process innovation: Reengineering work through information technology*. Boston, MA: Harvard Business School Press.

Davenport, T.H., & Prusak, L. (1998). *Working knowledge: How organizations manage what they know*. Boston, MA: Harvard Business School Press.

Drucker, P.F. (1969). *The age of discontinuity*. London: Heinemann.

Earl, M. J. (1994). The new and the old of business process redesign. *The Journal of Strategic Information Systems, 3*(1), 5-22.

Edwards, J.S., Shaw, D., & Collier, P.M. (2004 (to appear)). Knowledge management systems: Finding a way with technology. *Journal of Knowledge Management*.

Fisher, G. (2001). A framework for deploying knowledge management in an EDRM environment. *Information Management 2001*. Olympia, London.

Hamel, G., & Prahalad, C.K. (1994). *Competing for the future*. Boston, MA: Harvard Business School Press.

Hammer, M. (1990). Re-engineering work: Don't automate, obliterate. *Harvard Business Review, 68*(4 (July/August)), 104-112.

Hammer, M., & Champy, J. (1993). *Reengineering the corporation: A manifesto for business revolution*. London: Nicholas Brealey.

Hansen, M.T., Nohria, N., & Tierney, T. (1999). What's your strategy for managing knowledge? *Harvard Business Review, 77*(2), 106-116.

Holsapple, C.W., & Joshi., K.D. (2000). An investigation of factors that influence the management of knowledge in organizations. *Journal of Strategic Information Systems, 9*, 235-261.

Johannsen, G., & Alty, J.L. (1991). Knowledge engineering for industrial expert systems. *Automatica, 27*(1), 97-114.

Johnson, G., & Scholes, K. (2001). *Exploring corporate strategy, text and cases* (6th ed.). Harlow: Financial Times Prentice Hall.

Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company, how Japanese companies create the dynamics of innovation*. New York and Oxford: Oxford University Press.

Polanyi, M. (1966). *The tacit dimension*. Garden City, NY: Doubleday.

Porter, M.E. (1985). *Competitive advantage: Creating and sustaining superior performance*. New York, London: Collier Macmillan.

Probst, G., Raub, S., & Romhardt, K. (1999). *Managing knowledge: Building blocks for success*. Chichester: Wiley.

Skyrme, D.J. (1999). *Knowledge networking: Creating the collaborative enterprise*. Oxford: Butterworth-Heinemann.

Wiig, K.M., De Hoog, R., & Van der Spek, R. (1997). Supporting knowledge management: A selection of methods and techniques. *Expert Systems with Applications, 13*(1), 15-27.

Wizdo, L. (2001, November 8th-9th). Organisational models for enterprise knowledge management. Paper presented at the *Second European Conference on Knowledge Management*, Bled, Slovenia.

Zhou, A.Z., & Fink, D. (2003). Knowledge management and intellectual capital: An empirical examination of current practice in Australia. *Knowledge Management Research & Practice, 1*(2), 86-94.

KEY TERMS

Business Process: A structured, measured set of activities designed to produce a specified output for a particular customer or market (Davenport, 1993, p. 5).

Business Process Reengineering: The fundamental rethinking and radical redesign of business processes to achieve dramatic improvements in critical, contemporary measures of performance, such as cost, quality, service and speed (Hammer & Champy, 1993, p. 32).

Demand-Driven View of Knowledge: A view of knowledge stemming from the requirements of the organization; for example, what knowledge is needed to carry out a particular activity and how can it be applied?

Explicit Knowledge: Knowledge that has been (or can be) codified and shared with others.

Knowledge Management: Supporting and achieving the creation, sharing, retention, refinement, and use of knowledge (generally in an organizational context).

Knowledge Management Software: Software specifically intended for knowledge management, such as data mining and "people finder" software.

Knowledge Management System: A combination of people, processes and technology whose purpose is to perform knowledge management in an organization.

Supply-Driven View of Knowledge: A view of knowledge stemming from the knowledge itself rather than its uses. Often related to a continuum data-information-knowledge.

Tacit Knowledge: Knowledge that is difficult or impossible to express, except by demonstrating its application.

B

Business Strategy, Structure and IT Alignment

Petros Theodorou
Technological Educational Institution of Kavala, Greece

INTRODUCTION

The structure of production during the last decade has been changed radically. The importance of design, flexibility, quality, and dependability increased. The main motive behind that restructuring seems to be the qualitative and quantitative diversification of demand (regulatory, technology, and other issues can be mentioned as well). The diversification of demand forced production to be more flexible in order to accommodate the environment's variations. Information systems (IS) proved to be a perfect 'partner' for the enterprise who wants to move along at this pace: to increase competitive advantage and retain its competitive position in the market. The capabilities of information technology (IT) for increasing competitive advantage forced management to consider IT in the strategic planning process and to consider IT not as a mere tool of bolts and nuts, but as a strategic partner. The key for the successful IT-IS adoption is the strategic IT alignment model. According to this, IT should be aligned with organizations' structure and employees' culture in order to avoid resistance and increase core competence at the strategic level. But the strategic options offered by advanced IT investments are not appraised by using the usual hard financial criteria. Instead, Black and Scholes developed a financial formula to valuate derivative financial products and open the road to valuate options offered by real investments. Thus, the application of Black and Scholes' formula offers an opportunity to valuate financially strategic IT investment.

The importance of alignment among IT-business structure and strategy-environment is increased strategically, but the relationship among all those variables is complicated. Generally, in the classical strategy school, strategy is determined by environmental uncertainty. Structure (as well as technology) should follow the needs of strategy, and advanced information technology must be aligned with business strategy and structure. This alignment is a prerequisite for the creation of competitive advantage. Information technology is a key element of structure to bypass environmental uncertainty. In this model, other contingencies like size, age of organization, political forces, regulatory framework, and so forth must be taken into account as the interaction among all those factors forms the alignment model.

BACKGROUND

The concept of alignment has been widely examined in strategy and organizations' theory literature, underlying the contingency theories and constituting the groundwork of management of technology and strategy of information systems. Strategic alignment and strategic fit have been among the top concerns of business executives (Rodgers, 1997; Brancheay, Janz & Wetherbe, 1996) and the core concept in the normative models. Alignment literally means an arrangement of groups or forces in relation to one another. The determination of those forces is an important aspect of strategic alignment research. The following aggregate variables are commonly found in many alignment models: environment uncertainty, business structure, information technology (IT), and business strategy. Among those variables, Miles and Snow (1984) determine strategy as the basic alignment mechanism, and organizational structure and management processes as the internal arrangements. In the classical school of strategy, environment has the most important role as a determinant of strategy. Structure follows the alterations of strategy, while the role of IT is collateral. Strategy follows environmental variations and determines the structural form. IT as a part of structure follows strategy as well. In contrary to this view, Mintzberg and Quinn (1996) proposed a bottom-up approach that is free of the 'environmental biased' (Theodorou, 1997). Practical experience shows that variables which should be aligned are interrelated in a multidimensional way; but for the sake of simplicity, a simple bivariate approach is usually used among:

environment↔strategy, structure↔strategy, IT ↔strategy and IT↔structure.

Moreover, it should be mentioned that the concept of alignment in business strategy has two dimensions (Horovits, 1984; Reich & Benbasat, 1999)-the intellectual (interrelated IT and business plans) and the social dimension (understanding and commitment of participants). Furthermore, alignment is more an ongoing process and not a stable state, thus sustainability can be gained only by continuous efforts. No strategy is universally superior, irrespective of the environmental or organizational context (Venkatraman, 1989). A flexible structure that follows demand capable to produce in low volume, low cost, and high quality needs long-term investment. Strategic alignment enables enterprises to avoid over-simplicity while targeting competitive and distinctive advantage assisting to dominate in competition (Drago, 1999). According to that concept, business strategy should evolve while taking into account the impact of IT. On the other hand, if information technology is applied without taking into account the strategic targets of cost, design, quality, flexibility, and dependability, then only operational benefits will be created without any impact on competitive advantage.

Generally the strategic alignment problem has multiple dimensions, but the main determinants are the following: environmental uncertainty, business structure, IT and business strategy. Except for the previously mentioned variables, we could refer business culture, managerial knowledge/processes and individual roles, the regulatory issues, political forces, and so forth. The alignment model should consider among other contingencies the impact of industry and competitive change (particularly under the influence of IT and the generic strategies in the strategy portfolio), the strategic changes induced by IT that affect the reengineering of internal and external value chain, the IT capabilities that shift or expand the business scope and domain, and finally, the size and the age of the firm. In the “management in ’90s” research framework, Morton (1991) illustrates some of the various aspects of the firm that must be considered in the alignment process. In this model, strategic fit is defined by looking at the IT potential to shape and support business strategy; alignment at the organizational level is defined by looking at organizational infrastructures and processes. Thus, four perspectives of alignment arise: the strategy execution, the technology, the competitive potential, and the service level. Moreover, four alignment mechanisms are identified: value management, governance, technological, and organizational capability. Misfit and misalignment is responsible for why many IT applications are not meeting expectations and do not cover the needs of the company.

MAIN THRUST OF THE ARTICLE

As previously mentioned, the concept of strategic alignment has various dimensions. The most important according to the bivariate approach are business strategy, business structure, information technology, and business environment. Thus, the interrelation among those variables captures the purpose of the alignment model. The external environment is characterized as dynamic, if the firm experiences often exogenous changes, such as: product design by competitors, variability in time of deliveries (by suppliers), high growth in market share, changes in relative market share, changes and innovation in information technologies adapted by competitors or offered in the marketplace, and so forth. Every source of uncertainty which the firm realizes from the external environment (capable to affect competitive advantage) is a main drive to reconsider the alignment mechanism. Through strategic analysis the firm realizes (if possible) the changes in the external environment to take action realigning the business structure in order to reply to competition. Information technology in this context offers the flexibility needed for quick and efficient response. Thus, the relation of information technology and business strategy and structure is of vital importance. Following, we will present the strategy-IT and structure-IT alignment.

Strategic Alignment of Advanced IT

An important aspect of strategy and IT alignment was to correspond the strategic content with IT capabilities, in other words, IT to strategic objectives, and examine how they are affected. That topic is basically investigated in strategic information system literature where matrix models developed in order to determine strategic alignment. Porter’s work comprised the basis for the development of strategic information systems even though his model is more familiar in strategic management literature. Neumann (1994), in his book about strategic information systems, denotes that Porter’s model of ‘5-forces’ is the framework of frameworks. Specifically, regarding the strategic objectives, Porter attempted a generalization of specific strategic targets, the so-called generic strategies, which are: cost, differentiation, and focus (niche) strategy. Furthermore, Wiseman (1998) extended that logic by adding the strategic targets of alliance, growth, and innovation. Generally, business strategy should take into account IT in order to generate strengths and attain opportunities that increase competitive advantage. Competitive advantage from IT can be derived by taking into account the opportunities that can be exploited from the 5-forces model; for example, how can the firm: build barriers to entry, build switching costs, change the basis of competition, change the balance of power in suppliers relationship, and generate new products? Moreover, the sustainability issue raised the importance of business strategy and IT integration. Strategic information systems literature tried to explain alignment among strategy and IT with the use of matrix models. Synnot made a composition of the Boston Consulting Group’s matrix and Porter’s generic strategies in order to identify IT strategic options and achieve strategic alignment. A well-known model for the attainment of strategic opportunities criticized for over-simplicity is the Strategic Grid Matrix developed by McFarlan (1984). The empirical research of Hirschheim (1982) indicates that regardless of the 'simplicity problem', an important benefit that arises is that the model forms the basis for a dialogue among management. Thus this model is helpful for consensus and agreement, which is a significant part of social alignment. An important upgrade of strategic grid was made by Ward (1987), who extended the hermeneutic capability of the model by incorporating IS/IT planning (for a composition of strategic grid and Ward matrix, see Theodorou, 2003). Based on this matrix, Parson (1983; also see Ward, Griffiths & Whitmore, 1994, p. 251) discussed alignment among generic and IT strategies. Parson's generic IS/IT strategies include: centrally planned, leading edge, free market, monopoly, scarce resource, and necessary evil. The conclusion is that an organization will need to adopt a full range of generic strategies if it is to manage an extensive portfolio (Ward et al., 1994). Parson attempted a classification of IT applications according to the organization's functions and Porter's generic strategies. An important set of generic strategies is also found in manufacturing strategy literature. Based on the distinction made by Skinner (1985, pp. 78, 80, 88), Swamidass and Newell (1987), Adam and Swamidass (1989), and Swink and Way (1995), strategy is determined by priorities, and priorities determine the strategic direction. As can be seen from this literature, the generic business and IT strategies have been split up to more than five competitive priorities, and those priorities can be further split up to many tactics (for a more detailed analysis, see Theodorou, 1996b). The basic priorities are: design, cost, quality, flexibility, dependability, and innovation (Theodorou, 1997). Theodorou used a Likert scale to estimate the priorities of a firm, and further to detect the impact of IT using VACOR algorithm and based on the ROIC criterion. It was observed that higher performing firms follow the target of flexibility as a first-order priority, and late in the list was cost. That doesn't mean that cost was not of interest, but on the contrary, that the cost target for those firms was achieved (Theodorou, 2001). Kathuria and Igbaria create a matrix in order to correlate organizations' structure (from a functional view) with strategic priorities and IT, bypassing the limits of a two-dimensional approach. In this approach IT technologies and priorities are aligned according to the business functions (Kathuria & Igbaria, 1997).

Structural Alignment of Advanced IT

The theme of fit and alignment is central to any contingency model developed by organization theorists. Structure should be aligned with environment, taking advantage of IT capabilities. The target is to shape new forms of structure flexible enough to respond quickly to environmental threats and opportunities (Morton, 1995). Flexibility enables structure to interface with environmental uncertainty while taking advantage of IT potential. Venkatraman's (1989) statistical conceptualization 'opened' the field of empirical research in testing strategic alignment, introducing the concepts of moderation, mediation, matching, covariation, profile deviation, and gestalt. Ivary (1992) references that the mediation perspective dominates in research, but omits the criterion of performance. Fiedler, Grover, and Teng (1996) attempted a taxonomy of IT in relation to structure without a performance criterion. Brown and Magill (1994) refer to the gestalts approach for the identification of the form of centralization. Furthermore, Bergeron, Raymond, and Rivard (2001) found that moderation and matching perspective confirm the performance implications for the structure-IT pair under the strategic IT management approach. Generally, it can be seen that the prevailing concepts in structure-IT fit are: the matching, mediation, and moderation perspective. Those concepts, according to Raymond, Pare, and Bergeron (1995), can be presented as follows:

Matching: ((structure⇔IT)⇒Performance)
Mediation: (IT⇒structure⇒Performance)
Moderation: (IT⇒(structure⇒Performance))

Theodorou (1997) used performance as a criterion variable for structural variable analysis and determined IT-structure fit, using the moderate perspective for certain contingencies. In this model '21' (structural) variables try to grasp elements of structure in order to determine the structural design (organizational form) of the enterprise. Structural clusters were determined using VACOR algorithm. The impact of IT was estimated using the Newton Raphson algorithm and neural networks. It was concluded that the matching approach confirmed that high-performing organizational subunits matched a decentralized and differentiated structure with non-routine (flexible) technology (Theodorou, 1997; Theodorou & Dranidis, 2001). Once the firm makes strategic choices on IT, matching structures must be defined if the resulting performance is to be high (Raymond et al., 1995). If structure is to be defined in a more detailed mode, then the main structural variables and parameters should be determined (Theodorou & Dranidis, 2001).

FUTURE TRENDS: ALIGNMENT A REAL OPTION

Generic options are embedded in business strategy and business structure choice. As presented previously, the generic options regarding the strategic priorities can be conceptualized in cost, quality, dependability, and flexibility (for an extended discussion, see Theodorou, 2003). Additionally, business structure can take various forms (Mintzberg & Quinn, 1996). Theodorou and Dranidis (2001) developed a neural network system to determine business structure according to a set of structural variables, where different combinations provide different structural forms. Thus, options are embedded in structural design as well as in strategic priorities. Information technology must be aligned with business strategy and structure according to the demands of the external environment. Theodorou (2004) developed a DSS to determine the alignment mechanism for the above-mentioned variables. Strategic and structural options must be aligned within the external environment, taking advantage of information technology capabilities. Balasubramanian, Kulatilaka, and Storck (2000) noticed the importance of the relation among the real options and the alignment. They developed an approach using business capabilities from operating drivers and real options. According to this capability-based approach, valuation is contingent on market conditions and the degree of success in attaining the capabilities set as target (Balasubramanian et al., 2000).

But even if this framework can work well, in cases of IT applications with certain impact on sales (like e-commerce), it is not clear how it can be applied in IT applications with only operational benefits, where sales impact is superficial and only cost benefits can be encountered. In the case of the alignment model, a performance criterion (like ROIC or better a factor describing many performance estimates) can be employed in order to test if business structure, strategy, and information technology are aligned. Then the real options approach can be used on this criterion in place of exercise and spot price. Usually, analytical methods are used in a real options approach and are based upon the Black and Scholes formula, to estimate the options of abandonment, growth, and postponement (Kim & Sanders, 2002). In the classical approach the price of the asset (spot and exercise) is replaced with the NPV of the project. A basic step in that estimation is to derive the standard deviation from shocks in exogenous factors. In the case of the alignment model, standard deviation is estimated from shocks in strategic priorities and structural variables.

Reversibly, in the case of Black and Scholes, variance is derived from shocks on the financial variables (like prices, interest rates, etc.). The variability of choice in strategic priorities and structural variables produce variance in the strategic alignment model, and this is transferred in the criterion of performance which is applied in the option formula, in order to estimate the options of abandonment, growth, and postponement. The alignment model estimates the various strategic and structural options that can be found in the real context of an enterprise (Theodorou, 2004). The difficulty in IT valuation (and mainly in systems that automate operations such as CAM and MRPI) is that we can not estimate the currency inflows of the system. Alternatively, in the case where operational benefits, through cost reduction, can be encountered (like a database), two things can be done: a) discount the free cash flows, or b) estimate the strategic impact of efficiency on market share and estimate the MIRR on the differential revenues.

Generally, despite those two solutions, the mechanism of the alignment model with a performance factor can be proven more efficient and appropriate for implementation and valuation, as it takes into account all the available options in relation to business strategy and structure. Only under that concept can the benefits of information technology be fully valuated.

CONCLUSION

Firms, nowadays, operate in an uncertain environment, due to the qualitative and quantitative diversification of demand. Firms are targeting in small niches, trying to serve specific market needs. In this circumstance, the need for a flexible structure and strategy to bypass environmental uncertainty is increased. Information technology proved to be a good ‘partner’ to attain the strategic targets of this new environment. Computer-aided design, finite element analysis, concurrent engineering, and other techniques help to decrease the time to market and the cost of the new product’s design. Computer-aided manufacturing, DNC, and CNC machines enable the decrease of setup cost and produce economically smaller lots (batches) in order to serve smaller market niches, attaining economies of scope. Materials requirement planning for the West and just-in-time systems for the East automate the orders of material, and cut cost and time to market. Most of those systems (for a classification, see Theodorou, 2003) were integrated and popularized as MRPII or ERP in intranets or through the Internet.

The integration of those systems, along with the human factor, led to the emergence of strategic alignment theory. Alignment and fit seems to be very important in decreasing resistance and increasing competitive advantage. Key elements for the alignment to be considered by any firm are external environment, business strategy, business structure, and information technology. Business strategy should be aligned with the external environment and incorporate the information technology capabilities in the strategic plan. Information technology should fit with business structure and vise versa in order to serve specific strategic targets and priorities, like cost, quality, dependability, design, and so forth. Also, from the successful IT and business structure fit, strategic advantages can be created in an emerging pattern capable of influencing environment and strategy. It should be mentioned that alignment is not a stable state since environment is continuously changing, and strategy should accommodate respectively. Other contingency factors have to be taken into account like the age of the organization, the size, the culture, the power, as well as other historical factors specific for each enterprise.

A new challenge is opened for the application of the alignment model in the financial valuation of advanced IT systems. Hard financial measures proved incapable of evaluating the strategic impact and strategic role of those systems. Based on the alignment model, a proper choice of strategic priorities and structural variables can be made in relation to environmental uncertainty in order to create competitive advantage. Structural equation modeling opens the ground for an application. Structural variables and strategic priorities are the main latent variables that offer the options to be aligned under the performance criteria. Then, the performance factor can be used to estimate the options of abandonment, growth, and postponement based on the Black and Scholes formula.

REFERENCES

Adam, E.E. & Swamidass, P.M. (1989). Assessing operations management from a strategic perspective. *Journal of Management*, *15*(2), 181-203.

Balasubramanian, P., Kulatilaka, N. & Storck, J. (2000). Managing information technology investments using a real-options approach. *Journal of Strategic Information Systems, 9,* 39-62.

Bergeron, F., Raymond, L. & Rivard, S. (2001). Fit in strategic information technology management research: An empirical comparison of perspectives. *Omega*, *29*, 125.

Brancheau, J.C., Janz, B.D. & Wetherbe, J.C. (1996). Key issues in information systems management. *MIS Quarterly,* (March), 59-84.

Brown, C.V. & Magill, S.L. (1994). Alignment of the IS function with the enterprise: Toward a model of antecedents. *MIS Quarterly*, *18*(4), 371-403.

Drago, W.A. (1999). Simplicity/complexity as a dimension of strategic focus: Effect on performance in different organizational domains. *Management Research News, 22*(7), 12.

Fiedler, K.D., Grover, V. & Teng, J.T.C. (1996). An empirically derived taxonomy of information technology structure and its relationship to organizational structure. *Journal of Management of Information Systems*, *13*(1), 9-34.

Hirschheim, R.A. (1982). *Information management planning in organizations.* LSE Working Paper.

Horovits, J. (1984). New perspectives on strategic management. *Journal of Business Strategy,* (Winter), 19-33.

Ivary, J. (1992). The organizational fit of information systems. *Journal of Information Systems, 2*(1), 3.

Kathuria, R. & Igbaria, M. (1997). Aligning IT applications with manufacturing strategy: An integrated framework. *International Journal of Operations and Production Management*, *17*(6), 611-629.

Kim, Y.J. & Sanders, L. (2002). Strategic actions in information technology investment based on real option theory. *Decision Support Systems, 33,* 1.

McFarlan, F.W. (1984). Information technology changes the ways you compete. *Harvard Business Review,* (May-June), 98-103.

Miles, R.E. & Snow, C.C. (1984). Fit, failure and the hall of fame. *California Management Review*, *26*(3), 10-28.

Mintzberg, H. & Quinn, J.B. (1996). *The strategy process*. Englewood Cliffs, NJ: Prentice-Hall.

Morton, S.M. (1991). *The corporation of the '90s: Information technology and organizational transformation.* Oxford.

Neumann, S. (1994). *Strategic information systems, competition through information technologies.* New York: McMillan.

Raymond, L., Pare, G. & Bergeron, F. (1995). Matching information technology and organizational structure: An empirical study with implications for performance. *European Journal of Information Systems*, *4*(3).

Reich, B.H. & Benbasat, I. (1999). Factors that influence the social dimension of alignment between business and information technology objectives. *MIS Quarterly*.

Rodgers, L. (1997). Alignment revisited. *CIO Magazine*, (May 15).

Skinner, W. (1985). *Manufacturing: The formidable competitive weapon.* New York: John Wiley & Sons.

Swamidass, P.M. & Newell, W.T. (1987). Manufacturing strategy, environmental uncertainty and performance: A path analytic model. *Management Science*, *33*(4), 509-522.

Swing, M. & Way, M.H. (1995). Manufacturing strategy: Propositions, current research, renewed directions. *International Journal of Operations and Production Management, 15*(7), 4-26.

Theodorou, P. (1996a). Flexible manufacturing and the IT applications: The manufacturing strategies map. *Proceedings of the 2nd ALPS Euroconference*, Bologna, Italy.

Theodorou, P. (1996b). The restructuring of production in SME: The strategy of manufacturing flexibility. In I. Siskos, K. Zoupouidis & K. Pappis (Eds.), *The management of SME*. Chania, Creta Island, Greece: University Publications of Creta.

Theodorou, P. (1997). *Restructuring, structure strategy and information technology alignment.* PhD Dissertation, Department of Applied Informatics, University of Macedonia, Thessaloniki, Greece.

Theodorou, P. (2001). An implementation of VACOR algorithm in the taxonomy of strategic clusters and alignment among strategic targets and ERP systems. *Operational Research in the Service of Regional Development: Proceedings of the 14th National Congress of Operational Research Society of Greece*, Department of Mechanical Engineering, University of Thrace. Janthi 1-3 November. Greece, Ors of Greece.

Theodorou, P. (2003). Strategic information systems: The concept of alignment. In L.A. Joia (Ed.), *IT-based management challenges and solutions.* Hershey, PA: Idea Group Publishing.

Theodorou, P. (2004). A DSS that aligns business strategy and business structure with advanced information technology: A case study. *Annals of Cases in Information Technology, 6*, 157.

Theodorou, P. & Dranidis, D. (2001, July 3-4). Structural design fit in garment production: A neural network approach. *Proceedings of the 2nd European Conference on Intelligent Management Systems in Operations.* Operational Research Society, University of Salford, UK.

Venkatraman, N. (1989). The concept of fit in strategy research: Toward verbal and statistical correspondence. *Academy of Management Review, 14*(3), 423.

Ward, J.P. (1987). Integrating information systems into business strategies. *Long Range Planning*, *20*(3), 19.

Ward, J.P., Griffiths, P. & Whitmore, P. (1994). *Strategic planning for information systems.* New York: John Wiley & Sons.

Wiseman, C. (1998). *Strategic information systems*. Homewood, IL: Irwin.

KEY TERMS

Information Technology Competitive Advantage (ITCA): The increased profitability and market share by a better positioning due to strategy and structure selection based on the alignment model.

Information Technology Fit: The best relation that can be achieved in a strategic alignment model among information technology, business structure, and business strategy.

Strategic Alignment Model: A model to help appropriate selection of strategic priorities and structural variables. Incorporates a dynamic interaction among business strategy, structure, information technology, and environment, while taking into account the most important contingencies like: the age, the size, the culture, etc.

Strategic Information Systems: Information systems that increase competitive advantage in a long-term horizon, with more than operational benefits, such as better positioning and increase in market share.

Structural Design: Forms of structure derived by optimal combinations of structural variables with information technology and business strategy in a way that increase competitive advantage.

Structural Variables: The most important elements of structure (variables) that characterize and form the structural design (span of control, levels of hierarchy, level of standardization, etc.).

Cache Management for Web-Powered Databases

Dimitrios Katsaros
Aristotle University of Thessaloniki, Greece

Yannis Manolopoulos
Aristotle University of Thessaloniki, Greece

INTRODUCTION

In recent years, the World Wide Web, or simply the Web (Berners-Lee, Caililiau, Luotonen, Nielsen, & Secret, 1994), has become the primary means for information dissemination. It is a hypertext-based application and uses the hypertext transfer protocol (HTTP) for file transfers.

During its first years, the Web consisted of static hypertext markup language (HTML) pages stored on the file systems of the connected machines. When new needs arose, e.g., database access, it was realized that we could not afford in terms of storage to replicate the data we want to publish in the Web server's disk in the form of HTML pages. So, instead of static pages, an application program should run on the Web server to receive requests from clients, retrieve the relevant data from the source, and then pack the information into HTML or extensible markup language (XML) format. Even the emerged "semistructured" XML databases that store data directly into the XML format need an application program that will connect to the database management system (DBMS) and retrieve the XML file or fragment. Thus, a new architecture was born: in the traditional couple of a Web client and a Web server, a third part is added, which is the application program that runs on the Web server and serves data from an underlying repository that, in most cases, is a database. This architecture is referred to as *Web-powered database* and is depicted in Figure 1. In this scheme, there are three tiers: the database back-end, the Web/application server, and the Web client.

BACKGROUND

Due to the existence of *temporal locality* in Web request streams, we can exploit the technique of *caching*, that is, temporal storage of data closer to the consumer. Caching can save resources, i.e., network bandwidth, because fewer packets travel in the network, and time, because we have faster response times. Caching can be implemented at various points along the path of the flow of data from the repository to the final consumer. So, we may have caching at the DBMS, at the Web server's memory or disk, at various points in the network (i.e., proxy caches), or at the consumer's endpoint. Web proxies may cooperate so as to have several proxies to serve each other's misses. All the caches present at various points comprise a *memory hierarchy*. The most important part of a cache is the mechanism that determines which data will be accommodated in the cache space and is referred to as the cache *admission/replacement policy*.

Requests for "first-time accessed" data cannot benefit from caching. In these cases, due to the existence of *spatial locality* in request streams, we can exploit the technique of preloading or *prefetching*, which acts complementary to caching. Prefetching may increase the amount of traveling data, but on the other hand, it can significantly reduce the latency associated with every request.

The role of a cache is to store temporally a set of objects that will most probably be requested by its clients. A cache replacement policy assigns a value to every cached object, called *utility value (UV)*, and evicts from cache the object with the least utility value. The aim of the replacement policy is to improve the cache's effectiveness by optimizing two performance measures: the *hit ratio* and the *cost-savings ratio (CSR)*. The former measure is defined as:

$$HR = \sum h_i / \sum r_i$$

and the latter is defined as

$$CSR = \sum c_i * h_i / \sum ci * r_i$$

where h_i is the number of references to object i satisfied by the cache out of the r_i total references to i, and c_i is the cost of fetching object i in cache. The cost can be defined either as the object's size s_i or as the downloading latency c_i. In the former case, the CSR coincides with the byte–hit ratio (BHR); in the latter case, the CSR coincides with the delay-savings ratio (DSR).

Figure 1. Architecture of a typical Web-powered database

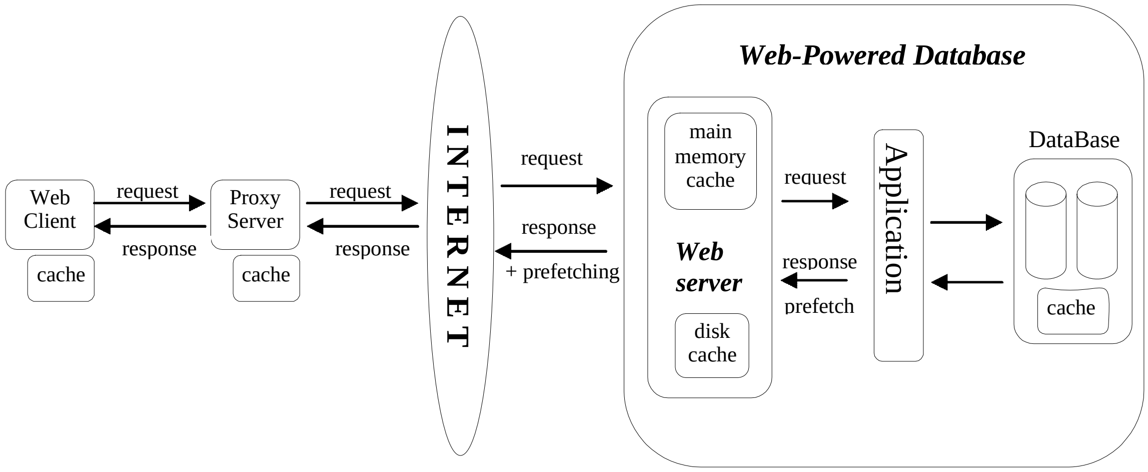


Table 1. A list of factors differentiating Web caching from traditional caching

1. Variable Object Size
The Web object's size varies considerably. It ranges from a few bytes (e.g., small HTML or text files) to several megabytes (e.g., large multimedia files). In contrary, the objects that move through the levels of the caching hierarchy in operating systems or database systems have fixed size, which is equal to the size of a disk block.

2. Variable Fetching Cost
The cost (time penalty) for retrieving a Web object varies significantly. Different objects may have different fetching costs, because they differ in size or in their distance from the client (in terms of network hops). Moreover, the same file may have different fetching costs at different time instances, depending on the server (e.g., heavy load) and network conditions (e.g., congestion).

3. The Depth of the Web Caching Hierarchy
Because caching can happen nearly anywhere, including on the server, on the user's machine, at Internet service protocols (ISPs), at telecommunication companies, at the peering points of national networks, etc., the depth of this hierarchy is significantly larger than the respective depth in computer systems. The large depth of this hierarchy significantly affects the characteristics of the request stream.

4. The Origin of the Web Request Streams
The requests seen by the Web caches, especially by the proxies and the reverse proxies, are not generated by a few programmed processes like the request streams encountered in traditional computer systems. They mainly originate from large human populations with diverse and varying interests.

CACHING IN WEB-POWERED DATABASES

Web cache replacement (Katsaros & Manolopoulos, 2002) is one of the most important areas of Web caching for several reasons. First, studies have shown that the cache HR and BHR grow in a *log-like fashion* as a function of cache size (Breslau, Cao, Fan, Phillips, & Shenker, 1999). Thus, a better algorithm that increases HR by only several percentage points would be equivalent to a several-fold increase in cache size. Second, the growth rate of Web content is much higher than the rate with which memory sizes for Web caches are likely to grow. Finally, the benefit of even a slight improvement in cache performance may have an appreciable effect on network traffic, especially when such gains are compounded through a hierarchy of caches. There are several factors that distinguish Web caching from caching in traditional computer architectures (see Table 1).

The majority of the replacement policies focus on the first two factors. The main drawback in the design of these policies is that they do not achieve a balance between HR and CSR. Some of them are called *recency-based policies* and favor the HR, e.g., the Greedy Dual-Size (Cao & Irani, 1997), the Size-Adjusted LRU (Aggrawal, Wolf, & Yu, 1999), whereas some others are called *frequency-based policies* and favor the CSR (BHR or DSR), e.g., LRV (Rizzo & Vicisano, 2000), LFU-DA (Diley & Arlitt, 1999), and LNC-R-W3 (Shim, Scheuermann, & Vingralek, 1999). Notable exceptions are the LUV (Bahn, Koh, Noh, & Min, 2002), GD* (Jin & Bestavros, 2001), and CRF (Katsaros & Manolopoulos, 2004), which try to combine recency and frequency. The drawback of LUV and GD* though, is the existence of some manually tunable parameters. CRF is a self-tunable policy that gracefully combines recency and frequency and addresses all the particularities of the Web caching environment.

PREFETCHING IN WEB-POWERED DATABASES

Prefetching is employed in order to cure caching's shortcomings. An effective and efficient prefetching scheme should maximize the number of cache hits due to its action and at the same time minimize the incurred cost due to the prefetched objects. This cost may represent cache space, network bandwidth, server overloading, etc. Although the implementation mechanism of prefetching is very important, the core issue in employing a prefetching scheme is the deduction of future object requests.

In general, there exist two alternatives for the deduction of the future client requests. The first is termed *informed prefetching*, and the second is termed *predictive prefetching*. In Table 2, we explain why the former is not appropriate for the Web environment.

A predictive prefetching algorithm is based on the notion of Markov predictors. Let $T_r = \langle tr_1, \ldots, tr_k \rangle$ be a sequence of consecutive requests (called a transaction) and $S = \langle d_1, \ldots, d_n \rangle$, $n \leq k$ be a sequence of accesses, which is a subsequence of T_r. Given a collection of nTr transactions, where the sequence S appears fr(S) times, then the appearance probability *P(S)* of S is equal to *P(S)* = fr(S)/nTr.

If $S = \langle d_1, \ldots, d_n \rangle$ is a sequence of accesses, then the conditional probability that the next accesses will be to $d_{n+1}, \ldots, d_{n+m}$ is $P(d_{n+1}, \ldots, d_{n+m} | d_1, \ldots, d_n)$, and it equals the following:

$$P(d_{n+1},\ldots,d_{n+m} \mid d_1,\ldots,d_n) = \frac{P(d_1,\ldots,d_n,d_{n+1},\ldots,d_{n+m})}{P(d_1,\ldots,d_n)}$$

Given a collection of transactions, rules of the form

$$d_1, \ldots, d_n \Rightarrow d_{n+1}, \ldots, d_{n+m} \tag{1}$$

can be derived, where $P(d_{n+1}, \ldots, d_{n+m} | d_1, \ldots, d_n)$ is not less than a user-defined value T_c. $P(d_{n+1}, \ldots, d_{n+m} | d_1, \ldots, d_n)$ is

Table 2. Informed and predictive prefetching

Informed Prefetching Can be applied in cases where the client program knows exactly the resources the client is going to request and reveals them into the cache (Patterson, Gibson, Ginting, Stodolsky, & Zelenka, 1995). Informed prefetching is actually a scheduling policy subject to a set of constraints regarding cache space, timeliness of prefetching, and available bandwidth. It requires the communication bandwidth between the applications and the cache to be stable. Thus, it can be implemented only in cases where the cache is embedded into the application, e.g., databases and operating systems.
Predictive Prefetching How will future references be deduced if not revealed by the client program? The only possibility is to take advantage of the spatial locality present in Web request streams. Spatial locality captures the coreference of some resources. The remaining issue is to "quantify" spatial locality, that is, to discover the dependencies between references for different data. Such dependencies can be discovered from past requests and can be used for making predictions about future requests.

the confidence of the rule. The left part of the rule is called the *head* and has size equal to *n*, whereas the right part is called the *body* and has size equal to *m*. The dependency of forthcoming accesses on past accesses defines a *Markov chain*.

Definition of a *n - m* Markov Predictor

An n - m Markov predictor calculates conditional probabilities $P(d_{n+1}, \ldots, d_{n+m}|d_1, \ldots, d_n)$ between document accesses and discovers rules of the form (1), which are used to predict future accesses.

A predictive prefetching algorithm is defined as a collection of n - 1, n - 2, …, n - m Markov predictors. Existing predictive Web prefetching mechanisms can be categorized into three families.

The algorithms of the first family use the concept of the *dependency graph* (DG) (Padmanabhan & Mogul, 1996; Jiang & Kleinrock, 1998). The DG has a node for each object that has ever been accessed and an arc from node X to node Y, if and only if Y was accessed within w accesses after X and the same client did both accesses. DG maintains the number of accesses to each node X and the number of transitions from X to Y. It calculates conditional probabilities $P(d_i|d_j)$ for all d_i, d_j belonging to a transaction. Therefore, DG discovers rules of the form $d_i \Rightarrow d_j$, or equivalently, it uses a 1-1 Markov predictor.

The algorithms belonging to the second category are based on the notion of a *k-order PPM predictor* (Fan, Cao, Lin, & Jacobson, 1999). A k-order PPM predictor maintains j - 1 Markov predictors, for all $1 \le j \le k$ (k is a user-specified constant). It employs Markov predictors that have the constraint that the preceding j requests must be consecutive in the request stream. These Markov predictors calculate conditional probabilities of the form $P(d_{n+1}|d_n), \ldots, P(d_{n+1}|d_{n-k+1}, \ldots, d_n)$ and determine the corresponding rules, which have head sizes equal to 1,2, …, k.

A generalization of the above two families was presented in Nanopoulos, Katsaros, and Manolopoulos (2003). That scheme uses *n - m* Markov predictors that calculate conditional probabilities of the form $P(d_{n+1}, \ldots, d_{n+m}|d_1, \ldots, d_n)$, though the document accesses need not be consecutive. Moreover, the maximum head size *n* and the maximum body size *m* are not set by the user but are adaptively estimated by the data.

FUTURE TRENDS

We all know that the Internet, and consequently the Web, faces performance problems. Performance problems arise in any of the following three general areas:

1. Web server processing delays
2. Internet delays
3. “Last-mile” delays (delays between the subscriber and the Internet, e.g., due to a slow dial-up modem connection)

The idea in alleviating these problems is to make content delivery from origin servers more “distributed,” moving some of their content to the “edge” of the Internet. Based upon this idea, the content delivery networks (CDN) emerged recently. CDN are designed to take advantage of the geographic locations of end users. Rather than serving content from the origin Web site, the content distribution model makes copies of “key” content on multiple content delivery servers sites distributed through the Internet, close to the users requesting that content.

A CDN, running on thousands of servers distributed across the Internet, contracts with content providers to deliver their content. A CDN addresses efficiently the aforementioned performance problems:

1. With the “hottest” content “outsourced,” the load on the origin server is reduced.
2. The connection from a local content delivery server is shorter than between the origin Web server and the user, thus reducing latency.
3. Because many users share the CDN servers, this service greatly increases the hit ratio.

The deployment of large-scale CDNs brings some new challenges to caching. One very important issue is that of object placement, that is, the selection of the edge servers where an object should be stored (Tang & Chanson, 2002). This problem could also be seen in combination with that of object replacement (Korupolu & Dahlin, 2002). Another very important issue is related to maintaining the coherency of the cached objects, especially in the case that these objects are dynamically generated (Tang & Chanson, 2004).

Prefetching is also challenging in the context of CDN because of the huge number of clients served. To deal with this problem, the cache satellite distribution systems (Armon & Levy, 2004), CSDS for short, have emerged as a technology for feeding the caches with the information clients are expected to request, ahead of time. In such a system, the participating caches periodically report to a central station about requests received from their clients. The central station selects a collection of Web documents that are “pushed” via a satellite broadcast to the participating caches, so that upon a future local request for the documents, they will already reside in the local cache and will not need to be fetched from the terrestrial network.

CONCLUSION

Web-powered databases have emerged to support numerous Web applications, like e-commerce systems, but due to the large number of visitors they attract, they face serious performance problems. To reduce or even alleviate such problems, the techniques of caching and prefetching have been employed. The different physical characteristics of the Web objects, the existence of hyperlinks, and the nature of the Web caching hierarchy call for different caching and prefetching solutions than those investigated in traditional operating and database systems. The present article recorded the most critical factors that affect the design of caching and prefetching policies for Web-powered databases and surveyed the major families and their representatives for these policies.

REFERENCES

Aggrawal, C., Wolf, J., & Yu, P. S. (1999). Caching on the World Wide Web. *IEEE Transactions on Knowledge and Data Engineering*, *11*(1), 94–107.

Armon, A., & Levy, H. (2004). Cache satellite distribution systems: Modeling, analysis, and efficient operation. *IEEE Journal on Selected Areas in Communications*, *22*(2), 218–228.

Bahn, H., Koh, K., Noh, S. H., & Min, S. L. (2002). Efficient replacement of nonuniform objects in Web caches. *IEEE Computer*, *35*(6), 65–73.

Berners-Lee, T., Cailiiliau, R., Luotonen, A., Nielsen, H. F., & Secret, A. (1994). The World Wide Web. *Communications of the ACM*, *37*(8), 76–82.

Breslau, L., Cao, P., Fan, L., Phillips, G., & Shenker, S. (1999). Web caching and Zipf-like distributions: Evidence and implications. In *Proceedings of the IEEE Conference on Computer Communications* (INFOCOM) (pp. 126–134). Washington: IEEE Press.

Cao, P., & Irani, S. (1997). Cost-aware WWW proxy caching algorithms. In *Proceedings of the USENIX Symposium on Internet Technologies and Systems* (USITS) (pp. 193–206). USENIX Press.

Dilley, J., & Arlitt, M. (1999). Improving proxy cache performance: Analysis of three replacement policies. *IEEE Internet Computing*, *3*(6), 44–55.

Fan, L., Cao, P., Lin, W., & Jacobson, Q. (1999). Web prefetching between low-bandwidth clients and proxies: Potential and performance. In *Proceedings of ACM International Conference on Measurement and Modeling of Computer Systems* (SIGMETRICS) (pp. 178–187). ACM Press.

Jiang, Z., & Kleinrock, L. (1998). An adaptive network prefetch scheme. *IEEE Journal on Selected Areas in Communications*, *16*(3), 358–368.

Jin, S., & Bestavros, A. (2001). GreedyDual* Web caching algorithm: Exploiting the two sources of temporal locality in Web request streams. *Computer Communications*, *24*(2), 174–183.

Katsaros, D., & Manolopoulos, Y. (2002). Cache management for Web-powered databases. In D. Taniar & W. Rahayou (Eds.), *Web-powered databases* (pp. 201–242). Hershey, PA: Idea Group Publishing.

Katsaros, D., & Manolopoulos, Y. (2004). Caching in Web memory hierarchies. In *Proceedings of the ACM Symposium on Applied Computing* (SAC) (pp. 1109–1113). ACM Press.

Korupolu, M. R., & Dahlin, M. (2002). Coordinated placement and replacement for large-scale distributed caches. *IEEE Transactions on Knowledge and Data Engineering*, *14*(6), 1317–1329.

Nanopoulos, A., Katsaros, D., & Manolopoulos, Y. (2003). A data mining algorithm for generalized Web prefetching. *IEEE Transactions on Knowledge and Data Engineering*, *15*(5), 1155–1169.

Padmanabhan, P., & Mogul, J. (1996). Using predictive prefetching to improve World Wide Web latency. *ACM SIGCOMM Computer Communication Review*, *26*(3), 22–36.

Patterson, H. R., Gibson, G. A., Ginting, E., Stodolsky, D., & Zelenka, J. (1995). Informed prefetching and caching. In *Proceedings of the ACM Symposium on Operating System Principles* (SOSP) (pp. 79–95). ACM Press.

Rizzo, L., & Vicisano, L. (2000). Replacement policies for a proxy cache. *IEEE/ACM Transactions on Networking*, *8*(2), 158–170.

Shim, J., Scheuermann, P., & Vingralek, R. (1999). Proxy cache algorithms: Design, implementation and performance. *IEEE Transactions on Knowledge and Data Engineering*, *11*(4), 549–562.

Tang, X., & Chanson, S. T. (2002). Coordinated en-route Web caching. *IEEE Transactions on Computers*, *51*(6), 595–607.

Tang, X., & Chanson, S. T. (2004). The minimal cost distribution tree problem for recursive expiration-based consistency management. *IEEE Transactions on Parallel and Distributed Systems*, *15*(3), 214–227.

KEY TERMS

Caching Proxy: A caching proxy or proxy server or proxy is a server that acts as an intermediary between a client and a content server. It intercepts the requests of the client and checks whether it can serve the client from its own cache, and if not, it forwards the requests to the content server.

Capacity Miss: This miss occurs because the cache cannot accommodate all requested objects.

CDN (Content Distribution Network or Content Delivery Network): This is a network of cache servers owned by the same Internet Service Provider that delivers content to users on behalf of content providers. CDN servers are typically shared, delivering content belonging to multiple Web sites, though all servers may not be used for all sites.

Compulsory Miss: A compulsory or cold-start miss occurs in the first access to a Web object.

CPN (Capacity Provisioning Network): This is a network of cache servers owned, operated, and coordinated through capacity trading by different Internet Service Providers. Unlike CDN, with the purpose of replicating content from specifically contracted content providers, CPN's goal is to cache whatever content users access from around the world of content servers. Qualitatively, a CDN services the supply side of content distribution; a CPN services the demand side.

Flash Crowd Problem: It occurs when the request load overwhelms some aspect of a Web site's infrastructure, such as the front-end Web server, network equipment or bandwidth, or the back-end transaction-processing infrastructure. The resulting overload can crash a site or cause unusually high response times.

Spatial Locality: Describes the effect that when an object is referenced, its "nearby" objects will tend to be referenced soon. The notion of "nearness" can include the minimum number of hyperlinks that link the two objects, or it can describe semantic nearness.

Temporal Locality: Describes the effect that an object, which has just been referenced, has a high (or increased) probability of being referenced again in the near future. Formally: let $p_i(k)$ denote the probability that, following a reference to object i, the next reference to object i occurs within k references. We say that a particular object shows temporal locality of reference if there exists a $k > 0$ such that $p_i(k) > 1 - (1 - 1/n)^k$.

Case-Based Learning in Computer Information Systems

Morgan Jennings
Metropolitan State College of Denver, USA

Charles H. Mawhinney
Metropolitan State College of Denver, USA

Janos Fustos
Metropolitan State College of Denver, USA

INTRODUCTION

How can we retain computer information systems (CIS) students? A decline in enrollment similar to that which occurred in the 80's (Mawhinney, Callaghan, & Cale, 1989) is the motivating factor for this question. A google™ search on declining enrollments in information systems brings up reports supporting this trend. DePaul University, for example, had increased undergraduate enrollments "in all colleges but the School for New Learning and the School of Computer Science, Telecommunications and Information Systems" (DePaul University, 2003). A report from the California Community College system listed the top 15 curricular areas of declining FTE's (Perry, 2003); Computer and Information Science and Computer programming made the list. Our own Computer Information Systems (CIS) and Computer Science programs have fewer students enrolled.

BACKGROUND

A recent comparison study (Jennings, Mahwinney, & Fustos, 2003) confirmed past research (Mawhinney et al., 1989) indicating that students perceive that an IS career provides less work-place interaction than they desire. Lack of understanding regarding an IS career is not uncommon. As Mawhinney et al. (1989) and von Hellens and Nielson (2001) found, students believe that CIS entails working alone and requires a high level of math. Because of the scarce or inaccurate knowledge regarding a CIS degree or career, students may avoid CIS programs.

In addition to a lack of understanding related to what an IS worker does in his/her job, retaining students within entry-level college IS courses is a problem (Myers & Beise, 2001). Many students feel they are computer literate until they enter an IS program. The skills that they possess and the skills needed within an IS degree are likely to be disparate (Easton & Easton, 2002; Karsten & Roth, 1998). Rather than a sink or swim attitude on the part of colleges and universities, time spent coaching and encouraging students on relevant computer skills for the IS degree may help them feel confident and able to complete the program (Compeau, Higgins, & Huff, 1999). This means more than showing students the benefits of technology or how to use a computer. It may require providing meaningful and relevant learning situations in which to use the technology (Venkatesh, 1999) that are similar to actual work-related scenarios (Gallivan, 2000).

Without an accurate picture, it is difficult for students to understand the work-style of information system professionals. The content of CIS courses is technically-oriented and many students struggle with the content, usually in isolation. This, however, belies the business environment where employees usually work together on projects. Learning should take place in a similar environment. In addition, research suggests that there is a synergistic learning effect within group environments (Ryan, Bordoloi & Harrison, 2000; Savery & Duffy, 1995). For example, better understanding of a system and development of more accurate mental models was found in a group environment (Gallivan, 2000). An unconstructive aspect of such a learning situation is that negative comments may affect the attitude of group members (Gallivan, 2000), or "group think" may stifle creativity.

MAKING THE CASE

Given that the team or group-based projects approach is a ubiquitous part of a career in this field, authentic learning environments could provide students with a more realistic model of the IS field. With authentic learning environments, the necessary information technology skills are embedded in the learning process. Also embedded in the learning process are the soft skills that managers say are lacking in employees, such as problem solving, communicating effectively and working in group environments. (Lee & Trauth, 1995; Todd, McKeen, & Gallupe, 1995).

There is a body of research and theory on the importance of providing authentic and relevant learning environments. This is often within the context of constructivism, an umbrella term that identifies a learning philosophy. Constructivism gained wide-spread acceptance in the educational technology discipline in the early nineties (Duffy & Jonassen, 1991; Lebow, 1993; Savery & Duffy, 1995; Vanderbilt Cognition and Technology Group, 1990), although it has been in existence in different forms for decades. It is gaining acceptance within information systems and computer science education as well (Ben-Ari, 2001; Boyle, 2000). The Web, in particular, has provided a platform where collaborative learning environments can be supported (e.g., Shang, Shi & Chen, 2001; Vat, 2001).

Cases (Harvard Business School, 2003; Potvin, 2000) or problem-based learning (PBL) (Barrows, 1993, 1985), situated cognition (Duffy & Jonassen, 1991), learner centered (Henson, 2003) and cooperative learning and anchored instruction (Vanderbilt Cognition and Technology Group, 1990) are some of the terms or models that fit within a constructivist paradigm. While some aspects may vary, the overriding principles include active engagement, collaboration and personal relevance. The reasons for learning are embedded within rich, authentic contextually relevant environments (Harvard Business School, n.d.; Lebow, 1993). The outcome is the ability to reason and problem-solve ill-structured situations. These types of environments are more likely to develop ownership on the part of the students involved (Savery & Duffy, 1995) and may promote what Agarwal & Karahanna (2000) call cognitive absorption and Jennings (2002) calls cognitive aesthetics.

The Harvard Case Method (Harvard Business School, 2003) and PBL are probably the most well known examples. Neither Harvard cases nor PBL are new concepts. The Harvard Case Method originated in 1913 (Harvard Business School, 2003). The objective of the case method was to enhance judgment and decision-making capability and it is widely used because of its success (Harvard Business School, 2003; Potvin, 2000). Problem-based learning was developed in the mid 50's for medical education (Savery & Duffy, 1995) and like the Harvard Case Method, provides a student with a problem that they must solve. Most often, when a student resolution has been reached, it is compared to the solution of the company or physician from which the case originated, although this is not always the situation. Whether or not a comparison is made, the goal is for students to acquire the knowledge and skills to solve problems in their respective fields.

Constructivist models are designed for ill-structured material and provide relevant, engaging, authentic contexts from which to draw conclusions. Teachers (or facilitators) are mentors and guides who challenge students to problem-solve; they model higher order thinking skills (Barrows, 2003; Savery & Duffy, 1995). These models all include interaction between peers which is a successful means to engage students (Vanderbilt Cognition and Technology Group, 1990).

They are however, not particularly easy to implement because of the time commitment and the difficulty in distilling a problem into an appropriate case or authentic environment in which to anchor instruction. The problems for Harvard cases come from the business world, such as a company implementing an e-business model. It can take between 1-4 months to develop a case and Harvard Business School (HBS) considers four factors when devising a case (Harvard Business School, 2003):

1. The issues on which the case will focus;
2. The analysis required on the part of students to address those issues;
3. The data required in the case to perform that analysis satisfactorily; and
4. Where the data can be obtained.

Two sites that provide many free cases are the Idea Group Inc. Web site (http://www.idea-group.com/cases_pp/login.asp) and the HBS Web site (http://harvardbusinessonline.hbsp.harvard.edu/b02/en/cases/cases_home.jhtml). Idea Group has information systems cases while HBS provides general business cases. In addition, the following links (retrieved February 12, 2004) have either example cases/problems or information regarding this type of learning environment.

- Multidisciplinary case-based learning for undergraduate students: http://www.blackwell-synergy.com/links/doi/10.1034/j.1600-0579.2000.040404.x/abs/
- The Center for Problem-Based Learning (PBL) at Samford University:http://www.samford.edu/pbl/
- University of Delaware (a list of links): http://www.udel.edu/pbl/others.html
- Problem-Based Learning: An Introduction: http://www.ntlf.com/html/pi/9812/pbl_1.htm

Another potential source for cases is a college's local business community. Working with a local company to develop a case provides a unique opportunity for students to interact with the company. Student feedback strongly indicates that this opportunity increases the relevance, engagement and desire to implement a worthwhile solution to a problem.

While Harvard cases are based on actual business situations, the conditions developed for problem-based learning come from well-formulated, realistic problems that mirror real-world situations. The following instructional principles (Savery & Duffy, 1995) provide a context from which to create appropriate problems:

1. Anchor all learning activities to a larger task or problem.
2. Support the learner in developing ownership for the overall problem or task.
3. Design an authentic task.
4. Design the task and the learning environment to reflect the complexity of the environment they should be able to function in at the end of learning.
5. Give the learner ownership of the process used to develop a solution.
6. Design the learning environment to support and challenge the learner's thinking.
7. Encourage testing ideas against alternative views.
8. Provide opportunity for and support reflection on both the content learned and the learning process.

Objectives for a problem include specific content learning. Student evaluation can be completed through peer and self-evaluation as well as traditional testing based on the objectives. Traditional testing may help alleviate an instructor's concerns regarding individual student learning.

FUTURE TRENDS

It is not enough to simply provide a case or problem to be solved. In order for this paradigm to work, students must become a team and team building must be taught. Wells (2002) uses Katzenbach and Smith's definition of team work: "A team is a small number of people with complementary skills who are committed to a common purpose, performance goals, and approach for which they hold themselves mutually accountable."

A cursory review of articles indicates that many faculty members are using a team approach in their courses. Fewer, it seems, include instruction on how to be a team member. As team work becomes prevalent within the college curriculum, we will likely see more instruction in team building. Without it, students often perceive group work with disdain: With guidance on team building, students are more positive in regard to the experience (Wells, 2002).

Companies have high expectations for CIS graduates (Shawyunm, 1999). It behooves us as educators to provide the business community with students who have had realistic CIS experience, which authentic learning environments provide. Team building concepts could be taught in the ubiquitous entry-level computing course so that students have the skills to work effectively in groups throughout their college term and careers.

CONCLUSION

We discussed the concepts of teamwork, using authentic problems or cases and how this type of model could be applied to courses in an undergraduate CIS program. Using this type of environment in CIS college programs would provide students with real-world skills and an accurate understanding of CIS. With authentic learning environments and team building instruction, the technology and soft skills are embedded in the learning process. By using this learning approach CIS majors experience interesting and varied problems during their student tenure. They enter the work force with immediately usable skills and a realistic understanding of their career of choice.

REFERENCES

Agarwal, R., & Karahanna, E. (2000). Time flies when you're having fun: Cognitive absorption and belief about information technology usage. *MIS Quarterly, 24*(4), 665-694.

Barrows, H.S. (1985). *How to design a problem-based curriculum for the preclinical years*. New York: Springer Publishing.

Barrows, H. (1993). An overview of the uses of standardized patients for teaching and evaluating clinical skills. *Academic Medicine*, *68*, 443-451.

Barrows, H. (2003). Response to: The problem with problem-based medical education: Promises not kept by R. H. Glew. *Biochemical Molecular Biology Education, 31*, 255–256.

Ben-Ari, M. (2001). Constructivism in computer science education. *Journal of Computers in Mathematics and Computer Science Teaching, 20*(1), 45-73. Retrieved January 24, 2004, from *http://dl.aace.org/638*

Boyle, T. (2000). Constructivism: A suitable pedagogy for information and computing sciences. *Proceedings of the LTSN-ICS 1st Annual Conference.* Retrieved October 29, 2003, from *http://www.ics.ltsn.ac.uk/pub/conf2000/papers/tboyle.htm*

Brown, J.S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-41.

Compeau, D., Higgins, C.A., & Huff, S. (1999). Social cognitive theory and individual reactions to computing technology: A longitudinal study. *MIS Quarterly, 23*(2), 145-158.

DePaul University. (2003). Enrollment matters. *Enrollment Management Newsletter*. Retrieved January 17, 2004, from *http://www.depaul.edu/em/forms/enrollmentmatters1003.doc*

Duffy, T., & Jonassen, D. (1991). Constructivism: New implications for educational technology? *Educational Technology, 31*(5), 7-12.

Easton, G., & Easton, A. (2003). Assessing computer literacy: A comparison of self-assessment and actual skills. In T. McGill (Ed.), *Current issues in IT education* (pp.238-254). Hershey, PA: Idea Group Publishing.

Gallivan, M.J. (2000). Examining workgroup influence on technology usage: A community of practice perspective. *Proceedings of the ACM SIGCPR*, Chicago, IL, 38, 49-60.

Harvard Business School (n.d.). *The case method at HBS*. Retrieved February 12, 2004, from Harvard School of Business Web site: *http://www.hbs.edu/mba/experience/learn/thelearningmodel/thecasemethod.html*

Harvard Business School (2003). Making a case: The birth of an HBS case study. *Enterprise Newsletter*. Retrieved January 17, 2004, from Harvard School of Business Web site: *http://www.hbs.edu/corporate/enterprise/case.html*

Henson, K.T. (2003). Foundations for learner centered education. *Education, 124*(1), 5-16. Retrieved January 15, 2004, from WilsonWeb database.

Jennings, M.M. (2002). What do good designers know that we don't? In M. Khosrow-Pour (Ed.), *Web-based instructional learning* (pp. 235-241). Hershey, PA: IRM Press.

Jennings, M.M., Mawhinney, C.H., & Fustos, J. (2003). Making the case for case-based learning in computer information systems. In T. McGill (Ed.), *Current issues in IT education* (pp.11-25). Hershey, PA: Idea Group Publishing.

Karsten, R., & Roth, R. (1998). The relationship of computer experience and computer self-efficacy to performance in introductory computer literacy courses. *Journal of Research on Computing in Education, 31*(1), 14-24.

Lambert, L., & Walker, D. (1995). *Learning and leading theory: A century in the making*. In L. Lambert, D. Walker, D.P. Zimmerman, J.E. Cooper, M.D. Lambert, M.E. Gardner, & P.J. Slack (Eds.), The constructivist leader (pp. 1-27). New York: Teachers College Press, Columbia University.

Lebow, D. (1993). Constructivist values for instructional system design: Five principles toward a new mindset. *Educational Technology Research and Development, 41*(3), 4-16.

Lee, D.M.S., & Trauth, E.D. (1995). Critical skills and knowledge requirements of IS professionals: A joint academic/industry investigation. *MIS Quarterly, 19*(3), 313-340.

Mawhinney, C.H. Callaghan, D.R. & Cale, E.G. (1989). Modifying freshman perceptions of the CIS graduates workstyle. *ACM SIC CSE Bulletin, 21*(1), 78-82.

Myers, M.E., & Beise, C.M. (2001). Nerd work: Attractors and barriers perceived by students entering the IT field. *Proceedings of the ACM SIGCPR*, San Diego, CA (pp.201-204).

Perry, P. (2003, September). *Access lost: An examination of supply constriction and rationing in the California community college system*. Retrieved January 15, 2004, from California Community College Chancellor's Office Web site: *www.cccco.edu*

Potvin, J. (2000, December). The case method. *Bellanet Advisor*. Retrieved January 10, 2004, from *http//www.bellanet.org/advisor/index.cfm?Fuseaction=view_article&TheArticle=38*

Ryan, S., Bordoloi, B., & Harrison, D.A. (2000). Acquiring conceptual data modeling skills: The effect of cooperative learning and self-efficacy on learning outcomes. *The DATA BASE for Advances in Information Systems, 31*(4), 9-24.

Savery, J.R., & Duffy, T.M. (1995). Problem-based learning: An instructional model and its constructivist framework. *Educational Technology, 35*, 31-37.

Shang, Y., Shi, H., & Chen, S. (2001). An intelligent distributed environment for active learning. *ACM Journal of Educational Resources in Computing, 1*(2), 308-315. Retrieved January 20, 2004, from ACM Digital Library.

Shawyunm, T. (1999). Expectations and influencing factors of IS graduates and education in Thailand: A perspective of the students, academics and business community. *Informing Science, 2*(1), 19-32.

Todd, P.A., McKeen, J.D., & Gallupe, R.B. (1995). The evolution of IS job skills: A content analysis of IS job advertisements from 1970 to 1990. *MIS Quarterly, 19*(1), 1-18.

Vanderbilt Cognition and Technology Group. (1990). Anchored instruction and its relationship to situated cognition. *Educational Researcher, 19*(6), 2-10.

Vat, K. (2001). Web-based asynchronous support for collaborative learning. *Journal of Computing in Small Colleges, 7*(2), 322-344. Retrieved October 12, 2003, from ACM Digital Library.

Venkatesh, V. (1999). Creation of favorable user perceptions: Exploring the role of intrinsic motivation. *MIS Quarterly, 23*(2), 239-260.

von Hellens, L., & Nielson, S. (2001). Australian women in IT. *Communications of the ACM, 44*, 46-56.

Wells, C.E. (2002). Teaching teamwork in information systems. In E.B. Cohen (Ed.), *Challenges of information technology education in the 21st century* (pp.1-24). Hershey, PA: Idea Group Publishing.

KEY TERMS

Active Learning: As a group, earners read, write, discuss, or are otherwise engaged in solving problems. Students engage in such higher-order thinking tasks as analysis, synthesis, and evaluation.

Anchored Instruction: Learning is anchored in a real world context. Within this context, learners engage in different activities to solve a problem, such as math calculations or geographic coordinates to travel across a continent or sale a ship.

Cognitive Apprenticeships: Students work in teams on projects or problems with close scaffolding of the instructor. Cognitive apprenticeships are representative of Vygotskian "zones of proximal development" in which student tasks are slightly more difficult than students can manage independently, requiring the aid of their peers and instructor to succeed.

Constructivism: A theory of learning and knowing that holds that learning is an active process of knowledge construction in which learners build on prior knowledge and experience to shape meaning and construct new knowledge (Lambert & Walker, 1995). Important concepts include: (a.) students are active rather than passive learners, (b.) students take responsibility for learning as an individual and within a social setting, (c.) learning must be embedded in contexts that have real-world relevance.

Harvard Case-Based Learning: Students are given a realistic case relevant to the course (information systems, accounting, management, etc.). Students work through the case, in or out of class and decide what should be done. They then meet with the entire class, or in groups and discuss and resolve the case.

Problem-Based Learning: Students are confronted with a well chosen, ill-structured problem that mirrors real-world problems. Students define problems, identify what information is needed and engage in solution generation and decision making.

Situated Cognition: A "situated cognition" theory of knowledge is characterized by the principle that the activity in which knowledge is developed and deployed is not separable from learning and cognition. Rather it is an integral part of what is learned (Brown, Collins & Duguid, 1989).

C

Census Data for Health Preparedness and Response

Jane L. Garb
Baystate Medical Center, USA

Richard B. Wait
Baystate Medical Center, USA

INTRODUCTION

The epidemiologist works with researchers in various disciplines as well as public and private health practitioners who are responsible for maintaining and improving the health of the population. Health is defined as "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity"(WHO, 1948). The role of the epidemiologist is to provide information on the extent, etiology and risk of disease or injury. Key to this role is knowledge of the location, size and demographic makeup of the population at risk for developing disease or injury. By integrating this information with the location of cases as well as risk factors, epidemiologists can make a vital contribution to disease and injury prevention, intervention and response. This applies both to endemic or "usual" levels of both chronic diseases like cancer or heart disease and infectious diseases like pneumonia or influenza, and injuries like gunshot wounds or motor vehicle accidents. It also applies to disease epidemics or outbreaks like SARS; attacks by biological or chemical weapons such as Anthrax, ricin or sarin; and inadvertent natural or technological disasters including earthquakes, transportation disasters or widespread power interruptions. This chapter explores the types of census data for disease surveillance, prevention and intervention.

BACKGROUND

Sources of Geographic and Population Data

The decennial census (*Census 2000*, 2001) and American Community Survey (*American Community Survey*, 2003; *Meeting 21st Century Demographic Data Needs Implementing the American Community Survey: May 2002*, 2002) are the major sources of population data in a GIS. The census is conducted every 10 years. It provides a complete count of the U.S. population based on the census "short form" given to every individual. In addition, it provides detailed socioeconomic and demographic information for individuals, families, and households and selected information about housing, based on the "long form", given to a 1/6 sample of the population.

The American Community Survey (ACS) is an ongoing survey by the U.S. Census Bureau on an approximate 2.5% sample of the population, with over-sampling for small governmental units such as American Indian Reservations. Data is collected monthly and reported annually using detailed demographic and economic questions from the Census long form, with additional questions such as those on grandparents as care-givers. Currently the American Community Survey is limited to an annual survey of 700,000 households in 31 sites across country. However, full implementation will begin in 2004. The first data products for the smallest areas and populations will be available in 2009. Once data products are produced for a population group or area, they will be updated each year (*American Community Survey Operations Plan*, 2003). ACS survey quality measures are high, and these help compensate for somewhat smaller sample size and larger confidence intervals of the ACS compared to the decennial census long form (Griffin & Obenski, 2002).

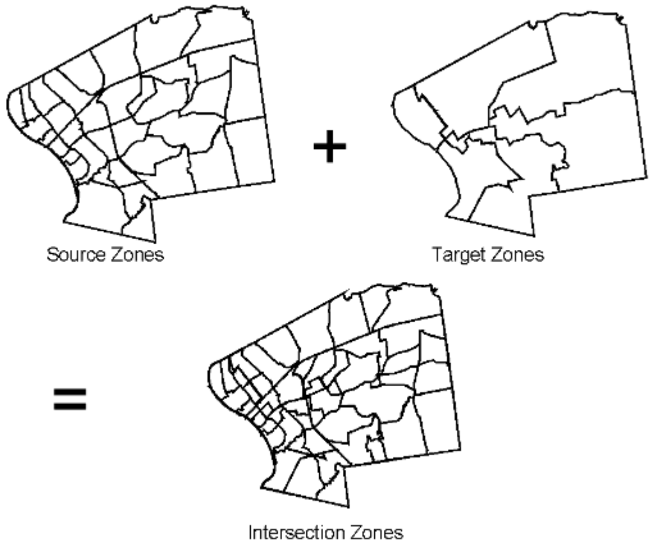
Census geographic data products are in the form of "TIGER"files (*TIGER/Line Files, Redistricting Census 2000*, 2001) which consist of street centerlines and address ranges as well as boundaries for census geographic units such as census tracts, block groups, cities and town boundaries, counties and regions.

All of these data products for the entire U.S. are publicly available from the Census Bureau Web site (http://www.census.gov) or on CD for a minimal price.

Daytime Population Estimates

The ACS holds the most promise for current population estimates because of its ongoing nature. However, other sources of daytime population estimates show utility when combined with additional data. The Census Transportation Planning Package (CTPP) (*Census Transportation and Planning Package (CTPP)*, 2003) by the Department of Transportation is a special tabulation of Census

Figure 1. Components of areal interpolation



2000 data (long form responses) that includes information according to place of work, place of residence and travel between home and work. Discussions are underway about CTPP products that would be based on ACS data.

Another survey is the Longitudinal Employer - Household Dynamics survey (LEHD) (*Longitudinal Employer - Household Dynamics*, 2003), based on a partnership between the Census Bureau and currently 10 states. This survey integrates state administrative data and census data products. It includes place of residence and place of work for workers by age, sex, and industry. Also an ongoing survey, it has ties to the ACS, which will improve its accuracy and potentially provide extremely accurate estimates of daytime populations in the future. Neither of these latter two surveys is fully implemented at this point, and this will depend on future congressional funding.

Data Integration in a GIS

We can integrate census data with health data in a GIS for prevention, intervention and response to diseases in all forms. This can be done for *endemic* or usual levels of chronic diseases like cancer or infectious diseases like pneumonia or influenza; infectious disease *epidemics* or outbreaks like the recent SARS epidemic; or sudden *attacks* with biological or chemical weapons. Census data allows us to perform critical GIS functions essential to mapping and analysis of spatial data in health.

To begin with, census geographic TIGER files are used as reference files for geocoding case or event locations. Census boundaries in these files are used in data aggregation of geographic areas and in the GIS function known as polygon overlay, where two map layers are overlaid on one another. This is the basis for a critical method necessary for utilizing census data for health called areal interpolation (Goodchild & Lam, 1980).

Both ACS and decennial census data are available only in aggregate form for fixed geographical units. These units – census tracts, block groups, counties – are administrative units used by the census to aid in census data collection, or by political entities to apportion representation and govern. However, the most useful geographic units for health applications, in which data are often available – neighborhoods, school districts, hospital service areas, and so forth – do not coincide with the census geography for which population counts are available.

Since 9/11, the scope of health has extended from preparedness against naturally occurring diseases and chronic conditions to biochemical weapons. The unique geography required for using census data in routine healthcare is even more of a challenge in the case of a bioterrorist attack, where the appropriate geographical units – police sectors, drainage sub-basins or the plume resulting from a chemical or biological release – may be unpredictable, and different depending on the nature of the biochemical weapon used.

Areal interpolation (AI) becomes a key method for addressing the challenge of estimating the population for the geography of healthcare. It starts with original geographic units for which data such as population counts is known. These are called "source zones" (see Figure 1). Units for which population counts or other information is unknown but desired, in this case, neighborhoods, schools districts, and so forth, are called "target zones".

A GIS can calculate the intersection of these two layers through the polygon overlay operation. The result is "intersection zones". The population totals or other attribute data for the source zones are allocated to each of the intersection zones according to one of several available statistical methods of varying complexity (Bracken & Martin, 1989; Burrough, 1986; Flowerdew & Green, 1994; Flowerdew, Green, & Kehris, 1991; Flowerdew & Openshaw, 1987; Green, 1988, 1990; Lam, 1983; Langford, Masser, & Unwin, 1989; Langford & Unwin, 1994; Martin & Bracken, 1991; Mugglin, Carlin, & Gelfand, 2000; Mugglin, Carlin, Zhu, & Conlon, 1999; Tobler, 1979). Then the population totals for the intersection zones are aggregated into the target zones. AI methods rely on the GIS primarily for calculating the area of the intersection zones. The simplest methods allocate the population into each intersection zone based on the proportion of the source zone area represented by the intersection zone (areal weighting) or whether the center of the target zone falls within the source zone (point-in-polygon). The LandView project (Census, 2004) of the U.S. Census Bureau provides easy-to-use software employing the latter type of method. Although these methods are relatively easy to calculate, they assume that the population is evenly distributed across the source zone, which is not usually the case. The more sophisticated AI methods do not make this assumption, but assume rather that the population is clustered around roads, houses, night-time lights, according to soil type or elevation, and so forth. The population is allocated into the target zones according to the presence of these factors. The LandScan Global Population Project (Laboratory) employs some of the latter methods in the process of estimating affected populations on a world-wide basis.

Using areal interpolation, census population data can be available for geographical units which are relevant to healthcare. Some examples of such units are hospital service areas, community health service areas, neighborhoods, school districts, police sectors, and water service areas. This allows much greater flexibility in using census data in health. The assumptions of the areal interpolation method used, its computational complexity and the degree of estimation accuracy needed must be kept in mind when choosing the appropriate areal interpolation method for a particular purpose.

Disease Surveillance and Community Profiles

We can assess the risk of disease to particular populations in a community by calculating disease rates. Rates tell us how prevalent endemic diseases are or what the risk is of getting the disease in an epidemic or biochemical attack. A rate is composed of a numerator divided by a denominator. For endemic or epidemic diseases, hospitals or public health departments have the numerators for disease rates – they know how many cases of disease there are and where they are located. For example, for "reportable" infectious diseases, physicians report cases

Figure 2. Rate of violent incidents by police sector in Springfield, MA

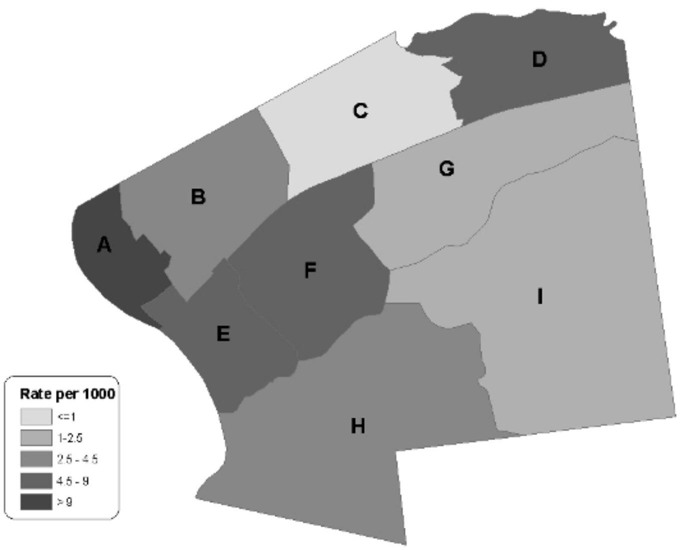
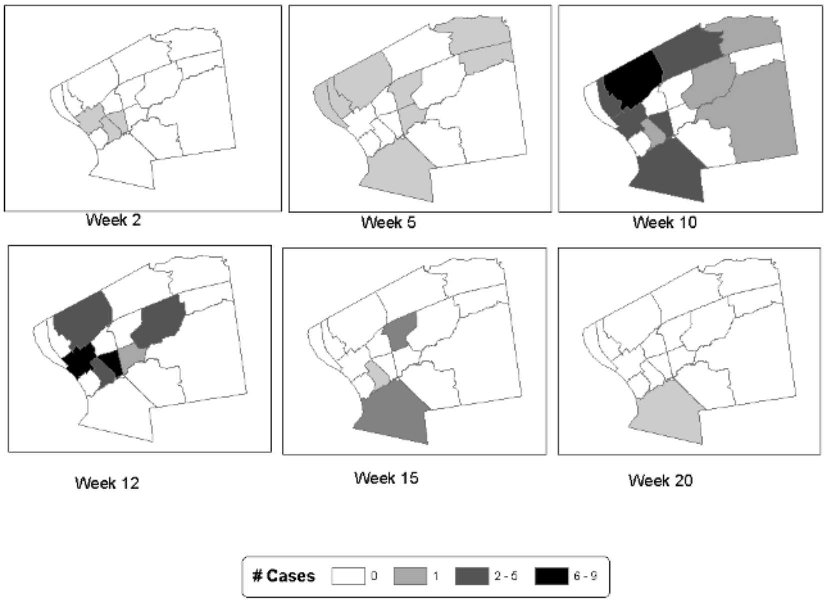


Figure 3. Time course of Shigellosis epidemic in Springfield, MA



to their local or state public health department, who then sends this information to the Centers for Disease Control. Incidence and prevalence rates for these diseases are published in local and state public health reports, and nationally in the *Morbidity and Mortality Weekly Report* (CDC). Many hospitals maintain disease registries for chronic diseases such as cancer, heart disease or neurological disorders. This data is often sent to state or national disease registries. Some of this data can obtained by directly contacting state public health departments.

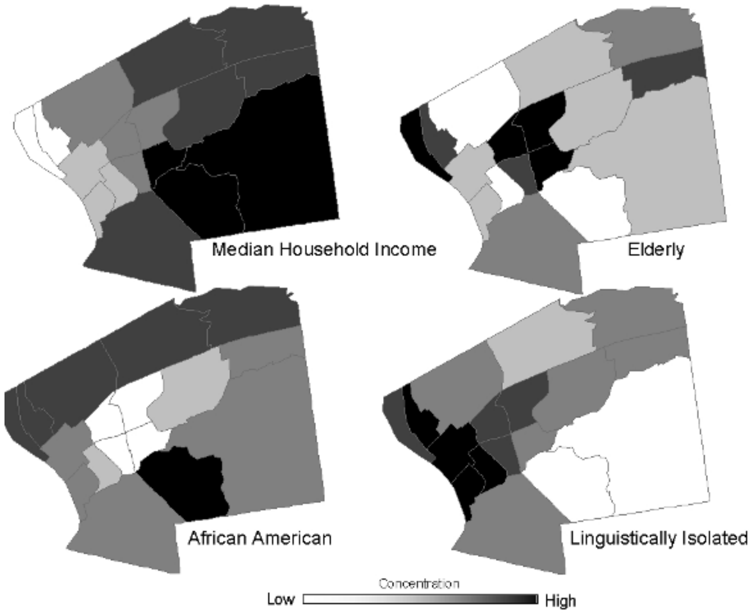
As discussed earlier, these types of data are usually available only in aggregate form, summarized as incidence rates for arbitrary geographic units like zip codes, which do not coincide with census geographic units.

Census data provides the denominators for these rates, so, through AI if necessary, we can estimate the risk of disease in our community for all residents or for different geographic areas or population subgroups. The map in Figure 2 is an example of how disease rates can be displayed for cancer surveillance. It represents the prevalence of late stage breast cancer in Springfield, Massachusetts – the number of existing cases of late stage disease for every 10,000 women by *neighborhood*. In this example, the exact location of breast cancer cases was obtained from the two hospital cancer registries in Springfield which represent 95% of reported cases in the city. Population totals by neighborhood were estimated from census block data using areal interpolation. This map is called a "choropleth" or color-coded map. The color intensity indicates the level of disease. The darker the shade, the higher is the risk of late stage breast cancer and visa versa.

Disease incidence rates can also be displayed temporally in a series of maps, to follow the course of disease over time. The time frame can be days or weeks for an epidemic or bioterrorist attack, or years for chronic diseases. This information on disease surveillance can be used to direct medical resources for intervention or prevention of endemic diseases or emergency treatment for the victims of an epidemic or biological attack. Figure 3 illustrates how GIS was used to combine local public health data on case locations with census population estimates to plot the course of a local shigella epidemic over time.

Census data also provides *both* numerators and denominators for demographic and economic rates, which allows us to visualize the socioeconomic character of a community. Figure 4 shows population concentrations for linguistically isolated, elderly, and African American residents as well as the distribution of median household income in Springfield using GIS software ArcGIS (ESRI, 1999a). In the case of a biological attack, identifying demographic subgroups known to be at increased risk could direct efforts to prevent the spread of the biological agent. For example, in a mass vaccination effort against smallpox, geographic areas with high concentrations of

Figure 4. Socioeconomic profile of Springfield, MA



the elderly could be targeted for the first wave of vaccination. In the event of a chemical weapon attack, areas with high concentrations of linguistically isolated residents would require foreign-language communication of emergency instructions.

Putting endemic disease rates and demographic rates together and mapping them in a GIS, we can visualize how the current geographic pattern of disease relates to demographic and economic patterns in the community. Pattern similarities between disease rates and census demographic rates indicate a possible relationship between the two. This forms the basis for hypotheses to be tested with spatial analysis.

Techniques such as spatial regression (Bailey & Gatrell, 1995) can be used to identify significant socioeconomic and demographic risk factors for disease – characteristics of population subgroups that put them at increased risk for disease. In this type of spatial analysis, census tracts, blocks, neighborhoods or other geographic areas become the unit of analysis. Disease rates in the different geographic areas are compared. Census population and housing characteristics are tested to see if they are related to the geographic distribution of disease. This type of analysis accounts for spatial proximity of geographic areas.

We can use the results of this type of analyses to design intervention and prevention programs for endemic diseases. It directs us to community groups with which to partner in implementing educational and outreach programs, and indicates what social and cultural factors should be taken into account in order to design more effective programs.

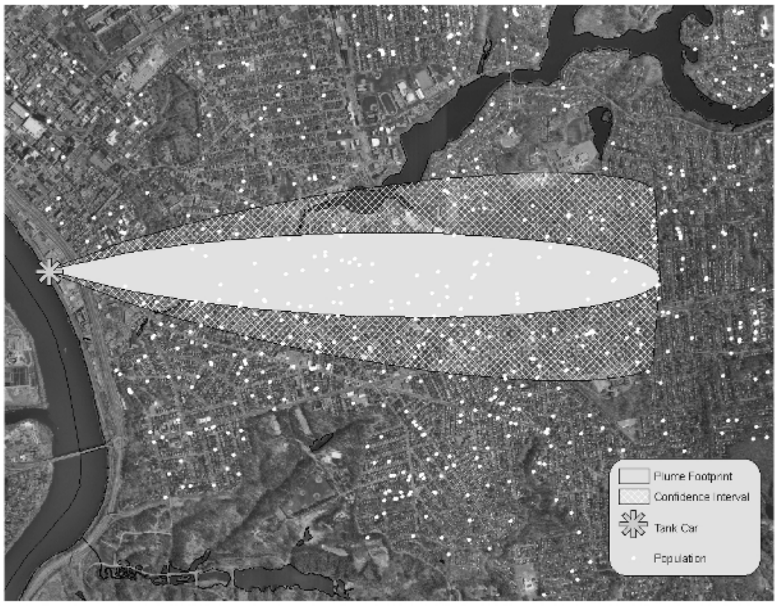
FUTURE TRENDS

Locating Affected Populations for Disaster Response

An increasingly important consideration in the context of homeland security and public health is determining the geographic location and size of the population at risk of exposure to biochemical weapons or other weapons of mass destruction. Special-purpose shareware programs such as ALOHA (Areal Locations of Hazardous Atmospheres, 2003), HAZUS (HAZUS(Hazards U.S.), 2003) or CATS (ESRI, 1999b) which model the dispersion of toxic chemicals, can be combined with areal interpolation methods in a GIS to help in this challenge.

For example, in a disaster drill at Baystate Medical Center in conjunction with the City of Springfield, a fictitious railroad tank car containing ammonia was hit by a terrorist attack and exploded. In addition to causing mass casualties at the site of the explosion, a plume of ammonia gas was released into the air. ALOHA was used to calculate the size and location of the plume of ammonia gas that was released based on knowledge of elevation and current wind and temperature conditions. This was

Figure 5. Footprint of amonia gas plume overlaid on population distribution in the affected area



mapped and overlaid on a map of the population distribution of the area, based on current American Community Survey data (Figure 5). The areal weighting method was used to estimate the size of the population affected by the plume which would likely be in need of medical care. This method was chosen because quick calculations could be computed and re-computed on-the-fly to accommodate the shifting plume, giving rough population estimates sufficient to determine the magnitude of response necessary. Without our calculations, hospital administrators were estimating about 10-12 casualties. Our calculations showed that the population inside the plume was approximately 25,000. In this case, ease-of-computation was more important than having the most accurate population estimates.

CONCLUSION

Census data has several key uses for the epidemiologist. TIGER/line files are integral to the GIS functions of data aggregation, address geocoding, areal interpolation. Population counts from the decennial census and the American Community Survey provide denominators for disease rates and show community demographic profiles which can be used in disease surveillance as well as the analysis of socioeconomic factors in disease, program design and locating high-risk populations in a disaster or disease epidemic. This article has provided some practical applications of how the epidemiologist can combine census and health data in a GIS in order to promote and protect the health of our population on a continual basis as well as in times of crisis.

REFERENCES

ALOHA (Areal Locations of Hazardous Atmospheres). (Version U.S. Dept. of Commerce)(2003). National Oceanic and Atmospheric Administration.

American Community Survey. (2003). Washington, DC: U.S. Census Bureau.

American Community Survey Operations Plan. (2003). Washington, DC: U.S. Census Bureau.

Bailey, T., & Gatrell, A.C. (1995). *Interactive spatial data analysis* (pp. 282-289, spatial regression; p. 252, definition of GIS). Crown Books.

Bracken, I., & Martin, D. (1989). The generation of spatial population distributions from census centroid data. *Environment and Planning A, 21*, 537-543.

Burrough, P.A. (1986). Principles of GIS for land resources assessment (pp. 30-32). Oxford: Clarendon.

Burrough, P.A., & McDonnell, R. A. (1998). *Principles of geographic information systems*. Oxford: Oxford University Press.

CDC, C. f. D. C. *Morbidity and Mortality Weekly Report*.

Census 2000. (2001). Washington, DC: U.S. Department of Commerce, Bureau of the Census, Geography Division.

Census Transportation and Planning Package (CTPP). (2003). Washington, DC: U.S. Census Bureau. Census, U.S. (2004). *LandView®, 6*

A Viewer for the Environmental Protection Agency, U.S. Census Bureau, and U.S. Geological Survey Data and Maps (2004) from *http://www.census.gov/geo/landview/*

ESRI. (1999a). ArcView GIS (Version 3.2). Redlands, CA: Environmental Systems Research Institute, Inc.

ESRI. (1999b). CATS for Emergency Response. *ArcNews*.

Flowerdew, R., & Green, M. (1994). Areal interpolation and types of data. In S. Fotheringham & P. Rogerson (Eds.), *Spatial analysis and GIS* (pp. 121-145). London: Taylor and Francis.

Flowerdew, R., & Openshaw, S. (1987). *A review of the problems of transferring data from one set of areal units to another incompatible set* (No. 4). Newcastle, UK: Northern Regional Research Laboratory.

Flowerdew, R., Green, M., & Kehris, E. (1991). Using areal interpolation methods in geographic information systems. *Papers in Regional Science, 70*, 303-315.

Goodchild, M.F., & Lam, N.S.-N. (1980). A real interpolation: A variant of the traditional spatial problem. *Geo-Processing, 1*, 297-312.

Green, M. (1988). *Statistical methods for areal interpolation: The EM algorithm for count data*. Lancaster, UK: North West Regional Research Laboratory.

Green, M. (1990). Statistical models for areal interpolation. Paper presented at the *EGIS '90: Proceedings, First European Conference on Geographical Information Systems*, Utrecht, The Netherlands.

Griffin, D.H., & Obenski, S.M. (2002). *A discussion of the quality of estimates from the American Community Survey for Small Population Groups* (No. 2). Washington, DC: U.S. Census Bureau, U.S. Department of Commerce.

HAZUS(Hazards U.S.). (2003). Washington, DC: Federal Emergency Management Agency (FEMA).

Laboratory, O.R.N. (2002). LandScan. A global population database for estimating populations at risk from *http://sedac.ciesin.columbia.edu/plue/gpw/landscan/*

Lam, N. (1983). Spatial interpolation methods: A review. *The American Cartographer, 10*, 129-140.

Langford, M., & Unwin, D. (1994). Generating mapping population density services within a geographical information system. *The Cartographic Journal, 31*, 21-26.

Langford, M., Masser, I., & Unwin, D. J. (1989). *Modeling population distribution using remote sensing and GIS* (No. 3): Midlands regional research laboratory.

Longitudinal Employer - Household Dynamics. (2003). Washington, DC: U.S. Census Bureau.

Martin, D., & Bracken, I. (1991). Techniques for modelling population-related raster databases. *Environment and Planning A, 23*(1069-1975).

Meeting 21st Century Demographic Data Needs Implementing the American Community Survey: May 2002. (2002). Washington, DC: U.S. Census Bureau.

Mugglin, A.S., Carlin, B.P., & Gelfand, A.E. (2000). Fully model-based approaches for spatially misaligned data. *Journal of the American Statistical Association, 95*(451), 877-887.

Mugglin, A.S., Carlin, B.P., Zhu, L., & Conlon, E. (1999). Bayesian areal interpolation, estimation, and smoothing: an inferential approach for geographic information systems. *Environment and Planning A, 31*(8), 1337-1352.

TIGER/Line Files, Redistricting Census 2000. (Version Redistricting Census 2000)(2001). Washington, DC: U.S. Department of Commerce, Bureau of the Census, Geography Division.

Tobler, W.R. (1979). Smooth pycnophylactic interpolation for geographic regions. *Journal of the American Statistical Association, 74*, 519-530.

WHO (1948). Text of the constitution of the World Health Organization. *Official Record of the World Health Organization, 2*, 100.

KEY TERMS

American Community Survey: An ongoing survey conducted by the U.S. Census Bureau, which collects detailed demographic and socioeconomic information on a sample of the population.

Areal Interpolation: A method of estimating counts or quantities for one configuration of geographic units (target zones) based on known counts from a different configuration of the same geographic space (source zones).

Choropleth Map: A color-coded map, also referred to as a "thematic" map, in which geographic areas are portrayed in different hues or intensity according to their value on some quantity being displayed.

Disease Rates: The level of disease in a given time period, population and geographic area. It represents the proportion of disease cases in the exposed population or population-at-risk.

Endemic: The usual level of disease in a geographic area in a given period of time

Epidemic: An excessive level of disease in a geographic area in a given period of time

Geographic Information System (GIS): "A computer-based set of tools for capturing, (collecting), editing, storing, integrating, analyzing and displaying spatially referenced data" (Bailey & Gatrell, 1995).

Health: A state of complete physical, mental and social well-being and not merely the absence of disease or infirmity (WHO, 1948).

Spatial Regression: A spatial analytic technique modeling the relationship of various factors to the geographical distribution of some attribute measured on a continuous scale.

Certifying Software Product and Processes

Hareton Leung
Hong Kong Polytechnic University, Hong Kong

INTRODUCTION

Software certification can not only greatly improve the overall quality of software; it can also help to control the cost of software development. Because software has traditionally been perceived as intangible, it has commonly been certified on the basis of the thoroughness of its development methodology. Certification of the development process is based on the assumption that the development process can assure that the developed product complies with specifications. However, software can also be certified at the product level. In this chapter, we will present a process model that captures the essential aspects of process and product certification (Silva, 2002).

BACKGROUND

Software can be certified at two levels: *process certification* and *product certification*. Process certification evaluates the process against a standard or model process to which it should conform. In Europe, the most popular process evaluation model is ISO9001 (Cianfrani, 2001). In America the focus of process quality improvement and evaluation is a process *maturity* model such as the Capability Maturity Model for Software (SW-CMM) and the Capability Maturity Model Integration (CMMI) (Chrissis, 2003; Humphrey, 1995; Paulk, 1993). Yet another example of process certification is Bootstrap from Europe (Card, 1993).

Product certification involves directly assessing the equivalence of key attributes of software at the level of its specifications (specified service), and its behaviour (actual service). A software product can be characterised by its functional and non-functional attributes. Examples of product certification include the Department of Defense Y2K certification process (DoD) and the Hong Kong Article Numbering Association Software Certification Program (Hong Kong Article, 1987).

CERTIFICATION MODELS

We first present two certification models, one for process certification and the other for product certification. We then present a generic certification model for both process and product certification. There are two key participants in the certification process, each playing a different role: the certification body and certificate applicant (or the software developer).

For the presentation of the certification process, we have adopted the IDEF0 notation (Klingler).

ProCess Certification (PCC)

Figure 1 shows the five major stages of the Process Certification (PCC) model.

PCC-0: Review of Certification Guidelines

The certification body studies the process requirements, analyses the relationships between the process requirements and the certification model, and identifies important requirements. It then issues and publishes the certification guidelines. An applicant will then study the certification model and process requirements and from that gathers important information about the implementation of the process. The applicant then includes the certification requirements in the production process before the process assessment.

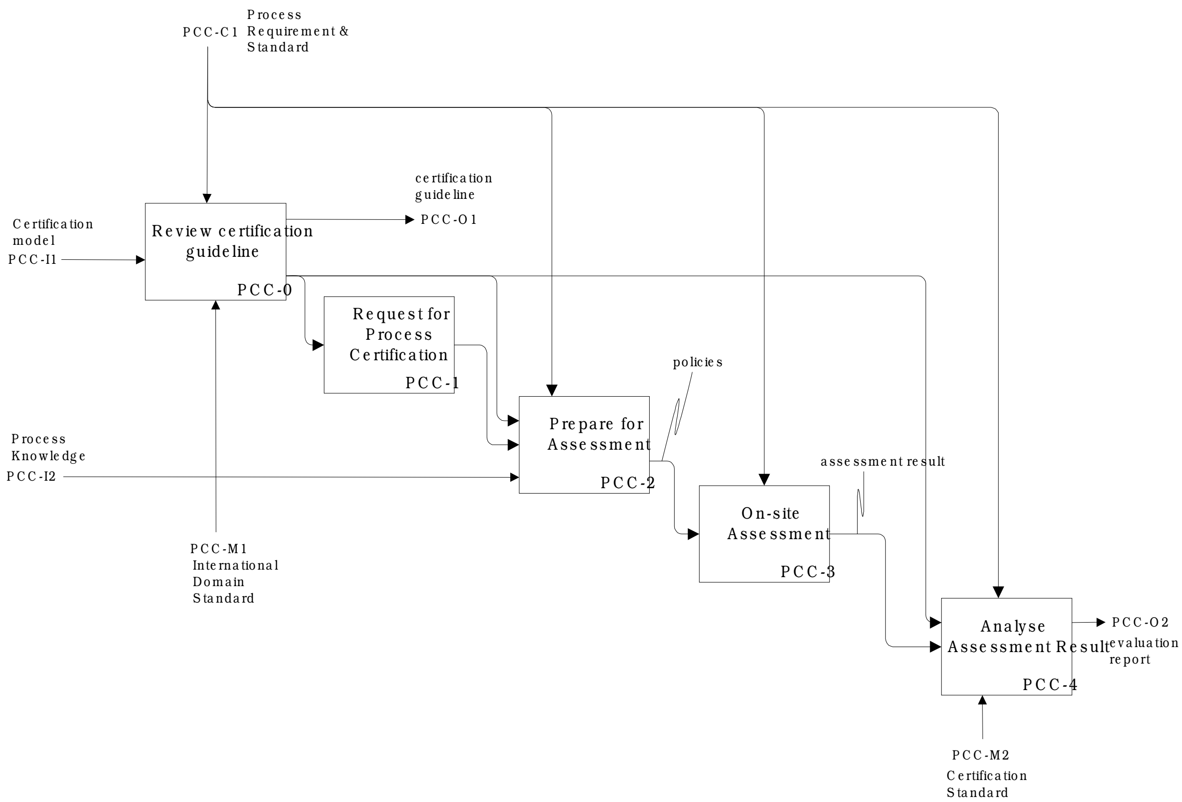
PCC-1: Request for Process Certification

After implementing the certification requirements in its production process, the applicant will apply to have the production process certified and will submit an application form to the certification body. The certification body will process the certification request and generate an application number to identify each certification.

PCC-2: Preparation for Assessment

The certification body will prepare the assessment guidelines and certification requirements to assess the applicant's production process. The body may familiarise the applicant's staff with the certification process by providing a pre-assessment service and audit training. The applicant's staff should ensure that the production processes fulfil the requirement of certification. The applicant then requests that the process be certified.

Figure 1. PCC model



PCC-3: On-Site Audit

The certification body will send assessors to the applicant's development site. Assessors will follow the assessment and certification guidelines to collect the process information and assess the applicant's production process. The applicant's staff should provide the necessary information to help assessors assess the production process. Finally, the assessors should produce an assessment report.

PCC-4: Analysis of Assessment Results

The certification body evaluates the assessment result to determine whether the production process passes the assessment and then returns the final result to the applicant. The applicant should analyse the assessment result to identify areas for improvement. Generally, to ensure that it keeps pace with the environmental and technological changes, the certification body should evaluate its certification guidelines after each certification.

Left-hand columns of Table 1 summarize the process stages of the PCC model.

ProDuct Certification (PDC)

Like the PCC model, the ProDuct Certification (PDC) model also consists of five major stages. The right-hand columns of Table 1 summarise the process stages of the PDC model.

The PCC and the PDC models not only both have five stages, they are also very similar in other ways. Indeed, although they differ in some details, most of their procedures have the same purposes and provide similar functions. For example, PCC-0, PCC-1, and PCC-4 are similar to, respectively, PDC-0, PDC-1, and PDC-4. The key difference between these models is that PCC focuses on the software production process, while PDC focuses on the software product.

Other specific differences between the PCC and PDC models are as follows.

Table 1. PCC model and PDC model description

	PCC Model *Certification body*	*Applicant*	**PDC Model** *Certification body*	*Applicant*
	PCC-0: *Review*	*Certification guideline*	**PDC-0:** *Review*	*Certification guideline*
	Gather the certification model information to issue, update and maintain certification guidelines.	Study the process requirement, and process certification model. Apply to the production process.	Gather all the domain knowledge, the industrial practice; issue and maintain a certification guideline.	Study the entire domain requirement, and domain certification model.
Input	Certification model	Certification model	Certification model	Certification model
Control	Process requirement	Process requirement	Domain requirement	Domain requirement
Support	International process standard	Nil	International domain standard	International domain standard
Output	Certification guideline	Quality policy	Domain certification model	Quality policy
	PCC-1: *Request for*	*Process certification*	**PDC-1:** *Request for*	*Product certification*
	Process the applications	Submit an application	Process the applications	Submit an application
Input	Certification guideline	Application form	Domain certification model	Quality policy
Control	Nil	Nil	Nil	Nil
Support	Nil	Nil	Nil	Nil
Output	Application no.	Certification process request	Application category	Certification request
	PCC-2: *Prepare for*	*Assessment*	**PDC-2:** *Prepare for*	*assessment*
	Help the applicant to produce documentation.	Implement quality policy, pre-assess before real assessment.	Study the certification guideline, arrange the certification schedule.	Study the certification guideline, prepare for the certification.
Input	Certification guideline	Certification guideline memo	Application category	Certification guideline
Control	Process requirement	Nil	Domain requirement	Domain requirement
Support	Nil	Nil	Nil	Nil
Output	Assessment schedule	Quality policy & documentation	Assessment material & schedule	Assessment schedule
	PCC-3: *On-site audit*		**PDC-3:** *Perform*	*Assessment*
	Perform on-site audit to assess the applicant's production processes.	Cooperate with the assessor to assess the production processes.	Following the assessment guideline, assess the product.	Submit the product and the product documentation.
Input	Assessment guideline	Company policies	Assessment material	Product and product information
Control	Process requirement	Process requirement	Domain requirement	Nil
Support	Nil	Nil	Nil	Nil
Output	Assessment result	Production processes information	Assessment result	Product and product information
	PCC-4: *Analyse*	*assessment result*	**PDC-4:** *Evaluation*	*assessment*
	Evaluate the assessment result.	Analyse the assessment result and identify improvement.	Evaluate the assessment result and generate an assessment report	Analyse the assessment result and identify improvement.
Input	Assessment results	Assessment result & its process	Assessment result	Assessment result & its product
Control	Process requirement	Process requirement	Domain requirement	Domain requirement
Support	Certification standard	Nil	Certification standard	Nil
Output	Evaluation result	Improvement list	Evaluation result	Improvement list

In the stage "Preparation for Assessment" PCC (PCC-2) requires the preparation of the assessment guidelines in order to assess the process. The corresponding stage of PDC requires only the preparation of assessment material in order to assess the product.

PCC-3 and PDC-3 differ in the assessment criteria that they use in the assessment stage. PCC involves some degree of subjectiveness, as the assessment is partly dependent on the assessors. In contrast, PDC often uses objective criteria based on the test execution result. Another difference is that as PCC requires that assessors appraise the staff's working procedures and the production methodologies, the process is assessed at the development site. PDC, however, will assess the product in a special laboratory, which is not usually at the applicant's own development site.

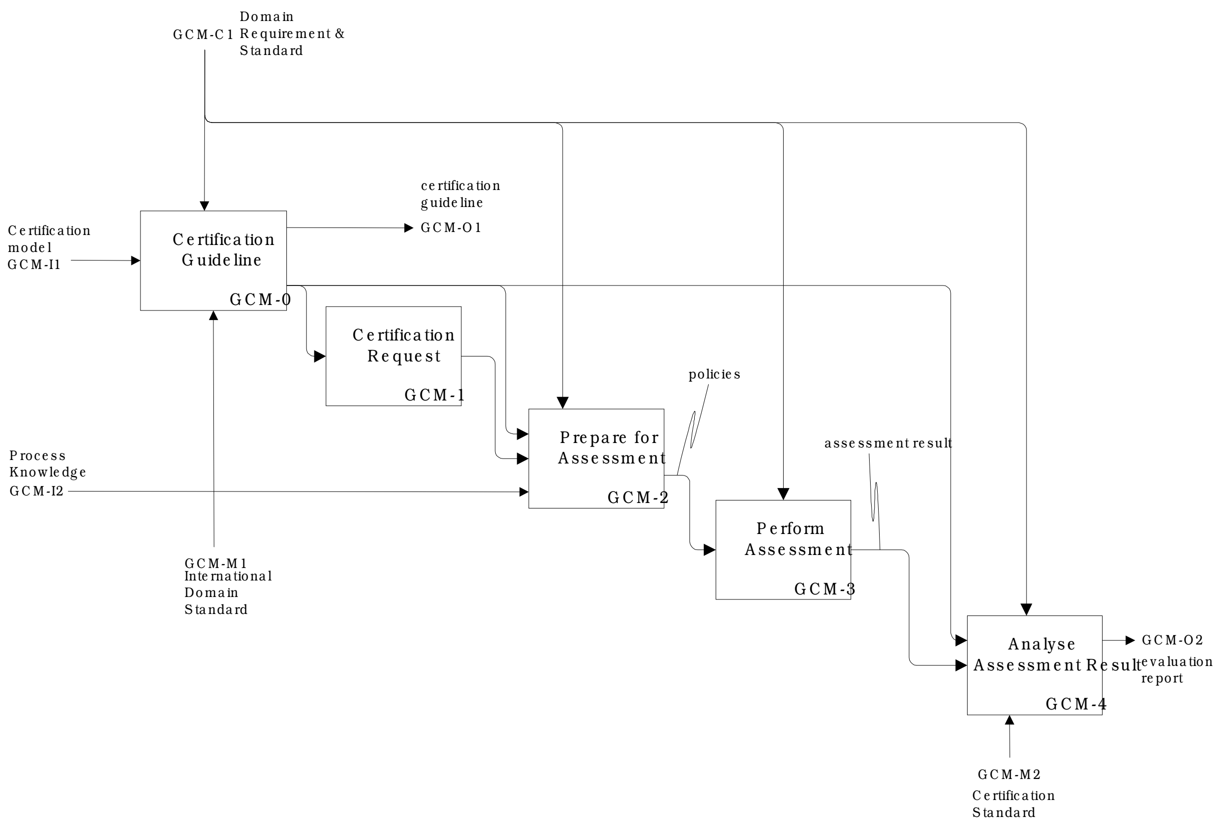
A Generic Certification Model (GCM)

A Generic Certification Model (GCM) for both process and product certification processes has been developed, as shown in Figure 2. The GCM is a natural and obvious adaptation of both PCC and PDC. Its model description is given in Table 2. GCM consists of five stages: establishment of certification guidelines, request for certification, preparation for assessment, performance of assessment, and analysis of assessment results.

FUTURE TRENDS

One area that has not been addressed in the general certification process model is the tailoring requirements. As some companies have their own specific needs and requirements, it would be advantageous that a tailoring

Figure 2. GCM



guideline be developed so that the GCM model can be adjusted to ease its adoption.

Newer process models, such as the project management processes promoted by the Project Management Institute (PMI, 2004) and various system outsourcing models proposed in the literature, are gaining acceptance. These new processes will have a positive impact on the final quality of a software system. We need more studies to understand the effect of implementing these models.

CONCLUSION

This article describes a general certification process model that includes both process certification and product certification. GCM captures the essence of most certification schemes and can be used as a basis for the development of various kinds of process and product certification methods. As the GCM components are reusable, they can increase the efficiency and effectiveness of the development of new certification methods. GCM can also serve as a basic checklist to ascertain whether a certification method has all the key elements required for carrying out certification. GCM has been evaluated by matching it against several process certification and product certification schemes (Leung & Li, 2002).

REFERENCES

Card, D. (1993, May). Bootstrap: Europe's assessment method. *IEEE Software*, 93-95.

Chrissis, M.B., Konrad, M., & Shrum, S. (2003). *CMMI guidelines for process integration and product improvement.* Addison Wesley.

Cianfrani, C.A., Tsiakals, J.J., West, J.E., & West, J. (2001). *ISO 9001: 2000 explained.* American Society for Quality.

DoD Y2k certification process. *http://www.army.mil/army-y2k/CertificationProcess.htm*

Hong Kong Article Numbering Association Software Certification Program Information Kit. (1987). Document No.: IK980713. Information Technology Division Hong Kong Productivity Council.

Humphrey, W.S. (1995). *The capability maturity model: Guidelines for improving the software process.* Software Engineering Institute. Addison Wesley.

Table 2. GCM description

	Certification body	Applicant
GCM-0: *Certification guideline*	Gather the certification model information in preparation for issuing, updating and maintaining certification guidelines.	Study the entire domain or process requirements, and the domain or process certification model.
Input	Certification model	Certification model
Control	Domain or process requirements & standards	Domain or process requirements & standards
Support	International domain standard	International domain standard
Output	Certification guidelines	Certification guidelines
GCM-1: *Certification request*	Process the application	Submit an application
Input	Domain or process certification model, certification guidelines	Application form, quality policy
Control	Nil	Nil
Support	Nil	Nil
Output	Application category	Application no., certification process request
GCM-2: *Prepare for Assessment*	Help the applicant to produce documentation to pass the assessment	Prepare and implement the certification criteria, pre-assess it before assessment.
Input	Certification guidelines, process knowledge, application category	Certification guidelines / memo
Control	Domain or process requirements & standards	Domain or process requirements
Support	Nil	Nil
Output	Assessment material, schedule, certification guidelines and documentation	Assessment schedule, quality policy and documentation
GCM-3: *Perform assessment*	Do the assessment	Cooperate with the assessors during the assessment, provide the necessary information.
Input	Assessment materials and certification guidelines	Company policies, process or product information
Control	Domain or process requirements & standards	Domain or process requirements & standards
Support	Nil	Nil
Output	Assessment result	Assessment information
GCM-4: *Analyse assessment result*	Evaluate the assessment result to see if anything fails the assessment.	Analyse the assessment results and identify improvements.
Input	Assessment result	Assessment results, its process or product
Control	Domain or process requirements	Domain or process requirements
Support	Certification standard	Nil
Output	Evaluation results	Improvement list

Klingler, C.D. (n.d.). A practical approach to process definition. *http://www.asset.com/stars/lm-tds/Papers/ProcessDDPapers.html*

Leung, H.K.N., & Li, V. (2002). A process model for certification of product and process. In S. Valenti (Ed.), *Successful software reengineering* (ch. 21). Hershey, PA: IRM Press.

Paulk, M. et al. (1993). *Capability maturity model for software, Ver. 1.1*. Pittsburgh, PA: Software Engineering Institute.

PMI Web site. (2004). *http://www.pmi.org*

Silva, T. et. al. (2002). The software process: Modelling, evaluation and improvement. *Handbook of software en-*

gineering and knowledge engineering (vol. I). World Scientific Publishing.

KEY TERMS

Assessment: The process of comparing the actual measurements of the characteristics of interest with the specifications of those characteristics.

BOOTSTRAP: BOOTSTRAP is a European method for software process assessment and improvement. It enhanced and refined the Capability Maturity Model (CMM) developed at the Software Engineering Institute (SEI) for software process assessment, and adapted it to the European software industry.

Certification: A procedure by which a third party gives written assurance that a product, process or service conforms to specified characteristics. Certification involves assessment.

CMMI: The Capability Maturity Model Integration (CMMI) models are an integration of best practices from proven process improvement models, including the SW-CMM, EIA731, and the Integrated Product Management CMM. They are tools that help organisations improve their ability to develop and maintain quality products and services.

IDEF0 Notation: Boxes within the diagram depict the sub-activities of the activity named by the diagram. Arrows between boxes depict availability of work products to activities. Arrows entering the left side of a box are inputs to the activity. Arrows exiting the right side of a box are outputs from the activity. Arrows entering the top of a box are controls that regulate the activity, and those entering the bottom are mechanisms that support the activity.

ISO9001: The ISO 9001 is an international standard that specifies the requirements of a quality management system. It is applicable to any organisation regardless of product, service, organisational size, or whether it is a public or private company.

SW-CMM: The Capability Maturity Model for Software (SW-CMM) describes a framework that software organisations can use to determine their ability to develop and maintain software. Its use can lead to organisational improvement. SW-CMM can be applied by acquisition organisations to select software contractor.

Challenges in M–Commerce

Pouwan Lei
University of Sussex, UK

Chris Chatwin
University of Sussex, UK

Rupert Young
University of Sussex, UK

M-COMMERCE

E-commerce activity is growing exponentially, and it is revolutionizing the way that businesses are run. In Asia, there is now an explosion of mobile wireless services. The number of 3G users in Korea has surpassed 1 million while in Japan it has grown from 50,000 in 2001 to 2 million in 2003. Mobile e-commerce (m-commerce) makes business mobility a reality; mobile users can access the Internet at any time, from anywhere and from their shirt pockets/purses using ubiquitous inexpensive computing infrastructure. It is estimated that the m-commerce market will grow to over USD200 billion by 2005 (Abbott, 2002). There are many definitions of m-commerce. One definition consists of all or part of mobile wireless services ranging from mobile phone networks to wireless local area networks. However, the service provided by mobile phone systems has achieved huge success. Mobile phone users originate from all walks of life and include almost all age groups, from teenagers to retired people. It creates a new method of personal communication without location constraints. Instead of briefly describing all mobile wireless services, we will concentrate on the mobile phone and PDA related to mobile telecommunication. Hence, m-commerce is defined as electronic commerce carried out in handheld devices such as the mobile phone and PDA through a mobile telecommunication network.

E-commerce is characterized by e-marketplaces, online auction systems which act as the intermediary between buyers and sellers, whereas, m-commerce is personalized and ideal for access to location-based services. Many new business models have been established around the use of mobile devices. Mobile devices have the characteristics of: portability, low cost, more personalization, global positioning system (GPS), voice, and so forth. The new business models include micro payment and mobile payment, content distribution services and business services. Figure 1 illustrates m-commerce applications. Because of their existing customer base, technical expertise and familiarity with billing, mobile telephone operators are the natural candidates for the provision of mobile and micro payment services. Micro payment involves small purchases such as vending and other items. In other words, the mobile phone is used as an ATM card or debit card. Consumers can pay for purchases at convenience stores or buy train tickets using their mobile phones.

Content distribution services are concerned with real-time information, notification (e.g., bank overdraft), using positioning systems for intelligent distribution of personalized information by location; for example, selective advertising of locally available services and entertainment. Real-time information such as news, traffic reports, stock prices, and weather forecasts can be distributed to mobile phones via the Internet. The information is personalized to the user's interests. By using a positioning system, users can retrieve local information such as restaurants, traffic reports and shopping information. Content distribution services with a greater degree of personalization and localization can be effectively provided through a mobile portal. Localization means to supply information relevant to the current location of the user. Users' profile such as past behavior, situation and location should be taken into account for personalization and localized service provision. Notification can be sent to the mobile device too. Mobile network operators (MNOs) have a number of advantages over the other portal players (Tsalgatidou & Veijalainen, 2000). First, they have an existing customer relationship and can identify the location of the subscriber. Second, they have a billing relationship with the customers while the traditional portal does not. MNOs can act as a trusted third party and play a dominant role in m-commerce applications. In addition, the mobile phone has become a new personal entertainment medium. A wide range of entertainment services are available, which consist of online game playing, ring tones download, watching football video clips, live TV broadcasting, music download and so on. Unsurprisingly, adult mobile services and mobile gambling services are among the fast growing services. According to Juniper research, the total revenue from adult mobile services and mobile gambling services could be worth US$1billion and US$15billion respectively by 2008 (Kwok, 2004). Law regulators have to stay ahead of the fast growing development.

Businesses also need to think across traditional boundaries in e-commerce. Interaction among businesses, consumers, and smart appliances creates a lot of new opportunities: first, appliance-to-appliance, that is, appliances interact with an automatic service scheduler; second, appliance-to-business can deliver smart appliance repair alerts; third, business-to-appliance can deliver remote activation of smart appliances. For instance, the built-in sensors in a car will inform the repair service which part is broken. M-commerce also has a great impact on business applications, especially for companies with remote staff. Extending the existing Enterprise Resource Planning (ERP) systems with mobile functionality will provide remote staff, such as sales personnel, with real-time corporate and management data. Time and location constraints are reduced and the capability of mobile employees is enhanced. The logistic related business also benefits from the use of mobile inventory management applications. One interesting application is "rolling inventory" (Varshney & Vetter, 2002). In this case, multiple trucks carry a large amount of inventory while on the move. Whenever a store needs certain items/goods, a nearby truck can be located and just-in-time delivery of goods can be performed. M-commerce offers tremendous potential for businesses to respond quickly in supply chains.

CHALLENGES IN M-COMMERCE

M-commerce has a number of inherent complexities as it embraces many emerging technologies: mobile wireless systems, mobile handheld devices, software, wireless protocols, and security (Ojanperä & Prasad, 2001). These technologies have rapid product cycles and quick obsolescence. M-commerce, which is more complex than e-commerce, faces a number of challenges see Figure 2. The challenges are:

(i) The delay in 3G mobile network operators (MNO) in implementing their systems infrastructure. The success of m-commerce in Japan changes the concept of "free" Internet to "paid" Internet. Users are willing to pay for the service. MNOs anticipate a huge profit in taking control of the backbone of m-commerce – the wireless infrastructure. In addition, MNOs also play a dominant position in providing m-commerce applications. This has created an unreasonably high expectation from the services. Big companies in Europe, such as Deutsche Telecom, France Télécom, Spain's Telefünica and the UK's Vodafone, spent an estimated USD125 billion to USD150 billion on 3G licenses (Garber, 2002). Many of them are burdened with high debts.

(ii) With the exception of Asia, there is a lack of interest in 3G mobile phone systems. The Western European market has reached saturation point, where mobile possession rate is close to 100% in some countries.

Figure 1. M-commerce applications

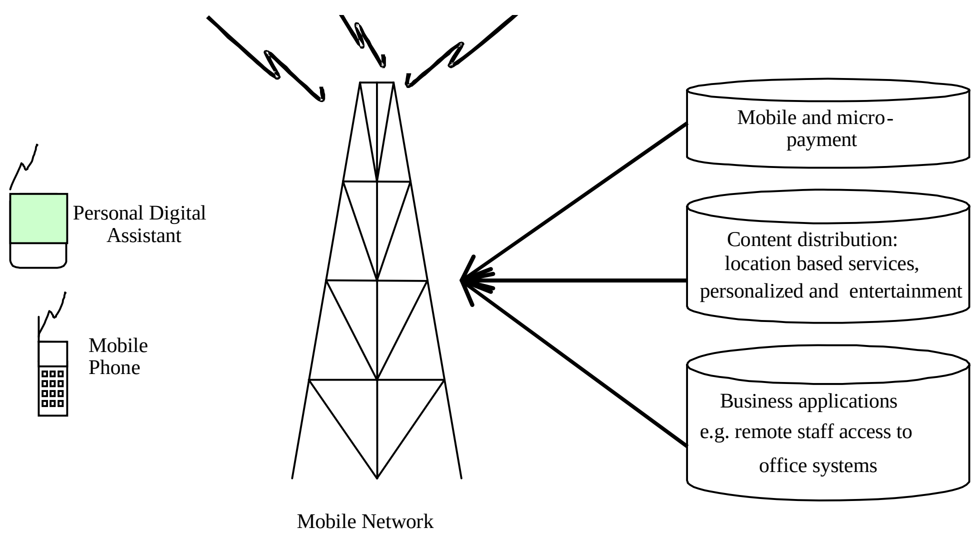
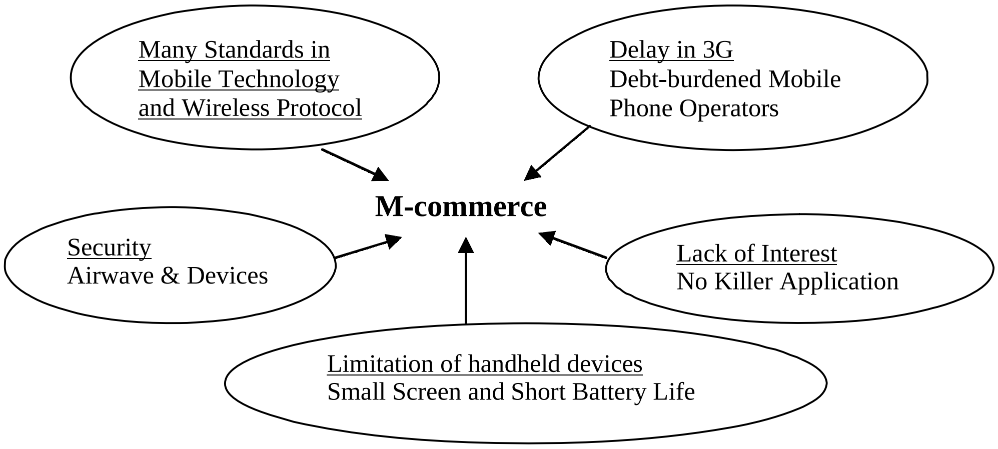


Figure 2. Challenges in m-commerce



In addition, mobile users have "upgrade fatigue", that is, they are reluctant to upgrade their mobile phones. A report from the World Market Research Center, released in May 2002, predicted that 3G sales are likely to be sluggish because of a lack of consumer interest in new services. The research done by the mobile consultancy group Detica also warned that there is no single "3G killer application" that will entice large numbers of people to embrace 3G. In 2002, the mobile phone business pushed very hard on picture messaging, which requires new expensive handsets. The response has been poor. The mobile revenue mainly comes from the voice calls and SMS messaging.

(iii) An immature market with many standards. 3G standards have some commonalities, but they are not fully compatible with each other at the air interface (radio-transmission) level. This produces uncertainty for the success of global roaming and hence limits the creation of mass-market volumes for terminals and infrastructure. In addition, the spectrum problem in the US casts a shadow over global roaming. The Department of Defense occupies the spectrum band needed for 3G services. After September 11th, the U.S. military is unlikely to give up the spectrum. This introduces further technical difficulties. The advances in technology have resulted in a convergence of the functions of mobile phone and PDA devices. The sophisticated mobile phones are commonly equipped with PDA-functionalities such as a large screen and easier methods of input, for example, pen with touch screen. On the other hand, the high-ended PDAs are more compact – thinner and lighter than their predecessors. Furthermore, they are affordable and packed with many features, such as: large bright screen, handwriting recognition software, media player software, good user interface, slots for different types of memory and various wireless technologies – Wi-Fi, Bluetooth and infra red. The market for handheld devices is different from the personal computer (PC) market. For instance, Nokia not only produces phone hardware but is also involved in developing the Symbian software (the operating system of mobile phone) together with other phone manufacturers such as Motorola. Each device comes with its own proprietary operating system. Many standards are competing with each other. However, this situation is changing as Microsoft and Linux are both involved in creating operating systems for handheld devices – that is, mobile phones and PDAs. Microsoft joined the competition by rolling out two operating systems, Pocket PC2002 and the smartphone system for PDAs and mobile phones respectively in 2002. Both systems can be synchronized to a desktop PC, a convenience for business users. Linux, the open-source code, has two popular versions of embedded Linux for handheld devices: MontaVista Linux and uLinux. The participation of Microsoft and Linux in the handheld devices market will have an impact on the dominant position of Symbian.

(iv) Security. Mobile communications offer users many benefits such as portability, flexibility and increased productivity. The most significant difference between wired networks and mobile communication is that the airwave is openly exposed to intruders. The intruder eavesdrops on signaling and data connections associated with other users by using a modified mobile phone. In addition, their small size, relatively low cost, and mobility means that they are likely to be lost or stolen. Sensitive information such as the "private-key" is thus vulnerable (Karygiannis & Owens, 2002). A 3G mobile device, when connected to an IP network, is in the "always-on" mode. Both this "always-on" mode and bluetooth's "on" mode make the device susceptible to attack. Moreover, it also provides the opportunity to track a user's activity, which may be a violation of privacy.

(v) The limitations of handheld devices. Technological developments will increase the computing power and storage in handheld devices. However, insufficient battery life and power consumption will impede the potential growth of m-commerce even when 3G is widely available. At present, the battery life is very short (e.g., two to four hours for surfing) and the small screen is another limitation. PDA's have a larger screen (13cm*8cm) whereas the screen of a mobile phone is 7cm*5cm, which poses difficulty when surfing the Web. A low power, inexpensive, high-resolution colour display would seriously increase the growth of m-commerce.

On the other hand, Wi-Fi, wireless area local networks, which allow users to surf the Internet while moving, are proliferating at astonishing speed on a global scale. Worldwide retail chains like Starbucks and McDonald's offer wireless Internet access to their customers. It offers a fast and stable connection; the data rate is several times faster than 3G. The WiFi is an important, new and disruptive technology to mobile telephone technology, and it may be a watershed for all other m-commerce investments by telecom and content providers in the world of the mobile Internet (Lamont, 2001). In making use of this technology, mobile phone manufacturer (Nokia) and wireless network manufacturer (Cisco) have been working together closely to produce the Wi-Fi phone. A USA mobile telecommunication operator has integrated a network of Wi-Fi hotspots with the existing mobile network systems. In such a way, the tariff of accessing the mobile Internet will be reduced to a budget price. More handheld device users will surf the Internet through their compact mobile phones or PDAs when on their move. At the time of writing this article, WiMax (Worldwide Interoperability for Microwave Access), a wireless broadband connection in wide area network (WAN), is at the stage of discussing standards (Cherry, 2004). When this technology is standardized and ready for deployment, will it pose a big threat or offer an opportunity to mobile phone operators in the future?

CONCLUSION

The mobile Internet is ideal for particular applications and has useful characteristics that offer a range of services and content. The widespread adoption of mobile commerce is fast approaching. In business, the order of the day is "change". Winning customers in today's highly competitive and demanding world is the key to survival. The mobile world is changing the logic of business; businesses have to implement effective strategies to capture and retain increasingly demanding and sophisticated customers. Business needs to think critically about how to integrate the mobile Web to the wired Web, this needs careful strategic thought, as not all applications are appropriate; businesses that take the correct decisions will control the future.

REFERENCES

Abbott, L. (2002). M-commerce. Retrieved on October 22, 2003 from *http://www.mobileinfo.com/Mcommerce / driving_factors.htm*

Cherry, S.M. (2004, March). WiMax and Wi-Fi: Separate and unequal. *IEEE Spectrum*.

Garber, L. (2002). Will 3G really be the next big wireless technology? *Computer, IEEE, 35*(1), 26-32.

Karygiannis, T. & Owens, L. (2002). Draft: Wireless network security: 802.11, Bluetooth™ and the handheld devices. National Institute of Standards and Technology, Technology Administration U.S. Department of Commerce, Special Publication 800-48.

Kwok, B. (2004). Watershed year for mobile phones. *Companies and Finance in South China Morning Post*, January 3.

Lamont, D. (2001). Conquering the wireless world: The age of m-commerce. Capstone Publishing Ltd. (A Wiley Company).

Ojanperä, T. & Prasad, R. (2001). WCDMA: Towards IP mobility and mobile Internet. Artech House Publishers.

Tsalgatidou, A. & Veijalainen, J. (2000). Mobile electronic commerce: Emerging issues. In *Proceeding of EC-WEB 2000, 1st International Conference on E-commerce and Web Technologies*, London, Greenwich, U.K., September 2000, Lecture Notes in Computer Science, 1875, Springer Verlag (pp.477-486).

Varshney, U. & Vetter, R. (2002). Mobile commerce: Framework, applications and networking support. *Mobile Networks and Applications, 7,* 185-198, Kluwer Academic Publishers.

KEY TERMS

2.5 Generation Wireless Service (2.5G): 2.5G changes wireless service to a packet-switched service that dramatically increases transmission speeds. It describes the state of wireless technology and capability usually associated with General Packet Radio Services (GPRS) - that is, between the second and third generations of wireless technology.

Global System for Mobile Communications (GSM): A world standard for digital cellular communications using narrowband time division multiple access (TDMA). It is the standard most commonly used in Europe and Asia, but not in the United States.

i-mode: The packet-based service for mobile phones offered by Japan's leader in wireless technology, NTT DoCoMo. The i-mode protocol uses compact HTML (cHTML) as its markup language instead of WAP's Wireless Markup Language (WML) to provide mobile phone voice service, Internet and e-mail.

Man-in-the-Middle: The intruders set up a base station transceiver with a modified phone and put themselves in between the target user and a genuine network. The intruders have the ability to eavesdrop, modify, delete, re-order, replay, and send spoof signaling.

Short Message Service (SMS): Has grown very rapidly and is very popular in Europe. SMS messages are two-way alphanumeric paging messages up to 160 characters that can be sent to and from mobile phones.

The First Generation (1G): The systems are analog, circuit-based, narrowband and are suitable for voice communication only.

The Second Generation (2G): Based on circuit-switched technology where each call requires its own cell channel, which makes transmission of data quite slow. 2G services include Code Division Multiple Access (CDMA), Time Division Multiple Access (TDMA), and GSM (Global System for Mobile Communication (GSM).

The Third Generation (3G): Will bring wireless transmission speeds up to 2Mbps, which permits high-quality wireless audio and video. It comprises three primary standards: W-CDMA (wide-band code division multiple access), CDMA2000, and TD-CDMA (time division CDMA).

Wireless Application Protocol (WAP): An open, global specification that empowers mobile users with wireless devices to easily access and interact with information and services instantly. The WAP is a standard for providing cellular phones, pagers, and other handheld devices with secure access to e-mail and text-based Web pages.

Wireless Fidelity (Wi-Fi): A popular term for 802.11b, a wireless local area network (WLAN) specified by the Institute of Electrical and Electronic Engineers (IEEE) and is based on the Ethernet protocol and CSMA/CA (carrier sense multiple access with collision avoidance) for path sharing. Wi-Fi supports a range of about 150 feet and data rates up to 11mbps.

Worldwide Interoperability for Microwave Access (WiMax): Called 802.16 in industry standard, a wireless broadband connection in wide area network (WAN). It offers fast wireless data communications over distance up to about 30 miles.

Challenges in Quality of Service for Tomorrow's Networks

Luiz A. DaSilva
Virginia Polytechnic Institute and State University, USA

INTRODUCTION

The original communication networks were designed to carry traffic with homogeneous performance requirements. The telephone network carried real-time voice, with stringent latency bounds, and therefore used circuit-switched technologies with fixed bandwidth allocated to each call. Original data networks were used for electronic mail and file exchange, and therefore employed packet switching and provided best-effort service.

Soon, however, it became clear that economies would accrue from utilizing a common network infrastructure to carry diverse kinds of traffic. Packet switching is now often employed for both real- and non-real-time traffic. This creates a problem: If the same network is to carry flows with diverse performance requirements, it is crucial that it support mechanisms to differentiate among these flows. For instance, real-time applications such as voice and video should be protected when competing for resources with nonreal-time applications such as file transfer or e-mail. Table 1 illustrates some of the quality of service (QoS) requirements of different classes of applications.

The desire to meet the needs of a broad range of applications that coexist in the same network is the primary motivation for the development of QoS mechanisms and architectures.

BACKGROUND

The concept of QoS arises from the need to allocate resources in order to maximize perceived quality on the basis of intrinsic application requirements (such as low latency for real-time voice), pricing (higher service quality for those willing to pay more), or policy (such as preferred access to network resources for users with mission-critical applications). A discussion of how pricing and policies relate to QoS can be found in DaSilva (2000a) and Flegkas, Trimintzios, and Pavlou (2002), respectively.

Either deterministic or probabilistic QoS guarantees may be associated with a given flow. When the network guarantees that the flow will be allocated bandwidth of at least B bits per second at every router on the source-destination path, it is providing a deterministic guarantee. On the other hand, if at least p% of packets belonging to the flow are guaranteed to encounter delay of less than D seconds, the network is providing a probabilistic guarantee.

It is important to note that QoS is not synonymous with performance. In the context of computer networks, the term QoS often implies that some sort of differentiation is made among disparate users. A variety of mechanisms are employed to this end. These include admission control, policing, congestion control, bandwidth reservation, marking, and classification. We briefly discuss each of these mechanisms next.

Table 1. QoS requirements of different classes of applications

Application	QoS Requirements
IP telephony	Low delay (on the order of ~ 100 ms)
Web surfing	Acceptable throughput
Streaming media	Low delay variation (jitter)
Networked virtual environments	Low delay in support of interactivity, high bandwidth in support of high-quality graphics
Online backup	High throughput, low packet losses (i.e., few retransmissions)
E-mail	High tolerance to delay, low to moderate data rate requirements

The decision of whether to accept a new call or flow is referred to as *admission control* (Breslau, Knightly, Shenker, Stoica, & Zhang, 2000); the objective is to ensure that the network can accommodate all of its current traffic flows with the desired QoS even after the new flow is accepted. In circuit-switched networks, incoming calls that cannot be accommodated are blocked; in packet switching, the flow may be denied access or the packets associated with it may be marked as lower priority and dropped when congestion occurs. This leads to *congestion control*; traffic supported by today's integrated networks tends to be bursty, and the admission decision is generally not made based on peak traffic conditions. It is therefore possible, even after admission control, that the network may experience congestion at times. Frost (2003) presents a study of the effects of temporal characteristics of congestion on user-perceived QoS. Measures to alleviate congestion include the dropping of packets at congested routers, as well as implicit or explicit signaling to the source to reduce its transmission rate. These measures may take into account QoS requirements, for instance, by requesting that one source reduce its transmissions while another (more critical) source is allowed to maintain its current rate.

Policing refers to measures taken by the network to ensure that the traffic being offered by a user conforms to a preagreed traffic contract. Excess traffic can be simply dropped at the ingress router or marked for best-effort delivery. Conversely, users may employ *shaping* to ensure their traffic conforms to preestablished parameters such as maximum data rate or maximum burst length. *Bandwidth reservation* may be used in packet-switched networks to provide minimum guarantees as to bandwidth availability to a flow. This requires a signaling phase to precede the transmission of packets, during which each router on the source-destination path agrees to reserve a portion of the available bandwidth to be used by the flow. Queuing and scheduling mechanisms such as weighted fair queuing are implemented by routers in order to meet such guarantees. Unlike in circuit switching, reserved bandwidth that is not being used by the reserving flow is generally made available to other flows through the use of work-conserving scheduling. To support service differentiation, packets are often *marked* using preassigned bit sequences in the packet header; this allows routers in the path to recognize the packet as part of a given flow and *classify* it accordingly (Gupta & McKeown, 2001).

QoS mechanisms can be provided at different layers of the protocol stack as well as by the application and the middleware (DaSilva, 2000b). At the physical and data link layers, prioritization, forward error correction, code, and slot assignment can be adopted for service differentiation. For instance, in random-access local area networks, nodes must back off in case of collision, picking an interval before they are allowed to attempt retransmission; by enforcing different back-off intervals for different nodes, we can achieve prioritization in access to the channel at times of heavy traffic. Scheduling, shaping, admission, and flow control are some of the mechanisms described above that may be adopted at the network layer. Middleware is sometimes developed to take care of classification of flows and marking of packets, and generation of resource-reservation requests. The application itself may employ prefetching and caching of information to improve performance experienced by selected users.

Asynchronous transfer mode (ATM) is one mature example of a packet-switched network providing QoS differentiation. In ATM, this is achieved by defining multiple service categories with associated QoS guarantees and traffic conformance definitions (Giroux & Ganti, 1999). Due to the ubiquity of the Internet protocol (IP) and current interest in real-time traffic using this protocol (voice over IP, for instance), recent research has focused on how to support QoS over the Internet (Armitage, 2000; Firoiu, Le Boudec, Towsley, & Zhang, 2002).

FUTURE TRENDS

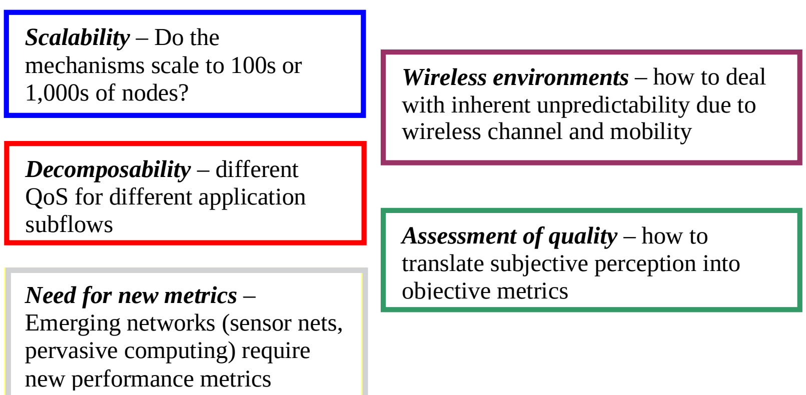
An important challenge in providing QoS differentiation in packet-switched networks has to do with the *scalability* of such mechanisms (Welzl & Muhlhauser, 2003). Stringent QoS guarantees require that all network nodes store information about individual flows in order to make scheduling decisions. While this is reasonable in an intranet, with a limited number of flows and complete control over the network by a single administrative unit, those types of approaches do not scale well. In particular, in the Internet, a core router may be routing packets belonging to many thousands of flows at any one time, and maintaining state information about each flow is infeasible. Stateless approaches achieve better scalability: They classify packets into a finite, reasonably small set of classes, marking each packet accordingly, and associate probabilistic service guarantees with each class. The Internet Engineering Task Force (IETF) has been studying QoS architectures for the Internet for several years. Any major change to the IP suite is, of course, always controversial (one must only look at the time it is taking for the widespread adoption of IPv6 for another example of the resulting inertia). While several important developments resulted from these working groups (Blake, Black, Carlson, Davies, Wang, & Weiss, 1998; Zhang, Deering, Estrin, Shenker, & Zappala, 1993), the ultimate goal of having QoS widely available in the Internet remains elusive.

Providing QoS guarantees in mobile wireless environments is another great challenge. The wireless medium is inherently subject to performance fluctuations due to fading, shadowing, noise, and interference. While techniques such as forward error correction, diversity, and power control reduce such fluctuations, it is still difficult to provide meaningful performance guarantees. Mobility compounds the problem. Still, third-generation (3G) cellular systems and beyond are providing differentiated service, including high-speed data service in addition to traditional telephony. In wireless local area networks (WLANs) such as IEEE 802.11, service differentiation will be achieved through prioritized (but, in most cases, still random) access to the wireless channel (Gu & Zhang, 2003). Physical-layer approaches to support QoS investigated in the literature include waveform selection and power control. These can be combined with slot assignment, prioritization, or other mechanisms for medium access control. QoS in mobile ad hoc networks is the subject of much research; the lack of infrastructure in such networks adds a layer of complexity to the problem (Chakrabarti & Mishra, 2001). In an ad hoc environment, it is difficult to guarantee the existence of a path between source and destination, much less the quality of communications using that path.

When link quality degrades or the network becomes congested, it is important to ensure that the inevitable QoS degradation occur in a graceful manner. We call an application *decomposable* if its data stream can be divided into two or more substreams that the application values differently. For instance, all quantized data is decomposable in the sense that the most significant bits of each sample are of greater "value" (in terms of the fidelity of the reconstructed signal) than the least significant ones. Similarly, albeit at a different level, multimedia applications are also decomposable: Each of the different components of a multimedia application (video, audio, image, text, animation) presents a different sensitivity to available bandwidth, as well as to delay and jitter, and each contributes differently to the perceived quality of the application. Both explicit and implicit adaptation may be employed during times of overload in order to minimize degradation of QoS as perceived by end users. In explicit adaptation, the application itself, upon receipt of feedback information from the network, reduces its transmission rate by transmitting only substreams that it considers most valuable. In implicit adaptation, the application negotiates different traffic contracts for each of its substreams; in times of overload, the network honors QoS assurances while discarding information in substreams for which only best-effort service has been negotiated. An early example of the latter is provided in Petr, DaSilva, and Frost (1989). There has also been some work in providing graceful performance degradation by taking advantage of the different impacts of the loss of intra, predictive, and bidirectional frames in Moving Pictures Experts Group (MPEG). However, taking into account application valuation of transmitted data in order to achieve graceful QoS degradation remains an open area of research.

An important open question is how to value the information that gets transmitted. The success of a QoS-enabled network should ultimately be assessed in terms of end-user perception of quality. Several of the QoS mechanisms deployed today act on a single layer of the protocol stack (say, the network or the data link layer) or on a single link in the path from source to destination. There has only been limited success in understanding how the deployment of differentiated scheduling or prioritization mechanisms at the different layers affects the end-to-end performance experienced by an application. Even more difficult is to translate user perception of quality (for instance, fidelity of video or audio, immersion in a networked virtual environment, satisfaction with an e-commerce experience) into metrics that are readily understood and optimized, such as throughput and av-

Figure 1. Some of the open areas of research in quality of service



erage delay. Subjective testing, one approach to assess user perception, is a lengthy and costly process that does not always generalize well.

In some emerging technologies, such as sensor networks and pervasive computing, traditional metrics (throughput, delay) do a poor job of characterizing performance. We need to define metrics that accurately assess how well the network fulfills its objective. For example, technology transparency is one of the desired features of pervasive computing environments, and performance metrics for the network should reflect how well this objective is achieved.

Figure 1 summarizes some of the current challenges in network quality of service.

CONCLUSION

A decade ago, networks that carried voice and data traffic were called *integrated networks*; today, *converged networks* is the more widely used term, and the diversity of traffic that must be carried increased significantly. Service differentiation is of vital importance if we are to ensure that a real-time video conference can share bandwidth with a massive file transfer (for instance, generated by an online backup job) and have both experience acceptable performance given the resources available in the network.

Although support for QoS in the Internet is taking longer than anticipated, most believe that it will eventually happen. In local networks and technologies at the edge of the internetwork, QoS support has been standardized faster. A recent example is IEEE 802.11e, which adds QoS features to the existing IEEE 802.11a and 802.11b standards for wireless LANs.

Significant challenges remain, including, among others, the scalability of QoS mechanisms for Internet deployment, application-layer QoS assessment, and support for QoS in mobile and ad hoc environments.

REFERENCES

Armitage, G. (2000). *Quality of service in IP networks: Foundations for a multi-service Internet.* MacMillan.

Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., & Weiss, W. (1998). *An architecture for differentiated services.* IETF RFC 2475.

Breslau, L., Knightly, E. W., Shenker, S., Stoica, I., & Zhang, H. (2000). Endpoint admission control: Architectural issues and performance. *ACM SIGCOMM Computer Communication Review, 30*(4), 57-69.

Chakrabarti, S., & Mishra, A. (2001). QoS issues in ad hoc wireless networks. *IEEE Communications Magazine, 39*(2), 142-148.

DaSilva, L. A. (2000a). Pricing for QoS-enabled networks: A survey. *IEEE Communications Surveys and Tutorials, 3*(2), 2-8.

DaSilva, L. A. (2000b). QoS mapping along the protocol stack: Discussion and preliminary results. *Proceedings of the IEEE International Conference on Communications (ICC 2000)*, (Vol. 2, pp. 713-717).

Firoiu, V., Le Boudec, J.-Y., Towsley, D., & Zhang, Z.-L. (2002). Theories and models for Internet quality of service. *Proceedings of the IEEE, 90*(9), 1565-1591.

Flegkas, P., Trimintzios, P., & Pavlou, G. (2002). A policy-based quality of service management system for IP DiffServ networks. *IEEE Network, 16*(2), 50-56.

Frost, V. S. (2003). Quantifying the temporal characteristics of network congestion events for multimedia services. *IEEE Transactions on Multimedia, 5*(3), 458-465.

Giroux, N., & Ganti, S. (1999). *Quality of service in ATM networks: State-of-the-art traffic management.* Prentice Hall.

Gu, D., & Zhang, J. (2003). QoS enhancement in IEEE802.11 wireless local area networks. *IEEE Communications Magazine, 41*(6), 120-124.

Gupta, P., & McKeown, N. (2001). Algorithms for packet classification. *IEEE Network, 15*(2), 24-32.

Petr, D. W., DaSilva, L. A., & Frost, V. S. (1989). Priority discarding of speech in integrated packet networks. *IEEE Journal on Selected Areas in Communications, 7*(5), 644-656.

Welzl, M., & Muhlhauser, M. (2003). Scalability and quality of service: A trade-off? *IEEE Communications Magazine, 41*(6), 32-36.

Zhang, L., Deering, S., Estrin, D., Shenker, S., & Zappala, D. (1993). RSVP: A new resource reservation protocol. *IEEE Network, 7*(9), 8-18.

KEY TERMS

Assured Forwarding (AF): A per-hop behavior defined in DiffServ. AF provides different levels of forwarding assurances depending on available resources, the current number of flows in that AF class, and the drop precedence associated with the IP datagram.

Converged Networks: Networks that carry diverse types of traffic, such as real-time voice and video, Web browsing, traffic generated in grid computing, networked virtual environments, and so forth.

Differentiated Services (DiffServ): Architecture for QoS differentiation in the Internet. It employs marking of IP datagrams to associate packets to predefined per-hop behaviors. The DiffServ architecture is described in the IETF RFC 2475.

Expedited Forwarding (EF): A per-hop behavior defined in DiffServ meant to provide a "virtual leased line" type of service with guarantees on delay and jitter.

Multiprotocol Label Switching (MPLS): Associates "labels" to flows in an IP network. Each router on the end-to-end path makes scheduling and routing decisions for a packet based on the contents of its label.

Per-Hop Behavior (PHB): Describes, within the DiffServ framework, how traffic belonging to a particular behavior aggregate should be treated in each network node.

Policing: Mechanism by which the network determines whether offered traffic meets the specifications in the service agreement; excess traffic may be dropped or marked as noncompliant. Policing is typically implemented at the entry point to a network.

Policy: Set of rules specifying the level of access to network resources to which a user or application is entitled.

Prioritization: Certain packets or data flows may be given higher priority in accessing the medium or in being queued for transmission.

Resource Management: Set of mechanisms to efficiently control and allocate scarce network resources (often bandwidth) to users and traffic flows.

Scalability: The degree to which a protocol or algorithm can be extended to an increasing number of users or network nodes and continue to function properly.

Shaping: Users or applications may shape traffic to comply with prespecified parameters such as maximum data rate or maximum burst length. Leaky bucket and token bucket algorithms are often used for traffic shaping.

Virtual Private Network (VPN): A virtual network constructed for users within a single administrative unit (e.g., a company or university), making use of public communication links. Encryption and bandwidth guarantees may be supported over a VPN.

Work-Conserving Scheduling: A work conserving scheduler allows the link to go idle only when there is no packet, of any service class, awaiting service.

Change Process Drivers for E-Business

H.D. Richards
MAPS, UK and Orion Logic Ltd, UK

C. Makatsoris
Orion Logic Ltd, UK

Y.S. Chang
HanKuk Aviation University, Korea & Orion Logic Ltd, UK

INTRODUCTION

The main focus of this article is a study of the transformation processes occurring in industry and business at large. It deals with the social and economic challenges and it explores the new concepts arising from an unprecedented technology revolution. In addition it sets the scene for a new era of industrial capitalism.

BACKGROUND

Over the last decade of the 20th century a large number of companies faced the future with trepidation while others lacked a good strategy (Ashkenas, 1997; Kidd, 1994; Possl, 1991). Many changes had taken place, including Just In Time (JIT) manufacturing and logistics, lean manufacturing (Womack et al., 1990), shorter product life cycles (Davenport, 1993), more intelligent approaches to IT (Drucker, 1992; MacIntosh, 1994; Nonaka, 1998), and costing (Ansari et al., 1997; Wilson, 1995), but making money was becoming more and more difficult. It was a time and climate for dramatic new approaches (Drucker, 1994; Goldman et al., 1995; Warnecke, 1993) with greater agility. New technologies were replacing old at a faster rate and information technology provided better management and control vision, albeit on a limited local scale (Arguello, 1994; Leachman et al., 1996; Makatsoris et al., 1996). And, push to pull manufacturing (Mertins, 1996) distinctly changed the approach to customers and service. Moreover, increased competitive and economic pressures resulted from the global reach of customers, manufactures and service providers keen to exploit the wealth of opportunities in both global markets and differences in worldwide regional markets (Bitran et al., 2003). Even players only operating in local markets (Bologni et al., 1996; Bonfatti & Monari, 2004; Zabel et al., 2000) could not resist the tide of change. As a result, many companies and economies (Hutton, 1995) were in a state of upheaval and as a consequence some fell by the wayside. This was a climate in which there was an uncertain outcome, and it was into this melting pot that the Internet and the World Wide Web (WWW) were to produce an environment for a much-needed revolutionary change in industrial approach. Later, broadband for landline and also wireless networking provided a much needed speedier access.

Businesses looked to the wider horizons and the dynamics of their supply chains as well as their markets to discover new ways of working with both customers and suppliers to grow and remain viable. The diverse industrial, commercial and operational practices and processes needed to be remolded. These targeted the collaborative aspects of external relationships to the advantage of company performance and the creation of new opportunities. This resulted in more and more use of new forms of communication and available multi-media. In this unsettled environment, once fear of change had been forced into the background, chaos became the domain of creative experimentation (Weiland-Burston, 1992). It is during this period of confusion and anxiety that the process of metamorphosis is taking place.

A surge of new software tool ideas have helped, including Enterprise Resource Planning (ERP) and Supply Chain Management (SCM) (Chang et al., 2001), Customer Relationship Management (CRM) (Greenberg, 2002), electronic commerce (e-commerce), electronic business (e-business) (CEC, 2000) and new forms of enterprise, amongst many others. These have stimulated the reformation of business attitudes to the flow of goods, services, information and knowledge (Bouet & Martha, 2000; Hardwick et al., 1996; Introna, 2001; Johnston, 2001; Richards et al., 1997; Zobel & Filos, 2002).

THE CHALLENGE

ICT (Information and Communication Technologies) tools and systems are important enablers (CEC, 2000) in the change process. They have played and will continue to play a major role in the emergence of new ways of conducting business and improving the economics of business. However, open global standards, protocols and interfaces, interoperable applications and platforms, trusted and sustainable infrastructure and compatibility between business practices have to be achieved before interconnection for broader based business is fully realized (Frick, 2000; Kidd, 2001).

The necessary social changes to business (McCarthy, 1996) are at least as significant as ICT. A web-like organizational network has emerged from the more loosely coupled supply chains. The value network and virtual enterprise permit new forms of communication, participation, leadership, and decision making to develop. In turn these create new economic balances, shared learning and procedures geared to people rather than institutions, in which ICT allows the collapse of space and time (Duttas, 2001; Franke, 2001). There is also resistance to change to be overcome (Deloitte & Touche, 2000; Hunziker & Sieber, 1999).

Three basic aspects to change have emerged, before smarter business is accomplished, to drive the change process and they are developing in parallel to carry business forward to the future:

- Organization: How organization and inter-company relations are developed to ensure greater collaboration; here meant as working jointly together, co-operating and co-ordinating, trusting each other, sharing information and knowledge where appropriate and refining the skills in the organization to cope with the economics, strategic aims, day-to-day operations and service excellence.
- ICT: How tools and systems are created, developed and introduced to ensure open, effective and efficient dynamic engagement between companies, using all the appropriate communication channels

Table 1. Some challenges to organizations

- Lack of awareness by very large sections of the business community:
    - How best to educate and train staff
- Lack of trust for successful e-business.
- Insufficient partners in an e-market.
- Lack of perceived benefit.
- Inequitable benefits:
    - How economics and risk are shared, which also takes into account a judgement of fairness.
- Lack of good business models.
- Limitation of collaboration to tier n suppliers.
- Needs to accelerate the business performance targets: for example--
    - Reduction of time to market;
    - Better precision for just-in-time production and logistics;
    - Provision of faster innovation cycles;
    - Working capital efficiency;
    - Increased resource utilization across the whole network;
    - Improvements to distributed inventory management;
    - Creation of new metrics for value network performance improvement.
- Demand for specialty products in small batches through intelligent automated production planning and logistics systems.
- Demand for astute fast mobile complex service engineering.
- Inter-company collaborative design of high-quality complex products.
- Lack of a standard business model for collaborative business.
- Lack of ability to manage and control large equipment/product integration produced in value networks to meet the necessary performance targets.
- Needs to meet the demand of mass customization and personalization through detailed models and processes and smart, agile order processing.
- Ability of companies to conceptualize a value network that is advantageous to them and to identify requirements for new skills, new metrics and changes to be made.
- Needs to have the ability to use shared knowledge and interpret shared information effectively among partners in a value network.
- How much transparency there should be between partners in a value network.
- Belonging to many value networks at the same time.
- Lack of standard business processes.

Table 2. Some challenges for ICT

• Lack of a cost-effective and affordable, real-time worldwide communication of complex business information to serve the globally distributed nature of business.
• Lack of interoperability between systems and applications.
• Lack of multi-lingual facilities.
• Lack of an affordable end-to-end high quality seamless integration.
• Lack of integrated workflow.
• Free flowing information flow in an end-to-end, timely and secure manner.
• Right economics for ICT suppliers.
• Lack of smart applications and good decision assistance tools for collaborative business.
• Need for data capture and filter systems that communicate in real time pertinent information to all partners in the value network.
• Knowledge sharing and usage across the value network.
• Cognitive decision making.
• Tracking individual ordered products from source to consumer and providing a succinct record of such. Human health in the human food chain is a critical issue.
• Development of intelligent agents as electronic proxies for individuals.
• Serve the special needs of SMEs.
• How to respond against unplanned/unexpected event with less cost.

Table 3. Some challenges for the business environment

• Too many standards developers and lack of co-ordination with sources of new work.
• Slowness of standards take-up by commercial ICT developers.
• Legal challenges:
• Contract law;
• Cross-border provision of services;
• Protection of intellectual property rights;
• Privacy;
• Consumer protection;
• Electronic payments.
• Regulation.
• Security:
• Trust;
• Cyber crime;
• Fraud;
• Unauthorized access;
• Computer attack.

open to them as necessary and of preferred choice. This applies both in business networks and supply chains and also in the marketplace, where new opportunities may be found. Such opportunities include competencies, products and services that may be useful to help create, enlarge or reduce value networks. Tools service amongst others co-operative planning, distributed event driven decision making and tracking.

- Environment: How exclusivity may be stripped away to provide global trade opportunities for all in the world of electronic trade not only buying but also selling to any region or nation irrespective of differences in language, culture or local laws or regulation.

Some of these challenges are outlined in Tables 1 to 3.

FUTURE TRENDS

Consolidation of ideas and rethinking of strategy has been helped not only by a political will (CEC, 2002; Timmers, 2002) and further research and development (CEC, 2003), but also by the economic down-turn[1] as well as a greater respect for the economics of change amongst all the industrial players.

ICT providers and consultants to date have provided much of the impetus for industrial change but industry has now to fully engage with strategic plans. One of the challenges highlighted in European statistics is that less than 20% of SMEs[2] had put in place any CRM or SCM software. Change through national incentives, joint projects and education through regional clusters of companies in the same industry sector is being encouraged.

Europe, one of the world's largest trading blocks, has a vision in line with the Lisbon objectives and eEurope for a Single European Electronic Market, SEEM (SEEM, 2002; SEEM, 2003), in contrast to today's separate and unconnected e-markets that serve different industrial or value network goals. It will solve many of the challenges illustrated previously and a research initiative for 2003 has been carefully orchestrated and heralded (CEC, 2003). Current CEC research has provided many strategies for future manufacturing (Geyer et al., 2003), as well as delivering results for biometrics (IBG, 2003) of particular relevance to future authentication, and specific industrial supply chain innovations. Other publications suggest that in the near future every single object will be connected to the Internet through a unique wireless address identifier allowing complete life cycle tracking (Datta, 2004; Murray, 2003; Sarma et al., 2000). A roadmap for intelligent agent research is also available (SEEM, 2002).

Studies and analysis of organizational webs (Tatnall & Gilding, 1999), value (Zobel & Filos, 2002) and inter-node relationships (Underwood, 2002) and new methodologies for design of networks (Soares et al., 2003) amongst many others have helped to develop the social processes in complex organizations and their symbiosis with new business systems. And the semantic Web may help to foster greater understanding between organizations (Berners- Lee, 2001).

In Europe the ICT Standards Board is set up to coordinate activities and sees that there is no duplication in the standards activities of European Standards Organizations (ESOs) and has provided input to standards (CENELEC, 2002). UN/CEFACT, OASIS, Open Source, and RosettaNet are working toward interoperability and effective processes for global e-business. Moreover, key industrial players through the organization, Open Mobile Alliance, OMA (SEEM 2002), are using interoperability and open standards as catalysts to ensure that mobile services develop more openly.

The European Union is acting to clarify its legislation and to provide new regulatory means for SEEM. A lot has already been dealt with and most of the issues should be cleared in the years up to 2005 (CSIRT, 2002; SEEM, 2002).

CONCLUSION

Metamorphosis will not be completed overnight and it will take many years before the process is finished. It is expected that a new end state for global business is characterized by: an expanding global economy; greater collaboration in smarter value networks; versatile virtual organizations; better deployment of knowledge; greater dynamism; versatility and unhindered opportunities in markets; highly dynamic processes; new standards of customer service excellence; better use of both capital and human resource; and better decision making. This will be brought about through better designed and correct deployments of ICT for structured, trusted and secure electronic exchanges between companies. Technological advances will continue unabated to feed the desires for improvement. Also, applications and tools will continue to improve with aids for helping organization building, revising or creating new inter-business processes, as well as providing aids for flexible sizing for smart collaborative value network configurations. Universal standards will make integration between applications and distributed databases and knowledge bases both easier and cheaper. And new telecommunications technology will improve both the speed and the volume of information flow per unit time from anywhere at any time. The workforce and end customers will experience a new way of living and fulfillment. The basic novelty of our age is the spirituality of consumerism and voracious appetite for information. How these may be blended with an increasing capacity for sympathy and mutual understanding will inspire both ways of life and collaborative business.

REFERENCES

Ansari, S.L., Bell, J.E., & the CAM-I Target Cost Core Group. (1997). *Target costing: The next frontier in strategic cost management*. New York: McGraw Hill.

Arguello, M., (1994). *Review of scheduling software, technology transfer 93091822A-XFER*. Austin, TX: Sematech.

Ashkenas, R. (1995). Capability: Strategic tool for a competitive edge. *Journal of Business Strategy, 16*(6), 13-14.

Bitran, G., Bassetti, P.F., & Romano, G.M. (2003). Supply chains and value networks: The factors driving change and their implications to competition in the industrial sector, *MIT Center for eBusiness Research Brief*, *II*(3). http://ebusiness.mit.edu

C

Bologni, L., Gozzi, C., & Toschi, E. (1996). *Selecting software for small companies: The SPI 072/2 experience*.

Bonfatti, F., & Monari, P.D. (2004). Special needs of SMEs and micro-businesses. In Y. Chang, C. Makatsoris, & H.D. Richards (Eds.), *Evolution of supply chain management*, Kluwer Academic Publisher.

Bouet, D., & Martha, J., (2000). *Value Nets: Breaking the supply change to unlock hidden profits*. John Wiley & Sons.

CEC. (2000, January). Directorate General Information Society, DGIS-C3-Electronic Commerce., *Developing a coherent policy and regulatory framework for advancing electronic commerce in Europe*.

CEC. (2002, May). *Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of Regions, eEurope 2005: An information society for all*. Brussels. 263 final.

CEC. (2003). Sixth framework program. *www.cordis.lu*

Chang, Y., McFarlane, D., & Shaw, A. (2001, August). *State-of-the-Art art system review, CUED/E-MANUF/TR.17* . Technical Report, University of Cambridge.

CSIRT. (2002). *Handbook of legislative procedures of computer and network misuse in EU countries for assisting computer security incident response teams*. European Commission, DG Information society Society C4.

Datta, S. (2004). Adaptive value networks: Emerging tools and technology as catalysts; value networks and ICT symbiosis. In Y. Chang, C. Makatsoris, & H.D. Richards (Eds.), *Evolution of supply chain management*. Kluwer Academic Publisher.

Davenport, T.H. (1993). *Process innovation: Re-engineering work through information technology*. Harvard Business School Press.

Deloitte & Touche. (2000). *Manufacturing with a small e: An account of ebusiness in UK and US manufacturers*. Manufacturing Group Report.

Drucker, P. (1992). *Managing for the future*. Oxford: Butterworth-Heinemann, Oxford.

Drucker, P. (1994). The theory of business. *Harvard Business Review, 72*, 95-104.

Frick, V., & Lill, A. (2000, August 4). *Ten imperatives for e-business success.*, Gartner Group report.

Geyer, A., Scapolo, F., Boden, M., Dory, T., & Ducatel, K. (2003). *The future of manufacturing in Europe 2015 to 2020: The challenge of sustainability*. Scenario Report. Joint Research Center European Commission.

Goldman, S.L., Nagel, R.N., & Preiss, K. (1995). *Agile competitors and virtual organizations – strategies for enriching the customer*. Van Nostrand Reinhold

Goldman, S.L., Nagel, R.N., & Preiss, K. (1995). *Agile competitors and virtual organizations – strategies for enriching the customer*. Van Nostrand Reinhold.

Greenberg, P. (2002). *CRM at the speed of light: Capturing and keeping customers in Internet real time* (2nd ed.). McGraw-Hill Osborne.

Hardwick, M., Spooner, D.L., Rando, T., & Morris, K.C. (1996). Sharing information in virtual enterprises., *Communications of the ACM, 39*(2), 46-54.

Hunziker, D., & Sieber P. (1999). Turbulence and the dynamics of Internet diffusion. *Special Issue of Electronic Journal of Organizational Virtualness, 1*(1), 237-261.

Hutton, W. (1995). *The state we're in*. Jonathan Cape.

Introna, L. (2001). Defining virtual organizations. In S. Barnes, & B. Hart (Eds.), *E commerce and v business: Business models for global success*. Butterworth-Heinman.

Kidd, P.T. (1994). *Agile manufacturing: Forging new frontiers*. Addison-Wesley.

Kidd, P.T. (2001). *e Business strategy: Case studies, benefits and implementations*. ISBN 1-901864-02-3.

Leachman, R.C., Benson R.F., Lui, C., & Raar, D.J. (1996). IMPReSS – An automated production-planning and delivery-quotation system at Harris Corporation- Semiconductor Sector. *Interfaces, 26*(1), 6-37.

Macintosh, A. (1994). Corporate knowledge management state of the art review. *Proceedings of ISMICK (Management of Industrial and Corporate Knowledge) Conference*, AIAI Edinburgh University AIAI-TR-151.

Makatsoris, C., Leach, N.P., & Richards, H.D. (1996). Addressing the planning and control gaps in semiconductor virtual enterprises. In J. Browne, R. Haendler Mas, & O. Hlodversson, (Eds.), *IT and manufacturing partnerships: Delivering the promise*. IOS Press.

McCarthy, E. (1996). Culture, mind and technology: Making the difference. In K.S. Gill (Ed.), *Human machine symbiosis*. London: Springer-Verlag.

Mertins, K. (1996). PULL-oriented synchronization of logistics and production flow in automobile industries. In J. Browne, R. Haendler Mas, & O. Hlodversson, (Eds.), *IT* *and manufacturing partnerships - delivering the promise.* IOS Press.

Murray, C.J. (2003, September 2003,). Network specifications released for every day products., *E.E.Times UK.*

Nonaka, I. (1998). The concept of "Baba". Building a foundation for knowledge creation. *California Management Review –Special Issue on Knowledge and the Firm, 40*(3).

Possl, G. W. (1991). *Managing in the new world of manufacturing.* Prentice-Hall Inc.

Richards, H.D., Dudenhausen, H.M., Makatsoris C., & de Ridder, L. (1997, August). Flow of orders through a virtual enterprise – their proactive planning, scheduling and reactive control. *IEE Computing & Control Engineering Journal*, 173-179.

Sarma, S., Brock, D.L., & Ashton, K. (2000). *The networked physical world – proposals for engineering the next generation of computing, commerce & automatic identification.* Technical Report MIT-AUTOID-WH-001, MIT Auto-ID Center.

SEEM. (2002, 2003). Workshop Reports Oct 1st 2002 & March 11th 2003 CEC Information Society. *Europa.eu.net/information_society/topics/ebusiness/ecommerce/seem/index_en.htm*

Soares, A., Sousa, J., & Barbedo, F. (2003). Modeling the structure of collaborative networks: Some contributions. In L. M. Camarinha-Matos, & H. Afsarmanesh (Eds.), *Processes and foundations for virtual organizations.* Kluwer Academic Publisher.

Tatnall, A., & Gilding,. A. (1999). Actor-network theory and information systems research. *Proceedings of the 10th Australasian Conference on Information Systems,* (pp. 955-966).

Underwood, J. (2002). Not another methodology: What ANT tells us about systems development. *Proceedings of the 6th Intl. Conf. on Information Systems Methodologies,* British Computer Society (*http://www-staff.mcs.uts.edu.au/~jim/papers/ismeth.htm*).

Warnecke, H.J. (1993). *The fractal company –a revolution in corporate culture.* Berlin: Springer-Verlag.

Wilson, R.M.S. (1995). Strategic management accounting. In D. Ashton, T. S. Hooper, & R. Scapens (Eds.), *Issues in management accounting* (pp. 159-190). Prentice Hall.

Womack, J.P., Jones, D.T., & Roos, D. (1990). *The machine that changed the world.* Rawson Associates.

Zabel, O., Weber, F., & Steinlechner, V. (2000). *Process reengineering and eBusiness models for efficient bidding and procurement in the tile supply chain.* ECPPM Lisbon, Portugal. *http://www.ebip.net*

Zobel, R., & Filos, E. (2002). Work and business in the eEconomy. In M. Stanford-Smith, E. Chiozzae, & M. Edin, (Eds.), *Technology and policy issues in challenges and achievements in Ebusiness and Ework.* Berlin: IOS Press.

KEY TERMS

Broker: A broker is a manager of a virtual enterprise. A broker acts as an intermediary with the buyer of goods or services. A broker may also manage and control the ICT and provide educational services to small and medium enterprises, SMEs, or micro-companies.

Electronic Business, E-Business: Electronic business is any form of business or administrative transaction or information exchange that is executed using information and communications technology, ICT. This may be transaction performed in a peer-to-peer fashion between companies or organizations or with a customer. Electronic business impacts on the way business is perceived.

Electronic Market, E-Market: An electronic market is a market free from inhibiting constraints and affordable for all businesses in any shape, form or size and to allow them to easily take part in e-business with beneficial returns. It is a market in which trust, security and dependability apply and in which regulatory and legal issues are unified. It is a market where buyers and sellers ubiquitously execute business transactions online. These may include searching and identifying competence, ability to identify the right product or service together with quality, price and quantity, and virtual auctions. It is also based on an open, secure and reliable collaborative platform for knowledge exchange, joint product design, production planning and logistics in stable customer-supplier relationships.

Hub Company: A hub company is a lead company in a supply chain or value network that organizes and manages the work for end products and/or services. This is, for example, an automotive manufacturer, an aircraft manufacturer, a clothing manufacturer, or major food supplier. These are usually large companies but can be small if the end product is in building construction, for instance. Usually the manufacturing and services chain is complex and contains many sub-products that have to be drawn together with processes and services to meet design to manufacture targets and delivery excellence. In a value network the tier *N* suppliers are part of the collaborative network. It is to be noted that communication in the old economy was EDI and in the new economy will be fully open system interoperable ICT.

Intelligent Software Agent: An intelligent software agent acts at speed over the electronic communication channel on behalf of human individuals or companies as their proxy; it is a program acting on behalf of another person, entity, or process. An intelligent software agent is an autonomous program that is capable of perceiving and interpreting data sensed from its environment, reflecting events in its environment and taking actions to achieve given goals without permanent guidance from its user. Agents have to have the intrinsic ability to communicate, cooperate, coordinate, negotiate and ability to learn, as well as have the capability to evolve through their interactions with other agents. Agents can be standalone or be part of a multi-agent system.

Metamorphosis: In application to economies of companies it is a process of change from an existing but obsolete way of doing business to a new form but non-predestined state.

Smart Organization: A smart organization is a further evolution of value networks and virtual corporations through use of more advanced business models taking account of human ICT symbiosis and utilizing more intelligent applications and tools for collaborative work and holistic development of both product and service engineering.

Supply Chain: In its basic form is a buyer-centric chain or network of independent companies that are loosely inter-linked by activity along a manufacturing, servicing and distribution channel of a product service specialty, from sources of raw material to delivery to an end customer. Supplementary to this supply chain management is a set of approaches utilized to integrate suppliers, manufacturers, warehouses and retail stores and so on. Consequently merchandise is produced and distributed in right quantities, and to right locations at the right time, to minimize system-wide costs while at the same time satisfying service level requirements.

Value Network: This term is ambiguous, as the analytical perspective colors its meaning. Nevertheless, the value network in general terms evolves from a supply chain through mutual use of ICT and more closely linked collaboration and mutual dependency between the partner organizations or independent companies. Collaboration means electronic communication via extranet, or Internet, co-operation and co-ordination of work flow, information and knowledge exchange, negotiation and dynamic trading and joint decision-making. Value is derived through the exchanges with partner organizations in the network and its shared knowledge. The value network also aims to deliver the highest value to the end consumer and to its stakeholders

Virtual Enterprise/Virtual Corporation: A virtual corporation or enterprise is formed from a pool of competencies and capabilities resulting from a club of pre-qualified partners that may be expanded or contracted through the mutual desires of the club. The management body for a virtual enterprise selects partners, from the pool of competence available, to provide products or comprehensive services to any industry in direct competition to single companies or other virtual enterprises. It is necessary to have strong collaborative relationships between partners in the club. The virtual enterprise may exist only on a temporary basis to take market chances, for example tendering. It may also exist for a longer term for optimization of a value network to service a market need.

ENDNOTES

1 Economic down-turn in the global economy in the first years of the 21st century.

2 SMEs together with micro-companies account for 70% of employment in Europe.

Changing Trends in the Preparation of Print Media

Mark Snyder
Clemson University, USA

INTRODUCTION

Traditionally, the main purpose of printing is to generate reproductions in quantity. Before printing existed, scribes reproduced manuscripts by hand. Frank Romano, in the foreword of *Pocket Guide to Digital Prepress*, tells the story of a monk who, shortly after Gutenberg's development of cast-metal movable type sparked the advent of printing, authored an essay titled "In Praise of Scribes." The essay advocated the continuation of copying manuscripts by hand because of the character-building values it instilled in the scribes. The ironic part of the story is that the monk decided to have his manuscript printed. The moral that Romano teaches us is that the monk "was caught in the paradox of transitioning technologies" (1996, p. iv) and that a similar situation is taking place as digital technology revolutionizes the printing industry. The objective of this section is to provide some historical perspective and background of printing technology, particularly in the area of preparing media for printing (prepress), and then identify current issues in the field as well as future trends that are developing.

BACKGROUND

Movable type existed before the time of Gutenberg. However, the concept of mass-reproduction of graphic images was not fully realized until Gutenberg created metal type and adapted a wooden screw-type olive press to invent printing around 1450 A.D.

Typesetting remained the primary form of "prepress" work for about the next 500 years. The industrial revolution brought several technological advancements to the printing industry and, in the middle of the twentieth century, phototypesetting became the norm for generating type. Artwork was assembled into paste-ups to be "shot" on a camera, and many people were involved in the process of getting a job ready for press. Designers, photographers, copywriters, proofreaders, editors, typographers, paste-up artists, camera operators, color separators, strippers, and platemakers all were occupations supported by this process.

The "imagesetter," a device developed around 1980, represents a transitional technology that uses digital information and a laser to create an analog product—film. At the same time, page-layout software was developed that allowed the merging of text, line art, and digital photographic images into a layout. Output the page layout from a computer to an imagesetter, add the ability to electronically separate colors, and we have process-color films that are already positioned, thus eliminating the need for several workers.

More recently, platesetters and other "direct-to-plate" technologies have eliminated film from the process. Now, even offset lithographic presses are available with platemaking systems right on the press. A file can be sent from the computer directly to the press. The invention of the silicon chip, and associated digital technologies, has virtually eliminated the majority of the occupations listed earlier.

In fact, "desktop publishing" makes it possible for one person to be all of the above and the printer too. Many do-it-your-selfers proceed to create printed matter, regardless of the fact that they have little or no experience with typography or formal layout and design principles. The result is typically less than pleasing to the graphics professional's eye.

Today, the majority of professionally printed jobs are created using computers. Often they include a mix of text, line art, and photographic images. All of these types of "copy" are typically created using different software applications that have evolved to become very powerful each in their own way. Once all of this various copy is generated, it can be merged using an electronic page layout application. This is the heart of professional digital prepress.

Digital technology, along with highly skilled personnel, has made it easy to do tasks that were once challenging. The rapidly increasing processing power of computers, coupled with ingenious software tools and useful peripheral devices, has enabled the printing industry to make improvements in speed and quality. But it has also increased the likelihood of mistakes and poor design. The technology continues to advance, but so do the problems that need to be solved.

ISSUES, CONTROVERSIES, AND PROBLEMS

Because the technology is readily accessible to non-professional designers, a lot of unprofessional design work is being submitted to printing companies. While many of those who submit a job to a printer will claim that it is "ready to go," unfortunately this is usually not the case. A preflight technician reviews data sent to a printer to make sure that all artwork, fonts, file formats, and so forth are included and will work with the systems at that printing firm. Other concerns that a preflight technician might delve into would include checking whether a job meets size limitations, whether trapping issues are resolved, imposition, and if finishing and bindery requirements are met.

Though most professional graphic designers have enough experience that they are familiar with the common issues related to preparing art and text for printing, sometimes they actually go beyond the limits of what can be achieved. While graphic designers and printers don't always see things the same way, they do know that they need each other—and their ability to communicate will have to continually improve as digital prepress technology advances. Since the traditional occupations held by craftsperson's such as typesetters, paste-up artists, color separators, and strippers are virtually obsolete, "designers are taking on more and more of the prepress process themselves" (Agfa, 1994, p. 1).

Most jobs, even jobs submitted by professional designers, rarely make it through a preflight check successfully the first time. Bob Atkinson, an electronic prepress consultant, claims that "only about 15% of clientele-supplied files are actually ready to output without problems. The other 85% have one or more problems—some are minor and hardly noticeable in the final piece; others have more serious problems that will jam an imagesetter or other output device or, worse, produce an unacceptable print job that the client won't pay for" (2001, p. 1).

The most common mistakes in jobs that are submitted to printers are problems that relate to fonts. And, like fonts, any graphics files that are placed into a page layout file must be included with a job. Nested files can also create a lot of problems when trying to output a job to films or plates.

File formats are another puzzle needing to be figured out. In its educational brochure titled *An Introduction to Digital Color Prepress*, Agfa Corporation stated, "bitmapped images may be saved improperly and print poorly when placed into page-layout…applications." The company continued: "to overcome these problems, a variety of graphics file formats have been devised which link text, graphics, and separation capabilities among different programs and computer platforms" (1997, p. 24). Understanding which file formats are appropriate for the planned end-use is important. Photos scanned on desktop scanners have proven to be a problem area for printers. Not many can answer the question "at what resolution should I scan this picture?"—that is because it depends on a few different variables. If we plan to use the images in printed media, we need to know what resolution we are capturing with that digital camera.

The key to the resolution issue is understanding halftones and the line screens that are being used for printing halftone images. A halftone converts a regular photographic image (referred to as continuous tone) into dots of various sizes. When printed on white paper, this provides the illusion of tones (grays or various colors). The line screen specifies the number and frequency of these dots. Even if a client submits a file at the right resolution, they may set screen frequency, angle, and dot shapes incorrectly which will cause problems with halftone images.

Trapping is actually an old concept that is just handled differently with digital technology. The general idea is that we want two adjacent colors to overlap just slightly when printing so it does not leave a white space if it is printed slightly out of register.

Color management is perhaps the most vexing and perplexing problem facing design and printing professionals today. The issue of color begins with the original art or image that is captured and continues all the way through the process of creating the printed piece. Ingram, a professor of graphic communications, stated, "a color management system is a set of tools that permits consistent and predictable color reproduction. This includes all equipment in the production workflow having an impact on the color files as they move from one device to another" (2001, p. 26). The people involved in this process must have an understanding of color theory and how to make different devices generate color the same way. In many cases, designers and digital prepress departments do not use color management, and simply rely on experience and/or guesswork to approximate color matching.

One of the most significant, and yet simplest problems with jobs today are typographic errors that seem to slip by and get printed on the final product. Traditionally, jobs went through the hands of many people in the process of getting printed. Now, the chain of human workers that touch a job has become so short that many minor errors go unnoticed. Direct-to-press technology now makes it possible for jobs to go right from a designer's computer to a press. While this is more efficient than in the past, it also enables us to make mistakes faster than before!

SOLUTIONS AND RECOMMENDATIONS

Starting from the design stage, perhaps the most important concept for a designer to remember is that printing is primarily all about the reproduction of large quantities of media. Designing something that requires handwork makes mass reproduction difficult. In most cases, this is the type of issue that causes a design concept never to be realized.

If we design with the fewest number of fonts possible—without compromising the design itself—and use more common fonts, we will save a lot of time. Otherwise, designers need to be organized, and include all fonts and linked graphics files into a folder that contains everything for that job. Preflighting will likely turn up some problems, and it is the designer and/or client's responsibility to either provide missing resources, correct errors, or approve charges for correction by the printer.

Capturing images with a digital camera or scanning them at the proper resolution is essential to generating quality bitmapped images. Following the Nyquist Theorem for digital sampling of analog information, we typically try to sample at 1.5 to 2 times the screen frequency, or lines per inch, at which we intend to print the image. This should be multiplied by the scaling factor for the job**.** Using images captured at lower resolutions will impair the quality of the reproduction, and image files with higher resolutions than necessary will be too big and require excessive processing time.

Printing companies are beginning to see value in offering training to their customers and designers that send them problem files. Many are also providing Web sites that explain how to capture images at the proper resolution, upload files, use preferred file formats, prepare graphics files, and so forth. These specifications are typically customized to meet the requirements of the equipment owned by the printing company.

Outputting files with stochastic screening is an alternative to conventional halftones. This technique eliminates the possibility of moiré, an optical phenomenon that occurs when screens overlap at close angles, and provides excellent reproduction of details in a printed image. Stochastic screening is available for postscript output devices. "The film separations can be proofed using conventional methods and plates can be printed on standard Web and sheetfed offset presses" (Agfa, 1997, p. 17).

Digital data transmission has improved the way that customers transport jobs to printers. In the past few years, companies that use the Internet to provide private network services for the printing industry have emerged. File transfer services are big time savers, but they still don't eliminate the fact that all needed files, fonts, and links must be included in a job folder.

Digital workflows for the printing industry have improved during the past five years. Sean Riley stated in *Package Printing* magazine: "One often overlooked stage that printers have begun to explore in an effort to improve workflow, and therefore costs, is the prepress facet of the industry. Improvements in prepress workflow and/or data transmission have been available for quite a while now but [printers] have been slow to take advantage of these opportunities" (2002, p. 14).

Proprietary systems that automate the digital prepress workflow are offered as packages by vendors who also manufacture equipment such as platesetters. Trapping, color management, and raster image processing (RIPing) can all be quickly and easily accomplished via seamless solutions on powerful workstations.

Open Prepress Interface (OPI) is another digital workflow solution that improves prepress performance by minimizing the amount of data handled while generating page layouts. OPI servers employ an extension to Postscript "that automatically replaces low-resolution placeholder images with high-resolution images" (International Paper, 2000, p. 224). Basically, the OPI server reduces the amount of information traveling over a local area network until needed for output.

Color management is another process that has improved digital prepress. "At this point it is apparent that moving to a color managed workflow is not a simple process handled by a single individual. A color managed workflow includes everyone in the color reproduction food chain" (Ingram, 2002, p. 29). It is a scientific approach to what is thought of as the domain of an artist—color. It does not achieve perfect results, but it is amazing to see how much of an improvement there is over a non-color managed environment.

Proofing is another area that benefits from color management. It is rarely advisable to use soft proofs (viewing from a monitor) for contractual color approval. It is better to choose a proofing system that utilizes media similar to the finished product and is known to match output from press systems through a color-managed workflow. Printers favor proofing systems that produce colors using a subtractive colorant system that produces some type of dots to approximate halftone systems.

FUTURE TRENDS

One trend in the printing industry is a shorter route from design to press. Digital prepress peripheral systems, such as scanners and imagesetters, are being bypassed by digital cameras, direct-to-plate, and direct-to-press systems. As these systems become more automated, and direct-to-press technology evolves, it seems possible that the printing press will soon be under control of the prepress department of printing companies and viewed as just another output device.

Digital printing is a new category in the hierarchy of printing processes. One way to identify digital printing processes "is to refer to any printing device that inputs a digital data stream and outputs printed pages as a *digital printer*, and then further distinguish between printers that first build a physical master from which multiple copies are printed and those that do not" (Cost, 1997, p. 79).

Digital printing devices fall into one of the following five categories: photographic, thermal, ink jet, dye sublimation, and electrostatic. The advantage of digital printing is that it is used for reproducing smaller numbers of reproductions more cost effectively. "Virtually every market analyst places flexographic printing growth second only to that of digital printing" (Hogenson, 2002, p. 1). The development of high-quality color digital printing devices has made digital printing the fastest growing printing segment today.

Digital printing brings the realization of a completely digital workflow.

Another technology that is simplifying digital workflows is Portable Document Format (PDF). Huff and West report that "the goal of a PDF workflow system is to deliver electronic files to a film or plate recorder, or to a digital press or printer, with a minimum amount of operator intervention" (2000, p. 42). PDF files are independent of authoring and system software, so they can be opened on any platform. "PDF files serve as a digital master throughout the entire continuum of communication—from posting on the Web to outputting on a press" (Agfa, 1998, p. 1).

Printers especially like what PDF has to offer them. "For print-on-demand service providers, PDF erases many of the current variables introduced by the variety of platforms, application programs, and printers involved in any corporate system" (Young, 1998, p. 57). Perhaps the biggest advantage of using PDF for prepress is embedded fonts and graphics; prepress operators will no longer need to spend time spent searching for missing image files (Snyder, 1999, p. 35). If your document can be "distilled" into a PDF file, it is likely to be output with no problems.

Variable data printing is another trend that is taking advantage of the speed of digital processing. This type of printing produces products that can be individually customized. Most variable data printing is fast, low-quality ink-jet marking, but higher-quality systems that can print bitmap images do exist—although they are slower and targeted for short-run markets. Variable address data can be simply added on to preprinted forms and then shipped. Customized labels are another popular example of this technology. In variable data printing, database management becomes part of the responsibilities of digital prepress personnel. As digital printing technology advances, variable data processing will become faster and more commonplace.

CONCLUSION

In 1978, F.W. Lancaster, a noted library futurist, authored a book titled *Toward Paperless Information Systems* in which he raised, and answered, the question, "Will paper be as important in the information systems of the year 2000 as it is today? Almost certainly not" (p. 1). This has proven to be true in most library systems—paper is not as important as it used to be—but it is still there. Further into his book Lancaster generalizes his vision to ponder a broader perspective. He stated, "The paperless society is rapidly approaching, whether we like it or not" (p. 166).

Yet, the same technology has also made it possible to produce more printed media and packaging than before. According to Mercker, "The so-called paperless society that was projected to evolve with the computer age has actually led to a 40 percent increase in office printing in the United States." He also estimates that paper production will increase by 30 percent to keep up with global demands (1999, p. 1).

Packaging is the second-fastest growing print market. Fruscione reported that growth in the flexographic printing industry in the United States and Mexico has been tremendous because of digital prepress and computer-to-plate advancements. "Flexographic printing held 22 percent of all printing volume last year and 70 percent of the packaging industry" (2000, p. 2).

Publishing is another area that has improved with the use of computers. In the 1960s and 1970s, magazines were published with a colorful cover and black text inside. Process-color images were difficult to reproduce in a magazine and were considered something special. That has changed to the point where, today, most magazines are printed in full color from front to back. Realistic color images are now taken for granted.

A review of the history of printing should point out that a very gradual evolution in printing processes occurred during the course of about 500 years until, in the late twentieth century, digital technology revolutionized the industry. Since then, the process of creating print media has undergone rapid changes and technological advancements that have brought new problems, and new solutions, to the forefront. Ultimately, it seems that we have arrived, as Gutenberg did, at another technological crossroad. Much of what is done in digital prepress today can be repurposed for use in a variety of media forms. Although printing was the primary form of graphic communication for hundreds of years, it is now simply one of a number of ways to output digital information.

REFERENCES

Agfa. (1998). *Agfa digital roadmaps, no. 3*. Ridgefield Park, NJ: Bayer Corporation.

Agfa-Gavaert, N.V. & Bayer Corporation. (1994). *Working with prepress and printing suppliers*. Randolph, MA: Agfa Educational Publishing.

Agfa-Gavaert, N.V. & Bayer Corporation. (1997). *An introduction to digital color prepress*. Randolph, MA: Agfa Educational Publishing.

Atkinson, B. (2001). 10 prepress troubleshooting tips. *Graphic Monthly*. Retrieved October 5, 2002, from *www.graphicmonthly.ca/DigitalReports/novdec01/10ppt.htm*

Cost, F. (1997). *Pocket guide to digital printing*. Albany, NY: Delmar.

Fruscione, E. (May, 2000). FFTA forum tackles flexo's future head-on. *Converting*, *18*(5), 10.

Hogenson, B. (2001). Analysis of digital flexo opportunities for trade shops. Retrieved September 25, 2002, from *unix.barco.com/graphics/packaging/analysis_digital_flexo.htm*

Huff, R. & West, K. (2000). PDF workflow systems. *High Volume Printing, 18*(2), 42-46.

Ingram, S.T. (2001). Color management: A system for improved color reproduction. *SGIA Journal*, (Second Quarter).

International Paper. (2000). *Pocket pal: Graphic arts production handbook*. Memphis, TN: International Paper Company.

Lancaster, F.W. (1978). *Toward paperless information systems*. New York: Academic Press.

Mercker, D.C. (1999). *Landowners dare to say "P" word.* University of Tennessee Agricultural Extension Service. Retrieved September 15, 2002, from *www.utextension.utk.edu/greenTN/2001/026.htm*

Riley, S. (2002). Old habits die hard. *Package Printing*, (August).

Romano, F.J. (1996). *Pocket guide to digital prepress.* Albany, NY: Delmar.

Snyder, M. (1999). Portable Document Format: A versatile communications tool. *Visual Communications Journal*. Sewickley, PA: International Graphic Arts Education Association.

Young, A. (1998). PDF: It's not just for publishing anymore. *Print on Demand Business, 4*(4), 57-60.

KEY TERMS

Desktop Publishing: Creating pages for print media using a standard home or office computer and commercial software, and then outputting them on a simple desktop printer.

Direct-to-Plate: Technology that enables digital layouts to be sent directly to a system that images plates appropriate to the printing process being used.

Flexography: A printing process that uses a raised, or relief-type, image on a flexible photoploymer or rubber plate. Commonly used for printing corrugated and flexible packaging, newsprint, and other media.

Halftone: A method to reproduce photographic images in printed media by creating a pattern of dots that change in size to create the perception of lighter and darker tones.

Imagesetter: An output device used primarily to generate films for the production of printing plates.

Nyquist Theorem: An equation to determine appropriate resolution when scanning an image digitally for graphic reproduction. The equation multiplies the desired printed halftone dot lines per inch (LPI) by 1.5 to 2 times, and then multiplies that value by the percentage of magnification of the reproduction from the original size of the artwork.

PDF (Portable Document Format): A file format that is capable of transferring designs across many different forms of media. PDF files are universal and independent of the computer systems they are on.

Postscript: A text-based page description language developed by Adobe Systems, Inc. that uses precise methods to communicate text and images to an output device.

Stochastic Screening: A method to reproduce photographic images in printed media by creating a random placement of very small dots that change in frequency to create the perception of lighter and darker tones.

Variable Data: Information that changes on sequential pages being output from a printing device.

Chief Knowledge Officers

Richard T. Herschel
Saint Joseph's University, USA

INTRODUCTION

A chief knowledge officer (CKO) is a senior executive who is responsible for ensuring that an organization maximizes the value it achieves through one of its most important assets-knowledge. Knowledge is often defined as information exercised for problem solving, understanding, and benefit. By adopting a CKO, firms formally recognize that knowledge is an asset that needs to be captured, disseminated, and shared to enhance firm performance and value creation. And most of all, they realize it is an asset that must be managed. Knowledge management is seen as essential, because firms today are valued in part on market perceptions of expertise as expressed through their processes, products and services (Choo, 1998).

The "intangibles" that add value to most products and services are knowledge-based — for example, technical know-how, product design, marketing presentation, understanding the customer, personal creativity, and innovation. Critical success factors for organizations today – the need for speed, management of complexity, a sense of history and context, effective judgment, and organizational flexibility — are all related to and dependent upon organizational knowledge (Herschel & Nemati, 2001).

Hence, the fundamental objective of a CKO-driven knowledge management program is to maximize return of investment of intellectual capital expended in the creation and delivery of the firm's products and services. To ensure continuous improvement, knowledge management programs must continuously pursue ongoing organizational learning so as to repeat successes, minimize repeating past failures, and to instill best practices and improve innovation. David Skyrme (2003) reports that the benefits of these programs can be quite significant. He cites the following examples:

- Dow Chemical has realized $125 million through better exploitation of its patent portfolio.
- Texas Instruments has saved investment in a new plant by sharing knowledge of best practice from its existing plants.
- Knowledge management at BP Amoco delivered $260 million to their bottom line in 1998.
- From 1997-2000, Ford Motor Company saved $914 million, mainly due to knowledge management programs.
- Chevron's knowledge management practices have enabled it to save $650 million since 1991.

BACKGROUND

The role of a CKO is to develop the firm's knowledge infrastructure, to promote knowledge capture, storage, and distribution, and to act as a symbol that employees look to for guidance in a knowledge management culture. Bontis (2002) states that a CKO can help a firm to leverage its intellectual capital by:

- promoting stability in a turbulent business environment,
- enabling the speedy delivery of products and services,
- creating high efficiency in the knowledge value chain by sharing of resources and realization of synergies, and
- enabling the separation of work so that specialization is feasible.

The CKO job description can encompass a number of responsibilities. For example, the CKO might be responsible for leading executive management to develop an enterprise knowledge strategy, validating this strategy across the enterprise, and then ensuring that its evolution complements and integrates with business strategy. The CKO may also be charged with setting priorities and securing funding for knowledge management (KM) programs as well as defining policies for security, usage, and maintenance of intellectual capital. Depending on the organizational culture, the CKO may also act as the chief advocate for KM as a discipline — walking and talking the program throughout the enterprise and assisting executives and senior management in building and communicating personal commitment and advocacy for KM (Davenport & Prusak, 1998).

In any scenario, the CKO must at least be responsible for managing and administering the day-to-day activities of the KM program and the KM infrastructure. This means overseeing the development of KM technology and information exchange architecture and assessing specific knowledge needs of business processes. Moreover, the CKO must ensure the integration of KM into employees' job activities, into key processes, and across communities of practice. They should also lead the development of formal programs that foster knowledge sharing and innovation, define expectations for individual and community participation in the KM program, and ensure participation and knowledge contribution by all levels of experts (including executive management). In addition, they should create a process to measure benefits and progress against program goals including competitive advancements, knowledge created, innovations, cost savings, speed of response, development of experts, sharing and participation, and then communicate achievements and shortcomings of the program. Finally, the CKO should manage relationships with external providers of information and knowledge and negotiate contracts with them (Harris, 1998).

While CKO job descriptions vary among firms, David Skyrme (2003) states that most CKO responsibilities include:

- developing an overall framework that guides knowledge management,
- actively promoting the firm's knowledge agenda within and beyond the company,
- overseeing the development of a knowledge infrastructure for the firm (e.g., via procedures and technology), and
- facilitating connections, coordination, and communication to enable knowledge exchange.

In a White Paper published at its Web site, DestinationKM (2000) states that identifying the need for a knowledge management officer or CKO and choosing that person may be among the most difficult endeavors a company undertakes. They argue that one way to approach these decisions is to ask at what stage in the corporate evolution a company recognizes the need for a knowledge management officer. Often, they claim, this realization comes after failures that affect the viability of a company. For example, companies that rely exclusively on paper trails and file cabinets to retain customer information could lose information, fail to note valuable customer characteristics and misrepresent the financial activity of the business. In many cases, they assert, company owners and top executives have been unaware of problems until they begin to lose customers and suffer dramatic losses in revenue. Other companies report implementing knowledge management and/or establishing a CKO position after competitors have hired away key employees who take the company's product knowledge, customer knowledge and confidential information with them.

DestinationKM states that identifying candidates for the CKO position can happen in a number of ways. In some cases, the potential CKO may already be part of the organization that utilizes knowledge management to improve business processes, but the CKO title or official position does not exist. For example, if a chief financial officer or senior business executive is already on board, the knowledge management role may become part of that individual's title and responsibility. Because business leaders tend to look in the direction of higher-ranking personnel and forward-thinking company officials to fill the CKO post, Destination KM asserts that it is not unusual for a member of the top management team to be sought for this role because these individuals have extensive knowledge of the company and already hold significant responsibilities

Davenport (1994) claims that to be successful, certain personal characteristics are critical for the CKO. These characteristics include:

- deep experience in some aspect of knowledge management, including its creation, dissemination, or application,
- familiarity with knowledge-oriented companies and technologies, such as libraries and groupware, and
- the ability to set a good example by displaying a high level of knowledgeability and success.

In a study of CKOs, TFPL (1998) finds that the majority of CKOs tend to possess these attributes. Their research shows that most CKOs emerge from planning teams and that they possess one of three main backgrounds – information technology (IT), human resources, or a core business function. But, TFPL notes, their common strength is their understanding of the organization and its business drivers, combined with an ability to take a holistic view of the company, and to understand the mix of hard and soft skills necessary to create, sustain and utilize the knowledge base.

MANAGING DIFFERENT FORMS OF KNOWLEDGE

A CKO's implementation of knowledge management programs often depends on his or her attitudes about explicit knowledge and implicit knowledge.

Explicit knowledge is defined as knowledge that can be expressed formally using a system of symbols, and can therefore be easily communicated or diffused. It is either object-based or rule-based. It is object-based when the knowledge is codified in strings of symbols (e.g., words, numbers, formulas) or in physical objects (e.g., equipment, documents, models). Object-based knowledge may be found in examples such as product specifications, patents, software code, computer databases, technical drawings, and so forth. Explicit knowledge is rule-based when the knowledge is codified into rules, routines, or standard operating procedures (Choo, 1998; Nonaka & Takeuchi, 1995).

CKOs preoccupied with explicit knowledge apparently tend to promote the use of technology to enable information distribution and sharing. Technologies that CKOs can employ in the explicit knowledge management process include text storage and retrieval systems, databases, expert systems, Lotus Notes groupware, and intranets. While use of technology to facilitate explicit knowledge transfer is a major concern for these organizations, Schreiber and Carey (2003) note that the interplay between task complexity, experience effects, and the type of data employed can affect its success.

Organizational members perform their work and make sense of their world using implicit knowledge as well. Implicit knowledge is knowledge that is uncodified and difficult to diffuse. It is hard to verbalize because it is expressed through action-based skills and cannot be reduced to rules and recipes. Implicit knowledge is learned through extended periods of experiencing and doing a task, during which the individual develops a feel for and a capacity to make intuitive judgments about the successful execution of the activity. Implicit knowledge is vital to the organization because organizations can only learn and innovate by somehow leveraging the implicit knowledge of its members. Despite its being uncodified, implicit knowledge can be and is regularly taught and shared. It can be learned by example. Implicit knowledge becomes substantially valuable when it is turned into new capabilities, products, or services (Choo, 1998).

It is important to note that implicit knowledge serves a critical role in the innovation process. It facilitates problem definition, problem solving, and predicting the outcomes of potential solutions (Nonaka & Takeuchi, 1995). However, it is often problematic to express implicit knowledge in a manner that others understand and hence formal intellectual exchanges as well as apprenticeships should be used to tap implicit knowledge. In two experiments examining the exchange of implicit knowledge, Herschel, Nemati, and Steiger (2001, 2003) find that the use of knowledge exchange protocols to structure the delivery of implicit information results in better understanding and retention by those with whom implicit knowledge is shared.

CKOs who emphasize implicit knowledge seem to spend a great deal of their time enabling, facilitating, and promoting informal dialogues. These CKOs appear to accept the premise that managing implicit knowledge means accepting chaos, ambiguity, and uncertainty and they feel that capturing implicit knowledge is expensive because it can only be revealed through conversations. For these CKOs, technology cannot be a driving force in implicit knowledge management because it lacks the media richness of face-to-face conversations — though some CKOs do promote video conferencing, because it enables non-verbal communication (Earl & Scott, 1999).

A decision by a CKO to emphasize one type of knowledge over another is often based more on pressures to produce demonstrable results than on the logic of facilitating organizational "knowing". However, such a strategy can be problematic. Choo (1998) and Nanaka and Takeuchi (1995) contend that failing to deal with both forms of knowledge and the conversion of one form to the other adversely impacts the effectiveness of knowledge management efforts. They argue that both forms of knowledge are critical, as is the interaction between them. These knowledge management experts believe that if a CKO sidesteps this issue, the CKO is inevitably compromising the long-term success of the knowledge management effort itself.

RESOLVING THE CKO CONTROVERSY

Twenty-five percent of Fortune 500 companies have a CKO and 80% of Fortune 500 companies have a knowledge management staff. Moreover, 42% of Fortune 500 companies anticipate appointing a CKO within the next 3 years (Flash, 2001).

Nevertheless, there are firms who believe the CKO function is not needed. Many of them feel that having a CKO is the wrong way to harness corporate know-how. Instead, these firms prefer a more grassroots approach, in which a team of knowledge management experts works closely with — or even as part of — the business units (Cole-Gomolski, 1999).

CKOs are often supposed to straddle business and information technology (IT) with a mandate to convince workers that it is good to share information and to work with IT to build applications to support such sharing. The problem is that some companies find that putting more control of knowledge management in the hands of end users themselves makes it an easier sell because knowledge sharing happens within business units. These firms believe that centralizing knowledge management under a CKO sends out the wrong message (Cole-Gomolski, 1999).

In firms where CKOs exist, Pringle (2003) notes that many of these have survived by judiciously distancing themselves from the original "craze," while still exploiting knowledge management concepts. This oftentimes means that CKOs do not talk about knowledge management per se. Instead, the CKO pursues activities that encourage employees to talk to one another or that allow workers to reuse already existing materials or information. Pringle notes that these CKOs typically imbed knowledge management activities within performance management systems that give staff members the incentive to learn and to share their expertise. That is, assessments of employee sharing of information efforts as well as demonstrable learning activities become an integral part of employee annual performance reviews.

Earl and Scott (1999) argue that appointing a CKO is indeed one way of galvanizing, directing, and coordinating a knowledge management program. However, they also feel that it is unlikely to be sufficient in itself; nor is it likely to be universally necessary. They concede that some organizations are investing in aspects of knowledge management without appointing a CKO. They believe that knowledge management, like total quality management, will become embedded in organizations and knowledge will become an obvious imperative source of value creation and competitiveness. In this scenario, all members of firms will own and drive knowledge management and, hence, the importance of and need for a CKO may decline.

FUTURE TRENDS

Contingency theory is useful in explaining company behavior relative to CKO adoption and deployment. Contingency theory states that desired organizational outcomes — such as performance — are the consequence of "fit" or match between two or more factors (e.g., strategy, culture, structure, tasks, environment and leadership style). The best way to organize depends upon the characteristics and the interplay between these factors (Scott, 1987).

Hence, the role of a CKO in a particular organization may inevitably depend on organizational factors. For example, in mature, knowledge-sharing organizational cultures, the CKO function may not need to be formalized as a distinct management position. However, in other firms, the relationship between knowledge management and other factors may not be as well integrated. In these firms, a CKO may be needed to act as a driver for organizational change. The point is that the focus of knowledge management will continue to vary significantly from one organization to another, and therefore so may the required leadership skills.

A multi-client study conducted by The Delphi Group revealed that, in practice, knowledge leadership spans a broad domain of positions and types of individuals. However, regardless of the differences in title or organizational placement, knowledge leaders appear to exhibit similar qualifications. The most notable have hybrid business/information technology experience, at least 10 years (and often much more) of line-of-business experience, an entrepreneurial attitude and a relatively high level of interest in career definition and development. They also display acumen in corporate communication throughout the formal hierarchy reflected in most organization charts, as well as in the informal networking that occurs in communities of practice or the "corporate underground". It is through all these channels that knowledge leaders may introduce methods and systems to match knowledge seekers with knowledge providers (Capshaw & Koulopoulos, 2004).

Hence, there does appear to be some common ground about what it will take to manage knowledge programs. In a recent survey at various companies, McKinsey (2004) asked CKOs for their views about make-or-break factors when it comes to managing knowledge. Although the CKOs had different experiences, all concurred that success will depend on two things: first, on the ability of senior management to agree about what it hopes to gain from managing knowledge explicitly and from creating a performance culture (which raises the demand for knowledge) and, second, on how well the CKO develops and executes a knowledge-management agenda.

CONCLUSION

According to Tom Davenport (1996), CKOs often oversee efforts to use technology for knowledge capture and distribution. However, CKOs must also foster sharing of implicit knowledge, which is less amenable to technology transfer. In this instance, CKOs must facilitate an organizational climate of knowledge sharing by encouraging such behaviors as mentoring and open dialogue. Hence, CKOs have three critical responsibilities: creating a knowledge management infrastructure, building a knowledge-sharing culture, and making it all pay off. Because of the complexity of the CKO job, these individuals necessarily need to have extensive knowledge of the company and already have significant organizational responsibilities.

REFERENCES

Bontis, N. (2002). The rising star of the chief knowledge officer. *Ivey Business Journal*, 20-25.

Capshaw, S., & Koulopoulos, T.M. (2004). Knowledge leadership. *DM Review. http://www.dmreview.com/portal_ros.cfm?NavID=91&EdID=20&PortalID=17*

Choo, C.W. (1998). *The knowing organization.* New York: Oxford University Press.

Cole-Gomolski, B. (1999). Knowledge 'czars' fall from grace. *Computerworld, 33*(1),1,13.

Davenport, T. (1996, April 1). Knowledge roles: The CKO and beyond. *CIO,* 24.

Davenport, T., & Prusak, L. (1998). *Working knowledge.* Boston: Harvard Business School.

DestinationKM. (2000). *Knowledge management officer basics.* White Paper. *http://www.destinationkm.com/articles/default.asp?ArticleID=838*

Earl, M., & Scott, I. (1999). Opinion: What is a chief knowledge officer? *Sloan Management Review, 40*(2), 29-38.

Flash, C. (2001, February 20). Who is the CKO? *Knowledge Management Magazine.*

Harris, K. (1998). Chief knowledge officer: Managing intellectual assets. *Gartner Group Document #KA-03-8437.* Boston, MA.

Herschel, R., & Nemati, H. (2001). Chief knowledge officers: Managing knowledge for organizational effectiveness. In *Knowledge management and business model innovation* (pp. 414-425). Hershey, PA: Idea Group Publishing.

Herschel, R., Nemati, H., & Steiger, D. (2001) Tacit to explicit knowledge conversion: Knowledge exchange protocols. *Journal of Knowledge Management, 5*(1), 107-116.

Herschel, R., Nemati, H., & Steiger, D. (2003). Knowledge exchange protocols: A second study. *Journal of Information and Knowledge Management: Special issue of JIKMS on "Knowledge Management in Context and Context for Knowledge Management," 2*(2), 153-163.

McKinsey Quarterly. (2004). *http://www.mckinseyquarterly.com/article_abstract.asp?tk=:1075:18&ar=1075&L2=18&L3=30*

Nonaka, I., & Takeuchi, H. (1995). *The knowledge creating company.* New York: Oxford University Press.

Pringle, D. (2003, January 7). Chief knowledge officers adapt to remain relevant, employed. *Wall Street Journal, 1.*

Schreiber, C., & Carey, K. (2003). The impact of databases on knowledge transfer: Simulation providing theory. *NAACSOS Conference Proceedings,* Pittsburgh, PA.

Scott, R.W. (1987). *Organizations: Rational, natural, and open systems.* Englewood Cliffs, NJ: Prentice-Hall.

Skyrme, D. (2003). Do you need a CKO? *Insights, 1-2.*

TFPL. (1998). What is a CKO – a concept or a team. *TFPL's knowledge management research – interim report,* 1-3. *http://www.tfpl.com/consult/what_is_a_cko.htm*

KEY TERMS

Chief Knowledge Officer (CKO): A senior level executive responsible for managing a firm's knowledge management initiative.

Contingency Theory: States that desired organizational outcomes — such as performance — are the consequence of "fit" or match between two or more factors (e.g., strategy, culture, structure, tasks, environment and leadership style). The best way to organize depends upon the characteristics and the interplay between these factors.

Explicit Knowledge: Knowledge that can be expressed formally using a system of symbols, and can therefore be easily communicated or diffused. It is either object based or rule based.

Implicit Knowledge: Knowledge that is uncodified and difficult to diffuse. It is hard to verbalize because it is expressed through action-based skills and cannot be reduced to rules and recipes. Implicit knowledge is the same as tacit knowledge.

Intellectual Capital: Can be divided into three categories:

- *Human Capital* - that in the minds of individuals: knowledge, competences, experience, know-how, and so forth
- *Structural Capital* - "that which is left after employees go home for the night": processes, information systems, databases and so forth
- *Customer Capital* - customer relationships, brands, trademarks and so forth

Knowledge Exchange Protocols: A method of structuring tacit knowledge sharing so as to make its conversion to explicit knowledge more efficient. It emphasizes the use of dialogue segmenting accompanied by structural content headings to deliver implicit content.

Knowledge Management: A program for managing a firm's intellectual capital by systematically capturing, storing, sharing, and disseminating the firm's explicit and tacit knowledge.

Knowledge Management Architecture: The technology and procedural platforms employed by a firm to enable knowledge capture, sharing, retention, and distribution.

Citizenship and New Technologies

Peter J. Smith
Athabasca University, Canada

Elizabeth Smythe
Concordia University College of Alberta, Canada

BACKGROUND

Not long ago globalization had only one face, that of a restructured capitalist economy employing new information technologies to operate on a global scale (Castells, 2000). According to this interpretation of globalization, the global is represented as space dominated by the inexorable and homogenizing logic of global markets (Steger, 2002). In this neoliberal model, the market replaces the state and the individual, the community thus posing a bleak future for citizenship.

Today, however, this one-dimensional view of globalization no longer holds sway. Transnational protests over the proposed Multilateral Agreement on Investment (MAI) and the meetings of the World Trade Organization (WTO) in Seattle are now seen as critical moments that speak to a new critical citizen activism that relies heavily upon new technologies and means of horizontal communication to operate on a global scale (Khagram, Riker & Sikkink, 2002; Ritchie, 2002). These means of communicating have provided an alternative voice and capacity to mobilize those excluded from an increasingly narrow spectrum of public discourse.

INTRODUCTION

This article examines this other side of globalization, in particular, the development of new processes and means of communicating and organizing which have enabled new forms of expression and connection among citizens which are not easily controlled by states and ruling elites. It does so by using two case studies of the failed attempt to negotiate the MAI at the Organization for Economic Cooperation and Development (OECD) in 1995-1998 and the failed effort to launch a new round of trade negotiations at the WTO Ministerial meetings in Seattle in 1999. These are good cases since both were based on a set of core neoliberal values shared by state elites of the 29 OECD member countries and the dominant members of the WTO. Despite having virtually no access initially to mainstream media and limited resources, a broad coalition of non-governmental organizations (NGOs) were able to challenge these elites and organize an effective opposition to the MAI and the Seattle meeting.

GLOBALIZATION, NEW INFORMATION TECHNOLOGIES, AND GLOBAL CIVIL SOCIETY

To understand how globalization might empower citizens, we must recognize that the information revolution made globalization possible. As Kobrin notes, "The emerging global world economy is electronic, integrated through information systems and technology rather than organizational hierarchies" (Kobrin, 1998, p. 362). However, while the information revolution breaks down hierarchies, it facilitates the creation of new power structures, redistributing not only economic, but also political power. Where economic globalization appears to be closing public spaces for state-centered citizenship, it may be opening them up elsewhere.

As a number of studies have indicated, information and communication technologies (ICTs), especially the Internet, have become important means of communication for many NGO networks (Warkentin, 2001; Smythe & Smith, 2003). Indeed, the Internet has been compared to the 'Swiss Army Knife' of political advocacy and organization (Hillwatch, 2001). This is because the Internet has become a multi-faceted tool providing civil society organizations with a number of advantages including:

- The facilitation of internal communication and provision of services to members—Strategy discussions and input of members at a distance are all made easier.
- The opening up of new public spaces for civil society organizations—The Internet can be used to educate the broader public about public policy and frame policy debates.
- The production, dissemination, and sharing of large amounts of information—Small, resource-poor organizations can, via the Internet, quickly and easily access research done by the organizations that do have these resources.

- The facilitation of public participation and mobilization—Thanks to the Internet civil society, organizations can quickly inform, organize, and mobilize their supporters.

The Internet thus possesses a variety of features that promote citizen engagement at a global level. The most frequently cited examples of how forces in civil and global civil society have combined to resist the traditional closed-door, top-down, multilateral processes are the MAI and the ministerial meeting in Seattle, and its failed attempt to launch the Millennium Round of trade negotiations (Ayres, 2003).

WHY THE MAI AND SEATTLE CASES?

Why use these two campaigns to empirically examine the argument about globalization's impact on citizens, democracy, and states? Primarily because many have argued that the root of the constraints states currently face in their capacity to make policy is globalization (McBride, 2003). The MAI, for example, was clearly designed and intended to limit state discretionary authority to discriminate between domestic and foreign investment in regulation.

Similarly, the expansion of areas of economic activity covered by trade rules have led to concerns that the WTO's rules are severely constricting the scope of state policies in favor of markets and corporations. If ICTs have facilitated contestation of globalization, it should be most apparent in the case of organizations like the OECD and the WTO.

The methodology used in examining these cases involved a variety of tools, including the analysis of thousands of Web pages and sites, and detailed interviews with a number of the organizations most actively involved in the campaigns, including participant observation in Seattle with an accredited NGO, along with the standard review of primary and secondary sources available.

THE MAI AND THE INTERNET

The MAI negotiations, formally launched in May, 1995 at the OECD Ministerial meeting, were drawn out over three-and-a-half years and ended in failure on December 2, 1998. While, at first, negotiations progressed rapidly in matters of general principles, they began to stumble in the winter of 1997 over the question of which economic sectors or state policies would be exempted from these general principles. Around the same time a draft of the February 1997 draft negotiating text was leaked and ended up very quickly on the Web sites of two public policy advocacy groups in North America. Once the text became known, public pressure—facilitated by the Internet—began to mount in a number of countries. In mid-October 1998, the French government, under pressure from the Green and Communists within its coalition, withdrew from the negotiations (Riché, 1998), thereby ending any real hope of agreement.

How important was the Internet in the politicization and mobilization against the MAI? To answer this question the authors began by surveying the presence on the Internet of MAI Web sites in February and March 1999, less than three months after the cessation of negotiations

Table 1. Web sites by type of sponsoring organization

Type of Group Sponsoring Site	Number of Sites	Per Cent of Total Sites
Public policy advocacy	53	15.1%
Political parties and MP sites	45	12.8
Media organizations	37	10.5
Government agencies—all levels	35	10
Individual/personal	30	8.5
Business organizations (including law offices)	26	7.4
Broad, anti-MAI coalitions	20	5.7
Environmental organizations	19	5.4
Trade unions	16	4.6
International organizations	17	4.8
Research institutes/centers	15	4.3
Student groups	9	2.6
Other (unable to classify)	9	2.6
Arts/cultural organizations	8	2.3
Church/religious	5	1.4
Total	**352**	**100%**

Table 2. Top 10 organizations appearing as links of Web sites

Name of Organization	Frequency	Rank
OECD	95	1
MAI Not (Flora)–OPIRG-Carleton University	87	2
National Center for Sustainability (Victoria BC)	25	3
Appleton (law office, Toronto)	23	4
YUCC (personal site, York University law student)	22	5
Public Citizen (Washington, DC)	21	6
Preamble Center (Washington, DC)	21	6
Friends of the Earth-U.S. (Washington, DC)	19	7
Multinational Monitor (U.S.—linked to Public Citizen)	17	8
Council of Canadians	16	9
Canadian Center for Policy Alternatives (Ottawa)	12	10
Stop MAI–Australia	12	10

on December 2. Many of the Web sites had not been updated and were still on the Net a year later; the survey is broadly indicative of what citizens searching for information would have found, for example, in the winter of 1998, when public debate about the MAI was at its height.

Given that much of the anecdotal discussion about the Internet and the MAI claimed that the Web was mostly a tool of NGOs, the authors were surprised by the number of Web sites sponsored by members of legislatures, political parties (in most case, opposition parties), media organizations, governments, business and other organizations. Interviews suggested that many of the media and government Web sites followed in response to those of the advocacy, environmental, and development NGOs.

The authors also tried to get a sense of the countries where organizations were located which were active on the Web. The data indicated, not surprisingly, that the United States had the most Web sites, 31.9% of the 400 coded. However, the relatively large number of Canadian sites (17.6% of total) is striking and reflective of both the high level of connectedness in Canada and of the controversy the agreement generated in that country.

How is information linked and shared on the Web? Commentators have pointed to two aspects that make it a speedy and effective way to share and gather information—these are the use of hypertext links, and the extent to which information may be reproduced and shared in a variety of Web sites based around the world. As Table 2 indicates, virtually all of the Web sites had links (650) to other sites that provided information.

As indicated, 10 organizations accounted for over half of all of the hypertext links that appeared on the 400 Web sites. Virtually all of the other sites are North American, with the dominance of Canadian sites.

Interviews indicated that NGOs were strikingly similar in the way that they used the Internet, and what variation there was often reflected their size or restrictions on their activities dictated by their funding sources. Virtually all of the organizations used a Web site, e-mail, and a listserv as part of their anti-MAI activities. Web sites were generally targeted at a broader public than just a group's main membership. Their main function was to provide a means of gathering and sharing information and mobilizing those concerned about the agreement.

In the end, did the Internet really have an impact on the resulting failure of the negotiations? Interviews indicated that while none of the NGOs felt that the Internet campaign alone was responsible for the defeat of the MAI, all felt that the Internet campaign set the terms of public debate over the MAI.

THE CASE OF SEATTLE: THE ROAD TO THE WTO MINISTERIAL MEETING

The MAI and the WTO-Seattle case have direct links in the concern of many NGOs that the failed attempt to negotiate investment rules at the OECD would simply migrate to the WTO if a few powerful actors had their way. It is also linked in the lessons learned by both NGOs and governments and organizations like the WTO about the power and potential of ICTs. But concerns about the WTO in the globalization process also have roots in the earlier rounds of trade negotiations. The last and most significant was the Uruguay Round. It created the WTO in 1995 and left its members two key issues, services and agriculture, with built-in mandates for further negotiations.

The likelihood for severe conflict among WTO members meeting in Seattle was already very high given the presence of agriculture on the agenda. There was also

C

Table 3. Web sites by organization types

Type of Organization	Number of sites	% of Total
Public advocacy	94	18.3
Media organizations	88	17.2
Other (includes anarchist and feminist organizations)	75	14.6
Government agencies (all levels)	52	10.1
Environmental & conservation organizations (including animal welfare)	44	8.6
Business organizations	41	7.8
Trade unions	33	6.4
Development organizations	32	6.2
Political parties	20	3.9
Farming organizations	18	3.5
International organizations (e.g., WTO, OECD)	16	3.1
TOTAL	513	100

Table 4. Top 10 organizations appearing as links of Web sites

Name of Organization	Frequency	Rank
World Trade Organization	2,129	1
Seattle Post-Intelligence	732	2
Org. Econ. Coop. Dev.	322	3
OneWorld Online	348	4
Washington Council on International Trade (host organization to *wtoseattle.org*)	127	5
Financial Times	125	6
Seattle Times	123	7
Institute de Recherche pour le Développement (IRD)	122	8
Third World Network	116	9
Institute for Global Communications	111	10

evidence of real discontent on the part of many developing countries who, now, along with economies in transition, vastly outnumbered the developed country members. Many of these developing countries had taken on new and onerous obligations as a result of the Uruguay Round, in return for little real results in improved market access for their goods in developed economies.

NGOs, as a result of the 18-month campaign of opposition to the MAI, were already organized. The WTO campaign, however, was even broader in scope, both in terms of countries, issues, and types of organizations involved. Opponents' preparations were underway more than a year in advance of the Seattle meeting, and went into high gear once the city had been chosen as the meeting venue (Brecher, Costello & Smith, 2000).

One major difference from the anti-MAI campaign was that this campaign sought to use the meeting venue itself as a site for a public display of opposition. Protests and direct action were designed to disrupt the meeting itself and stop the launch of a round of negotiations.

THE WTO MINISTERIAL AND THE INTERNET

What role did the Internet play in this campaign? The role of the Internet included three aspects: organizing, educating, and mobilizing. The Internet proved invaluable in coordinating and organizing the activities of thousands of groups and individuals. In terms of educating the public, detailed analyses and research were distributed worldwide. In regards to mobilization, many North American organizations also used their sites to encourage and call on citizens to come to Seattle.

In their second analysis of sites, the authors used 40 search engines and identified 4,089 Web sites with material specific to the Seattle round of the WTO. Of these sites 513 were selected to examine and classify.

As Table 3 indicates, those using the Web included a broad range of civil society. However, the significant presence of governments reflected the extent to which they too had learned important lessons and turned to the Internet to get their message out.

The authors also looked at the distribution of sites in terms of origin. What is striking is the fact that the sites included virtually every region of the world. The North still dominated with a majority in the United States. Not surprisingly, most sites contained links to the WTO and the Seattle host organization, the key place for information about the Ministerial, NGO registration, schedules, housing, and the city.

According to public opinion polls, the campaign also had an impact in raising public awareness about the WTO (Agnus Reid, 2002). However, it is debatable whether the civil society campaign can take sole credit for turning Seattle into a fiasco. Differences among member states on key issues, such as agriculture and the discontent of many developing countries with the implementation of the Uruguay round, were also important contributing factors.

FUTURE TRENDS

This examination of the use the Internet to mobilize against the MAI and a new round of trade negotiations in Seattle indicates that the growth of a global civil society has the potential to enhance citizenship globally and domestically. What is particularly striking is the extent to which the use of the Net has leveled the playing field in terms of who can now participate, decentering public discussion away from government agencies and business organizations. Clearly, civil society organizations have become particularly adept at horizontal communications, thus eroding the monopoly of information possessed by traditional economic, broadcast media, and political hierarchies. Indeed, the ability to publicize and disseminate information quickly has made closed-door, secretive trade negotiations difficult, if not impossible.

CONCLUSION

While the use of the Internet enhances the potential for citizenship in a number of respects, those interviewed were perfectly aware of its limitations. A digital divide does exist. Not all information provided is good, accurate, or timely. Often there is too much information, and it is impossible to analyze and respond to all of it. Furthermore, communication on the Net is often intemperate and tangential to the debate. The Net, then, is not a panacea for all that ails citizenship in a global era. It does, however, have the potential to enhance and enrich it.

REFERENCES

Agnus Reid. (2002, April 19). *Seattle protests against globalization*. Retrieved from *www.agnusreid.com/media/*

Ayres, J.M. (2003). Contesting neoliberalism: The political economy of transnational protest. In M.G. Cohen & S. McBride (Eds.), *Global turbulence: Social activists' and state responses to globalization* (pp. 89-105). Burlington, VT: Ashgate Publishing.

Brecher, J., Costello, T. & Smith, B. (2000). *Globalization from below*. Cambridge, MA: South End Press.

Castells, M. (2000). *The rise of the network society* (2nd edition). Oxford: Blackwell Publishers.

Hillwatch. (2001, April). *The Internet, transparency and public advocacy*. Submission to the Canadian House of Commons Industry Committee. Retrieved from *www.hillwatch.com/submissioneng.htm*

Keck, M.E. & Sikkink, K. (1998). *Activists beyond borders: Advocacy networks in international politics*. Ithaca, NY: Cornell University Press.

Khagram, S., Riker, J.V. & Sikkink, K. (2002). From Santiago to Seattle: Transnational advocacy groups restructuring world politics. In S. Khagram, J.V. Riker & K. Sikkink (Eds.), *Restructuring world politics: Transnational social movements, networks, and norms* (pp. 3-24). Minneapolis, MN: University of Minnesota Press.

Kobrin, S. (1998). Back to the future: Neo-medievalism and the post-modern digital economy. *Journal of International Affairs, 51*(2), 361-386.

McBride, S. (2003). International economic agreements as systems of power. In M.G. Cohen & S. McBride (Eds.), *Global turbulence: Social activists' and state responses to globalization* (pp. 26-41). Burlington, VT: Ashgate Publishing Company.

Riché, P. (1998). Jospin: Adieu l'AMI, salut les copains. *Liberation*, (October 15), 112-127.

Ritchie, M. (2002). A practitioner's perspective. In S. Khagram, J.V. Riker & K. Sikkink (Eds.), *Restructuring world politics: Transnational social movements, networks, and norms* (pp. 294-298). Minneapolis, MN: University of Minnesota Press.

Smith, P.J. & Smythe, E. (2004). Globalization, citizenship and new information technologies: From the MAI to Seattle. In M. Mälkiä, A.-V. Anttiroiko & R. Savolainen (Eds.), *E-transformation in governance*. Hershey, PA: Idea Group Publishing.

Smythe, E. & Smith, P.J. (2003). New technologies and networks of resistance. In E. Potter (Ed.), *Cyberdiplomacy*. Kingston: McGill-Queens Press.

Steger, M.B. (2003). *Globalism: The new market ideology*. Lanham, MD: Rowman and Littlefield.

Warkentin, C. (2001). *Reshaping world politics: NGOs, the Internet and global civil society*. Lanham, MD: Rowman and Littlefield.

Waters, M. (1995). *Globalization*. London: Routledge.

KEY TERMS

Civil Society: Refers to the sphere of associative activity between the individual and the state. Analysts differ between those who include the economy and those who do not. The authors include the economy.

Global Civil Society: Consists of the myriad groups and networks of action and knowledge that can, but do not necessarily, extend across state borders.

Globalization: Commonly refers to the process whereby the capitalist economy becomes increasingly integrated and organized on a global scale. More broadly, it refers to a social process whereby "the constraints of geography on social and cultural arrangements recede and in which people become increasingly aware that they are receding" (Waters, 1995, p. 3).

Multilateral Agreement on Investment: A draft agreement designed to protect foreign investors by limiting host state regulation. The 30 members of the Organization for Economic Cooperation and Development failed to reach agreement in 1998 after three years, and negotiations ceased.

Networks: Defined by Keck and Sikkink as "forms of organization characterized by voluntary, reciprocal, and horizontal patterns of exchange" (1998, p. 8).

Non-Governmental Organizations (NGOs): Non-profit, self-governing organizations with voluntary membership.

World Trade Organization: Includes 148 member states and was established in 1995 as successor to the General Agreement on Tariffs and Trade. Its primary purpose is to liberalize international trade through rounds of multilateral negotiations and to resolve trade disputes among member countries.

Classification-Rule Discovery with an Ant Colony Algorithm

Rafael S. Parpinelli
CEFET-PR, Brazil

Heitor S. Lopes
CEFET-PR, Brazil

Alex A. Freitas
University of Kent, UK

INTRODUCTION

Ant colony optimization (ACO) is a relatively new computational intelligence paradigm inspired by the behaviour of natural ants (Bonabeau, Dorigo & Theraulaz, 1999). The natural behaviour of ants that we are interested in is the following. Ants often find the shortest path between a food source and the nest of the colony without using visual information. In order to exchange information about which path should be followed, ants communicate with each other by means of a chemical substance called pheromone. As ants move, a certain amount of pheromone is dropped on the ground, creating a pheromone trail. The more ants follow a given trail, the more attractive that trail becomes to be followed by other ants. This process involves a loop of positive feedback, in which the probability that an ant chooses a path is proportional to the number of ants that have already passed by that path.

Hence, individual ants, following very simple rules, interact to produce an intelligent behaviour – a solution to a complex problem – at the higher level of the ant colony. In other words, intelligence is an emergent phenomenon; that is, "the whole is more than the sum of the parts".

In this article we present an overview of Ant-Miner, an ACO algorithm for discovering classification rules in data mining (Parpinelli, Lopes & Freitas, 2002a, 2002b). In essence, in the classification task each case (record) of the data being mined consists of two parts: a goal attribute, whose value is to be predicted, and a set of predictor attributes. The aim is to predict the value of the goal attribute for a case, given the values of the predictor attributes for that case.

To the best of our knowledge, the use of ACO algorithms (Bonabeau, Dorigo & Theraulaz, 1999; Dorigo et al., 2002) for discovering classification rules is a very under-explored research area. There are other ant algorithms developed for the data mining task of clustering – see for example Monmarché (1999) – but that task is very different from the classification task addressed in this article. Note that Ant-Miner was designed specifically for discovering classification rules, rather than for solving other kinds of data mining tasks.

In other research areas ACO algorithms have been shown to produce effective solutions to difficult real-world problems. A detailed review about many other ACO algorithms (designed to solve many other different kinds of problems) and a discussion about their performance can be found in Bonabeau, Dorigo and Theraulaz (1999) and Dorigo et al. (2002).

A typical example of application of ACO is network traffic routing, where artificial ants deposit "virtual pheromone" (information) at the network nodes. In essence, the amount of pheromone deposited at each node is inversely proportional to the congestion of traffic in that node. This reinforces paths through uncongested areas. Both British Telecom and France Telecom have explored this application of ACO in telephone networks.

ANT COLONY OPTIMIZATION

An ACO algorithm is essentially a system based on agents that simulate the natural behavior of ants, including mechanisms of cooperation and adaptation.

ACO algorithms are based on the following ideas:

- Each path followed by an ant is associated with a candidate solution for a given problem;
- When an ant follows a path, the amount of pheromone deposited on that path is proportional to the quality of the corresponding candidate solution for the target problem;

- When an ant has to choose between two or more paths, the path(s) with a larger amount of pheromone have a greater probability of being chosen by the ant.

As a result, the ants eventually converge to a short path, hopefully the optimum or a near-optimum solution for the target problem.

In essence, the design of an ACO algorithm involves the specification of (Bonabeau, Dorigo & Theraulaz, 1999):

- An appropriate representation of the problem, which allows the ants to incrementally construct/modify solutions through the use of a probabilistic transition rule, based on the amount of pheromone in the trail and on a local, problem-dependent heuristic;
- A method to enforce the construction of valid solutions;
- A problem-dependent heuristic function (h) that measures the quality of items that can be added to the current partial solution;
- A rule for pheromone updating, which specifies how to modify the pheromone trail (t);
- A probabilistic transition rule based on the value of the heuristic function (h) and on the contents of the pheromone trail (t) that is used to iteratively construct a solution.

Artificial ants have several characteristics similar to real ants, namely:

- Artificial ants have a probabilistic preference for paths with a larger amount of pheromone;
- Shorter paths tend to have larger rates of growth in their amount of pheromone;
- The ants use an indirect communication system based on the amount of pheromone deposited on each path.

MOTIVATIONS FOR USING ACO

ACO possesses a number of features that are important to computational problem solving (Freitas & Johnson, 2003):

- The algorithms are relatively simple to understand and implement, whilst also offering emergent complexity to deal effectively with challenging problems;
- They can be readily hybridized with other techniques and/or problem-dependent heuristics in a synergistic fashion;
- They are compatible with the current trend towards greater decentralization in computing;
- The algorithms are highly adaptive and robust, enabling them to cope well with noisy data.

Two more features of ACO are particularly useful in data mining applications:

- Many projects in the field of data mining were developed using deterministic decision trees or rule induction algorithms. These algorithms are hill climbing like and are susceptible to finding only locally optimal solutions instead of the global optimum. The utilization of ACO to induce classification rules tries to mitigate this problem of premature convergence to local optima, since ACO algorithms have a stochastic component that favors a global search in the problem's search space;
- Unlike classical methods for rule induction, the ACO heuristic is a population-based one. This characteristic has advantages over other methods because it allows the system to search many different points in the search space concurrently and to use the positive feedback between the ants as a search mechanism.

REPRESENTING A CANDIDATE CLASSIFICATION RULE

In Ant-Miner each artificial ant represents a candidate classification rule of the form:

- IF <*term1* AND *term2* AND … > THEN < *class* >.

Each term is a triple <*attribute*, *operator*, *value*>, where *value* is one of the values belonging to the domain of *attribute*. An example of a term is: <*Sex* = *female*>. *Class* is the value of the goal attribute predicted by the rule for any case that satisfies all the terms of the rule antecedent. An example of a rule is:

- IF <*Salary* = *high*> AND <*Mortgage* = *No*> THEN <*Credit* = *good*>.

In the current version of Ant-Miner the *operator* is always "=", so that Ant-Miner can cope only with categorical (discrete) attributes. Continuous attributes would have to be discretized in a preprocessing step.

DESCRIPTION OF ANT-MINER

The pseudocode of Ant-Miner is described, at a very high level of abstraction, in Algorithm 1. Ant-Miner starts by initializing the training set to the set of all training cases, and initializing the discovered rule list to an empty list. Then it performs an outer loop where each iteration discovers a classification rule.

The first step of this outer loop is to initialize all trails with the same amount of pheromone, which means that all terms have the same probability of being chosen by an ant to incrementally construct a rule. This is done by an inner loop, consisting of three steps. First, an ant starts with an empty rule and incrementally constructs a classification rule by adding one term at a time to the current rule. In this step a $term_{ij}$ – representing a triple $<Attribute_i = Value_j>$ – is chosen to be added to the current rule with probability proportional to the product of $h_{ij} \times t_{ij}(t)$, where h_{ij} is the value of a problem-dependent heuristic function for $term_{ij}$ and $t_{ij}(t)$ is the amount of pheromone associated with $term_{ij}$ at iteration (time index) t. More precisely, h_{ij} is essentially the information gain associated with $term_{ij}$ – see Cover and Thomas (1991) for a comprehensive discussion on information gain. The higher the value of h_{ij} the more relevant for classification $term_{ij}$ is and so the higher its probability of being chosen. $t_{ij}(t)$ corresponds to the amount of pheromone currently available in the position i,j of the trail being followed by the current ant. The better the quality of the rule constructed by an ant, the higher the amount of pheromone added to the trail positions ("terms") visited ("used") by the ant. Therefore, as time goes by, the best trail positions to be followed – that is, the best terms to be added to a rule – will have greater and greater amounts of pheromone, increasing their probability of being chosen. Please refer to Algorithm 1.

The second step of the inner loop consists of pruning the just-constructed rule, that is, removing irrelevant terms – terms that do not improve the predictive accuracy of the rule. This is done by using a rule-quality measure, the same one used to update the pheromones of the trails, as defined later. In essence, a term is removed from a rule if this operation does not decrease the quality of the rule. This pruning process helps to avoid the overfitting of the discovered rule to the training set.

The third step of the inner loop consists of updating the pheromone of all trails by increasing the pheromone in the trail followed by the ant, proportionally to the quality of the rule. In other words, the higher the quality of the rule, the higher the increase in the pheromone of the terms occurring in the rule antecedent. The quality (Q) of a rule is measured by the equation:

$$Q = \text{Sensitivity} \times \text{Specificity},$$

where $Sensitivity = TP / (TP + FN)$ and $Specificity = TN / (TN + FP)$. The meaning of the acronyms TP, FN, TN and FP is as follows:

- TP = number of true positives, that is, the number of cases covered by the rule that have the class predicted by the rule;
- FN = number of false negatives, that is, the number of cases that are not covered by the rule but that have the class predicted by the rule;
- TN = number of true negatives, that is, the number of cases that are not covered by the rule and that do not have the class predicted by the rule; and
- FP = number of false positives, that is, the number of cases covered by the rule that have a class different from the class predicted by the rule.

See Lopes (1997) for a discussion about these variables and their use to estimate predictive accuracy.

The inner loop is performed until some stopping criterion(a) is(are) satisfied, for example, until a maximum number of candidate rules have been constructed.

Algorithm 1. High-level pseudocode of Ant-Miner

```
TrainingSet = {all training cases};
DiscoveredRuleList = [];   /* initialized with an empty list */
REPEAT
    Initialize all trails with the same amount of pheromone;
    REPEAT
        An ant incrementally constructs a classification rule;
        Prune the just-constructed rule;
        Update the pheromone of all trails;
    UNTIL (stopping criteria)
    Choose the best rule out of all rules constructed by all the ants;
    Add the chosen rule to DiscoveredRuleList;
    TrainingSet = TrainingSet – {cases correctly covered by the chosen rule};
UNTIL (stopping criteria)
```

Once the inner loop is over, the algorithm chooses the highest-quality rule out of all the rules constructed by all the ants in the inner loop, and then it adds the chosen rule to the discovered rule list. Next, the algorithm removes from the training set all the cases correctly covered by the rule, that is, all cases that satisfy the rule antecedent and have the same class as predicted by the rule consequent. Hence, the next iteration of the outer loop starts with a smaller training set, consisting only of cases that have not been correctly covered by any rule discovered in previous iterations. The outer loop is performed until some stopping criterion(a) is(are) satisfied, for example, until the number of uncovered cases is smaller than a user-specified threshold.

Hence, the output of Ant-Miner is the list of classification rules contained in the discovered rule list.

A SUMMARY OF COMPUTATIONAL RESULTS

We have performed computational experiments comparing Ant-Miner with two well-known rule induction algorithms, namely CN2 (Clark & Niblett, 1989) and C4.5 (Quinlan, 1993) in several public-domain data sets often used as a benchmark in the machine learning literature. More precisely, the data sets used in the experiment are Ljubljana breast cancer, Wisconsin breast cancer, tic-tac-toe, dermatology, hepatitis, and Cleveland heart disease. A detailed description of all these data sets is available online from: *http://www.ics.uci.edu/~mlearn/MLRepository.html*. C4.5 and CN2, as well as many other classification algorithms, are also described in Witten and Frank (2000). C4.5 and CN2 were chosen for comparison because they are, in general, two of the most used algorithms belonging to the rule induction paradigm, where discovered knowledge is expressed by IF-THEN rules. Hence, they are a natural choice to be compared with Ant-Miner, since Ant-Miner also discovers knowledge expressed by IF-THEN rules. The results of the experiments summarized here are described in detail in Parpinelli, Lopes and Freitas (2002a, 2002b).

The results showed that Ant-Miner is competitive with both C4.5 and CN2 concerning predictive accuracy on the test set used to measure the generalization ability of the discovered rules. More precisely, predictive accuracy was measured by the accuracy rate, that is, the ratio of the number of correctly classified test cases divided by the total number of test cases (correctly-classified plus wrongly-classified test cases), as usual in the classification literature (Witten & Frank, 2000). Ant-Miner obtained a considerably better accuracy rate than CN2 in the Ljubljana breast cancer and the dermatology data sets, and a considerably better accuracy rate than C4.5 in the hepatitis data set. However, both CN2 and C4.5 obtained a considerably better accuracy rate than Ant-Miner in the tic-tac-toe data set. In the other data sets the difference in the predictive accuracy of Ant-Miner and the other two algorithms was quite small.

However, concerning rule set simplicity (measured by the number of discovered rules and the number of terms per rule), Ant-Miner discovered rule sets much simpler (i.e., smaller) than the rule sets discovered by both C4.5 and CN2. This is an important advantage in the context of data mining, where discovered knowledge is supposed to be shown to a human user in order to support his/her decision making (Fayyad, Piatetsky-Shapiro & Smyth, 1996).

REFERENCES

Bonabeau, E., Dorigo, M., & Theraulaz, G. (1999). *Swarm intelligence: From natural to artificial systems.* Oxford University Press.

Clark, P., & Niblett, T. (1989). The CN2 rule induction algorithm. *Machine Learning, 3*(4), 261-283.

Cover, T.M., & Thomas, J.A. (1991). *Elements of information theory*. New York: Wiley.

Dorigo, M., Gambardella, L.M., Middendorf, M., & Stutzle, T. (2002). Guest editorial: Special section on ant colony optimization. *IEEE Transactions on Evolutionary Computation, 6*(4), 317-319.

Fayyad, U.M., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery: An overview. In U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth & R. Uthurusamy (Eds.), *Advances in knowledge discovery & data mining* (pp. 1-34). Cambridge, MA: MIT Press.

Freitas, A.A., & Johnson, C.G. (2003). Research cluster in swarm intelligence. *EPSRC Research Proposal GR/*S63274*/01 – Case for Support*. Computing Laboratory, University of Kent.

Lopes, H.S., Coutinho, M.S., & Lima, W.C. (1997). An evolutionary approach to simulate cognitive feedback learning in medical domain. In E. Sanchez, T. Shibata & L.A. Zadeh (Eds.), *Genetic algorithms and fuzzy logic systems* (pp. 193-207). Singapore: World Scientific.

Monmarché, N. (1999). On data clustering with artificial ants. In A.A. Freitas (Ed.), *Data mining with evolutionary algorithms*, *Research directions – Papers from the AAAI Workshop* (pp. 23-26). Menlo Park, CA: AAAI Press.

Parpinelli, R.S., Lopes, H.S., & Freitas, A.A. (2002a). Data mining with an ant colony optimization algorithm. *IEEE Transactions on Evolutionary Computation, Special Issue on Ant Colony Algorithms*, *6*(4), 321-332.

Parpinelli, R.S., Lopes, H.S., & Freitas, A.A. (2002b). An ant colony algorithm for classification rule discovery. In H. Abbass, R. Sarker & C. Newton (Eds.), *Data mining: A heuristic approach* (pp. 191-208). London: Idea Group Publishing.

Quinlan, J.R. (1993). *C4.5: Programs for machine learning*. San Mateo, CA: Morgan Kaufmann.

Witten, I.H., & Frank, E. (2000). *Data mining: Practical machine learning tools with Java implementations*. San Mateo, CA: Morgan Kaufmann.

KEY TERMS

Data Mining: An interdisciplinary research field, whose core is at the intersection of machine learning, statistics, and databases. We emphasize that the goal – unlike, for example, classical statistics – is to discover knowledge that is not only accurate, but also comprehensible for the user.

Overfitting: Term referring to the situation where the discovered rules fit too much to the training set peculiarities. Overfitting usually leads to a reduction of the predictive accuracy rate on the test cases.

Rule List: An ordered list of IF-THEN rules discovered by the algorithm during training. When the rules are applied to classify cases in the test set, they are applied in order. That is, a case is matched with each of the rules in the list in turn. The first rule whose antecedent (conditions in the IF part) matches the attribute values of the case is then used to classify the case; that is, the case is assigned the same class as the class predicted by the first matching rule found in the discovered rule list.

Testing Case: Each of the cases (records) of the test set.

Test Set: A set of cases unseen during the training phase of the algorithm and used to compute the predictive accuracy of the list of rules discovered during the training phase.

Training Case: Each of the cases (records) of the training set.

Training Set: A set of cases used by the algorithm to discover the classification rules. At each iteration of Ant-Miner only one rule is discovered. The training cases that are covered correctly by the discovered rule (i.e., cases satisfying the rule antecedent and having the class predicted by the rule consequent) are removed from the training set. This process is performed iteratively while the number of uncovered training cases is greater than a user-specified threshold.

Classroom Communication on a Different Blackboard

Jennifer Paige Nightingale
Duquesne University, USA

INTRODUCTION

Employing technology that students find enticing and will prepare them well for the future, may be easier than preparing the teachers who will integrate this technology into learning activities. Competition among educational institutions for students is increasing exponentially, which requires a greater sensitivity to student needs (Heckman & Guskey, 1998). Course management systems (CMSs) can build a framework for using technology as a technique of providing this responsiveness to students, predominantly in the form of communication. The aim is to ultimately create an environment conducive to teaching and learning via a technological community where students, faculty, adjunct instructors, staff, and administration can work together to achieve a common goal—produce quality graduates.

BACKGROUND

The Impact of Technology

According to some educational technology resources, technology clearly has had, and will continue to have, a major impact on education. Many institution segments are moving transactions and shared information into online communication channels, such as the Internet. The U.S. Department of Education explained that the development of the National Education Technology Plan is part of a long-range national strategy and guide for using technology effectively to improve student academic achievement—directly or through integration with other approaches to overall reform. The *No Child Left Behind Act* is raising standards for student achievement, giving students more information and more choices, requiring more accountability from schools, and funding education at record levels. The goals of the *No Child Left Behind Act* are as follows (Kozberg, 2002):

1. To improve student achievement using technology in elementary and secondary schools
2. To assist students in becoming technologically literate by the time they finish the eighth grade
3. To ensure that teachers are able to integrate technology into the curriculum to improve student achievement

Rod Paige, U.S. Secretary of Education, believes that "by harnessing technology, we can expand access to learning and close the achievement gap in America."

As stated in Vallone (2000):

Education is the fuel for the New Economy, and demand for online teaching and learning resources has already reached critical mass. Web-based and Web-enhanced courses are…already popular e-learning platforms in higher education today. According to Student Monitor, 90% of college students used the Internet last fall and 66% were connecting once a day or more often. According to Campus Computing, over 53% of college courses used e-mail last year, almost 39% used Internet resources, and almost 28% used Web pages for class materials and resources… (p. 1).

As education and technology continue to merge and evolve at rapid speed, institutions will find an enormous array of effective solutions to augment their educational offerings and build deeper relationships with current and prospective students, alumni, and administrators (Blackboard® Incorporated, 2001). CMSs are modifying the way instructors disseminate information. A growing number of campuses identify CMSs as "very important" in their institutional information technology planning, and approximately one-fifth (20.6%) of all college courses now use course management tools, up from 14.7% in 2000 (Green, 2001). A leading e-learning industry analyst firm projects that the higher education e-learning market will grow from $4 billion today to $11 billion by 2003 (Stokes, 2000).

Choosing the Technology

Tomei (2002, p. 5) explained that to succeed in technology implementation, we need to understand that the technology itself is not the goal. In deciding what technology will support the objectives or if a CMS can satisfy determined needs, consider the following:

- The benefit of a CMS to faculty and an institution is that it should provide instructors with a pre-made course Web template on which they can easily place, especially if they possess the computing skills and knowledge identified above, their course materials and activities.
- A CMS should service the majority of instructors and their curriculum needs and yet be flexible enough to provide advanced users with the basics upon which they can add further functionality and meet the needs of special and unique requests.
- A CMS should be cost-effective for the institution in terms of operating and support costs and resources.
- Not only are developer skills and understanding important, an institution must commit to effective infrastructure and strategic support of the CMS by ensuring that adequate support for users exists and that resources are available for hardware and software improvements and maintenance necessary for reliable operation of the CMS.

Currently, Blackboard has been able to meet the above criteria, and as its increased use by faculty and programs across campus indicates, it has become the preferred CMS at Duquesne University (see Table 1). The fourth column indicates the growing trend of professors who use Blackboard over the previous semester. Furthermore, when SOBA-Net was born, less than 12% of the School of Business faculty was using Blackboard to supplement their face-to-face courses. Now we have 88.4% of faculty participation. Regardless of which technology you choose, you must realize the implementation is an ongoing process that demands time, attention, and dedication.

METHODOLOGY

The History of SOBA-Net

The Duquesne University School of Business (SOBA) in Pittsburgh, Pennsylvania, is bridging the communication gap between students, faculty, and staff with an online community called **SOBA-Net** (School of Business Administration Network). Using the CMS, Blackboard®, as a portal, users find important information, communication, and resources for their programs. SOBA-Net is quickly becoming a tool on which students, faculty, and staff depend.

There is no doubt that technology has had a major impact on our daily lives. In summer 1999, Duquesne University adopted the Blackboard® CMS as a means to provide technological advancements to education. This system traditionally allows for teaching online, where colleges and universities all over the world are diving into the online education arena. For example, an instructor can use a CMS course site simply and easily for increasing the availability of course materials, assignments, resources, and grades. In an effort to meet student demands, this CMS was modified to build human relationships that allow students to take an active role in, and be held accountable for, their academic careers. Ultimately, the creation of SOBA-Net has taken the foundation of a CMS and altered it to become a school management system (SMS).

What is a School Management System and Portal?

A management information system is, according to Ox (2002), a computer-based system that supports managers and their subordinates in their daily work and strategic planning….managerial decisions and daily operations. An SMS manages a school's key functional data, including items such as staff and faculty data, class schedules, school guidelines, special programs, and other school information, and, depending upon the type of SMS implemented and the desired outcomes, the system can range from low-end to high-end capabilities. Many schools depend on integrated information systems to manage their data and provide detailed information that aids in the communication process (Blackboard® Incorporated, 2001).

A portal is not a new concept to the technology industry; in fact, www.yahoo.com and www.msn.com are two of the many organizations that have used portals from the beginning. Building an online community in Blackboard® has made it convenient for students to find the tools and resources they need to be successful in their academic careers. "In the Internet economy, people, process, and technology must be inextricably linked," explained Greg Baroni (1999, [Online]), national partner-in-charge of the KPMG higher education practice. "Portals

Table 1. Statistics for nondistance learning courses

Term	Total Courses	Unique Faculty	Increase
Fall 2003	255	142	15%
Spring 2003	221	121	12%
Fall 2002	197	109	

are the latest iteration of enabling technology that universities can integrate to create strategic advantage."

Why Blackboard®?

Blackboard® makes it relatively easy to provide students with course materials in HTML, PDF, Word, PowerPoint, or Excel formats. It provides student home page creation, built-in chat and online discussion capabilities, e-mail, collaborative or group work facilitation and management areas, assessment creation and automatic grading, grade book distribution, class management features (such as calendaring or tasks), and a digital drop box for students to use to submit assignments. Course sites are password protected and are available only to registered students.

Although DU supports both WebCT® and Blackboard® CMSs, Blackboard® was selected because of its low learning curve, visual appeal, and secured environment, as well as the ability to access information via the Internet. In addition, some of the school's faculty participated in training when Blackboard® was initially implemented campus-wide in summer 1999, bringing a familiarity component into play and encouraging faculty involvement in implementation.

SOBA-Net reinforces both Duquesne University's and SOBA's directive of integrating technology across the curriculum. For many faculty members, the shear thought of integration can be greatly intimidating. By using Blackboard®, faculty members can observe how a CMS is used to aid in disseminating materials and communicating with students; in moving from observing to participating, the intimidation can be significantly reduced.

What were the Initial Objectives?

It was first thought that academic advisors would benefit most, as they constantly searched for a better method of communicating with the students and distributing information relevant to their academic careers. Thus, they embraced and encouraged this improvement and rather than merely adding to the student's knowledge bank, the objective was to set students up for success by requiring them to take an active role in their education. Students and advisors would be able to work smarter and more efficiently, help improve overall communication between faculty, staff, and students, validate concepts taught throughout the curriculum (i.e., teamwork), and greatly reduce the paper trail.

Getting Started

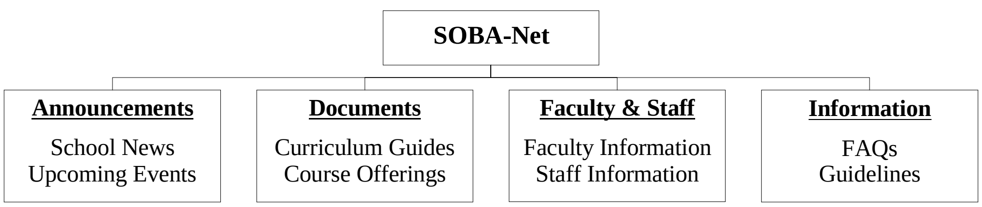
Before documents and information are uploaded, specific objectives and guidelines must be established. White and Weight (2000, p. 185) have used technology as a tool to enable human relationships, and have found that people are the real reason for success in any online learning environment. This environment requires a value-added approach, where technology is used to enhance the complex private and social activities of the learning process (p. vii). In this stage, three factors must be considered:

- Content management: What should be online? Who will provide it? Who will maintain and update it? How will it be organized? Is security an issue? What areas of the CMS should be enabled or disabled?
- Interactivity: What limitations will be imposed? Do all users have the ability to participate in threaded discussions? Should specific guidelines and standards be established and posted?
- User management: Who will manage the users and passwords (if applicable)? Should different security settings exist for different types or classifications of users (e.g., faculty, staff, students, etc.)?

Building the Virtual Community

Once the guidelines are established, the construction can begin. This is the longest part of the process, and obtaining the information can be a challenge. Consider holding CMS training sessions to involve faculty and staff—your colleagues add valuable feedback and offer assistance to feel like a valuable part of the process. After all, the virtual community encompasses the entire school, so demonstrate the need for and value of teamwork from

Figure 1. Snapshot of SOBA-Net's schema



its inception. In addition, draw an initial schema before uploading any content (see Figure 1). It is critical that the content be organized and easy to navigate, or users will not return.

If You Build It, They Will Come

This is not exactly the case. The overall marketing and enthusiasm generated will determine its success, and like the implementation phase, marketing SOBA-Net is an ongoing process that demands time, attention, and dedication. To generate interest and maintain momentum, remember that behavior breeds behavior—if you are excited about it, then others will follow suit. Users must have a reason to log in, so offer incentives or hold contests to create initial excitement. The results indicate that both the faculty and academic advisors proved to be a valuable communication medium and significant marketing resource.

Where are We Now?

At the time of SOBA-Net's initial construction, we did not have the skills or technology needed to develop such a site and host it on our Web server. Its success, along with improved technical resources and knowledge during the past year, has enabled us to centralize the information. Sections now exist for both prospective and current students as well as faculty, staff, administration, alumni, and friends (http://www.business.duq.edu).

Regardless of the technology used, it is important to understand how the fundamental concepts, through the construction and use of such a portal, will aid in building relationships and enhancing communication skills.

CONCLUSION

SOBA-Net provides connectivity via the Internet, preserves the School of Business, Duquesne University, meets President George W. Bush's initiative for technology integration and enhancement, and appeals to the students' desire for online resources.

While it may be too early to determine SOBA-Net's effectiveness in fostering academic success, it has encouraged students to become more self-sufficient and assume some responsibility and accountability for their educational outcomes. In addition, there has been a notable increase in faculty use of Blackboard® in their classes, and students are encouraging those who do not use it to adopt it. The information superhighway transports teachers and students beyond the walls of their classrooms by providing access to an immeasurable quantity of information (Tripathi, 2000). As educators, we must utilize the tools at hand to foster students' communication skills and accountability. The School of Business's experience has had a positive affect not only school-wide, but also throughout campus, and, as a result, several other programs are now developing similar sites for their faculty, staff, students, and administration.

REFERENCES

Baroni, G. (1999, October). *Portal technology—Into the looking glass.* Paper presented at Portal Technology Symposium in Baltimore, Maryland. Retrieved from *http://www.convergemag.com/SpecialPubs/Portal/creating.shtm*

Blackboard Incorporated. (2001). Course and portal solutions: Blackboard 5. Retrieved from *http://products.blackboard.com/cp/bb5/index.cgi*

Bober, M. (2001). School information systems and their effects on school operations and culture. *Journal of Research on Technology in Education, 33*(5).

The context of postsecondary education, *The Condition of Education 2001.* In *U.S. Department of Education, National Center for Education Statistics: Vol. NCES 2001-072.* Retrieved September, 2003, from *nces.ed.gov: http://nces.ed.gov/pubs2001/2001072.pdf*

Digest of Education Statistics 2002. (2002). In *U.S. Department of Education, National Center for Education Statistics.* Retrieved from *nces.ed.gov: http://nces.ed.gov/pubs2003/digest02/ ch_3.asp*

Green, K. (2001, October 29). *2001 national survey of information technology in U.S. higher education.* Paper presented at Educause 2001 in Indianapolis, Indiana.

Greene, B., Cattagni, A., & Westat, E. F. (2000). Internet access in U.S. public schools and classrooms: 1994–2000. In *U.S. Department of Education, National Center for Education Statistics: Vol. NCES 2000-071.* Retrieved August, 2003, from nces.ed.gov: *http://nces.ed.gov/pubs2000/2000071.pdf*

Heckman, R., & Guskey, A. (1998). The relationship between alumni and university: Toward a theory of discretionary collaborative behavior. *Journal of Marketing Theory and Practice, 6*(2), 97–113.

Kozberg, L. (2002, January 8). Paige joins President Bush for signing of historic No Child Left Behind Act of 2001: Bold new Federal Law to improve student achievement. Retrieved August, 2003, from www.ed.gov: *http://*

www.ed.gov/news/pressreleases/2002/01/01082002.html

Oz, E. (2002). *Management information systems* (3rd ed.) MA: Course Technology.

Rogoff, B. (1994). Developing understanding of the idea of communities of learners. *Mind, Culture, and Activity, 1*(4), 209–229.

Stokes, P. (2000, June). E-Learning: Education businesses transform schooling (chap.) Retrieved August, 2001, from www.Eduventures.com: *https://www.eduventures.com/pdf/doe_elearning.pdf*

Teacher Quality: A Report on the Preparation and Qualifications of Public School Teachers (1999). In *U.S. Department of Education, National Center for Education Statistics: Vol. NCES 1999–080*. Retrieved September, 2003, from *nces.ed.gov: http://nces.ed.gov/pubs1999/1999080.pdf*

Tomei, L. A. (2002). *The technology façade: Overcoming barriers to effective instructional technology*. Reading, MA: Allyn & Bacon.

Tripathi, A. K. (1998). The Internet in education. Retrieved October, 2003, from www.techLEARNING.com: *http://www.techlearning.com/db_area/archives/WCE/archives/tripathi.htm*

Vallone, C. (2000, May 9). *On the Internet, distance learning and the future of the research university*. Presented to the Subcommittee on Basic Research, Committee on Science, U.S. House of Representatives. Retrieved October 26, 2003, from www.house.gov: *http://www.house.gov/science/vallone_050900.htm*

KEY TERMS

Computer-Based Information System (CBIS): A computer-based information system is an information system that uses computer technology to perform input, processing, and output activities. A CBIS consists of people, procedures, data, hardware, and software.

Groupware: This software can be used by a group of people who are working on the same information at separate workstations. It includes features such as e-mail, shared calendar, task list, and address book, and is used for project management, brainstorming, and action planning.

Hypertext Markup Language (HTML): A markup language used to structure text and multimedia objects as well as set up links between documents. This language is used extensively on the Internet and can be viewed in a Web browser.

Internet: A worldwide system of networks that has transformed communications and methods of commerce by allowing various computer networks to interconnect. Sometimes referred to as a "network of networks," the Internet materialized in the United States in the 1970s but did not become overtly visible until the early 1990s.

Portable Document Format (PDF): According to Adobe Systems Incorporated, a PDF is a "universal file format that preserves the fonts, images, graphics, and layout of any source document, regardless of the application and platform used to create it."

Portal: A portal is any Web site considered an entry point to other Web sites, often by being or providing access to a search engine. Some examples include www.Yahoo.com, www.Google.com, and www.MSN.com.

Push Technology: Push technology enables people to subscribe to and have news and information delivered directly to their e-mail, releasing them from having to search the Internet.

Schema: A diagram, outline, or model.

Client/Server and the Knowledge Directory

Stuart D. Galup
Florida Atlantic University, USA

Ronald Dattero
Southwest Missouri State University, USA

Richard C. Hicks
Texas A&M University, USA

INTRODUCTION

Data, information, and knowledge are three related but not interchangeable concepts. Data are a set of discrete, objective facts about events. Information is organized data presented in context. Data become information when their creator adds meaning or value. Similarly, knowledge derives from information as information derives from data. Knowledge can be viewed as information in context, together with an understanding of how to use it. Knowledge can be either explicit (knowledge for which a person is able to make available for inspection) or tacit (knowledge for which a person is unable to make available for inspection) (Brooking, 1999; Davenpot & Prusak, 1998). In Table 1, a list of knowledge that is particularly critical for business organizations is given.

Knowledge is an intellectual property that although paid for in part by the employer is a difficult asset to control, as it is fragmented in documents, policies, procedures, and other data storage mediums. Another challenge for management is to retain this knowledge in a form that is easily retrievable. This is not an easy task, since the enterprise must first identify the location of all needed knowledge, and second, determine the easiest way to retrieve it.

There are many definitions of knowledge management, but the Gartner Group (1999) description seems most appropriate for the perspective expressed in our article: "Knowledge management promotes an integrated approach to identifying, capturing, retrieving, sharing, and evaluating an enterprise's information assets. These information assets may include databases, documents, policies and procedures, as well as the un-captured tacit expertise and experience stored in individual workers' heads."

This definition implies that information assets are plentiful and are stored in numerous locations throughout the organization. Storage options include documents, documents in document management systems, groupware such as Lotus Notes, and expert or knowledge-based systems (Brooking, 1999). Physically these information assets can be electronically stored on compact disk, laser disk, mechanical hard drives, microfilm, microfiche, and embedded in computer programs. Further, information assets are also stored in books, documents, and other paper-based medium.

BACKGROUND

In a world of multiple computer languages, database management systems, assorted collaborative and group support software, network technologies, and data storage methods, it can be a difficult and complex problem to

Table 1. Knowledge that is particularly critical (Brooking, 1999)

- Knowledge of a particular job
- Knowledge of who knows what in a company
- Knowledge of how to get things done in a company using the corporate culture
- Knowledge of who is best to perform a particular job or task
- Knowledge of corporate history (how and why)
- Knowledge of business customs
- Knowledge of a particular customer account
- Knowledge of how to put together a team that can achieve a particular task
- Knowledge of how to approach a particular problem which is difficult to solve

locate and retrieve enterprise knowledge. If KM promotes an integrated approach to identifying, capturing, retrieving, sharing, and evaluating an enterprise's information assets, then the challenge is to create a knowledge management system in order to get the right information to the right person at the right time.

"An integrated and integrative technology architecture is a key driver for Knowledge Management Systems (KMS) … KMS seem to require a variety of technologies: database and database management, communication and messaging, and browsing and retrieval. The need for seamless integration of the various technologies may lead to the dominance of the Internet and Internet-based KMS architectures" (Alavi & Leidner, 1999). Alavi and Leidner (1999) also note that "organizational intranets will also play a dominant role in support of internal knowledge management activities due to cost-effective technical capabilities including: access to the legacy systems, platform independence, access to multimedia data formats, a uniform and easy-to-use point-and-click interface, and capability for easy multi-media publication for knowledge sharing."

A CLIENT/SERVER ARCHITECTURE FOR KMS

Computing sources for the first 30 years of the information technology revolution were dominated by isolated hardware and network environments. Mainframes, mini-computers and local area networks were initially set up to support specific business functions. Each computing complex was installed with a separate physical data network. IBM mainframes used coaxial cable and 3270 terminal emulation and the IBM System 38 mini-computer used twin-axial cable and 5250 terminal emulation. Local area networks used their own respective cabling medium and data network architecture. As a result, these environments were isolated and data sharing was almost impossible (Kern et al., 1998).

Information systems written for these monolithic computer complexes contain three basic components: a presentation layer, a processing layer, and a data storage layer (Boar, 1996; Borthick & Roth, 1994). All three layers execute on one hardware platform. During the 1980s and 1990s, multiple protocol support between different platforms across inexpensive connections became more common. This connectivity enhancement helped the development of client/server technologies, which distributed these layers across hardware and operating systems platforms (Boar, 1996; Duchessi & Chengalur-Smith, 1998; Schulte, 1995). A client/server architecture consists of three layers: the presentation layer, the business logic layer, and the data layer. A two-tier client/server architecture places the presentation layer on the client and data layer on the server. The business layer may reside on either the client or server, or both. A three-tier client/server architecture places the presentation layer on the client, the business logic layer on a middle tier, and the data layer on a server. Although there are many variation of the client/server model, two-tier and three-tier are the two basic deployments (Edwards, 1999; Weston, 1998).

The three-tier client/server architecture has many advantages over the two-tier client/server architecture, including less complexity, higher security, higher encapsulation of data, better performance efficiency, excellent scalability, excellent application reuse, a good server-to-server infrastructure (via server-side middleware), legacy application integration (via gateways), excellent internal support, support for heterogeneous databases, rich communication choices, excellent hardware architecture flexibility, and excellent availability. The main disadvantage is the difficulties in development, but these difficulties are getting less and less over time (Orfali et al., 1999). One way around this development problem is to adopt a component-based architecture design.

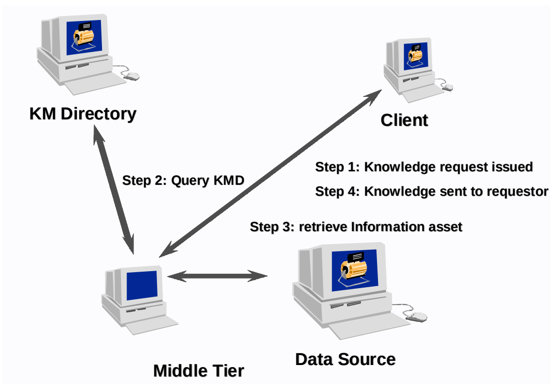
Employing a three-tier client/server architecture would provide a good flexible architecture for KMS. The structure would be very similar to the three-tier client/server architecture detailed by Orfali et al. (1999), except it would also have a knowledge directory (covered in the next section). The client layer would have at least the following: a nice GUI, a Web browser, the client operating system, and any required client-side applications for KM (such as Lotus Notes). The middle layer would contain a network operating system, and transport stack (such as TCP/IP) and service specific middleware for: databases (such as ODBC, JDBC, and SQLJ), internet/intranets (such as HTTP and CGI), e-mail (such as SMTP and POP3), storing and accessing multimedia documents, coordinating group conferencing, linking individuals in group scheduling, and workflow processes. The server layer would contain the server operating system and specific server-based applications such as database management systems, document management systems, server side groupware, and so forth.

KNOWLEDGE MANAGEMENT DIRECTORY

Achieving the connection to numerous data and information sources is a serious challenge for the development of KMS. Organizational knowledge management strives to cultivate, nurture and exploit knowledge at different levels and in different contexts throughout the organization (Handzic & Bewsell, 2001). The demands placed on the

Figure 1. Knowledge management directory process (Galup & Dattero, 2000, 2002)

Step 1: The Client issues an inquiry for knowledge.
Step 2: An inquiry is sent to the Knowledge Management Directory and a summary listing similar to an Internet search is presented to the Client. The KMD presents the Client with an inventory of the KM assets organized by context.
Step 3: When a KM asset is selected, the KMD launches the appropriate software and retrieves the knowledge.
Step 4: The knowledge is sent to the requester.



information technology infrastructure to support an integrated approach to identifying, capturing, retrieving, sharing, and evaluating an enterprise's information assets is a major undertaking; especially since these information assets may include databases, documents, policies and procedures stored in electronic forms.

One of the major requirements of KM is the ability to locate the knowledge and present it to the knowledge requester. The need for a directory of information assets is imperative. A facility, such as the Knowledge Management Directory (KMD), supports the central task of organizational knowledge management and knowledge mapping (Galup, Dattero, & Hicks, 2003). Knowledge mapping provides a continuously evolving organization memory by capturing and integrating knowledge sources and communicating knowledge in an understandable way (Wexler, 2001).

A knowledge directory (some sort of list or picture) points to knowledge (in people, documents, or databases) but does not contain it. The KMD resolves the client request to connect to an information asset by translating the keyword into a data source location, thereby permitting the customer to navigate the numerous information asset resources and locate the best match.

The KMD categorizes the context of the knowledge and provides the computational tools to provide the user with access to the requested knowledge. To use the KMD, a user selects a topic from the user interface screen. When the topic is selected, a list of all resources of each type (expert/practice, decision support systems/warehouse, database, expert, and document) is displayed. The user then selects and evaluates sources until the knowledge is acquired. When multiple knowledge sources exist for a topic, the user will have to choose which knowledge source is most accurate and best suited to his or her needs. The knowledge seeker should be influenced to select the most complete and accurate knowledge available. In the KMD design, this is accomplished by ordering the knowledge sources by their inherent verification characteristics. The most accurate sources are placed at the top of the display, implying their desirability to the user. Expert systems and best practices represent complete, accurate solutions to specific problems, so they are the most desirable knowledge source and are ordered first. Data warehouses and decision support systems applications are the next highest in accuracy, followed by experts and documents (Hicks, Dattero & Galup, 2002, 2003).

CONCLUSION

Our three-tier client/server architecture for KMS has traits of the network model for KMS (Alavi & Leidner, 1999; Bowman, 2002). A network KMS "… focuses on building knowledge directories. This involves defining the knowledge categories that are relevant to the organization, identifying knowledge owners for each category, and creating a searchable directory to help others in the organization identify and locate knowledge owners" (Bowman, 2002). Our three-tier client/server architecture for KMS also has traits of the repository model for KMS (Alavi & Leidner, 1999; Bowman, 2002), because our KMS architecture retrieves explicit organizational knowledge. Therefore, our three-tier client/server architecture for KMS has the traits of the two major models for KMS. Further, it provides a seamless integration of the variety of technologies required for a KMS such as database management, document management, groupware, and e-mail.

REFERENCES

Alavi, M., & Leidner, D. (1999). Knowledge management systems: Issues, challenges, and benefits. *Communications of the Association for Information Systems, 1*(7).

Boar, B. (1996). *Cost-effective strategies for client/server systems.* John Wiley & Sons, Inc.

Borthick, A.F., & Roth, H.P. (1994). Understanding client/server computing. *Management Accounting, 76*(2), 36-41.

Bowman, B.J. (2002). Building knowledge management systems. *Information Systems Management, 19*(3), 32-40.

Brooking, A. (1999). *Corporate memory.* International Thomson Business Press.

Davenport, T., & Prusak, L. (1998). *Working knowledge.* Harvard Business School Press.

Duchessi, P., & Chengalur-Smith, I. (1998). Client/server benefits, problems, best practices. *Communications of the ACM, 41*(5), 87-94.

Edwards, I. (1999). *3-tier client/server at work* (rev. ed.). Wiley.

Galup, S., & Dattero, R. (2000). Client/server and the knowledge directory: A natural relationship. *Proceedings of the Information Resource Management Association,* Anchorage, Alaska.

Galup, S., & Dattero, R. (2002). Client/server and the knowledge directory: A natural relationship. In D. White (Ed.), *Knowledge mapping and management* (pp. 187-194). Hershey, PA: IRM Press.

Galup, S., Dattero, R., & Hicks, R. (2003). Enterprise knowledge dictionary. *Knowledge Management Research & Practice, 1*(2), 95-101.

Gartner Group. (1999). *White papers on knowledge management.*

Handzic, M., & Bewsell, G. (2003, May 18-21). Corporate memories: Tombs or wellsprings of knowledge. *Proceedings of the Information Resources Management Association International Conference (IRMA 2003),* Philadelphia (pp. 171-173).

Hicks, R., Dattero, R., & Galup, S. (2002/2003). A verification-based conflict resolution strategy for knowledge management systems. *Journal of Computer Information Systems, 43*(2), 36-41.

Kern, H., Johnson, R., Galup, S., & Horgan, D. (1998). *Building the new enterprise: People, processes, and technology.* CA: Sun Microsystems Press – Prentice Hall.

Orfali, R., Harkey, D., & Edwards, J. (1999). *Client/server survival guide* (3rd ed.). Wiley.

Schulte, R. (1995). Middleware and three-tier architectures. *GartnerGroup Symposium/Itxpo 95: Managing the revolution.*

Weston, R. (1998). Client/server evolution: Thin clients, fat servers. *Computerworld, 32*(6), 20.

Wexler, M. (2001). The who, what and why of knowledge mapping. *Journal of Knowledge Management, 5,* 249-263.

KEY TERMS

Client/Server: Client/server technologies distributed these layers across hardware and operating systems platforms.

Knowledge: Knowledge is an intellectual property that although paid for in part by the employer is a difficult asset to control as it is fragmented in documents, policies, procedures, and other data storage mediums. Another challenge for management is to retain this knowledge in a form that is easily retrievable. This is not an easy task, since the enterprise must first identify the location of all needed knowledge, and second, determine the easiest way to retrieve it.

Knowledge Management: Knowledge management promotes an integrated approach to identifying, capturing, retrieving, sharing, and evaluating an enterprise's information assets. These information assets may include databases, documents, policies and procedures, as well as the un-captured tacit expertise and experience stored in individual workers' heads.

Knowledge Management Directory: A knowledge directory points to knowledge (in people, documents, or databases) but does not contain it.

Knowledge Management System: An integrated and integrative technology architecture consisting of a variety of technologies: database and database management, communication and messaging, and browsing and retrieval, that seamlessly integrates knowledge sources in an organization.

Three-Tier Client/Server: Three-tier splits the logic between 1) clients that run the graphical user interface (GUI) logic, 2) the application server running the business logic, and 3) the database and/or legacy application.

Two-Tier Client/Server: Two-tier splits the processing loading in two. The majority of the applications run on the client (often referred to as "fat client"), which typically sends SQL requests to a server-resident database.

Cluster Analysis Using Rough Clustering and k-Means Clustering

C

Kevin E. Voges
University of Canterbury, New Zealand

INTRODUCTION

Cluster analysis is a fundamental data reduction technique used in the physical and social sciences. The technique is of interest to managers in information science because of its potential use in identifying user needs though segmenting users such as Web site visitors. In addition, the theory of rough sets is the subject of intense interest in computational intelligence research. The extension of this theory into rough clustering provides an important and potentially useful addition to the range of cluster analysis techniques available to the manager.

Cluster analysis is defined as the grouping of "individuals or objects into clusters so that objects in the same cluster are more similar to one another than they are to objects in other clusters" (Hair, Anderson, Tatham & Black, 1998, p. 470). There are a number of comprehensive introductions to cluster analysis (Arabie, Hubert & De Soete, 1994; Cramer, 2003; Everitt, Landau & Leese, 2001). Techniques are often classified as hierarchical or nonhierarchical (Hair et al., 1998), and the most commonly used nonhierarchical technique is the *k*-means approach developed by MacQueen (1967). Recently, techniques based on developments in computational intelligence have also been used as clustering algorithms. For example, the theory of fuzzy sets developed by Zadeh (1965), which introduced the concept of partial set membership, has been applied to clustering (Dumitrescu, Lazzerini & Jain, 2000). Another technique receiving considerable attention is the theory of rough sets (Pawlak, 1982), which has led to clustering algorithms referred to as rough clustering (do Prado, Engel & Filho, 2002; Voges, Pope & Brown, 2002).

This article provides brief introductions to *k*-means cluster analysis, rough sets theory, and rough clustering, and compares *k*-means clustering and rough clustering. The article shows that rough clustering provides a more flexible solution to the clustering problem, and can be conceptualized as extracting *concepts* from the data, rather than strictly delineated subgroupings (Pawlak, 1991). Traditional clustering methods generate *extensional* descriptions of groups (i.e., which objects are members of each cluster), whereas clustering techniques based on rough sets theory generate *intensional* descriptions (i.e., what are the main characteristics of each cluster) (do Prado et al., 2002). These different goals suggest that both *k*-means clustering and rough clustering have their place in the data analyst's and the information manager's toolbox.

BACKGROUND

k-Means Cluster Analysis

In the *k*-means approach, the number of clusters (*k*) in each partition of the data set is decided *prior to* the analysis, and data points are randomly selected as the initial estimates of the cluster centres (referred to as centroids). The remaining data points are assigned to the closest centroid on the basis of the distance between them, usually using a Euclidean distance measure. The aim is to obtain maximal homogeneity within clusters (i.e., members of the same cluster are most similar to each other), and maximal heterogeneity between clusters (i.e., members of different clusters are most dissimilar to each other).

K-means cluster analysis has been shown to be quite robust (Punj & Stewart, 1983). Despite this, the approach suffers from many of the problems associated with all traditional multivariate statistical analysis methods. These methods were developed for use with variables that are normally distributed and that have an equal variance-covariance matrix in all groups. In most realistic data sets, neither of these conditions necessarily holds.

Rough Sets

The concept of rough sets (also known as approximation sets) was introduced by Pawlak (1982, 1991), and is based on the assumption that with every record in the information system (the data matrix in traditional data analysis terms), there is associated a certain amount of information. This information is expressed by means of attributes (variables in traditional data analysis terms), used as descriptions of the objects. For example, objects could be individual users in a study of user needs, and attributes could be characteristics of the users such as gender, level of experience, age, or other characteristics considered relevant. See Pawlak (1991) or Munakata (1998) for comprehensive introductions.

In rough set theory, the data matrix is represented as a table, the information system. The complete information system expresses all the knowledge available about the objects being studied. More formally, the information system is a pair, $S = (U, A)$, where U is a non-empty finite set of objects called the universe and $A = \{a1, \ldots, aj\}$ is a non-empty finite set of attributes describing the objects in U. With every attribute $a \in A$ we associate a set Va such that $a : U \rightarrow Va$. The set Va is called the domain or value set of a. In traditional data analysis terms, these are the values that each variable can take (e.g., gender can be male or female; users can have varying levels of experience).

A core concept of rough sets is that of indiscernibility. Two objects in the information system about which we have the same knowledge are indiscernible. Let $S = (U, A)$ be an information system; then with any subset of attributes B, ($B \subseteq A$), there is associated an equivalence relation, $IND_A(B)$, called the B-indiscernibility relation. It is defined as:

$$IND_A(B) = \{ (x, x') \in U^2 \mid \forall\ a \in B\ \ a(x) = a(x') \}$$

In other words, for any two objects (x and x') being considered from the complete data set, if any attribute a, from the subset of attributes B, is the same for both objects, they are indiscernible (on that attribute). If $(x, x') \in IND_A(B)$, then the objects x and x' are indiscernible from each other when considering the subset B of attributes.

Equivalence relations lead to the universe being divided into partitions, which can then be used to build new subsets of the universe. Two of these subsets of particular use in rough sets theory are the lower approximation and the upper approximation. Let $S = (U, A)$ be an information system, and let $B \subseteq A$ and $X \subseteq U$. We can describe the set X using only the information contained in the attribute values from B by constructing the B-lower and B-upper approximations of X, denoted $B_*(X)$ and $B^*(X)$ respectively, where:

$$B_*(X) = \{ x \mid [x]_B \subseteq X \},\ \text{and } B^*(X) = \{ x \mid [x]_B \cap X \neq \emptyset \}$$

The set $BN_B(X)$ is referred to as the boundary region of X, and is defined as the difference between the upper approximation and the lower approximation. That is:

$$BN_B(X) = B^*(X) - B_*(X)$$

If the boundary region of X is the empty set, then X is a crisp (exact) set with respect to B. If the boundary region is not empty, X is referred to as a rough (inexact) set with respect to B. The important insight of Pawlak's work is his definition of a set in terms of these two sets, the lower approximation and the upper approximation. This extends the standard definition of a set in a fundamentally important way.

Rough Clustering

Rough clusters are a simple extension of the notion of rough sets. The value set (Va) is ordered, which allows a measure of the distance between each object to be defined, and clusters of objects are then formed on the basis of their distance from each other. An object can belong to more than one cluster. Clusters can then be defined by a lower approximation (objects exclusive to that cluster) and an upper approximation (all objects in the cluster which are also members of other clusters), in a similar manner to rough sets.

Let $S = (U, A)$ be an information system, where U is a non-empty finite set of M objects $(1 \leq i \leq M)$, and A is a non-empty finite set of N attributes $(1 \leq j \leq N)$ on U. The j^{th} attribute of the i^{th} object has value $R(i, j)$ drawn from the ordered value set Va .

For any pair of objects, p and q, the distance between the objects is defined as:

$$D(p,q) = \sum_{j=1}^{N} | R(p,j) - R(q,j) |$$

That is, the absolute differences between the values for each object pair's attributes are summed. The distance measure ranges from 0 (indicating indiscernible objects) to a maximum determined by the number of attributes and the size of the value set for each attribute.

One algorithm for producing rough clusters is as follows (Voges et al., 2002). Initially, a distance matrix for all paired object comparisons is calculated. All object pairs at interobject distance D, where D steps from 0 to a determined maximum, are identified. Each object pair (a_i, a_j) can be in one of three situations in relation to current cluster membership, with the following consequences:

1. Both objects have not been assigned to any prior cluster. A new cluster is started with a_i and a_j as the first members.
2. Both objects are currently assigned to clusters. Object a_i is assigned to object a_j's earliest cluster, and object a_j is assigned to object a_i's earliest cluster. The earliest cluster is the first cluster the object was assigned to.
3. One object, a_i, is assigned to a cluster and the other object, a_j, is not assigned a cluster. Object a_j is assigned to object a_i's earliest cluster.

Voges and Pope (2004) have developed a new technique for generating rough cluster descriptions using an evolutionary algorithm.

FUTURE TRENDS

The theory of rough sets continues to generate numerous edited books and conferences extending Pawlak's original insight into new areas of application and theory (e.g., Lin & Cercone, 1997; Polkowski & Skowron, 1998; Polkowski, Tsumoto & Lin, 2000; Wang, Liu, Yao & Skowron, 2003; Zhong, Skowron & Ohsuga, 1999).

CONCLUSION

In rough clustering an object can belong to more than one cluster and therefore necessarily produces different solutions to *k*-means clustering, where an object can belong to only one cluster. This section briefly outlines a comparison between rough clustering and *k*-means clustering. A more detailed comparison can be found in Voges et al. (2002).

In business, one of the most common applications of cluster analysis is the segmentation of a market by identifying homogeneous groups of buyers (Punj & Stewart, 1983; Wedel & Kamakura, 2000). Segmentation can also be applied to groups of users to assist in identifying user needs. In a market segmentation study that compared *k*-means and rough clustering analyses of shopping orientations of Internet shoppers, Voges et al. (2002) found that the two clustering techniques resulted in some clusters that were identified by both techniques, and some clusters that were unique to the particular technique used. The rough clustering technique also found clusters that were "refined" sub-clusters of those found by *k*-means clustering, and which identified a more specific sub-segment of the market.

Rough clustering produces more clusters than *k*-means clustering (Voges et al., 2002), with the number of clusters required to describe the data dependent on the interobject distance (D). It was found that the lower approximation of each cluster was dependent on the number of clusters selected for the solution. More clusters means an object has a higher chance of being in more than one cluster, which moves the object from the lower approximation to the boundary region and reduces the size of the lower approximation. This suggested that a number of factors needed to be considered when determining the best maximum value for D and the best number of clusters to include in the solution. A solution with too few clusters does not provide a useful interpretation of the partitioning of the data. On the other hand, too many clusters make interpretation difficult. In addition, the degree of overlap between the clusters needs to be minimised to ensure that each cluster provided information to aid in interpretation. Determining a good rough cluster solution requires a trade-off between these factors.

REFERENCES

Arabie, P., Hubert, L., & De Soete, G. (Eds.). (1994). *Clustering and classification*. River Edge, NJ: World Scientific.

Cramer, D. (2003). *Advanced quantitative data analysis*. Philadelphia, PA: Open University Press.

do Prado, H.A., Engel, P.M., & Filho, H.C. (2002). Rough clustering: An alternative to find meaningful clusters by using the reducts from a dataset. In J.J. Alpigini, J.F. Peters, J. Skowronek & N. Zhong (Eds.), *Rough sets and current trends in computing, Third International Conference RSCTC 2002*. Berlin: Springer.

Dumitrescu, D., Lazzerini, B., & Jain, L. C. (2000). *Fuzzy sets and their application to clustering and training*. Boca Raton, FL: CRC Press.

Everitt, B.S., Landau, S., & Leese, M. (2001). *Cluster analysis* (4th ed.). New York: Oxford University Press.

Hair, J.E., Anderson, R.E., Tatham, R.L., & Black, W.C. (1998). *Multivariate data analysis* (5th ed.). London: Prentice-Hall International.

Lin, T.Y., & Cercone, N. (Eds.). (1997). *Rough sets and data mining: Analysis of imprecise data*. Boston: Kluwer.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In L.M. Le Cam & J. Neyman (Eds.), *Proceedings of the 5th Berkeley symposium on mathematics, statistics and probability* (vol. 1, pp. 281-298). Berkeley, CA: University of California Press.

Munakata, T. (1998). *Fundamentals of the new artificial intelligence: Beyond traditional paradigms*. New York: Springer-Verlag.

Pawlak, Z. (1982). Rough sets. *International Journal of Information and Computer Sciences, 11*, 341-356.

Pawlak, Z. (1991). *Rough sets: Theoretical aspects of reasoning about data*. Boston: Kluwer.

Polkowski, L., & Skowron, A. (Eds.). (1998). *Rough sets and current trends in computing (First International Conference, RSCTC98)*. Berlin: Springer.

Polkowski, L., Tsumoto, S., & Lin, T.Y. (Eds.). (2000). *Rough set methods and applications: New developments in knowledge discovery in information systems*. New York: Physica-Verlag.

Punj, G., & Stewart, D.W. (1983). Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research*, *20*, 134-148.

Voges, K.E., & Pope, N.K.Ll. (2004). Generating compact rough cluster descriptions using an evolutionary algorithm. In K. Deb (Ed.), *Genetic and evolutionary computation - GECCO2004*. New York: Springer-Verlag.

Voges, K.E., Pope, N.K.Ll., & Brown, M.R. (2002). Cluster analysis of marketing data examining on-line shopping orientation: A comparison of *k*-means and rough clustering approaches. In H.A. Abbass, R.A. Sarker, & C.S. Newton (Eds.), *Heuristics and optimization for knowledge discovery* (pp. 207 - 224). Hershey, PA: Idea Group Publishing.

Wang, G., Liu, Q., Yao, Y., & Skowron, A. (Eds.). (2003). *Rough sets, fuzzy sets, data mining, and granular computing: Proceedings 9th international conference, RSFDGrC 2003*. New York: Springer.

Wedel, M., & Kamakura, W.A. (2000). *Market segmentation: Conceptual and methodological foundations* (2nd ed.). Boston: Kluwer Academic.

Zadeh, L.A. (1965). Fuzzy sets. *Information and Control*, *8*, 338-353.

Zhong, N., Skowron, A., & Ohsuga, S. (Eds.). (1999). *New directions in rough sets, data mining, and granular-soft computing*. Berlin: Springer.

KEY TERMS

Approximation Set: An alternative (and more technically correct) name for a rough set, which is defined by two sets, the lower approximation and the upper approximation.

Boundary Region: Those objects that may or may not be in the approximation set. It is the difference between the upper approximation and the lower approximation. If the boundary region is empty, the set is said to be crisp. If the boundary region is not empty, the set is rough.

Cluster Analysis: A data analysis technique involving the grouping of objects into sub-groups or clusters so that objects in the same cluster are more similar to one another than they are to objects in other clusters.

k-Means Clustering: A cluster analysis technique in which clusters are formed by randomly selecting k data points as initial seeds or centroids, and the remaining data points are assigned to the closest cluster on the basis of the distance between the data point and the cluster centroid.

Lower Approximation: In rough sets theory, one of the two sets used to define a rough, or approximate set. The lower approximation contains objects that are definitely in the approximation set.

Market Segmentation: Market segmentation is a central concept in marketing theory and practice, and involves identifying homogeneous sub-groups of buyers within a heterogeneous market. It is most commonly conducted using cluster analysis of the measured demographic or psychographic characteristics of consumers. Forming groups that are homogenous with respect to these measured characteristics segments the market.

Rough Classification: Rough classification finds mappings from the partitions induced by the equivalence relations in the condition attributes to the partitions induced by the equivalence relations in the decision attribute(s). These mappings are usually expressed in terms of decision rules. It performs the same type of classification function as discriminant analysis or logistic regression, where there is a known sub-grouping in the data set, which is identified by the decision attribute.

Rough Clustering: Rough clustering is a simple extension of rough sets theory, and is analogous to traditional cluster analysis. The information table has no pre-existing subgroups, and clusters of objects are formed based on a distance measure. Clusters are defined by a lower approximation (objects exclusive to that cluster) and an upper approximation (all objects in the cluster which are also members of other clusters), in a similar manner to rough sets. An object can belong to more than one cluster.

Rough Set: The concept of rough, or approximation, sets was introduced by Pawlak, and is based on the single assumption that information is associated with every object in an information system. This information is expressed through attributes that describe the objects, and objects that cannot be distinguished on the basis of a selected attribute are referred to as indiscernible. A rough set is defined by two sets, the lower approximation and the upper approximation.

Upper Approximation: In rough sets theory, one of the two sets used to define a rough, or approximate set. The upper approximation contains objects that may or may not be in the approximation set. It can be formally defined as the union of the lower approximation and the boundary region. It is the complement of the set of objects definitely not in the set.

Cognitive Research in Information Systems

Felix B. Tan
Auckland University of Technology, New Zealand

M. Gordon Hunter
The University of Lambridge, Canada

INTRODUCTION

The existence and significance of cognition in organizations and its influence on patterns of behaviour in organizations and organizational outcomes are increasingly accepted in information systems (IS) research (Barley, 1986; DeSanctis & Poole, 1994; Griffith, 1999; Griffith & Northcraft, 1996; Orlikowski, 1992, 1994 #208). However, assessing the commonality and individuality in cognition and eliciting the subjective understanding of research participants either as individuals or as groups of individuals remain a challenge to IS researchers (Orlikowski & Gash, 1994). Various methods for studying cognition in organizations have been offered - for example, clinical interviewing (Schein, 1987), focus groups (Krueger, 1988), discourse-based interviewing (Odell, Goswami & Herrington, 1983). This article proposes that cognition applied to making sense of IT in organizations can also be explored using Kelly's (1955) Personal Construct Theory and its methodological extension, the Repertory Grid (RepGrid). The RepGrid can be used in IS research for uncovering the constructs research participants use to structure and interpret events relating to the development, implementation, use and management of IS in organizations.

In the context of this article, cognition is considered to be synonymous with subjective understanding: "the everyday common sense and everyday meanings with which the observed human subjects see themselves and which gives rise to the behaviour that they manifest in socially constructed settings" (Lee, 1991, p. 351). Research into cognition in organizations investigates the subjective understanding of individual members within the organization and the similarities and differences in the understandings among groups of individuals (Jelinek & Litterer, 1994; Porac & Thomas, 1989). In IS research, it is the personal constructs managers, users and IS professionals use to interpret and make sense of information technology (IT) and its role in organizations. The discussion here outlines the myriad of ways the RepGrid can be employed to address specific research objectives relating to subjective understanding and cognition in organizations. It illustrates, from a variety of published studies in IS (see Table 1), the flexibility of the RepGrid to support both qualitative and/or quantitative analyses of the subjective understandings of research participants.

BACKGROUND

We propose to use a framework to facilitate this discussion (see Figure 1) that presents a two-dimensional view of the types of research using the repertory grid. The examples in Table 1 are mapped along these two dimensions.

Figure 1. Distinguishing research using the repertory grid

Theory-Focused	Hunter (1997) *	Latta and Swigger (1992) †
Method-focused	Moynihan (1996) *	Phythian and King (1992) †

* Idiographic (i.e. individual interpretations – unique grids)
† Nomothetic (i.e. group interpretations – common grids)

Theory-Focused vs. Method-Focused

On one dimension, we distinguish research that applies Kelly's (1955) personal construct theory (theory-focused) from those applying the repertory grid method, without delving into the conceptual underpinnings of the theory (method-focused). When introduced some 45 years ago, the repertory grid technique served as the methodological extension of the personal construct theory. It operationalizes key aspects of Kelly's fundamental postulate and corollaries. IS researchers interested in the subjective understandings of individuals will find the repertory grid a powerful tool that permits the study of the individual's construct system and provides richer cognitive insights into research findings. For example, Latta and Swigger (1992) validated the use of the repertory grid for representing commonality of construing among participants regarding the design of intelligent user interfaces. The study lent strong support to the commonality corollary in grids, which can be confidently used to represent a consensus of knowledge around a problem domain. Hunter (1997) used the laddering technique to elicit what Kelly termed as super-ordinate constructs – constructs that are core to the individual's system of interpretation.

In contrast, there is research that has accepted Kelly's theory and employed the repertory grid solely as a data gathering technique. These works have employed the utility of the technique purely for its methodological strengths. Stewart and Stewart (1981) suggest, "At its simplest, Grids provide a way of doing research into problems – any problems – in a more precise, less biased way than any other research method" (pp. vii). These authors further contend that the repertory grid "….enables one to interview someone in detail, extracting a good deal of information … and to do this in such a way that the input from the observer is reduced to zero" (p. 5). Two of the examples in Table 1 have taken the method-focused approach to the use of the repertory grid technique. For instance, Moynihan (1996) was purely interested in using the repertory grid technique to collect data and to compare the results with the extant literature. Moynihan argued that the free-ranging responses resulting from the non-prohibitive nature of the technique permitted the participants to apply the "theories-of-action" (theories individuals use to guide their actions) they employ daily – resulting in the identification of themes and issues over and above the extant literature. In the studies by Phythian and King (1992), the repertory grid was used to explore the similarity and differences in the views between individual managers. No direct references were made to Kelly's personal construct theory, as the focus was to identify key factors influencing tender decisions and the relationships among these factors by interviewing two managers closely involved in such tender activities.

Qualitative vs. Quantitative

On the second dimension, we distinguish research that is either qualitative or quantitative. The identification of emerging themes from elicited constructs is common in a qualitative approach using the repertory grid. For example, Hunter (1997), when investigating how certain groups of individuals interpreted the qualities of "excellent" systems analysts, employed content analysis of the data gathered from individual interviews conducted using the repertory grid technique. The numeric component of the grid was only employed to conduct visual focusing at the end of each interview as a means of quickly assessing what had transpired during the interview and whether the research participant agreed with this initial assessment. Similarly, Moynihan (1996) employed the repertory grid technique as a method to elicit interpretations from research participants of what aspects were considered important when deciding upon an approach to adopt for projects to be conducted for external clients. Unique grids were developed for each research participant. Then the data were analyzed from a qualitative perspective via content analysis at the construct level, where emerging themes were identified and categorized. In these examples, the researchers took an open view toward gathering data and allowed themes or categories to emerge from the data as the investigation proceeded.

In contrast, the quantitative approach utilizes mathematical and/or statistical analyses of grid data (Daniels, Markoczy & de Chernatony, 1994). These techniques are commonly used to explore the structure and content of an individual's construct systems or make comparisons between groups of individuals (Ginsberg, 1989). This approach was adopted by two of the examples in Table 1. For instance, in Phythian and King (1992), statistical analyses (specifically, cluster analysis and correlation analysis) were conducted on individual and combined grids. These data were used to support the development of an expert support system. Similarly, Latta and Swigger (1992) applied cluster analysis and Spearman's rank order correlation to analyze the grids. The study revealed an overall correlation between the students' and the instructor's grids, promoting the utility of the repertory grid technique in modeling knowledge relating to the design of information systems.

Idiographic vs. Nomothetic

Within both the qualitative and quantitative perspectives, research using the repertory grid technique is either

Table 1. Examples of IS research using the RepGrid

	Hunter (1997)	Moynihan (1996)	Phythian and King (1992)	Latta and Swigger (1992)
Research Objectives	Explore the qualities of "excellent" systems analysts	Identify the situational factors considered in the planning/running of new systems development projects	Develop rules for an expert system to support customer tender evaluations	Validate the RepGrid in modeling communal knowledge regarding design of system interfaces
Research Perspective	Qualitative	Qualitative	Quantitative	Quantitative
Nature of RepGrid	Idiographic	Idiographic	Nomothetic	Nomothetic
Key Findings	Several themes considered as qualities of "excellent" systems analysts	Identified themes over and above literature. Differences in project managers' construction of project contexts	Identified key factors and rules influencing tender decisions. Expert system improved consistency	Commonality of constructions support the use of the RepGrid to model group knowledge
Research Design:				
Element Selection	Systems analysts with whom participant has interacted	Systems development projects on which participant has worked	Previous customer tender enquiries	Components of online bibliographic retrieval systems
	Elicited	Elicited	Supplied	Supplied
Construct Elicitation	Qualities of "excellent" systems analysts	Situational factors influencing risks in new systems projects	Key factors and rules influencing tender decisions	Attributes of system interface design
	Minimum context form (triadic sort) and laddering	Minimum context form (triadic sort)	Minimum context form (triadic sort) and laddering	Minimum context form (triadic sort) and supplied constructs
Linking	Rating	None	Rating (Grid) Ranking (Elements)	Rating
RepGrid Analysis	Content analysis Visual focusing COPE and VISA	Content analysis	Cluster analysis (FOCUS), correlation, mathematical modeling	Cluster analysis, correlation
Sample and Size	53 (users and IT professionals) from two insurance companies	14 systems development project managers	Two manager-experts involved in assessing tender enquiries	Instructor and students who completed an "information search and retrieval" course

idiographic or nomothetic in nature. The idiographic approach focuses on the subjective experiences of the individual and presents results in expressions and terms used by the individual. The resulting grid is considered unique in that there are no common elements or constructs employed in the elicitation process. For example, in the study of systems analysts, each participant was asked to name up to six systems analysts with whom s/he had interacted. In this project, Hunter (1997) provided a role description (i.e., system analysts interacted with) and asked each participant to specify examples that fit this category. The analysts named were not common among participants and as such the resulting grids were not common in terms of the elements used. Similarly, Moynihan (1988) asked each participating project manager to make a list of systems development projects s/he had worked on as a project manager. If the project manager named more than nine projects, s/he was then asked to choose the three that were most successful, the three that were the least successful, and three in between. Moynihan's research objective was to identify the situational factors project managers regard as important when planning new development projects and not to compare the subjective understandings of different project managers. As such, he did not supply a common set of systems development projects that would have permitted a comparative analysis of individual grids.

In contrast, research comparing the grids of individuals or groups of individuals requires different decisions to be made concerning the elements and constructs in the repertory grid process. This nomothetic approach necessitates the use of a common set of elements and/or constructs to permit comparisons to be made between grids (Easterby-Smith, 1980). Such research also tends to be quantitative in nature. For example, research to identify the similarities and differences in two managers' views on tender enquiry evaluations imposed a common set of tender enquiries as elements in the repertory grid process (Phythian & King, 1992). This permitted the comparison of the construct systems of the two managers based on their personal experiences of similar events. In another example, a set of constructs was elicited from an instructor and then supplied as common constructs for a group of students to evaluate against a prescribed set of elements representing the features of online bibliographic retrieval systems (Latta & Swigger, 1992). This permitted the commonality of construing among students and between students and the instructor to be tested. In these examples, the reason for the use of common components in the repertory grid process was to compare grids.

Finally, none of the examples discussed in this section approached their studies using both qualitative and quantitative approaches. This does not imply that the repertory grid cannot lend itself to both qualitative and quantitative analysis of the collected data.

CONCLUSION

This article has demonstrated the flexibility of the repertory grid technique to support both qualitative and/or quantitative approaches to investigating the subjective understanding of human subjects in complex and socially dynamic organizations. With this technique, both qualitative and quantitative approaches need not be mutually exclusive. For instance, Hunter (1997) approached his study of systems analysts from a grounded theory perspective. The data collected using the repertory grid were analysed to identify themes or qualities of "excellent" systems analysts. In an extension of this investigation, the characteristics of excellent systems analysts were statistically tested to highlight the similarities and differences in the subjective understandings of Canadian and Singaporean research participants (Hunter & Beck, 1996).

We would like to conclude by reiterating two important points. The first is a word of caution. The repertory grid should not be considered a panacea to investigating the subjective understanding of individuals in an organizational setting. It can be used in conjunction with other methods – as a means of validating other techniques or as a preliminary phase to further interpretive or positivist investigations. The second is that the personal construct theory is one of several theories in cognitive science (Berkowitz, 1978). The repertory grid is one of several cognitive mapping methods available to the IS researcher (Huff, 1990). This article was written in an attempt to stimulate interest in the IS research community of the need for more cognitive emphasis in our field. Hopefully, IS researchers will be encouraged to further reflect on the virtues of applied theories and methods that can deliver utilizable and consumable outcomes to research and practice in IS.

REFERENCES

Barley, S.R. (1986). Technology as an occasion for structuring: Evidence from observations of CT scanners and the social order of radiology departments. *Administrative Science Quarterly*, *31*, 78-108.

Berkowitz, L. (1978). *Cognitive theories in social psychology*. New York: Academic Press.

Daniels, K., Markoczy, L., & de Chernatony, L. (1994). Techniques to compare cognitive maps. *Advances in*

Managerial Cognition and Organizational Information Processing, 5, 141-164.

DeSanctis, G., & Poole, M.S. (1994). Capturing the complexity in advanced technology use: Adaptive structuration theory. *Organization Science, 5*(2), 121-147.

Easterby-Smith, M. (1980). The design, analysis and interpretation of repertory grids. *International Journal of Man-Machine Studies, 13,* 3-24.

Ginsberg, A. (1989). Construing the business portfolio: A cognitive model of diversification. *Journal of Management Studies, 26*(4), 417-438.

Griffith, T.L. (1999). Technology features as triggers for sensemaking. *Academy of Management Review, 24*(3), 472-488.

Griffith, T.L., & Northcraft, G.B. (1996). Cognitive elements in the implementation of new technology: Can less information provide more benefits? *MIS Quarterly, 20,* 99-110.

Huff, A.S. (1990). *Mapping strategic thought.* Chichester: John Wiley & Sons Ltd.

Hunter, M.G. (1997). The use of RepGrids to gather interview data about information systems analysts. *Information Systems Journal, 7*(1), 67-81.

Hunter, M.G., & Beck, J.E. (1996). A cross cultural comparison of 'excellent' systems analysts. *Information Systems Journal, 6,* 261-281.

Jelinek, M., & Litterer, J.A. (1994). Toward a cognitive theory of organizations. In C. Stubbar, J.R. Meindl & J.F. Porac (Eds.), *Advances in managerial cognition and organizational information processing* (pp. 3-41). Greenwich, CT: JAI Press.

Krueger, R.A. (1988). *Focus groups: A practical guide for applied research*. Newbury Park, CA: Sage Publications.

Latta, G.F., & Swigger, K. (1992). Validation of the repertory grid for use in modeling knowledge. *Journal of the American Society for Information Science, 43*(2), 115-129.

Lee, A.S. (1991). Integrating positivist and interpretive approaches to organizational research. *Organization Science, 2*(4), 342-365.

Moynihan, J.A. (1988). Current issues in introducing and managing information technology: The chief executive's perspective. In *Information technology for organisational systems*. Brussels-Luxembourg: Elsevier Science.

Moynihan, T. (1996). An inventory of personal constructs for information systems project risk researchers. *Journal of Information Technology, 11,* 359-371.

Odell, L., Goswami, D., & Herrington, A. (1983). The discourse-based interview: A procedure for exploring tacit knowledge of writers in nonacademic settings. In *Research on writing: Principals and methods* (pp. 220-236). New York: Longman.

Orlikowski, W.J. (1992). The duality of technology: Rethinking the concept of technology in organizations. *Organization Science, 3*(3), 398-427.

Orlikowski, W.J., & Gash, D.C. (1994). Technological frames: Making sense of information technology in organizations. *ACM Transactions on Information Systems, 12*(2), 174-201.

Phythian, G.J., & King, M. (1992). Developing an expert system for tender enquiry evaluation: A case study. *European Journal of Operational Research, 56*(1), 15-29.

Porac, J.F., & Thomas, H. (1989). Competitive groups as cognitive communities: The case of Scottish knitwear manufacturers. *Journal of Management Studies, 26*(4), 397-416.

Schein, E. (1987). *The clinical perspective in fieldwork.* Newbury Park, CA: Sage.

Stewart, V., & Stewart, A. (1981). *Business applications of repertory grid.* UK: McGraw-Hill.

KEY TERMS

Cognition: Cognition is considered to be synonymous with subjective understanding, "the everyday common sense and everyday meanings with which the observed human subjects see themselves and which gives rise to the behaviour that they manifest in socially constructed settings" (Lee, 1991, p. 351).

Construct: Constructs represent the research participant's interpretations of the elements. Further understanding of these interpretations may be gained by eliciting contrasts resulting in bi-polar labels. Using the same example, research participants may come up with bi-polar constructs such as "high user involvement – low user involvement" to differentiate the elements (i.e., IS projects). The labels represent the CSFs of IS projects.

Elements: Elements are the objects of attention within the domain of investigation. They define the entities upon which the administration of the RepGrid is based.

For example, to explore the critical success factors (CSFs) of IS projects, IS researchers can use IS projects as elements in the RepGrid.

Idiographic: The idiographic approach focuses on the subjective experiences of the individual and presents results in expressions and terms used by the individual. The resulting RepGrid is considered unique in that there are no common elements or constructs employed in the elicitation process across the sample.

Links: Links are ways of relating the elements and constructs. The links show how the research participants interpret each element relative to each construct. Further, the links reveal the research participant's interpretations of the similarities and differences between the elements and constructs.

Nomothetic: The nomothetic approach permits the comparison of RepGrids of individuals or groups of individuals. It necessitates the use of a common set of elements and/or constructs to permit comparisons to be made between RepGrids.

Repertory Grid (RepGrid): The RepGrid is a cognitive mapping technique that can be used to describe how people think about a phenomenon in their world. The RepGrid technique, for IS, entails a set of procedures for uncovering the personal constructs individuals use to structure and interpret events relating to the development, implementation, use, and management of IT in organizations. The RepGrid contains three components – elements, constructs and links.

Collaborative Learning On–Demand

Giancarlo Fortino
DEIS - Università della Calabria, Italy

INTRODUCTION

Advances in communications and software technology are leading the Internet to become an open and distributed computing platform able to provide diversified services which can be ubiquitously accessed by many users.

Electronic learning-oriented services are well suited to be supported by and delivered through the Internet. They are strategic for enabling virtual universities, enhancing the skills of the employees of a company, and facilitating auto-training.

Nowadays, a multitude of networked e-learning systems and applications, which can be usefully employed to support distance learning, have been developed. They are basically based either on asynchronous learning models or on synchronous learning models (see Table 1). Specifically, "virtual or real" collaborative learning environments are particularly interesting in the education research area since they are aimed at creating computer-based and multimedia-oriented learning processes where learners, that belong to a collaborative and interactive group, cooperatively construct knowledge (Costantini & Toinard, 2001).

In fact, it has been proved that instructional methods promoting interpersonal discourse and social construction of knowledge (i.e., collaborative learning techniques) are more effective than methods which simply rely on the broadcast of information (classroom transmission metaphor) or on the asynchronous, self-paced access to online training materials (Cohen, 1994).

In this article, an overview of the *Collaborative Learning On-Demand (CLOD)* paradigm along with its supporting technology (Fortino & Nigro, 2003) is presented.

In particular, the VICRO^{C} system, which fully supports the CLOD paradigm, is described. Finally, the usability evaluation of VICRO^{C} is shown. The carried out experimentations confirmed that the CLOD paradigm has the potential to improve the learner's productivity.

BACKGROUND

Collaborative Learning on-Demand (CLOD) is a virtual collaborative learning paradigm which enables a self-tutored, interactive and cooperative learning process where a small group of remote students requests, watches and controls the playback of an archived lecture by exchanging questions with each other (Fortino & Nigro, 2003).

CLOD borrows some of the ideas of:

- *Tutored Video Instruction* (TVI), which is a face-to-face collaborative learning methodology in which a small group of students driven by a tutor goes over a videotape of a lecture;
- *Distributed Tutored Video Instruction* (DTVI), which is a fully virtual version of TVI, in which each student has a networked computer equipped with audio (microphone and headset) and video (camera) facilities to communicate within a group.

Table 1. Asynchronous vs. synchronous distance learning

	Synchronous	Asynchronous
Model	The teacher and the student interact with each other in "real time".	The teacher may deliver the instruction via video, computer, or other means, and the students respond at a later time.
Technology	VideoConferencing, AudioConferencing, Internet Chat, Desktop Videoconferencing, Whiteboard	Video Tape, Broadcast Video, E-mail, CD-ROM, WWW-based courses
Example	Using a two-way video conferencing tool, students interact with "live" video of an instructor.	Instruction may be delivered via the Web or videotapes, and the feedback could be sent via e-mail messages.

Table 2. A comparison between CLOD and DTVI

FEATURE	CLOD	DTVI
LEARNING TYPE	SELF-TUTORED	TUTORED
GROUP SIZE	SMALL/MEDIUM (<=20)	SMALL (<=5)
CONTROL OF THE VCR	SHARED AMONG STUDENTS	APPLIED BY THE TUTOR
SUPPORTING TECHNOLOGY	COOPERATIVE PLAYBACK SYSTEMS	CONFERENCING SYSTEMS

TVI and DTVI have proven real effectiveness in that the students involved in their experimentation have been shown to outperform students who physically attended the lectures (Sipusic, Pannoni, Smith, Dutra, Gibbons, & Sutherland, 1999).

The substantial difference between CLOD and DTVI (see Table 2) is that CLOD methodology does not assume the presence of a tutor which guides students to construct knowledge. This fact has a direct impact on the technical implementation of CLOD because, while in DTVI only the tutor has control of the videoconference recorder (VCR), in CLOD each participant to the playback session uses a shared VCR remote controller. In addition, being the learning service on-demand, CLOD needs to be supported by a video on-demand system (VoD).

The CLOD paradigm is supported by the Cooperative Playback Systems (CPS; Fortino & Nigro, 2000) which are media on-demand systems providing cooperative playback sessions. In a cooperative playback session, the participants, who are explicitly grouped, share the vision and the control of a multimedia session streamed by a media server, and interact with each other by means of a question board.

The construction of CPS is efficiently enabled by the IP multicast technology which concurs to save network resources (e.g., bandwidth) and improve scalability (Kumar, 1996). IP multicast allows for the transmission of a packet to a group of hosts which are identified by a multicast address belonging to the class D of IP addresses. The worldwide testbed of IP-multicast is MBone (Multicast Backbone; Kumar, 1996) which, to date, cannot be accessed by all the users of Internet. However, private IP multicast-enabled networks (campus networks or intranets) can be easily set-up. IP multicast has promoted the proliferation of a rich set of multimedia applications, systems and protocols able to support synchronous distance learning over the Internet according to the real/virtual classroom metaphor (Costantini & Toinard, 2001; Crowcroft, Handley, & Wakeman, 1999; Kumar, 1996).

In particular, a CPS can be developed by integrating two enabling MBone-based technologies:

- *VideoConference Recording on-Demand (VCRoD) systems*. They are VoD-like systems which allow a user to connect to a Media Server (MS) and request two kinds of services: recording and playback (Fortino & Nigro, 2003; Holfelder, 1997; MLB, 2001; Parnes, Synnes & Schefstrom, 2000; Shuett, Raman, Chawathe, McCanne & Katz, 1998). Upon requesting a recording service, users can either select a media session being transmitted over an IP-multicast address or send their own media session directly to the MS. This way, the MS archives the media session in a multimedia repository. On the other hand, the playback service allows users to browse the list of the archived media sessions, select a particular media session and control its playback by means of a VCR remote controller. In Table 3, the description of three main VCRoD systems is given.
- *MBone tools*. They are multimedia applications enabling a group of users to interactively and synchronously exchange audio/video "live" streams, text-based messages, and to cooperatively share whiteboards and document editors (MASH Consortium, 2003; MBT, 2003; Parnes, Synnes & Schefstrom, 2000).

To date, a few trials have been devoted to building CPS. The most significant contributions in this direction are the VICRO^C system (Fortino & Nigro, 2003), and the MASH Rover system (MASH Consortium, 2003; Shuett, Raman, Chawathe, McCanne & Katz, 1998).

In particular the VICRO^C system addresses all the features of a CPS whereas the MASH Rover system only provides basic services.

VICRO^C: A COOPERATIVE PLAYBACK SYSTEM FOR CLOD

The main functionalities of the VICRO^C system can be summarized as follows.

C

Table 3. VCRoD systems

SYSTEM	DESCRIPTION
MASH Rover	It consists of two components: the media server and the MASH streaming player. The media server is implemented by using the TACC (Transformation, Aggregation, Customization and Control) toolkit. An RTSP-like protocol is used to remotely control media streams. Mechanisms of advertising and discovering new contents are provided such as rich-description hyperlinks to the archived sessions and automatic detection of significant instants during a session. The client part consists of the MASH streaming player that can operate as media browser or as helper application within a Web browser. It allows the user to bookmark specific sessions and specific instants within a session. The bookmark file can be shared among several participants. The client is implemented in C++ and TCL/TK. It was developed at the University of Berkeley (USA).
mMOD	It consists of three separate parts: the VCR, the data translator and the Web-controller. The VCR is a stand-alone program for recording and playing back IP-packets on either UDP or RTP level. The Data-Translator translates the traffic in various ways (recoding, mixing and switching techniques) to allow users with different bandwidth to access the service. The Web-controller is a program that acts as a Web-interface of mMOD. Using this interface, a new session can be started and controlled and information about running sessions can be listed so that a user can join them. mMOD, which is completely written in Java, was developed at the University of Lulea (SVE).
MVoD	It consists of three basic components: the Manager and the VideoPump, that form the logical unit called MVoD Server, and the MVoD Client. The interaction between the system components is regulated by four protocols: the VCR Service Access Protocol (VCRSAP), the VCR Stream Control Protocol (VCRSCP), the VCR Announcement Protocol (VCRAP), and the VCR Client Message Protocol (VCRCMP). The Manager and the Client are implemented in Java. The VideoPump is implemented in C++. The MVoD Client can be started either as a Java applet within a Java-enabled browser, or as a stand-alone Java application. It was developed at the University of Mannheim (GER).

Group organization, which contains group formation and group management. The former enables the creation of a group of users wishing to work on and control the same playback session. The latter deals with the following issues: (i) how to share the starting time of a playback session for the synchronization of the group members; (ii) how to expel from the group a member who interrupts the others using an improper behavior. To join a cooperative playback session, a Media Client (MC) has to contact the Media Server (MS) in order to authenticate itself and receive the session key (or *Secret Share k*). The scheme adopted is based on asymmetric cryptography and a Certificate Authority (CA; Fortino, Russo & Zimeo, 2003).

Media streaming, which allows for the multicast transmission of encrypted media streams based on the Real-time Transport Protocol (RTP; Schulzrinne, Casner, Frederick & Jacobson, 1996) to the cooperative group which is tuned on a Multicast Media Group (MMG). Encryption of media streams is based on an efficient technique centered on the Blowfish symmetric encryption algorithm (Fortino, Russo & Zimeo, 2003).

Control sharing, which enables a group of users to cooperatively control a media streaming session.

Joint-work, which allows the group participants to collaborate with each other by questioning on the contents of the cooperative session.

The architecture of ViCRO^C is depicted in Figure 1, which shows the basic components along with their protocol-based interactions and highlights the security-enhanced blocks.

Media streaming and playback are respectively supported by the streamer and player components which are based on the Java Media Framework (JMF) in which security was integrated, and are respectively located at the Media Server (MS) and Media Client (MC) sites.

Cooperative playback control is enabled by the Multicast Archive Control Protocol (MACπ) which is a multicast version of the Real Time Streaming Protocol (RTSP; Schulzrinne, Rao & Lanphier, 1998).

Collaboration among users is based on the COllaborative Protocol (COπ) which allows for the multicast-based exchange of questions and annotations. Both MACπ and COπ are based on the Lightweight Reliable Multicast Protocol (LRMP; Liao, 1998).

The Multimedia Archive keeps stored MPEG and RTP-based media files.

The exchange of the secret session keys is enabled by the security-based join tool and the session key dispenser which communicate through SDP/SIP messages (Session Description Protocol/Session Initiation Protocol; Crowcroft, Handley & Wakeman, 1999).

EVALUATION OF VICRO^C

The usability of a CPS is mainly influenced by general factors, which are related to the network and to the desktop hardware and software environment, and by specific factors such as media on-demand service availability, easiness to set up a cooperative playback session, and degree of appeal and simplicity-to-use of the graphical user interface.

The usability evaluation of ViCRO^C was carried out by using three different group of users in different scenarios based on a local cluster of PCs (Fortino & Nigro, 2003). Each user group was small (about four to five persons) and formed by (i) professors and researchers, (ii) graduate students and (iii) students.

In order to have useful feedback by a user, a simple questionnaire was required to be filled out at the end of a cooperative playback session.

The questions were about:

- the perceived *media quality* including video and audio quality, and the quality of the lip synchronization;
- the perceived *interactivity* fluidity when playback control is applied and questions are sent to and received from companions;
- the *friendliness* and *look & feel* attractiveness of the GUI;
- the *degree of breakdown* due to:
  (a) the lack of monitor space to layout all the GUIs;
  (b) the frequent pop-ups of the *Answer* dialog, which allows replying to a question, and the *Voting* dialog, which allows voting a control command.

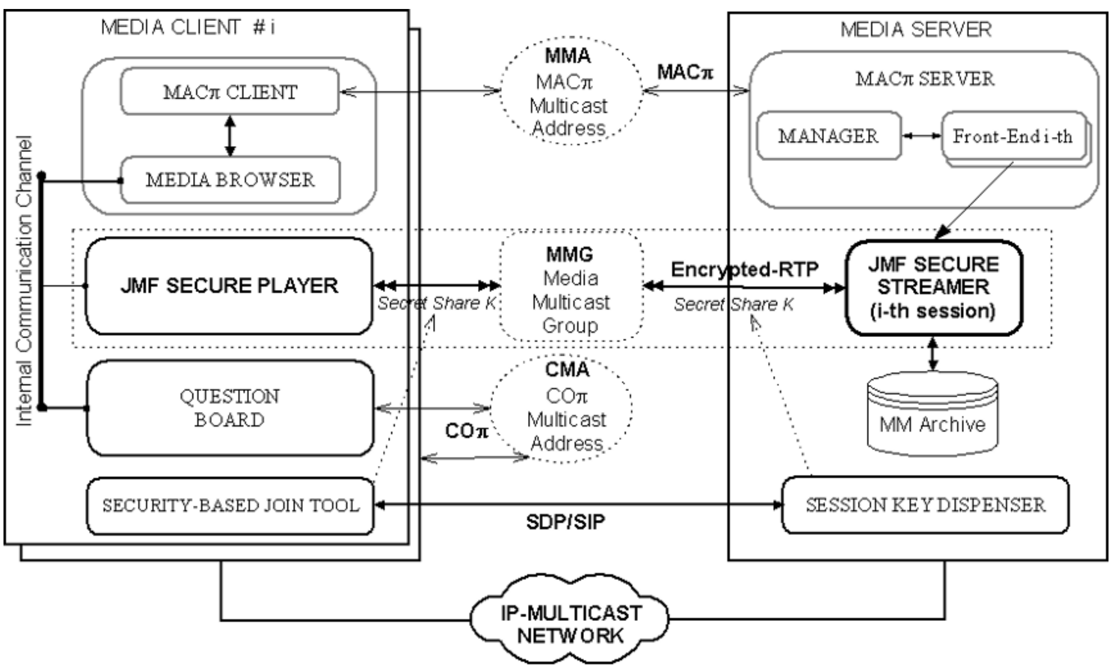
Each question was scored from 1 (lowest) to 5 (highest) points.

The analysis of the results (Table 4) showed that:

- the quality of service of the audio/video streams was considered good by the user;
- the control GUI was considered intuitive and simple to use;
- when the interaction among users increases and the control actions are more frequent, user's attention is taken away from the playback session and focuses on the interaction with the *Answer* and *Voting* dialogs.

The overall evaluation indicated that the CLOD paradigm, featured by the ViCRO^C system, increases the productivity of learners organized in small homogeneous groups.

Figure 1. Architecture of ViCRO^C



C

Table 4. Results of the usability evaluation

Group	Media Quality	Interactivity	Friendliness and Look&Feel	Degree of Breakdown
Professors (mean values)	4.0	3.8	4.2	4.0
Graduate Students (mean values)	3.8	3.6	4.2	4.1
Students (mean values)	3.7	3.4	4.1	4.2

FUTURE TRENDS

The future CPS can greatly benefit from the following advances of the Internet technology: (i) increase of available bandwidth, (ii) "non best effort" guarantees of quality of service and (iii) worldwide exploitation of IP multicast. The first two points can improve usability from a network point of view, whereas the third point can enable geographically distributed cooperative playback sessions. Appealing research opportunities were envisaged as follows: (i) integration of distributed virtual environments with CPS so bringing to life new collaborative distributed virtual environments; (ii) creation of personal and lightweight interfaces to CPS so enabling a client which uses a handheld device (or personal digital assistant) to participate to a cooperative playback session.

CONCLUSION

This article has proposed an overview of the Collaborative Learning On-Demand paradigm and technology. CLOD is an original learning paradigm which enables a group of students to jointly work on and share a playback session. CLOD is fully supported by the Cooperative Playback Systems (CPS); in particular, ViCRO^{C}, a full-fledged CPS developed at University of Calabria which also provides secure cooperative playback sessions, has been described. Usability evaluation of ViCRO^{C} was carried out on the basis of three kinds of tests. Results indicated that media quality is considered good and the GUI easy-to-use. However, under conditions of high transmission frequency of questions and control commands, the attention of the session participants shifts from the view of the playback to coping with frequent pop-ups of the dialog windows. Based on this experience, ViCRO^{C} is being upgraded by using selective filters, which limit undesired questions, and timeouts, which regulate control command transmissions.

REFERENCES

Cohen, E. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research*, *64*(1), 1-35.

Costantini, F., & Toinard, C. (2001). Collaborative learning with the distributed building site metaphor. *IEEE Multimedia*, *8*(3), 21-29.

Crowcroft, J., Handley, M., & Wakeman, I. (1999). *Internetworking multimedia*. San Francisco, CA: Morgan Kaufmann Publishing.

Fortino, G., & Nigro, L. (2000). A cooperative playback system for on-demand multimedia sessions over Internet. *Proceedings of the IEEE Conference on Multimedia and Expo*, New York, USA, August.

Fortino, G., & Nigro, L., (2003). Collaborative learning on-demand on the Internet MBone. In C. Ghaoui (Ed.), *Usability evaluation of online learning programs* (pp.40-68). Hershey, PA: Information Science Publishing.

Fortino, G., Russo W., & Zimeo, E. (2003). Enhancing cooperative playback systems with efficient encrypted multimedia streaming. *Proceedings of the IEEE Conference on Multimedia and Expo*, Baltimore, USA, July.

Holfelder, W. (1997). Interactive remote recording and playback of multicast videoconferences. *Proceedings of IDMS'97*, Darmstadt, Germany, September.

Kumar, V. (1996). *MBone: Interactive multimedia on the Internet*. Indianapolis, IN: New Riders Publishing.

Liao, T. (1998). Light-weight reliable multicast protocol. Documentation and libraries at the World Wide Web: *http://webcanal.inria.fr/lrmp/*

MASH Consortium. (2003). Mash streaming media toolkit and distributed collaboration applications based on the Internet MBone tools and protocols. University of Berkeley (CA). Software and documentation at the World Wide Web: *http://www.openmash.org/*

MBT (MBone Tools). (2003). Department of Computer Science, University College London. Software and documentation at the World Wide Web: *http://www-mice.cs.ucl.ac.uk/multimedia/index.html*

MLB (Multimedia Lecture Board). (2001). Documentation and tools at World Wide Web: *http://www.informatik.uni-mannheim.de/informatik/pi4/projects/mlb/*

Parnes, P., Synnes, K., & Schefstrom, D. (2000). mSTAR: enabling collaborative applications on the Internet. *IEEE Internet Computing, 4*(5), 32-39.

Schulzrinne, H., Casner, S., Frederick, R., & Jacobson, V. (1996). RTP: A transport protocol for realtime applications. Request for Comments, No. 1889, Internet Engineering Task Force, January.

Schulzrinne, H., Rao, A., & Lanphier, R. (1998). Real Time Streaming Protocol (RTSP). Request for Comments, No. 2326, Internet Engineering Task Force.

Shuett, A., Raman, S., Chawathe, Y., McCanne, S., & Katz, R. (1998). A soft state protocol for accessing multimedia archives. *Proceedings of the NOSSDAV'98*, Cambridge, UK, July.

Sipusic, M.J., Pannoni, R.L., Smith, R.B., Dutra, J., Gibbons, J.F., & Sutherland, W.R. (1999). Virtual collaborative learning: A comparison between face-to-face tutored video instruction (TVI) and distributed tutored video instruction (DTVI) (Report No. SMLI TR-99-72). Sun Microsystems Laboratories, Palo Alto (CA), USA.

KEY TERMS

Certificate Authority: An entity (typically a company) that issues digital certificates to other entities (organizations or individuals) to allow them to prove their identity to others.

CLOD: A virtual collaborative learning paradigm enabling a self-tutored, interactive and cooperative learning process where a small group of remote students requests, watches and controls the playback of an archived lecture by exchanging questions with each other.

Java Media Framework (JMF): A Java library for the development of stand-alone and networked multimedia systems.

MBone: The Virtual Internet Backbone for IP Multicast.

Media Encryption: Any procedure used in cryptography to convert media objects into cipher media objects in order to prevent any but the intended recipient from reading that data.

Multimedia Internetworking: Refers to network infrastructures, protocols, models, applications and techniques which are being currently deployed over the Internet to support multimedia applications such as videoconferencing, VoD, shared workspaces, and so forth.

Usability Evaluation: Can be defined as the act of measuring (or identifying potential issues affecting) usability attributes of a system or device with respect to particular users, performing particular tasks, in particular contexts.

Video On-Demand (VoD): A planned system using compressed video streams to supply programs to viewers when requested, via a network.

VideoConference Recording On-Demand (VCRoD): A VoD system which is also able to dump and archive networked media streams.

C

Collective Intentional Action in Virtual Communities

Richard P. Bagozzi
Rice University, USA

Utpal M. Dholakia
Rice University, USA

INTRODUCTION

The Internet is an important innovation in information science and technology and profoundly affects people in their daily lives. To date, these effects have been construed in overly individualistic ways and often all too negatively. For example, the Internet is seen by many as an individual means for obtaining or sending information flexibly and efficiently (e.g., Dreyfus, 2001). Some researchers also claim that participation on the Internet often leads to feelings of isolation and depression and even negatively affects relationships with one's family members and friends (Kraut et al., 1998; cf. Kraut et al., 2002; UCLA Internet Report, 2003). Likewise, Dreyfus (2001) takes a generally pessimistic tone with regard to Internet usage and worries that when we engage the Internet, it "diminishes one's sense of reality and of the meaning in one's life" and "…we might…lose some of our crucial capacities: our ability to make sense of things so as to distinguish the relevant from the irrelevant, our sense of the seriousness of success and failure that is necessary for learning, and our need to get a maximum grip on the world that gives us our sense of the reality of things."

In contrast to individualistic construals of Internet usage, we claim that the Internet often is a medium for group action ("collective intentional action") whereby people function in virtual communities to fulfill not only individual but jointly conceived and mutually beneficial ends (e.g., Bagozzi & Dholakia, 2002). Furthermore, participation on the Internet need not function negatively to harm people and their relationships with others. Instead, participation, particularly in virtual communities, can enhance one's personal well-being and promote positive social values and outcomes.

Our purpose in this article is to consider the emerging ideas and research concerning collective intentional action in virtual communities. We begin with a discussion of collective intentions of virtual community participants, and then study their antecedents and consequences. We conclude with a discussion regarding the importance of studying participants' emotions in enabling collective intentional action in virtual communities. Figure 1 sketches the primary variables and processes under consideration.

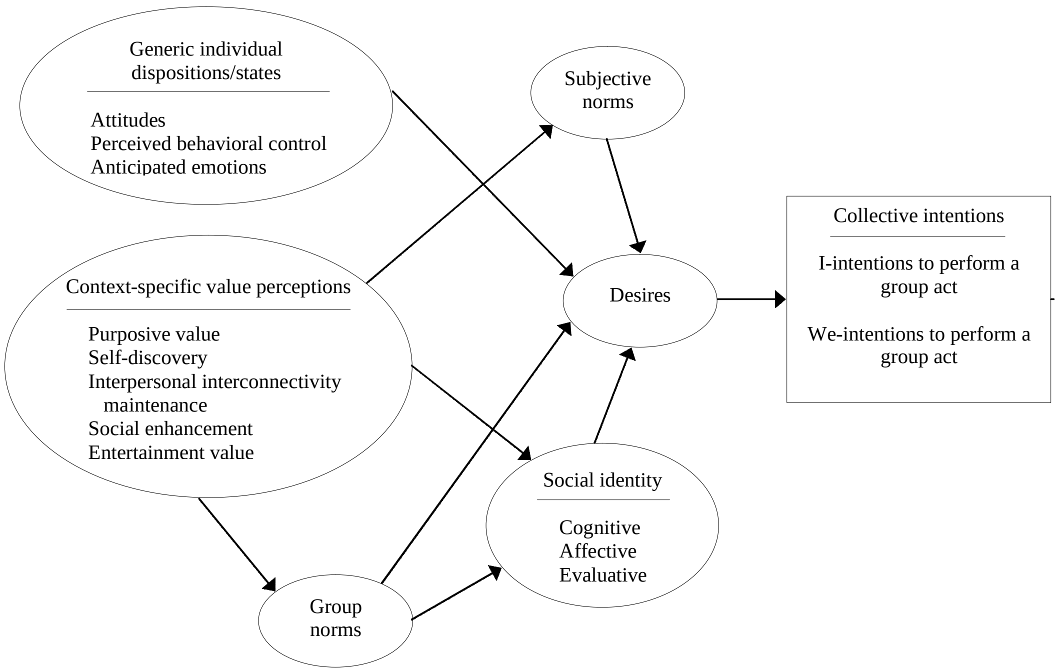
COLLECTIVE INTENTIONS OF VIRTUAL COMMUNITY PARTICIPANTS

Virtual communities are mediated social spaces in the online environment that allow collections of people to form and be sustained through ongoing communication processes. Two kinds of virtual communities are the network-based and small-group-based virtual communities (Dholakia, Bagozzi & Klein Pearo, 2004). The former is a specialized, geographically dispersed virtual community based on a structured, relatively sparse and dynamic pattern of relationships among participants sharing a common focus. Common examples include e-mail lists, Web site bulletin boards, and Usenet newsgroups. Members of network-based virtual communities typically view the community in terms of venue and only superficially identify with particular individuals within the community.

The small-group based virtual community is a handful or so of persons with a dense web of relationships, interacting together online in order to accomplish a wide range of jointly-conceived goals and to maintain the functioning of the group. Common instances of such communities are real-time, online-chat systems, Web-based chat-rooms, multi-player virtual games, and multi-user domains (MUDs). Members of small-group-based virtual communities typically engage often and intensely with the same specific individuals whom they know by name and to a certain extent personally, and they may even on occasion also meet face-to-face in certain communities.

Many behaviors involved in virtual communities, and which form the focus of the present article, are described by their actors through the use of collective concepts, based on a strong sense of "we-ness." Gilbert

Figure 1. Intentional social action and social identity in virtual communities plus key antecedents and consequences



(1992) suggests that "we" can mean the self and one or more others "that share in the action of a verb." Volitionally, we may equate a plural subject with "the concepts of a pool or sum of wills, dedicated, as one, to a certain 'cause'," e.g., promotion of a group goal (Gilbert, 1992). Then, "our wanting X is a reason for me to exercise my will in order to get X" (Gilbert, 1992).

Recently, philosophers have developed elaborate expositions of the logical foundations of collective intentions (Bagozzi, 2000). In abbreviated forms, two formulations of note are Tuomela's (1995) notion of a we-intention ("a commitment of an individual to participate in joint action [that] involves an implicit or explicit agreement between the participants to engage in that joint action" and Bratman's (1999) idea of shared intentions("for you and me to share an intention" means "first, we each recognize the other as a participant whose intentions in favor of the shared activity are partly constitutive of the shared intention. Second, we each intend that the other person's relevant intentions be effective in our joint activity. Third, we each intend that we each fill in and execute each of our individual plans for participating in the joint activity in ways that mesh with and do not thwart each other).

Philosophers have understandably not been concerned with operationalizing collective intentions and developing hypotheses for empirically testing the relationships of collective intentions to antecedents and consequences. Bagozzi and Dholakia (2002) used two measures of collective intentions, which they termed "we-intentions," in their investigation of virtual chat room participation: "We (i.e., I and the group of online friends that I regularly chat with) intend to chat in the virtual chat room together sometime during the next two weeks" and "I intend that our group (i.e., I and the group of online friends that I regularly chat with) chat in the virtual chat room together sometime during the next two weeks." Although not investigated to date in virtual community research, a third operationalization of collective intentions is possible to study: namely, a member's "I-intention" to do his/her part in a joint action of a virtual community. This latter intention might be termed a group-oriented I-intention to contrast it with both the more common I-intention (i.e., an individual person's intention to perform an individual act by him or herself alone and not as part of a group) and we-intentions and shared intentions, as defined above (see Bagozzi & Lee, 2002).

In sum, collective intentions capture a central aspect of purposive social interactions in virtual communities. Members of the community see themselves as either (a) acting as an agent of the group or as an intrinsic part of the whole group which itself acts or (b) acting as a person contributing individually to a group goal or action. We turn now to an analysis of the antecedents of collective intentions.

ANTECEDENTS OF COLLECTIVE INTENTIONS

Research to date has examined virtual community participation as a function of social and individual motives for acting. Consider first the social determinants of participation. Three categories of social factors have been investigated: compliance (subjective norms), group norms, and social identity.

Subjective norms have long been studied within the context of the theory of planned behavior (Ajzen, 1991) but not until recently in virtual communities. As felt expressions from significant others to act, subjective norms are aspects of interpersonal influence and reflect a type of compliance based on the need for approval. Most virtual communities are characterized by little or no opportunities for direct mediation of rewards or punishments, low barriers to exit from the community, and generally inconspicuous participation in the sense that the ability to monitor other people's compliance is relatively low. As a consequence, subjective normative pressure from other members of the virtual community has generally been found to be absent (Bagozzi & Dholakia, 2002; Bagozzi, Dholakia, & Mookerjee, 2005), or modest (Dholakia et al., 2004). Studies to date have examined the indirect effect of subjective norms on intention mediated by desire (see presentation below on individual reasons for acting). A direction for future research would be to study felt normative pressure to participate or not participate in virtual communities, as a function of expectations of significant others outside the virtual community (e.g., family members, authority figures, close friends, coworkers).

Subjective norms function in interpersonal ways but are not group processes *per se*. A more social type of influence on collective intentions can be found through group norms. Group norms consist of shared values or goals among community members with respect to common expectations of group members' conduct and become internalized as a function of developmental processes, socialization, and role modeling (Bagozzi & Lee, 2002). Group norms have been found to influence collective intentions to participate in the virtual community indirectly through desires (Bagozzi, Dholakia, & Klein Pearo, 2004) and directly (Bagozzi et al., 2005; Dholakia et al., 2004). In addition, group norms have been found to influence social identity with one's virtual community (Dholakia et al., 2004).

Social identity in a virtual community refers to one's self-conception based on membership in the community and has three distinct components (Bagozzi & Lee, 2002; Bergami & Bagozzi, 2000; Dholakia & Bagozzi, 2004). The cognitive component captures the degree of self-awareness of membership in the virtual community and is related to the question, "Who am I?" To the extent that one defines him or herself by group membership, he/she can be said to identify with the group. Cognitive social identity is reflected in both the degree of overlap between one's personal identity and the identity of the virtual community, and the extent of perceived congruence between one's attributes and the attributes of the virtual community. A second component of social identity addresses one's affective commitment to the virtual community. This is typically measured by the degree to which one feels attachment and belongingness to the virtual community. The third component of social identity concerns the positive or negative value connotations attributed to virtual community membership. It is manifest frequently by collective or group-based self-esteem and measured by how important or valued one feels as a member of the virtual community. Social identity in the virtual community has been found to strongly influence the desire to participate in network- and small-group-based communities (Bagozzi & Dholakia, 2002; Bagozzi et al., 2004; Dholakia et al., 2004) but not in collaborative browsing groups (Bagozzi et al., 2005).

Collective intentions to participate in virtual communities are driven most directly by the felt desire to participate in the community. Desire is a motivational state that instigates intentions (and in turn, behavior), and is influenced by reasons to act (Bagozzi, 1992). Either deliberative (i.e., volitive) or automatic (i.e., appetitive), desires serve to transform felt plausible bases for acting into actual motivation to act (Perugini & Bagozzi, 2001). Desires are experienced as felt urges to act. A number of studies show that desires strongly influence the commitment and intention to participate in virtual community activities (Bagozzi & Dholakia, 2002; Bagozzi et al., 2004; Dholakia et al., 2004). Future research should explore the possibility that, in addition to personal desires, shared desires might govern intentional social action in virtual communities.

More distal personal reasons for participating in virtual communities can be found in individual dispositions to act, certain mental states or events, and value perceptions (see Figure 1). Actually, such reasons for acting constitute alternative perspectives on how psychological variables function in virtual communities. That is, what we loosely classify here as individual dispositions or mental states/events refers to general psychological reactions from the literature on attitudes and perceived behavioral control (Ajzen, 1991) and anticipated emotions (Bagozzi, Baumgartner, & Pieters, 1998), which are applied to a particular virtual community experience. Attitudes are evaluative reactions toward participating in the virtual community. Perceived behavioral control refers to how easy or difficult it is to participate or how much confidence one has that he/she can participate. Anticipated (positive and negative) emotions are feelings prefactually tied to successfully or unsuccessfully accomplishing a planned goal of participating in a specific virtual community. All three of the above are reasons for acting. Alternatively, reasons for acting might be represented in more context-specific ways. Value perceptions refer to specific attributes of participation in a virtual community that potentially provide reasons for participation. These include purposive value (i.e., informational and instrumental utility), self-discovery, maintenance of interpersonal connectivity, social enhancement (i.e., increase in status), and entertainment value (e.g., fun and relaxation).

Bagozzi and Dholakia (2002) discovered that positive anticipated emotions influenced the desire to participate in virtual communities (chat rooms), and Bagozzi et al. (2004) found that both positive and negative emotions influenced the desire to participate in (small-group- and network-based) virtual communities. The effects of attitudes on desire have been mixed, with one study showing no effect (Bagozzi & Dholakia, 2002), and one study revealing an effect for small-group-based, but not network-based, virtual communities (Bagozzi et al., 2004). Neither of the above mentioned studies examined the effects of attitudes and anticipated emotions on group norms and social identity. Perceived behavioral control has been found to have mixed effects. Bagozzi and Dholakia (2002) discovered no effects for perceived behavioral control on either desires or intentions. Dholakia et al. (2004) revealed that perceived behavioral control influenced desires for network-based virtual communities but not small-group based virtual communities. They also found that perceived behavioral control had direct effects on intentions for both network-based and small-group based virtual communities (not shown in Figure 1). Bagozzi et al. (2005) showed that perceived behavioral control influences intentions to participate in collaborative browsing but not chat-rooms for Americans, while perceived behavioral control influenced intentions to participate in chat-rooms for Indians. In the only study found that examined value perceptions, Dholakia et al. (2004) showed that purposive value affects both group norms and social identity, self-discovery influences group norms, and entertainment value determines social identity. Since group norms and social identity ultimately had effects on desires, the influence of value perceptions on collective intentions is mediated through desires. Dholakia et al. (2004) also found that entertainment value directly affects participation in virtual communities. Finally, within the context of the theory of planned behavior (Ajzen, 1991), Bagozzi et al. (2005) discovered that attitudes influence intentions to participate in collaborative browsing and chat-room groups.

CONSEQUENCES OF COLLECTIVE INTENTIONS

The primary outcome of collective intentions with respect to participation in virtual communities is, of course, actual engagement with the community. Only three studies have investigated participation behavior in virtual communities as a function of collective intentions. Bagozzi et al. (2004) found that collective intentions influence actual participation for both network-based and small-group-based virtual communities. Likewise, Dholakia et al. (2004) showed that collective intentions influence participation behavior in their study of virtual communities, and Bagozzi et al. (2005) found similar results for both collaborative browsing and chat-room groups.

What about second-order effects of involvement with virtual communities? Bagozzi et al. (2004) examined the effects of collective intentions to engage in virtual community activities on off-line behavioral outcomes. For small-group-based virtual communities, the stronger the intention to participate in these communities, the less telephone and radio were used and the more people actually engaged in reading books. For network-based virtual communities, the stronger the collective participation intention, the less television was watched, and the more people interacted with family members, friends, and neighbors, and the more they engaged in such hobbies as sports teams and book clubs.

A neglected area of research is the study of the effects of virtual community participation on general well-being. While some commentators have warned about such negative effects of participation on the Internet as alienation, disembodiment, and depression (e.g., Dreyfus, 2001; Kraut et al., 1998), negative outcomes are not inevitable. Researchers need to focus on what happens in one's personal and social life on and off the Internet and how the two modes of experience relate to each other and one's well-being. The Internet, in general, and participation in virtual communities, in particular, can enhance human well-being under certain conditions. This might happen when social identity and personal identity are fostered and people give and receive in meaningful ways to the relationships within which they are engaged. Moreover, virtual communities and the members within them may even participate in fulfilling outreach activities, altruistic acts, group citizenship behaviors (Bergami & Bagozzi, 2000), and charity work, which also potentially adds to human flourishing.

CONCLUSION

A key to the formation and ongoing vitality of virtual communities is the involvement and commitment of people to the community and the challenge and joy they receive from participation. Human emotion plays an essential part in the life of virtual communities and their relationship to its members. Figure 1 summarizes places in collective intentional action system where this happens. First, positive and negative anticipated emotions are tied to the fulfillment or thwarting of virtual community participation as a personal goal and thus function to orient and motivate the individual to act in a purposive way toward one's virtual community. Second, rational, emotional, and social reasons for participating in virtual communities are integrated and transformed through formation of desires to participate or not. A third way that emotions appear in the collective intentional action system is in affective commitment (i.e., feelings of attachment and belongingness to the community). People potentially receive happiness and satisfaction from identification with and participation in their virtual communities, as well as give back emotion, time, and effort to the communities and people in it. Finally, emotions reside in general well-being that is psychologically and socially constructed as people reflect upon and discuss their experiences in their virtual communities with members and others outside the community. A full range of individual and social emotions are to be expected in virtual communities, depending on the level of involvement and commitment demanded: empathy, trust, affection, pride, joy, happiness, guilt, shame, embarrassment, frustration, disgust, anxiety, envy, and jealousy. Obviously, these positive and negative emotions can make participation in a virtual community a rewarding/punishing or pleasant/unpleasant experience, depending on the nature of the experience and its meaning for the person embedded in it.

REFERENCES

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50, 179-211.

Bagozzi, R.P. (1992). The self-regulation of attitudes, intentions, and behaviors. *Social Psychology Quarterly*, 55, 178-204.

Bagozzi, R.P. (2000). On the concept of intentional social action in consumer behavior. *Journal of Consumer Research*, 27, 388-396.

Bagozzi, R.P., & Dholakia, U.M. (2002). Intentional social action in virtual communities. *Journal of Interactive Marketing*, 16, 2-21.

Bagozzi, R.P., & Lee, K-H. (2002). Multiple routes for social influence: The role of compliance, internalization, and social identity. *Social Psychology Quarterly*, 65, 226-247.

Bagozzi, R.P., Baumgartner, H., & Pieters, R. (1998). Goal-directed emotions. *Cognition and Emotion*, 12, 1-26.

Bagozzi, R.P., Dholakia, U.M., & Klein Pearo, L.R. (2004). Antecedents and consequences of online social interactions. Unpublished working paper, Rice University.

Bagozzi, R.P., Dholakia, U.M., & Mookerjee, A. (2005). Individual and group-bases of social influence: Interactions in digital environments. *Media Psychology*, in press.

Bergami, M., & Bagozzi, R.P. (2000). Self-categorization, affective commitment, and group self-esteem as distinct aspects of social identity in the organization. *British Journal of Social Psychology*, 39, 555-577.

Bratman, M.E. (1999). *Faces of intention: Selected essays on intention and agency*. Cambridge: Cambridge University Press.

Dholakia, U.M., & Bagozzi, R.P. (2004). Motivational Antecedents, Constituents, and Consequences of Virtual Community Identity. In S. Godar & S. Pixie-Ferris (Eds.), *Virtual and Collaborative Teams: Process, Technologies, and Practice* (pp. 253-268). Hershey, PA: Idea Group Publishing.

Dholakia, U.M., Bagozzi, R.P., & Klein Pearo, L.R. (2004). A social influence model of consumer participation in network- and small-group-based virtual communities. *International Journal of Research in Marketing, 21*, 241-263.

Dreyfus, H.L. (2001). *On the internet*. London: Routledge.

Gilbert, M. (1992). *On social facts*. Princeton, NJ: Princeton University Press.

Kraut, R., Kiesler, S., Boneva, B., Cummings, J., Helgeson, V., & Crawford, A. (2002). Internet paradox revisited. *Journal of Social Issues*, 58, 49-74.

Kraut, R., Patterson, M., Lundmark, V., Kiesler, S. Mukophadhyay, T., & Scherlis, W. (1998). Internet paradox: A social technology that reduces social involvement and psychological well-being? *American Psychologist*, 53, 1017-1031.

Perugini, M., & Bagozzi, R.P. (2001). The role of desires and anticipated emotions in goal-directed behaviors: Broadening and deepening the theory of planned behavior. *British Journal of Social Psychology*, 40, 79-98.

Tuomela, R. (1995). *The importance of us: A philosophical study of basic social notions*. Stanford, CA: Stanford University Press.

UCLA Internet Report: Surviving the digital future year three. (2003, January). University of California, Los Angeles, Center for Communication Project.

KEY TERMS

Collaborative Browsing: Online navigation on the Internet by a group of individuals that assembles in the same physical setting and browses online using the same or adjacent interfaces in the attainment of task-oriented objectives.

Compliance: Social influence process whereby a person conforms to the expectations of others based upon the motivation of a need for approval. Also known as subjective norm.

Collective Intentional Action: Mutual or joint behaviors performed by a group of persons and that are explained by collective intentions (e.g., we-intentions) and such social influence as compliance, group norms, and social identity and such personal mental states as anticipated emotions, desires, and value perceptions (e.g., purposive value, self-discovery, maintenance or personal connectivity, social enhancement, and entertainment value).

Network-Based Virtual Community: A specialized, geographically dispersed virtual community based on a structured, relatively sparse and dynamic pattern of relationships among participants sharing a common focus.

Small-Group-Based Virtual Community: A handful or so of persons with a dense web of relationships, interacting together online in order to accomplish a wide range of jointly conceived goals and to maintain the functioning of the group.

Social Identity: A conception of the self consisting of (a) self-awareness of membership in a specific group, (b) emotional attachment or involvement with the group, and (c) a positive or negative value connotation attached to this group membership (i.e., group-based or collective self-esteem). These are sometimes referred to, respectively, as the cognitive, affective, and evaluative components of social identity.

Subjective Norms: Social influence process whereby a person acts on the basis of internalized values or goals shared with members of a reference or friendship group. Similar to compliance.

Virtual Communities: Mediated social spaces in the online environment that allow collections of people to form and be sustained through ongoing communication processes.

We-Intention: Decision of a person to participate in joint action and which involves an implicit or explicit agreement between the participants to engage in the joint action.

Combining Local and Global Expertise in Services

Hannu Salmela
Turku School of Economics and Business Administration, Finland

Juha Pärnistö
Fujitsu Services, Finland

INTRODUCTION

Since the 1990s, services characterized by a considerable geographical distance between the service person and the customer have become increasingly commonplace. Banks and insurance companies are introducing call centers or service centers to complement, or even replace, the old regional service organization. In the information and communication technology (ICT) sector, companies such as Fujitsu and IBM provide part of the end-user support for their clients from a few centralized call centers. Telecommunications operators have established call centers to serve their customers in conducting basic business transactions. To a large extent, the change in the 1990s can be attributed to ICT development. As call centers and local offices have equal access to all the information, many of the services that previously had to be provided locally can now come from a call center. Furthermore, this decade will bring new technologies that will further enhance capabilities to serve customers over long distances. They will, for instance, provide increasingly rich media for interaction between customers and remote service personnel.

This article investigates factors that need to be considered when moving service production from regional offices to service centers. The empirical part of the study comprises a longitudinal analysis of the ways how Fujitsu Invia, a European IS company within Fujitsu Group, has transformed its service organization. The company has moved a long way from local, site-specific service units to national service centers, and ultimately to a few global centers that provide services to thousands of computer users worldwide. In retrospect, it can be said that the decision to centralize service production turned out to be very successful. However, the reasons why Fujitsu Invia decided to return part of the end-user support closer to customer sites illustrates the complexities associated with centralizing services that were previously produced locally.

BACKGROUND

The ability to centralize services appears to provide a cure for some of the traditional problems of service organizations. Managers of distributed service organizations are painfully aware of the difficulties to maintain an equal level of knowledge among all individual service persons in all regional offices. Centralizing the services to a call center or service center seems like an easy solution for ensuring that all customers receive equal service.

The more complex the services are, the more difficult it becomes to maintain equal knowledge in all regional offices. Hence, the analysis of forces for specialization among service staff is one of the key issues when considering the potential advantages of centralizing service production. Because of the special expertise necessary to solve the specific problems they encounter, professional service providers need a high level of specialization (Koelemeijer & Vriens, 1998). Factors that increase service complexity and thus pressure for specialization include (Mäkelin & Vepsäläinen, 1989), for instance,

1. diversity in customer needs and requests,
2. variety in the services available,
3. the number of situational factors that need to be considered,
4. uncertainty related to customer needs and circumstances,
5. the ability of the customer to define the services, and
6. the complexity of contracts used for governing the transactions.

In essence, complexity makes it difficult for a generalist service person to be able to handle all possible inquiries from all customers adequately. While generalists can deal with routine cases, specialists are needed to handle the difficult and unique cases. The main problem for producing complex services in regional offices is that it is difficult to maintain highly specialized knowledge in every regional office. The most forceful argument for establishing a service center is that the customer with a unique problem can talk with a global specialist rather than with a local generalist. In addition, the cost savings that can be achieved in regional offices (office space, service personnel) are often sufficient to make the projects acceptable in terms of financial profitability measures.

In this chapter we suggest, however, that managers should also pay considerable attention to the opposite forces as well, that is, forces for providing local service. An obvious reason for providing services locally is that the service has a physical component and thus requires presence close to the customer. There are, however, many soft issues that may also make the customers prefer a local and personal service. In face-to-face discussions, information can be communicated with multiple cues like the sound of voice, facial expressions, and body language. Thus, the richness of communication media is very high (Daft & Lengel, 1986; Huang, Watson, & Wei, 1998). In this respect, the need to rely on conversations over phone or e-mail may have a negative impact on the quality of service.

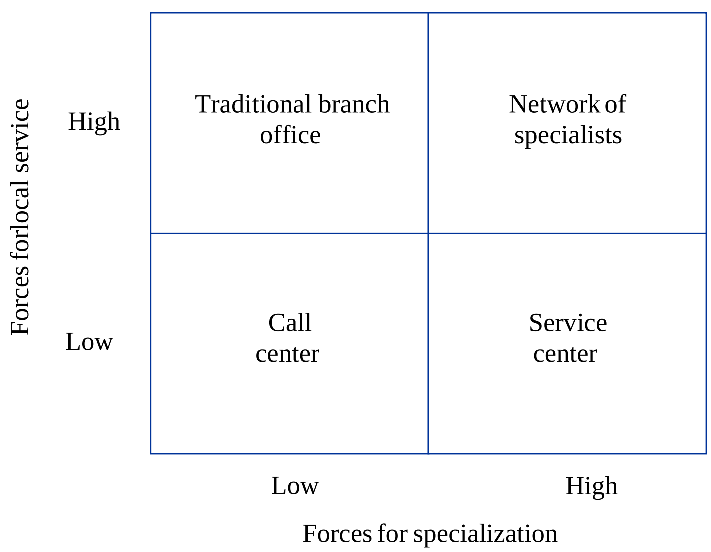
A TYPOLOGY OF SERVICE ORGANIZATIONS

The typology that is suggested in this article is based on forces for providing local service and forces for specialization among service staff. Figure 1 provides a typology of four ideal types of service organizations, each designed to take into account the particular nature of service situations. Each organizational type has its strengths as well as its typical ways of using technology. More often than not, service organizations provide a mix of services for their customers. Some of the services provided may require local presence and some of them specialization. Thus, the objective is not to locate the whole company to a single quadrant. Rather, it is essential to identify all services that are provided to customers and to locate each of them to the right quadrant.

A *traditional branch office* is best suited to relatively simple service situations that require local presence. In simple services, the local generalist service persons are able to provide sufficient service quality. The quality and profitability of service is ensured by replicating and controlling precisely defined activity cycles, personal selling approaches, inventory control patterns, and counter display techniques (Quinn & Paquette, 1990). The possibilities for more stringent specialization are limited because a relatively small number of service personnel in one office have to deal with a full variety of customer inquiries.

In a *call center*, service personnel are centralized to one physical location. Part of the service personnel is moved from regional offices to a physical location to answer a definite set of requests from customers. The change is usually motivated by a managers' observation that some of the simple services are such that customers don't really expect to be served locally. The objectives for

Figure 1. Simple typology of service organizations



establishing a call center are a reduced cost of service, extended contact hours, and standardized quality. As the services provided are fairly simple, there is no need for specialization between service personnel: Any service person can solve any incoming customer inquiry.

A *service center* type of organization takes advantage of the fact that one centralized unit handling inquiries from a large number of customers allows greater specialization of service personnel. Each service person specializes in a particular type of customer problem, and incoming customer inquiries are routed to the best specialist in that area. The assumption is that a specialist with experience in solving similar problems with many companies can provide a more accurate service, particularly with regard to complex customer problems.

A *differentiated network* is the most challenging type of service organization as the service should be simultaneously local and specialized. Neither a local service office nor a centralized service center alone can provide the service. Finding a perfect organization in this quadrant is difficult. Often, organizations in this quadrant rely on a combination of traditional branch offices and centralized service centers (which deal with more complex cases). The potential for using sophisticated information and communication technologies is clearly highest in a differentiated network.

Even if the differentiated network may seem like a less attractive organizational solution, one should keep in mind that in service situations requiring both a local presence and specialization, there may be no alternative. In this situation, a service center may not be able to respond to local needs.

In fact, writings about multinational manufacturing enterprises often favor a differentiated network, which allows local responsiveness in different areas but possesses the control mechanisms to ensure global integration (Bartlett & Ghoshal, 1998; Nohria & Ghoshal, 1997; Prahalad & Doz, 1987). Rather than choosing a fully centralized or decentralized structure, the assets and resources of the company are widely dispersed but mutually supportive to achieve global-scale efficiency, the roles and responsibilities of organizational units are differentiated but interdependent to maximize national flexibility, and its knowledge and initiatives are linked through a worldwide learning capability that assures the efficient development and diffusion of innovations (Bartlett & Ghoshal).

With such practices, the organization provides a context for employees to create and share both tacit and explicit knowledge (Nonaka & Takeuchi, 1995; Polanyi, 1966). The relationships between individual actors are seen as a form of social capital that constitutes a valuable resource (Castells, 1996; Nahapiet & Ghoshal, 1998). In essence, it is asserted that creating and sharing knowledge in a geographically distributed organization is not only possible, but it may even be more effective than knowledge creation in one centralized point.

It seems apparent that many service organizations could benefit from similar practices. In fact, predictions of the impact of IT on service organizations have discussed the “spider’s web” type of organization, where local offices are independent but are also able to use each other’s knowledge resources as and when needed (Quinn & Paquette, 1990). Similarly, a recent study concluded that the competence to do global product development is both collective and distributed (Orlikowski, 2002).

DEVELOPMENT OF SERVICES IN FUJITSU INVIA

Fujitsu Invia (formerly ICL Invia) is a North-European vendor of information technology products and services. It belongs to Fujitsu Group, which operates in more than 100 countries and has almost 200,000 employees. Because the case describes developments mainly in the Fujitsu Invia, this brand name will be used below hereafter. According to a recently launched strategy, Fujitsu Invia has two main businesses: solutions and services. Solutions business refers to consulting, designing, and constructing information systems. Services business mainly involves operating customer’s IT infrastructure as well as business applications. The case description focuses on service business and its evolution during the past few years.

Stage 1: All Services Produced Locally

Only a few years ago, Fujitsu Invia was mainly a hardware vendor. The main service was naturally hardware maintenance, which is typically a local service. The customers were provided with a service number to which they could call and report a hardware failure. The maintenance organization also delivered large installation projects, but not actual end-user support or administrative services (e.g., opening of a new e-mail address). These were typically delivered by the customer’s own IT department. Maintenance was organized in regional business units that were quite independent. Customer-call management and the information system for managing service processes were typically shared by business units on a national level. Service coordination between countries was not really required, although certain issues such as pricing and service levels were negotiated on a corporate level.

Stage 2: Initiation of Centralized Service Production

In the second stage, customers became interested in acquiring relatively large service packages so that one vendor takes responsibility for a relatively large and logical part of the IT services. The change was initiated by customers who found managing IT and particularly the IT infrastructure increasingly difficult. These packages typically include a selection of IS services, for example (Weill & Broadbent, 1998),

- on-site support (i.e., solving end users' problems locally),
- help desk (i.e., a contact point where end users' requests are registered and escalated; also attempts to solve requests),
- hardware maintenance (i.e., fixing broken IT equipment),
- systems management (i.e., monitoring and operating systems typically from a centralized service center), and
- service management (i.e., customer reporting, maintaining contracts, improving service quality).

It was not possible to deliver these services by a maintenance organization, so in order to respond to this request, Fujitsu Invia established new business units. Centralized services (mainly help desk and systems management) are used to support the IT infrastructure. Centralization seemed like an obvious solution because of cost-effectiveness requirements and also because the technical expertise is easier to acquire and maintain in specialized teams. The aim was to minimize local service. Maintenance is the only local service. New technologies, like remote management, further improved the quality of service as end users no longer needed to describe the problems in detail. The help-desk specialist simply opened a remote connection to the end user's workstation, diagnosed the problem, and changed the parameters or trained the end user.

Stage 3: End-User Services Return to be Produced Locally

Although a centralized service center is probably the most cost-effective way to organize this type of services, there are also some problems. Customer organizations are not standardized really well even though the technological solutions are. Information systems can be installed in many ways, a situation that has implications on their operation. The quality of service perceived by end users is an even more significant issue. It seems that people prefer to be served locally. The barrier to call to a distant help desk and to an unknown specialist is rather formidable, even within a single nation. Furthermore, transferring the necessary knowledge about local environments to a help desk is not easy. Therefore, a local service organization is frequently required to manage situations like this. The end user easily feels that he or she has to call a help desk, which mainly escalates problems to a local organization. The added value of the service can be questioned.

Because of the problems described above, the help desk has been reorganized in some customer cases so that the first line is either integrated with a local service organization, or the help-desk first line is actually on the customer site. Some members of the on-site support team receive customer service requests, and other members undertake installations and problem solving. The roles within the team can be changed, which means that the on-site support team knows the local customer environment very well, and also the end users know on-site support people. This makes communication easier and the barrier to calling is lower. The solution rate has improved significantly in the pilot cases (in one case, from a level of 50% to 80%), which makes it possible to deliver the same service with fewer resources. At the same time, customer satisfaction improved (in the example case, from 6.5 to 8.4 when measured on a 1 to 10 scale). Their second-line and third-line support is still delivered from a centralized service center during the night and weekends.

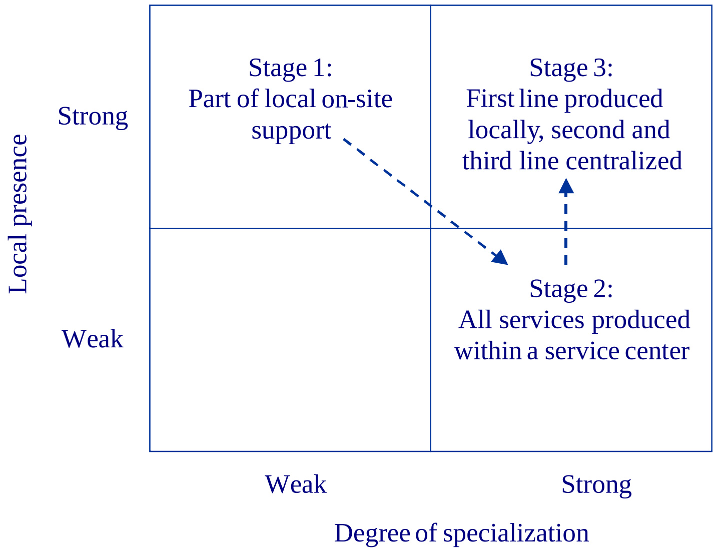
Analysis of the Nature of Services

The most problematic service for Fujitsu Invia has been how help-desk services should be organized. What makes it problematic is that on the one hand, end users in customer companies prefer to contact local service persons instead of a distant call center. On the other hand, solving the problems often requires a specialized knowledge that the local service persons lack. The fact that the service should be both simultaneously local and also specialized has made the service difficult to organize, and the attempt to find the best means of organization has been a trial-and-error process (Figure 2).

In the first stage, the idea of providing help-desk services to end users was more or less informal. While end-user satisfaction for services was probably high, the customers felt that the cost of producing services locally was too high. Outsourcing these services to Fujitsu Invia and simultaneously centralizing them was considered an interesting option. The new technologies appeared to provide excellent tools for centralized production of help-desk services. When the service was centralized, however, more was learned about its nature. Service turned out to be more local than what the customers and Fujitsu

C

Figure 2. Organizational arrangements for help-desk services in different stages



Invia had expected. In the third stage, help-desk services for large-site customers are based on a combination of local and centralized services. While the experiences from this organizational form are highly positive, it is also more costly and can be applied only to customers with relatively large sites.

FUTURE TRENDS

The sociological theory of the postindustrial society was elaborated more than 3 decades ago (Bell, 1973; Touraine, 1969). In essence, the theory combined three postulates (Castells & Aoyama, 1994): (a) The source of productivity and growth lies in the generation of knowledge, (b) economic activity shifts from goods production to services delivery, and (c) the importance of occupations with a high information and knowledge content will increase.

Existing postindustrial societies do not fully reflect all predictions made within this theory (Castells, 1996; Castells & Aoyama, 1994). Nevertheless, the significance of various types of services, for example, capital management services, services produced for industrial companies, and social and health services, is increasing. As firms in advanced economies have off-shored manufacturing jobs, the society as a whole has transformed toward a "service economy" model (Castells & Aoyama). Hence, the significance of services in general and knowledge-intensive services in particular is increasing.

The point raised in this paper is that information and communication technology will also induce a major transformation inside the service sector. Increasing telecommunications bandwidth together with its falling cost will enable new applications that are particularly useful for geographically dispersed production of services. In the future, the possibility to use wide computer screens for face-to-face interaction is likely to have a major impact on the way services are organized. Furthermore, the penetration of innovations such as integrated customer databases, e-services, global scheduling systems, call center technology, and virtual learning environments is likely to continue.

The emergence of call centers and service centers provides first visible evidence that the transformation is already taking place. However, as the Fujitsu Invia case illustrates, a call center can be too remote for providing services that require local presence. Hence, the organizational forms required in many services may be more complex than mere centralization of service production. The pressures for local responsiveness are likely to force service organizations to adopt similar complex network structures as some of the advanced manufacturing companies are already using.

This research was, however, based on only one project in one service organization. As service organizations both in the private and public sector have been slow in adopting new technologies, previous IS research has paid relatively little attention to these sectors. The need for further research about the ways in which organizational structures, processes, and information infrastructures are being transformed is evident.

CONCLUSION

This paper provides a simple typology of service situations that can be used when planning IT investments in service organizations. The use of the typology is illustrated through a longitudinal analysis of the ways in which Fujitsu Invia, a European IS service company, has transformed its service organization to better meet customer expectations and needs. Practicing managers can use the typology to analyze organizations' services, to investigate where services ought to be produced, and how ICT technology should be employed to support customer service situations. The authors hope that this article will be useful for people working in companies either providing a service business or aiming to enter a service business in an international context. For research, the typology, its dimensions, and the background theories identified provide some early steps on the way toward a more comprehensive understanding about the impact of ICT on service organizations.

REFERENCES

Bartlett, C. A., & Ghoshal, S. (1998). *Managing across borders: The transnational solution* (2nd ed.). Boston: Harvard Business School Press.

Bell, D. (1973). *The coming of postindustrial society: A venture in social forecasting.* New York: Basic Books.

Castells, M. (1996). *The rise of the network society*. Malden, MA: Blackwell Publishers Inc.

Castells, M., & Aoyama, Y. (1994). Paths towards the informational society: Employment structure in G-7 countries, 1920-90. *International Labour Review, 133*(1), 5-31.

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science*, *32*(5), 554-571.

Huang, W., Watson, R. T., & Wei, K. (1998). Can a lean e-mail medium be used for rich communication? *European Journal of Information Systems*, *7*, 269-274.

Huber, G. P. (1991). Organizational learning: The contributing processes and the literatures. *Organization Science, 2*, 88-115.

Koelemeijer, K., & Vriens, M. (1998). The professional services consumer. In M. Gabbott & G. Hogg (Eds.), *Consumers and services* (pp. 163-184). Chichester, England: John Wiley & Sons.

Mäkelin, M., & Vepsäläinen, A. (1989). *Palvelustrategiat: Palveluorganisaation kehittäminen ja tietotekniikka*. Jyväskylä: Gummerus kirjapaino Oy, Finland.

Nahapiet, J., & Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage. *Academy of Management Review*, *23*(2), 242-266.

Nohria, N., & Ghoshal, S. (1997). *The differentiated network: Organizing multinational corporations for value creation*. San Francisco: Jossey-Bass Inc.

Nonaka, I., & Takeuchi, H. (1995). *The knowledge creating company: How Japanese companies create the dynamics of innovation*. Oxford, England: Oxford University Press.

Orlikowski, W. J. (2002). Knowing in practice: Enacting a collective capability in distributed organizing. *Organization Science*, *13*(3), 249-273.

Polanyi, M. (1966). *The tacit dimension*. London: Routledge & Kegan Paul.

Prahalad, C. K., & Doz, Y. L. (1987). *The multinational mission: Balancing local demands and global vision*. New York: Free Press.

Quinn, J. B., & Paquette, P. C. (1990). Technology in services: Creating organizational revolutions. *Sloan Management Review, 33*(2), 67-78.

Sanchez, R., Heene, A., & Thomas, H. (1996). Introduction: Towards the theory and practice of competence-based competition. In R. Sanchez, A. Heene, & H. Thomas (Eds.), *Dynamics of competence based competition: Theory and practice in the new strategic management* (pp. 1-35). Exeter: Pergamon.

Storbacka, K., Strandvik, T., & Grönroos, C. (1994). Managing customer relationships for profit: The dynamics of relationship quality. *International Journal of Service Industry Management*, *5*(5), 21-38.

Touraine, A. (1969). *La societe post-industrielle.* Paris: Denoel.

Weill, P., & Broadbent, M. (1998). *Leveraging the new infrastructure: How market leaders capitalize on information technology*. Boston: Harvard University Press.

KEY TERMS

Capabilities: Are repeatable patterns of action in the use of assets to create, produce, and/or offer products to a market (Sanchez et al., 1996).

Competence: An ability to sustain the coordinated deployment of assets in a way that helps a firm to achieve its goals (ability here is used in the ordinary language meaning "of a power to do something"; Sanchez, Heene, & Thomas, 1996).

Information: Refers to data that give meaning by reducing ambiguity, equivocality, or uncertainty, or data that indicate that conditions are not presupposed (Huber, 1991).

Knowledge: Refers to interpretations of information, know-how, and beliefs about cause-effect relationships (Huber, 1991).

An entity **learns:** If, through its processing of information, the range of its potential behaviors is changed. The information processing can involve acquiring, distributing, or interpreting information (Huber, 1991).

Perceived Service Quality: Refers to customers' cognitive evaluation of the service across episodes compared with some explicit or implicit comparison standard (Storbacka, Strandvik, & Grönroos, 1994).

Skill: Understood as a special form of capability, with the connotation of a rather specific capability useful in a specialized situation or related to the use of a specialized asset (Sanchez et al., 1996).

Communication Management for Large Modules

Elayne Coakes
University of Westminster, UK

Dianne Willis
Leeds Metropolitan University, UK

INTRODUCTION

This article addresses issues concerning the suitability of particular media as mass communication tools in an UK higher education setting. It looks firstly at the use of e-mail as a communication method whilst managing large modules. The article then goes on to examine the use of a Virtual Learning Environment (VLE) to provide a mass communication method more suited to the needs of both the staff and the students.

An in-depth case study will look at how a large module is managed through the use of a Virtual Learning Environment, discussing the pros and cons of using such technology. For the purposes of this article, distance learning will not be considered, as the university in question has no plans to follow or adopt such an approach for the courses that undertake these modules.

BACKGROUND TO THE STUDY

The University of Westminster is a large, modern university operating on four sites across the centre of London and West London. It was Britain's first polytechnic, established in 1838 in Regent Street, and it still uses these original buildings for lectures and seminars. It became a university in 1992 and is now recognised in the *Financial Times* annual survey as the top English 'new' university.

The university has a population of both full-time and part-time students drawn from the local surrounding areas, but also has a large population of international students. Many of these students are studying post-graduate or post-experience/professional updating courses.

The Westminster Business School operates out of the Marylebone campus, which is situated close to Regents Park and is thus a very popular location for students to attend courses, as not only is it in the centre of London, but is also easily accessible by public transport.

COMMUNICATION MEDIA

This section discusses the theoretical background to communication studies looking at 'fitness for purpose' for the media.

To begin, a brief overview of the communication process and its constituents would be useful. A basic theoretical model of the communication process states that messages are 'sent' and 'received', confirmation of receipt and interpretation of the message indicates that it is a two-way communication process (Warner, 1996). The efficiency of this process is subject to a variety of issues such as:

- *Effectiveness:* How do we measure this when using e-mail or a Web site?
- *Simultaneous reception of information by recipients:* The information is available simultaneously, but how do we measure if students have accessed it?
- *Acknowledgement of receipt:* With e-mail, a receipt can be provided, but how is this addressed using a Web site?
- *Speed:* How quickly does the information reach the recipients?
- *Cost of the process:* Is it cost effective?

All of these factors will be considered in the case study section of the article.

It can be said that one of the major benefits of the growth in the use of the Internet has been the ability to communicate using e-mail as a medium. E-mail is still the most regularly used aspect of the Internet, and the level of usage is growing. Sands (2003) argues that e-mail is now as accepted a communication method as the telephone within just a few years of its introduction. E-mail has become a common mode of communication for many people, though exceptions must be made for those who do not have easy and regular access to the technology required. Adequate access to the technology is assumed in this article. However, communications theorists argue there are problems with the use of e-mail as the sole communication medium. Culnan and Markus (1987) suggested that a lack of face-to-face communication changes the intra- and interpersonal variables because of a lack of social context. Sproull and Kiesler (1986) argued that e-mail was devoid of social cues, and this would seriously affect communication patterns. E-mail provides neither audible nor visible cues to the communication process and as such can be seen as a relatively impoverished communication style. Hirschheim (1985) argued that one of the major benefits of using e-mail is to support communication between people who are geographically distant. This leads us to the initial premise that although widely utilised, there are significant problems with the use of e-mail as the sole communication medium.

There are further problems with an e-mail system in that it is not always possible to ensure that the recipient has received and read the message in an appropriate timeframe. This may be caused by a variety of problems, both technical and human. These issues are discussed in more detail and possible solutions posited in Willis and Coakes (2000).

Further analysis of the current situation indicates that e-mail lacks the collaborative dimension that is needed in today's world. In an academic setting, the ability to discuss matters in 'real time', rather than waiting for a reply from a tutor via an e-mail system, may be of great importance.

IS E-MAIL A SUITABLE MASS COMMUNICATION MEDIUM?

A traditional view of the mass communication process would have the university providing content and it passing to the students via the e-mail medium as shown in Figure 1.

This model allows no interaction. We posit that what is required in the modern communication environment is a more interactive model which allows students to influence the content, as well as be recipients of, information. In our article, the aim of mass communication is to provide information simultaneously to a large number of students. We have to seriously question whether e-mail is the correct option given the problems outlined.

Further analysis leads us to the model suggested below which, while offering limited interaction, we feel is likely to be more effective and more satisfactory than the traditional model outlined above.

The model outlined as Figure 2 gives students the option of simply gaining information via the medium, or engaging and interacting with it by influencing the content. From the preceding arguments, it is doubtful that e-mail can fulfil the role of the appropriate medium in the new model, so an alternative has to be found.

The ability to interact is an aspect that has acquired great importance in the communication process, and a method must be found that does not have the weaknesses of e-mail. What is required is a medium that can have input from both staff and students, and which can change rapidly to respond to these needs. As the staff and students are already familiar with e-mail and the Internet, a move to a more directly Web-based approach using a VLE is likely to provide a viable solution. This is supported by the fact that students have better access to Web sites and are more used to using this medium now than in the past. Additionally, many organisations limit the size of e-mails that can be sent across the network, which can cause real problems when disseminating large amounts of information, which can cause issues with part-time students who use their workplace e-mail systems for their prime communication access. The authors feel the use of a Web site accessible from any location and requiring only a browser offers a realistic alternative.

In terms of the ability to interact, asking students to access a Web site is not necessarily more effective than the simple use of e-mail - it is the add-ons of lecture notes, seminar notes, and feedback that will make the medium more successful.

Figure 1. A traditional mass communication model

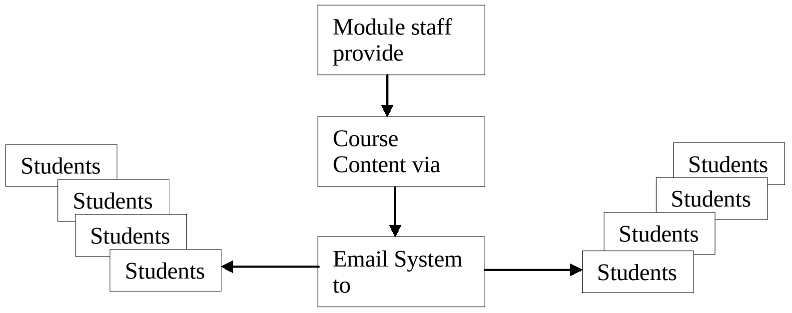
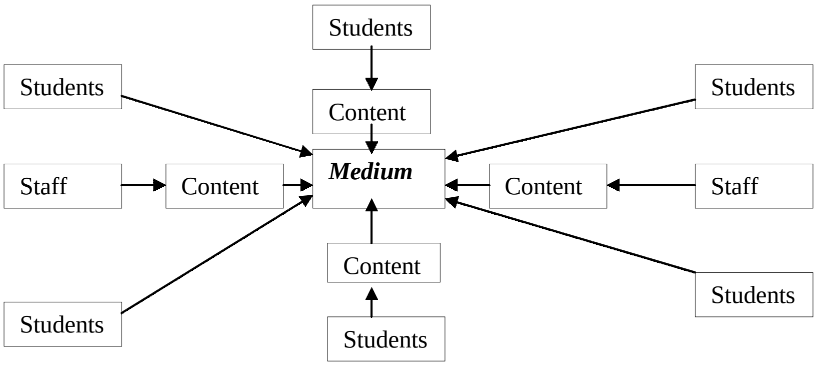


Figure 2. An interactive mass communication model (Adapted from Hoffman, Novak & Chatterjee, 2000)



Kling (1996) looked at the benefits of electronic newsgroups, bulletin boards, conferences, and distribution lists, and the ways in which these change the communication model. There is some evidence to suggest that electronic communities help to foster a sense of community amongst geographically isolated people, and this would be of obvious benefit to the groups

being studied, as they are unable to meet physically, but may wish to feel part of a group (Kling, 1996). Therefore, it can be seen that in order to provide a sense of community for the large numbers involved in the case study, some communication medium needs to be added to foster a sense of group identity. As described below the VLE provides these capabilities.

ARE VLEs SUITABLE AS A MASS COMMUNICATION MEDIUM?

Much of the literature on VLEs is written from the point of view of their use as a teaching tool, considering what adjustments need to be made in order to present material in an online environment (Collins, 2000; Johnson, Aragon, Shatk & Palma-Rivas, 2000; Mishra, 2002). In most cases, the use of a VLE as an educational tool is not separated from its use as a communication tool; communication is one aspect of the range of options provided. This study comments on the use of a VLE as the definitive source of information for a particular module, and contrasts the capabilities for communication and tracking of student engagement. Work by Kekkonen-Moneta and Moneta (2002) indicates results that are better than traditional classroom teaching when a VLE is used. However, Mishra's (2002) work indicates that new possibilities in terms of the learning experience online are not being addressed in most cases.

This study focuses on the benefits to students of having one source for all information and the advantages this gives over multiple source enquiry points. According to Barbera (2004) the main issue is quality and how this is addressed in a virtual learning environment. VLEs tend to focus on technological and aesthetic criteria over educational ones, and there is often confusion between their use as a mere supply of information and their capabilities for knowledge-building processes. Quality will be seen as not merely offering Web pages full of information, but in establishing filters to ensure materials are reliable and worthy.

A further dimension of VLEs is their humanism (McFazdean, 2001). The instructor is no longer the orator of information, but needs to encourage learners to feed their own curiosity—that is, 'spoon-feeding' is inappropriate in this environment. Guidelines for a productive VLE include maintaining a non-authoritarian style and explicitly stating what is required of the students at all times. A further issue raised by McFazdean (2001) is that of the capability of the technology to support the demands of the group; aspects such as not owning a powerful enough computer to run the software and poor levels of computer literacy form technical barriers to the successful use of the learning environment.

A similar use of VLEs was undertaken at the University of Hull (Pennie, Barnett, Chin & Dolphin, 2001) which focuses on the experiences there of migrating two modules to a VLE, in response to increasing numbers of full-time and part-time students. The VLE approach was selected to cope with rapid change in online information resources, coupled with the need to make students more aware of what resources are best for them and how to use them effectively.

THE CASE STUDY

Strategies for Information Management Module

This is a large postgraduate module that is compulsory for students on two master's programmes and available for students on other master's programmes as an option.

Originally this module had a Web site that was used as a method of ensuring that communication is kept consistent and constantly available, but was in addition used to provide additional teaching materials around and between formal lectures and seminars (Coakes & Willis, 2002). Students on this module are assertive about what they want from a Web site, and have been instrumental in suggesting changes and additions to content and layout; as a result the standard Web site was recently replaced by a Virtual Learning Environment.

The VLE (BlackBoard™ is a standard, pre-programmed Web site brought in from a commercial supplier. It is a Web-based server software platform for course management that is gradually replacing lecturer-built Web sites across the university, and has a number of advantages and disadvantages over the standard (FrontPage) designed Web site. These are listed below.

A 2004 survey of 1,000 students at the university showed that modules using a VLE produced a more positive view (of the module) from the students; in particular, enhanced communication was considered a positive outcome as was the ability to retrieve information and announcements.

PROS AND CONS OF USING A WEB SITE FOR DISSEMINATING INFORMATION

The major pro in the use of this type of communication is the secure knowledge that you have done your best to make all the necessary information available, to all the stakeholders — staff and students — in one place that can be accessed by all. It makes it much harder for students to claim 'but x told me something else', as all students are automatically referred to the Web site when querying something. In addition, having the information provided in written form, rather than verbal, attempts to provide some security that, for those students for whom English is not their first language, consistency of interpretation of instructions is provided. The issue of closure and understanding that what has been achieved is complete and unambiguous is discussed in Coakes and Willis (2000), and also the potential for misinterpretation, misunderstanding, and misinformation through the written word in Willis and Coakes (2000), where we also discuss the often slap-dash, and haphazard way that e-mails are written. Responses to communications are 'dashed off' without due care, or at least the care required of a written, paper-based reply that a Web site environment can provide. A particular issue with using only e-mail as a means of communication is the lack of guarantee of delivery. Not just physical delivery through the technology, but also a guarantee that the e-mail has been accessed and acted upon by the recipient. As an experiment a number of e-mails with read receipts were sent over a period of months to a selected population of IT-oriented students. The read receipts were often returned up to six weeks after the e-mail was sent!

The cons are threefold: firstly the time involved in developing and managing the site can be considerable, especially at the beginning of a semester, and is a constant drain as the site has to be updated regularly; secondly, students cannot be forced to use to the Web site, although we can make it very difficult for them if they don't; and thirdly the technology—in order to rely on such technology for student support, it is essential that it is reliable and easily available. Additionally, many staff members will be reluctant to use the new technology and may take some time to adjust.

FUTURE TRENDS

It is clear from the literature that the use of VLEs will increase over time as pressure on both lecturers and students increases for flexible forms of course delivery. This may in time be supplemented by a financial pressure to increase staff/student ratios in times of falling student numbers. The new debate in the use of VLEs is not around using such environments for giving lectures, but for creating online learning environments with support groups and group-work, fostering a different type of learning environment from that encountered in standard lectures and seminars.

As lecturers become more familiar with the use of this type of technology, it can become increasingly integrated into 'normal' courses through use of the discussion facility to encourage critical thinking and debate amongst students and as a supplement to the face-to-face seminars. Group-work can also be facilitated through these environments, which can overcome the distance issue between part-time and full-time students working together on group projects for instance. Indeed, we are currently considering how students who are undertaking the same module, but in universities sited across the world

Table 1. The VLE and students

Students	*Advantages*	*Disadvantages*
	The provided interface is the same across all modules.	As the area titles are largely pre-programmed, lecturers have to be creative as to where they place the items for their modules. Thus some students 'lose' material or fail to access it appropriately or at the correct time.
	The inbuilt e-mail facility allows easy communication with all students on the module.	The e-mail addresses listed are those given by the students when they register at the university. Unfortunately many students do not use the standard university e-mail, but a Yahoo account or similar, and forget to check their university account!
	The announcements page permits students to see any urgent messages that the lecturer needs to bring to their notice.	
	The site is hosted on the university intranet, but is accessible on the Internet with the student's standard password.	
	The Web site actually develops interactively with the student comments; it is enhanced with new references, updates to materials, answers to questions raised in seminars, and through student e-mails. This interactive aspect is considered by students to be an important point in the Web site's favour, in particular, as it shows that lecturers are listening to them and are involved in the teaching of the modules.	
	The Web sites are available to potential and forthcoming students (entering as guests if the lecturer permits) as a 'taster' of what will be coming.	
	Ensuring essential materials such as assessments are on the Web site reduces student stress, as they know they always have access to vital details should they lose the originals.	Printing issues especially with Adobe files.
	Students can compare their work against posted assessment comments and see where they may have fallen short without talking to lecturers.	
	The use of a Web site also provides a learning experience for students in Web navigation and the utilisation of essential 21st century skills such as file downloading. These may be new skills for some international students.	
	Forums are available for students to discuss issues relating to seminars and lectures.	Students are reluctant to use the discussion areas due to the *time* factor involved and also perhaps due to the exposure (of what they might see as their weakness if they ask questions).
	Group work is able to be performed, as the site permits group areas to be created for joint working. The discussion areas additionally have chat rooms.	
	A link to the university's electronic resources is provided.	

Table 2. The VLE and lecturers

Lecturers	*Advantages*	*Disadvantages*
	As the interface is pre-programmed, the lecturer is not required to know any programming skills or to make any decisions on layout.	There is no opportunity to provide the information in a unique manner, and many of the section titles are inappropriate but cannot be customised.
	The e-mail facility allows easy communication with all students on the module.	The e-mail addresses listed are those given by the students when they register at the university.
	The announcements page permits students to see urgent messages.	
	As the material is contained within the intranet, intellectual property issues are no longer such a worry as when Internet access permitted anyone access to all module materials.	
	All module teaching staff also access this material and thus can ensure that they all talk with 'one voice' in relation to queries; also ensures that they have all the vital module information readily available in good time for preparation of seminars and are kept informed of what has been covered in lectures and so forth.	
	Putting answers to student queries on the Web sites saves lecturer time; it can also be shown that using such Web sites saves staff time in other ways. For instance, by ensuring that all essential information such as assessment details are on the site, duplicate copies do not have to provided in office hours to students who have lost their originals.	The time issues with updating and monitoring the Web site and providing the additional items such as tests.
	Another saving in staff time has been provided through posting a set of generic comments as feedback on formative assessments.	
	The Web also permits lecturers to monitor student activity on the site, and to see which areas have been accessed and when by individual students. Students who have not been accessing materials can be followed up.	
	Online assessments (tests of various types) are able to be created for assessment or reinforcement of seminar activities.	

(Europe and South East Asia for instance), can undertake group-work supported in just such a way.

The capability of the VLEs to enable online assessment and automated marking may additionally provide both lecturers and students with quick assessment and results for those topic areas which are suitable. One issue that needs careful consideration is that of engagement with modules. Unless students are required to access the VLE on a regular basis, there may be difficulty in getting all students assessed at any one time. It is particularly useful in those subjects where online collaboration forms part of the assessment regime.

In a recent course undertaken on e-tutoring by one of the authors (on BlackBoard), it was noted that:

"…the power of online learning is the potential to put experiences together that otherwise would never encounter one another, i.e., on this course we have a number of people across a relatively small 'world' (the university) who would probably never had had the opportunity to share experiences and therefore create better understanding (or knowledge)." (Gunter, 2004)

It is this power that may provide the boost to VLEs and their future use.

CONCLUSION AND PLANS FOR THE FUTURE

As discussed earlier, these Web sites are not static. We have talked about the need for interactivity, and this is a factor for future development. Issues such as speed of response and download time are important considerations if moving towards a multimedia approach, and this may negate some of the positives of the change to a Web-based solution. There are always issues and enhancements, however, now that the basics are working, that could be provided.

Hit counters have proved not to work successfully in FrontPage, however the new VLE can be monitored for students accessing selected pages. However, these are not a good measure of the extent to which participation is leading to learning. This is better monitored by feedback from the students on the effectiveness of the sites and the future development needs.

Enhancements will be provided as time and software capabilities permit and as the students suggest. The obvious development is to move towards a multimedia-based Web site with enhanced capabilities, but there are issues with media storage facilities here, as well as copyright.

In conclusion, it has to be said that the use of a VLE has been very successful and has the potential to allow delivery of such large modules on other courses with the ability to manage the process—a truly enabling technology.

REFERENCES

Barbera, E. (2004). Quality in virtual education environments. *British Journal of Educational Technology, 35*(1), 13-20.

Coakes, E. & Willis, D. (2000). Computer mediated communication in universities and further education establishments—a comparison of use and utility. In M. Khosrow-Pour (Ed.), *IRMA Conference Proceedings: Challenges of Information Technology Management in the 21st Century* (pp. 202-206). Hershey, PA: Idea Group Publishing.

Coakes, E. & Willis, D. (2002). Managing large modules: E-mail or Web site? In M. Anandarajan & C. Simmers (Eds.), *Managing Web usage in the workplace: A social, ethical and legal perspective* (pp. 235-253). Hershey, PA: Idea Group Publishing.

Collins, M. (2000). Comparing Web, correspondence and lecture versions of a second-year non-major biology course. *British Journal of Educational Technology, 31*(1), 21-27.

Culnan, M.J. & Markus, M.L. (1987). Information technologies. In F.M. Janlin (Ed.), *Handbook of organisational communication: An interdisciplinary perspective.* London: Sage Publications.

Gunter, S. (2004). Retrieved May 2004 from *learning.wmin.ac.uk/Webapps/portal/frameset.jsp?tab=courses&url=/bin/common/course.pl?course_id=_7902_1*

Hirschheim, R.A. (1985). *Office automation: Concepts, technologies and issues* Wokingham: Addison-Wesley.

Hoffman, D.L., Novak, T.P. & Chatterjee, P. (2000). Commercial scenarios for the Web: Opportunities and challenges. *Journal of Computer Mediated Communication, 11*(3). Retrieved from *www.ascusc.org/jcmc/vol11/issue3/hoffman.html*

Johnson, S.D., Aragon, S.R., Shatk, N. & Palma-Rivas, N. (2000). Comparative analysis of learner satisfaction and learning outcomes in online and face-to-face learning environments. *Journal of Interactive Learning Research, 11*(1), 29-49.

Kekkononen-Moneta, S. & Moneta, G.B. (2002). E-learning in Hong Kong: Comparing learning outcomes in online multimedia and lecture versions of an introductory computing course. *British Journal of Educational Technology, 33*(4), 423-433.

Kling, R. (1996). *Computerization and controversy: Value conflicts and social choices.* San Diego: Academic Press.

McFazdean, E. (2001). Supporting virtual learning groups: Part 1: A pedagogical perspective. *Team Performance Management: An International Journal, 7*(3), 53-62.

McFazdean, E. (2001). Supporting virtual learning groups: Part 2: An integrated approach. *Team Performance Management: An International Journal, 7*(5/6), 77-92.

Mishra, S. (2002). A design framework for online learning environments. *British Journal of Educational Technology, 33*(4), 493-496.

Pennie, D., Barnett, K., Chin, P. & Dolphin, I. (2001). From virtuous to virtual: The collaborative development of information skills at the University of Hull. *VINE, 31*(1), 17-21.

Sands, M. (2003). Integrating the Web and e-mail into a push-pull strategy. *Qualitative Market Research: An International Journal, 6*(1), 27-37.

Sproull, L. & Kiesler, S. (1986). Reducing social context cues: Electronic mail in organisational communication. *Management Science, 32*(11), 1492-1512.

Warner, T. (1996). *Communication skills for information systems*. UK: Pitman.

Willis, D. & Coakes, E. (2000). Enabling technology for collaborative working¾a socio-technical experience. In S. Clarke & B. Lehaney (Eds.), *Human-centered methods in information systems: Current research and practice* (pp. 119-130). Hershey, PA: Idea Group Publishing.

KEY TERMS

Bulletin Board: An electronic message database where people can log in and leave messages.

Conferences: Virtual synchronised communication across multiple sites.

Distribution Lists: E-mail lists of all participants in the virtual environment who can be contacted as a totality or divided into specified sub-groups.

Electronic Communities: Online forums to discuss topics of mutual interest.

Electronic Newsgroup: A collection of messages posted online by individuals on a specified subject matter.

Module: A discrete educational item that covers a topic such as Statistics. It would typically consist of a set number of weeks of learning material delivered through lectures and seminars.

Virtual Learning Environment: Technology-mediated venue for learning; typically includes asynchronous as well as synchronous facilities.

Comparing Conventional and Non–Parametric Option Pricing

Paul Lajbcygier
Monash University, Australia

INTRODUCTION

Once, the seminal Black–Scholes (Black & Scholes, 1973) model was thought to be the last word on option pricing: all that was needed, it was thought, was some adjustments and it could be applied to price options on any financial instrument.

In the past two decades, increases in the bias of these Black–Scholes style models (Rubinstein, 1985) have led researchers to develop new models, coined modern parametric option pricing models. The underlying logic of this modern parametric option pricing program is that by loosening the "unrealistic" assumptions of the conventional Black–Scholes style option pricing models (e.g., including jumps in stochastic processes describing underlying asset prices (Merton, 1973); incorporating skewness and/or kurtosis describing the underlying asset distribution (Corrado & Su, 1996); considering the effects on demand and/or supply (Follmer & Sondermann, 1986); and the effects of tax (Scholes, 1976)), it is possible to mitigate the bias associated with them.

However, recently, many authors (e.g., Bakshi, Cao & Chen, 1997; Bakshi & Chen, 1997) have found that the modern parametric models are found to:

- be too complex;
- have poor out of sample performance;
- have implausible/inconsistent implied parameters.

Perhaps the final word on modern parametric models should be left to Fischer Black (Black, 1998): "The Black–Scholes formula is still around, even though it depends on at least 10 unrealistic assumptions. Making the assumptions more realistic hasn't produced a formula that works better across a wide range of circumstances".

FUTURE TRENDS: NON-PARAMETRIC OPTION PRICING

If we concede that the modern parametric option pricing program has failed, where does this leave us? There remains significant, persistent and systematic bias in modern parametric option pricing models (Rubinstein, 1985). The ideal option pricing model would not only provide unbiased option prices, be consistent with the underlying process and distributions of asset returns, and have minimal assumptions and parameters to estimate, it would also incorporate a statistical estimate of option pricing error. This ideal model may never be found by generalizing the unrealistic assumptions of the conventional option pricing models. What alternative is left?

Either we must accommodate option pricing model error explicitly and surrender any notion of improvement, or some new, alternative approach must be used. Non-parametric techniques represent such an alternative approach.

There are three scenarios where non-parametric approaches are particularly useful:

- when conventional/modern parametric solutions lead to bias in pricing;
- when conventional/modern parametric solutions exist but are too complex to use;
- and, when conventional/modern parametric solutions do not exist.

Non-parametric option pricing models can be divided into two separate strands: the model free and the semi-parametric. These two strands can be characterized by their dependence on finance theory. Model free non-parametric techniques have no reliance on finance theory whatsoever; semi-parametric approaches have a high reliance on finance theory (e.g., approaches that augment parametric option pricing models with non- parametric techniques).

Numerous technologies have been used to estimate model free non-parametric option pricing, including: genetic programming (Chen, Yeh & Lee, 1998), kernel regression (Ait-Sahalia & Lo, 1995; Broadie, Detemple, Ghysels & Torres, 1996) and artificial neural networks (Malliaris & Salchenberger, 1993).

The fundamental problem of non-parametric model free option pricing models lies with their greatest strength: their independence from the assumptions of finance theory. On the one hand, this independence is a great strength because being free of these assumptions means that the persistent, systematic and significant bias found in parametric option prices may be eliminated. On the other hand, this independence is a great weakness. The fact that non-parametric model free option pricing approaches do not rely on any finance theory for their derivation means that there is no guarantee that the prices obtained from these models will not conform to rational pricing[1].

So, the cost of the complete flexibility of the model free non-parametric option pricing approaches is that there can be no guarantee of rational pricing. Ghysels et al. (Ghysels, Patilea, Renault & Torres, 1997) state:

"non parametric model free option pricing becomes quickly infeasible since it is not able to capture a large set of crucial restrictions implied by arbitrage."

The aim of semi-parametric option pricing models is to estimate a portion of the option pricing model non-parametrically while retaining the conventional option pricing model framework to guarantee rational pricing.

There are three main branches of semi-parametric option pricing. Hybrid approaches model conventional option pricing residual error in the hope of reducing this error (Lajbcygier, 2003). General volatility models use non- parametric techniques to estimate the volatility used in conventional parametric option pricing (Dumas, Fleming & Whaley, 1996). Finally, equivalent martingale measures (or risk neutral pricing) (Campbell, Lo & Mackinlay, 1997) use non- parametric techniques to estimate the risk neutral probability distribution, which in turn can be used to price options.

CRITICAL ISSUES

Despite all its promise, non-parametric option pricing approaches are considered an emerging technology with the potential to help improve option pricing. There are still many open critical issues (some of which are discussed in Table 1).

CONCLUSION

The ideal option pricing model would not only provide unbiased option prices, be consistent with the underlying process and distributions, and have minimal assumptions and parameters to estimate, it would also incorporate a statistical estimate of option pricing error. No such model exists today. It is interesting to speculate as to whether such a model will exist in the future, and if the model will be parametric, non-parametric, or a mixture of the two.

As Campbell, Lo and Mackinlay (1997) have stated, non-parametric option pricing approaches hold promise:

"Although it is still too early to tell if these non parametric and highly data intensive methods will offer improvements over their parametric counterparts, the preliminary evidence is quite promising."

REFERENCES

Ait-Sahalia, Y., & Lo, A.W. (1995). *Nonparametric estimation of state-price densities implicit in financial asset prices.* Working Paper No. LFE-1024-95. MIT-Sloan School of Management, Laboratory for Financial Engineering.

Table 1. Summary of critical issues of non- parametric option pricing

Elimination of Option Pricing Bias To eliminate persistent, systematic and significant bias in option pricing is a critical requirement of any model.
Poor out of sample fit of modern parametric option pricing models The fact that many modern parametric models provide inaccurate pricing out of sample (i.e., for data which were not used for parameter estimation) presents a major flaw.
Implausible Implied Parameters Some parameters implied by a modern parametric model fitted to a data set (e.g., skewness) may be nonsensical.
Arbitrage and Option Pricing The process by which "riskless profit" can be earned when certain pricing conditions do not hold.

Bakshi, G., Cao, C., & Chen, Z. (1997). Empirical performance of alternative option pricing models. *The Journal of Finance, LII*(5), 2003-2049.

Bakshi, G.S., & Chen, Z. (1997). An alternative valuation model for contingent claims. *Journal of Financial Economics, 44*, 123-165.

Black, F. (1998). How to use the holes in Black-Scholes. In R. Kolb (Ed.), *Practical readings in financial derivatives* (pp. 451-461). Cornwall: Blackwell.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy, 81*, 637-654.

Broadie, M., Detemple, J., Ghysels, E., & Torres, O. (1996). *Nonparametric estimation of american options exercise boundaries and call prices* (Scientific Series No. 96s-24). Montreal: CIRANO.

Campbell, J., Lo, A., & Mackinlay, G. (1997). *The econometrics of financial markets*. Princeton, NJ: Princeton University Press.

Chen, S., Yeh, C., & Lee, W. (1998). Option pricing with genetic programming. *Proceedings of the 3rd Annual Genetic Programming Conference*, University of Wisconsin (pp. 32-37).

Corrado, C.J., & Su, T. (1996). Skewness and kurtosis in S&P 500 index returns implied by option prices. *The Journal of Financial Research, XIX*(2), 175-192.

Dumas, B., Fleming, J., & Whaley, R.E. (1996). Implied volatility functions: Empirical tests. *The Journal of Finance, LIII*(6), 2059-2106.

Follmer, H., & Sondermann, D. (1986). Hedging of non-redundant contingent claims. In W. Hildenbrand & A. Mas-Colell (Eds.), *Contributions to mathematical economics* (pp. 205-223).

Ghysels, E., Patilea, V., Renault, E., & Torres, O. (1997). *Nonparametric methods and option pricing* (No. 97s-19). Montreal: CIRANO.

Lajbcygier, P. (2004). Improving option pricing with the product constrained hybrid neural network. *IEEE Transactions on Neural Networks, 15*(2), 465-476.

Malliaris, M., & Salchenberger, L. (1993). A neural network model for estimating option prices. *Applied Intelligence, 3*(3), 193-206.

Merton, R.C. (1973). Theory of rational option pricing. *Bell Journal of Economics and Management Science, 4*, 141-183.

Rubinstein, M. (1985). Nonparametric tests of alternative option pricing models using all reported trades and quotes on the 30 most active CBOE option classes from August 23, 1976 through August 31, 1978. *The Journal of Finance, XL*(2), 455-480.

Scholes, M.S. (1976). Taxes and the pricing of options. *The Journal of Finance, 31*(2), 319-332.

KEY TERMS

Call Option: The right but not the obligation to purchase a stock (or financial instrument) at a pre-determined price at some time in the future.

Financial Future: The right and obligation to purchase an underlying financial instrument at a pre-determined price at some time in the future.

Modern Parametric Option Pricing Models: Those models that extend the Black–Scholes work and relax some of their assumptions in the hope of eliminating bias.

Non-Parametric Option Pricing Models: Pricing options using inductive or statistically oriented techniques.

Put Option: The right but not the obligation to sell a stock (or financial instrument) at a pre-determined price at some time in the future.

Stochastic Process: A mathematical description of the random movement of a stock or financial instrument such as a future.

Volatility: The most important input to the Black–Scholes model (a.k.a. standard deviation of the underlying asset price returns).

ENDNOTES

[1] Rational pricing refers to option pricing constraints that must be obeyed.

Complex Adaptive Enterprises

Anet Potgieter
University of Cape Town, South Africa

Kurt April
University of Cape Town, South Africa

Judith Bishop
University of Pretoria, South Africa

INTRODUCTION

In a world where the market, customer profiles and demands change constantly and the events in the global marketplace are unpredictable, it becomes increasingly difficult for an enterprise to sustain its competitive advantage. Under these conditions of uncertainty, complexity and constant change, it becomes very important for an enterprise to be able to learn from its experience and to adapt its behavior in order to constantly outperform its competitors. An enterprise that has these characteristics is a complex adaptive enterprise.

The interrelationships between resources in a complex adaptive enterprise and its global behavior within the marketplace can be numerous and mostly hidden, and can affect many different resources throughout the enterprise. One of the main challenges of the modern enterprise is to understand this complex web of interrelationships and to integrate this understanding into its business processes and strategies in such a way that it can sustain its competitive advantage.

BACKGROUND

The Chain of Sustainability

According to the resource-based theory, there are dynamic relationships between enterprise resources, the capabilities of the enterprise and the competitive advantage of the enterprise. The complex adaptive enterprise maintains a chain of sustainability that constantly evolves from the interactions between the individual resources and the interactions between the resources and the dynamically changing marketplace.

Resources or assets are the basic components in the chain of sustainability. Example resources are products, employee skills, knowledge, and so forth. These resources are combined into complementary resource combinations (CRCs) according to the functionality that these resources collectively achieve. CRCs are the unique inter-relationships between resources and are the source of competitive advantage in an enterprise, as these relationships cannot be duplicated by competitors. The behaviors of the CRCs define the strategic architecture of an enterprise, which is defined as the capabilities of an enterprise, when applied in the marketplace.

Social complexity refers to the complex behavior exhibited by a complex adaptive enterprise, when its CRCs are embedded in a complex web of social interactions. These CRCs are referred to as socially complex resource combinations (SRCs). In social complexity, the source of competitive advantage is known, but the method of replicating the advantage is unclear. Examples include corporate culture, the interpersonal relations among managers or employees in an enterprise and trust between management and employees. SRCs depend upon large numbers of people or teams engaged in coordinated action such that few individuals, if any, have sufficient breadth of knowledge to grasp the overall phenomenon.

Casual ambiguity refers to uncertainty regarding the causes of efficiency and effectiveness of an enterprise, when it is unclear which resource combinations are enabling specific capabilities that are earning the profits.

The Complex Adaptive Enterprise

A complex adaptive enterprise is an enterprise that can function as a complex adaptive system. A complex adaptive system can learn from and adapt to its constantly changing environment. Such a system is characterized by complex behaviors that emerge as a result of interactions among individual system components and among system components and the environment. Through interacting with and learning from its environment, a complex adaptive enterprise modifies its behavior in order to maintain its chain of sustainability.

It is impossible for an enterprise that cannot learn from experience to maintain its chain of sustainability. The learning process involves perception of environmental inputs, understanding the perceived inputs (making meaning out of these inputs), and turning this understanding into effective action (Senge, Kleiner, Roberts, Ross & Smith, 1994). The Soft Systems Methodology (Checkland, 2004) is a methodology that was developed that involves perception, understanding and acting in an enterprise.

Understanding Emergence

Self-awareness in a complex adaptive enterprise is instrumental in the maintenance of the chain of sustainability. Enterprises need to understand the interrelationships between the individual behaviors of the resources and the emergent behaviors of the CRCs and SRCs. This will enable the enterprise to understand its own social complexity and causal ambiguity.

Emergence, the most important characteristic of a complex adaptive enterprise, is the collective behavior of interacting resources in the CRCs. Emergence is the same as holism (Baas & Emmeche, 1997). Holism in a complex adaptive system means that the collective behaviour of the system components is more than the sum of the behaviours of the individual system components, for example, a flock is more than a collection of birds and a traffic jam is more than a collection of cars (Odell, 1998).

What does it mean to understand something? According to Baas & Emmeche (1997), understanding is related to the notion of explanation. All complex adaptive systems maintain internal models (Holland, 1995). These mechanisms are used for explanation and understanding.

The human mind is self-aware and capable of self-observation and self-interaction. Consciousness may be seen as an internal model maintained by the mind. In Minsky's *Society of Mind*, internal observation mechanisms called A-Brains and B-Brains maintain internal models consisting of hyperstructures called K-Lines. Each K-Line is a wire-like structure that attaches itself to whichever mental agents are active when a problem is solved or a good idea is formed (Minsky, 1988). Minsky describes how a system can watch itself, using its B-Brain.

Gell-Mann (1994) refers to the information about the environment of a complex adaptive system and the system's interaction with the environment as the "input stream" of the system. A complex adaptive system creates and maintains its internal model by separating "regularities from randomness" in its input stream (Gell-Mann, 1994). These regularities are represented using hyperstructures, which in turn constitute the internal model of the complex adaptive system. The observation mechanism of a complex adaptive system is responsible for the identification of regularities in its input stream, as well as for the progressive adaptation of the hyperstructures to include these regularities.

In the complex adaptive enterprise, the hyperstructures encode the knowledge of the enterprise, and are distributed throughout the enterprise. This knowledge belongs to one of the following component knowledge types:

- knowledge related to internal relationships within the company;
- knowledge related to products and services;
- knowledge related to business processes and business units;
- knowledge related to specific projects and project implementations;
- knowledge related to customers;
- knowledge related to the marketplace.

Component knowledge consists of both tacit and explicit knowledge. Tacit knowledge is usually defined as that which cannot be written down or specified. This knowledge is embedded within the interrelationships between the local behaviors of resources within the CRCs and the emergent behaviors of the CRCs. Knowledge, particularly tacit knowledge, is the most important strategic resource in an enterprise (April, 2002).

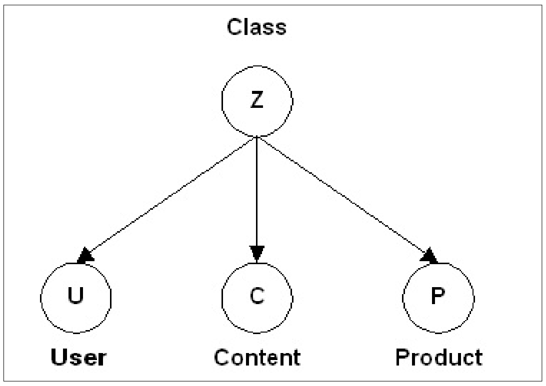
Bayesian Hyperstructures

Bayesian networks provide the ideal formalism to be used as hyperstructures in the complex adaptive enterprise. These networks can be used to encode beliefs and causal relationships between beliefs and provide a formalism for reasoning about partial beliefs under conditions of uncertainty (Pearl, 1988). These networks can be used to learn a probabilistic model of what the emergent effects are of certain interactions and behaviors in response to certain environmental states (the causes). Such a causal model can then be queried by an arbitration process to decide which action(s) are most relevant given a certain state of the environment.

A Bayesian network is a directed acyclic graph (DAG) that consists of a set of nodes that are linked together by directional links. Each node represents a random variable or uncertain quantity. Each variable has a finite set of mutually exclusive propositions, called states. The links represent informational or causal dependencies among the variables, where a parent node is the cause and a child node, the effect. The dependencies are given in terms of conditional probabilities of states that a node can have given the values of the parent nodes (Pearl, 1988). Each node has a conditional probability matrix to store these conditional probabilities, accumulated over time.

Figure 1 illustrates a simple Bayesian network, which we adapted from the user-words aspect model proposed by Popescul, Ungar, Pennock & Lawrence (2001). Our

Figure 1. A simple Bayesian network



network models the relationship between three observable variables, namely users (U), the contents of browsed web pages characterized in terms of concepts (C), products bought from these pages (P) and one hidden variable, namely the class variable (Z).

In Figure 1 above, the states of the hidden class variable Z are mined from historical data (observations of U, P and C). The class variable Z is the single cause influencing multiple effects (U, P and C). This probability distribution is called a naïve Bayes model or sometimes called a Bayesian classifier (Russell & Norvig, 2003).

Bayesian learning can be described as the "mining" of the structure of a Bayesian network and the calculation of the conditional probability matrices from history data. The data may be incomplete and the structure of the Bayesian network can be unknown.

Bayesian inference is the process of calculating the posterior probability of a hypothesis H (involving a set of query variables) given some observed event e (assignments of values to a set of evidence variables),

$$P(H \mid e) = \frac{P(e \mid H)P(H)}{P(e)}, \text{ where}$$

$P(H \mid e)$ represents the belief in H given e,

$P(e \mid H)$ represents the belief in e given H, and

$P(H)$ and $P(e)$ represent the beliefs in H and e respectively.

Bayesian inference is NP-hard (Pearl, 1988; Dechter, 1996). In order to simplify inference, Bayesian networks are simplified to trees or singly-connected polytrees. A tree is a DAG in which each node has only one parent (Pearl, 1988). A singly-connected polytree is a DAG in which the nodes can have multiple parents, but with the restriction that there is only one path, along arcs in either direction, between any two nodes in the DAG (Nilsson, 1998; Pearl, 1988).

SELF-AWARENESS AND SUSTAINABLE COMPETITIVE ADVANTAGE USING BAYESIAN AGENCIES

Adaptive agents are the basic building blocks of a complex adaptive system. The collective behavior of the agents, the interactions between the agents and the environment as well as the interactions between the agents themselves comprise a complex set of causal relationships.

We implement complex adaptive systems using Bayesian agencies that collectively implement Bayesian behavior networks. These networks are Bayesian networks that model the regularities in the input stream of a complex adaptive system. The nodes in a Bayesian behavior network are grouped into what we call competence sets, where each competence set has an associated set of actions that must be performed by the Bayesian agencies depending on the states of the nodes in the competence set. These actions are usually part of a business process or workflow in the enterprise.

Complex adaptive systems generate their internal models from re-usable building blocks (Holland, 1995). As an example, the quarks of Gell-Mann (1994) are combined into nucleons, nucleons are combined into atoms, atoms are combined into molecules, and so forth. It is essential that the knowledge in the internal model of the enterprise be represented using re-usable building blocks, in order for the enterprise to be able to function as a complex adaptive system.

Our Bayesian agencies consist of simple re-usable software components, distributed throughout the enterprise. There are two types of Bayesian agencies, namely belief propagation agencies and competence agencies. Belief propagation agencies consist of a collection of components, where each component can be one of three re-usable components, namely node components, link components and belief propagation agents. Collectively these simple components capture the knowledge throughout the enterprise by collectively implementing distributed Bayesian behavior networks. Each node component implements a Bayesian behavior network node. Each network link is implemented by a queue, together with a link component that participates in the synchronization of messages flowing to the child, or to the parent node via the queue. For each queue, a belief propagation agent is deployed that listens on that queue for messages from the child or parent node of the associated network link.

The belief propagation agents collectively perform Bayesian inference by localized message passing in response to the environmental evidence in order to update the beliefs of network nodes. The competence agencies use the beliefs of selected network nodes to determine if certain business components must be activated or not. Business components are re-usable components containing parts of business processes or workflow processes. Each competence agency monitors a set of constraints on the beliefs of a subset of nodes – the constraint set. If all the constraints in a constraint set are met, the competence agency can activate its associated business component.

Node components are deployed throughout the enterprise to collect evidence from disparate data sources within the enterprise or from external data sources. The Bayesian agencies incrementally learn from this experience.

The Bayesian agencies are observation mechanisms that enable the enterprise to be self-aware. Belief propagation agencies are connected to the real world. As soon as evidence is received from the environment, the belief propagation agents collectively perform Bayesian inference by using local message passing. The competence agencies inspect the beliefs of nodes and act upon these beliefs and possibly change the state of the environment, influencing the collective Bayesian inference of the belief propagation agencies.

The flexibility, adaptability and reusability of automated business processes (enterprise software) determine the ability of an enterprise to evolve and survive in the marketplace (Sutherland & van den Heuvel, 2002). The belief propagation agencies enable the re-usable business components in the competence agencies to be flexible and adaptable.

We have successfully implemented prototype Bayesian agencies using Sun's Enterprise JavaBeans™ component architecture. We developed prototype node and link components and belief propagation agents that are assembled into distributed Bayesian behavior networks, collectively performing Bayesian learning and Bayesian inference in singly-connected Bayesian behavior networks with known structure and no hidden variables.

FUTURE TRENDS

Future research will involve a full implementation of Bayesian learning, where Bayesian agents collectively and incrementally discover structure from data in the presence of known values for variables as well as in the presence of missing data. We will also complete the collective belief propagation capabilities of the Bayesian agencies in order to cope with multiply-connected Bayesian behaviour networks.

CONCLUSION

Our Bayesian agencies can be distributed throughout an enterprise, enabling it to function as a complex adaptive enterprise. These agencies will assist the enterprise to be self-aware by collectively modeling the complex interrelatedness of local behaviors of resources and emergent behaviors of CRCs, from which the enterprise's tacit knowledge, social complexity and causal ambiguity emerges – the source of its competitive advantage. The enterprise can then use this self-understanding to adapt its business processes and to formulate new knowledge or business strategies in response to the ever-changing marketplace in order to sustain its competitive advantage.

REFERENCES

April, K. (2002). Guidelines for developing a k-strategy. *Journal of Knowledge Management, 6*(5), 445-456.

Baas, N. A., & Emmeche, C. (1997). On emergence and explanation. *Intellectica*, 25, 67-83. Retrieved March 22, 2001, from *http://www.nbi.dk/~emmeche/coPubl/97d.NABCE/ExplEmer.html*

Checkland, P. (2004). Soft systems methodology. Retrieved June 25, 2004, from *http://www.onesixsigma.com/_lit/white_paper/petercheckandssm1.pdf*

Dechter, R. (1996). Bucket elimination: A unifying framework for probabilistic inference. Uncertainty in Artificial Intelligence, UAI96, 211-219. Retrieved October 8, 2000, *http://www.ics.uci.edu/~dechter/publications/*

Gell-Mann, M. (1994). *The quark and the jaguar* (2nd ed.). London: Little, Brown and Company.

Holland, J. H. (1995). *Hidden order: How adaptation builds complexity*. MA: Addison-Wesley Publishing Company.

Minsky, M. (1988). *The society of mind* (First Touchstone ed.). New York: Simon & Schuster.

Nilsson, N. J. (1998). *Artificial intelligence: A new synthesis* (1st ed.). San Francisco: Morgan Kaufmann.

Odell, J. (1998). Agents and beyond: A flock is not a bird. *Distributed Computing,* 52–54. Retrieved June 25, 2004, from *http://www.jamesodell.com/publications.html*

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference* (2nd ed.). San Mateo: Morgan Kaufmann Publishers.

Pearl, J., & Russell, S. (2000). *Bayesian networks* (Technical Report R-277). UCLA, Cognitive Systems Laboratory. Retrieved May 5, 2001, from *http://bayes.cs.ucla.edu/csl_papers.html*

Popescul, A., Ungar, L. H., Pennock, D. M., & Lawrence, S. (2001). *Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments.* Retrieved January 28, 2002, from *http://www.cis.upenn.edu/~popescul/publications.html*

Russell, S. J., & Norvig, P. (2003). *Artificial intelligence: A modern approach* (2nd ed.). New Jersey, USA: Prentice Hall.

Senge, P. M., Kleiner, A., Roberts, C., Ross, R. B., & Smith, B. J. (1994). *The fifth discipline fieldbook*. New York: Double Day.

Sutherland, J., & van den Heuvel, W. (2002). Enterprise application integration encounters complex adaptive systems: A business object perspective. *Proceedings of the 35th Hawaii International Conference on System Sciences*. Retrieved June 25, 2004, from *http://jeffsutherland.com/papers/hicss2002/eai_hicss2002.pdf*

KEY TERMS

Bayesian Agencies: Agencies consisting of simple agents that collectively implement distributed Bayesian behavior networks. The agents are organized into agencies, where each agency activates one or more component behaviour depending on the inference in the underlying Bayesian behaviour network.

Bayesian Behavior Networks: Specialized Bayesian networks, used by the Bayesian agents to collectively mine and model relationships between emergent behaviours and the interactions that caused them to emerge, in order to adapt the behaviour of the system.

Bayesian Hyperstructures: Bayesian Behavior Networks are Bayesian hyperstructures that in turn constitute the internal model of the complex adaptive system.

Causal Ambiguity: Refers to uncertainty, by competitors, regarding the causes of efficiency and effectiveness of a company, when it is unclear which resource combinations are enabling specific competitive capabilities that are earning the company profits.

Chain of Sustainability: An evolving, dynamic and matched mix between company resources (arranged in value-generating combinations) and the changing marketplace that gives the company a competitive edge.

Competence Sets: The nodes in a Bayesian behavior network are grouped into competence sets, where each competence set has an associated set of actions that must be performed by the Bayesian agencies depending on the states of the nodes in the competence set.

Competitive Advantage: A company is said to have a competitive advantage when, based on its strategic architecture and complementary resource combinations (CRCs), it is able to implement a strategy that generates returns and benefits in excess of those of its current competitors – who simultaneously are implementing strategies, similar or otherwise – because of the perceived value in the marketplace. The definition therefore also depends on what the company, its management and its stakeholders, define as what the required returns and benefits should be (because even though many would list it as financial, clearly this does not apply to all companies, i.e., an advantage could be something other than financial). One could reasonably expect, though, that companies within similar industries would define similar variables as the required returns and benefits. A company is said to have a sustained competitive advantage when it is implementing a value-creating strategy, which generates returns and benefits at a level not enjoyed by current competitors and when these other companies are unable to reach an "equilibrium level" with the company enjoying the advantage. In this sense, the definition of sustained competitive advantage adopted here does not imply that it will "last forever," and does not depend upon the period of time during which a company enjoys a competitive advantage (rather, the equilibrium level is critical in this definition).

Complementary Resource Combinations (CRCs): Are not factor inputs, but are complex combinations of interrelated configurations, or networks of assets, people, and processes that companies use to transform inputs to outputs. Many of these configurations are a blend of "hard" tangible resources and "soft" intangible resources which simply cannot be recreated by another company. Finely honed CRCs can be a source of competitive advantage.

Resources: Also known as "assets", come in many forms, from common factor inputs that are widely available and easily purchased in arms-length transactions, to highly differentiated resources, like brand names, that are developed over many years and are very difficult to replicate. Resources come in two main forms: "tangible resources", which are the easiest to value, and often are the only resources that appear on a company's balance sheet. They include real estate, production facilities and raw materials, among others. Although tangible resources may be essential to a company's strategy, because of their standard nature they rarely constitute a source of competitive advantage; and "intangible resources" - include such things as company reputations, brand names, cultures, technological knowledge, know-how shared among employees, patented process and design, trademarks, accumulated learning and/or knowledge, as well as experience. These resources often play important roles in competitive advantage (or disadvantage) and company value. Intangible resources also have the important property of not being consumed in usage.

Social Complex Resource Combinations (SRCs): Depend upon large numbers of people, or teams, engaged in coordinated action such that few individuals, if any (both outside the company, as well as inside the company), have sufficient breadth of knowledge to grasp the overall phenomenon.

Social Complexity: Is when the source of advantage is known, but the method of replicating the advantage is unclear, e.g., corporate culture, the interpersonal relations among managers in a company, or trust between management and labor.

Strategic Architecture: Refers to a company's capabilities, when applied in the marketplace.

Component-Oriented Approach for Designing Enterprise Architecture

Zoran Stojanovic
Delft University of Technology, The Netherlands

Ajantha Dahanayake
Delft University of Technology, The Netherlands

INTRODUCTION

One of the main challenges enterprises face today is how to manage complexity of systems being developed, effectively utilize the power of the Internet, and be able to rapidly adapt to changes in both technology and business. The new paradigm of component-based development (CBD) has been introduced as an excellent solution for building complex Internet-enabled enterprise information systems (Brown, 2000; Szyperski, 2002). The basic idea of CBD originates from the strategy successfully applied in other engineering disciplines that a system developed from components is more flexible and easier to develop. CBD provides higher productivity in system development through reusability, more effective system maintenance, higher quality of solutions and the possibility for parallel work. Moreover, it provides better system adaptability through replaceability of parts, localization and better control of changes, system scalability, and the possibility of using legacy assets.

CBD has often been presented as a new silver bullet for complex, enterprise-scale system development in the Internet age (Udell, 1994). However CBD inherits many concepts and ideas from the earlier encapsulation and modularization, "divide-and-conquer" initiatives in information technology (IT). The NATO Conference in 1968 recognized that producing software systems should be treated as an engineering discipline providing system assembling from software components (McIlroy, 1968). Parnas (1972) defines concepts and requirements for decomposing system into modules. These principles of separation of concerns, encapsulation, and plug-and-play building blocks have been applied in different ways through the concepts of functions, subroutines, modules, units, packages, subsystems, objects, and now components.

The CBD paradigm has been first introduced at the level of implementation and deployment. CBD middleware technologies, such as CORBA Components (Siegel, 2000), Sun's Enterprise Java Beans and Microsoft's COM+/.NET, are now used as standards for the development of complex enterprise distributed systems. While the technology solutions are necessary in building the system, one cannot simply program and deploy components using a component middleware without any prior plan to follow from business requirements towards implementation. For the effective use of the CBD paradigm and in order to gain real benefits of it, the component way of thinking must be applied in earlier phases of the development lifecycle, such as system analysis and design. CBD methods and approaches proposed so far do not provide a complete and consistent support for various component concepts. Components are often treated as implementation concepts—packages of binary or source code that can be deployed over the network nodes. During the system analysis and design, components, if used, are often represented as larger-grained business objects. This suggests using components mainly at the system implementation and deployment as software code packages, while still following the principles of object-oriented modeling, analysis and design. At the same time, the role and usefulness of the component concept as a bridge between business and technology issues have not been truly recognized yet. Components can be identified very early in the system lifecycle, namely derived from business requirements, and then used as central artifacts of the whole development process. In this way, the whole process would be structured and organized around the same set of component concepts. That can provide a necessary alignment and traceability from business services to implementation assets. The benefit from the separation of concerns using components will be gained at all levels of system development.

BACKGROUND

Component technologies are now widely used in the development of complex Internet-enabled systems. First, VBX controls, DCOM/COM, CORBA and Java Beans, and now COM+/.NET, CORBA Components and Enterprise Java Beans (EJB) represent the standard component-based implementation solutions. The physical perspective on components as binary packages of software is still predominant. The standard unified modeling language (UML) treats components as packages of binary code and uses them in describing system implementation through component and deployment diagrams (Booch, Rumbaugh, & Jacobson, 1999). Components in UML represent physical things that can be deployed over network nodes. The Catalysis approach defines a component as a package of software code as well as other software artifacts (D'Souza & Wills, 1999). According to Szyperski, a software component is a unit of composition with contractually specified interfaces and explicit context dependencies (Szyperski, 2002). A software component can be deployed independently and is subject to composition by third parties. Gartner Group (1997) defines a runtime software component as a dynamically bindable package of one or more programs managed as a unit and accessed through documented interfaces that can be discovered at runtime. Definition of a business-oriented component concept can be found in Herzum and Sims (2000), where a business component is the software implementation of an autonomous business concept or business process, and in Andersen Consulting (1998), where a business component is a means for modeling real-world concepts in the business domain. When introducing components, the question about similarities and differences between objects and components are naturally arising. According to Udell (1994), components represent a new silver bullet for system development in the Internet age, while objects have failed to provide higher level of reusability. In the UML, components are nothing else than larger-grained objects deployed on the network nodes. In Szyperski (2002), a component comes to life through objects, and therefore, it would normally contain one or more classes, as well as traditional procedures and even global variables. In a debate over this topic (Henderson-Sellers, Szyperski, Taivalsaari, & Wills, 1999) granularity has been seen as the main issue in distinguishing components and objects. According to Catalysis, components are often larger-grained than traditional objects, and can be implemented as multiple objects of different classes.

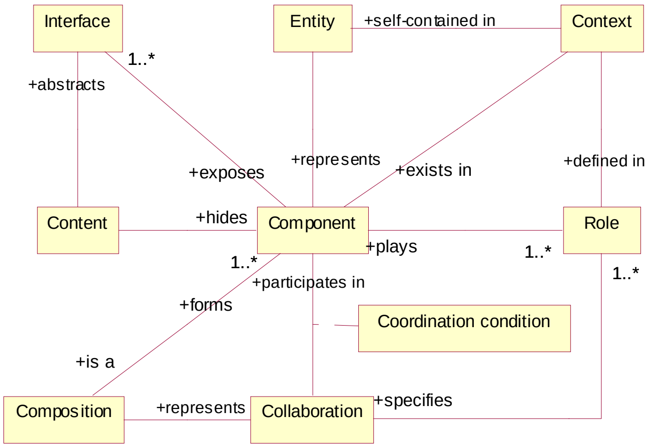
The academia and industry have just started to recognize the importance of new CBD methods, processes, techniques and guidelines. The methods and approaches are often greatly influenced by the object-oriented concepts, constructs and principles, dictated by the use of the standard UML. The Rational Unified Process (RUP) (Jacobson, Booch, & Rumbaugh, 1999), Catalysis, and the Select Perspective (Allen & Frost, 1998) can be considered as the first generation of the CBD methods. The business component factory (BCF) (Herzum & Sims, 2000), the UML components approach (Cheesman & Daniels, 2000), and the KobrA approach (Atkinson et al., 2002) represent the second generation of the CBD methods. These methods are more focused on components concepts than previous ones. They provide a comprehensive support to CBD throughout the system lifecycle, and represent remarkable achievements in the field. On the other hand, there are certain shortcomings. The analysis and evaluation of CBD methods can be found in Dahanayake, Sol, and Stojanovic (2003).

SERVICE-BASED COMPONENT CONCEPT

Components have been so far used mainly as implementation artifacts. However, the components are equally useful and important if used as modeling and design artifacts for building the logical architecture of the system (Stojanovic & Dahanayake, 2002). The essence of the component approach is the explicit separation between the outside and the inside of the component. This means that only the question WHAT is considered (what useful services are provided by the particular building block to the context of its existence?) not the HOW (how are these services actually implemented?). A component fulfills a particular role in the context, by providing and requiring services to/from it. A component has a hidden interior and exposed interface. It participates in a composition with other components to form a higher-level behavior. At the same time, every component can be represented as a composition of lower-level components. Well-defined behavioral dependencies and coordination of activities between components are of a great importance in achieving the common goal. The metamodel of the core component concepts is shown in Figure 1.

According to the role(s) a component plays in the given context, it exposes corresponding behavior by providing and requiring services to/from its context, or by emitting and receiving events. The services a component provides and requires form the basic part of its contract. Services can be of different types such as performing computation, providing information, communication with the user, and so forth. They are fully specified in a contract-based manner using pre-conditions, post-conditions and other types of constraints. A component must handle, use, create or simply be aware of certain information to provide its services properly. In order to be used in a different context or to be adaptable to the changes in its context, a component can possess configuration parameters that can adapt the component according to new requirements coming from the outside. A component can possess a set of non-functional parameters that characterize the "quality" of its behavior. Figure 2 shows the component specification concepts.

Figure 1. Core component concepts



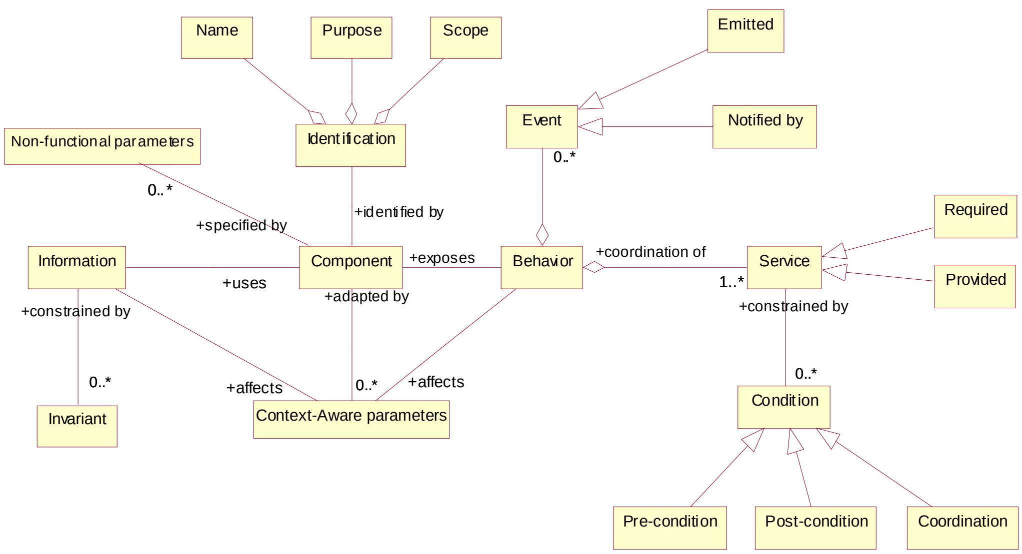
BUSINESS-DRIVEN COMPONENT ARCHITECTURE

In order to manage complexity and ensure completeness in the specification of components and component-oriented system architecture we use the ISO standard reference model of open distributed processing (RM-ODP) as an underlying idea (ODP, 1996). RM-ODP defines a framework for specifying architectures for distribution, interoperability and portability of applications based on object-oriented technology. The RM-ODP specification of a system consists of five different specifications, corresponding to five separate, but related and consistent viewpoints – enterprise, information, computational, engineering, and technology. Based on ODP viewpoints, we propose the following three architectural models:

- business architecture model (BAM);
- system architecture model (SAM); and
- distribution architecture model (DAM).

These different but consistent and correlated architectural models provide us a way of separation of concerns through different architectural views (Stojanovic, Dahanayake, & Sol, 2001). The relation between our architectural models and the RM-ODP viewpoints, as well

Figure 2. Component specification concepts

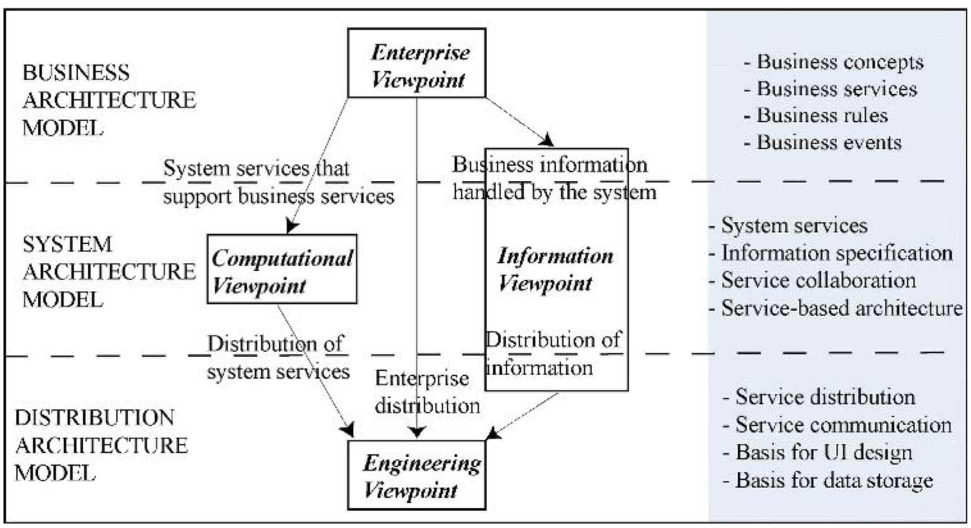


as the main relations between the models are shown in Figure 3. Our main idea is to incorporate the component concept into each of the RM-ODP viewpoints. In this way, we can define component-oriented viewpoints on the system, completely focused on the component concept and organized around it, as powerful means for specifying component architecture. The same consistent set of component concepts used in different viewpoints represents the point of consistency and integration between them.

Proposed models are defined as following:

- *Business architecture model (BAM)* specifies the behavior of the system in the context of the business for which it is implemented. It represents business context and processes in which the system should participate, services that must be offered by the system to fulfill business purposes, business domain concepts and information that must be handled by the system, and business rules that must be satisfied and supported by the system. Defined services are the basis for identification of higher-level business components as we have seen in the previous section. Each business component, as well as their collaboration, must be specified in the business architecture model, through the common set of component properties. Business domain conceptual information should be defined and cross-referenced by the specification of business services offered by components. This shows how conceptual information types are used and handled by particular business components in providing services. Each business component is typically a collaboration of several system components. This means that identified business components are used as an entry for defining business-driven system architecture and correspondent system components.
- *System architecture model (SAM)* defines the structure of the system in terms of configurations of system services and the interactions among them to provide necessary business services related to the business requirements on the system. This model defines information that should be handled by the system, rules for information dynamics, and information processing that should be performed by the system. SAM takes higher-level components identified in the previous model as an input. This model further specifies system components that in collaboration realize required business services. Components of different granularity are defined and fully specified. This model specifies also information managed by the system, that is, by each of its components. Collaboration of components inside the system architecture is specified as a potential for future distribution of components. System components provide lower grained services that in collaboration actually realize business services. Each system service component actually represents one or a collaboration of several distributed components.
- *Distribution architecture model (DAM)* specifies a distributed infrastructure of the system in terms of distribution of system services, their allocation and communication over the tiers in the n-tier architecture. This model defines in what way information is distributed over the network nodes, as a basis for future data storage and user interface design. It

Figure 3. Proposed architecture models and RM-ODP viewpoints



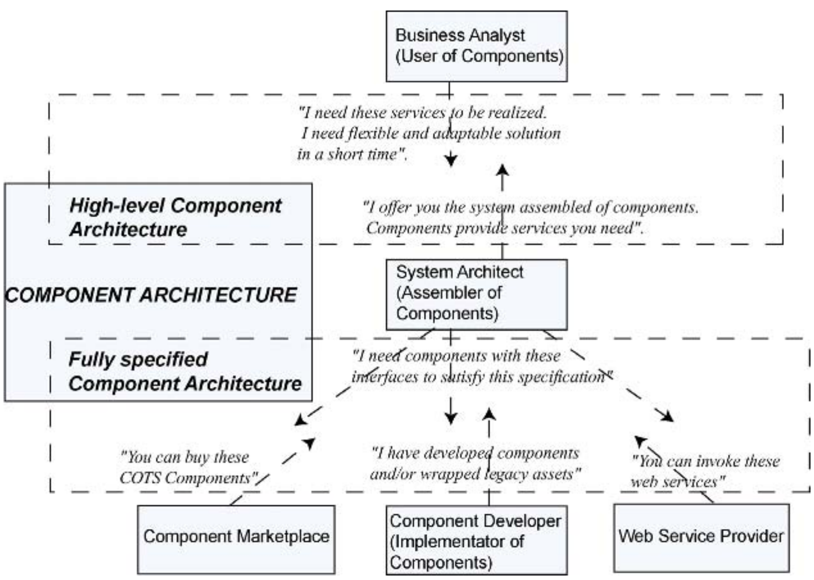
describes also the way of eventual distribution of the enterprise, for example in the case of so-called virtual enterprises. DAM uses collaboration of components inside the specified architecture as a starting point, and uses particular distributed settings to precisely define complete component-based multi-tier system architecture. Distributed components are actually system components in the distribution context. Typically business and system components can have user-interface, business logic, and permanent storage facets (one, two or all three of them), so that they can be represented as a collaboration of their subcomponents (i.e., distributed components) attached to different architecture tiers.

FUTURE TRENDS

The approach presented in this article is in line with the current OMG strategy in establishing the model driven architecture (MDA) (OMG-MDA, 2004). The MDA suggests first creating a high-level UML description of how applications will be structured and integrated, independently of any implementation details (Platform Independent Model"PIM), then moving toward more constrained UML design according to a chosen platform (Platform Specific Model"PSM), and, finally, converted into language code for a specific platform. Our service-based component approach actually aims at fully specified component-oriented PIM that can be easily mapped first to component-oriented PSM and then to a particular component middleware. In this way, components are not only implementation units but also the main artifacts in system analysis and design. The component concept is the focus of the whole development process and the point of integration and consistency between different perspectives, viewpoints and abstraction levels of the system.

The approach presented here represents a paradigm shift from components as objects to components as services. Defining components as basically service managers provides both business and IT people with a powerful and easily understandable concept for representing their views on the enterprise distributed system. Thus, a business analyst represents a system through the set of collaborating business-oriented services, a system architect specifies this service assembly in the form of the component system architecture, while a software developer provides a realization of these services using lower-level software artifacts. The approach can be effectively applied for modeling, analysis and design of systems using a new paradigm of Web services (W3C, 2004). Web services are self-contained, self-describing, modular units providing a location independent business or technical service that can be published, located and invoked across the Web. From a technical perspective, a Web service is essentially an extended and enhanced component interface constructs. Web services are natural extension of component thinking and further convergence of business and tech-

Figure 4. A role of component architecture in a dialog between actor roles



nology. By using the approach presented here, Web services become just one way of possible realization of the service-based component specification. In this way, the approach provides a smooth transition from the standard object-oriented way of thinking to the advanced Web-service paradigm.

CONCLUSION

Defined component architecture actually represents a point of negotiation between different actors in component-based system development (Figure 4). It provides a clear connection to the business processes and business domain that are the basis for the requirements of the system being developed. At the same time, this architecture provides an easy and effective mapping into appropriate implementation architecture of the system.

The higher-level component architecture, not overloaded with unnecessary details, can be used for the purpose of negotiation between the business user as the user of component system and the system architect as the assembler of components. At this level, it should be decided how business requirements should be mapped into the system solution in the best and most effective manner. It is possible to exactly determine what system parts are responsible for satisfying particular business needs, and therefore be able to replace/improve/change those parts if business is changed. Business users can easily recognize and articulate their needs through the business component architecture that hides implementation details from them.

On the other hand, the completely specified component architecture can serve as a point of negotiation between a system architect on one side, and a component developer, a provider of COTS components, or a provider of Web services, on the other side. At this level, the decision about an appropriate implementation solutions for components specified in the architecture should be made. The way of concrete realization of the component specification is not important as long as the contract between the component and its context is completely fulfilled.

REFERENCES

Allen, P., & Frost, S. (1998). *Component-based development for enterprise systems: Applying the select perspective.* Cambridge University Press.

Andersen Consulting Co., now Accenture. (1998, September 20). Understanding components. White paper. Eagle project. Available: *http://www.accenture.com*

Atkinson, C. et al. (2002). *Component-based product line engineering with UML.* Addison-Wesley.

Booch, G., Rumbaugh, J., & Jacobson, I. (1999). *The unified modeling language user guide*. Addison-Wesley.

Brown, A.W. (2000). *Large-scale component-based development*. Prentice-Hall PTR.

Cheesman, J., & Daniels, J. (2000). *UML components: A simple process for specifying component-based software.* Addison-Wesley.

Dahanayake, A., Sol, H., & Stojanovic, Z. (2003, January-March). Methodology evaluation framework for component-based system development. *Journal of Database Management, 14*(1).

D'Souza, D.F., & Wills, A.C. (1999). *Objects, components, and frameworks with UML: The Catalysis approach*. Addison-Wesley.

Gartner Group. (1997, December 5). Componentware: Categorization and cataloging. Applications Development and Management Strategies Research Note. Available: *http://www.gartnergroup.com*

Henderson-Sellers, B., Szyperski, C., Taivalsaari, A., & Wills, A. (1999). Are components objects? *OOPSLA'99 - Panel Discussion.*

Herzum, P., & Sims, O. (2000). *Business component factory: A comprehensive overview of business component development for the enterprise*. John Wiley & Sons.

Jacobson, I., Booch, G., & Rumbaugh, J. (1999). *The unified software development process*. Reading MA: Addison-Wesley.

McIlroy, M.D. (1968). Mass produced software components. Report NATO Conference on Software Engineering, October.

ODP. (1996). International Standard Organization (ISO). Reference model of open distributed processing: Overview, foundation, architecture and architecture semantics. ISO/IEC JTC1/SC07. 10746-1 to 4. ITU-T Recommendations X.901 to 904.

OMG-MDA (2004, April 15). Object Management Group – Model Driven Architecture. Information available at *http://www.omg.org/mda/*

Parnas, D.L. (1972). On the criteria to be used in decomposing systems into modules. *Communication of the ACM, 15* (12), 1053-1058.

Siegel, J. (2000). *CORBA 3: Fundamentals and Programming*. OMG Press, John Wiley & Sons.

Stojanovic, Z., & Dahanayake, A.N.W. (2002). A new approach to components. *IRMA 2002 International Conference*, Seattle Washington, USA, May 19-22.

Stojanovic, Z., Dahanayake, A.N.W., & Sol, H.G. (2001). Integration of component-based development concepts and RM-ODP viewpoints. *First Workshop on Open Distributed Processing WOODPECKER 2001,* Setubal, Portugal, July 6-7 (pp. 98-109).

Szyperski, C. (2002). *Component software* (2nd ed.). Addison-Wesley Pub Co.

Udell, J. (1994, May). *Cover story: Componentware*. Byte Magazine.

W3C World-Wide-Web Consortium (2004, April 15). Source for XML, SOAP, WSDL. Information available at: *http://www.w3c.org/*

KEY TERMS

Architecture: The fundamental organization of a system, embodied in its components, their relationships to each other and the environment, and the principles governing its design and evolution (ANSI/IEEE Standard 1471-2000).

Architecture Viewpoint: An abstraction of a set of concerns on a system derived by using a set of concepts and rules of structure (RM-ODP).

Component: An encapsulated, autonomous, service-based software unit that delivers useful services through the well-specified interface to its environment.

Component-Based Development: A software development approach where all aspects and phases of the development lifecycle are based on components.

Component Interface: The behavior of a component along with constraints at a subset of component's interactions, data types used in exposing the behavior, configuration and quality parameters of the behavior.

Component Middleware: A commercially available component technology and its associated connectivity capabilities; includes, for example, CORBA components, EJB and COM+/.NET.

Model: Represents the system from a set of concerns. A model can be informal, semi-formal and formal. A system is typically represented by a set of models, each addressing some particular area of concern.

Composition in Object-Relational Database

David Taniar
Monash University, Australia

Eric Pardede
La Trobe University, Australia

J. Wenny Rahayu
La Trobe University, Australia

INTRODUCTION

Object-Relational Database (ORDB) is increasingly popular as the database storage. Its popularity is based on its ability to capture the object-oriented modeling semantic and the maturity of relational implementation.

Many works have proposed the design method for ORDB. The design includes different data structures and relationships. One type of relationship is *composition*. It is not the same as aggregation. While aggregation is identified as a relationship in which a composite object ("whole") consists of other component objects ("parts") (Rumbaugh et al., 1991), composition has two additional constraints mentioned below.

- *Non-shareable*. This is the case when one class can only be the part of one and only one other class. Treating this type as aggregation enables other classes to own the "part" class and thus, violate the conceptual semantic.
- *Existence-dependent*. This is the case when one class can only exist with the existence of another class. Treating this as aggregation will enable a "part" class to remain in existence even though the "whole" class has been removed.

These reasons have motivated us to differentiate between the composition and the aggregation. This work will focus to preserve the composition hierarchy.

In ORDB, composition can be implemented as the attribute of row type. Row type is the constructed data type that contains a sequence of attribute names and their data types (Fortier, 1999; Melton, 2002). Row type attribute will be fully dependent and exclusive to the object that owns it.

This article aims to propose models for preserving a composition in ORDB, and in particular, we introduce the use of row type. We also propose the main queries required for the composition hierarchy.

BACKGROUND

In this section, we show a brief overview on SQL row type. We also show the related work on composition relationship mapping in object-relational database.

Row Type: An Overview

Before we start using row type in ORDB, we will briefly discuss this data type. Row type is one data type available in the Structured Query Language (SQL). This language was introduced in 1970 and has emerged as the standard language for Relational Database (RDB) (Melton, Simon, & Gray, 2001). It is used for database definition and manipulation.

Along with the establishment of RDB and the emergence of ORDB, SQL has undergone many changes, one of which is the additional data types. It is required to accommodate complex data structures. One of the data type is Row Type.

Fortier (1999) and Melton (2002) define row type as constructed type that contains a sequence of attribute names and their data types. This type is actually not a new data type in the database system. It has been used even since the legacy data model era (CODASYL Database Task Group, 1971).

After the emergence of relational model, there is also a data model that is aimed at capturing the nested structure such as row type in relations. The model is called Nested Relational Model (Jaeschke & Schek, 1982; Roth & Korth, 1987). Nevertheless, traditional relational model still dominates the database community. Even until recently there is no commercial DBMS which has chosen to implement the Nested Relational Model even in its original form (Elmasri & Navathe, 2002).

Not until the release of SQL 1999, relational model recognizes row type as one data type that can enrich its data structure (Fortier, 1999). In SQL4, it is even possible to have varying levels of row type (Melton, 2002). It has become a powerful means to capture real world problems that can rarely be represented by a simple flat table. General syntax for the SQL4 row type is shown as follows:

SQL4 Row Type

CREATE TABLE <table schema>
($attr_1$ data type
CONSTRAINT $attr_1$ **PRIMARY KEY**, ...,
$attr_i$ **ROW** ($attr_{i1}$ data type,...,
$attr_{ij}$ data type));

Existing Mapping Methods for Composition in ORDB

Mapping method can be defined as a formal process of transforming a schema level (such as conceptual schema) to another schema level (such as logical schema) in a database system. A mapping method works effectively if the schema result is complied with the requests. If the result schema does not preserve all semantics of the requests, the mapping method is not effective.

Some works on mapping methods have tried to preserve composition hierarchy in ORDB. However, no work utilizes the row data types that were introduced by SQL 1999 (Fortier, 1999) and is enriched in SQL4 (Melton, 2002). Very often the "part-of" relationship is either flattened or split into an entirely separate table (see Figure 1).

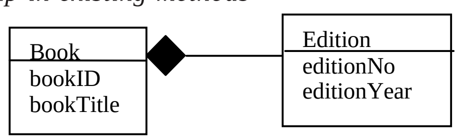
The most common practice of implementing composition relationship is by separating the "part" component in another table with composite PKs. This practice only uses object-oriented paradigm for conceptual modeling. The implementation is purely relational (Ambler, 1997; Hsieh & Chang, 1993).

Marcos, Vela, Cavero and Caceres (2001) propose a design of aggregation and composition in ORDB. In the logical and the implementation level, they use the nested table for the "part" component. Two problems arise. Firstly, nested table is a vendor-specific feature. It is not recognized in SQL standard. Secondly, the implementation of nested table cannot represent the composition type because actually the "part" component is stored in a separate table.

Rahayu and Taniar (2002) use two different ways of preserving this type of relationship. It uses index clustering for relationships that requires ordering semantics, and nested tables for the one which does not. However, this work can only be applicable for regular aggregation relationship.

We find that the existing works either have not preserved the composition semantic or have preserved it as

Figure 1. Composition relationship in existing methods



Composition flattened into separate tables

Book	
PK	
book_id	book_title
EN1	Fundamentals of Database System

Book-Edition		
PK, FK	PK	
book_id	ed-no	ed-year
EN1	1	1989
EN1	2	1997
EN1	3	2000

Composition split in nested table

Book		
book_id	book_title	edition
EN1	Fundamental of Database System	♦

edition
•
•
•

ed-no	ed-year
1	1989
2	1997
3	2000

Composition flattened using cluster

Book	
Cluster	
book_id	book_title
EN1	Fundamentals of Database System

Edition		
Cluster		
book_id	ed-no	ed-year
EN1	1	1989
EN1	2	1997
EN1	3	2000

a usual aggregation. Therefore, we aim to propose new methods that can preserve the composition hierarchy in ORDB.

PROPOSED MAPPING METHODS

We propose the mapping methods of composition relationship into ORDB using row type. Our mapping method aims to transform the composition relationship into ORDB tables.

To model the conceptual schema, we use the Unified Modeling Language (UML) notation (Booch, Rumbaugh & Jacobson, 1999). It is a powerful notation to specify, visualize, and construct conceptual schema. Using UML, we can distinguish aggregation and composition easily. For example, in Figure 2a, composition is shown by a full diamond shape symbol. If it is an aggregation relationship, the diamond shape symbol will not be filled.

We map the "whole" class as the table and the "part" class as the row type attribute. For multiple-row, we use multiset data types (Melton, 2002). Other types of collection are also possible depending on the requirements for ordering and duplication semantic.

This work has increased the usage of ORDB for different database domains. In Pardede, Rahayu, and Taniar (2004), a subset of our proposed mapping methods has been used to accommodate XML data storage. It shows the utility and the practical importance of the work.

Single Row Composition

If the "part" class of the composition is single, we can map the class as a single row (see Figure 2a).

Rule1: For two classes namely C_1 and C_2 with elements/attributes (A,B) and (M,N) respectively, if C_1 is composed by at most one C_2, implement C_2 as a single row attribute of table C_1. Transformation result is:

$$\text{Table } C_1 \, (A, B, \overset{1}{\underset{0}{Row}} \; C_2 \, (M, N)).$$

Example 1: Class AUTHOR is composed by at most one type ADDRESS (see Figure 2b). The composition type will be mapped into a single row attribute in ORDB table. The transformation into ORDB is shown in the following:

```
CREATE TABLE Author
(AuthorID CHARACTER VARYING(20)
 CONSTRAINT authorid_pk PRIMARY KEY,
 AuthorName CHARACTER VARYING(2),
 Address ROW (Street CHARACTER VARYING(50),
City CHARACTER VARYING(30),
Country CHARACTER VARYING(5)));
```

The proposed method has captured the real semantic of composition hierarchy. In ORDB, the "part" class is physically implemented inside the "whole" class. By doing this, we avoid another class sharing the particular "part" class (non-shareable constraint). We also ensure that on removal of the "whole" class, we remove all "part" classes that are defined inside it (existence-dependent constraint).

In Example 1, the *address* data has a composition relationship to the *author*. It means that the *address* can only be attached within the *author* data and its existence is dependent on the *author* data.

Multiple Row Composition

If the "whole" class can have more than one "part" class of the same type, we use a multiple row (see Figure 3a).

Figure 2. Composition for single row

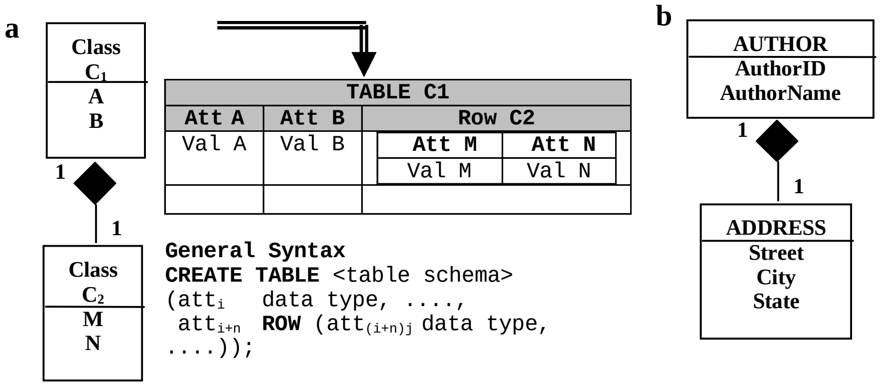


Figure 3. Composition for multiple row

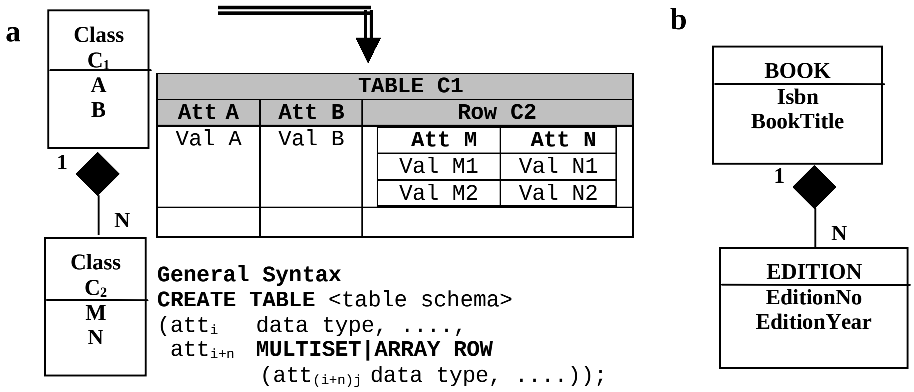
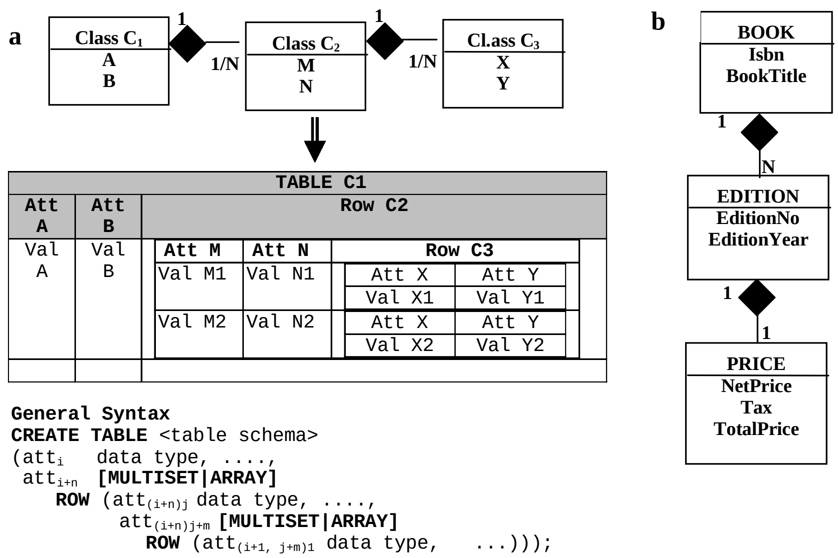


Figure 4. Composition for multilevel row



Rule 2: For two classes namely C_1 and C_2 with elements/attributes (A,B) and (M,N) respectively, if C_1 can be composed by more than one C_2, implement C_2 as a multiple row attribute of table C_1. Transformation result is:

$$Table\ C_1\ (A, B, \overset{n}{\underset{0}{Row}}\ C_2\ (M, N))$$

Example 2: Class BOOK can be composed by more than one class EDITION (see Figure 3b). The composition class will be mapped into a multiple-row attribute in ORDB table. The transformation into ORDB is shown in the following:

CREATE TABLE Book
(Isbn CHARACTER VARYING(5)
CONSTRAINT isbn_pk PRIMARY KEY,
BookTitle CHARACTER VARYING(50),
Edition MULTISET (**ROW** (EditionNo NUMBER,
EditionYear NUMBER)));

This example shows that EDITION class is mapped as a collection of row attribute in table *Book*. If we remove a *book* details, all details of the book *edition* will also be removed.

Multilevel Row Composition

In ORDB, supported by SQL4, we can have a row inside another row attribute. The data structure will be more complex than the flat relationship in a conventional model, but sometimes it resembles the real world problem.

If the "part" class of the composition can be composed by other "part" class, we use multi-level row (see Figure 4a).

Rule 3: For three classes namely C_1, C_2 and C_3, with elements/attributes (A,B), (M,N), and (X, Y) respectively, if C_1 is composed by C_2 and C_2 is composed by C_3, implement the last two types as a multi-level row attribute of C_1. Transformation result is *Table* $C_1(A, B, \underset{0}{\overset{1/n}{Row}}\ C_2(M, N, \underset{0}{\overset{1/n}{Row}}\ C_3\ (X, Y)))$

Example 3: Continuing Example 2, now class EDITION can be composed by at most one class PRICE (see Figure 4b). The composition class will be mapped into a multi-level row attribute in ORDB table.

```
CREATE TABLE Book
(Isbn CHARACTER VARYING(5)
 CONSTRAINT isbn_pk PRIMARY KEY,
 BookTitle CHARACTER VARYING(50),
 Edition MULTISET
(ROW (EditionNo NUMBER,
EditionYear NUMBER,
Price ROW (NetPrice DECIMAL(3,2),
Tax DECIMAL(3,2),
TotalPrice DECIMAL(3,2)))));
```

In this example, the class PRICE is mapped as *price* row attribute inside *edition* row attribute. And we have *edition* as the row attribute of *Book* table. If we remove a *book,* all details of the *book edition* and its price will also be removed.

COMPOSITION QUERIES

Our previous section describes the proposed methods for preserving composition in ORDB using row types. This section shows how the queries can be performed in the data structures.

We describe two type of queries namely *part queries* and *whole queries*. Part query is a query used to retrieve the information of "part" classes with selection predicates originated at the "whole" class. Whole query is a query used to retrieve information of "whole" class with selection predicates originated at the "part" classes.

Single Row Composition Query

Part query and whole query general syntax for single row composition is shown as follows. SQL 1999 provides double dot notation (Fortier, 1999) to specify a subpart of a table such as row type element.

Part Query:

```
SELECT <alias1..”part”_class_attributes>
FROM <“whole”_class_table alias1>
WHERE <alias1.”whole”_class_attribute = &input_class_selection_predicates>
```

Whole Query:

```
SELECT <alias1.”whole”_class_attributes>
FROM <“whole”_class_table> alias1
WHERE <alias1..”part”_class_attribute = &input_class_selection_predicates>
```

An example of a part query is to retrieve the street and the city of a given author id (e.g., AuthorID = 'RE01'). The SQL statement for this query is shown in the following:

```
SELECT A..street, A..city
FROM Author A
WHERE A.AuthorID = ‘RE01’;
```

An example of a whole query is to retrieve an author name who lives in Melbourne, Australia. The SQL statement for this query is shown in the following:

```
SELECT A.AuthorName
FROM Author A
WHERE A..city = ‘Melbourne’
AND A..country = ‘Australia’;
```

Multiple Row Composition Query

Part query and whole query general syntax for multiple row composition is shown as follows. TABLE keyword is used to treat collection types such as multiset as a table, allowing querying in the same way that any base table is accessed.

Part Query:

```
SELECT <alias2..”part”_class_attributes>
FROM <“whole”_class_table alias1, TABLE(alias1.”part”_class_name) alias2>
WHERE <alias1.”whole”_class_attribute = &input_class_selection_predicates>
```

Whole Query:

```
SELECT <alias1.”whole”_class_attributes>
FROM <“whole”_class_table alias1,
```

TABLE(alias1.”part”_class_name) alias2>
WHERE <alias2..”part”_class_attribute = &input_class_selection_predicates>

An example of a part query is to retrieve the edition details of a given book title (e.g., BookTitle = ‘Database Systems’). The SQL statement for this query is shown in the following:

SELECT E..EditionNo, E..EditionYear
FROM Book B, TABLE(b.edition) E
WHERE B.BookTitle = ‘Database Systems’;

An example of a whole query is to retrieve the book title of given edition (e.g., EditionNo = ‘5’) and published after 2002. The SQL statement for this query is shown in the following:

SELECT B.BookTitle
FROM Book B, TABLE(b.edition) E
WHERE E..EditionNo = 5
AND E..EditionYear > 2002;

Multilevel Row Composition Query

As we know, multi-level row is formed by more than one single row and/or multiple row attribute. Therefore, the general syntax is really dependent on the combination of the row in different levels. For example, if the first level of row is multiple and the second level is a single row, the general syntax is as follows.

Part Query:
SELECT <alias2..”part2"_class_attributes>
FROM <“whole”_class_table alias1, **TABLE**(alias1.”part1"_class_name) alias2>
WHERE <alias1.”whole”_class_attribute = &input_class_selection_predicates>
AND <alias2..”part1"_class_attribute = &input_class_selection_predicates>

Whole Query:
SELECT <alias1.”whole”_class_attributes>
FROM <“whole”_class_table alias1, **TABLE**(alias1.”part2"_class_name) alias2>
WHERE <alias2..”part1"_class_attribute = &input_class_selection_predicates>
AND <alias2..”part2"_class_attribute = &input_class_selection_predicates>

An example of a part query is to retrieve the total price of a given book title (e.g., BookTitle = ‘Database Systems’) and a specific edition number (e.g., EditionNo = 5). The SQL statement for this query is shown in the following:

SELECT E..TotalPrice
FROM Book B, TABLE(b.edition) E
WHERE B.BookTitle = ‘Database Systems’
AND E..EditionNo = 5;

An example of a whole query is to retrieve the book title of given maximum price $50 and oldest edition accepted (e.g., Edition No = 3) and published after 2002. The SQL statement for this query is shown in the following:

SELECT B.BookTitle
FROM Book B, TABLE(b.edition) E
WHERE E..EditionNo > 2
AND E..TotalPrice < 50;

FUTURE TRENDS

In the future, we will require database storage that can capture the semantic of the applications with a high level of correctness. This article demonstrates how composition semantics should be distinguished from the aggregation semantic. We expect future works utilizing many new SQL data types to accommodate different semantic constraints.

At the time of writing, the queries for row components are not sophisticated since we have to treat the row components as a table. In the future, the SQL standardization body will need to enable users to perform a direct query of the row components.

Finally, more applications are expected to use row data type. So far, row data type is typically used for single row composition (such as address attribute). This article has demonstrated how this data type can be applicable for more complex structures.

CONCLUSION

It is evident that the previous works on ORDB transformation have not captured the semantic of composition in their mapping results. It has limited the usage of ORDB in some database problems, especially those that require specific relationship constraints to be met.

In this article, we have shown how the composition relationship can be preserved in the implementation using ORDB. We propose the usage of single-row, multiple-row, and multi-level rows for different cardinality and different levels in the composition hierarchy. We also provide two main important queries for the composition hierarchy, the part queries and the whole queries.

Unlike other works in composition transformation, our method preserves the "part" component as the part of the "whole" component. By doing this, we maintain the non-shareable and existence dependent constraints from the conceptual level.

Finally, this work provides general methods for ORDB. It can be used for any database domain that uses ORDB for its storage solution. A subset of this work has been implemented for XML documents storage.

REFERENCES

Ambler, S.W. (1997). *Mapping objects to relational databases.* AmbySoft Inc.

Booch, G., Rumbaugh, J. & Jacobson, I. (1999). *The unified modeling language users guide.*

CODASYL Database Task Group. (1971). Data base task group report.

Elmasri, R. & Navathe, S.B. (2002). *Fundamentals of database systems*.

Fortier, P. (1999). *SQL3 implementing the SQL foundation standard*.

Hsieh, S.Y. & Chang, C.K. (1993). Capturing the object-oriented database model in relational form. *Proceedings of the IEEE International Computer Software and Application Conference.*

Jaeschke, G. & Schek, H,-J. (1982). Remarks on the algebra of non first normal form relations. *Proceedings of ACM Symposium on Principles of Database Systems*, (pp.124-138).

Marcos, E., Vela, B., Cavero, J.M., & Caceres, P. (2001). Aggregation and composition in object-relational database design. *Proceedings of East-European Conference on Advances in Database and Information Systems*, (pp.195-209), Springer.

Melton, J. (Ed.). (2002). *Database language SQL – Part 2 foundation*. ISO-ANSI WD 9072-2.

Melton, J., Simon, A.R., & Gray, J. (2001). *SQL: 1999 – Understanding relational language components.*

Pardede, E., Rahayu, J.W., & Taniar, D. (2004). On using collection for aggregation and association relationships in XML object-relational storage. *Proceedings of ACM Symposium on Applied Computing.*

Rahayu, J.W., & Taniar, D. (2002). Preserving aggregation in an object-relational DBMS. *Proc. International Conference on Advances in Information Systems*, (pp.1-10).

Roth, M.A., & Korth, H.F. (1987). The design of ¬1NF relational databases into nested normal form. *Proceedings of ACM SIGMOD International Conference on Management of Data*, (pp.143-159).

Rumbaugh, J. et al. (1991). *Object-oriented modelling and design*.

KEY TERMS

Aggregation: A relationship in which a composite object ("whole") consists of other component objects ("parts").

Composition: A specific form of an aggregation. It is a relationship in which a composite object consists of non-shareable objects, and the latter are existence-dependent to the former.

Mapping Method: A formal process of transforming a schema level (such as conceptual schema) to another schema level (such as logical schema) in a database system.

Object-Relational Database: Database that is developed using object-oriented concept and is implemented in relational environment. This database was introduced in the 90s and since then has been widely used for database applications.

Part Query: Query to retrieve the information of "part" classes with selection predicates originated at the "whole" class.

Row Type: Constructed data type that contains a sequence of attribute names and their data types.

Structured Query Language (SQL): A standard interactive and programming language for getting information from and updating a database. It was developed by IBM in the mid-1970s for relational database management systems.

Unified Modeling Language (UML): A notation to specify, visualize, construct real-world objects. It is a standard modeling technique for object-oriented design methodology. It is also frequently used for object-relational database since this data model has object-oriented features.

Whole Query: Query to retrieve information of "whole" class with selection predicates originated at the "part" classes.

C

Computer Attitude and Anxiety

Pieter Blignaut
University of the Free State, South Africa

Andries Burger
University of the Free State, South Africa

Theo McDonald
University of the Free State, South Africa

Janse Tolmie
University of the Free State, South Africa

INTRODUCTION

Computers in the workplace are a given. Although the advantages of computers are well-known and proven, many people still try to avoid using them. It is extremely important to find out which factors influence the success of end-user computing. What are the reasons that some people excel on a computer while others have problems and even build up a resistance toward the use of computers?

This chapter provides a literature-based overview of computer attitude and computer anxiety as factors that influence a user's resistance, commitment, and achievement. A graphic model, according to which the interactions between computer attitude and anxiety, their causes, indicators, and impacts may be understood, is proposed. It is put forth that external strategies to deal with anxiety and a negative attitude are imperative to break down a snowballing effect of cause and effect and to ensure effective end-user computing.

BACKGROUND

Computer Attitude

Gordon Allport (1935) defined the concept of attitude, in general, as follows: "An attitude is a mental and neural state of readiness, organized through *experience*, exerting a directive or dynamic *influence* upon the individual's *response* to all objects and situations with which it is related" (p. 810). In other words, attitude is determined by experience and impacts upon the individual's behavior.

A person's attitude toward a computer is influenced by a variety of aspects, e.g., the social issues relating to computer use (Popovich et al., 1987), computer liking, computer confidence, computer anxiety or comfort (Delcourt & Kinzie, 1993; Loyd & Gressard, 1984a), achievement (Bandalos & Benson, 1990), usefulness, and value (Francis-Pelton & Pelton, 1996).

Computer Anxiety

According to Henderson et al. (1995) anxiety is viewed as "a drive that motivates the organism to avoid the stimulus for anxiety" (p. 24). This implies that an individual will avoid the use of a computer in the presence of computer anxiety and if possible.

Kaplan and Sadock (1998) referred to anxiety as "a diffuse, unpleasant, vague sense of apprehension, often accompanied by autonomic symptoms" (p. 581). Specifically, computer anxiety involves an array of emotional reactions, including fear, apprehension, uneasiness, and distrust of computer technology in general (Negron, 1995; Rohner & Simonson, 1981).

Computer anxiety is also influenced by a variety of aspects, e.g., general anxiety and confidence (Harrison & Rainer, 1992), computer liking (Chu & Spires, 1991; Loyd & Gressard, 1984b), impact of computers on society (Raub, 1981), equipment-related anxiety (Marcoulides, 1989), comfort and value (Violato et al., 1989), and corporate pressure.

The Relationship between Computer Attitude and Computer Anxiety

Computer anxiety is often included as a component of attitude (Delcourt & Kinzie, 1993; Loyd & Gressard, 1984a). Jawahar and Elango (2001) reported, however, that previous studies used the concepts of computer anxiety and negative attitudes toward computers interchangeably. Computer anxiety is, however, not solely responsible for a negative attitude. A person can have a negative attitude toward computers even though he or she is not overly anxious about using them. This may be because of a negative experience, e.g., an apologizing clerk blaming an erroneous account statement on the computer.

Furthermore, attitude allows for both a negative and a positive grading, whereas anxiety is, by definition, either negative or absent.

MAIN THRUST OF THE CHAPTER: A MODEL FOR INTERACTION

In order to indicate the various influences on the mental states of computer attitude and computer anxiety and the effect they have on a user's ability to execute computer-related tasks effectively, a model for interaction was developed (Figure 1).

The model shows interaction on three levels of abstraction. The right-hand column resembles a typical flow diagram but with an adapted convention. It shows the sequence of mental and operational events when a user is confronted with a task to be done on the computer. The diamond symbols do not represent conscious decisions but rather indicate general user behavior as determined by the user's current levels of computer attitude, computer anxiety, knowledge, and pressure experienced.

As an example of how to read the flow diagram, consider a user who has to perform a computer task. If his or her level of computer anxiety is not above a specific critical level (*D1*), he or she has a positive attitude toward computer tasks (*D2*). If he or she knows how to perform the task (*D4*), he or she will do the task (*P2*). If the user's knowledge is inadequate, this person will go through a process of learning (*P1*) until he or she can do the task. If the anxiety level is high (*D1*), the user will only use the computer if forced to do so (*D3*), or else he or she will opt out of the task or do it without a computer (*P3*).

The middle column in Figure 1 indicates the user's current levels of computer anxiety and computer attitude. The influence that computer anxiety and computer attitude have on each other as well as their influence on user behavior and the processes of learning and task execution is indicated with curved arrows (*E5–E11*). It is also clear that task execution (computer experience, *P2*) impacts computer attitude and anxiety in return (*E12–E13*).

The left-hand column shows the external processes and factors that influence a user's levels of computer attitude and anxiety.

Further discussion in this chapter serves to substantiate the claimed influences from the literature.

Factors that Determine Computer Attitude

Several studies have been undertaken to explore potential factors associated with a positive attitude toward computers (Brodt & Stronge, 1986; Scarpa et al., 1992; Sultana, 1990; Schwirian et al., 1989; Bongartz, 1988; Burkes, 1991). Some of the factors that were considered were level of education, years of experience in the work environment, computer experience, age, gender, and job title (*E3* and *E13*). The only factor that was repeatedly, although not consistently, found to have a positive effect on computer attitude, was computer experience (*E13*).

Causes of Computer Anxiety

According to Torkzadeh and Angulo (1992), there are three perspectives of computer anxiety: psychological (*E1*), sociological (*E1*), and operational (*E1* and *E12*). From a psychological perspective, users may fear that they will damage the computer, feel threatened when having to ask younger workers for help, or feel that they are losing control because computers are perceived as a threat to one's power and influence. From a sociological perspective, people have a need for social contact with other people, and because computers can change existing social patterns, they find the situation unbearable. People may also have a fear of computers replacing them. From an operational point of view, people want to avoid embarrassment connected with their inability to type or to use the keyboard. An initially confident user might be disillusioned with the complexity and sophistication of computer systems and procedures after a first experience (*E12*).

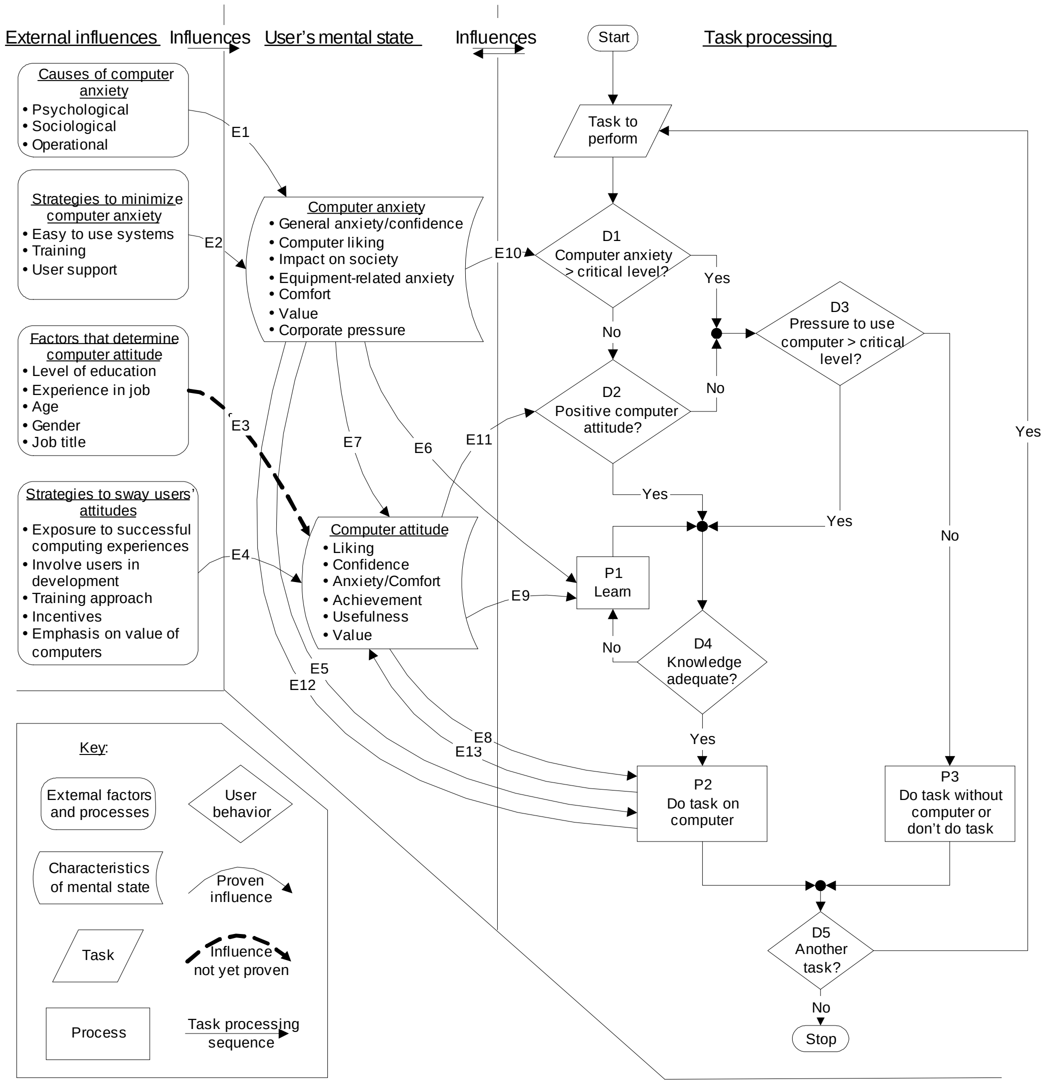
Effects of a Positive Attitude

According to Ngin et al. (1993), individuals with work excitement express creativity, receptivity to learning, and have the ability to see opportunity in everyday situations. Positive attitudes enhance the learning process (Shneiderman, 1980) (*E9*), specifically the motivation to learn and the ability to retain information in a given situation (Jawahar & Elango, 2001).

A negative attitude may lead to computer resistance (Sheiderman, 1980) (*D2*, *D3*, and *P3*), a phenomenon that can be found among experienced as well as inexperienced users (Negron, 1995). A negative attitude may even lead to defamation or sabotage of computer technology (Gibson & Rose, 1986).

A person's attitude toward computers and related technology could determine his or her performance with

Figure 1. Model for interaction



the technology and the satisfaction he or she draws from the experience (*E8*), although contradictory results are reported in the literature. Nickell and Pinto (1986) as well as Jawahar and Elango (2001) found a positive correlation between scores on a computer attitude scale and the final course grades of students enrolled in an introductory computer class. Other studies on students report an inverse relation between attitudes and performance (Hayek & Stephens, 1989; Marcoulides, 1988; Mawhinney & Saraswat, 1991), while still other researchers report no relationship between these two constructs (O'Quin et al., 1987; Kernan & Howard, 1990; Szajna & Mackay, 1995).

It was indicated above that scales to measure computer attitude differ with regard to the components they include. Jawahar and Elango (2001) also indicated that some tests measure attitudes toward working with computers, while others have components measuring general attitudes toward computers. These might be possible reasons for the contradictory results regarding the impact of computer attitude on computing performance. One cannot expect to obtain consistent results if the instruments used differ substantially from one another. This is one area in this field of research where agreement and subsequent standardization are yet to be achieved.

Impact of Computer Anxiety

Anxiety affects people's thinking, perception, and learning (Kaplan & Sadock, 1998). It also produces confusion and distortions of perception relating to time, space, people, and the meanings of events. These distortions usually have a negative effect on learning ability by lowering concentration, reducing recall ability, and impairing the ability to make associations (*E6*).

Specifically, computer anxiety can be recognized by a fear expressed regarding present or future interactions with computers or computer-related technology, negative global attitudes about computers (*E7*), or self-critical internal dialogues during computer interactions (Rosen & Weil, 1995).

People with computer anxiety experience the incidence of physical, phobic symptoms, such as stomach cramps and cold sweats, as well as psychological symptoms, such as a resistance to use and a negative attitude toward a system (Shneiderman, 1980).

Computer anxiety seems to be a good predictor of computer achievement (*E5*). Marcoulides (1988) found that computer anxiety influenced the effectiveness with which college students could utilize the computer. Rosen et al. (1987) also found that computer-anxious undergraduate students enrolled in computer-based courses were twice as likely to drop out of the course and received lower course grades than nonanxious students.

Although anxiety usually has a negative connotation, there appear to be optimal levels of anxiety that help people to function effectively (Higgins, 1989) and that make them more alert and aware of what is going on (Lugo & Hershey, 1981; Beck & Emery, 1985) (*E5*).

Strategies to Sway Users' Attitudes

Previously, it was indicated that a negative computer attitude might lead to computer resistance. It is, therefore, important to identify strategies to enhance or sway users' attitudes toward computer use. Possible strategies are as follows (*E4*):

- Expose the users to successful and positive computer experiences (Yaghmaie et al., 1998).
- Make users aware of the fact that their computer literacy would be an advantage in any future work environment (Emmet, 1988).
- Involve users actively in the implementation of computer systems (Barry & Gibbons, 1990).
- Emphasize the ease with which work can be accomplished using a computer function (Marasovic et al., 1997).
- In training, stress the fact that computer use leads to desired outcomes (Marasovic et al., 1997).
- Include in training programs activities aimed at increasing self-efficacy (Henderson et al., 1995).
- Introduce incentives in order to convince and motivate users (Sultana, 1990).
- Consider the user's attitude and motivation when system designers design the user interface (Galitz, 1997).

Strategies to Minimize Computer Anxiety

Computer anxiety is something that will not just disappear on its own, but it is something that has to be dealt with. It is considered to be a temporal emotional state rather than a permanent personality trait and, therefore, can be remedied through positive computing experiences (Cambre & Cook, 1987). Possible strategies are as follows (*E2*):

- Easy-to-use computer systems (Appelbaum & Primmer, 1990)
- Basic training for the most routine tasks (Cambre & Cook, 1987; Flaughler, 1986; Lewis, 1988)
- Advanced training, which is essential to keep users interested as they progress (Appelbaum & Primmer, 1990)
- User support in the form of manuals, consultants, user groups, and computer specialists (Appelbaum & Primmer, 1990)

FUTURE TRENDS

It was argued above that the variety with regard to components and focus of the scales used by researchers to measure computer attitudes might be responsible for the contradictory results regarding the impact of computer attitude on computing performance. If a measuring scale for both computer attitude and computer anxiety can be standardized, much of the current confusion that exists in the literature could be solved. For example, the definition of computer comfort as it is given under "Terms and Definitions" below is merely the opposite of that for computer anxiety. This means that a standardization of measuring scales would include a consolidation of terminology.

CONCLUSION

The use of a graphic model provides an efficient way of understanding the interaction between computer attitude and computer anxiety, together with their causes, indicators, impacts, and strategies to deal with them.

From this model, it can be deduced that there are two-way interactions between computer attitude and computer anxiety on the one hand and computer task execution on the other. A vicious circle can be established where anxiety inhibits performance (*E5*), which in turn, produces a negative attitude toward the system (*E13*) and further slows the process of learning to use the system (*E9*).

External strategies can be applied to reduce computer anxiety and sway users' attitudes toward the use of computers (*E1–E4*). This will then lead to better achievement (E5 and *E8*), which will, in turn, lead to self-confidence (*E12*) and appreciation of the value of computers in general (*E13*). Less anxiety and a positive attitude will support the learning process (*E6* and *E9*), which will eventually result in better achievement.

Because of the proven effect of computer experience on attitude, moderate pressure can be applied to ensure that users do not avoid the use of computers and, in this way, never get the opportunity to gain experience (*D3*). This pressure can be explicit by means of assignment or by applying measures to create intrinsic motivation so that the user drives him- or herself.

REFERENCES

Allport, G. W. (1935). Attitudes. In C. M. Murchison (Ed.), *Handbook of social psychology* (pp. 796–834). Worcester, MA: Clark University Press.

Appelbaum, S. H., & Primmer, B. (1990). An RHx for computer anxiety. *Personnel, 87*(9), 8–11.

Bandalos, D., & Benson, J. (1990). Testing the factor structure invariance of a computer attitude scale over two grouping conditions. *Educational and Psychological Measurement, 50*(1), 49–60.

Barry, C. T., & Gibbons, L. K. (1990). Information systems technology: Barriers and challenges to implementation. *Journal of Nursing Administration, 20*(2), 40–42.

Beck, A. T., & Emery, G. (1985). *Anxiety disorders and phobias: A cognitive perspective*. New York: Basic Books.

Bongartz, C. (1988). Computer-oriented patient care: A comparison of nurses' attitudes and perceptions. *Computers in Nursing, 6*, 204–210.

Brodt, A., & Stronge, J. H. (1986). Nurses' attitudes toward computerization in a mid-western community hospital. *Computers in Nursing, 4*, 82–86.

Burkes, M. (1991). Identifying and relating nurses' attitudes toward computer use. *Computers in Nursing, 9*, 190–201.

Cambre, M. A., & Cook, D. L. (1987). Measurement and remediation of computer anxiety. *Educational Technology, 27*(12), 15–20.

Chu, P. C., & Spires, E. C. (1991). Validating the computer anxiety rating scale: Effects of cognitive style and computer courses on anxiety. *Computers in Human Behavior, 7*, 7–21.

Delcourt, M. A. B., & Kinzie, M. B. (1993). Computer technologies in teacher education: The measurement of attitudes and self-efficacy. *Journal of Research and Development in Education, 27*(1), 35–41.

Emmet, A. (1988). Overcoming computer resistance. *Personal computing, 7*(12), 80–98.

Flaughler, P. O. (1986). Computer training for nursing personnel. *Computers in nursing, 4*(3), 105–108.

Francis-Pelton, L., & Pelton, T. W. (1996). *Building attitudes: How a technology course affects preservice teachers' attitudes about technology*. Retrieved from *http://www.math.byu.edu/~lfrancis/tim's-page/attitudesite.html*

Galitz, W. O. (1997). *The essential guide to user interface design: An introduction to GUI design principles and techniques*. New York: John Wiley & Sons.

Gibson, S. E., & Rose, M. A. (1986). Managing computer resistance. *Personal Computing, 4*(5), 201–204.

Harrison, A. W., & Rainer, K. (1992). An examination of the factor structures and concurrent validities for the computer attitude scale, the computer anxiety scale, and the computer self-efficacy scale. *Educational and Psychological Measurement, 52*, 735–745.

Hayek, L. M., & Stephens, L. (1989). Factors affecting computer anxiety in high school computer science students. *Journal of Computers in Mathematics and Science Teaching, 8*, 73–76.

Henderson, R. D., Deane, F. P., & Ward, M. (1995). Occupational differences in computer related anxiety: Implications for the implementation of a computerised patient management information system. *Behaviour and Information Technology, 14*(1), 23–31.

Higgins, D. L. (1989). Anxiety as a function of perception: A theory about anxiety and a procedure to reduce symptoms to manageable levels. In M. D. Yapko (Ed.), *Brief therapy approaches to treating anxiety and depression* (pp. 245–263). New York: Brunner/Mazel, Inc.

Jawahar, I. M., & Elango, B. (2001). The effect of attitudes, goal setting and self-efficacy on end user performance. *Journal of End User Computing, 13*(2), 40–45.

Kaplan, H. I., & Sadock, B. J. (1998). *Synopsis of psychiatry: Behavioural sciences/clinical psychiatry* (8th ed.). Baltimore, MD: Lippencott Williams & Wilkins.

Kernan, M. C., & Howard, G. S. (1990). Computer anxiety and computer attitudes: An investigation of construct and predictive validity issues. *Educational and Psychological Measurement, 50*, 681–690.

Lewis, L. H. (1988). Adults and computer anxiety: Facts or fiction? *Lifelong Learning, 11*(8), 6–12.

Loyd, B. H., & Gressard, C. (1984a). The effects of sex, age, and computer experience on computer attitudes. *Association for Educational Data Systems, 18*(2), 67–77.

Loyd, B. H., & Gressard, C. (1984b). The reliability and factorial validity of computer attitude scales. *Educational and Psychological Measurement, 44*, 501–505.

Lugo, J. O., & Hershey, G. L. (1981). *Living psychology*. New York: Macmillan.

Marasovic, C., Kenney, C., Elliott, D., & Sindhusake, D. (1997). Attitudes of Australian nurses toward implementation of a clinical information system. *Computers in Nursing, 15*(2), 91–98.

Marcoulides, G. A. (1988). The relationship between computer anxiety and computer achievement. *Journal of Educational Computing Research, 4*, 151–158.

Marcoulides, G. A. (1989). Measuring computer anxiety: The Computer Anxiety Scale. *Educational and Psychological Measurement, 49*, 733–739.

Mawhinney, C. H., & Saraswat, S. P. (1991). Personality type, computer anxiety, and student performance. *Journal of Computer Information Systems, 8*, 110–123.

Negron, J. A. (1995). The impact of computer anxiety and computer resistance on the use of computer technology by nurses. *Journal of Nursing Staff Development, 11*(3), 172–175.

Ngin, P., Simms, L., & Erbin-Roesemann, M. (1993). Work excitement among computer users in nursing. *Computers in Nursing, 3*, 127–133.

Nickell, G. S., & Pinto, J. N. (1986). The computer attitude scale. *Computers in Human Behavior, 2*, 301–306.

O'Quin, K., Kinsey, T. G., & Beery, D. (1987). Effectiveness of a microcomputer-training workshop for college professionals. *Computers in Human Behaviour, 3*, 85–94.

Popovich, P. M., Hyde, K. R., & Zakrajsek, T. (1987). The development of the attitudes toward computer usage scale. *Educational and Psychological Measurement, 47*, 261–269.

Raub, A. C. (1981). *Correlates of computer anxiety in college students*. Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia.

Rohner, D. J., & Simonson, M. R. (1981). *Development of an index of computer anxiety*. Paper presented at the annual convention of the Association of Educational Communications and Technology, Philadelphia, PA.

Rosen, L. D., Sears, D. C., & Weil, M. M. (1987). Computerphobia. *Behavior Research Methods, Instruments, and Computers, 19*, 167–179.

Rosen, L. D., & Weil, M. M. (1995). Adult and teenage use of consumer, business, and entertainment technology: Potholes on the information highway? *Journal of Consumer Affairs, 29*(1), 55–84.

Scarpa, R., Smeltzer, S. C., & Jasion, B. (1992). Attitudes of nurses toward computerisation: A replication. *Computers in Nursing, 10*, 72–80.

Schwirian, P., Malone, J. A., Stone, V. J., Nunley, B., & Francisco, T. (1989). Computers in nursing practice: A comparison of the attitudes of nurses and nursing students. *Computers in Nursing, 7*, 168–177.

Shneiderman, B. (1980). *Software psychology*. Cambridge, MA: Winthrop.

Sultana, N. (1990). Nurses' attitudes toward computerization in clinical practice. *Journal of Advanced Nursing, 15*, 696–702.

Szajna, B., & Mackay, J. M. (1995). Predictors of learning performance in a computer-user training environment: A path-analytic study. *International Journal of Human–Computer Interaction, 7*(2), 167–185.

Torkzadeh, G., & Angulo, I. E. (1992). The concept and correlates of computer anxiety. *Behavior and Information Technology, 11*(2), 99–108.

Violato, C., Marini, A., & Hunter, W. (1989). A confirmatory factor analysis of a four-factor model of attitudes toward computers: A study of preservice teachers. *Journal of Research on Computing in Education*, Winter, 199–213.

Yaghmaie, F., Jayasuriya, R., & Rawstorne, P. (1998). Computer experience and computer attitude: A model to predict the use of computerized information systems. *Human Computer Interaction*, 895–899.

KEY TERMS

Computer Anxiety: A diffuse, unpleasant, and vague sense of discomfort and apprehension when confronted by computer technology or people who talk about computers.

Computer Attitude: A complex mental state that affects a human's choice of action or behavior toward computers and computer-related tasks.

Computer Comfort: The user does not experience any suffering, anxiety, pain, etc., when using a computer.

Computer Confidence: The user is confident that he or she would be able to master a required skill to solve a particular problem using a computer, e.g., learning how to use a specific facility of an application program or learning a programming language.

Computer Liking: The use of a computer to solve problems is enjoyable, stimulating, and even addictive.

Self-Efficacy: Measure of a person's belief and confidence that he or she can perform a certain task.

Usefulness: The user is convinced that the use of computers in the workplace is an efficient and effective means to solve problems.

Computer-Mediated Learning Groups

Charles R. Graham
Brigham Young University, USA

Melanie Misanchuk
University of Guelph, Canada

INTRODUCTION

Online learning has seen tremendous growth over the past decade in both the corporate and higher education sectors of society. This has been facilitated by rapid increases in the availability of computer- and network-based technologies for communication and sharing of information. The U.S. National Center for Educational Statistics (2003) recently reported that for the 2000-01 academic year, 2- and 4-year institutions offered over 127,000 different distance education (DE) courses and had over three million enrollments. Of the institutions offering DE courses, 90% reported using the Internet and asynchronous communication as an instructional delivery mode (National Center for Educational Statistics, 2003). In the corporate sector, the American Society for Training & Development reported record levels technology-mediated training (or e-learning) accompanied by slight decreases in face-to-face classroom training (Thompson, Koon, Woodwell, & Beauvais, 2002). At the same time, there has been an increased awareness among distance educators and researchers regarding the importance of human interaction in the learning process. These two trends have driven the study of computer-mediated communication (CMC) and computer support for collaborative learning (CSCL). Groupwork has long been an important instructional strategy used in face-to-face learning environments and is now being explored in computer-mediated environments. This article will define critical aspects of computer-mediated groupwork and outline benefits and challenges to using computer-mediated groups as an instructional strategy. Additional details for the research presented in this article can be found in full-length publications by the authors (Graham, 2002a, 2002b, 2003; Graham & Misanchuk, 2003).

BACKGROUND

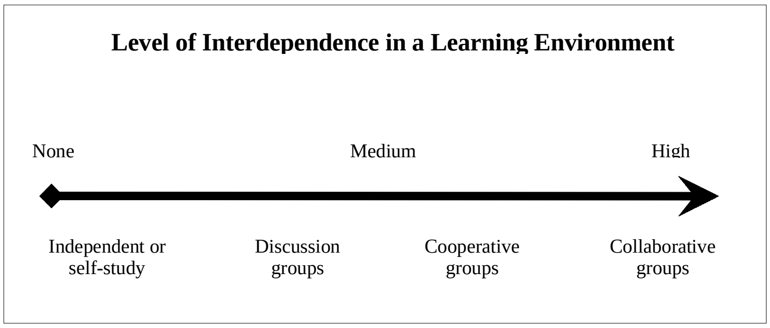
The first step to understanding the challenges and benefits of using computer-mediated learning groups is understanding the term itself. There are three important concepts embedded in the term *computer-mediated learning groups* that will be addressed in this section of the article. First is understanding what groups are and the role of interdependence in groupwork. Second is understanding the differences between *computer-mediated* groups and face-to-face groups. Lastly is the difference between *learning* groups and other types of groups such as *work* groups, an often overlooked distinction that has important implications for learning.

Interdependence in Groups

In a learning environment, there is a wide range of instructional strategies that emphasize interaction between individuals involved in the learning process. Interdependence can be defined *as the level of dependence between group members required to accomplish the desired learning outcome*, and is one of the most important characteristics defining the nature of the group interaction. Figure 1 depicts the spectrum of interdependence and some examples of different groups across that spectrum. At the low end of the spectrum, there is independent learning, which involves little or no interaction or dependence on others in the learning process. At the high end of the spectrum, instructional activities are highly dependent on collaborative interactions between group members.

Discussion groups appear at the lower end of the spectrum because participants are usually assessed based on their individual contributions and insights rather than the ability of the group to negotiate and come to a consensus or a common understanding of the discussion topic. Cooperative groups also have a lower level of interdependence than collaborative groups because they tend to divide the work into chunks that can be accomplished independently. There may be a high level of interaction at the administrative level in chunking and assigning tasks, but the majority of the work is done individually and then pieced together at the end, which also leads to a sense of individual accountability. Groups working collaboratively are more interdependent than cooperative groups. Rather than assigning individual tasks in a "divide and conquer" approach, they work through each of the tasks together as a group. Each

Figure 1. Different levels of interdependence in learning environments



participant's learning success depends on all other group members' efforts. Paulus (2004) asserts that students often default to a cooperative framework unless explicitly instructed to employ collaborative ways of working (Hathorn & Ingram, 2002; Kitchen & McDougall, 1999). This default behavior is mirrored in learning groups versus work groups, as explained in the following section.

Computer-Mediated vs. Face-to-Face Groups

The most common model of groupwork is in a face-to-face situation, whether it be in the workplace, classroom, committees, clubs, or other organizations. Indeed, many people have trouble envisioning successful interaction when group members are not co-located. However, the use of computer-mediated communication (CMC) tools to facilitate work has become more and more common in global business ventures, and is gaining acceptance in educational settings as well. Table 1 lists several of the most common tools used by distributed groups for communicating.

Groups may not necessarily need the most "enriched" mode of communication: often, a telephone conversation can accomplish as much as an elaborate videoconference.

Learning Groups vs. Work Groups

Much of the research literature related to computer-mediated groupwork looks at *work* groups as opposed to *learning* groups. There are important differences between the two that have an impact on how we understand

Table 1. Communication tools used by groups working at a distance

Synchronous	**Asynchronous**
• telephone (including multi-party calls) • instant messaging or other computer-mediated text (or voice) chat • two-way videoconferencing • electronic white-board tools	• email or listservs • voice mail messages • electronic bulletin boards/ discussion forums • commenting on documents using word processor tracking features • WebLogs (Blogs)

Table 2. Differences between typical work and learning groups

Work Group Characteristics	**Learning Group** Characteristics
• Hierarchical leadership structure • Clear role definitions • Cooperation maximizes productivity • Goals are product-oriented • Team members take on tasks that reflect skills and strengths already acquired • Focus is on the product or outcome	• Flat leadership structure • No role definitions • Collaboration maximizes learning • Goals are learning-oriented • Team members may accept tasks to gain skills they have not already acquired in order to learn • Focus is on the process or learning

this body of research. Table 2 outlines some common differences between *work* groups and *learning* groups.

For example, because the focus of a work group is typically product-oriented rather than learning-oriented, work groups often function more like cooperative groups (with the goal of maximizing productivity) than collaborative groups. Projects are often divided into tasks and assigned to the group members who already have the skills to accomplish the task. For example, a design group might have a project manager, an instructional designer, and a programmer, each with expertise and experience in their area. Learning teams, on the other hand, typically have a more process-oriented focus because they are interested in learning outcomes; indeed, in the ideal learning team, members who are inexperienced in a certain area will take on challenging tasks in order to improve their proficiency, which works directly against the productivity model. Work groups often have a hierarchical leadership structure with clear role definitions. In learning groups, leadership structures tend to be less hierarchical and formal because group members are peers with comparable levels of knowledge and expertise and roles may be arbitrarily assigned.

BENEFITS OF COMPUTER-MEDIATED LEARNING GROUPS

The use of learning groups has proven to be a powerful instructional tool in face-to-face contexts. There is an increasing awareness that learning groups can also have a positive impact on learning in a distributed, computer-mediated context. Table 3 summarizes three major benefits to using learning groups.

Challenges Associated with Computer-Mediated Learning Groups

There are three major areas that are important for successful implementation of groups in a computer-mediated environment (Graham, 2002a; Graham & Misanchuk, 2003):

1. Creating the groups
2. Structuring the learning activities
3. Facilitating group interactions

Table 4 identifies challenges associated with each of these major areas. This section will briefly address each of these challenges.

1. **Choosing Appropriate Group Size.** There is no ideal group size because appropriate group size is highly dependent on the learning context. Larger groups have a greater capacity for diversity of ideas and input at the increased cost of coordinating the group efforts. Prominent cooperative learning researchers have suggested that for learning effectiveness “the smaller [the group] the better” (Johnson, Johnson, & Holubec, 1994, p. 24). Size is a particularly important consideration in a computer-mediated environment where communication takes longer than in face-to-face environments (Bordia, 1997; Walther, 1996).
2. **Determining Group Composition.** Determining whether to create heterogeneous or homogeneous groups is an important consideration. Homogeneous groups tend to be more cohesive (Perrone & Sedlacek, 2000) while heterogeneous groups have the advantage of bringing different perspectives to the group discussion and work which can enrich

Table 3. Three major benefits for using computer-mediated learning groups

Benefits	Description
Learning Theory Perspective	Collaboration creates synergistic problem solving, making the group more than the sum of is parts (Brown et al., 1989) and requires students to present and defend their ideas, negotiating and constructing meaning.
Cooperative Learning Research	Johnson and Johnson (1996) found that group learning results in “significantly higher achievement and retention than does competitive and individualistic learning” (p.1022).
Learner Motivation	Cooperative experiences encourage more task orientation and personal social support than competitive or individualist environments and encourage students to monitor each others’ work (Johnson & Johnson, 1996).

Table 4. Six challenges to creating effective computer-mediated learning groups

Creating the Groups	Structuring Learning Activities	Facilitating Group Interactions
Choosing appropriate group size	Establishing an appropriate level of interdependence	Developing cooperative group skills
Determining group composition	Creating learner accountability	Establishing group norms

the learning experience (Cohen, 1994). Some differences that may be considered when forming groups are: demographic, content knowledge, technical skill level, location (time zones), and so forth. In a computer-mediated environment, spreading the individuals with high technical skills across groups can: (1) prevent the creation of some groups with all of the technical expertise and others with no technical expertise; and (2) allow skilled group members to mentor other group members in developing needed technology skills. Diversity of location (meaning different time zones) can present challenges for teams trying to coordinate work, especially if synchronous communication tools are used.

3. **Establishing an Appropriate Level of Interdependence.** One of the most difficult challenges with online learning groups is establishing an appropriate level of interdependence among group members. The tension that exists is between efficiency and learning. Higher levels of interdependence mean higher levels of collaboration (and thus communication) among learners. The learning task will dictate what portion of that communication will be directly related to learning and what portion will be required for task coordination and administration. The nature of the learning task will also determine whether interaction can be predominantly asynchronous or whether there needs to be a synchronous component. An appropriate goal would be to try to minimize time required for coordination efforts and maximize time spent directly engaged with the learning content.
4. **Creating Learner Accountability.** Closely linked to the issue of interdependence is the challenge of creating learner accountability in learning groups. Accountability should be addressed at both the group and individual level and determining the appropriate balance between the two is the challenge. Over-emphasizing group accountability can lead to problems such as social loafing or free-riding (Latane, Williams, & Harkins, 1979; Wagner, 1995) where certain individuals pull back from fully contributing to the group efforts. On the other hand, over-emphasizing individual accountability can undermine the cohesiveness of the group. The structure of assessment in the learning environment will have a big impact on learner accountability.
5. **Developing Cooperative Group Skills.** Students who have developed cooperative group skills are more likely to gain from their collaborative learning experience than others (Johnson & Johnson, 1987; Johnson & Johnson, 1999). Some important skills talked about in the literature are: decision making, consensus building, dealing with conflict, and basic communication skills (Graham, 2002a). One might expect adult learners to have already developed many of these skills or at least that the skills might transfer directly from prior experiences with face-to-face groups. However, communicating in an online environment without the body language and visual cues available requires different styles of communication and increased awareness of potential misunderstandings.
6. **Establishing Group Norms.** Norms are shared expectations that drive and inhibit the actions of group members. There are many types of norms including value norms, leadership and decision-making norms, communication norms, and logistical norms that are important for computer-mediated learning groups (Graham, 2002b). Establishing communication norms is particularly important for computer-mediated groups because without effective communication, the group cannot function properly. Online communication can be a challenge because of the many differences between computer-mediated and face-to-face communication, including lower fidelity interactions and slower communication speeds. Three areas that should be addressed in establishing explicit group norms are: (1) choosing modes of communication, (2) communicating the unseen, and (3) acknowledging communication (Graham & Misanchuk, 2003).

FUTURE TRENDS

Over the past decade there has been increased emphasis placed on the social nature of learning. The proliferation of tools that support human interaction and collaboration are on the rise. Tools that foster collaboration such as asynchronous discussion boards, instant messaging, personal blogs, wikis, groupware, and so forth, are becoming commonplace in both industry and educational settings. Groupwork is an important dimension of collaborative learning. Future research will continue to explore how to make learning in groups more effective. Additionally, research will continue to support the development of tools that facilitate group collaboration and work.

CONCLUSION

Groupwork is an important instructional strategy used in face-to-face learning environments. Groupwork is also being used with increased frequency in computer-mediated environments in both corporate and educational sectors of society. This article discussed three important characteristics of groups (interdependence, computer-mediated versus face-to-face interaction, and learning groups versus work groups). Also examined were both benefits and challenges of computer-mediated learning groups. This article identifies six challenges associated with forming and working in computer-mediated groups. Appropriate group size and composition must be considered in order to provide a synergistic learning environment. Learner interdependence and accountability must be balanced in order to maximize participants' investment in the learning process and responsibility for their own contributions. Learners must have experience—and oftentimes, instruction—in appropriate cooperative learning techniques for a computer-mediated environment. The establishment of group norms can help all group processes run more smoothly, especially in the situation of limited visual and non-verbal cues of online interaction.

REFERENCES

Bordia, P. (1997). Face-to-face versus computer-mediated-communication: A synthesis of the experimental literature. *The Journal of Business Communication, 34*(1), 99-120.

Cohen, E.G. (1994). *Designing groupwork: Strategies for the heterogeneous classroom* (2nd ed.). Teachers College, NY: Teachers College Press.

Graham, C.R. (2002a). Factors for effective learning groups in face-to-face and virtual environments. *Quarterly Review of Distance Education, 3*(3), 307-319.

Graham, C.R. (2002b). *Understanding and facilitating computer-mediated teamwork: A study of how norms develop in online learning teams.* Unpublished Dissertation, Indiana University, Bloomington, IN.

Graham, C.R. (2003). A model of norm development for computer-mediated teamwork. *Small Group Research, 34*(3), 322-352.

Graham, C.R., & Misanchuk, M. (2003). Computer-mediated teamwork: Benefits and challenges of using teamwork in online learning environments. In T.S. Roberts (Ed.), *Online collaborative learning: Theory and practice* (pp.181-202). Hershey, PA: Idea Group Publishing.

Hathorn, L.G., & Ingram, A.L. (2002). Cooperation and collaboration using computer-mediated communication. *Journal of Educational Computing Research, 26*(3), 325-347.

Herring, S.C. (2001). Computer-mediated discourse. In D. Schiffrin, D. Tannen, & H. Hamilton (Eds.), *The Handbook of Discourse Analysis* (pp.612-634). Oxford: Blackwell Publishers.

Johnson, D.W., & Johnson, F.P. (1987). *Joining together: Group theory and group skills* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Johnson, D.W., & Johnson, R.T. (1996). Cooperation and the use of technology. In D.H. Jonassen (Ed.), *Handbook of Research for Educational Communications and Technology* (pp.1017-1044). New York: Macmillan Library Reference.

Johnson, D.W., & Johnson, R.T. (1999). *Learning together and alone: Cooperative, competative, and individualistic learning* (5th ed.). Boston, MA: Allyn and Bacon.

Johnson, D.W., Johnson, R.T., & Holubec, E.J. (1994). *Cooperative learning in the classroom*. Alexandria, VA: Association for Supervision and curriculum development.

Kitchen, D., & McDougall, D. (1999). Collaborative learning on the internet. *Journal of Educational Technology Systems, 27*(3), 245-258.

Latane, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: Causes and consequences of social loafing. *Journal of Personality and Social Psychology, 37*, 822-832.

National Center for Educational Statistics. (2003). *Distance education at degree-granting postsecondary institutions: 2000-2001* (NCES 2003-17). Washington DC: U.S. Department of Education.

Paulus, T. (2004). Collaboration or cooperation? Small group interactions in a synchronous educational environment. In T.S. Roberts (Ed.), *Computer-supported collaborative learning in higher education* (pp. 100-124). Hershey, PA: Idea Group Publishing.

Perrone, K.M., & Sedlacek, W.E. (2000). A comparison of group cohesiveness and client satisfaction in homogeneous and heterogeneous groups. *Journal for Specialists in Group Work, 25*(3), 243-251.

Stahl, G. (2002, January). Introduction: Foundations for a CSCL community. Paper presented at the *Computer Support for Collaborative Learning 2002*, Boulder, CO.

Thompson, C., Koon, E., Woodwell, W.H.J., & Beauvais, J. (2002). *Training for the next economy: An ASTD state of the industry report on trends in employer-provided training in the United States*. American Society for Training and Development. Retrieved March 19, 2003, from the WWW: *http://www.astd.org/virtual_community/research/pdf/SOIR2002_Training_summary.pdf.pdf*

Wagner, J.A. (1995). Study of individualism-collectivism: Effects on cooperation in groups. *Academy of Management Journal, 38,* 152-172.

Walther, J.B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research, 23*(1), 3-43.

KEY TERMS

Blog: Short for Weblog. A form of Web-based communication that has increased in popularity in the last several years. Blogs facilitate easy publishing and sharing of ideas between individuals.

Computer-Mediated Communication (CMC): "The communication produced when human beings interact with one another by transmitting messages via networked computers" (Herring, 2001, p. 612).

Computer Supported Collaborative Learning (CSCL): "CSCL is a field of study centrally concerned with meaning and the practices of meaning-making in the context of joint activity, and the ways in which these practices are mediated through designed artifacts" (Stahl, 2002, p. 1).

Cooperation & Collaboration: Although the terms cooperation and collaboration are often used interchangeably, cooperative interaction implies more of a "divide and conquer" approach while collaborative interaction implies a more integrated process where shared meaning or understanding of an event or product is created.

Interdependence: The level of dependence that one group member has on other group members in order to complete the learning tasks.

Learning Group: A group of people, often with a fluid hierarchy and shifting roles, whose emphasis is learning before efficiency. The focus is on the process (the product being secondary), with the idea that the best learning outcomes will occur when participants take on challenges and unfamiliar tasks in order to gain competency in new areas.

Norms: Shared expectations that constrain and drive the action of group members.

Virtual Team: A team with members that are not co-located and so use technology to communicate and conduct their work. (This term is predominantly used in the corporate literature and means the same thing as computer-mediated team.)

Wiki: A Website to which many authors can contribute using a simple browser interface that allows them to modify existing pages.

Work Group: A group of people, often with a distinct hierarchy and role structure, whose emphasis is efficiency before learning. Focus is on the outcome and productivity, with the idea that the best product will be most efficiently created when everyone works using their present strengths.

Computing Curriculum Analysis and Development

Anthony Scime
State University of New York College at Brockport, USA

INTRODUCTION

Information technology (IT) is an umbrella term that encompasses disciplines dealing with the computer and its functions. These disciplines originated from interests in using the computer to solve problems, the theory of computation, and the development of the computer and its components.

Professionals from around the world with similar interests in IT came together and formed international professional organizations. The professional organizations span the disciplines of computer engineering (CE), computer science (CS), software engineering (SE), computer information systems (CIS), management information systems (MIS), and information technology (IT) (Freeman & Aspray, 1999). Note that information technology is both an umbrella term and a specific discipline under that umbrella.

These organizations exist to promote their profession and one method of promotion is through education. So, these professional organizations defined bodies of knowledge around the computer, which have been formalized and shaped as model curriculums. The organizations hope that colleges and universities will educate students in the IT disciplines to become knowledgeable professionals.

Because of the common interest in computing, there is a basic theory and a common technical core that exists among the model curricula (Denning, 1999; Tucker et al., 1991). Nevertheless each of the model curricula emphasizes a different perspective of IT. Each fills a different role in providing IT professionals. It falls upon the colleges and universities to select and modify the corresponding curriculum model to fit their needs.

BACKGROUND

Currently, there are a number of model curricula for computing (Table 1). A Joint Task Force on Computing Curricula created by the Association for Computing Machinery (ACM), and the IEEE Computer Society (IEEE-CS) developed Computing Curricula 2001 (CC 2001). This model focuses on programs in theoretical and applied computer science with various areas of emphasis in all areas of computing including computer engineering (CE), the engineering of computer hardware, and computer science (CS), the theory and design of hardware and software (Computing Curricula, 2001).

The field of information systems (IS) can be divided into the management of information systems (MIS), the engineering of computer information systems (CIS) and the use of existing commercial software applications to solve organizational problems or information technology (IT). The Information Resource Management Association (IRMA) and the Data Administration Managers Association (DAMA) have a curriculum model for MIS known as the Information Resource Management (IRM) model. It takes a management of data approach to information systems (Cohen, 2000). For a strong accounting and management MIS orientation, the Information Systems Auditing and Control Foundation has developed an interdisciplinary curriculum known as the Information Systems Auditing at the Undergraduate and Graduate Levels (ISA) model (ISACF, 1998).

IS 2002 (Information Systems, 2002) is a model curriculum developed through the joint efforts of the ACM, the Association for Information Systems (AIS), and the Association of Information Technology Professionals (AITP). This curriculum model focuses on information systems development as well as on management (Gorgone, Davis, Valacich, Topi, Feinstein, & Longenecker, 2002). The Information Systems Centric Curriculum (ISCC '99) model was developed by a task force that included members from academe and industry. It is oriented to large-scale system design and implementation. The focus is on the construction of the tools necessary for information management (Lidtke, Stokes, Haines, & Mulder, 1999).

The IT education special interest group of the ACM (SIGSITE) has developed a curriculum proposal. This curriculum is oriented toward the use of computing applications to solve organizational problems (IT Curriculum Proposal – Draft, 2002). An existing IT model is the Organizational and End User Information Systems (OEIS) model developed by the Office Systems Research Association. It is aimed at IT support of end-users. (Office Systems Research Association, 1996).

The Software Engineering Institute (SEI) has developed a model that follows the engineering approach of

Table 1. IT professional organizations and curriculum models

Professional Organization	Curriculum Models
Association for Computing Machinery (ACM)	Computing Curricula 2001 (CC 2001) Information Systems 2002 (IS 2002)
Association for Information Systems (AIS)	Information Systems 2002 (IS 2002)
Association of Information Technology Professionals (AITP)	Information Systems 2002 (IS 2002)
Computing Society of the Institute of Electrical and Electronic Engineers (IEEE-CS)	Computing Curricula 2001 (CC 2001)
Data Administration Managers Association (DAMA)	Information Resource Management (IRM)
Information Resources Management Association (IRMA)	Information Resource Management (IRM)
Information Systems Audit and Control Association (ISACA)	Information Systems Auditing at the Undergraduate and Graduate Levels (ISA)
International Federation for Information Processing (IFIP)	Informatics Curriculum Framework 2000 (ICF-2000)
Office Systems Research Association (OSRA)	Organizational and End User Information Systems (OEIS)
Software Engineering Institute (SEI)	Software Engineering Institute (SEI)
Special Interest Group in Information Technology Education (SITE) of the ACM	Information Technology Curriculum Proposal (IT)
An independent group of Academics and Professionals	Information Systems Centric Curriculum '99 (ISCC '99)

design first in the construction of software for embedded, large, and critical systems. The model strongly suggests specialization in a specific application domain (Bagert, Hilburn, Hislop, Lutz, McCracken, & Mangal, 1999).

Internationally, the International Federation for Information Processing (IFIP) in coordination with the United Nations Educational, Scientific and Cultural Organization (USESCO) has developed a framework within which schools can develop an IT curriculum (Mulder & van Weert, 2000). The Informatics Curriculum Framework 2000 (ICF-2000) considers the needs of developing countries for IT knowledgeable workers. These needs are balanced against the country's educational resources. The result is a tailored IT curriculum based on the established models.

CONSIDERATIONS IN DEVELOPING A CURRICULUM

To select or develop a curriculum, a school needs to assess their objectives and capabilities in providing graduates to the IT work force. A school with a strong liberal arts tradition has a different philosophy than a technically oriented school. A large university may have schools of computing and business; the focus of each may produce different information technology professionals. Some schools prepare students for further study while others are oriented to the job market. Schools with an international or national focus have different objectives than schools providing entry-level professionals locally.

The resources of a school may limit the curriculum as well. Time is a critical resource. Some model curricula require a student begin studying information technology courses in the first semester, others require only 4 or 5 semesters of IT and begin in the 3rd year of study. IT disciplines vary on the requirement for facilities. Courses requiring hands-on hardware to test theory or practice application require laboratories similar to those in an electrical engineering department. Some curricula stress practical application in commercial settings. This requires the school have close working relationships with local sponsors. The interests of the IT faculty also have an impact on curriculum development. Ideally, IT departments are well balanced in faculty expertise. However, it

Table 2. Technical core and role emphasis by discipline and curriculum model

Discipline	CIS	MIS			IT		SE	CE	CS
Technical Core — **Model**	ISCC '99	IS 2002	IRM	ISA	IT	OEIS	SEI	CC 2001	CC 2001
Role	CDM	CDS	C	DS	DS	CS	DM	CDM	CDM
Computer Literacy and Use of Software Tools		2R		1R 2E	1R 2E	1R 1E		1R	1R
Overview of IT and the Relationship to Business	2R	3R	2R 2E	1E	3E	2R 1E			
Computer Organization and Architecture		1R			3E			1R	2R 3E
Operating Systems and Systems Software								1R	1R
Programming, Algorithms and Data Structures	2R	1R	1R	1R	1R 2E		2R	3R	3R
Networking and Telecommunications	2R	1R	1E	1R 1E	1R 2E	1R		1R	
Systems/Software Analysis & Design	2R	2R	1R	2R 3E	3E	2R 3E	4R	3E	1R
Database and Information Retrieval	1R	1R	1R	1R 1E	1R 2E			1R	1R
Project Management		1R		2E	3E	1R	1R		
Intelligent Systems	1R		1R 1E	1E		1E		1R	1R
Social, Ethical & Professional Issues	1R					1E	1R	1R	
Internship/Capstone Project	1R					2E	1R	1R	1R

Key: C – Conceptualizer D – Developer M – Modifier S – Supporter xR – Number of Required Courses xE – Number of Elective Courses

is possible for a balanced department to have a greater interest in the development of IT artifacts versus the management of those artifacts. IT is a very large discipline, for one small department to be able to provide expertise in all facets of IT is unlikely.

Having considered the previously mentioned considerations, a school should also consider the role for which they are preparing graduates. Students require different knowledge dependent upon the role they will perform within IT. The curriculum studied helps determine the graduate's role. All information technology education consists of various technical, computer-oriented topics ranging from a theoretical understanding of computing, through the design and support of practical applications for complex computer systems. The depth of knowledge in these topics varies with the specific IT discipline and model curriculum (Scime, 2002b).

The fundamental knowledge in all the information technology disciplines involves the development, modification, support, conceptualization, and management of software and hardware artifacts (Freeman & Aspray, 1999; Information Technology Association of America, 1997). Developers work with programming and modeling languages. Developers need multiple courses in at least one of these areas. Modifiers need strength in programming or application tools. Supporters need to understand the end-users as well as the technology. Although conceptualizers typically have graduate degrees, at the undergraduate level this thinking begins with a strong theory component and by relating IT to the real world. Finally, IT managers are also conceptualizers by bringing together the IT professionals to satisfy an organization's need for information.

Not emphasizing the same common technical core of IT knowledge is what makes the IT disciplines differ from one another. Some areas of the core are emphasized in different model curricula, and provide an orientation of knowledge toward one or more of the professional roles. By considering the number of required and elective courses for an element of the technical core in a model the strengths of the model can be estimated. The relationship of the emphasis of technical topics, and the IT disciplines and professional roles supported by the curriculum models is of value in selecting a model as a beginning to program development (Table 2) (Scime, 2002b).

FUTURE TRENDS

The ACM and IEEE-CS are currently working on extending their model (CC 2001) to include other computing disciplines. They expect to create four volumes of curriculum models. These will include a model for computer engineering, software engineering, and information systems, as well as the current model for computer science. The software engineering work is currently in first draft. The computer engineering model is currently a preliminary draft. The information systems model is expected to closely match AIS/AITP/ACM's current IS 2002 model.

As computing becomes more and more ubiquitous, the use of the computer and its theoretical basis will continue to expand and infiltrate other disciplines. Expansion will manifest itself as the creation and clear definition of sub-disciplines, such as CE, SE and IS are today. Infiltration is the inclusion of IT into the sciences, arts, and humanities, for example, the use of graphical information systems in earth science and criminal justice.

CONCLUSION

The constantly changing world of computing and the constantly changing world of business leads to the enviable weakness of computing education (Lidtke, Stokes, Haines, & Mulder, 1999). Each school needs to assess their educational philosophy and student needs to create the curriculum best for them. By closely following a model, a school's prospective students, student's potential employers, and graduate schools know the type of education received by the graduates. The school administration is assured that the IT department is providing a curriculum, which covers all of the central topics of IT and emphasizes the chosen information technology discipline (Scime, 2001, 2002a, 2002b).

Although the disciplines differ in emphasis, all businesses that use information (and they all do) will need information technologist from each of the disciplines. All are necessary, but not all need to be provided from the same source.

REFERENCES

Bagert, D.J., Hilburn, T.B., Hislop, G., Lutz, M., McCracken, M., & Mangal, S. (1999). Guidelines for Software Engineering Education Version 1.0 (Technical Report CMU/SEI-99-TR-032). Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA.

Cohen, E. (Ed.) (2000). IRMA/DAMA curriculum model, IRMA, Hershey. Retrieved on November 15, 1999 from *http://gise.org/IRMA-DAMA-2000.pdf*

Computing Curricula 2001 (CC2001). (2000). The joint task force on computing curricula. *IEEE Computer Society and Association of Computing Machinery*, March 2000.

Denning, P.J. (1999, March). Our seed corn is growing in the commons. *Information Impacts Magazine*. Retrieved on September 19, 2000 from *http://www.cisp.org/imp/march_99/denning/03_99denning.htm*

Freeman, P., & Aspray, W. (1999). *The supply of information technology workers in the United States*. Washington, DC: Computing Research Association.

Gorgone, J.T., Davis, G.B., Valacich, J.S., Topi, H., Feinstein, D.L., & Longenecker, H.E., Jr. (2002). Model curriculum and guidelines for undergraduate degree *Programs in Information Systems Association for Information Systems*.

Information Technology Association of America (ITAA) (1997). Help wanted: The workforce gap at the dawn of a new century, Arlington, VA, (p.9).

ISACF (1998). ISACF Task Force for Development of Model Curricula in Information Systems Auditing at the Undergraduate and Graduate Levels, Academic Relations Committee and Research Board (1998). Model curricula for information systems auditing at the undergraduate and graduate levels. Information Systems Audit and Control Foundation.

IT Curriculum Proposal – Draft (2002). SITE Curriculum Committee. *Proceedings of the 2002 Conference for In-*

formation Technology Curriculum, Rochester, NY, September.

Lidtke, D.K., Stokes, G.E., Haines, J., & Mulder, M.C. (1999). ISCC'99, An information systems-centric curriculum '99 program guidelines for educating the next generation of information systems specialists. In *Collaboration with industry*.

Mulder, F., and van Weert, T. (2000). ICF-2000 Informatics Curriculum Framework 2000 for Higher Education. Paris: UNESCO. Retrieved on February 7, 2004 from http://poe.netlab.csc.villanova.edu/ifip32/icf2000.htm

Office Systems Research Association (1996). Organizational & end-user information system curriculum model, OSRA. Retrieved on December 3, 2000, from *http://pages.nyu.edu/~bno1/osra/model_curriculum/*

Scime, A. (2001). Information systems draft accreditation criteria and model curricula. *Proceedings of the 18th Annual Information Systems Conference (ISECON 2001)*, Cincinnati, OH, November. Retrieved on November 10, 2003, from *http://colton.byuh.edu/don/isecon/*

Scime, A. (2002a). Information systems and computer science model curricula: A comparative look. Chapter 18 in A. Saber, S. Saber, & M. Dadashzadeh (Eds.), *Information technology education in the new millennium* (pp.146-158). Hershey, PA: IRM Press.

Scime, A. (2002b). Information technology model curricula analysis. Chapter 12 in E. Cohen (Ed.), *Challenges of information technology education in the 21st century* (pp.222-239). Hershey, PA: Idea Group Publishing.

Tucker, A.B., Barnes, B.H., Aieken, R.M., Barker, K., Bruce, K.B., Cain, J.T., Conry, S.E., Engel, G.L., Epstein, R.G., Lidtke, D.K., Mulder, M.C., Rogers, J.B., Spafford, E.H., & Turner, A.J. (1991). Computing Curricula 1991: Report of the ACM/IEEE-CS Joint Curriculum Task Force, Association of Computing Machinery.

KEY TERMS

Computer Engineering (CE): The engineering of computer hardware.

Computer Information Systems (CIS): Concerns information systems with an emphasis on information as an enterprise resource, and its design, development, implementation, and maintenance of information systems. Sometimes referred to as information systems.

Computer Science (CS): Hardware and software theory and design.

Computing Curricula 2001 (CC 2001): Developed by the Joint Task Force on Computing Curricula created by the ACM and the IEEE-CS. This model focuses on programs in theoretical and applied computer science (Computing Curricula, 2001).

Information Resource Management (IRM): Curriculum model 2000 of the IRMA and the DAMA focuses particularly on the disciples of information resource management and management information systems. It takes a management of data approach (Cohen, 2000).

Information Systems (IS): Use data to create information and knowledge to assist in operational, management, and strategic organizational decision-making. It is also an umbrella term for computer information systems, management information systems and information technology.

Information Systems 2002 (IS 2002): Model curriculum developed by the efforts of the ACM, AIS, and AITP. This curriculum model focuses on information systems development and management (Gorgone, Davis, Valacich, Topi, Feinstein, & Longenecker, 2002).

Information Systems Auditing (ISA) at the Undergraduate and Graduate Levels: Model developed by the ISACF Task Force for Development of Model Curricula. It is an interdisciplinary approach with a strong accounting and management orientation. (ISACF, 1998).

Information Systems Centric Curriculum (ISCC '99): Model developed by a task force that included members from academe and industry. It is oriented toward large-scale system design and implementation as opposed to automata theory and programming. The focus is on the construction of the tools necessary for information management (Lidtke, Stokes, Haines, & Mulder, 1999).

Information Technology (IT): Uses existing commercial software applications to solve organizational problems. Sometimes refer to as information systems. It is also the umbrella term for all the disciplines involved with the computer.

IT Curriculum Proposal: Being developed by SIGSITE, the curriculum is oriented toward the use of computing applications to solve organizational problems (IT Curriculum Proposal – Draft, 2002).

Management Information Systems (MIS): The management of information systems and data including management of the design, development, implementation, and maintenance of information systems. Sometimes referred to as information systems.

Organizational and End User Information Systems (OEIS): Model developed by the OSRA is aimed at IT support of end-users (Office Systems Research Association, 1996).

Software Engineering (SE): The engineering of software for embedded, large, and critical systems.

Software Engineering Institute (SEI) Model: Model developed at SEI that follows an engineering approach of design first. The model suggests specialization in a specific domain (Bagert, Hilburn, Hislop, Lutz, McCracken, & Mangal, 1999).

Concepts and Dynamics of the Application Service Provider Industry

Dohoon Kim
Kyung Hee University, Korea

INTRODUCTION: SOFTWARE AS A SERVICE

The enterprise intelligence through e-transformation is one of the cornerstones of the next-generation e-business era where the Internet constitutes the core business resource. Furthermore, the severe competitive landscape of e-business makes firms focus on their core capability and farm out staffing functions such as IT. Under this circumstance, enhancing intelligence and synergy through e-transformation will be accomplished by IT outsourcing via ASPs (application service providers). The ASP industry now provides an essential infrastructure for the Internet-based e-business transactions, thereby accelerating corporate e-transformation.

An ASP is generally defined as a third-party service firm that deploys, manages, and/or remotely hosts a software application through centrally located servers in a lease agreement. ASPs started their business by providing online application programs such as ERP (enterprise resource planning) and CRM (customer relationship management) solution packages to corporate customers. The first customers were small companies or local branches of multinational companies where IT outsourcing was the only option to deploy IT resources due to financial or regional constraints. As seen in these cases, the biggest merit of employing ASPs is that corporate customers do not have to own the applications and take responsibilities associated with initial and ongoing support and maintenance. Consequently, ASPs are differentiated from the existing IT services in that ASPs provide IT resources to multiple corporate clients on a one-to-many basis with a standardized service architecture and pricing scheme.

BACKGROUND: INDUSTRY VALUE CHAIN

The industry value chain does not allow a single service provider to control the entire service delivery process. Even if we confine our attention to the software delivery process in the value chain, the complexity does not reduce significantly. In order to deliver applications over the Internet, we need a mechanism to establish and maintain collaboration among independent functional divisions. Analysis of this nature of the value chain shows how the industry is likely to evolve and gives some insights into the strategic meaning of special types of convergence. In particular, we should point out two critical aspects of the value chain, which are required to survive in the market: a large customer base and stable relationship with other functional divisions. The structure of partnership among the players in the value chain is one of the major elements to classify emerging ASP business models. Table 1 summarizes key players in the ASP value chain.

There are a number of factors that are frequently cited as fueling or dashing the growth of the ASP market (Burris, 2001; Factor, 2002; Kim, 2002; Sparrow, 2003; Toigo, 2001). One of the striking characteristics observed so far is that immaturity of the industry is the most representative challenge in terms of the market factor: for example, the uncertainty as to whether existing and emerging ASPs are winning enough customers to validate an ASP business model for highly sophisticated enterprise applications. While some ASPs are gaining momentum with early adopters, there are many client companies that are unwilling to rent ERP applications due to the lack of trust in the industry itself in Korea (Kim & Choi, 2001).

Table 1. Key players in the ASP value chain model

- Software Vendors: including ISVs (independent software vendors), content providers (CPs), and so forth
- Network Infrastructure Providers: including telecommunication operators, ISPs (Internet service providers), and so forth
- Application Service Providers: as an intermediary or an organizer between software vendors and customers
- Individual and Corporate Customers: subscribers (end users) of the ASP services

Table 2. Drivers and challenges of the ASP industry

Category	Drivers	Challenges
Technology	◆ Reduce risk of technological obsolescence due to rapidly changing IT ◆ Provide a chance to utilize best-of-breed applications ◆ Avoid IT staffing shortage	◆ Unsolved security concerns ◆ Emerging, new technological requirements from the clients: e.g., SLA with client participation ◆ Unproved service reliability: e.g., network problems, system scalability and performance
Market	◆ Minimize up-front TCO (total cost ownership) ◆ Provide predictable cash flows	◆ Unproved client momentum ◆ Failure in giving clients sufficient trust due to unstable ASP industry

Moreover, it is security control and remote monitoring systems, SLA (service level agreement; Lee & Ben-Natan, 2002; Sturm, Morris, & Jander, 2000) management, and the global standardization process that should be further developed to support proliferation of ASPs. In the end will survive only a few successful ASPs that adapt themselves to the market requirements and take the most advantage of the competitive landscape.

ASP BUSINESS MODELS

The industry's short history raises the following questions. What changes will happen? Who will be the winners and losers? To answer these questions, Table 3 clarifies different types of the ASP business domains that are currently emerging. ASP's common value proposition to improve total benefits from IT outsourcing has been giving rise to various trials in designing the service

Table 3. ASP business models and capability profiles

Basic Types	Characteristics and Value-Added Components	Basic Capability
H-ASP (Horizontally Specialized ASP)	◆ Develop deep expertise within a given functional area (as opposed to one-stop shop): Substantial consulting services are possible ◆ ISV's need of partnership with systems integration and distribution companies ◆ Should be Web-based software provider ◆ Either own the software or develop proprietary integration in a specific field	◆ Well positioned to expand customer basis quickly ◆ Hard to copy the domain-specific knowledge
V-ASP (Vertically Specialized ASP)	◆ Industry-specific applications (in contrast to one-stop shop) ◆ Vertically oriented template methodology: easily deploy across multiple clients in the same industry	◆ Strong advantage in customized solutions ◆ Hard to copy the industry-specific knowledge
AIP (Application Infrastructure Provider)	◆ Originated from telecommunication company that owns networks and has operations experience ◆ Provide infrastructure management to ASPs ◆ Provide system management services including SLA ◆ Alleviate client concerns regarding network reliability, etc.	◆ High investment costs as an entry barrier: easy to protect their market share
XSP (Extended Service Provider)	◆ Provide total services from front end to back end with systems integration consulting ◆ Create new business process by rearranging suppliers and customers ◆ Help customers and even other service providers enter new markets, deploy services, and improve profitability easily while minimizing risk ◆ Build and integrate customized applications, thereby enabling clients to avoid the need to handle multiple ASP solutions	◆ Going back to one-stop-shop idea: Improved flexibility will be the core competitive edge for XSP

delivery processes, each of which corresponds to a business model suggested in the table. Therefore, this classification plays a key role in identifying and analyzing the collaborative networking structure in the ASP value chains.

FUTURE TRENDS: INDUSTRY DYNAMICS AND EVOLUTION

The guiding principles of the industry evolution, which have been leading the industry to face proliferation of ASP business models, are summarized into (a) economies of scale through positive feedback from the market and (b) integration across the value chain for attaining cost reduction and differentiation.

First, it is the economies of scale or increasing return that serves as the core economic guiding principle for ASPs. A survey on the Korean IT outsourcing market reveals that, in terms of the TCO (total cost ownership) of a typical ERP package, IT outsourcing through ASPs enables clients to save roughly 20% of their up-front license fee and 80% of the implementation and maintenance service costs (Kim & Choi, 2001). Accordingly, ASPs that host these applications basically seek to lower this 80% portion of the TCO upon the notion of a one-to-many relationship between an ASP and its clients. An ASP is usually able to leverage standardized solutions across multiple clients. Attaining client momentum and reducing the overall costs per client are the major economic motivations for ASPs to compete with each other, thereby creating a positive feedback mechanism through network externality on the demand side. In sum, the competition keeps going for the expansion of a customer base or market share, which provides a good surrogate measure of profit for this case.

Second, the competitive landscape is also defined by the unique nature of a service system market where independent hardware and software resources are combined and reorganized into a new package in alignment with partners along the value chain and even a customer's business process. These offerings aim at designing a seamless and proprietary service delivery process to sharpen the competitive edge while raising the entry barrier. This essential feature of integration in the service delivery process makes the various possible business models reduce into the different types of service product combinations along the value chain as presented in Table 1. The integration, however, should be verified by achieving savings in TCO, though it is not easy to measure the amount of cost reduction by a certain partnership structure. Accordingly, the cutthroat competition fueled by business domain integration not only drives down the price to an acceptable market price, but also creates diverse market segmentations based on the service product differentiation.

Furthermore, increasing switching costs and rising entry barriers, two basic phenomena regarding the guiding principles and competitive landscape, are common to all the business models. As a result, efforts to penetrate into different market segments and build new customer relationships at the niche will inevitably run into strong resistance from the incumbents. Some events like technological breakthroughs will be required in order for a specific ASP model to consolidate another. Therefore, various business models will thrive over a period of time before some giant players in each business model emerge. Despite a unique coexistence of diverse ASP business models, some hypotheses on the industry structure change and evolution can be derived from those observations together with the guiding principles.

First, the general trend will be that the total number of ASPs in the industry will reduce since the customer base is not large enough to keep all the incumbent ASPs alive. Cash flows generated from the market give winners resilience to possible occasional failures and allow them to better manage risk by diversifying a portfolio of value components to open a new market niche. It is this kind of positive feedback loop (Arthur, 1989; Nelson & Winter, 1978) from the economies of scale that accelerates the exit of losers from the market and shapes the industry structure (Shy, 2002).

Second, the industry has been concentrating more around horizontally and vertically specialized ASPs than around the pure ASPs (that is, a simple partnership with an ISV). The primary concern of the emerging ASPs is to build some value-added components to the service architecture, thereby making it hard for competitors to replicate their business model and for customers to replace the current provider. However, reliance on third-party ISVs could make it more difficult to resolve underlying performance issues that have been the subject of customer scrutiny. On the other hand, looking up the capability profiles of the ASP business models, we can conclude that both H-ASPs and V-ASPs hold a dominant position from this standpoint. If some technical constraints such as SLA and security requirements come to rise to the surface, AIP will gain technological competitiveness since delegating the control of core enterprise applications to an external provider requires ASPs to prove their capability of reliable and stable operations.

Last, we predict that the rate-of-demand increase will affect the industry structure: the pattern of market segmentation, the market share of each ASP type, and so forth. The speed of the market expansion will affect ASPs' selection of competitive priorities.

CONCLUSION

The ASP industry will shape the future e-business transactions, providing a great flexibility in redeploying a firm's resources. Although the industry is currently at its early stage of the industry life cycle, much attention is now paid to vertical or domain-specific expertise and flexible capabilities in addition to the basic offerings. In order to assess both the market and the supply side, classified are emerging ASP business models together with some driving forces shaping the evolutionary path. Some careful observations disclosed that (a) the capability of an ASP model hinges on the differentiation of service products to a large degree and (b) economies of scale play a key role in the dynamically evolving market mechanisms. ASPs that originally developed their proprietary solutions will be better positioned in terms of ultimate performance and scalability. Those ASPs will increase the chance to succeed in the market irrespective of how critical a given solution is to their client's day-to-day operations. Last, some technical factors that may affect the evolution path (for example, SLA regulation and security) should be considered in the near future.

REFERENCES

Arthur, W. B. (1989). Competing technologies, increasing returns, and lock-in by historical events. *Economic Journal*, (99), 116-131.

Burris, A. M. (2001). *Service provider strategy: Proven secrets for xSPs*. Upper Saddle River, NJ: Prentice Hall.

Church, J., & Gandal, N. (2000, Spring). Systems competition, vertical merger and foreclosure. *Journal of Economics and Management Strategy*, (Vol. 9, pp. 25-51).

Factor, A. (2002). *Analyzing application service providers*. Upper Saddle River, NJ: Sun Microsystems Press.

Farrell, J., Monroe, H., & Saloner, G. (1998, Summer). The vertical organization of industry: Systems competition versus component competition. *Journal of Economics and Management Strategy*, (Vol. 7, pp. 143-182).

Harney, J. (2002). *Application service providers (ASPs): A manager's guide*. Boston, MA: Addison-Wesley.

Katz, M. L., & Shapiro, C. (1994, Spring). Systems competition and network effects. *Journal of Economic Perspectives*, (Vol. 8, pp. 93-115).

Kim, D. (2002). ASP and collaborative network infrastructure for global enterprise intelligence: An explanatory approach to identify prerequisites and challenges. In J. Chen (Ed.), *Global supply chain management* (pp. 166-170). Beijing, China: International Academic Publication.

Kim, M. S., & Choi, Y. C. (2001). *The current status of ASP market development*. Korea Information Society Development Institute, Seoul, Korea.

Lee, J. J., & Ben-Natan, R. (2002). *Integrating service level agreements: Optimizing your OSS for SLA delivery*. Indianapolis, Indiana: Wiley.

Nelson, R. R., & Winter, S. G. (1978). Force generating and limiting concentration under Schumpeterian competition. *Bell Journal of Economics*, (9), 524-548.

Shy, O. (2002). *The economics of network industries*. New York: Cambridge University Press.

Sparrow, E. (2003). *Successful IT outsourcing: From choosing a provider to managing the project*. London: Springer.

Sturm, R., Morris, W., & Jander, M. (2000). *Foundations of service level management*. Indianapolis, Indiana: Sams.

Toigo, J. W. (2001). *The essential guide to application service providers*. Upper Saddle River, NJ: Prentice Hall.

Zuscovitch, E., & Justman, M. (1995). Networks, sustainable differentiation, and economic development. In D. Batten, J. Casti, & R. Thord (Eds.), *Networks in action* (pp. 269-285). New York: Springer-Verlag.

KEY TERMS

AIP: An application infrastructure provider (AIP) is a type of ASP, which is usually originated from telecommunication operators that run their own networks and IDCs (Internet data centers). AIP focuses on server hosting and network infrastructure management for other ASPs and corporate clients, and provides value-added services based on its technology leadership, for example, online security and e-payment services.

ASP: An application service provider (ASP) is a third-party service firm that deploys, manages, and/or remotely hosts a software application through centrally located servers in a lease agreement.

Economies of Scale: Economies of scale are the achievement of lower average cost per unit through increased production, or the decrease in the marginal cost of production as a firm's extent of operations expands.

Horizontal ASP: Horizontal ASPs provide online applications for a specific business function such as human resource management, procurement, customer relations, and so forth.

IT Outsourcing: IT outsourcing is the outsourcing of enterprise information systems and management to computer manufacturers or software companies (the term outsourcing stems from using an outside resource). Companies can save purchasing cost, maintenance cost, and labor cost by outsourcing and paying for those services. Outsourcing has become a common practice in the US where companies are faced with uncertain returns of massive investment in IT resources.

SLA: A service level agreement (SLA) is a contract between a supplier and a customer that identifies (a) services supported at each of three layers—application, host (system), and network—(b) service parameters for each service, (c) levels of service quality, and (d) liabilities on the part of the supplier and the customer when service quality levels are not met.

Vertical ASP: Vertical ASPs provide online applications customized for a specific industry such as staff or operations scheduling for hospitals, material procurement for the steel industry, and so forth.

XSP: An extended service provider (XSP) provides total IT-related services from online applications through maintenance and reengineering of IT resources to business process consulting for its clients. Success of the XSP model should presume the rapid proliferation of the ASP services in an overly complex market. If the ASP service demand grows explosively in a short period of time, the XSP model will debut in the market earlier on, increasing the possibility of XSPs dominating the industry as they have scale advantage in terms of cost.

Value Chain: A value chain is a chain of activities in a group of collaborators who are designed to meet market demand. They are vendors involved in value chains across purchasing, procurement, manufacturing, warehousing, distribution, and sales of components, equipments, raw materials, and so forth to manage a series of resource and information flow.

C

Concepts of Emergence Index in Image Databases

Sagarmay Deb
University of Southern Queensland, Australia

INTRODUCTION

Images are generated everywhere from various sources. It could be satellite pictures, biomedical, scientific, entertainment, sports and many more, generated through video camera, ordinary camera, x-ray machine and so on. These images are stored in image databases. Content-based image retrieval (CBIR) technique is being applied to access these vast volumes of images from databases efficiently. Some of the areas where CBIR is applied are weather forecasting, scientific database management, art galleries, law enforcement and fashion design.

Initially image representation was based on various attributes of the image like height, length, angle and was accessed using those attributes extracted manually and managed within the framework of conventional database management systems. Queries are specified using these attributes. This entails a high level of image abstraction (Chang, 1988; Gudivada & Raghavan, 1995). Also there was feature-based object-recognition approach where the process was automated to extract images based on color, shape, texture and spatial relations among various objects of the image.

Recently, combinations of these two approaches, efficient image representation and query-processing algorithms, have been developed to access image databases. Recent CBIR research tries to combine both of these approaches and has given rise to efficient image representations and data models, query-processing algorithms, intelligent query interfaces and domain-independent system architecture.

As we mentioned, image retrieval can be based on low-level visual features such as color (Pass et al., 1996; Smith & Chang, 1996; Srisuk & Kurutach, 2002; Sural et al., 2002; Swain & Ballard, 1991; Traina et al., 2003), texture (Manjunath, 1996; Sheikholeslami, 1997; Smith & Chang, 1994; Zhou et al., 2001), shape (Safar, 2000; Shahabi & Safar, 1999; Tao & Grosky, 1999), high-level semantics (Forsyth et al., 1996; Torratha & Oliva, 1999) or both (Zhao & Grosky, 2001).

But most of the works done so far are based on the analysis of explicit meanings of images. But image has implicit meanings as well, which give more and different meanings than only explicit analysis provides. In this article we provide the concepts of emergence index and analysis of the implicit meanings of the image, which we believe should be taken into account in analysis of images of image or multimedia databases.

BACKGROUND

Concepts of Emergence

A feature of an image that is not explicit would be an emergent feature if it could be made explicit. There are three types of emergence: computational emergence, thermodynamic emergence and emergence relative to a model (Cariani, 1992). We use the latter one in our article.

Whenever we shift our focus on an existing shape, in other words an image, a new shape emerges. The representation of the new shape is based upon our view of the original shape. The new shape emerges as we change our view of the original shape. This is the most important idea of emergence.

Two classes of shape emergence have been identified: embedded shape emergence and illusory shape emergence (Gero, n.d.; Gero & Maher, 1994). These procedures could be based on geometrical, topological or dimensional studies of the original shape.

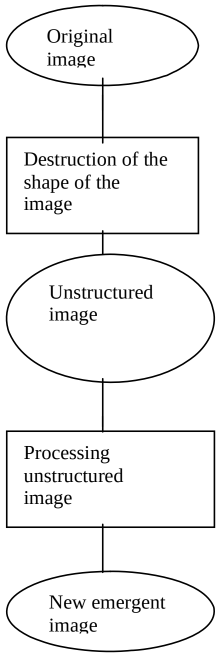
Model of Emergence

To extract emergent shape from an image, first we have to destroy the original shape of the image. This would give us an unstructured image. Now we take the unstructured image and find out the extra or implicit meaning out of it in addition to the original meaning, and this process gives rise to emergent image with implicit meaning, making explicit and emergent images generated. This can be defined in a model in Figure 1 (Gero & Yan, 1994):

Definition of Emergence Index

For image retrieval where the hidden or emergence meanings of the images are studied and based on those hidden meanings as well as explicit meanings, where there is no hidden meaning at all, an index of search is defined to retrieve images is called emergence index.

Figure 1. Model of emergence



When images are retrieved based on textual information, then various parameters and descriptions might define the input and the images of the database. Whenever there would be symmetry of parameters and descriptions, the image could be retrieved. As mentioned earlier, in CBIR, color, texture and shape are widely used as index to retrieve images. But in our studies, we can find the hidden meanings of the images and whenever those hidden meanings match with the input given, although the original image may not match at all with the input, we can retrieve that image.

When an input would come in the form of an image, the image could be studied based on features, constraints, variables and domains and converted into parametric form. Then the image database would be accessed and each image would be interpreted considering the items mentioned previously and also emergence and converted into parametric form like the input image. Whenever there would be a match between parameters of the input and the images of the database, these records would be selected. In other words, indexing would be decided by the outcome of emergence, which means more meaningful images could be found hidden in an image which would otherwise not be understood.

Many images of the database may not have any apparent similarities with the input, but emergence could bring out the hidden meaning of the image and could establish similarities with the input image. So emergence outcomes of the images would form the index structure of the search.

Analyses of Works Done

Attempts have been made to give rise to symbolic representation of shape where shape is defined as

$$S = \{N; \text{Constraints}\}$$

where N is the cardinality, that is, the number of infinite maximal lines constituting shape S, and the constraints limit the behaviors or properties resulting from the infinite maximal lines, based upon which particular shape is defined. Lines have been defined as Ik, Ij and so on with their intersection as Ikj. Then topological, geometric and dimensional properties are defined (Gero, 1992). Also symmetry has been found through the corresponding relevant positions of the lines and coordinates of one shape with that of the other, and in the process, emergence of the shapes are studied (Jun, 1994).

There is no direct approach to solve the problem of emergent index other than the ones mentioned previously. There is only an indirect approach where this conception has been applied. In a model named Copycat involving computer programs, the program makes all possible sets of consistent combinations of pairings once all plausible pairings have been made. In other words, it gives rise to something explicit that was implicit earlier, which is the essential feature of emergence phenomenon (Mitchell & Hofstadter, 1994).

MAIN THRUST OF THE ARTICLE

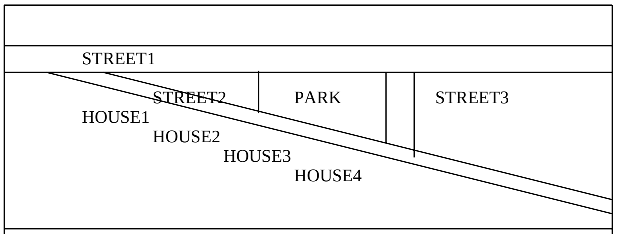
We attempt to study the problem of image query where a query made would be searched through the database to select those records where a similar shape has been found. But in addition to that we pick up records based on the emergence phenomena where the query input may not have an apparent match in a particular image of the database, but emergence phenomena could give rise to a similar structure in the same image and as such this image should be selected as a query result. For example, a square with single diagonal can be observed as two triangles. So whenever search intends to find a triangle, this image, which apparently is much different than triangle, would be selected because of emergence.

We calculate emergence index of images of image databases based on features, constraints, variables, domains and emergence.

Various mathematical tools that could be used in defining the image are:

- Geometric property
- Topological property

Figure 2. Geographic location



- Dimensional property
- Statistical properties

Structure of Emergence Index

Emergence indexes can be defined out of five factors

EI = f(D,F,V,C,E)

where EI stands for emergence index, D for domain where the image belongs, F for features, V for variables that can define the feature's constraints under which the features are defined, C for constraints and E for emergence characteristics of images.

We believe any image, static or in motion, could be expressed semantically in terms of the previously mentioned five parameters (Deb & Zhang, 2001).

Application of Emergence Index in Geographic Location

If we have a map of a geographic location like the one in Figure 2, then we find there are three streets, namely, STREET1, STREET2 and STREET3. There is a park between STREET1 and STREET2 and HOUSE1, HOUSE2, HOUSE3, HOUSE4 are four houses on STREET2.

We also notice that STREET1, STREET2 and STREET3 form a triangle surrounding PARK. In normal map interpretation this may not surface. But when hidden shape is searched we get a triangle. This is the emergence outcome of the search. This would help us to locate the places more accurately by referring to the triangle in the map. Also if there is an input in the form of a triangle, then this image, although a map, would be selected because of emergence.

CONCLUSION

Emergence is a phenomenon where we study the implicit or hidden meaning of an image. We introduced this concept in image database access and retrieval of images using this as an index for retrieval. This would give an entirely different search outcome than ordinary search where emergence is not considered, as consideration of hidden meanings could change the index of search. We discussed emergence, emergence index and approach as to how to apply this concept in image retrieval in this article.

REFERENCES

Cariani, P. (1992). Emergence and artificial life. In C. Langton, C. Taylor, J.D. Farmer & S. Rasmussen (Eds.), *Artificial life II* (pp. 775-797). Reading: Addison-Wesley.

Chang, S.K, Yan, C.W., Dimitroff, D.C., & Arndt, T. (1988). An intelligent image database system. *IEEE Transactions on Software Engineering, 14*(5).

Deb, S., & Zhang, Y. (2001). Emergence index structure in image retrieval. *Tamkang Journal of Science and Engineering, 4*(1), 59-69.

Forsyth, D. et al. (1996). Finding pictures of objects in large collections of images. *Report of the NSF/ARPA Workshop on 3D Object Representation for Computer Vision, 335.*

Gero, J.S. (n.d.). *Visual emergence in design collaboration.* Key Center of Design Computing, University of Sydney.

Gero, J.S. (1992). *Shape emergence and symbolic reasoning using maximal lines.* Unpublished notes. Design Computing Unit, Department of Architectural and Design Science, University of Sydney, Sydney.

Gero, J.S., & Maher, M.L. (1994, September). Computational support for emergence in design. *Information Technology in Design Conference,* Moscow.

Gero, J.S., & Yan, M. (1994). Shape emergence by symbolic reasoning. *Environment and Planning B: Planning and Design, 21,*191-212.

Gudivada, V.N., & Raghavan, V.V. (1995, September). Content-based image retrieval systems. *IEEE.*

Jun, H.J. (1994). *Emergence of shape semantics in CAD system.* PhD thesis proposal. Design Computing Unit, Department of Architectural and Design Science, University of Sydney, Sydney.

Manjunath, B.S., & Ma, W.Y. (1996, August). Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 18*(8).

Mitchell, M., & Hofstadter, D. (1994). The Copycat project: A model of mental fluidity and analogy-making. *Fluid* *concepts & analogies: Computer models of the fundamental mechanisms of thought.* New York: BasicBooks.

Pass, G., Zabih, R., & Miller, J. (1996). *Comparing images using color coherence vector*. Boston: ACM Multimedia.

Safar, M., Shahabi, C., & Sun, X. (2000). Image retrieval by shape: A comparative study. *Proceedings of IEEE International Conference on Multimedia and Exposition* (ICME), USA.

Shahabi, C., & Safar, M. (1999). Efficient retrieval and spatial querying of 2D objects. *IEEE International Conference on Multimedia Computing and Systems* (ICMCS99), Florence, Italy (pp. 611-617).

Sheikholeslami, G., & Zhang, A. (1997). An approach to clustering large visual databases using wavelet transform. *Proceedings of the IEEE International Conference on Image Processing* (pp. 407-411).

Smith, J.R., & Chang, S. (1994). Transform features for texture classification and discrimination in large image databases. *Proceedings of the IEEE International Conference in Image Processing* (pp. 407-411).

Smith, J.R., & Chang, S.F. (1996). VisualSeek: A fully automated content-based image query system. *Proceedings of ACM Multimedia 96*, Boston, MA (pp. 87-98).

Srisuk, S., & Kurutach, W. (2002, March 8-13). An efficient algorithm for face detection in color images. *Proceedings of 6th Joint Conference on Information Sciences*, Research Triangle Park, NC (pp. 688-691).

Sural, S., Qian, G., & Pramanik, S. (2002, March 8-13). A histogram with perceptually smooth color transition for image retrieval. *Proceedings of 6th Joint Conference on Information Sciences,* Research Triangle Park, NC (pp. 664-667).

Swain, M., & Ballard, D. (1991). Color indexing. *International Journal of Computer Vision*, *7*(1), 11-32.

Tao, Y., & Grosky, W. (1999). Delaunay triangulation for image object indexing: A novel method for shape representation. *Proceedings of the Seventh SPIE Symposium on Storage and Retrieval for Image and Video Databases,* San Jose, CA (pp. 631-942).

Torratha, A., & Oliva, A. (1999). Semantic organization of scenes using discriminant structural templates. *International Conference on Computer Vision* (ICCV99) (pp. 1253-1258).

Traina, A.J.M., Traina, C., Jr., Bueno, J.M., & Chino, F.J.T. (2003). Efficient content-based image retrieval through metric histograms. *World Wide Web Internet and Web Information Systems, 6*, 157-185.

Zhao, R., & Grosky, W.I. (2001). Bridging the semantic gap in image retrieval. *Distributed multimedia databases: Techniques and applications* (pp. 14-36). Hershey, PA: Idea Group Publishing.

Zhou, P., Feng, J.F., & Shi, Q.Y. (2001). Texture feature based on local Fourier transform. *International Conference on Image Processing*, *2*, 610-613.

KEY TERMS

Computational Emergence: Here it is assumed that computational interactions can generate different features or behaviors. This is one of the approaches in the field of artificial life.

Content-Based Image Retrieval: In this kind of retrieval, symmetry between input image and images of database are established based on contents of the images under consideration.

Embedded Shape Emergence: In embedded shape emergence, all the emergent shapes can be identified by set theory kind of procedures on the original shape under consideration. For example, in a set S={a,b,c,d,e}, we can find subsets like S1={a,b,c}, S2={c,d,e}, S3={a,c,e} and so on.

Emergence Index: Image retrieval where the hidden or emergence meanings of the images are studied and based on those hidden meanings as well as explicit meanings. Where there is no hidden meaning at all, an index of search is defined to retrieve images, called emergence index.

Emergence Relative to a Model: In this case, deviation of the behavior from the original model gives rise to emergence.

Illusory Shape Emergence: In illusory shape emergence, contours defining a shape are perceived even though no contours are physically present. Here set theory procedures are not enough and more effective procedures have to be applied to find these hidden shapes.

Thermodynamic Emergence: This is of the view that new stable features or behaviors can arise from equilibrium through the use of thermodynamic theory.

Conducting Ethical Research in Virtual Environments

Lynne D. Roberts
University of Western Australia, Australia

Leigh M. Smith
Curtin University of Technology, Australia

Clare M. Pollock
Curtin University of Technology, Australia

INTRODUCTION

The rapid growth of the Internet has been accompanied by a growth in the number and types of virtual environments supporting computer-mediated communication. This was soon followed by interest in using these virtual environments for research purposes: the recruitment of research participants, the conduct of research and the study of virtual environments. Early research using virtual environments raised a number of ethical issues and debates. As early as 1996, a forum in the *The Information Society* (vol. 12, no. 2) was devoted to ethical issues in conducting social science research online. The debate has continued with more recent collaborative attempts to develop guidelines for ethical research online (Ess & Association of Internet Researchers, 2002; Frankel & Siang, 1999).

BACKGROUND

The basic principles of ethical research with humans are integrity, respect, beneficence and justice (National Health & Medical Research Council, 1999). Based on these principles, many professional associations provide ethical guidelines, or codes, for the conduct of research. However, these codes have typically been developed for use in offline settings, prior to consideration of research being conducted online[1]. While these codes contain guiding principles for research generally, the translation of these principles into actions for conducting research in virtual environments is open to interpretation. The process of translating ethical guidelines into ethical practice online involves a deliberation of the options available to the researcher and the likely impact on research participants, their communities and the research process. Central concerns in this process are maintaining respect for individuals, their online identities and the ownership of words.

PUBLIC VS. PRIVATE SPACE

Research online can take place within a range of virtual environments that vary in terms of purpose, synchronicity, access, number of users and norms. A major issue in developing ethical research procedures for use within a particular virtual environment is determining whether the setting represents a private or public "space". Various attempts have been made to distinguish between the public and the private in virtual environments (see, for example, Lessig, 1995) but little agreement has been reached. There are currently no clear guidelines for researchers on what constitutes private versus public space in virtual environments, yet the distinction is important as it affects the rights of participants to be advised of the research and to give or withhold their informed consent.

The defining of public versus private space cannot be reduced to the single dimension of accessibility to the virtual environment. Interactions that occur within publicly accessible virtual environments may be perceived by participants to be private. Newsgroups can be accessed without restriction, yet newsgroup postings can be, and frequently are, high in self-disclosure and are perceived by many users to be private (Witmer, 1997). Similarly, support groups on sensitive issues may be conducted in publicly accessible sites with participants adhering to norms of confidentiality and privacy (Elgesem, 2002).

Some ethical codes exempt naturalistic observations and archival research from requiring informed consent where no harm or distress is likely to come to those researched and where their confidentiality is protected. King (1996) highlighted the potential for psychological harm to members of online groups where research is conducted and published without the prior knowledge and informed consent of participants. Where there has been the expectation of privacy within a group (however misinformed that expectation may be), the individual may feel violated upon hearing of, or reading, the results of that research.

Where the presumption is made that online communication occurs in public space simply because it is accessible without restriction, an anomaly may result in how research participants are treated in equivalent settings in on- and off-line research. For example, research on support groups off-line requires the informed consent of research participants, while similar research online may occur without the knowledge or informed consent of the participants, on the grounds that all postings are public documents (see, for example, Salem, Bogat & Reid's 1997 study of a depression support group).

Table 1 summarizes possible dimensions against which the public/private nature of a virtual environment can be assessed. Virtual environments where all dimensions fall on the left-hand side of the continuums may be deemed as public environments for research purposes and subject to guidelines for research in public settings. Virtual environments where all dimensions are on the right should be deemed as private environments, requiring informed consent from research participants. The difficulty arises with the majority of settings that do not fall clearly into public or private spaces. Researchers do not have the right to define virtual environments as public or private to meet their own research needs (Waskul & Douglass, 1996). Rather, account should be taken of the size and nature of the online forum and the intrusiveness of the study. Consideration should be made of the likely effect of the request to conduct research and the research itself on research participants and their communities. The process of requesting consent to research may in itself alter group dynamics (Sixsmith & Murray, 2001).

INFORMED CONSENT

Research conducted in virtual environments that have been conceptualized as private settings requires the informed consent of research participants. Obtaining informed consent in virtual environments is more problematic than in off-line research as participants are frequently geographically dispersed. In addition, research participants may be reluctant to divulge details of off-line identities required for the signing of consent forms. A range of options has been suggested for obtaining informed consent in online research (Bruckman, 1997; Flicker, Haans & Skinner, 2004; Jacobson, 1999; Roberts, Smith & Pollock, 2004; Smith & Leigh, 1997) and these have been summarized in Table 2. Selection of a method for obtaining informed consent will necessarily be dependent upon the type of virtual environment, the level of anonymity required by research participants, and their access to high-level computing facilities. Regardless of the method used, the information about the research should be presented in a format that the research participants can keep and refer back to at any time before, during, or after their research participation. Care needs to be taken to fully inform potential research participants of the possible consequences of their research involvement (Reid, 1996).

In addition to seeking consent from research participants in virtual environments, it may be advisable to also seek the consent of gatekeepers of the community and advise the community as a whole of the research being undertaken. Advising communities of a research project requires the public identification of the researcher. In some circumstances, the decision to research within a particular virtual environment may be made after the researcher has been either an active participant or "lurker" within that environment. We recommend that researchers make their researcher status overt as soon as the research process begins. This may include identifying as a researcher in pseudonyms (Roberts et al., 2004), descriptions (Allen, 1996) or objects (Reid, 1996); linking between research and social identities (Roberts et al., 2004); and posting information about the research.

Advising communities of a research project may take ongoing effort in public virtual environments without membership boundaries. Identifying oneself as a researcher once within an online group does not mean that absent or future members of the group are also informed of the researcher's role (Sixsmith & Murray, 2001). There may be a need to re-identify researcher status and restate and clarify the role of the researcher on an ongoing basis.

PROTECTING ANONYMITY AND CONFIDENTIALITY

Individuals typically adopt a pseudonym (or pseudonyms) for use in virtual environments, providing a level of anonymity. While it has been argued that research involv-

Table 1. Dimensions of public and private space in virtual environments

Accessibility:	Accessible to all	➔	Restricted membership
Users' perceptions:	Public	➔	Private
Community statement:	Research permitted	➔	Research prohibited
Topic sensitivity:	Low	➔	High
Permanency of records:	Public archives	➔	Private logs only

Table 2. Methods of obtaining informed consent in online research

	Format of information	*How consent obtained*
Signed consent:	Hard copy or electronic	Post, fax or email
Implied consent:	Electronic	Gateway WWW page Logging of consent Use of password protected site

ing pseudonymous characters is exempt from regulations governing human subjects as "true" or offline identities are not known (Jacobson, 1999), there are often links between the individual and their pseudonyms that decrease the level of anonymity a pseudonym provides (Allen, 1996; Bruckman, 2002; Jacobson, 1996). These are presented in Table 3. The combination of these factors means that researchers cannot assume that pseudonyms provide adequate protection for offline identities. The degree of anonymity conferred in virtual environments does not reduce the ethical requirements for researchers to protect the anonymity of research participants and virtual interaction settings (Waskul & Douglass, 1996).

Protecting the anonymity of the individual extends to protecting the anonymity of their pseudonym(s), as representations of the individual online. Pseudonyms themselves gain reputations over time (Bruckman, 2002). Researchers can provide varying levels of protection to research participants' anonymity (see Table 4). The practice of replacing existing pseudonyms with other pseudonyms in research materials confers little additional protection to the existing pseudonym when text searches can be used to identify source documents (Allen, 1996). Further, other community members may seek to identify disguised identities and may share this information with others (Bruckman, 2002).

In addition to protecting the anonymity of research participants in research reports, the data collected needs to be kept secure in order to protect confidentiality. Maintaining the security of data collected in computer-mediated research poses unique difficulties. Confidentiality may be breached at the site of data collection, during transmission of data, or in the storage of data. Sites at which data is collected may not be secure and may be subject to surveillance by gate-keepers or "hackers" (Rhodes, Bowie & Hergenrather, 2003). Confidentiality of data may be breached during data transmission where another party intercepts data (Nosek, Banaji & Greenwald, 2002). This may include the Internet service provider of the research participant or researcher. Employers may also monitor employees' email (Sipior & Ward, 1995; Weisband & Reinig, 1995). Confidentiality of data may be

Table 3. Factors that decrease the anonymity afforded by pseudonyms

- Use of name, derivation of name or nickname
- Use of same pseudonym across virtual environments with differing requirements for identification
- Self-disclosure
- Active seeking of identifying information by others

Table 4. Levels of anonymity (site, pseudonym & quotations)

- Identify site, use online pseudonym and directly quote
- Identify site, use pseudonym of online pseudonym and directly quote
- Identify site, use pseudonym of online pseudonym and paraphrase
- Do not identify site, use online pseudonym and directly quote
- Do not identify site, use pseudonym of online pseudonym and directly quote
- Do not identify site, use pseudonym of online pseudonym and paraphrase

breached during storage of data by hackers, employers or as a result of "open records" legislation (Pittenger, 2003). Online researchers need to provide the most secure forms of data collection, transmission and storage possible, aiming to minimize the risks of unauthorized persons gaining access to research data at any stage of the research process. The procedures used to ensure this will differ according to the type of virtual media used.

OWNERSHIP OF WORDS

The ownership of electronic messages has been contested. It is still unclear whether the individual who authored a message, the community to which it was sent, or anyone who has access to the message is the owner of the electronic message. Electronic postings may be considered original works protected by copyright, although this has not yet been legally tested (Sixsmith & Murray, 2001). If informed consent is not obtained to use electronic messages, copyright provisions suggest that they are subject to "fair dealing" for research purposes, and should be attributed to the author (Australian Copyright Council, 2001). Researchers who neither obtain informed consent, nor reference the material they are quoting, risk violating both ethical and copyright standards. With the consent of the research participant, quotes in research may be attributed to the individual, their online pseudonym, or used anonymously. Respect for research participants is demonstrated through asking, and abiding by, their preferences for anonymity, pseudonymity or identification. However, this is not without its problems. Individual preferences for identification may not be consistent with the norms or wishes of the community. There can be tensions between respecting copyright entitlements of individuals and protecting the privacy of other participants within a virtual environment. Roberts et al. (2004) highlighted the potential for negative impacts on the privacy of other virtual environment members when one research participant's work is fully attributed, including information on the virtual environment.

RETURNING RESEARCH FINDINGS TO THE COMMUNITY

A requirement in some ethical codes is to provide research participants with information about the outcome of the research. In addition to being a requirement, this can also demonstrate respect for the individuals who participated in the research. A summary of research findings can be provided to research participants in hardcopy or electronic format. Where research participants are reluctant to provide contact information that may link their on- and off-line identities the summary can be placed on a Web site or sent through the messaging system of the virtual community (Roberts et al., 2004).

FUTURE TRENDS

Rapidly changing technologies will result in the development of an increasing range of virtual environments that may be used for research purposes. The precise characteristics of these new virtual environments may vary greatly from the virtual environments available today. Before conducting research within each new type of environment researchers will need to address the intrusiveness of the proposed research, the perceived privacy of the research setting, the vulnerability of the community, the potential for harm to individuals and/or the community, and how confidentiality will be maintained and intellectual property rights respected in their research proposals (Eysenback & Till, 2001). This requires a consideration of the likely impacts of the research on both research participants and the communities in which the research is conducted. It should be guided by researcher's knowledge and adherence to the "netiquette" and social norms of the virtual environments concerned. Guided by the principles outlined in their ethical codes, researchers will need to develop ethically defensible strategies that balance the needs of research participants and their online communities and offers protection to both.

CONCLUSION

Our approach to the conduct of ethical research in virtual environments is based on a human research perspective, explicitly recognizing that communication online is conducted by individuals who interact via their online identities. Our focus is therefore on individuals rather than texts. We privilege the rights of individuals to make informed consent about whether or not traces of their interaction online (e.g., logs, postings) can be used for research purposes. We believe this approach is consistent with general ethical guidelines for human research in the social sciences. Alternative perspectives to the conduct of ethical research in virtual environments place a stronger emphasis on the cultural production of texts and performance (Bassett & O'Riordan, 2002; White, 2002) reflecting calls for an ethical pluralism in Internet research that recognizes a range of ethical perspectives as legitimate (Ess, 2002). Regardless of the perspective adopted, all research should comply with the principles of ethical research as outlined in professional associations' codes of ethics or by institutional review boards. In the absence of specific guidelines for online research and where review committees are unfamiliar with online research issues (Keller & Lee, 2003), we recommend researchers are guided by the principles outlined in their code, adapting the guidelines for use in virtual environments as necessary.

REFERENCES

Allen, C.L. (1996). What's wrong with the "golden rule"? Conundrums of conducting ethical research in cyberspace. *The Information Society, 12*, 175-187.

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. Retrieved October 29, 2003, from *http://www.apa.org/ethics/code2002.html*

Australian Copyright Council. (2001). Information sheet G53. Copying for research or study. Retrieved October 22, 2003, from *http://www.copyright.org.au/*

Bassett, E.H., & O'Riordan, K. (2002). Ethics of Internet research: Contesting the human subjects research model. *Ethics and Information Technology, 4*, 233-247.

Bruckman, A. (2002). Studying the amateur artist: A perspective on disguising data collected in human subjects research on the Internet. *Ethics and Information Technology, 4*, 217-231.

Bruckman, A.S. (1997). MOOSE Crossing: Construction, community and learning in networked virtual world for kids (Doctoral dissertation, MIT Media Lab, 1997). Dissertation Abstracts International, DAI-A 58/11, 4241.

Elgesem, D. (2002). What is special about the ethical issues in online research? *Ethics and Information Technology, 4*, 195-203.

Ess, C. (2002). Introduction. *Ethics and Information Technology, 4*, 177-188.

Ess, C., & AoIR Ethics Working Committee (2002). Ethical decision-making and Internet research: Recommendations from the aoir ethics working committee. Retrieved October 22, 2003, from *http://www.aoir.org/reports/ethics.pdf*

Eysenbach, G. & Till, J.E. (2001). Ethical issues in qualitative research on Internet communities. *British Medical Journal, 323*, 1103-1105.

Flicker, S., Haans, D., & Skinner, H. (2004). Ethical dilemmas in research on Internet Communities. *Qualitative Health Research, 14*(1), 124-134.

Frankel, M.S., & Siang, S. (1999). Ethical and legal aspects of human subjects research on the Internet: A report of a workshop June 10-11, 1999. Retrieved October 22, 2003, from *http://www.aaas.org/spp/dspp/sfrl/projects/intres/main.htm*

Jacobson, D. (1996). Contexts and cues in cyberspace: the pragmatics of names in text-based virtual realities. *Journal of Anthropological Research, 52*, 461-479.

Jacobson, D. (1999). Doing research in cyberspace. *Field Methods, 11*(2), 127-145.

Keller, H.E, & Lee, S. (2003). Ethical issues surrounding human participants research using the Internet. *Ethics & Behavior, 13*(3), 211-219.

King, S. (1996). Researching Internet communities: Proposed ethical guidelines for the reporting of the results. *The Information Society, 12*(2), 119-127.

Lessig, L. (1995). The path of cyberlaw. *Yale Law Journal, 104*, 1743-1755.

National Health and Medical Research Council. (1999). National statement on ethical conduct in research involving humans. Commonwealth of Australia. Retrieved March 5, 2003, from *http://www.health.gov.au/nhmrc/publications/humans/contents.htm*

Nosek, B.A., Banaji, M.R., & Greenwald, A.G. (2002). E-research: Ethics, security, design, and control in psychological research on the Internet. *Journal of Social Issues, 58*(1), 161-176.

Pittenger, D.J. (2003). Internet research: An opportunity to revisit classic ethical problems in behavioral research. *Ethics & Behavior, 13*(1), 45-60.

Reid, E. (1996). Informed consent in the study of on-line communities: A reflection on the effects of computer-mediated social research. *The Information Society, 12*, 169-174.

Rhodes, S.D., Bowie, D.A., & Hergenrather, K.C. (2003). Collecting behavioural data using the World Wide Web: Considerations for researchers. *Journal of Epidemiology and Community Health, 57*(1), 68-73.

Roberts, L.D., Smith, L.M., & Pollock, C.M. (2004). Conducting ethical research online: Respect for individuals, identities, and the ownership of words. In E. Buchanan (Ed.). *Readings in virtual research ethics: Issues and controversies* (pp. 159-176). Hershey, PA: Idea Group Publishing.

Salem, D.A., Bogat, G.A., & Reid, C. (1997). Mutual help goes on-line. *Journal of Community Psychology, 25*, 189-207.

Sipior, J.C., & Ward, B.T. (1995). The ethical and legal quandary of email privacy. *Communications of the ACM, 38*(12), 48-54.

Sixsmith, J., & Murray, C.D. (2001). Ethical issues in the documentary data analysis of internet posts and archives. *Qualitative Health Research, 11*(3), 423-432.

Smith, M.A, & Leigh, B. (1997). Virtual subjects: Using the Internet as an alternative source of subjects and research environment. *Behavior Research Methods, Instruments, & Computers, 29,* 496-505.

Waskul, D., & Douglass, M. (1996). Considering the electronic participant: Some polemical observations on the ethics of on-line research. *The Information Society, 12,* 129-139.

Weisband, S.P., & Reinig, B.A. (1995). Managing user perceptions of email privacy. *Communications of the ACM, 38*(12), 40-47.

White, M. (2002). Representations or people? *Ethics and Information Technology, 4*, 249-266.

Witmer, D.F. (1997). Risky business: Why people feel safe in sexually explicit on-line communication. *JCMC, 2*(4). Retrieved March 19, 1997, from *http:// jcmc.huji.ac.il/* vol2/issue4/witmer2.html

KEY TERMS

Computer-Mediated Communication: Communication between two or more individuals that occurs via computer networks. Computer-mediated communication may be text, audio, graphics or video based and occur synchronously (in "real time") or asynchronously (delayed).

Informed Consent: An individual's freely given consent to participate in research based on information provided by the researcher(s) about the research, possible risks associated with the research and the voluntary nature of participation. Informed consent must be obtained without coercion or undue influence.

Netiquette: The etiquette, or social rules, associated with communicating online. Netiquette may vary across virtual environments.

Private Space: Off-line, private space refers to geographical areas that are not for general or public use (e.g., your home). Online, the term private space is commonly used to refer to virtual environments, or parts of virtual environments that have restrictions on who may access them.

Pseudonym: The fictitious name adopted for use within a virtual environment. An individual may consistently use the same pseudonym or adopt several pseudonyms for use within and between virtual environments.

Public Space: Off-line, public space refers to geographical areas that are accessible to the general public (e.g., streets). Online, the term public space is commonly used to refer to virtual environments that do not have restrictions on access.

Virtual Identity: Representation of the individual in a virtual environment. The form of representation varies across virtual environments and may range from a pseudonym only (Internet Relay Chat), a pseudonym combined with a character description (Multi-User Dimensions) through to graphical representations, or avatars, in graphics based environments. An individual may have multiple virtual identities.

ENDNOTE

[1] Over time as codes are updated, consideration of research conducted online may be included. See, for example, the revised *Ethical Principles of Psychologists and Code of Conduct* (American Psychological Association, 2002).

C

Consistent Queries Over Databases with Integrity Constraints

Sergio Greco
DEIS Università della Calabria, Italy

Ester Zumpano
DEIS Università della Calabria, Italy

INTRODUCTION

The aim of data integration is to provide a uniform integrated access to multiple heterogeneous information sources, which were designed independently for autonomous applications and whose contents are strictly related.

Integrating data from different sources consists of two main steps: the first in which the various relations are merged together and the second in which some tuples are *removed* (or *inserted*) from the resulting database in order to satisfy integrity constraints.

There are several ways to integrate databases or possibly distributed information sources, but whatever, integration architecture we choose, the heterogeneity of the sources to be integrated, causes subtle problems. In particular, the database obtained from the integration process may be inconsistent with respect to integrity constraints, that is, one or more integrity constraints are not satisfied. Integrity constraints represent an important source of information about the real world. They are usually used to define constraints on data (functional dependencies, inclusion dependencies, etc.) and have, nowadays, a wide applicability in several contexts such as semantic query optimization, cooperative query answering, database integration and view update.

Since, the satisfaction of integrity constraints cannot generally be guaranteed, if the database is obtained from the integration of different information sources, in the evaluation of queries, we must compute answers which are consistent with the integrity constraints. The following example shows a case of inconsistency.

Example 1. Consider the following database schema consisting of the single binary relation *Teaches (Course, Professor)* where the attribute *Course* is a key for the relation. Assume there are two different instances for the relations *Teaches*, D1={(c1,p1),(c2,p2)} and D2={(c1,p1),(c2,p3)}.

The two instances satisfy the constraint that *Course* is a key but, from their union we derive a relation which does not satisfy the constraint since there are two distinct tuples with the same value for the attribute *Course*.

In the integration of two conflicting databases simple solutions could be based on the definition of preference criteria such as a partial order on the source information or a majority criteria (Lin and Mendelzon, 1996). However, these solutions are not generally satisfactory and more useful solutions are those based on 1) the computation of 'repairs' for the database, 2) the computation of consistent answers (Arenas et al., 1999).

The computation of repairs is based on the definition of minimal sets of insertion and deletion operations so that the resulting database satisfies all constraints. The computation of consistent answers is based on the identification of tuples satisfying integrity constraints and on the selection of tuples matching the goal.For instance, for the integrated database of *Example 1*, we have two alternative repairs consisting in the deletion of one of the tuples *(c2,p2)* and *(c2,p3)*. The consistent answer to a query over the relation *Teaches* contains the unique tuple *(c1,p1)* so that we don't know which professor teaches course *c2*.

Therefore, it is very important, in the presence of inconsistent data, to compute the set of consistent answers, but also to know which facts are unknown and if there are possible repairs for the database.

TECHNIQUES FOR QUERYING AND REPAIRING DATABASES

Recently, there have been several proposals considering the integration of databases as well as the computation of queries over inconsistent databases. Most of the techniques work for restricted from of constraints and only recently have there been proposals to consider more general constraints. In this the following we give an informal description of the main techniques proposed in the literature.

- In Agarwal et al. (1995) it is proposed an extension of relational algebra, called *flexible algebra*, to deal with data having tuples with the same value for the key attributes and conflicting values for the other attributes. The technique only considers constraints defining functional dependencies and it is sound only for the class of databases having dependencies determined by a primary key consisting of a single attribute.
- In Dung (1996) it is proposed the Integrated Relational Calculus, an extension of flexible algebra for other key functional dependencies based on the definition of *maximal consistent subsets* for a possibly inconsistent database. Dung proposed extending relations by also considering null values denoting the absence of information with the restriction that tuples cannot have null values for the key attributes. The Integrated Relational Calculus overcomes some drawbacks of the flexible relational algebra. Anyhow as both techniques consider restricted cases the computation of answers can be done efficiently.
- In Lin and Mendelzon (1996), it is proposed an approach taking into account the majority view of the knowledge bases in order to obtain a new relation which is consistent with the integrity constraints. The technique proposes a formal semantics to merge first-order theories under a set of constraints.

Example 2. Consider the following three relation instances which collect information regarding author, title and year of publication of papers:

- Bib1={(John,T1,1980),(Mary,T2,1990)},
- Bib2={(John,T1,1981),(Mary,T2,1990)},
- Bib3={(John,T1,1981), (Frank,T3,1990)}

From the integration of the three databases Bib1, Bib2 and Bib3 we obtain the database Bib={(John,T1,1980), (Mary,T2,1990), (Frank,T3,1990)}.

Thus, the technique, proposed by Lin and Mendelson, removes the conflict about the year of publication of the paper T1 written by the author John observing that two of the three source databases, that have to be integrated, store the value 1980; thus the information that is maintained is the one which is present in the majority of the knowledge bases.

However, the 'merging by majority' technique does not resolve conflicts in all cases since information is not always present in the majority of the databases and, therefore, it is not always possible to choose between alternative values. Thus, generally, the technique stores disjunctive information and this makes the computation of answers more complex (although the computation becomes efficient if the 'merging by majority' technique can be applied); moreover, the use of the majority criteria involves discarding inconsistent data, and hence the loss of potentially useful information.

- In Arenas et al. (1999) it is introduces a logical characterisation of the notion of consistent answer in a possibly inconsistent database. The technique is based on the computation of an equivalent query $T_\omega(Q)$ derived from the source query Q. The definition of $T_\omega(Q)$ is based on the notion of residue developed in the context of semantic query optimization.

More specifically, for each literal B, appearing in some integrity constraint, a residue *Res(B)* is computed. Intuitively, *Res(B)* is a universal quantified first order formula which must be true, because of the constraints, if B is true. Universal constraints can be rewritten as denials, i.e., logic rules with empty heads of the form $\leftarrow B_1 \wedge \ldots \wedge B_n$.

Let A be a literal, r a denial of the form $\leftarrow B_1 \wedge \ldots \wedge B_n$, B_i (for some $1 \le i \le n$) a literal unifying with A and θ the most general unifier for A and B_i such that variables in A are used to substitute variables in B_i but they are not substituted by other variables. Then, the residue of A with respect to r and B_i is:

$$\begin{aligned} Res(A,r,Bi) &= not((B_1 \wedge \ldots \wedge B_{i-1} \wedge B_{i+1} \wedge \ldots \wedge B_n)\ \theta) \\ &= not\ B_1\theta \vee \ldots \vee not\ B_{i-1}\theta \vee not\ B_{i+1}\theta \vee \ldots \vee not\ B_n\theta. \end{aligned}$$

The residue of A with respect to r is $Res(A,r) = \wedge_{B_i | A=B_i\theta}$ $Res(A,r,B_i)$ consisting of the conjunction of all the possible residues of A in r whereas the residue of A with respect to a set of integrity constraints IC is $Res(A) = \wedge_{r \in IC} Res(A,r)$.

Thus, the residue of a literal A is a first order formula which must be true if A is true. The operator $T_w(Q)$ is defined as follows:

$T_0(Q) = Q$;
$T_i(Q) = T_{i-1}(Q) \wedge R$ where R is a residue of some literal in $Ti\text{-}1$.

The operator T_w represents the fixpoint of T.

Example 3. Consider a database D consisting of the following two relations:

Supplier	Department	Item
c1	d1	i1
c2	d2	i2

Item	Type
i1	t
i2	t

with the integrity constraint, defined by the following first order formula:

$$\forall(X, Y, Z)\ [\ Supply(X,Y,Z) \wedge Class(Z,t) \supset X{=}c1\ ]$$

stating that only supplier *c1* can supply items of type *t*.

The database *D = { Supply(c1, d1, i1), Supply(c2, d2, i2), Class(i1, t), Class(i2, t) }* is inconsistent because the integrity constraint is not satisfied (an item of type *t* is also supplied by supplier *c2*).

This constraint can be rewritten as:

$$\leftarrow\ Supply(X,Y,Z) \wedge Class(Z,t) \wedge \mathrm{X} \neq \mathrm{c1},$$

where all variables are (implicitly) universally quantified. The residue of the literals appearing in the constraint are:

$$Res(Supply(X,Y,Z)\ =\ not\ Class(Z,t) \vee\ \ X = c1$$
$$Res(Class(Z,t))\ =\ not\ \ Supply(X,Y,Z) \vee\ X = c1$$

The iteration of the operator *T* to the query goal Class(Z,t) gives:

$$T0(Class(Z,t))\ =\ Class(Z,t),$$
$$T1(Class(Z,t)) = Class(Z,t) \wedge not\ Supply(X,Y,Z) \vee X = c1),$$
$$T2(Class(Z,t)) = Class(Z,t) \wedge (not\ Supply(X,Y,Z) \vee X = c1).$$

At Step 2 a fixpoint is reached since the literal *Class(Z,t)* has been "expanded" and the literal *not Supply(X,Y,Z)* does not have a residue associated to it. Thus, to answer the query *Q = Class(Z,t)* with the above integrity constraint, the query $T_\omega(Q){=}Class(Z,t) \wedge (\ not\ Supply(X,Y,Z) \vee X = c1)$ is evaluated. The computation of $T_\omega(Q)$ over the above database gives the result *Z=i1*.

The technique, more general than the previous ones, has been shown to be complete for universal binary integrity constraints and universal quantified queries. However, the rewriting of queries is complex since the termination conditions are not easy to detect and the computation of answers generally is not guaranteed to be polynomial.

- In Arenas et al. (2000) it is proposed an approach consisting in the use of a Logic Program with Exceptions (LPe) for obtaining consistent query answers. An LPe is a program with the syntax of an extended logic program (ELP), that is, in it we may find both logical (or strong) negation ($\neg$) and procedural negation (not). In this program, rules with a positive literal in the head represent a sort of general default, whereas rules with a logically negated head represent exceptions. The semantic of an LPe is obtained from the semantics for ELP's, by adding extra conditions that assign higher priority to exceptions. The method, given a set of integrity constraints ICs and an inconsistent database instance, consists in the direct specification of database repairs in a logic programming formalism. The resulting program will have both negative and positive exceptions, strong and procedural negations, and disjunctions of literals in the head of some of the clauses; that is it will be a disjunctive extended logic program with exceptions. As in (Arenas et al., 1999) the method considers a set of integrity constraints, IC, written in the standard format $\vee_{i=1}^{n} P_i(x_i) \vee \vee_{i=1}^{m}(\neg Q_i(y_i) \vee \varphi$ where φ is a formula containing only built-in predicates, and there is an implicit universal quantification in front. This method specifies the repairs of the database, D, that violate IC, by means of a logical program with exceptions, Π^D. In Π^D for each predicate *P* a new predicate P' is introduced and each occurrence of *P* is replaced by *P'*. More specifically, Π^D is obtained by introducing:

Persistence Defaults: For each base predicate *P*, the method introduces the persistence defaults:

$$P'(x)\ \leftarrow\ P(x),$$
$$\neg P'(x) \leftarrow\ not\ P(x).$$

The predicate *P'* is the repaired version of the predicate *P*, so it contains the tuples corresponding to *P* in a repair of the original database.

Stabilizing Exceptions: From each *IC* and for each negative literal *not* Q_{i0} in *IC*, the negative exception clause is introduced:

$$\neg Q'_{i0}(y_{i0}) \leftarrow \wedge_{i=1..n} \neg P'_i(x_i), \wedge_{i\neq 0} Q'_i(y_i), \varphi'$$

where φ' is a formula that is logically equivalent to the logical negation of φ. Similarly, for each positive literal P_{i1} in the constraint the positive exception clause:

$$P'_{i1}(x_{i1}) \leftarrow \wedge_{i\neq i1} \neg P'_i(x_i), \wedge_{i=1..m} Q'_i(y_i), \varphi$$

is generated. The meaning of the Stabilizing Exceptions is to make the ICs be satisfied by the new predicates. These exceptions are necessary but not sufficient to ensure that the changes the original subject should be subject to, in order to restore consistency, are propagated to the new predicates.

Triggering Exceptions: From the IC in standard form the disjunctive exception clause:

$$\vee_{i=1..n} P'_i(x_i) \vee \vee_{i=1..m} Q'_i(y_i) \leftarrow \wedge_{i=1..n} \; not\ P_i(x_i), \wedge_{i=1..m} Q_i(y_i), \varphi'$$

is produced.

The program Π^D constructed as shown above is a 'disjunctive extended repair logic program with exceptions for the database instance D'. In Π^D positive defaults are blocked by negative conclusions, and negative defaults, by positive conclusions.

Example 4. Consider the database $D = \{p(a), q(b)\}$ with the inclusion dependency *ID*: $p(X) \supset q(X)$

In order to specify the database repairs the new predicates *p'* and *q'* are introduced. The resulting repair program has four default rules expressing that *p'* and *q'* contain exactly what p and q contain, resp.:

$p'(x) \leftarrow p(x);$
$q'(x) \leftarrow q(x);$
$\neg p'(x) \leftarrow not\ p(x)$ *and*
$\neg q'(x) \leftarrow not\ q(x);$

two stabilizing exceptions :

$q'(x) \leftarrow p'(x);$
$\neg p'(x) \leftarrow \neg q'(x);$

and the triggering exception:

$\neg p'(x) \vee q'(x) \leftarrow p(x), not\ q(x).$

The answer sets are $\{ p(a), q(b), p'(a), q'(b), \neg p'(a) \}$ and $\{ p(a), q(b), p'(a), q'(b), q'(b) \}$ that correspond to the two expected database repairs.

The method can be applied to a set of domain independent binary integrity constraints *IC*, that is the constraint can be checked w.r.t. satisfaction by looking to the active domain, and in each *IC* appear at most two literals.

- In Greco and Zumpano (2000) a general framework for computing repairs and consistent answers over inconsistent databases with universally quantified variables was proposed. The technique is based on the rewriting of constraints into extended disjunctive rules with two different forms of negation (negation as failure and classical negation). The disjunctive program can be used for two different purposes: compute 'repairs' for the database, and produce consistent answers, i.e., a maximal set of atoms which do not violate the constraints. The technique is sound and complete (each stable model defines a repair and each repair is derived from a stable model) and more general than techniques previously proposed.

More specifically, the technique is based on the generation of an extended disjunctive program *LP* derived from the set of integrity constraints. The repairs for the database can be generated from the stable models of *LP*, whereas the computation of the consistent answers of a query *(g,P)* can be derived by considering the stable models of the program $P \cup LP$ over the database *D*.

Let *c* be a universally quantified constraint of the form:

$$\forall X \;\; [\, B_1 \wedge \ldots \wedge B_k \wedge not\ B_{k+1} \wedge \ldots \wedge not\ B_n \wedge \phi \supset B_0 \,]$$

then, *dj(c)* denotes the extended disjunctive rule

$$\neg B'_1 \vee \ldots \vee \neg B'_k \neg B'_{k+1} \wedge \ldots \vee B'_n \vee B'_0 \leftarrow (B_1 \vee B'_1), \ldots, (B_k \vee B'_k), (not\ B_{k+1} \vee \neg B'_{k+1}), \ldots, (not\ B_n \vee \neg B'_n), \phi, (not\ B_0 \vee \neg B'_0),$$

where B'_i denotes the atom derived from B_i, by replacing the predicate symbol *p* with the new symbol p_d if B_i is a base atom otherwise is equal to false. Let *IC* be a set of universally quantified integrity constraints, then $DP(IC) = \{ dj(c) \mid c \in IC \}$ whereas *LP(IC)* is the set of standard disjunctive rules derived from *DP(IC)* by rewriting the body disjunctions.

Clearly, given a database *D* and a set of constraints *IC*, $LP(IC)_D$ denotes the program derived from the union of the rules *LP(IC)* with the facts in *D* whereas $SM(LP(IC)_D)$ denotes the set of stable models of $LP(IC)_D$ and every stable model is consistent since it cannot contain two atoms of the form *A* and $\neg A$. The following example shows how constraints are rewritten.

Example 5. Consider the following integrity constraints:

$\forall X\ [\ p(X) \wedge not\ \ s(X) \supset q(X)\ ]$
$\forall X\ [\ q(X) \supset r(X)\ ]$

and the database D containing the facts p(a), p(b), s(a) and q(a).

The derived generalized extended disjunctive program is defined as follows:

$$\neg\, p_d(X) \vee s_d(X) \vee\; q_d(X) \leftarrow\; (p(X) \vee\; p_d(X)) \wedge$$
$$(\text{not}\;\; s(X) \vee \neg\, s_d(X)) \wedge (\text{not}\; q(X) \vee\; \neg\, q_d(X)).$$
$$\neg\, q_d(X) \neg\, r_d(X) \quad \leftarrow (q(X) \neg\, q_d(X)) \wedge (\text{not}\; r(X) \vee \neg r_d(X)).$$

The above rules can now be rewritten in standard form. Let P be the corresponding extended disjunctive Datalog program. The computation of the program P_D gives the following stable models:

$M_1 = D \cup \{\neg p_d(b), \neg q_d(a)\}$,
$M_2 = D \cup \{\neg p_d(b), r_d(a)\}$,
$M_3 = D \cup \{\neg q_d(a), s_d(b)\}$,
$M_4 = D \cup \{r_d(a), s_d(b)\}$,
$M_5 = D \cup \{q_d(b), \neg q_d(a), r_d(b)\}$ *and*
$M_6 = D \cup \{q_d(b), r_d(a), r_d(b)\}$.

A (generalized) extended disjunctive Datalog program can be simplified by eliminating from the body rules all literals whose predicate symbols are derived and do not appear in the head of any rule (these literals cannot be true). As mentioned before, the rewriting of constraints into disjunctive rules is useful for both (i) making the database consistent through the insertion and deletion of tuples, and (ii) computing consistent answers leaving the database inconsistent.

CONCLUSIONS AND FUTURE TRENDS

In the integration of knowledge from multiple sources two main steps are performed: the first in which the various relations are merged together and the second in which some tuples are removed (or inserted) from the resulting database in order to satisfy integrity constraints.

The database obtained from the merging of different sources could contain inconsistent data. In this chapter we investigated the problem of querying and repairing inconsistent databases. In particular we presented the different techniques for querying and repairing inconsistent databases (Agarwal et al., 1995; Lin and Mendelzon, 1996; Arenas et al., 1999; Greco and Zumpano, 2000).

As a future trend an interesting topic consists in specifying preference criteria so that selecting among a set of feasible repairs the preferable ones, i.e., those better conforming to the specified criteria. Preference criteria introduce desiderata on how to update the inconsistent database in order to make it consistent, thus they can be considered as a set of desiderata which are satisfied *if possible* by a generic repair. Therefore, informally a preferred repair is a repair that better satisfies preferences.

REFERENCES

Agarwal, Argaval, S., Keller, A.M., Wiederhold, G. and Saraswat, K. (1995). Flexible relation: An approach for integrating data from multiple, possibly inconsistent databases. *Proc. of the IEEE Int. Conf. on Data Engineering*, (pp. 495-504).

Arenas, M., Bertossi, L., and Chomicki, J. (1999). Consistent query answers in inconsistent databases. *Proc. Int. Conf. on Principles of Database Systems*, (pp. 68-79).

Arenas, M., Bertossi, L., and Chomicki, J. (2000). Specifying and querying database repairs using logic programs with exceptions. *Proc. Int. Conf. on Flexible Query Answering*, (pp. 27-41).

Baral, C., Kraus, S., Minker, J., and Subrahmanian,V. S. (1991). Combining knowledge bases consisting of first order theories. *Proc. Int. Symp. on Methodologies for Intelligent Sistems*, (pp. 92-101).

Bry, F. (1997). Query answering in information system with integrity constraints. *IFIP WG 11.5 Working Conf. on Integrity and Control in Inform. System*.

Dung, P. M. (1996). Integrating data from possibly inconsistent databases. *Proc. Int. Conf. on Cooperative Information Systems*, (pp.58-65).

Greco, G., Greco, S., and Zumpano, E. (2001). A logic programming approach to the integration, repairing and querying of inconsistent databases. *Proc. Int. Conf. on Logic Programming*.

Greco, G., Sirangelo, C., Trubitsyna, I., and Zumpano, E. (2003). Preferred repairs for inconsistent databases, *Proc. Int. Conf. on Database Engineering and Applications Symposium*.

Greco, S. and Zumpano, E. (2000). Querying inconsistent databases. *Proc. Int. Conf. on Logic Programming and Automated Reasoning* (pp. 308-325).

Greco, S. and Zumpano E. (2000b). Computing repairs for inconsistent databases. *Proc. Int. Symp. on Cooperative Database System for Advanced Applications*, (pp. 33-40).

Lin, J. (1996). A semantics for reasoning consistently in the presence of inconsistency. *Artificial Intelligence*, *86*(1), 75-95.

Lin, J. and Mendelzon, A. O. (1996). Merging databases under constraints. *Int. Journal of Cooperative Information Systems*, *7*(1), 55-76.

Lin, J. and Mendelzon, A. O. (1999). Knowledge base merging by majority. In R. Pareschi and B. Fronhoefer

(Eds.), *Dynamic worlds: From the frame problem to knowledge management.* Kluwer.

Ullman, J.K. (2000). *Information integration using logical views. 239*(2), pp. 189-210.

Wiederhold, G. (1992). Mediators in the architecture of future information systems. *IEEE Computer*, *25*(3), 38-49.

TERMS AND DEFINITIONS

Consistent Answer: A set of tuples, derived from the database, satisfying all integrity constraints.

Consistent Database: A database satisfying a set of integrity constraints.

Data Integration: A process providing a uniform integrated access to multiple heterogeneous information sources.

Database Repair: Minimal set of insert and delete operations which makes the database consistent.

Disjunctive Datalog Program: A set of rules of the form:

$$A_1 \vee \ldots \vee A_k \leftarrow B_1, ..., B_m, \text{not } B_{m+1}, \ldots, \text{not } B_n, \; k+m+n>0$$

where $A_1,\ldots,A_k, B_1,\ldots, B_n$ are atoms of the form $p(t_1,..., t_h)$, p is a predicate symbol of arity h and the terms $t_1,..., t_h$ are constants or variables.

Inconsistent Database: A database violating some integrity constraints.

Integrity Constraints: Set of constraints which must be satisfied by database instances.

C

Constructionist Perspective of Organizational Data Mining

Isabel Ramos
Universidade do Minho, Portugal

João Álvaro Carvalho
Universidade do Minho, Portugal

INTRODUCTION

Scientific or organizational knowledge creation has been addressed from different perspectives along the history of science and, in particular, of social sciences. The process is guided by the set of values, beliefs and norms shared by the members of the community to which the creator of this knowledge belongs, that is, it is guided by the adopted paradigm (Lincoln & Guba, 2000). The adopted paradigm determines how the nature of the studied reality is understood, the criteria that will be used to assess the validity of the created knowledge, and the construction and selection of methods, techniques and tools to structure and support the creation of knowledge. This set of ontological, epistemological, and methodological assumptions that characterize the paradigm one implicitly or explicitly uses to make sense of the surrounding reality is the cultural root of the intellectual enterprises. Those assumptions constrain the accomplishment of activities such as construction of theories, definition of inquiry strategies, interpretation of perceived phenomena, and dissemination of knowledge (Schwandt, 2000).

Traditionally, social realities such as organizations have been assumed to have an objective nature. Assuming this viewpoint, the knowledge we possess about things, processes, or events that occur regularly under definite circumstances, should be an adequate representation of them. Knowledge is the result of a meticulous, quantitative, and objective study of the phenomenon of interest. Its aim is to understand the phenomenon in order to be able to anticipate its occurrence and to control it.

Organizations can instead be understood as socially constructed realities. As such, they are subjective in nature since they do not exist apart from the organizational actors and other stakeholders. The stable patterns of action and interaction occurring internally and with the exterior of the organization are responsible for the impression of an objective existence.

BACKGROUND

The Rational and Emotional Nature of Personal Knowledge

Individual knowledge is actively constructed by the mind of the learner (Kafai & Resnick, 1996).

We make ideas instead of simply getting them from an external source. Idea making happens more effectively when the learner is engaged in designing and constructing an external artifact, which is meaningful for the learner, and he or she can reflect upon it and share it with others. From this constructionist description of the learning process, we can emphasize several elements associated with the creation of knowledge, namely, *cognition*, *introspection*, *action*, *interaction*, and *emotion*.

Through *cognitive* processes, humans construct mental representations of external and mental objects. *Introspection* is a specific type of cognition that permits the personal inquiry into subjective mental phenomena such as sensory experiences, feelings, emotions, and mental images (Damásio, 1999; Wallace, 2000). Through *action* and *interaction*, we create our experiences of the world we live in. The effective construction of personal knowledge requires the building of relationships between concepts and other mental constructs, in profoundly meaningful experiences (Shaw, 1996). All human experience is mediated by *emotions*, which drive our attention and concentration in order to help us to process external stimuli and to communicate with others.

The Historical and Socio-Cultural Context of Knowledge

A social reality is a construction in continuous reformulation that occurs whenever social actors develop social constructions that are external and sharable.

By the mere fact that people interact, influencing each other's mental constructs, social reality is in constant reconstruction. In this context, learning of new concepts and practices are happening continuously, either intentionally or unintentionally.

Learning happens inside specific mental and social spaces, meaning that what a group can learn is influenced by:

- The concepts, schemata, values, beliefs, and other mental constructs shared by the group.
- All knowledge we create about external things, events, and relationships, is based on and constrained by our mental constructs.
- The creation of knowledge is founded on the historical and socio-cultural context of its creators, providing a shared basis for the interaction inside a group. The continuous interaction of the group members, happening in a common environment, leads to similar mental constructs, a common interpretation of events, and the creation of shared meaning structures and external constructions.
- There is no viewpoint outside human subjectivity or historical and socio-cultural circumstances from which to study phenomena and to judge the inquiry process and the knowledge produced.

ODM AND KNOWLEDGE CREATION: PROBLEMS AND OPPORTUNITIES

ODM (also called Organizational Knowledge Discovery) has been defined as the process of analyzing organizational data from different perspectives and summarizing it into useful information for organizational actors who will use that information to increase revenues, reduce costs, or achieve other relevant organizational goals and objectives (Fayyad, Piatetsky-Shapiro, & Smyth, 1996; Matheus, Chan, & Piatetsky-Shapiro, 1993).

Data mining is a sub-process of the knowledge discovery. It leads to the finding of models of consumer behavior that can be used to guide the action of organizational actors. The models are built upon the patterns found out among data stored in large databases that are backed by statistical correlations among that data. Those patterns are extracted by specific mechanisms called data mining algorithms.

Attached to the discourse around the data mining tools, there is the idea that in the future, new and more powerful algorithms will be developed that will be able to find more valuable patterns and models, independently from human subjectivities and limitations. If it ever becomes possible to integrate the knowledge of the relevant business domain into the system, the algorithm would be able to decide the usefulness and validity of discovered patterns, correlations and models as well as to grow in sophistication by integrating these models in its knowledge of the business. The decision-making process would become extensively automated and guided by the objective reasoning of clear and rational rules implemented in a computer-based system.

However, this view has several drawbacks, namely:

1. Since all human knowledge has a tacit and non-expressible dimension, it will never be possible to integrate all relevant business knowledge in a repository to be analyzed by a data-mining algorithm.
2. The diversity of views about the business activities and their context is what allows for the emergence of organizational creativity and development, and the challenge of taken-for-granted concepts and practices (Bolman & Deal, 1991; Morgan, 1997; Palmer & Hardy, 2000). The stored knowledge representations are those around which there is some degree of consensus. This is important for the stability of work concepts and practices and to support organizational cohesion. However, they may also trap organizational actors in those concepts and practices, even when evidence shows they are threatening organizational success.
3. The relevance of knowledge representations stored in organizational repositories changes according to changes in the socio-cultural circumstances that offer the context for making sense of the representations. Only the organizational actors can understand those contexts and are able to give meaning to knowledge representations.
4. It is still believed that decision-making is or should be an essentially rational process, guided by cognitive processes such as planning, resolution of problems, and creativity (Sparrow, 1998). However, recent experiments in neurobiology show that emotion is an integral part of reasoning and decision-making (Damásio, 1999). Thus, only organizational actors can make decisions. The full automation of the process is not a realistic objective.

Instead of the present focus on the technological side of ODM, it would be interesting to adopt a constructionist approach and to focus on the social process of knowledge construction that makes ODM meaningful. With this new focus on people and the way they create and share knowledge, the main concern would be to mobilize the knowledge of organizational actors so the whole organization can benefit from it. This concern is justified by the awareness that the organization, seen as a community, is more intelligent than each one of its members, including any of its leaders.

LEVERAGING KNOWLEDGE CREATION IN ORGANIZATIONS: SOME CONSTRUCTIONIST GUIDELINES FOR ODM

With ODM, there is a special focus on knowledge about consumer behavior to support decision and action. ODM assists the organization in knowing the preferences of its customers and in anticipating their needs and reactions. The construction of this knowledge must be guided by the specific purposes of the several communities of practice that constitute the organization.

ODM and the knowledge it helps to create are social constructions. Repositories, data mining tools, and the resulting patterns, correlations and models are social artifacts that should be used to make ideas tangible, to negotiate meanings, and to facilitate communication between organizational actors. As such, they may become catalysts for the development of shared knowledge about consumer behavior, when they are used in the contexts of meaningful projects.

Data mining systems may become empowering tools in the sense that they make viable the analysis of large organizational repositories of knowledge representations. These knowledge representations are social constructions that connect organizational actors to a common view of the business concepts and practices that shape their intentions and interactions. Problems in the performance of organizational tasks or in organizational adaptation to environmental changes may reside in the inappropriateness of knowledge representations or in the tools used to extract rules and patterns from them. Knowledge representations were created and stored under specific historical and socio-cultural circumstances of which their readers must be aware in order to be able to understand their relevance or inadequacy.

Table 1 summarizes the constructionist guidelines for ODM, grouping them in two categories:

- guidelines that should be considered for the creation of rich learning environments in which data mining systems are used as social artifacts that leverage continuous learning, and
- guidelines that should be considered when using a specific data mining tool.

These guidelines are given from constructionist theories developed and applied in areas such as psychology, education, and organization theory.

FUTURE TRENDS

According to the assumptions of the constructionist perspective, ODM should be designed to involve organizational actors in the social construction of something

Table 1. A summary of constructionist guidelines for ODM

Creating rich learning environments	Using data mining tools
Work relationships must be strengthened in order to create the social cohesiveness needed for the ongoing production of shared constructions that engage the organization in developmental cycles.	Data mining results will support insight and creativity when organizational actors have enough time to reflect upon them and the opportunity to externalize and discuss their interpretations.
The construction of knowledge about customers' preferences and their future needs and reactions must be guided by the shared purposes of the specific communities of practice that constitute the organization.	Effective formal and informal communication must be fostered in order to become possible to discuss each other's interpretations of past experience in the light of the context in which it occurred.
Organizational repositories, data mining tools, and the results of data mining are social artifacts that should be used to make ideas tangible, to negotiate meanings, and to facilitate communication between organizational actors.	Theoretical tools, locally or externally developed, should be used to critically analyze the old meaning structures, facilitating the rearrangement of those structures.
Knowledge representations were created and stored under specific historical and socio-cultural circumstances of which their readers must be aware in order to be able to understand relevance or inadequacy of those representations.	The search and interpretation of patterns and models of consumer behavior should be guided by a multi-dimensional knowledge of the business domain, and work concepts and practices.

external and sharable. The designing of a marketing campaign, the making of a decision, the transformation of work concepts and practices are examples of social construction processes for which ODM could be viewed as relevant.

As a result of the process, the individual and shared knowledge will become more sophisticated, empowering the action of individuals and groups, and facilitating interaction. In this way, organizational actors consciously create cohesive and pluralist work environments, more prone to deal with problems and difficult decisions associated with consumer behavior. This perspective is more realistic than the traditional view of ODM as a process of making knowledge neutral and independent of the knower and social contexts in which it is created, in order to support decision-making processes idealized as inherently rational.

The tools used to support ODM fundamentally shape and define the process. Lack of appropriate tools impoverishes a social setting and makes social construction difficult. Future research is needed to study if current data mining systems facilitate organizational developmental activities. It will also be important to create practical experiences of designing and implementing the ODM process in specific organizational settings so that learning from a constructionist perspective can be supported.

CONCLUSION

This article describes ODM as a process for the social construction of knowledge. As such, the focus changes from the technology used to discover patterns in the stored data to the human and social issues surrounding knowledge creation in organizations.

Managers should provide the resources and the conditions for the emergence of rich learning environments in which data repositories and data mining tools sustain collective cognitive processes such as memory, reasoning, language and attention. In this way, ODM becomes a key organizational process in the construction of organizational representations of external realities. These representations will guide organizational decision and action. In accordance with this view, this article provides a summary of constructionist guidelines for ODM to help managers leveraging knowledge creation in organizations.

REFERENCES

Bolman, L.G., & Deal, T.E. (1991). *Reframing organizations: Artistry, choice, and leadership*. San Francisco: Jossey-Bass Publishers.

Damásio, A. (1999). *The feeling of what happens: Body and emotion in the making of consciousness*. New York: Harcourt Brace.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery: An overview. In U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, & R. Uthurusamy (Eds.), *Advances in knowledge discovery and data mining* (pp.1-34). MA: The MIT Press.

Kafai, Y., & Resnick, M. (Eds.). (1996). *Constructionism in practice: Designing, thinking, and learning in a digital world*. Mahwah, NJ: Lawrence Erlbaum Associates.

Lincoln, Y.S., & Guba, E.G. (2000). Paradigmatic controversies, contradictions, and emerging confluences. In N.K. Denzin & Y.S. Lincoln (Eds.), *Handbook of qualitative research* (pp.163-188). Sage Publications.

Matheus, C.J., Chan, P.K., & Piatetsky-Shapiro, G. (1993). Systems for knowledge discovery in databases. *IEEE Transactions on Knowledge and Data Engineering, 5*(6), 903-913.

Morgan, G. (1997). *Images of organization*. SAGE Publications.

Palmer, I., & Hardy, C. (2000). *Thinking about management.* SAGE.

Schwandt, T.A. (2000). Three epistemological stances for qualitative inquiry: Interpretivism, hermeneutics, and social constructionism. In N.K. Denzin & Y.S. Lincoln (Eds.), *Handbook of qualitative research* (pp.189-213). Sage Publications.

Shaw, A. (1996). Social constructionism and the inner city: Designing environments for social development and urban renewal. In Y. Kafai & M. Resnick (Eds.), *Constructionism in practice* (pp.175-206). Lawrence Erlbaum Associates Publishers.

Sparrow, J. (1998). *Knowledge in organizations: Access thinking at work.* London: Sage Publications.

Wallace, B. A. (2000). *The Taboo of Subjectivity: Toward a new science of consciousness.* New York: Oxford University Press.

KEY TERMS

Constructionism: A set of theories that defines the human beings as active constructors of their own learning and development. This learning and development of knowledge happens more effectively when individuals are involved in the construction of something external, something that can be shared, or both.

Objective Social Reality: It has an independent existence from any account of it.

Objectivism: A set of theories that views true knowledge about external realities, and the process of its creation, as neutral and independent of the knowledge creator.

Rich Learning Environment: Learning environments in which the learner is empowered to create a strong connection with the reality of interest by directly experiencing with it in order to develop mental constructs that are deep, complex, pluralist, and emotionally rich.

Social Constructions: External and sharable concepts, associations, artifacts, and practices that people actively develop and maintain in their social settings. An organization is an example of a social construction that interconnects its members in a specific social setting, in which many other social constructions are continuously being developed and maintained.

Socially Constructed Reality: It is created through purposeful human action and interaction. This reality is shaped by the individual's subjective conceptual structures of meaning. It is reconstructed by the human interactions that support continuous reinterpretations and change of meanings. The social institutions are the means through which meanings are stabilized and the social reality assumes an objective appearance.

Contemporary IT-Assisted Retail Management

Herbert Kotzab
Copenhagen Business School, Denmark

INTRODUCTION

Retailing can be defined as a set of specific business processes that add value to products and services sold to end users (e.g., Levy & Weitz, 2003). Such retail business processes refer to *marketing processes* (like assortment in order to provide end users with a required mix of products and services in terms of quantity as well as quality; and advice, advertising, and credit processes, which ease the purchase for end users, for example, offering special payment modes or informing end users about the offers) and *logistics processes* (such as transportation, breaking bulk, and inventory handling). The orchestration of these functions leads to various types of retail formats such as store-based retailers (e.g., supermarkets, hypermarkets, or category killers), non-store-based retailers (e.g., mail-order retailing or electronic commerce), and hybrid retailers (e.g., home delivery services or party sales services; Coughlan, Anderson, Stern, & El-Ansary, 2001).

Although retailing plays a vital role in today's economy, retailing companies also face economic pressure as many players operate mainly in mature and stagnant markets (e.g., Seth & Randall, 2001). In order to face these specific challenges, retailing companies adapted strategies that allow them to gain economies of scale by offering highly customized solutions to their customers (see Table 1).

These strategies are called IT-assisted retail management strategies as they are built upon the latest developments in information technology (IT). The following chapter presents an overview of contemporary IT-based retail business models and frameworks that show how IT has created a new mandate for retail management. IT is defined here as the hard- and software that collects, transmits, processes, and circulates pictorial, vocal, textual, and numerical data and information (e.g., Hansen & Neumann, 2001; Chaffey, 2004). The following are technologies of special interest used in IT-assisted retail management.

- Mobile data-capturing terminals, light pens, bar-code readers, bar-code labels, disks, chip cards, RFID (radio frequence identification), sensors to collect information
- Database systems, tapes, CDs, DVDs, optical disks, document-retrieval systems to store information
- PCs, information retrieval, decision support systems, expert systems (MIS, EIS, MSS, ESS to process information)
- Services (e.g., fax, e-mail, EDI [electronic data interchange], Web-EDI, FTP, WAIS, World Wide Web, SMTP, TCP/IP), networks (videoconferencing, teleconferencing, voice mail, ISDN, LAN, WAN, fiber optic, and intra-, inter-, and extranet), devices (e.g., phones, TV, radio, fax machine, PC, PDA [personal digital assistant]) to transmit information

The increasing use of these technological possibilities leads to major changes in the strategic management of distribution channels as the layers are compressed and the distances between the first and last echelon of the channel are reduced (e.g., Coughlan et al., 2001; Porter, 2001). Leading retailers are aware of these possibilities and implement customized POS-data-based marketing strategies (IT-based retail marketing) and demand syn-

Table 1. Cornerstones of contemporary IT-based retail management (see Kotzab, Schnedlitz, & Neumayer, 2003)

IT-based retail marketing strategies	IT-based retail logistics systems
• Reengineered, IT-driven retail formats allowing a customized shopping experience • Development of new retail channels, e.g., Internet-based retail formats to address new customer segments • Category management in order to offer client-oriented sets of products, resulting from a joint-planning process with manufacturers based on real-time accessed client data	• The implementation of just-in-time-oriented replenishment systems by connecting the electronic point-of-sale (EPOS) systems with the manufacturers' ERP systems • The execution of IT-driven distribution-center operations with no-inventory-holding transit terminal structures • The realization of vendor-managed inventory programs on a continuous–replenishment basis to reduce inventory levels and to improve order cycles

chronized replenishment systems (IT-based retail logistics).

BACKGROUND

IT-Based Retail Marketing Processes

Business practice shows a huge variety of IT-based retailing marketing strategies that includes the use of smart cards, theft prevention, self-checkout systems, Web kiosks, and/or merchandise-planning systems. The common goal of all strategies is to obtain better information on consumer behavior in order to improve customer service. IT-based retail marketing affects in that sense all retail areas from the sales floor to the back offices (Kotzab, Schnedlitz, et al., 2003).

Referring to store-based retailing, IT influences the layout and the atmosphere of a store by optimizing the link between sales productivity and consumer excitement (Nymphenburg, 2001). Metro introduced recently its future store concept that promotes technologically driven innovations in IT-assisted retail marketing as offered by the combined use of RFID, electronic shelf labels, self-checkout systems, personal shopping agents, in-store media such as info terminals, loyalty cards, personal shopping assistants for shoppers, personal digital assistants for employees, and intelligent scales (e.g., Metro, 2003; NCR, 2003; Ody, 2002).

Another prototype of a modern supermarket can be found in Austria (see Kotzab, Schnedlitz, et al., 2003), where Rewe Austria operates in Purkersdorf (nearby Vienna) an outlet where shoppers self-register their purchases via self-scanning devices. Rewe also uses the "talking" or "communicating" shopping cart WATSON, whose technology is based on radio frequency. Whenever passing a labeled shelf, the cart announces a message to the shopper (Atlas New Media, 2001).

Other recent examples are presented by Carter and Lomas (2003), who refer to a Sainsbury store in Hazelgrove (UK) and a Darty store in France that represent the state of the art of technology-driven store layout.

IT also changed the organizational setup from hierarchical to hybrid or borderless arrangements such as category management (CM; Gruen, 2002). CM is a joint retailer and manufacturer effort that involves managing product categories as business units and customizing them on a store-by-store basis to satisfy end-user needs (Dussart, 1998). The purpose is to identify those combinations of products that make up consumers' expectations. CM replaces traditional product-focused strategies (e.g., brand management) and allows retailers and suppliers to faster react to shifts in the marketplace (Schröder, 2003). The increasing use of data-warehousing and data-mining approaches helps to use the scanner data more efficiently in order to establish customer-oriented assortments (Chen et al., 2000).

The RFID technology (e.g., AIMGLOBAL, 2004; Öztürk, 2003) especially seems to change retailing marketing but also logistics dramatically. The Exxon Speedpass program might be the new way of cashless payment as immediate authorization is possible and it reduces the waiting time enormously (Speedpass, 2003). Metro's future store concept also shows that self-scanning processes can also be replaced by RFID. Wal-Mart, however, intends to implement RFID with about 100 vendors in order to improve the replenishment processes (Ashcroft, 2003).

IT-Based Retail Logistics Processes

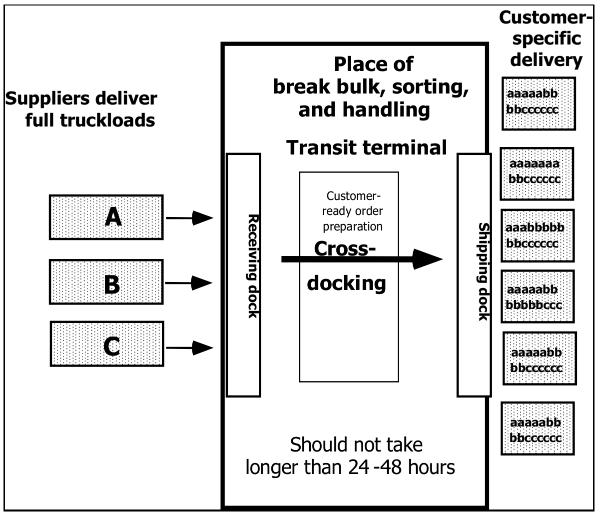
Logistics in a retailing context refers to multiechelon logistics systems with many nodes from the original supplier to the final store destination (Kotzab, 2002). The use of specific IT in retail logistics, such as EDI (e.g., EANCOM, EDIFACT, ANSI X.12), bar codes (e.g., EAN/UCC) and scanner technology, converted traditional retail logistics systems into just-in-time-oriented retail supply chain management systems. A channel-wide use of technology allows harmonization and synchronization of logistics operations between retailers and their suppliers, and give retailers additional profitability as such systems operate on a pull instead of a push base. Consequently, the total bullwhip effect in such channels is reduced (Lee & Whang, 2002).

The major IT-assisted retail logistics processes are cross-docking and continuous replenishment (Kotzab, Schnedlitz et al., 2003), and recently collaborative planning forecasting and replenishment (CPFR; Skjoett-Larsen, Thernoe, & Andresen, 2003).

Cross-docking is the metaterm for all IT-related flow-through activities within a distribution center that provide tailor-made deliveries on a just-in-time basis. Different vendors deliver full truckloads of their goods to a retailer's transit terminal (a reengineered distribution center; also called a transshipment point). The goods are there consolidated and/or broken to vendor-integrated, POS-required smaller delivery units (see Figure 1).

Cross-docking's basic idea is to avoid inventory at the distribution-center level, which leads to a lot of replacement of stock-holding activities through sorting, transportation, and handling activities, which are controlled by increased use of IT (e.g., EAN/UCC 128 in combination with EANCOM messages). The relevant literature offers various types of cross-docking operations depending whether vendors deliver prelabeled units

Figure 1. Basic cross-docking operation (Kotzab, 1997)



to the distribution center (complex cross-docking), full pallets or cases (full pallet cross-docking versus break-bulk cross-docking), or customized mixed or nonmixed pallets (e.g., Napolitano, 2000).

While cross-docking refers to IT-assisted logistics on a distribution-center level, vendor-managed inventory (VMI) or continuous replenishment (CRP) refer to all cooperative forms of an interfirm, automated replenishment program where the common goal is the automatic reinforcement of the supply of goods and the transfer of the burden of responsibility of storage from a retailer to a vendor.

Within any VMI/CRP setting, retailers retransfer the inventory competence back to their vendors by agreeing on certain average inventory levels on the distribution-center level, service levels, and/or other arrangements like the reduction or avoidance of out-of-stock situations (Raman, DeHoratius, & Zeynep, 2001). Within VMI/CRP, the former one-to-one relationship (where a seller and a buyer individually represented the goals of their companies) is replaced by interdepartmental, interorganizational teams, which are responsible for the ordering process (Waller, Johnson, & Davis, 1999).

A further development in IT-based retail logistics can be seen in the use of CPFR that is defined by the Voluntary Interindustry Commerce Standards (VICS) as "a collection of new business practices that leverage the Internet and electronic data interchange in order to radically reduce inventories and expenses while improving customer service" (VICS, 1998b). This definition suggests that the Internet and electronic data interchangeare substantial prerequisites of CPFR.

VICS also crafted a guideline in the form of a nine-step model detailing how to implement CPFR (VICS 1998a; see Table 2).

As Fliedner (2003, p. 15) points out, these steps can be seen as a "cyclic and iterative approach to derive consensus supply chain forecasts." The CPFR process builds to a large extent on the exchange of information among collaboration partners. This exchange of information can be carried out through the use of various technologies such as EDI, private networks, or the Internet (XML [extended markup language]). For the processing of information, a large number of software programs have

Figure 2. Basic continuous replenishment process (Glavanovits & Kotzab, 2002)

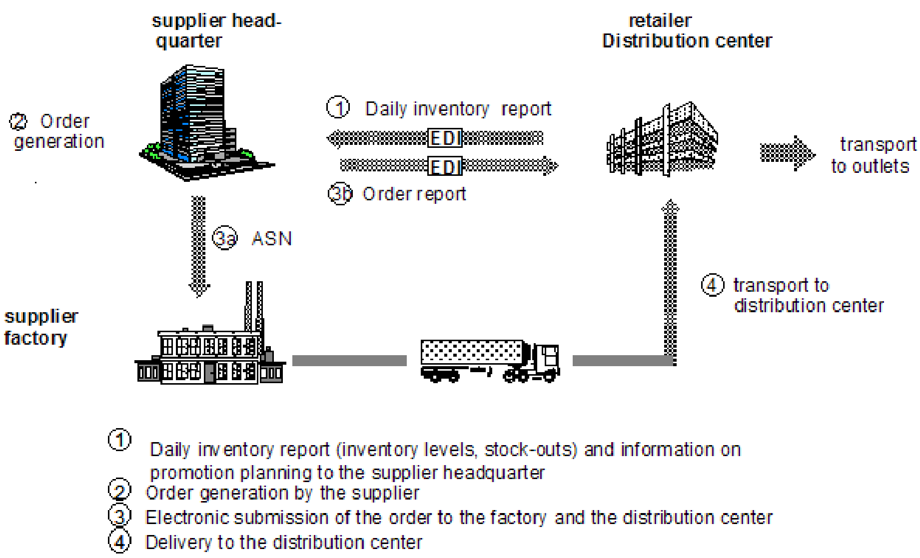


Table 2. CPFR implementation guideline (see Kotzab, Skjoett-Larsen, Andresen, & Thernoe, 2003)

Step	Activity	Description
1	Develop front-end agreement	A front-end agreement is developed. Criteria for success are established. Identification of the CPFR project owners in the companies. Financial reward and contribution system is agreed upon.
2	Create joint business plan	A joint business plan for the areas of collaboration is created. (plans regarding advertising campaigns, etc.)
3-5	Sales-forecast collaboration	The parties get together with each of their customer-demand prognoses to establish a common prognosis. In case of deviation from forecast, the partners meet to discuss deviations and to update the common forecast.
6-8	Order-forecast collaboration	The partners share replenishment plans and discuss deviations and constraints.
9	Order generation	The reordering process/goods flow is initiated. Result data is discussed (POS, orders, shipments). Forecast deviation and stock-level problems are identified and solved.

been developed to support the CPFR processes (e.g., Syncra, Logility, Manugistics; i2 Technologies, E-Millennium, E3, J.D. Edwards, Numetrix og Excentric, SAP, Oracle, Baan, or Peoplesoft; see Kotzab, Skjoett-Larsen, et al., 2003).

FUTURE TRENDS

A major revolution in retailing can be seen in Web-based retailing (e-tailing) as offered by pure virtual retailers (dot-coms, e.g., Amazon.com) or multichannel retailers ("bricks and clicks," e.g., Wal-Mart with Wal-mart.com). The business challenge of e-tailing can be seen rather in logistics than in order generating as the most prominent example of Webvan.com has shown. The concept of home delivery receives due to these developments a renaissance, for example, known as consumer direct services (Corbae & Balchandani, 2001). Experts estimate the total home delivery market in the grocery industry to be an over-100-billion-Euro business, including over 900 million deliveries per year to over 35 million households.

IT-assisted home delivery concepts try to optimize the so-called last mile to the consumer, which is the most expensive part of the total supply chain. There are plenty of home delivery systems, which either deliver the products directly to the homes or where customers collect their purchased goods at certain pickup points (e.g., Pflaum, Kille, Mirko, & Prockl, 2000, Punikavi, Yrjölä, & Holmström, 2001, Siebel, 2000). Prominent examples for pickup-point solutions are the Tower 24 concept or the pick-point concept (Burtscher & Köppel, 2002).

CONCLUSION

The chapter described the consequences of the use of IT in the field of retailing. It was shown how IT can be used in retail marketing to reengineer retail formats in order to allow a customized shopping experience. IT-assisted retail logistics refers to just-in-time-oriented demand-synchronized delivery systems. The increased use of such systems will lead in the future to more hybrid retail management strategies, where the organizational borders between retailers, suppliers, and even consumers will disappear.

REFERENCES

AIMGLOBAL. (2004). *What is radio frequency identification (RFID)?* Retrieved March 10, 2004, from *http://www.aimglobal.org/technologies/rfid/what_is_rfid.asp*

Ashcroft, J. (2003). *Wal-Mart announces RFID plans.* Retrieved October 10, 2003, from *http://logistics.about.com/cs/logisticssystems/a/aa060303.htm*

Atlas New Media. (2001). Watson, der sprechende einkaufswagen [Watson, the talking shopping cart]. Hamburg, Germany.

Burtscher, G., & Köppel, M. (2002). E-commerce: Treffen am pick point? [E-commerce: Meeting at the pick point?] *Austropack, 12,* 29-31.

Carter, D., & Lomas, I. (2003). *Store of the future*. Presentation at the Eighth Official ECR-Europe Conference, Berlin, Germany.

Chaffey, D. (2004). *E-business and e-commerce management* (2nd ed.). Prentice Hall, Financial Times.

Corbae, G., & Balchandani, A. (2001). Consumer direct Europa: Erfolg durch zusammenarbeit [Consumer direct Europe: Success by cooperation]. In D. Ahlert, J. Becker, P. Kenning, & R. Schütte (Eds.), *Internet & co. im handel: Strategien, geschäftsmodelle, erfahrungen* [Internet & co. in retail: Strategies, business models and experiences] (pp. 63-78). Berlin-Heidelberg, Germany: Springer.

Coughlan, A., Anderson, E., Stern, L., & El-Ansary, A. (2001). *Marketing channels* (6th ed.). Upper Saddle River, NJ: Prentice Hall.

Dussart, C. (1998). Category management: Strengths, limits and developments. *European Management Journal, 16*(1), 50-62.

Fliedner, G. (2003). CPFR: An emerging supply chain tool. *Industrial Management and Data Systems, 103*(1), 14-21.

Glavanovits, H., & Kotzab, H. (2002). *ECR kompakt* [ECR compact]. Wien, Austria: EAN-Austria.

Gruen, T. (2002). The evolution of category management. *ECR Journal, International Commerce Review, 2*(1), 17-25.

Hansen, H.-R., & Neumann, G. (2001). *Wirtschaftsinformatik* [Computer science in business and management] (8th ed.). Stuttgart, Germany: UTB.

Kotzab, H. (1997). *Neue konzepte der distributionslogistik von handelsunternehmen* [Recent concepts for retail logistics]. Wiesbaden, Germany: Deutscher Universtitätsverlag.

Kotzab, H., Schnedlitz, P., & Neumayer, K. (2003). Contemporary IT-assisted retail management. In L. Joia (Ed.), *IT-based management: Challenges and solutions* (pp. 175-203). Hershey, PA: Idea Group Publishing.

Kotzab, H., Skjoett-Larsen, T., Andresen, C., & Thernoe, C. (2003). Logistics managers' perception and viewpoint to interorganizational supply chain collaboration by CPFR. In T. Spengler, S. Voss, & H. Kopfer (Eds.), *Logistik management: Prozesse–systeme–ausbildung* [Logistics management: Processes–systems–education] (pp. 65-78). Berlin, Germany: Springer.

Lee, H., & Whang, S. (2001). Demand chain excellence: A tale of two retailers. *Supply Chain Management Review, 5*(2), 40-46.

Levy, M., & Weitz, B. (2003). *Retailing management* (5th ed.). Boston: McGrawHill-Irwin.

Metro. (2003). Retrieved May 25, 2003, from http://www.future-store.org

Napolitano, M. (2000). *Making the move to crossdocking*. Warehousing Education and Research Council.

NCR. (2003). *NCR store automation solutions*. Retrieved October 10, 2003, from http://www.ncr.com/solutions/store_automation/storeauto.htm

Nymphenburg. (2001). *Elektronische medien erobern den handel* [Electronic media conquer retail]. Retrieved August 14, 2001, from *http://www.nymphenburg.de*

Ody, P. (2002). *A lingering goodbye to the check out*. Retrieved October 10, 2003, from *http://specials.ft.com/ftit/april2002/FT3T7ST0HZC.html*

Öztürk, M. (2003). *RFID und logistik* [RFID and logistics]. Retrieved December 15, 2003, from *http://www.competence-site.de/pps.nsf/0/b1d6a9e977b62247c1256dcd004b7ae8?OpenDocument*

Pflaum, A., Kille, C., Mirko, W., & Prockl, G. (2000). *Heimlieferdienste für lebensmittel und konsumgüter des täglichen bedarfs im Internet: Die letzte meile zum kunden aus der logistischen perspektive* [Home delivery services for groceries and consumer goods for every day usage in the Internet: The last mile to the consumer from a logistical perspective].

Porter, M. (2001). Strategy and the Internet. *Harvard Business Review, 79*(3), 63-78.

Punakivi, M., Yrjölä, H., & Holmström, J. (2001). Solving the last mile issue: Reception box or delivery box? *International Journal of Physical Distribution & Logistics Management, 31*(6), 427-439.

Raman, A., DeHoratius, N., & Zeynep, T. (2001). Execution: The missing link in retail operations. *California Management Review, 43*(3), 136-152.

Schröder, H. (2003). Category management: Eine standortbestimmung [Category management: A positioning]. In H. Schröder (Ed.), *Category management: Aus der praxis für die praxis-konzepte-kooperationen–erfahrungen* [Category management: From business practice for business practice–concepts–cooperation–experiences] (pp. 11-38). Frankfurt am Main, Germany: LebensmittelZeitung.

Seth, A., & Randall, G. (2001). *The grocers* (2nd ed.). London/Dover: Kogan Page.

Siebel, L. (2000). *Food logistics: Lebensmittel via Internet* [Food logistics: Groceries via Internet]. Düsseldorf.

Skjoett-Larsen, T., Thernoe, C., & Andresen, C. (2003). Supply chain collaboration. *International Journal of Physical Distribution and Logistics Management, 33*(6), 531-549.

Speedpass. (2003). Retrieved December 9, 2003, from *http://www.speedpass.com/home.jsp*

Voluntary Interindustry Commerce Standards (VICS). (1998a). Collaborative planning forecasting and replenishment. *Voluntary guidelines.* Retrieved from http://*www.cpfr.org*

Voluntary Interindustry Commerce Standards (VICS). (1998b). *VICS helps trading partners collaborate to reduce uncertainty with CPFR guidelines.* Retrieved from *http://www.cpfr.org/19981008.html*

Waller, M., Johnson, M., & Davis, T. (1999). Vendor-managed inventory in the retail supply chain. *Journal of Business Logistics, 20*(1), 183-203.

KEY TERMS

Bar Codes: Simple form of optical character recognition, where information is encoded in printed bars of relative thickness and spacing. RFID combines this technology with radio frequency.

Electronic Data Interchange (EDI): Metaterm for a multitude of different electronic message standards that allow a computerized and highly structured, low-error communication between computers. A "merge" between EDI and Internet technology can be recently observed by the upcoming of Web-based EDI solutions, where on EDI partner does not have to install EDI, but uses common Web browsers to communicate via EDI.

Electronic Shelf Labels: Price tags that provide accurate pricing due to electronic linkage between the shelves and the checkout system. The technology is based on radio frequency, infra-red, and/or WLAN linking a master checkout with the shelves.

Intelligent Scale: A scale that is equipped with a special camera and identification software. Based on an object's structure, size, color, and thermal image, the scale automatically recognizes the item, weighs it, and prints out a price tag.

RFID (Radio Frequency Identification): Form of automated-radio-frequency-based identification of objects. RFID systems consist of an antenna, a transceiver for reading radio frequency and to transfer information, a processing device, and a transponder.

Scanner: Electronic devices that convert bar-code information into digitized electronic images.

Self-Checkout Systems: Self-checkout systems can occur at the end or during shopping processes whenever cash-desk operations are "outsourced" to consumers. In that case, consumers self-register their items with specific scanning devices.

Virtual Shopping Assistant: Small mobile computer with a touch screen and bar-code scanner that can be installed to a shopping trolley and can serve as a personal shopping advisor (e.g., offering a customized shopping list).

Content Description for Face Animation

Ali Arya
Simon Fraser University, Canada

INTRODUCTION

Face animation is a challenging area of computer graphics and multimedia systems research (Parke, 1996). Realistic personalized face animation is the basis for virtual software agents that can be used in many applications, including video conferencing, online training and customer service, visual effects in movies, and interactive games. A software agent can play the role of a trainer, a corporate representative, a specific person in an interactive virtual world, and even a virtual actor. Using this technology, movie producers can create new scenes including people who are not physically available. Furthermore, communication systems can represent a caller without any need to transmit high volume multimedia data over limited bandwidth lines. Adding intelligence to these agents makes them ideal for interactive applications such as online games and customer service. In general, the ability to generate new and realistic multimedia data for a specific character is of particular importance in cases where prerecorded footage is unavailable, difficult, or expensive to generate, or simply too limited due to the interactive nature of the application.

Personalized face animation includes all of the algorithms, data, modules, and activities required to create a multimedia presentation resembling a specific person. The input to such a system can be a combination of audio-visual data and textual commands and descriptions. A successful face animation system needs to have efficient yet powerful solutions for providing and displaying content, for example, a content description format, decoding algorithms, creating required content, and finally, an architecture to put different components together in a flexible way. Multimedia modeling and specification languages play a major role in content description for face animation systems both as general tools and in customized forms. Such descriptions can be used for authoring and also as run-time input to the animation system.

BACKGROUND

The diverse set of works in multimedia content description involves methods for describing the components of a multimedia presentation and their spatial and temporal relations. Historically, some of the first technical achievements in this area were related to video editing where temporal positioning of video elements is necessary. The SMPTE (Society of Motion Picture and Television Engineers) time coding that precisely specifies the location of audio/video events down to the frame level is the base for EDL (Edit Decision List) which relates pieces of recorded audio/video for editing (Ankeney, 1995; Little, 1994). Electronic program guide (EPG) is another example of content description for movies in the form of textual information added to the multimedia stream. More recent efforts by SMPTE are focused on a metadata dictionary that targets the definition of metadata description of content (see *http://www.smpte-ra.org/mdd*).

Motion Picture Expert Group (MPEG) is another major player in the field of standards for multimedia content description and delivery. MPEG-4 standard (Battista et al., 1999), introduced after MPEG-1 and MPEG-2, is one of the first comprehensive attempts to define the multimedia stream in terms of its forming components (objects such as audio, foreground figure, and background image). Users of MPEG-4 systems can use object content information (OCI) to send textual information about these objects. MPEG-7 standard, mainly motivated by the need for a better and more powerful search mechanism for multimedia content over the Internet, can also be used in a variety of other applications including multimedia authoring. The standard extends OCI and consists of a set of descriptors for multimedia features (similar to metadata in other works), schemes that show the structure of the descriptors, and a description/schema definition language based on eXtensible Markup Language (XML).

Most of these methods are not aimed at, and customized for, a certain type of multimedia stream or object. This may result in a wider range of applications but limits the capabilities for some frequently used subjects such as the human face. The study of facial movements and expressions started from a biological point of view by scientists such as John Bulwer and Charles Darwin (see Ekman & Friesen, 1978). More recently, one of the most important attempts to describe facial activities (movements) was the Facial Action Coding System (FACS). Introduced by Ekman and Friesen (1978), FACS defines 64 basic facial action units (AUs), such as raising brows, talking, and turning left and right. It should be noted though, that FACS does not provide any higher-level construct to describe the sequence of actions and spatial and temporal relations between facial activities. In other words, it is not designed to be a face animation description language.

MPEG-4 standard uses an approach similar to FACS to integrate face animation into multimedia communication, by introducing face definition parameters (FDPs) and face animation parameters (FAPs). FDPs define a face by giving coordinates and other information for its major feature points such as eyes' and lips' corners. They allow personalization of a generic face model to a particular face, and are more suitable for synthetic faces. FAPs, on the other hand, encode the movements of these facial features. There are more than 70 FAPs defined similar to FACS AUs.

Although MPEG-4 defines two sets of higher-level codes, that is, visemes and expressions, compared to low-level FACS AUs, but it still has only a set of animation commands and not an animation language. Synchronized Multimedia Integration Language (SMIL), an XML-based language, is designed to specify temporal relation of the components of a multimedia presentation, especially in Web applications (Bulterman, 2001).

There have also been different languages in the fields of virtual reality and computer graphics for modeling computer-generated scenes. Examples are Virtual Reality Modeling Language (VRML, *http://www.vrml.org*), its XML-based version known as X3D, and programming libraries like OpenGL (*http://www.opengl.org*). MPEG-4 standard includes Extensible MPEG-4 Textual format (XMT) framework to represent scene description in a textual format providing interoperability with languages such as SMIL and VRML.

None of these languages are customized for face animation, and they do not provide any explicit support for it, either. Recent advances in developing and using embodied conversational agents (ECAs), especially their Web-based applications, and growing acceptance of XML as a data representation language have drawn attention to markup languages for virtual characters. The basic idea is to define specific XML constructs related to agents' actions such as moving and talking. Virtual Human Markup Language (VHML, *http://www.vhml.org*) is an example in this regard (Marriott & Stallo, 2002). It comprises a number of special purpose languages, such as Emotion Markup Language (EML), Facial Animation Markup Language (FAML), and Body Animation Markup Language (BAML). A simple VHML document looks like this:

- <vhml>
  <person disposition=”angry”>
  <p>
  First I speak with an angry voice and look very angry,
  <surprised intensity=”50”>
  but suddenly I change to look more surprised.
  </surprised>
  </p>
  </person>
  </vhml>

Multimodal Presentation Markup Language (MPML, *http://www.miv.t.u-tokyo.ac.jp/MPML/en*) is another XML-based markup language developed to enable the description of multimodal presentation on the WWW, based on animated characters (Prendinger et al., 2002). MPML addresses the interactivity and decision-making not directly covered by VHML, but both suffer from a lack of explicit compatibility with MPEG-4 (XMT, FAPs, etc.).

Another important group of related works are behavioural modeling languages and tools for virtual agents. Behaviour Expression Animation Toolkit (BEAT) is an XML-based system, specifically designed for human animation purposes (Cassell et al., 2001). It is a toolkit for automatically suggesting expressions and gestures, based on a given text to be spoken. BEAT uses a knowledge base and a rule set, and provides synchronization data for facial activities, all in XML format. Although BEAT is not a general content description tool, it demonstrates some of the advantages of XML-based approaches together with the power of behavioural modeling.

Another group of researchers have proposed the concept of cognitive modeling for character animation (Funge et al., 1999). Their system is based on a set of geometric, behavioural, and cognitive models for the characters to be animated. In this approach not only the physical capabilities but also the behavioural and cognitive processes are defined and modeled. A special Cognitive Modeling Language (CML) is also developed to support this system. CML does not provide any explicit support for face animation and, unlike BEAT, is not XML-based. Neither is Parameterized Action Representation (PAR), another language proposed to describe and model actions of an agent, based on interaction with environment and the personality and emotions (Allbeck & Badler, 2002). In PAR, the agent personality is defined in terms of parameters such as openness, agreeableness, and extraversion. Similar parameters are defined for other aspects affecting the behaviour.

FACE MODELING LANGUAGE (FML)

Face Modeling Language (FML) is part of the ShowFace Personalized Face Animation Framework (Arya & Hamidzadeh, 2002). FML (*http://www.raminsoftworx.com/Research/fml_1.html*) is an XML-based language to provide structure content description, that is, a hierarchical description of animation events from low-level movements such as MPEG-4 FAPs to high-level stories. A sample FML document is shown here:

```
<fml>
 <model>    <!— Model Info —>
   <img src=”me.jpg” />
   <range dir=”0” val=”60” />
   <event id=”user” val=”-1” />
   <template name=”hello” >
       <seq begin=”0”>
           <talk begin=”0”>Hello</talk>
           <hdmv begin=”0” end=”5” dir=”0” val=”30” />
       </seq>
   </template>
 </model>
 <story>    <!— Story Timeline —>
   <action>
       <behavior template=”hello” />
       <excl ev_id=”user”>
           <talk ev_val=”0”>Hi</talk>
           <talk ev_val=”1”>Bye</talk>
       </excl>
       <par begin=”0”>
           <talk begin=”1”>Hello World</talk>
           <exp begin=”0” end=”3” type=”3” val=”50” />
       </par>
   </action>
 </story>
</fml>
```

Each FML module has two main parts: **model** and **story**. The former defines initialization, movement ranges, and behavioural templates. The latter describes the animation sequence in a hierarchical way. Each story consists of **actions** (independent meaningful sets of moves). At the lower level, there are time containers: **par**, **seq**, and **excl**. The first two construct parallel and sequential moves. **excl** is used for decision-making, that is, waiting for an event such as user input and taking different paths based on its value. The XML-based nature of this language allows the FML documents to be embedded in Web pages. Normal XML parsers can extract data and use them as input to an FML-enabled player through simple scripting. Such a script can also use an XML Document Object Model (DOM) to modify the FML document, for example, adding certain activities based on user input. This compatibility with Web browsing environments gives another level of interactivity and dynamic operation to FML-based systems.

Another aspect of FML is its compatibility with MPEG-4 face definition/animation parameters. This has been achieved by:

- Translation of FML documents to MPEG-4 codes by the media player;
- Embedded MPEG-4 elements (fap element is considered to allow direct embedding of FAPs in FML document); and
- Integrating FML into XMT framework by FML-to-MPEG converters.

FUTURE TRENDS

Rapid growth of “face-based” applications requires a more systematic and comprehensive approach to content description for facial animation. Some future trends in this regard can be:

- Behavioural modeling, that is, addition of language constructs that model and describe personality, mood, and behavioural rules;
- Better integration with MPEG-4 standard, for example, translation tools for dedicated face animation languages to/from MPEG-4 stream format; and
- Authoring tools, that is, user-friendly GUI environments to create scripts.

CONCLUSION

Content description, as a major issue in multimedia systems, is briefly discussed. Considering face animation systems as a challenging example, different approaches are reviewed. It can be seen that although general-purpose modeling, specification languages and coding systems are powerful means, a dedicated language for face animation is required to provide a higher level of abstraction, ease of programming, and compatibility with existing multimedia standards and technologies.

After reviewing some other languages, XML-based Face Modeling Language (FML) is introduced as a good example. FML allows event handling, and also sequential or simultaneous combination of supported face states, and can be converted to a set of MPEG-4 face animation parameters. Main contributions of FML are its hierarchical structure, animation configuration and modeling, flexibility for static and dynamic scenarios, and dedication to face animation. FML fully supports MPEG-4 FAPs using high-level constructs which can be translated to FAPS and also direct FAP embedding. Compatibility with MPEG-4 XMT and use of XML as a base are also among the important features in the language. Future extensions to FML can include more complicated behaviour modeling and better integration with MPEG-4 streams.

REFERENCES

Allbeck, J., & Badler, N. (2002). Toward representing agent behaviours modified by personality and emotion. *Proceedings of First International Conf Autonomous Agents & Multi-Agent Systems, Workshop on Embodied Conversational Agents,* Bologna, Italy.

Ankeney, J. (1995, May). Non-linear editing comes of age. *TV Technology*.

Arya, A., & Hamidzadeh, B. (2002). An XML-based language for face modeling and animation. *Proceedings of IASTED Int. Conf. on Visualization, Imaging and Image Processing,* Barcelona, Spain.

Battista, S. et al. (1999). MPEG-4: A multimedia standard for the third millennium. *IEEE Multimedia*, *6*(4).

Bulterman, D. (2001). SMIL-2. *IEEE Multimedia, 8*(4).

Cassell, J. et al. (2001). BEAT: The behavior expression animation toolkit. *Proceedings of ACM SIGGRAPH*.

Ekman, P., & Friesen, W.V. (1978). *Facial action coding system*. Consulting Psychologists Press Inc.

Funge, J. et al. (1999). Cognitive modeling: Knowledge, reasoning, and planning for intelligent characters. *Proceedings of ACM SIGGRAPH*.

Little, T.D.C. (1994). Time-based media representation and delivery. In J.F. Koegel Buford (Ed.), *Multimedia systems*. ACM Press.

Marriott, A., & Stallo, J. (2002). VHML: Uncertainties and problems. A discussion. *Proceedings of First International Conference of Autonomous Agents & Multi-Agent Systems, Workshop on Embodied Conversational Agents,* Bologna, Italy.

Parke, F.I., & Waters, K. (1996). *Computer facial animation*. A.K. Peters.

Prendinger, H. et al. (2002). Scripting affective communication with life-like characters in Web-based interaction systems. *Applied Artificial Intelligence, 16*(7-8).

KEY TERMS

Agent: A virtual representation of real or imaginary human beings in software systems.

Bandwidth: Maximum data transfer capacity of a communication channel.

Cognitive Modeling: Algorithms for representing and simulating the cognitive process, for example, to add intelligence to software agents.

Document Object Model (DOM): An object-oriented model representing different components of a system and their interfaces. In XML, DOM allows working with language elements as interactive objects with data, methods, and events.

Extensible Markup Language (XML): A general-purpose data representation language similar to HyperText Markup Language (HTML) used in Web pages.

Face Model: General information required for recreating faces in a computer graphics system, for example, geometrical parameters.

Metadata: Information about the content as opposed to the data themselves.

Multimedia Stream: Progressively transmitted/received data corresponding to a multimedia presentation, as opposed to a file containing all the data.

Virtual Reality: Computer-generated environment that simulates three-dimensional reality.

Content Management in Organizations

Anne Honkaranta
University of Jyväskylä, Finland

Pasi Tyrväinen
University of Jyväskylä, Finland

INTRODUCTION

Content management may be characterized as "*a variety of tools and methods that are used together to collect, process, and deliver content of diverse types*" (McIntosh, 2000, p. 1). At least three differing approaches on content management may be identified: 1) Web content management, 2) Document management, and 3) Utilization of structured documents.

The Web has become a widely used, yet more complex platform for disseminating a portion of organizational content. Hence new means for creating, publishing and managing content on the Web have been sought (Grossniklaus & Norrie, 2002). A great deal of current research efforts (e.g., Boiko, 2002; Grossniklaus & Norrie, 2002; McIntosh, 2000; Weitzman et al., 2002) are targeted specifically for Web content management. The underlying approaches and conceptual base used can be traced to electronic publishing (e.g., Boiko, 2002), and to database-oriented approaches (e.g. McIntosh, 2000). Continuous adoption of new technologies and system versions (Balasuramanian & Bashan, 1998), and technology-driven development (Weitzman et al., 2002) are also characteristic to the Web content management approach.

Documents as identifiable units of content, flexibly structured for human comprehension (Murphy, 1998; Salminen, 2000), have traditionally been considered as containers for organizational content. Research on document management in organizations has included a multitude of focuses, from document standardization (Salminen, 2000), document metadata (Murphy, 1998), and requirements elicitation (Lyytikäinen, 2003), to document and information retrieval (Blair, 2002). Approaches and methods, such as the Stair Step method (Koulopoulos & Frappaolo, 1995) and the RASKE methodology (e.g., Salminen, 2000) have been developed. The wide selection of document management systems available (Doculabs, 1998) includes systems that are evolving towards content management (Medina, Meyers, Bragg, & Klima, 2002). The essential features of document management systems cover:

- library services and version management,
- management of user roles and access rights,
- support for document life-cycle and related workflows,
- management of *metadata*, as information about documents, and
- multichannel publishing for a multitude of devices and print.

In *structured documents*, such as documents defined by Extensible Markup Language (XML; Bray, Paoli, Sperberg-McQueen, Maler, & Yergeau, 2004), the content itself, the logical structure, and the visual layout for the content are separated from each other and made visible both for humans and computer software by using markup delimiters. The logical content is described as a hierarchy of named elements and associated attributes. A *Document Type Definition* (DTD) or an *XML schema* (Fallside, 2001) may be used to define the markup vocabulary and the structure for a class of XML documents. The research on structured documents has principally focused on document grammars and transformations (Lindén, 1997), assembly (Heikkinen, 2000), and schema design (Maler & El Andaloussi, 1996). XML standardization in organizations (Salminen, 2000), and Enterprise Application Integration (Linthicum, 2000) have also been studied. For content management, the structured documents provide means to access logical content units that are portions of contents within documents, and to separate the content from its layout properties.

BACKGROUND

Challenges on content management in organizations include finding the right focus, breadth and width for content management development with the limited resources available. A content management development project may cover the bulk of content used in an organization (e.g., Karjalainen, Päivärinta, Tyrväinen, & Rajala, 2000; Munkvold, Päivärinta, Hodne, & Stangeland, 2003), it may be focused only to a specific set of documents and related content (e.g., Honkaranta & Lyytikäinen, 2003; Honkaranta, Salminen & Peltola, 2005), or to a certain process, such as multichannel publishing (Boiko, 2002). In practice, a number of variations of content management focuses exist. An organization may need to combine research findings and technologies, and overcome the deficiencies of the diverse approaches developed for content management. The limitations of the approaches are discussed on the remainder of this section.

The Web content management approach focuses on the multichannel and Web content publishing. This approach suffers from immature systems and fast technical evolution, which make the approach laborious to adopt for organizations. The complexity on systems and technologies may consume a great deal of resources, causing that social aspects of content management may be overlooked. The conceptual base for the approach is inconsistent and immature (Grossniklaus & Norrie, 2002). Differing terms, such as object, component, content item, and content unit, may all be used to refer to the content. The phases of a content life-cycle may be overloaded with a myriad of concepts which are new to the people in an organization. For example, according to McIntosh (2000, p.1), a content life cycle consists of three main phases; 1) content assembly, 2) content production, and 3) content delivery, whereas Boiko (2002) utilizes concepts such as content acquisition, aggregation and metatorial processing. The lack of a concise conceptual base may hinder requirements elicitation and cause communicational breakdowns between system analysts and the people in an organization.

Applicability of document management approach and systems for content management have been limited by orientation towards managing content as document units only. For example, long documents are difficult to browse thorough, portions of document content may need to be reused in other documents, and long documents are inconvenient for Web delivery (e.g., Honkaranta, Salminen & Peltola, 2005). These factors have resulted in a trend towards use of smaller units of content than that of a document. Contemporary document management systems include limited capability to customize workflows for complex life-cycles of organizational content and for incorporating enhanced or multiple metadata standards. Document management software may have a limited functionality for design and management of a Web site of an organization. Despite the aforementioned deficiencies, document management approach and systems developed may provide a consistent base for content management development.

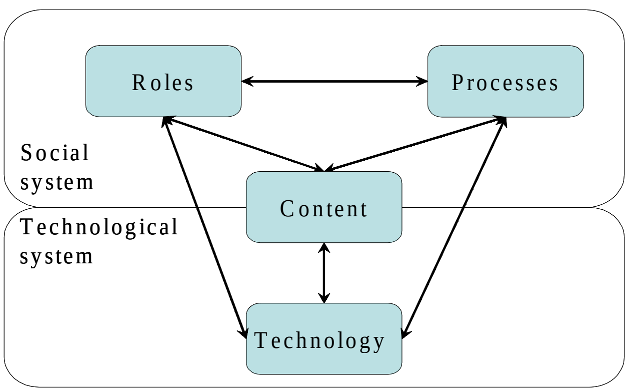
The benefits of the use of structured documents and related technologies have lately been recognized and adopted by both document management and Web content management approaches. Yet the use of XML poses multiple challenges. Even though there are systems and tools for managing XML documents as files or within databases (e.g., Bertino & Catia, 2001; Chaudhri, Rashid, & Zicari, 2003; Salminen & Tompa, 2001), the level of support for XML in document and content management systems is varying, and mostly missing from commonly used office software. In addition, developers lack knowledge about research findings and practical experience on adaptations and implementations of XML and related technologies in real-life organizations (e.g. Fahrenholz-Mann, 1999; Karjalainen & Tyrväinen, 2001; Salminen, 2000, Weitzman et al., 2002) since these are yet rarely reported.

ASPECTS AND CONCEPTS FOR DEVELOPING CONTENT MANAGEMENT IN ORGANIZATIONS

Content management in organizations is deeply intertwined with organizational work. Therefore development of content management should include the aspects of organizational work on which content is used and produced. We propose a framework for content management that includes both the technological and social system of an organization, allowing content management to be analyzed and evaluated from both aspects. The framework originates from the Leavitt's (1965) seminal framework for analyzing organizational change, and from contemporary research and empirical work. Figure 1 illustrates the framework while the rest of the chapter discusses the components of the framework and related concepts for content management development.

In the figure, the *social system* consists of processes and roles of people on the organization. Content is posi-

Figure 1. Framework for analyzing organizational content management



tioned between the social system and the *technological system*, the latter consisting of the technologies and the software used. The *content*, *i.e.* content units and metadata, must adhere to the operation of the social system with the technology available in the technical system of the organization. The components of the framework are interrelated. Whenever a change is implemented in one component, it will impact all other components of the framework. For example, adopting a Web content management system on the technology side will impact the publication processes, the roles of the people, and the content units used.

For content management, people's *roles* should be separated with regard to content management development vs. operational roles. Further, roles of content producers and users are needed both for content management *processes* and for operational processes, as they actually exist in the specific organization. For example, in new business models enhanced by digital rights management technology (Luoma & Tyrväinen, 2004; Rosenblatt, Trippe, & Mooney, 2002), content states related to content configurations, versions, revisions, and publicity of content should be taken into consideration both from the content management perspective and in the business process context.

Management of organizational *content* deals with complexity on identifying and designing logical, physical and processable units of content in organizations. The logical units of content are defined by social domain, and they have domain-oriented names; such as "Budget for Department of Computer Science and Information Systems". The *grain size* of logical content units may be varying, from content *aggregations* as collections of documents to documents, and to content units pertaining to a topical section of a document. For example, a budget document may consist of topical sections relating to different expense types. These topical sections may be used as independent units of content in one organizational process, and as an *assembly* which constitutes a document, or a Web page in another. Technical systems deal with physical units of content, such as folders and files, and with processable units of content, such as a field on a database or an XML element. Content management systems may use virtual folders for aggregating content units, and Web content management systems utilize *templates* for creating combinations of content and functionality for a Web site.

The amount of *metadata* related to the content increases while the size of the managed unit decreases. In overall, metadata is "*the sum total of what one can say about any information object at any level of aggregation*" (Gilliland-Swetland, 1998, p.1). Metadata may be used for defining properties of content unit instances and their mutual relationships, as well as for defining the organizational context, such as roles, processes, life-cycle states and audiences in which the content is used and produced. The metadata about organizational context is used for defining user roles and access rights for a system, as well as for workflow management. Documents may be independent content units providing a consistent whole for a human, while smaller content units need to be enriched with metadata for humans and especially for computer systems managing them. Specific purpose metadata may be used, e.g., for managing video and 3-dimensional content, and digital rights. Adoptation of a commonly used metadata standard, such as Dublin Core (1999), to an organization may include development of organization-specific metadata elements and attributes, defining optional and required values and pre-defined value lists, as well as tailoring the names of metadata elements.

The *technology* of organizational content management includes standards, architectures, technologies, tools, and applications. Contemporary content management systems may be roughly divided into three categories: platform systems, horizontal systems, and vertical systems. Relational and XML databases and other generic platforms provide a base-line onto which content management applications and functionalities may be built on. *Horizontal systems* are common-purpose document and content management systems, which already provide a base for managing content units as well as metadata and related processes and workflows. *Vertical systems* are systems that provide metadata and functionality related to management of specific content units or for a specific business processes. Examples of vertical systems are Customer Relationship Management (CRM), Supply Chain Management (SCM), and Product Data Management (PDM) systems, as well as patient information systems with specialized functionalities related their subject area.

FUTURE TRENDS

Content management combines approaches and findings from multiple historically separated approaches, such as document management, Web content management, and structured documents. These approaches have not yet been consolidated into a uniform research area. The conceptual base is inconsistent, and needs to be elaborated. Methods for content management need to be developed, and enhanced from existing ones.

Analysis and design of logical units of content in organizational settings is complex. The theory of genres as communicational, prototypical models for communication (Swales, 1999) provides a promising approach for identifying and analyzing logical content units on organizations. So forth, the genres have been operationalized for overcoming the shortcomings of the document concept while content units of an organization are identified (e.g., Karjalainen et al., 2000) and for analyzing the digitalization of organizational content (Tyrväinen & Päivärinta 2003), as an example.

CONCLUSION

Content management in organizations may consider the whole content of an organization, or just a specific portion of it. Content management combines approaches and findings from multiple historically separated approaches, such as document management, Web content management, and structured documents. These approaches have not yet been consolidated to a uniform research area. From socio-technical viewpoint the findings from, and systems developed for document management may provide a starting point for contemporary content management in organizations.

Characteristic for this emerging area of research and development is the move towards using units of content with varying grain sizes. A unit of content to be managed may be broader, similar, or smaller than traditional document. Due to increased popularity of content delivery on the Web, the use of content units smaller than a document has increased. Variance on content unit grain size is reflected to increasing number of kinds and amount of metadata needed for managing and processing the content. The content units defined by social systems are differing from those used and managed by technological systems. Social systems rely upon logical units of content with domain-oriented names, while technical systems use system-oriented terms, and deal with physical and processable units of content. Even though the use of XML (Bray et al., 2004) and related languages provide features for enhancing content management, there are still multiple challenges to be overcome.

REFERENCES

Balasuramanian, V., & Bashan, A. (1998). Document management and web technologies: Alice marries the mad hatter. *Communications of the ACM, 41*(7), 107-115.

Bertino, E., & Catia, B. (2001). Integrating XML and databases. *IEEE Internet Computing, 5*(4), 84-88.

Blair, D., C. (2002). The challenge of commercial document retrieval, Part I: Major issues, and a framework based on search exhaustivity, determinacy of representation and document collection size. *Information Processing & Management, 38*(3), 273-291.

Boiko, B. (2002). *Content management bible*. New York, U.S.A: Hungry Minds, Inc.

Bray, T., Paoli, J., Sperberg-McQueen, C. M., Maler, E., & Yergeau, F. (2004). Extensible markup language (XML) 1.0. (3rd Ed.). W3C Recommendation 4 February 2004. Retrieved June 15, 2004, *http://www.w3.org/TR/2004/REC-xml-20040204/*

Chaudhri, A. B., Rashid, A., & Zicari, R. (2003). *XML data management: Native XML and XML-enabled database systems*. Boston: Addison-Wesley.

Doculabs, I. (1998). *Special report on document management products (2nd ed., Tech. Rep. No. 2.1)*. Chicago, Illinnois: Doculabs, Inc.

DublinCore. (1999). Dublin Core Metadata Element Set (Version 1.1). Retrieved Feb. 17, 2000, from *http://purl.org/DC/documents/rec-dces-19990702.htm*

Fahrenholz-Mann, S. (1999). SGML for electronic publishing at a technical society - Expectations meets reality. *Markup Languages: Theory and Practice, 1*(2), 1-30.

Fallside, D. C. E. (2001, May 2). XML Schema Part 0: Primer. W3C Recommendation, 2 May 2001. Retrieved June 16, 2001, from *http://www.w3.org/TR/xmlschema-0/*

Gilliland-Swetland, A. J. (1998). Defining metadata. In M. Baca (Ed.), *Introduction to metadata* (pp. 1-8). Los Angeles: Getty Information Institute.

Grossniklaus, M., & Norrie, M. C. (2002). Information concepts for content management. In *Proceedings of the Third International Conference on Web Information Systems Engineering (Workshops)*: IEEE.

Heikkinen, B. (2000). *Generalization of document structures and document assembly*. PhD Thesis, University of Helsinki, Helsinki.

Honkaranta, A., & Lyytikäinen, V. (2003). Operationalizing a genre-based method for content analysis: A case of a church. In W. Abramowicz & G. Klein (Eds.), *The Proceedings of the 6th International Conference on Business Information Systems* (pp. 108-116), Colorado Springs, U.S.A.. Poland: Department of Management Information Systems at Poznan University of Economics.

Honkaranta, A., Salminen, A., & Peltola, T. (2005). Challenges in the redesign of content management: A Case of FCP. *International Journal of Cases on Electronic Commerce (IJCEC), 1*(1).

Karjalainen, A., & Tyrväinen, P. (2001). Defining genres and their features for studying information reuse: Preliminary findings of a case study of training materials. In M. Khosrowpour (Ed.), *Managing information technology* *in a global economy. Proceedings of Information Resources Management Association 2001 International Conference* (pp. 346-348), Toronto. Hershey, PA: Idea Group Publishing.

Karjalainen, A., Päivärinta, T., Tyrväinen, P., & Rajala, J. (2000). Genre-based metadata for enterprise document management. In R. H. Sprague (Ed.), *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences (HICSS)*. Los Alamitos, CA: IEEE Computer Society.

Koulopoulos, T., M., & Frappaolo, C. (1995). *Electronic document management systems: A portable consultant* (1st ed.). New York: McGraw-Hill, Inc., U.S.A.

Leavitt, H. J. (1965). Applied organizations change in industry: Structural, technological and humanistic approaches. In J. March (Ed.), *Handbook of organizations* (pp. 1144-1170). Chicago, U.S.A: Rand McNally & Co.

Lindén, G. (1997). *Structured document transformations*. PhD Thesis. Department of Computer Science, University of Helsinki, Helsinki.

Linthicum, D. (2000). *Enterprise application integration*. Reading, Massachusetts: Addison-Wesley Longman, Inc.

Luoma, E., & Tyrväinen, P. (2004). Integrated domain model for digital rights management research and system development. In Dean (Ed.), *E-commerce and M-commerce technologies*. Hershey, PA: Idea Group Publishing.

Lyytikäinen, V. (2004). Analysing requirements for content management. In O. Camp, J. Filipe, S. Hammoudi & M. Piattini (Eds.), *Enterprise Information Systems V (pp. 233-240)*. Dordrecht, Holland: Kluwe Academic Publishers.

Maler, E., & El Andaloussi, J. (1996). *Developing SGML DTDs: From text to model to markup*. Upper Saddle River, NJ: Prentice Hall.

McIntosh, M. (2000). *Content management using the rational unified process®. Rational software white paper*. Cupertino, CA: Rational Software Corporation.

Medina, R., Meyers, S., Bragg, J., & Klima, C. (2002). *Doculabs evaluates document management for enterprise content management*. Retrieved April, 2, 2003, from *http://www.transformmag.com/db_area/archs/2002/02/tfm0202f1.shtml?contentmanagement*

Munkvold, B. E., Päivärinta, T., Hodne, A. K., & Stangeland, E. (2003). Contemporary issues of enterprise content management: The case of statoil. In *Proceedings of the 11th European Conference on Information Systems (ECIS)*.

Murphy, L. D. (1998). Digital document metadata in organizations: Roles, analytical approaches, and future research directions. In R. H. J. Sprague (Ed.), *Proceedings of the 31st Annual Hawaii International Conference on System Sciences (HICSS)*. Los Alamitos, CA: IEEE Computer Society.

Rosenblatt, B., Trippe, B., & Mooney, S. (2002). *Digital rights management: Business and technology*. New York: M&T Books.

Salminen, A. (2000). Methodology for document analysis. In A. Kent (Ed.), *Encyclopedia of library and information science* (Vol. 67, Supplement 30, pp. 299-320). New York: Marcel Dekker, Inc.

Salminen, A., & Tompa, F. W. (2001). Requirements for XML document database systems. In E. V. Munson (Ed.), *ACM Symposium on Document Engineering* (pp. 85-94). Atlanta, Georgia: ACM Press, U.S.A.

Swales, J. M. (1999). *Genre analysis: English in academic and research settings* (6th ed.). Cambridge: Cambridge University Press.

Tyrväinen, P., & Päivärinta, T. (2004). In O. Camp, J. Filipe, S. Hammoudi & M. Piattini (Eds.), *Enterprise Information Systems V* (pp. 258-268). Dordrecht, Holland: Kluwer Academic Publishers.

Weitzman, L., Dean, S. E., Meliksetian, D., Gupta, K., Zhou, N., & Wu, J. (2002). Transforming the content management process at IBM.com. In *Conference on Human Factors in Computing Systems. Case Studies of the CHI2002/AIGA Experience Design FORUM, Minneapolis, Minnesota* (pp. 1-15). New York: ACM Press.

KEY TERMS

Content Aggregation (noun/verb)**:** A set of existing content units collected together for a specific use purpose. An aggregation may contain several versions of the same unit of content, and its creation may require human involvement.

Content Assembly (noun/verb)**:** A collection of existing or new units of content which may be manipulated to produce content for a publication or for a specific target audience. May be produced (semi-)automatically or involve manual processing. A portion of training content for specialists only may be an assembly.

Content Management: Tools, methods, and processes to develop, implement and evaluate the management of content intended mainly for human comprehension. Content has a life-cycle, and its management involves associated metadata.

Content Seeding: Adding identifiers and metadata to content units or their parts to enable computerized assemblies and aggregations on the content.

Content Unit: The object with which the management metadata is associated. May be "a document", "a file", "a component", or "a section of a document" among others.

Document Management: Managing the content consisting of independent logical content units referred to, or manipulated as documents.

Metadata: Data describing the content from the viewpoint of humans or computers. Metadata may describe the domain in which content is used, as well as collections, classes, units or processable portions of the content and content instances. Metadata enables content search, classification, and processing. Whereas humans may use and interpret metadata about documents, the metadata about the small units of content is primarily meant for computerized manipulation of content.

Rendering: A process in which the layout and typography for content is created for a specific human viewer, program, or output device.

Template: A combination of static content, references to units of existing content, and program code for producing the content and navigation aids for a Web site.

Web Content Management: Management of content intended primarily for web delivery. Grain size of the content unit is typically smaller (e.g., a page or a paragraph) than that of documents.

Content–Based Image Retrieval Query Paradigms

Colin C. Venters
University of Manchester, UK

Richard J. Hartley
Manchester Metropolitan University, UK

William T. Hewitt
University of Manchester, UK

INTRODUCTION

The proliferation in volume of digital image data has exacerbated the general image retrieval problem, creating a need for efficient storage and flexible retrieval of vast amounts of image data (Chang, 1989). Whilst there have been significant technological advances with image data capture and storage, developments in effective image retrieval have not kept pace. Research in image retrieval has been divided into two areas: concept-based image retrieval and content-based image retrieval. The former focuses on the use of classification schemes or indexing terms to retrieve images while the latter focuses on the visual features of the image, such as colour, shape, texture, and spatial relationships.

The field of content-based image retrieval has been a thriving area of research and development for over a decade. Its origins lie in the subject areas of artificial intelligence, computer vision, image processing, information retrieval, pattern recognition, and signal processing. Advancement of the approach has been attributed to the early experiments conducted by Kato, Kurita, and Shimogaki (1991) into the automatic retrieval of images by colour and shape feature (Eakins & Graham, 1999). The approach focuses on the semiautomatic or automatic extraction, indexing, and retrieval of images by their visual attributes and characteristics. The process involves a direct matching operation between a query image and a database of stored images, where a feature vector or global signature has been computed for the unique visual properties of an image. Similarity is calculated by comparing the vector or signature of the query image against those of images stored in a database. The visual distance D between images can be measured by functions such as Euclidean distance, L_1 and L_2 distances, weighted Euclidean distance, city-block distance, and Minkowsky distance (Figures 1 to 6).

Figure 1. Euclidean distance

$$d(s_1, s_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Figure 2. Weighted Euclidean distance

$$D_H(I_Q, I_D) = \left(H(I_Q) - H(I_D)\right)^t A\left(H(I_Q) - H(I_D)\right)$$

Figure 3. L_1 distance

$$D_H(I_Q, I_D) = \sum_{j=1}^{n} \left| H(I_Q, j) - H(I_D, j) \right|$$

Figure 4. L_2 distance

$$D_H(I_Q, I_D) = \left(\sum_{j-1}^{n} \left(H(I_Q, j) - H(I_D, j) \right)^2 \right)$$

Figure 5. City-Block distance

$$d(s_1, s_2) = \|x_2 - x_1\| + \|y_2 - y_1\|$$

Figure 6. Minkowsky distance

$$d(s_1, s_2) = \sum_{k-1}^{n} |x_k - y_k|$$

The result of this process is a quantified similarity score that measures the visual distance between the two images, represented by the feature vectors, in a 2-D feature space (Del Bimbo, 1999). The most common image-matching feature characteristics are colour, shape, and texture.

Research in the field of content-based image retrieval has investigated only a small number of the problems inherent in the development of visual information management systems, with a considerable amount of attention being directed at the development and advancement of retrieval algorithms (Venters & Cooper, 2000). However, many research topics in this field, which are considered fundamental to the advancement of content-based image retrieval as a viable retrieval tool, have been largely ignored. One such area is the user interface.

BACKGROUND

The user interface provides the bridge between the end user and the system and is crucial to effective human-computer interaction. Retrieval is commonly obtained through an interactive session (Del Bimbo, 1999). To initiate a search, the user provides or creates the visual representation of their information need and then selects the features, range of model parameters, and similarity measure that are important. Smeulders, Worring, Santini, Gupta, and Jain (2000) defined an abstract query space to represent user interaction in a content-based image retrieval system (Figure 7).

Figure 7. Abstract query space (Smeulders et al., 2000)

$$Q = \{I_Q, F_Q, S_Q, Z_Q\}$$

To start a query session, an instantiation of $Q = \{I_Q, F_Q, S_Q, Z_Q\}$ is created. Where I_Q is the selection of images from the image archive I, F_Q is the selection of features derived from the images in I_Q, S_Q is the similarity function, and Z_Q is the goal-dependent semantics. The query space forms the basis of user interaction in a content-based image retrieval system, specifying queries and displaying results. Similarly, Rui and Huang (2001) state that an image object O can be modelled as a function of the image data D, features F, and representations R (Figure 8).

Figure 8. Image object model (Rui & Huang, 2001)

$$O = O(D, F, R)$$

Where D is the raw image data, $F = \{f_i\}$, $i = 1, \ldots, I$ is a set of visual features associated with the image object, such as colour, texture, or shape, and $R_i\{r_{ij}\}$, $j = 1, \ldots, J_i$ is a set of representations for a given feature, such as colour histogram or colour moments. Each representation of r_{ij} is a vector consisting of multiple components where K_{ij} is the length of the vector r_{ij}, i.e., $r_{ij} = \lfloor r_{ij_1}, \ldots, r_{ij_k}, \ldots, r_{ij} k_{ij} \rfloor$. In general, the majority of content-based image retrieval systems exemplify these interaction models although research activity has only explored specific elements. The models emphasise that user interaction is a complex interplay between the user and the system, suggesting that only highly skilled and specialised users would be able to utilise the system effectively to obtain relevant retrieval results.

VISUAL QUERY FORMULATION

Query formulation is a core activity in the process of information retrieval and a number of paradigms have been proposed for specifying a query Q_n in an image database I. Several database query languages have been proposed to support image query expression in both first- and second-generation image retrieval systems and are based on either the relational or tabular data models. The most widespread languages are extensions of the Structured Query Language (SQL) proposed by Codd (1970). With this approach the user formulates the query and the system retrieves a set of results. Examples include ISQL (Assmann, Venema, & Hohne, 1986), PROBE (Orenstein & Manola, 1988), Spatial SQL (Egenhofer, 1991), and $\sum QL$ (Chang & Jungert, 2001). A major criticism levelled at these languages is the potential complexity of the query statement required to express a simple query (Catarci, Costabile, Levialdi, & Batin, 1997).

As an alternative to the relational-based query language approach, Zloof (1977) proposed Query by Example (QBE). This is based on a tabular data model and employs a tabular query structure to express queries. The user specifies an example output of the query by directly constructing entries into relational skeletons instead of writing lengthy queries, and the system identifies the goal by generalizing an example. QBE is a revision of Codd's relational calculus language SQL. Several examples have been based on this approach, including Aggregate by Example (Klug, 1981), Time by Example (Tansel, Arkun, & Ozsoyoglu, 1989), Generalized Query by Example (Jacobs & Walczak, 1983), Office by Example (Whang et al., 1987), and Natural Forms Query Language (NFQL; Embley, 1989). The languages share a commonality in the method of query expression and differ only in their expansion or revision of the QBE paradigm to serve different application areas. Two languages specifically developed for image databases were Query by Pictorial Example (Chang & Fu, 1980) and PICQUERY (Joseph & Cardenas, 1988).

Content-based image retrieval requires the user to provide or create a visual example or representation of their query image Q_n as the basis of the search from the image archive *I*. Query by Visual Example (QVE) is the general term used to describe this approach. Several examples of this approach to query specification have been reported in the literature. Del Bimbo (1999) and Lew and Huang (2001) distinguish between different methods of QVE: query by browsing, query by icon, query by image, query by painting, and query by sketch.

QUERY BY BROWSING

Query by browsing permits users to browse the database by three methods: unstructured, semi-structured, and structured browsing. The simplest method of query by browsing is unstructured browsing, where the user is presented with a complete view of all the images contained in the database. Generally, the images are presented as thumbnails ordered by filename that the user can scroll through. Semi-structured browsing allows the user to search through the image collection by choosing example images. The system computes a similarity ranking on image features and disregards the least similar images after each iteration until the user is satisfied with the set of images left in the retrieval loop. Examples of this approach were proposed by Laaksonen, Koskela, Laakso, and Oja (2000) and Vendrig, Worring, and Smeulders (2001). With structured browsing, images are clustered into predefined, hierarchal classes based on a visual property such as colour. The images are presented as thumbnails and the user can move through the hierarchy until satisfied. Examples of this approach were proposed by Lai et al. (1999) and Brodley et al. (1999). The concept of browsing has been recognised in the field of information retrieval as a fundamental part of human information-seeking behaviour (Ellis, 1989; Wilson, 1981). Marchionini (1995) states that browsing is an approach to information seeking that is both informal and opportunistic. Nevertheless, the viability of browsing as an access strategy is heavily dependent on the size of the image collection and the effectiveness of any classification by which a physical order is imposed (Enser, 1995). No empirical studies have measured the ceiling or threshold of the maximum number of images a user is willing to search through in order to find a suitable image.

QUERY BY ICON

Query by icon allows the user to select an icon that represents a high-level concept of a category or idea. Icons are commonly found in graphical user interfaces (GUIs) and are a graphic representation of an entity, metaphor, or artefact. Similar to iconic-based visual query languages, the user physically places object icon(s) on a query canvas to specify relative locations, orientations, and sizes of objects within the desired image (Gupta & Jain, 1997). Several examples of this approach have been proposed over the years by Angelaccio, Catarci, and Santucci (1990), Del Bimbo, Campanai, and Nesi (1993), Papantonakis and King (1995), Aslandogan, Thier, and Yu, (1996), Chavda and Wood (1997), and Lee and Whang (2001). Lew and Huang (2001) suggest that the iconic query paradigm is the most intuitive query method, but that it is restricted to a vocabulary of visual concepts that can be reliably understood by the VIR system and semantics by the user.

QUERY BY IMAGE

Query by image allows the user to provide an example image as a representation of their query. The example can be either an image selected externally or internally from the system. Di Lecce and Guerriero (1999) characterize this as query by external pictorial example and query by internal pictorial example. Query by external pictorial example permits the user to submit their own image to the system and is generally perceived as the simplest approach to query formulation. However, it is based on the assumption that the user has a suitable representative image to use as the basis of their query. Query by internal pictorial example is query by browsing where the user selects an image from the system's database. All the images contained in the database are presented or a selection is generated randomly in an *n*-dimensional matrix. Similar to the problems associated with query by browsing, the main disadvantages of this method are providing suitable access mechanisms to retrieve the internal example and the size of the image collection the user is willing to search through in order to find a suitable image.

QUERY BY PAINTING

Query by painting allows the user to manually specify the percentage or the distribution of colour values. For example, the user is able to specify the percentages of colour within a composition, such as 50% green, 25% yellow, and 25% red (Figure 9). Similarly, the user is able to specify the *x y* positions of the of colour values on the query canvas (Figure 10). Examples of this approach were proposed by Flickner et al. (1995) and Del Bimbo, Mugnani, and Turco (1998).

Figure 9. Colour percentages (QBIC)

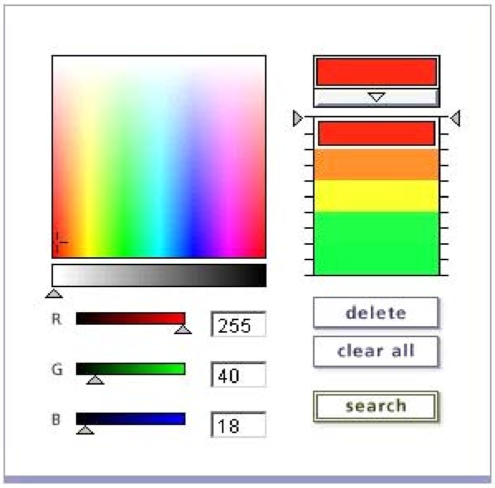
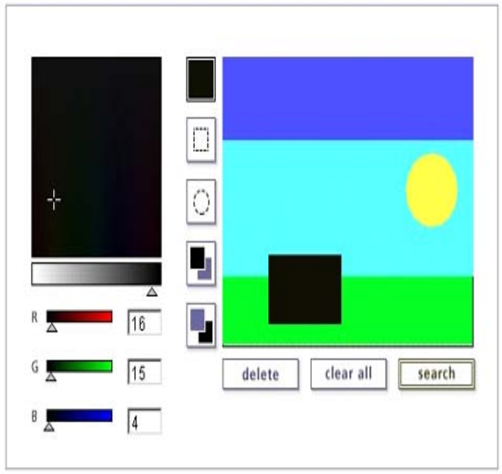


Figure 10. Colour distribution (QBIC)



QUERY BY SKETCH

Query by sketch allows the user to draw a sketch of the desired image by combining several features commonly found in computer graphic applications. The sketch represents a template of either a completed object or scene. Kato, Kurita, Otsu, and Hirata (1992) were the first to provide a query by sketch, but other examples of this approach have since been proposed by Chuah, Roth, and Kerpedjiev (1997), Del Bimbo and Pala (1997), Muller, Eickeler, and Rigoll (1998), and Di Sciascio and Mongiello (1999). As Korfhage (1997) indicates, queries formulated by this method are simplistic, relatively crude sketches of the desired query image, and the tool has a limited functionality for expressing more complex image queries. Eakins (1992) stressed that drawing a shape query is inherently time-consuming and requires a certain modicum of artistic ability. Similarly, the effectiveness of shape-matching features is highly sensitive to noise and pixel arrangement in the query image. Bird, Elliott, Hayward (1999) demonstrated that the user blamed their own artistic drawing ability when query results where unsuccessful.

Query by image, painting, and sketch are the most widespread interaction methods of query formulation found in content-based image retrieval systems. As Zachary and Iyengar (1999) suggest, a good query method will be natural to use as well as capturing enough information from the user to extract meaningful results. However, supporting the expression of image queries by these modes of interaction is a nontrivial problem. Eakins (1992) stresses, the more complex the query in terms of shape and structure, the more difficult it will be for the user to express and produce a meaningful visual example. Similarly, Chang and Jungert (2001) state that most visual query systems (VQSs) restrict human-computer interaction to only one kind of interaction paradigm. They suggest that the presence of several paradigms, each one with its different characteristics and advantages, will help both naïve and experienced users in interacting with the system.

FUTURE TRENDS

Alternative methods have been proposed to extend the QVE paradigm. Gupta and Jain (1997) note that while most systems are limited in the query types they can handle, i.e., colour, shape, and texture, image query specification should not be performed exclusively within the query by visual example paradigm but should be extended through a range of different tools that serve as a visual information retrieval query language. They proposed a nine-component query language framework that included: image processing, feature-space manipulation, object specification, measurement specification, classification, spatial arrangement, temporal arrangement, annotation, and data definition. Several visualization tools for image browsing that have extended approaches used in text-based information visualizations systems have been proposed by Cinque, Levialdi, Malizia, and Olsen (1998), Santini and Jain (1997), and Nakazato and Huang (2001). Although there have been many prototypes of visualization in support of information retrieval, there has been little systematic evaluation to distinguish the benefits of the visualization per se from the various accompanying features (Korfhage, 1997). Assfalg, Del Bimbo, and Pala (2000) argue that virtual reality (VR) provides a realistic reproduction of a user's natural environment, particularly where images represent real-world scenes. They propose that the use of a non-immersive 3-D environment is a natural and effective interface to perform query by content. Three-dimensional virtual reality (3DVR) potentially represents the next generation of user interfaces for information retrieval. How they can be effectively applied in an operational setting is unclear and introduces a new avenue for further research.

CONCLUSION

Several commentators have highlighted the need to design and develop new user-interface metaphors and paradigms for specifying visual queries (Aigrain, Zhang, & Petkovic, 1996; Besser, 1996; Chang & Hsu, 1992; Del Bimbo, 1999; Eakins, 2001; Gudivada & Raghavan, 1995; Gupta & Jain, 1997; Jain, 1996; Rui, Huang, & Chang, 1999; Venters, Hartley, & Hewitt, 2003). However, despite being identified as one of the major research areas of the field, few studies have investigated the user interface. The requirement for end users to communicate their information needs accurately and effectively is crucial in any information retrieval system. How this may be achieved in practice with content-based image retrieval systems is not readily apparent (Eakins & Graham, 1999). A number of research questions from the macro- to the micro-level have been postulated regarding research into the user interface and content-based image retrieval. These include: What is the most efficient user interface for querying an image database ("CBIR Challenges," n.d.; Jain, 1993; Venters et al., 2003)? Is access to visual data best resolved by allowing the user to switch between browsing, navigation, and querying (Aigrain et al., 1996; Del Bimbo, 1999)? What is the most effective method for specifying an initial query to the system, and the best to refine search parameters if the retrieval results fail to meet user information needs (Eakins, 2001; Rui et al., 1999;)? The validity of the query by visual example paradigm and its effectiveness in supporting image query expression remain open and valid research issues that now require critical attention. The problem of what type of user interface is required to support human-computer interaction in a content-based image retrieval environment remains unsolved.

REFERENCES

Aigrain, P., Zhang, H., & Petkovic, D. (1996). Content-based image retrieval of visual media: A state of the art review. *Multimedia Tools and Applications*, *3*, 179-202.

Angelaccio, M., Catarci, T., & Santucci, G. (1990). QBD*: A graphical query language with recursion. *IEEE Transactions on Software Engineering*, *16*, 1150-1163.

Aslandogan, Y. A., Thier, C., & Yu, C. (1996). A system for effective content based image retrieval. *Multimedia'96: Proceedings of the Fourth ACM International Conference on Multimedia*, 429-430.

Assfalg, J., Del Bimbo, A., & Pala, P. (2000). Virtual reality for image retrieval. *Journal of Visual Languages and Computing*, *11*, 105-124.

Assmann, K., Venema, R., & Hohne, K. H. (1986). The ISQL language: A software tool for the development of pictorial information systems in medicine. In S.-K. Chang, T. Ichikawa, & P. Ligomenides (Eds.), *Visual languages* (pp. 261-283).

Besser, H. (1996). Image databases: The first decade, the present, and the future. In *Digital Image Access and Retrieval: 1996 Clinic on Library Applications of Data Processing* (p. 17).

Bird, C., Elliott, P. J., & Hayward, P. M. (1999). Content-based retrieval for European image libraries. In *The Challenge of Image Retrieval: CIR99, The Second UK Conference on Image Retrieval* (pp. 25-26).

Brodley, C. E., Kak, A. C., Shyu, C., Dy, J., Broderick, L. S., & Aisen, A. M. (1999). Content-based retrieval from medical image databases: A synergy of human interaction, machine learning and computer vision. In *AAAI/IAAI'99: The Proceedings of the 16th National Conference on Artificial Intelligence and 11th Conference on Innovative Applications of Artificial Intelligence*, (pp. 760-767). MIT Press.

Catarci, T., Costabile, M. F., Levialdi, S., & Batini, C. (1997). Visual query systems for databases: A survey. *Journal of Visual Languages and Computing*, *8*(2), 215-260.

CBIR challenges. (n.d.). Retrieved from http://www.benchathlon.net/resources/challenges.html

Chang, N.-S., & Fu, K.-S. (1980). Query by pictorial example. *IEEE Transactions on Software Engineering*, *6*(6), 519-524.

Chang, S-K. (1989). Principles of pictorial information systems design. Prentice-Hall.

Chang, S.-K., & Hsu, A. (1992). Image information systems: Where do we go from here? *IEEE Transactions on Knowledge and Data Engineering*, *4*(5), 431-442.

Chang, S.-K., & Jungert, E. (2001). Query languages for multimedia search. In M. S. Lew (Ed.), *Principles of visual information retrieval* (pp. 199-217). Springer-Verlag.

Chavda, M., & Wood, P. (1997). Towards an ODMG-compliant visual object query language. In *VLDB'97: Proceedings of 23rd International Conference on Very Large Data Bases* (pp. 456-465).

Chuah, M. C., Roth, S. F., & Kerpedjiev, S. (1997). Sketching, searching, and custom visualizations: A content-based approach to design retrieval. In M. T. Maybury (Ed.), *Intelligent multimedia information retrieval* (pp. 83-111). AAAI Press.

Cinque, L., Levialdi, S., Malizia, A., & Olsen, K. A. (1998). A multidimensional image browser. *Journal of Visual Languages and Computing*, *9*(1), 103-117.

Codd, E. F. (1970). A relational model of data for large shared data banks. *Communications of the ACM*, *13*(6), 377-387.

Del Bimbo, A. (1999). *Visual information retrieval*. Morgan Kaufmann.

Del Bimbo, A. & Pala, P. (1997). Visual image retrieval by elastic matching of user sketches. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*(2), 121-132.

Del Bimbo, A., Campanai, M., & Nesi, P. (1993). A three-dimensional iconic environment for image database querying. *IEEE Transactions on Software Engineering*, *19*(10), 997-1011.

Del Bimbo, A., Mugnani, P. P., & Turco, F. (1998). Visual querying by colour perceptive regions. *Pattern Recognition*, *31*(9), 1241-1253.

Di Lecce, V., & Guerriero, A. (1999). An evaluation of the effectiveness of image features for image retrieval. *Journal of Visual Communication and Image Representation*, *10*, 353.

Di Sciascio, E., & Mongiello, M. (1999). Query by sketch and relevance feedback for content-based image retrieval over the Web. *Journal of Visual Languages and Computing*, *10*, 565-584.

Eakins, J. P. (1992). Pictorial information systems: Prospects & problems. In *Proceedings of the 14th British Computing Society Information Retrieval Specialist Group Research Colloquium* (pp. 102-123).

Eakins, J. P. (2001). Trademark image retrieval. In M. S. Lew (Ed.), *Principles of visual information retrieval* (p. 345). Springer-Verlag.

Eakins, J. P., & Graham, M. E. (1999). *Content-based image retrieval*. JISC, 27.

Egenhofer, M. (1991). Spatial SQL: A query and presentation language. *IEEE Transactions on Knowledge Data Engineering*, *5*(2), 161-174.

Ellis, D. (1989). A behavioral approach to information retrieval systems design. *Journal of Documentation*, *45*(3), 171-211.

Embley, D. W. (1989). NFQL: The natural forms query language. *ACM Transactions on Database Systems*, *14*(2), 168-211.

Enser, P. G. B. (1995). Progress in documentation pictorial information retrieval. *Journal of Documentation*, *51*(2).

Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., et al. (1995). Query by image and video content: The QBIC system. *IEEE Computer*, *28*, 23-32.

Gudivada, V. N., & Raghaven, V. V. (1995, September). Content-based image retrieval systems. *IEEE Computer*, 18-22.

Gupta, A., & Jain, R. (1997). Visual information retrieval. *Communications of the ACM*, *40*(5), 71-79.

Jacobs, B. E., & Walczak, C. A. (1983). A generalized query-by-example data manipulation language based on database logic. *IEEE Transactions on Software Engineering*, *SE-9*(1), 40-56.

Jain, R. (1993). NSF workshop on visual information retrieval management systems. In W. Niblack (Ed.), *Storage and Retrieval for Image and Video Databases: Proceedings of the International Society for Optical Engineering* (pp. 198-218).

Jain, R. (1996). Visual information retrieval in digital libraries. In *Digital Image Access and Retrieval: 1996 Clinic on Library Applications of Data Processing* (pp. 67-85).

Joseph, T., & Cardenas, A. F. (1988). PICQUERY: A high level query language for pictorial database management. *IEEE Transactions on Software Engineering*, *14*(5), 630-638.

Kato, T., Kurita, T., Otsu, N., & Hirata, K. (1992). A sketch retrieval method for full colour image database: Query by visual example. *Computer Vision and Applications: Proceedings of ICPR*, 530-533.

Kato, T., Kurita, T., & Shimogaki, H. (1991). Intelligent visual interaction with image database systems: Towards the multimedia personal interface. *Journal of Information Processing*, *14*(2), 134-143.

Klug, A. C. (1981). ABE: A query language for constructing aggregates-by-example. SSDBM: Proceedings of the First LBL Workshop on Statistical Database Management, Melno Park, California, USA, December 2-4, 1981, pp. 190-205.

Korfhage, R. R. (1997). Information Storage and Retrieval. John Wiley & Sons.

Laaksonen, J., Koskela, M., Laakso, S., & Oja, E. (2000). PicSOM: Content-based image retrieval with self-organising maps. *Pattern Recognition Letters*, *21*, 1199-1207.

Lai, T. S., Tait, J.I. & McDonald, S. (1999). Image browsing and navigation using hierarchical classification. CIR'99: The Challenge of Image Retrieval: Second UK Conference on Image Retrieval. Newcastle upon Tyne, UK, 25-26 February, 1999.

Lai, T.-S., Tait, J., & McDonald, S. (2000). A user-centred evaluation of visual search methods for CBIR. In *The Challenge for Image Retrieval: CIR2000, Third UK Conference on Image Retrieval*.

Lee, S. K., & Whang, K.-Y. (2001). VOQL. *Journal of Visual Languages and Computing*, *12*, 413-433.

Lew, M. S., & Huang, T. S. (2001). Visual information retrieval: Paradigms, applications, and research issues. In M. S. Lew (Ed.), *Principles of visual information retrieval* (p. 3). Springer-Verlag.

Marchionini, G. (1995). *Information seeking in electronic environments*. University of Cambridge Press.

Muller, S., Eickeler, S., & Rigoll, G. (1998). Image database retrieval of rotated objects by user sketch. In *CBAIVL: Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries* (pp. 40-44).

Nakazato M. and Huang, T. S. (2001). An interactive 3D visualization for content-based image retrieval. In: ICME2001: Proceedings of the International Conference on Multimedia and Expo, Waseda University, Tokyo, Japan, 22-25 August 2001.

Orenstein, J. A., & Manola, F. A. (1988). PROBE spatial data modeling and query processing in an image database application. *IEEE Transactions on Software Engineering*, *14*(5), 611-628.

Papantonakis, A., & King, P. J. H. (1995). Syntax and semantics of Gql, a graphical query language. *Journal of Visual Languages and Computing*, *6*, 3-25.

Rui, Y., & Huang, T. S. (2001). Relevance feedback techniques. In M. S. Lew (Ed.), *Principles of visual information retrieval* (p. 219). Springer-Verlag.

Rui, Y., Huang, T. S., & Chang, S.-F. (1999). Image retrieval: Current techniques, promising directions, and open issues. *Journal of Visual Communication and Image Representation*, *10*(1), 39-62.

Santini, S., & Jain, R. (1997). Visual navigation in perceptual databases. In *Visual'97; Proceedings of the Second International Conference on Visual Information Systems*.

Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(12).

Tansel, A. U., Arkun, M. E., & Ozsoyoglu, G. (1989). Time-by-example query language for historical databases. *IEEE Transactions on Software Engineering*, *15*(4), 464-478.

Vendrig, J., Worring, M., & Smeulders, A. W. M. (2001). Filter image browsing: Interactive image retrieval by using database overviews. *Multimedia Tools and Applications*, *15*(1), 83-103.

Venters, C. C., & Cooper, M. D. (2000). A review of content-based image retrieval systems. *JISC*.

Venters, C. C., Hartley, R. J., & Hewitt, W. T. (2003). Mind the gap: Content-based image retrieval and the user interface. In S. Deb (Ed.), *Multimedia systems and content-based image retrieval*. Hershey, PA: Idea Group Publishing.

Whang, K.-Y., Ammann, A., Bolmarcich, A., Hanrahan, M., Hochgesang, G., Huang, K.-T., et al. (1987). Office-by-example: An integrated office system and database manager. *ACM Transactions on Office Information Systems*, *5*(4), 393-427.

Wilson, T.D. (1981). On user studies and information needs. *Journal of Documentation,* 37,3-15.

Zachary, J., & Iyengar, S. (1999). Content-based image retrieval systems. In *ASSET'99: IEEE Symposium on Application-Specific Systems and Software Engineering & Technology*.

Zloof, M. M. (1977). Query-by-example: A data base language. *IBM Systems Journal*, *16*(4), 324-343.

KEY TERMS

Concept-Based Image Retrieval: A term used to describe the classical approach to information management that focuses on the use of classification schemes and their indexing terms to retrieve images.

Content-Based Image Retrieval: A general term used to describe the semiautomatic or automatic extraction, indexing, and retrieval of images by their visual attributes and characteristics.

Digital Image: Image $f(x,y)$ that has been discretized both in spatial coordinates and brightness and consists of a set of elements, defined on an n-dimensional regular grid, that have the potential for display. It can be considered a matrix whose row and column indices identify a point in the image, and the corresponding matrix element values identify the grey level at that point. The elements of such a digital array are called image elements, picture elements, pixels, or pels.

Query: Formal expression of an information need.

Query by Browsing: Approach that permits users to browse the database by three methods: unstructured, semi-structured, and structured browsing.

Query by Icon: An approach that allows users to construct queries by selecting an icon that represents a high-level concept of a category or idea.

Query by Image: An approach that allows users to provide an example image as a representation of their query. The example can be either an image selected externally or internally from the system.

Query by Painting: An approach that allows users to manually specify the percentage or the distribution of colour values from a colour picker or wheel, such as RGB (red, green, blue) or HIS (hue, intensity, saturation).

Query by Sketch: An approach that allows users to draw a sketch of the desired image by combining several features commonly found in computer graphic applications. The sketch represents a template of either a completed object or scene.

Query by Visual Example: General term used to describe how users provide or create a visual example or representation of their query image Q_n as the basis of the search from the image archive I.

Query Language: Set of formally defined operators that allow queries to be expressed to a database. The execution of the query produces a set of results that are extracted from the data.

Content-Based Retrieval Concept

Yung-Kuan Chan
National Chung Hsing University, Taiwan, R.O.C.

Chin-Chen Chang
National Chung Cheng University, Taiwan, R.O.C.

INTRODUCTION

Because of the demand for efficient management in images, much attention has been paid to image retrieval over the past few years. The text-based image retrieval system is commonly used in traditional search engines (Ratha et al., 1996), where a query is represented by keywords that are usually identified and classified by human beings. Since people have different understandings on a particular image, the consistency is difficult to maintain. When the database is larger, it is arduous to describe and classify the images because most images are complicated and have many different objects. There has been a trend towards developing the content-based retrieval system, which tries to retrieve images directly and automatically based on their visual contents.

A similar image retrieval system extracts the content of the query example ***q*** and compares it with that of each database image during querying. The answer to this query may be one or more images that are the most similar ones to ***q***. Similarity retrieval can work effectively when the user fails to express queries in a precise way. In this case, it is no longer necessary to retrieve an image extremely similar to the query example. Hence, similarity retrieval has more practical applications than an exact match does.

Content-Based Image Retrieval Systems

In a typical content-based image retrieval system, the query pattern is queried by an example in which a sample image or sketch is provided. The system then extracts appropriate visual features that can describe the image, and matches these features against the features of the images stored in the database. This type of query is easily expressed and formulated, since the user does not need to be familiar with the syntax of any special purpose image query language. The main advantage is that the retrieval process can be implemented automatically (Chen, 2001). The scope of this article is circumscribed to image abstraction and retrieval based on image content.

Human beings have a unique ability that can easily recognize the complex features in an image by utilizing the attributes of shape, texture, color, and spatial information. Many researchers analyze the color, texture, shape of an object, and spatial attributes of images, and use them as the features of the images. Therefore, one of the most important challenges in building an image retrieval system is the choice and representation of the visual attributes. A brief overview of the commonly used visual attributes shape, texture, color, and spatial relationship will be illustrated as follows.

Commonly Used Image Features in Content-Based Image Retrieval Systems

Shape characterizes the contour of an object that identifies the object in a meaningful form (Gevers & Smeulders, 2000; Zhang & Lu, 2002). Traditionally, shapes are described through a set of features such as area, axis-orientation, certain characteristic points, and so forth. These systems retrieve a subset of images that satisfy certain shape constraints. In the shape retrieval, the degree of similarity between two images is considered as the distance between the corresponding points.

Color attribute may simplify the object's identification and extraction in the image retrieval (Galdino & Borges, 2000; Gevers & Smeulders, 2000). Color may provide multiple measurements at a single pixel of the image, and often enable the classification to be done without complex spatial decision-making. Any resulting difference between colors is then evaluated as a distance between the corresponding color points. The color-based retrieval system measures the similarity of the two images with their distance in color space.

Texture attribute depicts the surface of an image object (Yao & Chen, 2002; Zhang & Tan, 2003). Intuitively, the term refers to properties such as smoothness, coarseness, and regularity of an image object. Generally, the structural homogeneity does not come from the presence of a single color or intensity, but it requires the interaction of various intensities within a region.

Retrieval by spatial constraints facilitates a class of queries based on the 2-D arrangement of objects in an image (Chang Erland & Li, 1989; Chang & Li, 1988; Chang, Shi & Yan, 1987; Lee & Hsu, 1992). The query is composed by placing sketches, symbols or icons on a plane where every symbol or icon is predefined for one type of objects in an image. The relationships between the objects can be broadly classified as either directional (also referred as projective) (Chang & Li, 1988; Chang, Shi & Yan, 1987) or topological (Lee & Hsu, 1992). Directional relationship is based on the relative location and the metric distance between two image objects. Topological relationships are based on set-theoretical concepts like union, intersection, disjunction and so forth. Spatial information is a higher-level attribute, which is increasingly more specific. For example, facial features are frequently presented in terms of spatial information (Sadeghi, Kittler & Messer, 2001).

Briefly, color attribute depicts the visual appearance of an image, characterized by the luminance and chrominance histograms of the image. Texture attribute refers to three components: bi-dimensional periodicity, mono-dimensional orientation, and complexity obtained through world decomposition. Shape attribute sketches the geometrical properties of objects in images. Spatial attribute represents the relative position relationships between objects of an image.

TYPICAL IMAGE RETRIEVAL SYSTEMS

This section briefly overviews the image retrieval systems based on the most commonly used image features: color, shape, texture, and spatial content.

The Color-Based Image Retrieval Systems

Generally, the color-based image retrieval system does not find the images whose colors are exactly matched, but images with similar pixel color information. This approach has been proven to be very successful in retrieving images since concepts of the color-based similarity measure is simple, and the convention algorithms are very easy to implement. Besides, this feature can resist noise and rotation variants in images.

However, this feature can only be used to take the global characteristics into account rather than the local one in an image, such as the color difference between neighboring objects in an image. For example, if a landscape image with blue sky on the top and green countryside at the bottom is employed as a query example, the system that retrieves the images with similar structures based on these global features often gives very unsatisfactory results. In addition, the color-based image retrieval system often fails to retrieve the images that are taken from the same scene in which the query example is also taken from under different time or conditions, for example, the images of a countryside taken at dusk or dawn under a clear or a cloudy sky. In another scenario, the same scene may be imaged by different devices. Using one image taken by one device as the query example may fail to find the same scene taken by other devices.

The Shape-Based Image Retrieval Systems

A shape-based image retrieval system is used to search for the images containing the objects, which are similar to the objects specified by a query. Since an object can be formed by a set of shapes in most cases (e.g., a car can be made of some little rectangles and circles), most similar objects have a high correlation in their set of shapes (Gevers & Smeulders, 2000; Zhang & Lu, 2002). The shape-based image retrieval system extracts the shapes of objects from images by segmentation, and classifies the shapes, where each shape has its own representation and variants to scaling, rotation, and transition.

Some criteria on shape representation and similarity measure for a well performing content-based image retrieval system should be achieved. Firstly, the representation of a shape should be invariant to scale, translation, and rotation. Secondly, the similarity measure between shape representations should conform to human perception; that is, perceptually similar shapes should have highly similar measures. Thirdly, the shape representation should be compact and easy to derive, and the calculation of similarity measure should be efficient.

However, how to locate and how to recognize objects from images is a real challenge. One of the obstacles is how to separate the objects from the background. Difficulties come from discrimination, occlusions, poor contrast, viewing conditions, noise, complicated objects, complicated backgrounds, and so forth. Moreover, the shape-based image retrieval system can only deal with the images that have simple object shapes. For complex object shapes, the region-based method has to build a binary sequence by using smaller grid cells, so that results that are more accurate can be obtained; nevertheless, the storage of indices and retrieval time may increase tremendously.

The Texture-Based Image Retrieval Systems

Literally, texture relates to the arrangement of the basic constituents of a material. In digital images, texture describes the spatial interrelationships of the image pixels.

Texture similarity can often be useful in distinguishing the areas of objects in images with similar color, such as sky and sea as well as leaves and grass. Texture queries can be formulated in the manner that is similar to the color queries by selecting an example of desired textures from a palette, or by supplying an example query image. The system then returns the images which are most similar to the query example in texture measures.

Making texture analysis is a real challenge. One way to perform content-based image retrieval using texture as the cue is by segmenting an image into a number of different texture regions and then performing a texture analysis algorithm on each texture segment. However, segmentation can sometimes be problematic for image retrieval. In addition, texture is quite difficult to describe and subject to the difference of human perception. No satisfactory quantitative definition of texture exists at this time.

The Spatial-Based Image Retrieval Systems

There are two kinds of spatial-based image retrieval systems: retrieval by spatial relationships (RSRs) and spatial access methods (SAMs). The RSR image retrieval system is to retrieve the images from a database that are similar to the query sample based on relative position relationships between the objects in the images. Hence, a physical image can be regarded as a symbolic image, each object of which is attached with a symbolic name. The centroid coordinates of the object with reference to the image frame are extracted as well. By searching for the logical images, the corresponding physical images can then be retrieved and displayed. Therefore, image retrieval can be simplified to the search of symbolic images.

Chang, Shi, and Yan (1987) used a 2D string representation to describe a symbolic image. Objects and their spatial relationships in a symbolic image can be characterized by a 2D string. An image query can be specified as a 2D string too. Consequently, the problem of image retrieval then turns out to be the matching of a 2D string. Subsequently, a great number of other image representations popped out that were derived from a 2D string, such as 2D G-string (Chang, Erland & Li, 1989), 2D B-string (Lee, Yang & Chen, 1992), 2D C-string (Lee & Hsu, 1992), and so forth. These representations adopt the description of orthogonal projection to delineate the spatial relationships between objects.

The SAM image retrieval systems are to manage large collections of points (or rectangles or other geometric objects) in the main memory or on the disk so that range queries can be efficiently answered. A range query specifies a region in the address space, requesting all the data objects that intersect it. They divide the whole space into several disjoint sub-regions, each with no more than P points (a point may represent a rectangle). P is usually the capacity of a disk page. Inserting a new point may result in further splitting a region. The split methods can be classified according to the attributes of the split (Gottschalk, Turney & Mudge, 1987).

Color attribute is most intuitive and straightforward for the user. Texture analysis systems are often developed to perform filtering in the transform domain in order to obtain feature images. The use of global color or texture features for the retrieval of images tends to be misleading, especially in homogeneous image collections. Though shape and texture are the essential visual attributes to derive potentially useful semantic information, there exists less understanding of the benefits to implement these attributes as compared to color, for efficient image retrieval. This approach apparently focuses on global frequency content of an image; however, many applications require the analysis to be localized in the spatial domain.

FUTURE TRENDS

Many visual attributes have been explored, such as color, shape, texture, and spatial features. For each feature, there exist multiple representations that model the human perception of the feature from different perspectives. There is a demand for developing an image content description to organize the features. The features should not only be just associated with the images, but also be invoked at the right place and the right time, whenever they are needed to assist retrieval.

Human beings tend to apply high-level concepts in their daily lives. However, most features, the current computer vision techniques automatically extracting from images, are low-level. To narrow down this semantic gap, it is necessary to link the low-level features to high-level concepts. On the high-level concept, it should allow the user to easily provide his or her evaluation of the current retrieval results to the computer. It is likely for different people to give an identical name; therefore, generating the representation by automatically extracting the objects from an original image is very difficult. Therefore, the spatial relationships between objects cannot be extracted automatically without human interaction with the current techniques of image understanding and recognition. More recent research emphasis is given to “interactive systems” and “human in the loop”.

Due to the perception subjectivity of image content, it is difficult to define a good criterion for measuring the similarity among images. That is, the subjectivity of image perception prevents us from defining objective evaluation criteria. Hence, it is urgent to find an appropriate way of evaluating the system performance guiding the research effort in the correct direction.

Establishing a well-balanced large-scale test bed is an important task too. A good test bed must be huge in scale for testing the scalability (for multidimensional indexing), and be balanced in image content for testing image feature effectiveness and overall system performance.

Human beings are the ultimate end users of the image retrieval system. This topic has attracted increasing attention in recent years, aiming at exploring how humans perceive image content and how one can integrate such a "human model" into the image retrieval systems. Recently, more studies of human perception focus on the psychophysical aspects of human perception.

CONCLUSION

A successful image retrieval system requires the seamless integration of the efforts of multiple research communities. To achieve a fast retrieval speed and make the retrieval system truly scalable to large-size image collections, an effective multidimensional indexing module is an indispensable part of the whole system. The interface collects the information from the users and displays back the retrieval results to the users in a meaningful way. To communicate with the user in a friendly manner, the query interface should be graphics-based.

In the iconic image representation, an icon is used to represent an object in an image. The iconic image representation has two advantages. First, once the images in the database are analyzed, it is not necessary to analyze the image again in a query processing. Secondly, since the size of the symbolic image is much smaller than that of the original image, this representation can be well suited to be distributed to database environments where a large number of image transmissions between distant nodes are required.

Typically, color, shape and texture attributes provide a global description of images, but they fail to consider the meaning of portrayed objects and the semantics of scenes. The descriptions of objects and the relative positions among objects provide a spatial configuration and a logical representation of images. Combining approaches for content-based image retrieval systems could be considered as complementary.

REFERENCES

Chang, S.K., & Li, Y. (1988). Representation of multiresolution aymbolic and binary images using 2D H-strings. *Proceedings of IEEE Workshop Languages for Automation*, 190-195.

Chang, S.K., Erland, J., & Li, Y. (1989). The design of pictorial database upon the theory of symbolic projections. *Proceedings of the 1st Symposium on the Design and Implementation of Large Spatial Databases*, 303-323.

Chang, S.K., Shi, Q.Y., & Yan, C.W. (1987). Iconic indexing by 2D strings. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 9*(3), 413-328.

Chen, H.L. (2001). An analysis of image retrieval tasks in the field of art history. Information Processing and Management, *37*(5), 701-720.

Galdino, L.L., & Borges, D.L. (2000). A visual attention model for tracking regions based on color correlograms. *Proceedings of 8th Brazilian Symposium on Computer Graphics and Image Processing*, 36-43.

Gevers, T., & Smeulders, A.W.M. (2000). PicToSeek: Combining color and shape invariant features for image retrieval. *IEEE Transactions on Image Processing, 9*(1), 102-119.

Gottschalk, P., Turney, J., & Mudge, T. (1987). Two-dimensional partially visible object recognition using efficient multidimensional range queries. *Proceedings of IEEE International Conference on Robotics and Automation, 4*, 1582 -1589.

Lee, S.Y., & Hsu, F.J. (1992). Spatial reasoning and similarity retrieval of images using 2D C-string knowledge representation. *Pattern Recognition, 25*(3), 305-318.

Lee, Y., Yang, M.C., & Chen, J.W. (1992). 2D B-string knowledge representation and image retrieval for image database. *Proceedings of 2nd International Computer Science Conference Data and Knowledge Engineering: Theory and Applications*, 13-16.

Owei, V., & Navathe, S.B. (2001). Enriching the conceptual basis for query formulation through relationship semantics in databases. *Information Systems, 26*(6), 445-475.

Ratha, N.K., Karu, K., Chen, S.Y., & Jain, A.K. (1996). A real-time matching system for large fingerprint databases. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 18*(8), 799 -813.

Sadeghi, M., Kittler, J., & Messer, K. (2001). Spatial clustering of pixels in the mouth area of face images. *Proceedings of 11th International Conference on Image Analysis and Processing*, 36-41.

Yao, C.H., & Chen, S.Y. (2002). Retrieval of translated, rotated and scaled color textures. Pattern Recognition, *36*(4), 913-929.

Zhang, D., & Lu, G. (2002). Shape-based image retrieval using generic Fourier descriptor. *Signal Processing: Image Communication, 17*(10), 825-848.

Zhang, J., & Tan, T. (2003). Affine invariant classification and retrieval of texture images. *Pattern Recognition, 36*(3), 657-664.

KEY TERMS

Color Feature: Analyzing the color distribution of pixels in an image.

Geometric Hashing: The technique identifying an object in the scene, together with its position and orientation.

Query by Example: The image retrieval system where a sample image or sketch can be provided as a query.

Shape Feature: Characterizing the contour of an object that identifies the object in a meaningful form.

Spatial Feature: Symbolizing the arrangement of objects within the image.

Symbolic Image: Consisting of a set of objects; each object stands for an entity in a real image.

Texture Feature: Depicting the surface of an image object.

C

Context-Aware Framework for ERP

Farhad Daneshgar
University of New South Wales, Australia

INTRODUCTION

Like many existing ERP models (e.g., Podolsky, 1998; Van Stijn & Wensley, 2001), the OOAB framework is also based on a widely accepted assumption that a corporate-wide information system consists of a set of potentially related subsystems; and as a result, information flows among these subsystems must be identified, and required resources planned, using an appropriate ERP methodology. However, up until now there existed no formalised framework that facilitates sharing of contextual knowledge in ERP processes. A unique attribute of the OOAB framework is that it treats ERP processes as a collaborative processes where various roles/actors collaboratively perform tasks in order to achieve a common overall goal. An object-oriented framework is presented in this article that facilitates sharing the contextual knowledge/resources that exist within ERP processes. Context is represented by a set of relevant collaborative semantic concepts or "objects". These are the objects that are localised/contextualised to specific sub-process within the ERP process.

BACKGROUND

From a purely object orientation perspective, a collaboration is defined as "the structure of instances playing roles in a behavior and their relationships" (OMG, 2001). The behaviour mentioned in this definition refers to an operation, or a use case, or any other behavioural classifier. This article provides an overview of a framework for analysing awareness requirements of the actors in ERP systems using an object-oriented awareness-based approach. A similar study was also conducted for developing a new version of this framework that takes into consideration the specific characteristics of virtual communities (Daneshgar, 2003). The proposed approach specialises the notion of collaboration and extends it to the ERP processes. This has roots in the activity network theory (Kaptilini et al., 1995) and is based on the fact that all ERP processes involve *multiple roles performing various tasks* using appropriate artefacts (e.g., departmental subsystems, databases, etc.) in order to achieve both their local as well as the overall organization-wide goals. Conceptually speaking, this will justify a frame-based object-oriented approach to analysis and design for ERP processes (Turban & Aaron, 2001). The conceptual model of the proposed framework is made of the following components:

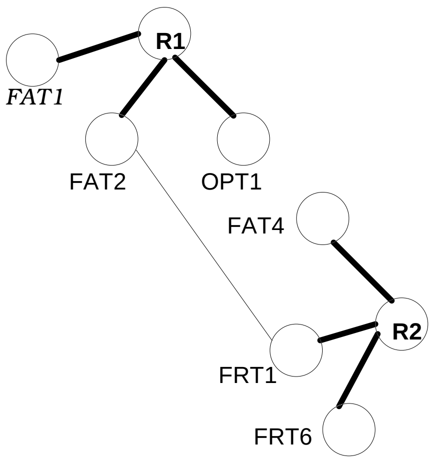
- a set of collaborative semantic concepts including roles, the tasks that these roles play within the process, and the artefacts that these roles use to perform various tasks within the process, and
- relationships among these semantic concepts.

This conceptual model can then be mapped directly to an object model and be used as an analytical tool for identifying awareness requirements of the actors within the ERP process. The fact that ERP is treated as a collaborative process calls for a mechanism for maintaining awareness requirements of the actors involved in this collaboration. Furthermore, due to its object orientation, the framework is capable of encapsulating all complications and dependencies in sub/local processes within individual tasks as well as resources required to perform those tasks, further relieving the ERP management and the associated software.

OOAB FRAMEWORK

A domain-specific conceptual model of a hypothetical ERP process that resembles an object diagram is shown in Figure 1. Use of a domain-specific conceptual model instead of a straight object diagram is justified by the fact that the ontological foundation of the framework prevents growth of the objects and relationships indefinitely, and as a result using an object model may hide such ontology. In Figure 1 there are two roles: R1 and R2; six tasks: FAT1, FAT2, OPT1, FAT4, FRT1 and FRT6 (all shown by circles). It also shows various resources by straight lines connecting tasks and roles. These lines represent rich ontological relationship between a pair of semantic concepts. Each task object requires certain resources for achieving its local/departmental goal or purpose (called *process resource*), as well as certain other resources for achieving the collaborative organization-wide goals of the ERP process (called *collaborative resource*). In Figure 1, a line

Figure 1. A representation of an ERP collaborative process model



connecting a role vertex to a task vertex is a process resource, whereas a line connecting two tasks together is a collaborative resource.

According to the framework, effective knowledge and/or resource exchange among actors is closely related to the level of awareness as defined in the awareness model that each actor possess about the ERP process. These awareness levels are defined in terms of the collaborative semantic concepts used within the ERP conceptual model as shown in Figure 1. Details of the proposed methodology for identifying awareness requirements of actors in ERP process follow:

STEP 1. Develop an ERP Process Model similar to that in Figure 1.
FA: Financial Accounting sub-process/task
OP: Order Processing sub-process/task
CS: Customer Service subprocess/task
FR: Financial Reporting subprocess/task
T1…T6: <appear as postfixes indicating various tasks>

STEP 2. Measure the actual levels of awareness for each role on the process model using the awareness model. In order to measure this level of awareness the actor must be exposed to all the objects on the ERP process model, and be asked to identify those objects that s/he is aware of. Selected pool of objects are then used by an awareness model in order to arrive at a number reflecting the actual level of awareness associated with that role.

STEP 3: The actor's actual level of awareness is then is compared against the required level of awareness; the latter is a parameter, provided by the task that the actor performs within the process. The difference between these two levels of awareness constitutes the collaborative requirement of the actor for that particular task. Factors that affect the required level of awareness of a task include organisational culture, and the nature of task itself. Without possessing such awareness level the actor will not be able to collaborate with others optimally. A comparison between the actual level of awareness of the actor and the required level of awareness of the task will result in one of the following two outcomes:

1. The task's required level of awareness is either equal to, or less than, the role's actual level of awareness. This indicates that the role is qualified, or has sufficient level of awareness for taking up the task, and the OOAB framework cannot enhance collaboration any further.
2. The task's required level of awareness exceeds the role's actual level of awareness. This indicates potential for enhancing collaboration. To do so it will be necessary to put the missing objects within the focus of the actor in a way that s/he can perceive these objects, receive required awareness, and perform that particular task successfully. This will require additional resources in order to enhance the actor's awareness level. These required resources may include one or more of process resources, collaborative resources, and other communication resources, for example resources that provide awareness about other roles and other tasks within the ERP process.

IMPLEMENTATION ISSUES

One method for integration of the OOAB framework with the existing ERP systems is by developing an organisational infrastructure that provides business intelligence to the users of the ERP system by maintaining contextual knowledge that these users/actors require for effective collaboration within the ERP process. The writer is in the process of developing an expert system that provides expert advice for answering the following two specific questions:

(i) In terms of awareness requirements, is an actor capable of performing certain tasks within the ERP process?
(ii) If not, what objects need to be put within his/her focus in order to enable the actor to perform the task properly?

The ERP collaborative process of Figure 1 consists of 15 objects, including two roles, six subprocesses/tasks, six role artefacts and one task artefact. Within each of these objects is encapsulated all relevant contextual knowledge as well as pointers to relevant objects as determined by the process map. Each task possesses a set of attributes and relevant methods; and each method consists of a set of steps that corresponds to codes describing a codified knowledge. These attributes will indicate to which subprocess the task belongs to. This will enable an actor to play various roles within different subprocesses without being permanently linked to a specific subprocess, a factor that can remove some complexities in existing ERP implementations.

FUTURE TRENDS AND CONCLUSION

It was shown that the OOAB methodology can enhance existing ERP systems by formally adhering to a framework that facilitates knowledge sharing among various actors in ERP processes in the following ways: This article introduces a novel concept for measuring collaboration in ERP processes; a concept that is non-existent in current ERP literature. This measure is based on a conceptual model of collaboration that can be directly transformed into an object model. This measure is used to determine required level of awareness of various actors within the ERP process. As far as the author is aware, this is a novel concept that results in treating the organisation-wide ERP process as a single collaborative process that consists of multiple sub-processes, called tasks, that are linked through formal artefacts. These artefacts utilise certain resources in order to carry/deliver contextual collaboration knowledge that is required for effective collaboration. Knowledge sharing occurs through various interaction acts such as exchanging the artefacts, creation of artefacts, and updating artefacts. As a result of its object-orientation and collaborativeness of the ERP process, and contrary to the existing process-based ERP frameworks that assume fixed roles within the ERP process, the OOAB framework enables an actor to assume multiple roles within different ERP subprocesses. This will provide another avenue for sharing contextual knowledge for sub-processes/tasks. The interdependency issue among various subprocesses is also simplified by encapsulating this knowledge within the task objects and relevant artefacts.

By reducing granularity of the collaborative process model, the same model representation can be applied for internal representation of the subprocesses, allowing smooth transition of modelling components from subprocesses to the ERP process, and vice-versa. This in turn will reduce much of the existing complexities in designing ERP systems where the system is permanently responsible to maintain such linkages, rather than delegating such responsibility to various objects within the system.

Work is in progress for incorporating communication and coordination dimensions to the existing collaboration dimesion in order to provide complete analysis of groupware systems that maintain awareness requirements of the actors within the ERP processes (Daneshgar et al., 2004; Sundarraj et al., 2002).

REFERENCES

Abramson, B.D. (1998). Translating nations: Actor-network theory in/and Canada. *Canadian Review of Sociology and Anthropology, 35*(1), 1-20.

Daneshgar, F. (2003). Context management of ERP processes in virtual communities. In G. Grant (Ed.), *ERP & datawarehousing in organizations: Issues and challenges* (pp. 119-130). Hershey, PA: IRM Press.

Daneshgar, F., Ray, P., Rahbi, F., & Godar, C. (2004). Knowledge sharing infrastructures for teams within virtual communities. In M. Fong (Ed.), *e-Collaborations and virtual organizations.* Hershey, PA: IRM Press.

Kaptelinin, V., Kuutti, K., & Bannon, L. (1995). Activity theory: Basic concepts and applications. In Blumenthal et al. (Eds.), *Human-computer interaction. Lecture Notes in Computer Science.* Springer.

Object Management Group. (2001). *OMG Unified Modeling Language Specification – Version 1.4* (pp. 3-100).

Podolsky, M. (1998, July 20–24). An integrated approach to object-oriented modeling of business processes. *ECOOP'98 Workshop on Object-Oriented Business Process Modeling,* Brussels, Belgium.

Sundarraj, R.P., & Sarkis, J. (2002). Implementation management of an e-commerce-enabled enterprise information systems: A case study at Texas Instruments. In L. Hossein, J.D. Patrick & M.A. Rashid (Eds.), *Enterprise resource planning: Global opportunities & challenges* (pp. 133-148). Hershey, PA: IRM Press.

Turban, E., & Aarons, J.E. (2001). *Decision support systems and intelligent systems*. NJ: Prentice Hall International Inc.

Van Stijn, E., & Wensley, A. (2001). Organizational memory and the completeness of process modelling in ERP systems: Some concerns, methods and directions for future research. *Business Process Management Journal - Special Issue on Enterprise Resource Planning, 7*(3).

KEY TERMS

Action: A sequence of goal-directed steps.

Actual Level of Awareness: The awareness that a role actually possesses within the ERP process. Actual awareness is represented by an integer number ranging from zero to four, representing various levels of awareness. Actual awareness is a property of an actor who performs one or more roles within the ERP process.

Awareness: A specialised knowledge about the objects that leads an actor to an understanding of various aspects of the ERP collaborative process. It is defined and measured in terms of the semantic concepts (*task, role, process resource,* and *collaborative resource*) used in the map.

Awareness Model: A model that represents various levels of awareness. Level-0 awareness consists of the concepts that lead an actor to knowledge about all the tasks that an actor performs within the process.

Example: In Figure 1, level-0 awareness for the role "R1" is a sub-graph that consists of the tasks "FAT1," "FAT2," and "OPT1," as well as the process resources shown by thick lines connecting "R1" to these tasks. Level-1 awareness is a subgraph that consists of the R1's level-0 awareness subgraph, plus awareness about the concepts that leads an actor to knowledge about other related roles within the process. In Figure 1, level-1 awareness for the role "R1" is the sum of its level-0 awareness subgraph, plus one collaborative resource linking "FAT2" to "FRT1," plus "FRT1" itself, plus the process resources shown by thick lines connecting "FRT1" to "RT2," plus "RT2" itself. A role's level-2 awareness is its level-1 awareness, plus an awareness about all other (or, the rest of) roles within the process. Level-2 awareness is knowledge about the human boundary of the process. In Figure 1, there are no other roles that have not been known to the R1 already, and therefore the R1's level-2 and higher levels of awareness are irrelevant and identical to its level-1 awareness. However, for the sake of completeness, their definitions are presented.

A role's level-3 awareness is its level-2 awareness, plus awareness about all the interactions (represented by the process resources used/shared) that occur between any two roles within the process. And finally, level-4 awareness is the highest level of awareness that a role can have in any ERP process. It is defined as the knowledge about how everything fits together to form the ERP process. In other words, having this level of awareness will bring all the remaining concepts used in the ERP process model of Figure 1 within the focus of the role.

Collaborative Resources: An object representing a resource used/shared/exchanged by a pair of collaborating roles in order to perform certain simple tasks in collaboration with one another.

Process Resource: An object that represents a resource used by a role in order to perform a task in isolation from other tasks.

Required Level of Awareness: A property of a task. It represents the expected awareness from any actor who performs the task. Its value also ranges from 0 to 4.

Role: A set of norms expressed in terms of obligations, privileges, and rights assigned to an actor.

Task: An object with a set of attributes and actions to achieve a specific process goal using certain resource called *process resource.*

Contextual Metadata for Document Databases

Virpi Lyytikäinen
University of Jyväskylä, Finland

Pasi Tiitinen
University of Jyväskylä, Finland

Airi Salminen
University of Jyväskylä, Finland

INTRODUCTION

Metadata has always been an important means to support accessibility of information in document collections. Metadata can be, for example, bibliographic data manually created for each document at the time of document storage. The indexes created by Web search engines serve as metadata about the content of Web documents. In the semantic Web solutions, ontologies are used to store semantic metadata (Berners-Lee et al., 2001). Attaching a common ontology to a set of heterogeneous document databases may be used to support data integration. Creation of the common ontology requires profound understanding of the concepts used in the databases. It is a demanding task, especially in cases where the content of the documents is written in various natural languages. In this chapter, we propose the use of contextual metadata as another means to add meaning to document collections, and as a way to support data integration. By *contextual metadata*, we refer to data about the context where documents are created (e.g., data about business processes, organizations involved, and document types). We will restrict our discussion to contextual metadata on the level of collections, leaving metadata about particular document instances out of the discussion. Thus, the contextual metadata can be created, like ontologies, independently of the creation of instances in the databases.

BACKGROUND

European legal databases offer an example of a new means for data integration. Due to the development towards European integration, the legal information needed in a particular situation often concerns not only regulations of the home country, but also European Union regulations and those in foreign countries. The information may be scattered in various European legal databases with varying retrieval techniques. The databases are organized in different ways, and their content is written in different languages. Differences in legal systems aggravate the retrieval problems. Similar problems and needs may be identified in other types of environments, as well. For example, the information needed during manufacturing processes may be created in a number of cooperating organizations and scattered in heterogeneous intranet and extranet repositories.

Where the creation of ontologies requires analysis and a description of concepts used on a domain, creation of contextual metadata requires analysis of the environment where documents are created. We will first describe methods for collecting contextual metadata, and then we will show how the metadata can be visualized to users in a graphical interface. The graphical models providing contextual metadata aid the users in understanding the context of documents and in locating information from correct sources.

COLLECTING THE CONTEXTUAL METADATA

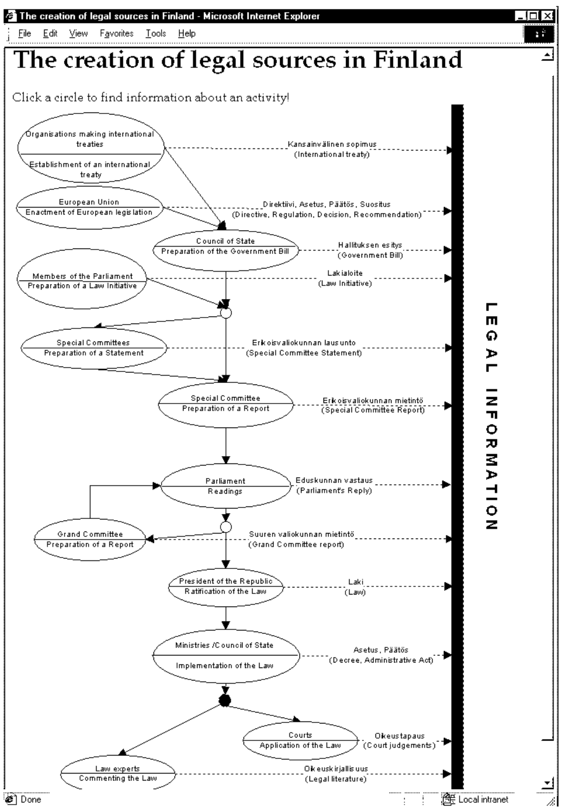
A special methodology called Rakenteisten Asiakirja Standardien KEhittäminen (RASKE) meaning “Development of standards for structured documents,” has been developed for analyzing and describing document management environments. The RASKE method was tailored from more general information systems analysis methods and tested in a project where document management practices in the Finnish Parliament and ministries were redesigned (Salminen, 2003; Salminen et al., 2000; Salminen et al., 2001). The major practical result of the project is that the Finnish Parliament currently creates all Parliamentary documents in SGML (Standard Generalized Markup Language) format (Goldfarb, 1990). The RASKE methodology offers tools for gathering and representing contextual and structural metadata about documents.

The modeling techniques of RASKE are intended for describing documents, processes where documents are created and manipulated, and actors and their roles in the activities of the processes. The most important origins of the techniques are in the object-based modeling methodology of Shlaer and Mellor (1992), in Information Control Nets (Ellis, 1979), and in the document structure specification methods of SGML standard and elm graphs (Goldfarb, 1990; Maler & El Andaloussi, 1996). From the various models created in a RASKE analysis, three kinds of models provide important contextual information. The *organizational framework model* describes the organizations involved and their roles on the analysis domain. The activities of the domain are described by process models. The *document output model* is a process model showing the activities and actors of a process together with the document types produced in the activities. The relationships of different document types are described by a *document-relationship diagram.* In addition to a set of models, the RASKE methodology produces textual descriptions (e.g., about document types and actor roles).

VISUALIZATION OF THE CONTEXTUAL METADATA

To support users in retrieving information created on a complex domain, we suggest using the models created

Figure 1. Process view describing the Finnish legal system



during the document analysis as metadata visualizations. The visualization of information itself is not a novel issue. Graphical models have been used, for example, in software engineering (e.g., UML) (Booch et al., 1999), business process redesign (Abeysinghe & Phalp, 1997), workflow management systems (Ellis, 1983), and computer supported cooperative work systems (Sarin et al., 1991). The visualizations we propose add, however, a new perspective for the uses of visual presentations. Our solution for visualizing contextual metadata by graphical models is intended to be used as a retrieval interface for distributed document repositories. The solution has been tested in the European User Views to Legislative Information in Structured Form (EULEGIS) project, whose main goal was to improve access to European legal information (Lyytikäinen et al., 2000). The EULEGIS prototype system demonstrated how a single-point access to various European legal databases could be implemented.

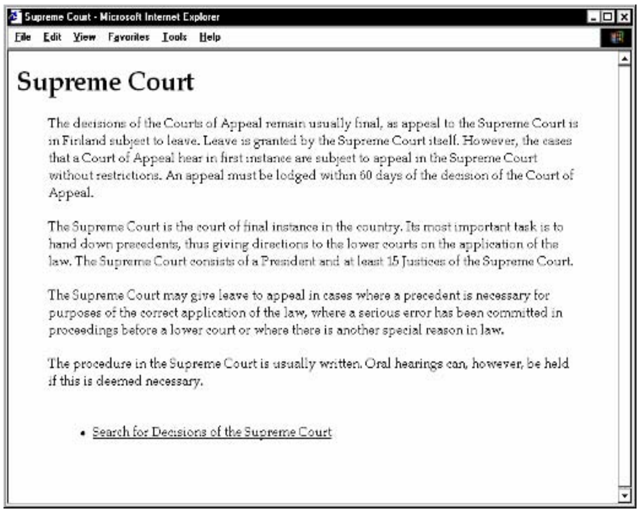
The three models discussed previously provide three different views to the information on the domain. The *actor view* describes the most significant actors on the domain; the *information source view* shows the different kinds of documents and their relationships; and the *process view* describes activities related to the information production process. In order to enable querying the document databases, links from each graphical element of the view to the query form of the relevant database can be defined.

The process view describing the Finnish legal system is shown in Figure 1. The graphical model originates from the document output model created during the RASKE analysis. The main activities of the process are depicted by circles. Each circle shows both the name of the activity and the actor(s) performing the activity. The order of activities is expressed by solid arrows between the activities. The information flow from an activity to the collective document repository is shown by a dashed arrow labelled with the type of the documents created in the activity. Additional textual information is linked to the graphical symbols. By clicking, for example, an activity symbol in the graph, the user can obtain more information about the activity. From this additional information, a link leads to a search form by which the user can search for the documents that originate from the selected activity. Similarily, the user can retrieve information about a particular actor and a search form for accessing documents created by the actor (see Figure 2).

The three views were implemented and tested in the EULEGIS prototype. The feedback from the users was encouraging. The views were seen as a good tool to become acquainted with a previously unfamiliar legal system and to limit the scope of a query statement.

For the dynamic generation of graphical views, the data pertaining to the models has to be formally defined.

Figure 2. Additional information about the Supreme Court in Finland



For this purpose, we have described the models in an XML (Extensible Markup Language) (Bray et al., 2000) document type definition (DTD). In the DTD, the data sources, documents, actors, and processes of a domain are defined, together with their relationships to each other (Lyytikäinen et al., 2000). The DTD is designed in a way that it enables the text in the models to be described in several languages.

FUTURE TRENDS

In the global and organizational networks, documents are increasingly created in complex interorganizational processes. At the same time, the improved authoring and communication technology increases document production. In Europe, the expansion of the European Union is creating a situation where organizations and citizens all over the Europe have a need for legal information, not only concerning their own country, but also in relation to foreign countries and the European Union. A similar trend also can be seen in the business world where corporations operate in global markets, and information related to a business process is produced in different countries.

To assure the accessibility of information created in complicated processes, different kinds of techniques are needed. The contextual metadata approach is a way to add meaning to document repositories. Another approach for the same purpose is to build ontologies describing concepts and their relationships on a domain. The techniques and methods for building and sharing ontologies are an active area of the current semantic Web research (Fensel et al., 2000).

CONCLUSION

On various domains, information needed by people is increasingly available in large and heterogeneous web repositories. The lack of the knowledge about the repositories, differences in the organization and user interfaces of the repositories, and the lack of knowledge about the context where information is created hinder the use of the repositories. In the article, we discussed how metadata about the context of the documents in the repositories can be used to improve information retrieval. Graphical information models were adopted to visualize the collected contextual metadata: information sources, their relationships to each other, actors, and activities in the documentation creation processes. The graphical interface has been tested in a prototype system. The test users regarded the interface as a valuable tool for retrieving information from heterogeneous databases. More systematic user testing is an important area of future research.

REFERENCES

Abeysinghe, G., & Phalp, K. (1997). Combining process modeling methods. *Information and Software Technology, 39*(2), 107-124.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic Web. *Scientific American, 284*(5).

Booch, G., Rumbaugh, J., & Jacobson, I. (1999). *The unified modeling language user guide*. Reading, MA: Addison-Wesley.

Bray, T., Paoli, J., Sperberg-McQueen, C.M., & Maler, E. (2000). *Extensible markup language (XML) 1.0 (2nd Edition).* Retrieved April l, 2003, from *http://www.w3.org/TR/2000/REC-xml-20001006*

Ellis, C.A. (1979). Information control nets: A mathematical model of office information flow. *Proceedings of the Conference on Simulation, Measurement and Modeling of Computer Systems* (pp. 225-239).

Ellis, C.A. (1983, September 19-23). Formal and informal models of office activity. In R.E.A. Mason (Ed.), *Proceedings of the IFIP 9th World Computer Congress*, Paris, France.

Fensel, D., Horrocks, I., van Harmelen, F., Decker, S., Erdmann, M., & Klein, M. (2000, October 2-6). OIL in a nutshell. *Proceedings of the European Knowledge Acquisition Conference (EKAW-2000)*, (pp. 1-16), Juan-les-Pins, France.

Goldfarb, C.F. (1990). *The SGML handbook*. Oxford, UK: Oxford University Press.

Lyytikäinen, V., Tiitinen, P., & Salminen, A. (2000). Graphical information models as interfaces for Web document repositories. In V. Di Gesú, S. Levialdi, & L. Tarantino (Eds.), *Proceedings of the Working Conference on Advanced Visual Interface* (pp. 261-265), New York.

Lyytikäinen, V., Tiitinen, P., Salminen, A., Mercier, L., & Vidick, J.-L. (2000). Visualizing legal systems for information retrieval. In M. Khosrowpour (Ed.), *Challenges of information technology management in the 21st century. Proceedings of 2000 Information Resources Management Association International Conference* (pp. 245-249). Hersley, PA.

Maler, E., & El Andaloussi, J. (1996). *Developing SGML DTDs. From text to model to markup*. Upper Saddle River, NJ: Prentice Hall.

Salminen, A. (2003). Document analysis methods. In C.L. Bernie (Ed.), *Encyclopedia of library and information science, second edition, revised and expanded* (pp. 916-927). New York: Marcel Dekker, Inc.

Salminen, A., Lyytikäinen, V., & Tiitinen, P. (2000). Putting documents into their work context in document analysis. *Information Processing & Management, 36*(4), 623-641.

Salminen, A., Lyytikäinen, V., Tiitinen, P., & Mustajärvi, O. (2001). Experiences of SGML standardization: The case of the Finnish legislative documents. In J.R.H. Sprague (Ed.), *Proceedings of the Thirty-Fourth Hawaii International Conference on System Sciences*. Los Alamitos,CA: IEEE Computer Society.

Sarin, S.K., Abbott, K.R., & McCarthy, D.R. (1991). A process model and system for supporting collaborative work. *SIGOIS Bulletin, 12*(2,3), 213-224.

Shlaer, S., & Mellor, S.J. (1992). *Object lifecycles: Modeling the world in states*. Englewood Cliffs, NJ: Yourdon Press.

UML-1.5. (2003). *OMG unified modeling language specification. Version 1.5, March 2003*. Retrieved December 12, 2003, from *http://www.omg.org/docs/formal/03-03-01.pdf*

Yourdon, E. (1989). *Modern structured analysis*. Englewood Cliffs, NJ: Prentice-Hall.

KEY TERMS

Contextual Metadata: Information describing the context where the object of the contextual metadata is created. The contextual metadata can cover, for example, information about the producers and production processes of documents.

Document Database: A collection of documents associated with a system to manage the documents and their content.

Document Type Definition (DTD): The capability of XML to specify constraints for a class of documents. A DTD defines element and attribute names and the hierarchic structure of the elements for the documents of a class.

Extensible Markup Language, XML: A markup language for representing information as structured documents. Developed from an earlier language called SGML (Standard Generalized Markup Language) for the purposes of information management on the Internet.

Legal System: The set of legal rules governing an entity. The entity in question can be, for example, a state, a group of states, an international organisation, a region, or a city.

Metadata: Data about data. In the context of document databases, metadata is data carrying information about documents, document collections, or parts of documents.

Semantic Web: An extension of the current Web where information resources are attached with metadata to support people and computers as they work in cooperation. The metadata is intended to give well-defined meaning to resources and to support automated reasoning about the meaning and trustworthiness of resources.

C

Contingency Theory, Agent-Based Systems, and a Virtual Advisor

John R. Durrett
Texas Tech University, USA

Lisa Burnell
Texas Christian University, USA

John W. Priest
University of Texas at Arlington, USA

INTRODUCTION

In this article, we investigate the potential of using a synthesis of organizational research, traditional systems analysis techniques, and agent-based computing in the creation and teaching of a Contingency Theoretic Systems Design (CTSD) model. To facilitate understanding of the new design model, we briefly provide the necessary background of these diverse fields, describe the conceptualization used in the integration process, and give a non-technical overview of an example implementation in a very complex design environment. The example utilized in this article is a Smart Agent Resource for Advising (SARA), an intelligent multi-agent advising system for college students. To test all of the potential of our CTSD model, we created SARA utilizing a distributed instructional model in a multi-university, multi-disciplinary cooperative design process.

Just as a dynamic task environment forces an organization to compress its management structure and to outsource non-core activities in order to become flexible, a dynamic software development environment forces designers to create modular software. Until now, cooperative development paradigms were too complex to facilitate inter-organizational cooperative development efforts. With the increasing popularity of standards-based Web services, the development of pervasive computing technologies, and the advent of more powerful rapid application development languages and IDEs, this limitation has been removed. Our purpose in this research is twofold: first, to test the viability of using Contingency Theory (CT), a sub-discipline of Management Organizational Theory (OT), in an agent-based system; and second, to use these new technologies in creating a distributed instructional model that will allow students to interact with others in diverse educational environments. As an example implementation, we create a virtual advisor that will facilitate student advising in distributed environments.

In the following sections, we outline the background theories involved in the conceptualization of our design model. We start with the shifts in systems design techniques and how CT can be applied to them and to various Multi-Agent Systems (MAS) to allow Contingency Theoretic Systems Design (CTSD). Once the necessary background is in place, we briefly discuss our new eLearning approach to cooperative distributed education. Finally, the structure of the SARA is discussed.

BACKGROUND

Multi-Agent Systems

Agents and communication protocols form the basic components of a multi-agent system. Agents exchange messages according to a protocol of expected messages delivered in a communication language in which the message content and format adhere to a shared standard. Individual agents make decisions, which may include contacting other agents for information, and perform processing to satisfy their goals

An agent is commonly defined as a program or collection of programs that lives for some purpose in a dynamic environment and can make decisions to perform actions to achieve its goals. In other words, agents are goal-based programs that must deal with changing access to resources, yet run continuously. Like the best administrative assistants, agents know and adapt to their master. Individual agents may be conceptualized as having beliefs, desires, and intentions that can communicate with other agents to satisfy their goals. Multi-agent systems are those in which multiple agents (usually) cooperate to perform some task. Agents may be independently developed and allow the decomposition of a complex task into a collection of interacting agents that together solve some problem. It is not necessary that an individual agent "understand" the overall system goals or structure.

Agent communication can be viewed at four distinct levels. The first level is the expected protocol for exchanging sequences of messages, like a script. For example, when negotiating, the parties expect bids to be offered, rejected, and counter-offered. The second level relates to the content or meaning of the messages. To enable inter-agent communication, an ontology is created. Examples of such concepts are things, events, and relationships. At the third level, a representation language defines the syntax for structuring the messages; The Knowledge Interchange Format (KIF) (Gensereth & Fikes, 1992) is one example. At the fourth level, an agent communication language (ACL) such as the Knowledge Query and Manipulation Language (KQML) or the Foundation for Intelligent Physical Agents (FIPA) ACL (Labrou, Finin, & Peng, 1999), defines message formats and message delivery. An example KQML message, in Sandia Lab's Java Expert System Shell (JESS) (Owen, 2004), that shows how an agent registers a service is shown below:

(register :sender student :receiver advisor :reply-with msg1 :language JESS :ontology SARA :content '(MajorCourses:Compliance Check Hours))

Just as human systems created to achieve complex goals are conceived of as organizations, multi-agent systems can be conceptualized as "organizations of agents". Individual components, whether human employees or software agents, need to be managed, guided toward a constructive goal, and coordinated toward the completion of the necessary individual tasks. In "empowered organizations", lower-level employees have the knowledge and authority to perform many tasks without the intervention of superiors. This conceptualization allows us to use well-established research from management organization theory (and Contingency Theory in particular) in creating guidelines for the design of agent-based systems.

ORGANIZATIONAL THEORY (OT)

While much of the background concerning OT is explained in the main chapter below, the following is a brief overview of the relevant research trends. OT examines an organization's structure, constituencies, processes, and operational results in an effort to understand the relationships involved in creating effective and efficient systems. A major division of OT, Contingency Theory (CT), postulates that no organization operates without constraints from environmental, personnel, technological, and informational influences (Andres & Zmud, 2001). This relationship is explained by the information processing theory (IPT) (Galbraith, 1973). IPT postulates that the more heterogeneous, unpredictable, and dependent upon other environmental resources a task is, the greater the information processing that the organization must be able to do in order to successfully accomplish it. As complexity and unpredictability increase, uncertainty increases due to incomplete information. As diversity of processes or outputs increases, inter-process coordination requirements and system complexity increase. As uncertainty increases, information-processing requirements increase. The basic premise of IPT is that the greater the complexity and uncertainty in the tasks in an organizational system, the greater the amount of information that the system must process (Galbraith, Downey, & Kates, 2001). A basic premise of our research is that this relationship is also true for information systems (Avgerou, 2001).

MAIN THRUST OF THE ARTICLE

Multi-Agent System Architectures Using CTSD

Contingency-theoretic system development (CTSD) adapts CT and IPT to the development and maintenance of software systems (Burnell, Durrett, Priest et al., 2002; Durrett, Burnell, & Priest, 2001, 2003). A business can organize employees in a number of different ways, for example by function or by project, and reorganize as the business environment changes. Software systems can benefit from this flexibility as well. The CTSD design approach is focused on design for maintainability, a crucial requirement for complex, dynamic systems.

Agent-based architectures are a means for structuring software systems that adhere to Contingency Theoretic principles. Each agent is viewed as an employee that has specific capabilities, responsibilities, and knowledge within an organization. Agents, like employees, are grouped into departments, as needed, to best satisfy the goals of the organization. Agents can communicate peer-to-peer within and across departments, and manager agents resolve conflicts and make resource allocation decisions.

Tightly interrelated tasks are grouped into one or more agents. Each of these groupings is referred to as a "software team", and parallels a department of employees that perform roughly equivalent jobs. For example, a set of agents that each handle one type of course requirement (e.g., lab, art appreciation) may be grouped into a team, where communication can occur quickly between these agents and with a "manager" agent that can resolve conflicts, exceptions, and course-independent tasks. An example agent in our system is encoded using JESS rules to check that student preferences (e.g., for afternoon courses) and constraints (e.g., no more than 12 hours per semester) are satisfied. Another agent offers heuristic advice as an actual advisor might. For example, a student may be able to enroll in 17 hours of math and science courses, but this may be strongly advised against, depending on the student's GPA and perhaps other factors.

Each agent in a multi-agent architecture has specific tasks to perform and communications requirements. Once an ontology and agent communication language has been specified, agents can be designed independently and integrated into the system to progressively add capabilities. Using CTSD principles, tasks that are dynamic and shared are grouped into support agents to enhance maintainability of the system. The primary architectural decision is to separate knowledge based on two criteria: the degree of dynamicism and the degree of complexity. Dynamicism is exemplified by how frequently knowledge is expected to change over time, while complexity determines how abstract that knowledge is. A simple example describing the latter is the fact that "a student must be a junior with a GPA greater than 3.0 in all math and science courses". Dynamicism within the SARA domain was determined by advisor interviews, and analysis of degree requirements changes over time at three universities, both large and small and private and public.

Using the above CTSD conceptualizations in previous research (Burnell, Priest, & Durrett, 2002, 2003; Durrett, Burnell, & Priest, 2000), we have developed and tested the following CTSD guidelines for creating MAS:

1. *Describe business activity and identify tasks*: Allow management and developers to refine the overall purpose of the software being designed.
2. *Determine task predictability*: Since a basic premise of CT is that the control structure of a business process must match the environment in which it operates, we must identify the predictability of each task.
3. *Assign tasks to employees*: Once the level of predictability has been estimated for each task, the granularity of the employees being created can be determined and component designs finalized.
4. *Group employees into teams*: As with human organizations, our employees can be grouped along any of several dimensions, including task, workflow, product, manager, or communication requirements, as required by the operating environment.
5. *Identify communications needs:* Once teams are determined, the communication requirements of individual employees, and of teams, can be determined.
6. *Construct management groups*: In software systems operating in a dynamic environment, management is required only when employees are unable to handle events

.

We show the application of these guidelines in a distributed instructional environment in a multi-university, multi-disciplinary cooperative design process.

Multi-Agent System Architectures Using CTSD

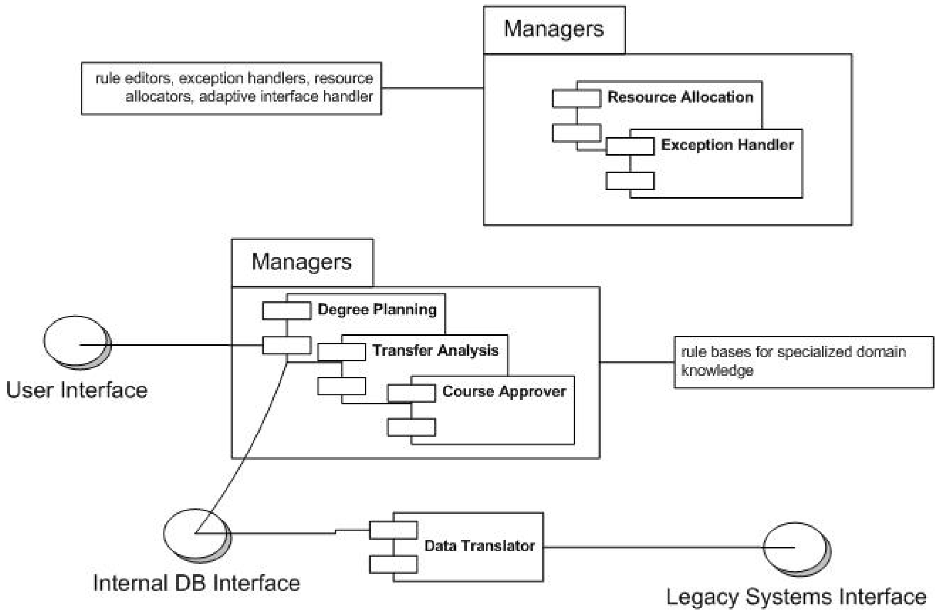
Our example implementation of the above CTSD guidelines is SARA, an interactive tool intended to aid in the planning, scheduling, and advising process to complete a college student's degree. SARA, along with the instructional environment in which it was created, is described very briefly next; for more detailed information on SARA or our distributed instructional model please see any of the Burnell, Durrett, or Priest references in the reference section.

As described previously, CTSD is most effective in complex dynamic environments. To create this environment, SARA was designed in a collaborative education project among three major universities, the University of Texas at Arlington (UTA), Texas Christian University in Fort Worth (TCU), and Texas Tech University in Lubbock (TTU). Teams of students at UTA are in industrial engineering and software design courses. The TCU students are in computer science and have had coursework in artificial intelligence, and the TTU teams are in courses that study application servers and server-based JAVA databases. In creating SARA, responsibilities have been purposefully segregated to create the necessity of interdisciplinary cooperation. Current trends in industry are toward outsourcing in many major Fortune 500 companies (Fox, 2004). In order to do this effectively, the designer and the coders must communicate in an online forum and adapt to rapidly changing conditions. Therefore, we have attempted to emulate this CTSD type environment through our interdependent classes. Also, previous research has shown that a major hurdle to software development is the communication among the programmers, the users, the executives, and the domain experts. A major component of successful software development is overcoming this communication hurdle. We have attempted to emulate these trends in the design of SARA.

Design of SARA

SARA gives students the resources and knowledge to troubleshoot issues with their current degree plan. A profile of student preferences and a transcript of completed courses allow the system to generate customized degree plans. Students can also manually customize degree plans and create schedules course-by-course that are checked for requirements compliance. Errors in

Figure 1. Prototype architecture (from Priest et al., 2002)



plans and schedules are explained so that the student can make corrections. Advisors can review degree plans on the system and send comments back to the students, without their having to physically meet. Thus, some of the most time consuming tasks in advising are eliminated (Priest, Burnell, & Durrett, 2002).

The current prototype of SARA that is depicted in the following diagrams and text is the work of several semesters work at all 3 universities. As a result of these cooperative efforts, a basic system prototype (shown in Figure 1) has been developed, initial system databases have been designed and implemented (shown in Figure 2), and smart user interfaces that utilize MAS (shown in Figure 3) were created. The user interfaces were designed primarily by the TCU teams, the basic CTSD-based analysis and thus overall system requirements were created by the UTA students, and the backend database design and application server support were provided by the TTU students.

The architecture depicted previously (in Figure 1) follows the CTSD guidelines discussed. Given the dynamic, complex environment in which the system will operate, we have segregated tasks into individual software teams and empowered the constituent agents with as much decision making ability as possible. Exceptional situations are handled by "manager agents" and by other exception handlers.

The prototype SARA database module depicted next (in Figure 2) was created to provide the flexibility required by the MAS system architecture. Initial implementations are on MySQL, and migration to IBM DB2 to provide for more system automation is planned.

The MAS-based user interfaces (one example of which is depicted in Figure 3) were created using data from TCU's course schedule. They allow students to do most of the routine work in degree planning, freeing up advising time for more meaningful discussions. Once proposed plans are created, they can be evaluated and approved off-line.

FUTURE TRENDS

Our future plans include broadening the scope of the project just described using an open source software development model. Our intent is to research the potential in open source design models for relieving some of the coordination issues inherent in a resource limited project such as ours. We hope that this new model will relieve some of the faculty coordination required and still allow (or force) students to cooperate with each other even though in different disciplines and locations.

CONCLUSION

In the project just described, we have developed a multi-agent system for college advising using the contingency-theoretic system development (CTSD) process. Multi-

Figure 2. SARA database entity diagram

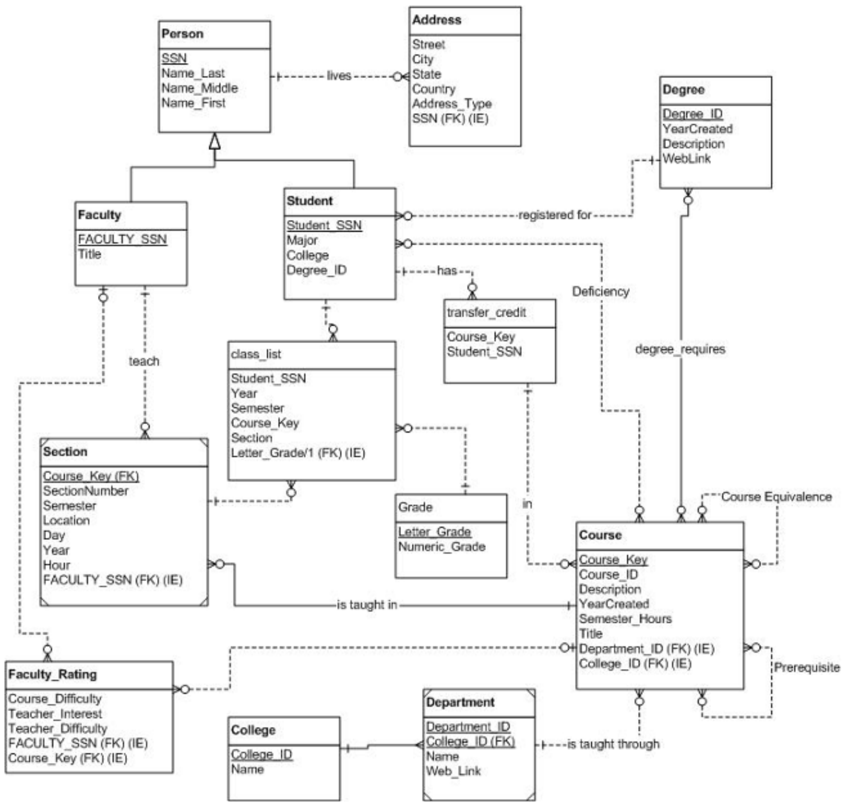
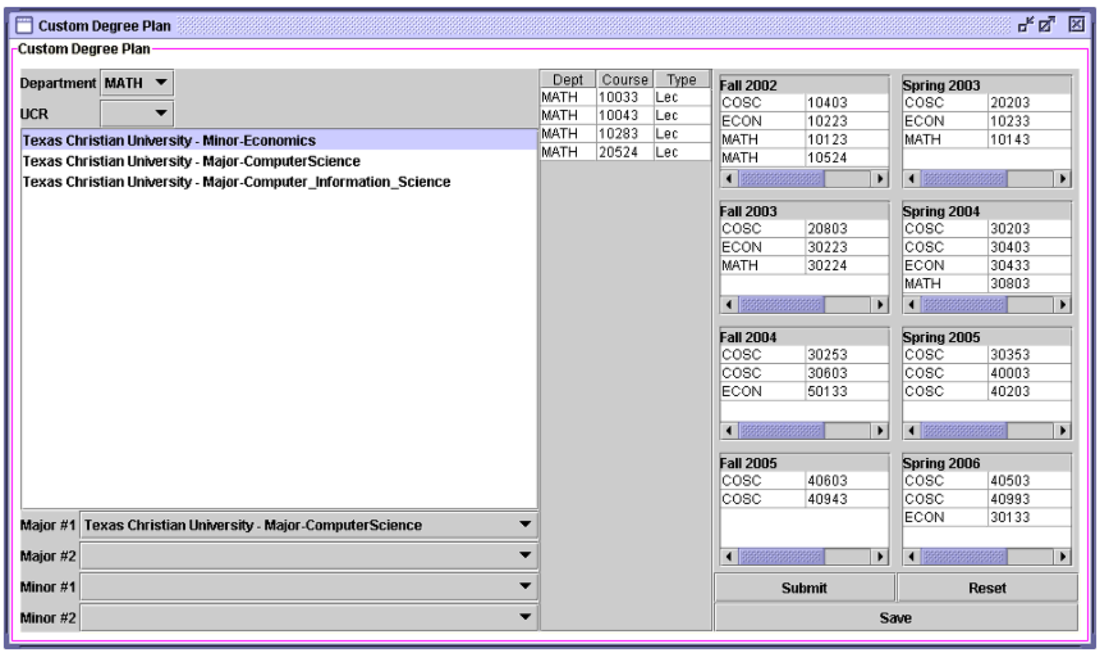


Figure 3. Custom degree planner interface



disciplinary teams of students at three universities were employed to create the system. Once the ontology and agent communication language are defined, individual agents are constructed incrementally. The resulting system is a hierarchy of intelligent agents that together provide support for course scheduling and degree planning and that are adaptable to changing environments. The CTSD process applied to the creation of intelligent multi-agent systems results in maintainable systems operating in complex, dynamic domains. As such domains become increasingly automated, training and support tools for distributed CTSD will be needed. We have created an approach for such training that has been successfully applied to SARA and other projects over the last three years. Support tools will need to be created that capture not only the system design, but provide knowledge-based analysis of designs based on the principles we have defined.

REFERENCES

Andres, H.P., & Zmud, R.W. (2001). A contingency approach to software project coordination. *Journal of Management Information Systems, 18*(3), 41-70.

Avgerou, C. (2001). The significance of context in information systems and organizational change. *Information Systems Journal, 11*, 43-63.

Burnell, L.J., Durrett, J.R., Priest, J.W., et al. (2002). A business rules approach to departmental advising. Paper presented at the *Proceedings of the 15th International Florida Artificial Intelligence Research Society Conference (FLAIRS-2002)*, Pensacola Beach, FL.

Burnell, L.J., Priest, J.W., & Durrett, J.R. (2002). Teaching distributed collaborative software development. *IEEE Software, 19*(5), 86-93.

Burnell, L.J., Priest, J.W., & Durrett, J.R. (2003). Assessment of a resource limited distributed multidisciplinary process for teaching software development in a university environment. *ACM Inroads*.

Durrett, J.R., Burnell, L., & Priest, J. (2001, July 29-August 2). An organizational behavior approach for managing change in information systems. Paper presented at the *Proceedings of PICMET*, Portland, Oregon.

Durrett, J.R., Burnell, L.J., & Priest, J.W. (2000, November 2-8). Contingency theoretic methodology for agent-based, web-oriented manufacturing systems. Paper presented at the *SPIE: Photonics East 2000*, Boston, MA USA.

Durrett, J.R., Burnell, L.J., & Priest, J.W. (2003, December). A hybrid analysis and architectural design method for development of smart home components. *IEEE Personal Communications*, 2-9.

Fox, S. et al. (2004, March 8). Offshoring. *InfoWorld, 26*, Special Issue.

Galbraith, J., Downey, D., & Kates, A. (2001). *Designing dynamic organizations: A hands-on guide for leaders at all levels*. New York, NY: Amacom.

Galbraith, J.R. (1973). *Designing complex organizations*. Reading, MA: Addison-Wesley.

Gensereth, M.R., & Fikes, R.E. (1992). *Knowledge interchange format, version 3.0 reference manual*. Santa Clara, CA: Stanford University.

Labrou, Y., Finin, T., & Peng, Y. (1999). The current landscape of agent communication languages. *Intelligent Systems: IEEE Computer Society, 14*(2), 45-52.

Owen, J. (2004, March 15). Budget-minded BRMS: JESS and OPSJ are faster, cheaper, and harder to use. *InfoWorld, 26*, 24-26.

Priest, J., Burnell, L., & Durrett, J.R. (2002, May 19-22). SARA: Smart, agent-based resource for virtual advising. Paper presented at the *International Resource Managers Association*, Seattle, Washington.

KEY TERMS

Contingency Theoretic Software Development (CTSD): A new model for MAS design using tenets from CT and IPT. The CTSD design approach is focused on design for maintainability, a crucial requirement for complex, dynamic systems.

Contingency Theory (CT): A research branch of organizational theory that suggests that an organization's structure reflects its adaptation to the environment in which it operates. Hierarchical organizations operate best in stable, simple environments while flat, team-based organizations are better adapted to dynamic, complex task environments.

E-Learning: Any form of education or training that utilizes online media and remote connectivity for all or part of its curricula. This model includes both purely online courses and those in brick-and-mortar universities facilitated by email, the Internet, newsgroups, or chat.

Information Processing Theory (IPT): An explanation for the organization structure-environment relationship suggested by CT. IPT suggests that the information processing requirements dictated through interactions with the environment force certain structures in order to be efficient and effective.

Multi-Agent Systems (MAS): Multi-agent systems are those in which multiple agents (usually) cooperate to perform some task.

Ontology: An ontology is a well-defined set of concepts that are ordered in some manner to create an agreed-upon vocabulary for exchanging information.

Smart Agent: A program or collection of programs that lives for some purpose in a dynamic environment and can make decisions to perform actions to achieve its goals. Individual agents may be conceptualized as having beliefs, desires, and intentions that can communicate with other agents to satisfy their goals.

Software Team: Groups of agents which have tasks that are tightly interrelated; these teams roughly parallel a department of employees.

Contract-Based Workflow Design Patterns in M-Commerce

V.K. Murthy
University of New South Wales, Australia

E.V. Krishnamurthy
Australian National University, Australia

INTRODUCTION

This overview describes an object-based workflow paradigm to support long and short duration transactions in a mobile e-commerce (or m-commerce) environment. In this environment, the traditional transaction model needs to be replaced by a more realistic model, called a "workflow model" between clients and servers (peers) that interact, compete and cooperate, realizing an intergalactic client-server program. The various types of task patterns that arise in m-commerce, for example, reservation, purchasing, require a "what if" programming approach consisting of intention and actions for trial-error design before an actual commitment is made. Such an approach enables us to take care of the unpredictable nature of connectivity of the mobile devices and the networks and also provide for the trial and error program design required in m-commerce.

BACKGROUND

A mobile computing environment consists of fixed host computers and mobile client computers that are linked together by a network (wired or wireless) so that these computers can communicate among themselves using messages. The mobile clients (MC) are capable of connecting to the fixed network via a wireless link. Fixed host (FH) provides mobile application services and information to mobile clients. MC supports query invoking and information filtering from FH to provide personal information service. We model a mobile computing environment as a collection of different types of objects by identifying each computer (mobile and stationary) as an object. This model is called Mobile Object Programming System (MOPS).

A mobile object programming system (MOPS) (Vitek & Tschudin, 1997) is interpreted as a collection of objects interacting through messages. Each object maintains its own share of data and has its own program piece to manipulate it. That is each object combines datastructure and functionality. The objects are active and behave like actors in a movie, each following its own script and interacting with other objects.

A task submitted from a mobile client is called a mobile transaction. It is a distributed task that can be executed partly within that mobile client (MC) as an internal transaction (Intran) and partly in other fixed hosts (FH) as external transactions (Extran). Each FH has a coordinator FHC that receives external transaction operations from mobile hosts and monitors their execution in the database servers within the fixed host. Similarly each MC has a coordinator MCC.

Conventionally, transactions are assumed (Bacon, 1993; Krishnamurthy & Murthy, 1992) to satisfy the ACID properties, namely:

- **Atomicity:** All or none of transaction happens;
- **Consistency:** A transaction preserves the consistency in database before and after its execution;
- **Isolation:** Intermediate results are not externally made visible until commitment;
- **Durability:** The effects are made permanent when a transaction succeeds and recovers under failure.

The ACID properties turn out to be restrictive for mobile transactions and need to be relaxed as illustrated by the following example on flight reservation.

Here, one needs to have the following tasks: select a suitable airline that offers cheaper fares, ensure there is vacancy, and make an advanced booking (that is to be confirmed later). These individual steps are really not traditional transactions, but well-defined program pieces that can be carried out concurrently and may need to follow a predefined partial order to satisfy certain predicates (e.g., seat availability) invariant criteria (number of seats is determined by the size of aircraft) and synchronization (one step requires the completion of other step). Therefore, the steps need not be atomic, need not be immediately consistent, and need not satisfy isolation property since intermediate non-commit states are to be made visible (called externalization). Furthermore, such steps are liable to be cancelled eventually and require a suitable rollback preserving some local states. Also, in a cooperative mobile environment it may be necessary to allow data exchange between some transactions during their execution thereby necessitating the relaxation of the isolation property.

Thus in a mobile environment, we need a new model where:

(i) The isolation property is removed making the intermediate results visible.
(ii) The precedence order in execution and other dependencies are taken care by modifying the atomicity requirement.

This model is called "a workflow" between the MC and FH. The workflow model, its contract based formulation and the e-commerce patterns described here result from the basic work on contract model of Wachter and Reuter (1995), the design by contract by Meyer (1992a, 1992b) and the design patterns and contracts by Jezequel, Train, and Mingins (2000).

MAIN THRUST OF THIS ARTICLE

The main thrust of this article is to describe the workflow paradigm for applications in m-commerce. Unlike the transactional paradigm, which is severely restricted by the ACID properties, the workflow permits more general properties that are suitable for m-commerce applications. We illustrate some of the important applications and the design of suitable protocols. We also classify some workflow patterns arisng in e-commerce and the language support needed.

WORKFLOWS: EXTERNAL AND INTERNAL

A global workflow (we call it an external workflow or *Extran*) T(ij) is defined as a workflow between two objects O(i) and O(j); this consists of a message sent from O(i) to execute a desired workflow in O(j). This message is received by O(j). O(j) has a behavior specified by: Pre(T(ij)), G(j), C(j), Post (T(ij))S(j), where Pre() and Post() are respectively the pre and post states that are active before and after the workflow T(ij). G(j) is a guard of O(j) to signal when the required precondition is met, and C(j) is the command function; S(j) signals when the post condition is achieved. Here, the script specifies what message O(j) can accept and what actions it performs when it receives the message while in state Pre(T(ij)) to satisfy the post condition post(T(ij)). The Extran T(ij) can trigger in O(j) numeric, symbolic or database computations.

Each Extran T(ij) from object i to object j triggers a set of serializable computations in O(j) either in a total order or in a partial order depending upon whether parallelism, concurrency and interleaving are possible locally within O(j). If the object O(j) is "made up" of subobjects, we may have to execute a workflow consisting of several local workflows (called internal workflow or Intran). After executing Intran, each object reaches a new state from an old state using its internal command set C(j); before executing the commands, the required precondition is met, and after completion of the command set, the post condition is ensured in the new state. This is the design by contract approach (Meyer, 1992a, 1992b) and widely used in the language Eiffel. The precondition is specified by "require" and post condition by "ensure", see Meyer (1992a, 1992b).

We shall see in the next section that the Extran and Intran have more general properties than the ACID properties mentioned earlier.

CHEMICAL REACTIVITY-LIKE PROPERTIES

The Extran and Intran have more general properties called "chemical reactivity properties", since they resemble chemical reactions: molecularity, contractual obligation, opacity during a molecular action, and retry or rescue through a recovery protocol bringing the system back into the invariant state. These are defined as below:

(i) **Molecularity:** If there is a crash during a composite operation, all of the effects of the sub-operation are lost. If there is no crash, the composite or molecular operation is complete. That is, a molecule is synthesised fully or not at all.
(ii) **Contractual obligation:** Invocation of a single composite operation takes the program from one consistent state to another. This means precondition and post condition of a contract holds. Thus, conventional consistency is replaced by contractual obligation.
(iii) **Opacity:** The results of the sub-operations of composite operation should not be revealed until the composite operation is complete.
(iv) **Durability:** If a crash occurs and contract fails and a component cannot meet its obligation and fails in its contract, an exception is raised. Then we have three possibilities:
    a. Exception is not justified: It is a false alarm we may ignore.

b. If we have anticipated the exception when we wrote the routine and provided an alternative way to fulfill the contract, then the system will try that alternative. This is called resumption.

c. If, however, we are still unable to fulfill the contract, we go into graceful degradation or surrender with honour. Then bring all the objects to an acceptable state (pre- committed-state) and signal failure. This is called organized panic. This should restore the invariant. At this point, we initiate retry. The effect of retry is to execute the body of the routine again.
Remark: In Eiffel (Jezequel et al., 2000; Meyer, 1992a, 1992b; Wiener, 1996) the rescue clause does all of the above (this is essentially RESTART after recovery in transaction processing)
Also in MOPS, the chemical reactivity properties are generalized further: Since a number of remote objects are invoked, the coordinators FHC and MCC should ensure all the remote actions and the local actions are complete. If anyone fails the whole program has to be abandoned, and we need to retry, rescue and bring the system to its invariant state. Contractual obligation is extended to all objects under concurrent invocation and partial failures.
No results are revealed from any objects until all the actions are complete and the coordinator commits.

(v) **Retry/Rescue and Reset:** If false alarm then retry; else rescue and restart so that all the invariants in all objects are rest to their pre-action state.

Recall that we have two types of transactions " Extran (external transaction) between a MC and an FH, and Intran (internal transaction) that take place locally within each MC or in FH. We split the Extrans into intention and action transactions, where the intention transactions have the ACID properties, and the action transactions are long duration transactions supported by a recovery protocol. The intention transactions are again local to each MC or FH and based on the decision in this phase, the action transaction takes place through a protocol (called intention-action protocol (IAP)) provided with a time-out strategy and recovery to cope up with failures of disconnection. We now illustrate the applications of the intention-action protocol.

HOTEL RESERVATION PROBLEM

Consider the hotel reservation problem. The hotel database has a set of rooms and each room has a status available or unavailable on specified dates. The object Alloc stores the confirmed allocation of rooms to named guests and Int-alloc holds the intentional reservations intention-committed. We assume that each guest has a unique ID and is not permitted to hold multiple reservations in the same hotel.

The hotel database has two integrity constraints:

(i) The number of confirmed + intentional reservations do not exceed the number of rooms on any date.
(ii) The set of rooms that are unavailable on any date are those already allocated to the guests in Alloc plus those in Int-Alloc on those dates.

The following transactions act on the hotel database.

(i) Reservation request from guest within a deadline (Extran);
(ii) Reservation availability from hotel and intention commit (Intran);
(iii) Send negative intent or (positive intent + deadline) to guest to confirm (Extran);
(iv) Reservation acceptance or cancellation by the guest (Extran);
(v) After cancellation by the guest or FH-deadline, the intent reservations are cancelled and rooms are made available and the Int-Alloc is suitably modified and committed (Intran); and
(vi) After the confirmation by guest within FH-deadline the Int-Alloc rooms are transferred to Alloc and committed (Intran).

Note that transaction 2, 5, 6 are Intrans with ACID properties while 1,3,4 are Extrans using IA-protocol.

E-CHECKING

The electronic check (e-check) was introduced by Financial Services Technology Consortium (FSTC) [www.fstc.org]. This system has functions, such as: deposit and clear, cash and transfer. It has entities: payee, payer, payer bank, payee bank and clearing house. E-check can be an effective system over integrated networks. E-check can be used for high value transactions, business-to-business (B2B) payments, and so forth. At every stage of this system, the intention-action protocol and time-tolerant and time-critical transactions arise (Dani & Radha Krishna, 2000; Wang & Das, 2001). Using intention-action, log and time stamps, we can incorporate fairness, non-repudiation and timeliness of financial transactions.

When a check is presented, the bank verifies the details, signature and validates the check and security codes and accepts or rejects a check within a time limit. Hence, the time stamp also plays an important role in clearing and depositing checks in time order of issue whenever there is not adequate balance in the account, in order to resolve the conflicts and appropriately inform the payee and payer. Also using logs non-repudiation can be achieved so that the participants cannot deny a particular intention, event or action.

FUTURE TRENDS

Two important aspects require further research and development: The analysis of workflow patterns that arise in e-commerce and the development of appropriate software tools for the design of contract-based workflow patterns. These aspects are described in the following section.

WORKFLOW PATTERNS IN E-COMMERCE

Certain special behavioral patterns (Gamma et al., 1995; Jezequel et al., 2000; Shalloway & Trott, 2002) of workflows occur in e-commerce applications. A pattern enforces a problem solving discipline for a design architecture. It consists of three parts: a context, a problem and a solution. The context is a description of the environment before the pattern is applied. It outlines the preconditions under which the problem and its solution appear. The context shows where and when the pattern will work. It is usually introduced with the help of a scenario. Problem describes the goals and objectives of a situation. Solution describes the means to achieve the goals and objectives through a suitable protocol.

In e-commerce, we encounter the following patterns; there can be other workflow patterns which are combinations of these patterns.

(i) **Conventional transactions with ACID properties:** These occur at the *lowest level of a workflow*.

(ii) **Supply-chain pattern:** Here the context is *a sequence of conventional transactions between a customer and a seller through a set of intermediary agents*. Here, the workflow is successful, if and only if each individual transaction in a chain is successful. An abort in any transaction will be considered unsuccessful, restoring the consistency of the initial state.

(iii) **Mutually exclusive- collectively exhaustive pattern:** Here, the context is *a single buyer trying to acquire goods from several suppliers to complete a given task through several mutually exclusive transactions* and collectively exhaustive transactions. Here, the workflow is considered successful only if all the mutually exclusive transactions are all committed. If any one is aborted this workflow is unsuccessful and the initial states are restored.

(iv) **Negotiated choice pattern:** Here, the context is that *a customer bargains and negotiates with several suppliers* simultaneously to obtain a particular product at a bargain price. This workflow is successful, if and only if one of the transactions is committed and soon after that time the rest of the transactions are aborted respecting the contract. In the latter case, the states are properly restored.

(v) **Auction Pattern:** Auction process is *a controlled competition among a set of clients and auctioneer* coordinated by the auctioneer. In the auction pattern, the transactions happen between an auctioneer and several clients through communication and these are successively revised. The rules of the English auction pattern are as follows:

a. The auctioneer-agent begins the process and opens the auction.
b. At every step, decided by a time stamp, only one of the client-agent is permitted to bid; the auctioneer relays this information. The bidding client agent is called active, and it does not bid more than once and this client becomes inactive until a new round begins.
c. After the auctioneer relays the information, a new client becomes active and bids a value strictly greater than a finite fixed amount of the earlier bid. (This is English auction; it can be modified for other auctions).
d. If within a time-out period no client-agent responds, the last bid is chosen for the sale of the goods and the auction is closed.

Table 1 summarizes the patterns.

LANGUAGE SUPPORT

Eiffel, Java, and UML are powerful languages to implement mobile object programming systems. They provide for software contracts that capture mutual obligations through program constructs such as: “require [else]” for precondition and “ensure [then]”for post condition, assertions, invariants. Eiffel provides for a good semantics

Table 1. Workflow patterns in e-commerce

Name of Pattern	Properties
Conventional Transaction	Has ACID properties: ***one customer, and one supplier*** ; commitment. or abort.
Supply Chain	Workflow between ***a customer and several sellers consisting of a set of transactions through intermediary agents***. The workflow is successful if and only if each individual transaction in the chain is successful.
Mutually Exclusive-Collectively Exhaustive	Workflow consisting of several ***mutually exclusive and collectively exhaustive transactions between a customer and several suppliers*** to acquire different items. The workflow is successful if and only if all the transactions are committed.
Negotiated Choice	Workflow takes place between a ***customer and several suppliers; the customer bargains and negotiates with several suppliers*** simultaneously to obtain a particular product at a bargain price. This workflow is successful, if and only if one of the transactions is committed .
English Auction	A ***controlled competition among a set of clients and an auctioneer*** coordinated by the auctioneer. Here the last bid with the highest value is chosen for the sale of the goods and the auction is closed with one successful bidder.

for exception handling through a "rescue" clause and "retry" clause for dealing with the recovery and resumption. The tool called "iContract" (Kramer, 1998) provides developers with supports for design by contract using Java. UML (Unified Modeling Language) and OCL (Object constraint Language) (Warmer & Kleppe, 1999) has also been gaining importance.

CONCLUSION

We described the mobile-object based workflow paradigm for mobile e-business by splitting the transactions into external and internal transactions. Also, we described the intention-action protocol and how to implement them by contractual design using Eiffel, i-contract tool of Java and UML. We also described several kinds of behavioral patterns that arise in e-commerce and how the contract-based workflow scheme becomes a useful vehicle for implementing these patterns through software.

REFERENCES

Bacon, J. (1993). *Concurrent systems*. Reading: Addison Wesley.

Clark, A. & Warmer, J. (2002). Object modeling with the OCL. *Lecture Notes in Computer Science*, 2263. New York: Springer Verlag.

Dani, A.R. & Radha Krishna, P. (2001). An e-check framework for e-payment systems in Web-based environment, 91-100, In *E-commerce and Web Technologies*, *Lecture Notes in Computer Science*, 2115. New York: Springer Verlag.

Gamma, E. et al. (1995). *Design patterns*. Boston: Addison Wesley.

Jezequel, M., Train, M., & Mingins, C. (2000). *Design patterns and contracts*. Reading, MA: Addison Wesley.

Kramer, R. (1998). iContract The Java design by contract tool. *26th Conference on Technology of Object Oriented Systems,(TOOLS USA'98)*, Santa Barbara.

Krishnamurthy , E.V., & Murthy, V.K. (1992). *Transaction processing systems*. Sydney: Prentice Hall.

Meyer, B.(1992a). *Eiffel - The language.* New York: Prentice Hall.

Meyer, B. (1992b). Applying design by contracts. *IEEE Computer, 25*(10), 40-52.

Shalloway, S., & Trott, J.R. (2002). *Design patterns explained*. New York: Addison Wesley.

Thomas, P., & Weedon, R. (1998). *Object-oriented programming in Eiffel*. Reading, MA: Addison Wesley.

Vitek, J., & Tschudin, C. (1997). Mobile object systems. *Lecture Notes in Computer Science*, 1222. New York: Springer Verlag.

Wachter, H., & Reuter, A.(1995). The ConTract model. In A.K. Elmagarmid (Ed.), *Database transaction models* (pp.219-264). San Mateo,California: Morgan Kaufmann.

Wang, G., & Das, A. (2001). Models and protocol structures for software agent based complex e-commerce transactions. In *E-commerce and Web Technologies, Lecture Notes in Computer Science*, 2115, 121-131. New York: Springer Verlag.

Warmer, J., & Kleppe, A. (1999). *The object constraint language*. Reading, MA: Addison Wesley.

Wiener, R.S. (1996). *An object oriented introduction to computer science using Eiffel.* NJ: Prentice Hall.

KEY TERMS

Acid Properties:

- **Atomicity:** All or none of transaction happens;
- **Consistency:** A transaction preserves the consistency in database before and after its execution;
- **Isolation:** Intermediate results are not externally made visible until commitment;
- **Durability:** The effects are made permanent when a transaction succeeds and recovers under failure.

Contract: A consistent and fault tolerant execution of a program from one consistent state to another consistent state achieving the required precondition and ensuring that the post condition is met .

Intergalactic Client-Server Program: In the mobile computing environment, the traditional transaction model is replaced by a more realistic model called a "workflow model" between several clients and servers that interact, compete and cooperate. This is called "an intergalactic client-server program (ICSP)".

Intran: A transaction that is executed within a mobile host (MH) as an internal transactions (for brevity, Intran) using its own internal operations and data.

Extran: An external transaction that is requested by a mobile host to be performed in another external computing environment, using message communication.

Pattern: Enforces a problem solving discipline for design architecture. It consists of three parts: a *context*, a *problem* and a *solution*. The terms context, problem and solution are defined below:

- **Context:** Describes the environment before the pattern is applied. It outlines the preconditions under which the problem and its solution appears. Thus the context shows where and when the pattern will work. It is usually introduced with the help of a scenario.
- **Problem:** Describes the goals and objectives of a situation.
- **Solution:** Describes the means to achieve the goals and objectives.

Script: A condition-event structure that describes a stereotyped sequence of events in a particular context. Scripts are useful because, in the real world, there are patterns in the occurrence of events. Events described in a script form a causal chain.

Workflow: A collection of simpler tasks organized to accomplish some well-defined total task. Here each simple task defines some unit of work to be carried out as in a transaction. A workflow ties together a group of tasks by specifying execution dependencies and the dataflow between tasks. Also, there may be a constraint that a particular task cannot begin until some other task ends.

Contracting Mechanisms in the IS/IT Outsourcing Phenomenon

Abdulwahed Mohammed Khalfan
College of Business Studies, Kuwait

Majed Z. Al-Hajery
College of Business Studies, Kuwait

Khamis Al-Gharbi
Sultan Qaboos University, Oman

INTRODUCTION

As organisations need diverse and high quality IS/IT services to survive and excel in the rapidly changing business environment (Lacity & Willcocks, 2001), regular attempts have been made by organisations to sustain competitiveness through increasing their investments in IS/IT operations and resources. However, this has pushed for pursuing different options and alternatives as organisations strive to maximise their flexibility and control (Lacity, Willcocks, & Feeny, 1996). To accomplish this objective, more and more organisations are constantly looking to IS/IT outsourcing by external vendors as a viable option rather than pursuing in-house IT developments (Lacity & Willcocks).

As the central key issue in the successful IS/IT outsourcing operation is the right contract (Lacity & Hirschheim, 1995), the purpose of this chapter is to gain a deeper understanding of the contracting mechanisms that surround IS/IT outsourcing operations.

IS/IT OUTSOURCING DEFINITION AND BACKGROUND

Precise definitions of IS/IT outsourcing differ in the literature, but there is a general agreement that it is the carrying out of IS/IT functions by a third party (Ketler & Walstrom, 1993). In the context of this chapter, however, IS/IT outsourcing refers to information systems/ information technology functions only.

Cheon, Grover, and Teng (1995, p. 209) define IS/IT outsourcing as "the organisational decision to turn over part or all of an organisation's IS functions to external services provider(s) in order for an organisation to be able to achieve its goals. This definition includes the following external services: applications development and maintenance, systems operations, network/communications management, end-user computing support, systems planning and management, and purchase of application software, but excludes business consulting services, after-sale vendors services, and the lease of telephones lines. An organisation can obtain these services through complete outsourcing, facilities management, systems integration, time-sharing, and other contracts (including rental, installation and procurement, and maintenance and programming)."

GLOBAL BUSINESS SCOPE OF THE IS/IT OUTSOURCING MARKET

International Data Corporation (IDC, July 24, 2001) indicated that global spending on IT services would reach US$700.3 billion by 2005, up from US$439.9 billion in 2001. IDC found that the United States would remain the biggest spender on IT services at US$335 billion in 2005, up from US$206.9 billion this year. The Outsourcing Institute (2000) expects a growth in small business outsourcing contracts exceeding 25%. IDC (July 24, 2001) found that spending in Western Europe will grow from US$127.5 billion to US$192.4 billion in the same period, and spending in Japan will increase from US$53.2 billion to US$75.2 billion. Spending on IT services in the rest of the world will amount to US$52.5 billion to US$97.7 billion.

Even though the above-mentioned forecasts implicitly point to increasing size and scope of contracts, Currie (2000, p. 177) observes, "most companies tend to prefer to enter outsourcing contracts which are smaller (in money), shorter (in duration) and more narrow in scope (range of IT services)." However, the megadeals which involve total outsourcing continue to gain media attention due to the dollar value of such deals (Lacity & Willcocks, 2001).

IDC (July 12, 2001) indicated that spending on IT services in Asia-Pacific (excluding Japan) is predicted to rise from US$18 billion by the end of 2001 to US$42 billion by 2005. China will have the biggest growth in IT services spending, up to US$9 billion in 2005 from US$900 million in 2000.

IDC (2003) reveals the fastest growth for IT outsourcing services will be in government, financial markets, services, and discrete manufacturing. At the same time, the market opportunity for IT outsourcing services can differ significantly depending on the vertical industry. The report (IDC, 2003), which presents the 5-year forecast for the U.S. IT outsourcing market within 16 vertical markets, indicates a range in the compound annual growth rate (CAGR) from 3.5% in communications/media and education to 7.4% in government.

IS/IT OUTSOURCING IN PRACTICE

There are different types of IS/IT outsourcing. For example, Grover, Cheon, and Teng (1994) have proposed the following taxonomy: (1) complete outsourcing, (2) facilities management outsourcing, (3) systems integration outsourcing, (4) time-sharing outsourcing, and (5) other types of outsourcing. Additionally, Lacity and Hirschheim (1995) provide a taxonomy of IT outsourcing options. They are divided into three categories of IS/IT outsourcing: body shop, project management, and total outsourcing.

INSOURCING

According to Lacity and Willcocks (1998), total insourcing is "the decision to retain the management and provision of more than 80% of the IT budget internally after evaluating the IT services market" (p. 371).

SELECTIVE IS/IT OUTSOURCING

According to Lacity and Willcocks (1998), selective outsourcing is "the decision to source selected IT functions from external provider(s) while still providing between 20% and 80% of the IT budget internally. This strategy may include single or multiple vendors" (p. 371). Currie and Irani (1999) believe that "selective sourcing" is a variant of multiple-supplier sourcing.

TOTAL IS/IT OUTSOURCING

According to Lacity and Willcocks (1998), total outsourcing is "the decision to transfer the equivalent of more than 80% of the IT budget for IT assets, leases, staff, and management responsibility to an external IT provider" (p. 371).

CONTRACT DEFINITION, APPROACH AND THEORY

A contract is defined as: "An agreement between two or more people or organisations, 'parties' to the contract, which creates rights and obligations for those parties. These obligations can be legally enforced by a party who is entitled to their benefits" (Klinger & Burnett, 1994, p. 58). It is normally "a bargain made between at least two parties with an offer by one party which has been accepted by the other party with some benefits gained" (Klinger & Burnett, p. 59). A contract enables the different parties to achieve their strategic and commercial aims. It regulates the relationship of the different parties, pointing out risk and costs. It also can provide a framework for continuing to work together in the computing environment with consideration given to "certain wrongs" such as negligence, "deceit or defamation are legally known as torts" (Klinger & Burnett, p.59). A good contract is often the central key to a successful IT outsourcing relationship (Lee, 1995). As Lacity and Hirschheim (1995, p. 243) explain, "the contract is the only mechanism that establishes a balance of power in the outsourcing relationship." The contract usually defines the rights, liability, and expectations of both the outsourcing parties (i.e., the vendor and the organisation) and is often the only solid mechanism for regulating the relationship between them. However, research on IT outsourcing negotiations and contract building tend to be mostly theoretical and very limited in scope (Lee). Currie (1995) points out that one of the most significant problems of any IT outsourcing deal is "defining the legal contract that underpins the client-supplier relationship" (p. 194).

Contracts are an important and critical part of outsourcing decisions and can be very complicated due to the fact that there is often a transfer of assets, including people, equipment, software, and buildings from the client to the vendor. For the terms of the contract to be enforceable, they must be "verifiable." Based on their empirical research, Diromualdo and Gurbaxani (1998) identified elements "critical" to the success of contractual agreements, including: (1) the contract type, (2) performance measurement and evaluation scheme, (3) the compensation systems, and (4) assignment of decision-making rights to the vendor.

IS/IT CONTRACT TYPES

For each given transaction, IT managers can choose between two governance mechanisms to govern the transaction:

- **Hierarchical governance:** produce the transaction internally
- **Market governance:** purchase the transaction from an IT service provider

Also, Fitzgerald and Willcocks (1994) distinguish six types of contracts, each with a different price mechanism (see Figure 1).

Moreover, outsourcing contracts range from fixed fee to incentive types with penalties in case of unsatisfactory performance. Further, many contracts also contain early termination provisions and incentives schemes for technology change. Penalties and incentives are important features of any type of IS/IT outsourcing contract, as they serve as "inducements to the vendor and as mechanisms by which the outsourcer can manage shirking in the relationship" (Ngwenyama & Bryson, 1999, p. 353).

IS/IT outsourcing contracts had evolved from the buyer/supplier perspectives of service bureau and facilities management contracts to the partnership perspective of system integration outsourcing. At the same time, Lacity and Willcocks (1998) name the IT contracts as "fee-for-service contracts," where detailed information about requirements, service levels, performance metrics, penalties for nonperformance, and price are discussed and documented.

In addition, Diromualdo and Gurbaxani (1998) note that different types of contracts apply in different circumstances. For example, the strategic alliance or joint venture type would be chosen if the organisation required a specialized expertise and knowledge of the business process as well as tight control over the resources involved in the work. In general, the literature distinguishes between two types of contracts: (1) comprehensive contracts and (2) incomplete contracts.

1. Comprehensive (Complete) Contracts

A comprehensive (complete) contract specifies every possible contingency. Indeed, such a contract never needs to be revised or complemented (Richmond, Seidman, & Whinston, 1992). A contract for the purchase of computer supplies can be an example of a comprehensive contract.

2. Incomplete Contracts (Flexibility)

Most business contracts are incomplete because when a contingency arises the different parties must bargain to resolve the conflict. In this sense, the contract must provide for "arbitration and renegotiation" (Richmond et al., 1992). Additionally, Harris, Guiunipero, and Hult (1998) suggest that the term so-called "incomplete contract" is synonymous with a flexible contract. Most outsourcing contracts are incomplete since rapidly changing technology, uncertainty (i.e., complexity), and organisational environments make it so difficult to specify every contingency in the contract. Moreover, using incentive contracts is another mechanism to provide the agreement of IT outsourcing with flexibility. This means that the contract can include some method by which the vendor shares in any cost savings or increased profits made under the agreement.

CRITICAL CRITERIA FOR CONTRACT NEGOTIATION

The key clauses and schedules that constitute the main contractual issues are:

Before a thorough discussion of the criteria of contract negotiation starts, there is a need to discuss:

Figure 1. Types of outsourcing contracts (Adapted from Fitzgerald & Willcocks, 1994)

Contract	Basis of Payment
Time and material fixed fee	Actual use of personnel and material lump sum for a defined workload or service
Fixed fee	A lump sum for a defined workload or service
Fixed fee plus variable element	Predicted changes in, for example, workloads on business circumstances
Cost plus management fee	Real costs incurred by vendor plus a percentage
Fee plus incentive scheme	Some benefits that accrue to client company, or performance over and above an agreed baseline
Share of risk and reward	How well client company or joint venture performs

Need for Negotiation

Negotiations may be defined as: "the bargaining process between two or more parties seeking to secure predetermined objectives, usually with a clear sense of purpose" (Klinger & Burnett, 1994, p. 13). In fact, compromise is an essential feature of most successful negotiations. Additionally, negotiations can be costly and can range from a telephone discussion on some key issues through to meetings. In the same time, negotiations require much skill in how to negotiate and compromise. To build a strong contract, much negotiation is needed between the different parties to discuss all possible contingencies and also the service level agreement (SLA) of the contract.

Service Level Agreement (SLA)

The SLA is usually a major part of any outsourcing contract and has been considered as the most important component of any outsourcing deal (Richmond et al., 1992). In short, the SLA refers to the "details and/or quantifies in schedules the minimum level of service which the customer (i.e., client organisation) will obtain" (Parry, 1997). In addition, the SLA contains the technical details of the proposed systems, performance levels and standards, methods of monitoring, the output (i.e., deliverables), procession priorities, response time, termination notice, and so on. It is to be noted that there is mounting evidence that poor IS/IT outsourcing decisions are the direct result of an inadequate definition of customer requirements or the SLA (Richmond et al.).

Furthermore, Kern and Willcocks (2000) identify typical clauses and exhibits for service enforcement and monitoring:

- Application development measure performance (i.e., allows organisation to monitor development).
- Benchmarking (i.e., the organisation can compare at any time IT services with other services).
- Customer satisfaction survey.
- Performance reporting requirements (i.e., reports on service to be provided by the vendor in time intervals).
- Performance implies that the supplier provides products or services and the customer (i.e., client) pays for them. And also the work must be completed before the payment has to be made.

Price Mechanism and Flexibility

Pricing scheme is the central key to any outsourcing agreement. In fact, Diromualdo and Gurbaxani (1998) call it "vendor compensation" (p. 71). The pricing schedule is either specified in advance for the duration of the contract or negotiated each year. A major task in determining the future cost (i.e., price schedule) is quite difficult given the uncertainty in future technologies and business conditions. At the same time, any additional services outside the previously agreed service scope (e.g., overtime) will be charged according to the agreed price rates. Also, the timing of the payments has to be agreed on in advance.

Contract Duration

The performance of an obligation should be done in a "reasonable time" that the two parties have agreed on.

It is widely believed that in order to achieve a flexible contract, the contract duration should be short (Fitzgerald & Willcocks, 1994). However, it is difficult to define a short contract precisely, but it could be argued that a contract that is tied to the "technological life cycle" would be a short contract (Harris et al., 1998).

Termination Clause

The contract itself often contains termination conditions. There are two different types of terminations, early and exit. In addition, the contract may permit either party to terminate on notice to the other party.

Early Termination

If the parties involved in the agreement decide during the course of the contract that they wish to terminate the agreement, they can agree to discharge it (Klinger & Burnett, 1994). Sometimes, one party wants to release the second party from completing the contract. There have been some high-profile outsourcing disasters, where some organisations terminated the contract with the IT vendors for a number of reasons (Currie & Irani, 1999). However, it should be borne in mind that switching costs are expensive, whether the organisations choose to move to another vendor or rebuild an in-house department.

Exit Termination (Provisions)

According to Shepherd (1999, p. 79), exit provisions define the "responsibilities of each of the parties when the contract is terminated." It may also include arrangements of services while handing over to another vendor or the internal IS department. The importance of the "exit provision" stems from the fact that the organisation is potentially exposed to risk from the IT services in case of contract termination.

Key Vendor Personnel

It has been argued that it is of fundamental importance to the organisation to know who will be responsible for the IT services or technology from the vendor side. In other words, the key vendor's employees should be explicitly listed within the scope of the contract, with their expertise and experience, as remaining with the account for specified time periods.

Dispute Resolution

Kern and Willcocks (2000) assume that the contract should point out clearly what is the "resolution mechanism" to deal with any dispute among the different parties. Further, if the dispute cannot be resolved in the first instance, the contract should be able to direct to a further "escalation procedure," where the appropriate people or group should be contacted (e.g., arbitration). As a final resolution, a neutral third-party advisor will be appointed to solve the problem.

Post-Contract Management

The post-contract management agenda is mainly concerned with enforcing the contract and achieving the stipulated terms after it has been signed by the different parties.

FUTURE TRENDS

The growth in IS/IT outsourcing is being fueled by IT technological advances and new business opportunities. The technological trends behind the next generation of IS/IT outsourcing include the growth of universal IP networking, development of server-based computing, and better distributed systems management (Wreden, 1999, p. 128). The market has fragmented into a confusing mix of application service providers (ASPs), Webhosting, and e-commerce providers and network-integration experts.

CONCLUSION

IS/IT outsourcing contracts are expected to increase dramatically in number in coming years, forcing IT managers to face an array of complicated choices in picking IS/IT vendors and monitoring contracts.

One main aspect of this chapter, an aspect often overlooked by organisations, has been to illustrate the importance of the contracting mechanisms throughout the IS/IT outsourcing process and operation.

The success of an IT outsourcing relationship depends primarily on both parties' ability to negotiate and sign up a successful contract, which ultimately eliminates the risk factors mainly in three areas: complexity, governance, and flexibility.

Bearing in mind these considerations, the most critical contracting lines and issues have been identified, and the theoretical guidelines for successful IT contracts have been shown throughout this chapter.

REFERENCES

Cheon, M., Grover, V., & Teng, J. (1995). Theoretical perspectives on the outsourcing of information systems. *Journal of Information Technology*, *10*(4), 209-220.

Currie, W. (1995). Managing IT in the private and public sectors: Strategy and outsourcing. In W. Currie (Ed.), *Management strategy for IT: An international perspective*. London: Pitman.

Currie, W. (2000). *The global information society*. Chichester, UK: Wiley.

Currie, W., & Irani, Z. (1999). *Evaluating the benefits, costs, and risks of IT/IS outsourcing in a maturing market*. Paper presented at the Sixth European Conference on IT Evaluation, London.

Diromualdo, A., & Gurbaxani, V. (1998). Strategic intent for IT outsourcing. *Sloan Management Review*, *39*(4), 67-80.

Fitzgerald, G., & Willcocks, L. (1994). Contracts and partnerships in the outsourcing of IT. *Proceedings of 15th International Conference on Information Systems*.

Grover, V., Cheon, M., & Teng, T. C. (1994). An evaluation of the impact of corporate strategy and the role of information technology on IS functional outsourcing. *European Journal of Information Systems*, *3*, 179-190.

Harris, A., Guiunipero, L., & Hult, G. T. (1998). Impact of organizational and contract flexibility on outsourcing contracts. *Industrial Marketing Management*, *27*, 373-384.

International Data Corporation. (2001, July 12). *IT services spending to boom in Asia*. Retrieved July 25, 2003, from *http://www.nua.com/surveys/index.cgi?f=VS&art_id=905356970&rel=true*

International Data Corporation. (2001, July 24). *IT services spending to soar*. Retrieved July 25, 2003 from the Web: *http://www.nua.com/surveys/index.cgi?f=VS&art_id=905357011&rel=true*

International Data Corporation. (2003). *Market coverage*. Retrieved May 18, 2003, from *http://www.idcresearch.co.nz/market/default.htm*

Kern, T., & Willcocks, L. (2000). Contracts, control and presentation in IT outsourcing: Research in thirteen UK organizations. *Journal of Global Information Management*, *8*, 15-29.

Ketler, K., & Walstrom, J. (1993). The outsourcing decision. *International Journal of Information Management*, *13*, 449-459.

Klinger, P., & Burnett, R. (1994). *Drafting and negotiating computer contracts*. London: Butterworth.

Lacity, M., & Hirschheim, R. (1995). *Beyond the information systems outsourcing bandwagon: The insourcing response*. Chichester, England: Wiley.

Lacity, M., & Hirschheim, R. (1996). The role of benchmarking in demonstrating IS performance. In L. Willcocks (Ed.), *Investing in information systems: Evaluation and management*. London: Chapman & Hall.

Lacity, M., & Willcocks, L. (1998). An empirical investigation of information technology sourcing practices: Lessons from experience. *MIS Quarterly*, *22*(3), 364-408.

Lacity, M., & Willcocks, L. (2001). *Global information technology outsourcing: In search of business advantage*. Chichester, England: Wiley.

Lacity, M., Willcocks, L., & Feeny, D. (1996). The value of selective IT sourcing. *Sloan Management Review*, *37*(3), 13-25.

Lee, M. (1995). IT outsourcing contracts: Practical issues for management (Working Paper No. 95/05). City University of Hong Kong, Information Systems Department.

Ngwenyama, O., & Bryson, N. (1999). Making the information systems outsourcing decision: A transaction cost approach to analyzing outsourcing decision problems. *European Journal of Operational Research*, *115*, 351-367.

Outsourcing Institute. (2000). *Outsourcing index 2000: Strategic insights into US outsourcing*. Retrieved September 8, 2001, from *http://www.outsourcing.com/Common/Index2000.jsp*

Parry, B. (1997). *IS outsourcing*. Unpublished notes on IS outsourcing, BP Consultancy, Kuwait.

Richmond, W., Seidman, A., & Whinston, A. (1992). Incomplete contracting issues in information systems development outsourcing. *Decision Support Systems*, *8*, 459-477.

Shepherd, A. (1999). Outsourcing IT in a changing world. *European Management Journal*, *17*(1), 64-84.

Willcocks, L., & Fitzgerald, G. (1994). Contracting for IT outsourcing: Recent research evidence. *Proceedings of 15th International Conference on Information Systems*, 91-98.

Wreden, N. (1999). Outsourcing flexes its muscles. *InformationWeek*, (761).

KEY TERMS

Comprehensive (Complete) Contracts: A comprehensive (complete) contract specifies every possible contingency and pays attention to all different and tiny issues.

Contract: "An agreement between two or more people or organisations, 'parties' to the contract, which creates rights and obligations for those parties."

Incomplete Contracts (Flexibility): Most outsourcing contracts are incomplete since rapidly changing technology, uncertainty (i.e., complexity), and organisational environments make it so difficult to specify every contingency in the contract.

Insourcing: "The decision to retain the management and provision of more than 80% of the IT budget internally after evaluating the IT services market."

IS/IT Outsourcing: "The organisational decision to turn over part or all of an organisation's IS functions to external services provider(s)."

Selective IS/IT Outsourcing: "The decision to source selected IT functions from external provider(s) while still providing between 20% and 80% of the IT budget internally."

Service Level Agreement (SLA): "Details and/or quantifies in schedules the minimum level of service which the customer (i.e., client organisation) will obtain."

Total IS/IT Outsourcing: "The decision to transfer the equivalent of more than 80% of the IT budget for IT assets, leases, staff, and management responsibility to an external IT provider."

Cooperation of Geographic and Multidimensional Databases

Elaheh Pourabbas
National Research Council, Rome, Italy

INTRODUCTION

In recent years, the enormous increase of independent databases widely accessible through computer networks has strongly motivated the interoperability among database systems. Interoperability allows the sharing and exchange of information and processes in heterogeneous, independent, and distributed database systems. This task is particularly important in the field of decision support systems. These systems through the analysis of data in very large databases identify the unusual trends in particular applications for creating opportunities for new business or for forecasting production needs.

Currently, in the research community, geographic information systems (GISs) and multidimensional databases (MDDBs) are seen as the most promising and efficient information technologies for supporting decision making. Geographic information systems, which are geographic-database- (GDB) dependent, through graphic display functionalities and complex spatial data structures, facilitate the storage and manipulation of geographic data.

Multidimensional databases refer either to statistical databases (Chan & Shoshani, 1981; Rafanelli & Shoshani, 1990), which mostly represent applications in the socio-economic area, or OLAP (online analytical processing) databases (Agrawal, Gupta & Sarawagi, 1997; OLAP Council, 1997), which emphasize business applications. Similar to statistical databases, OLAP databases have a data model that represents one or more "summary measures" over a multidimensional space of "dimensions," where each dimension can be defined over a hierarchy of "levels" (Shoshani, 1997). In this area, mostly the aspects of handling the multidimensional data and summarizations over the dimensions have been largely investigated. However, unlike OLAP databases, statistical databases may have only the summarized data available for reasons of privacy. These databases are often referred to by the term "summary databases."

The common key elements between geographic and multidimensional data that allow effective support in data cooperating are basically *time* and *space*. In literature, space has been considered as a bridge element for cooperating GDB and MDDB, on which our attention will be focused.

A feature notably lacking in most GDBs is the capability of accessing and manipulating business data, which are stored in MDDBs. We tackle this task by a novel approach that shares a number of characteristics and goals with the approaches proposed in literature. They aimed at defining a set of operators applicable to either spatial or summary data without dealing with the "logical organization" of databases at all. Similar to these models, our approach is addressed for cooperative query answering but it provides a data model for the summary data manipulation in the context of GDB. In addition, the above-mentioned models are based on multidimensional data formed by solely one location dimension, whereas in our approach we also consider data defined by more than one location dimension and we analyze their effect on data modeling and query answering.

BACKGROUND

In the database community, the cooperation between GDB and MDDB is indicated by taking into account the notion of map generalization. Map generalization is intended to consider the impact of scale and resolution on spatial data querying (see Muller, Lagrange, & Weibel, 1995). In this context, some attempts have been made to look for a standard set of multidimensional (or statistical) operators based on aggregation and disaggregation.

Gargano, Nardelli, & Talamo (1991), for instance, have extended the relational algebra essentially by defining two algebraic operators that are able to manipulate either the spatial extension of geographic data or summary data. They are named *G-Compose* and *G-Decompose*. The first operator is denoted by $G-Compose_X(F_y;Y)$, where X and Y are two nonintersecting subsets of attributes of a relation R. In the case of spatial data, it "merges" all tuples of R which are already projected on Y in a single one whose Y value is generated by the application of the spatial fusion function. This function, which is represented by F_y, takes a subset of geometric attributes of a given type and returns a single geometric attribute of the same type. In the case of summary data, F_y aggregates the numeric value of Y attributes. The effect of G-Decompose is that all tuples of R projected on Y are "decomposed." In this proposal, the fundamental issues of hierarchies and data aggregation for either spatial or summary data have been omitted.

These issues are discussed later in an approach proposed by Rigaux and Scholl (1995). In this work, the authors make the bridge between the geographic and statistical disciplines by defining an aggregation technique over a hierarchy of space partitions. Their model is based on the concept of "partition" that is used for partitioning either geometric space or other sets (e.g., a set of people). They introduced a relation called *cover*, which is represented by the schema $O = \{A_1, \ldots, A_n, A_g\}$, such that $\pi_{A_g}(O)$ is a partition and there is a bi-univocal functional dependency between the attributes $A_1, \ldots, A_n$ and A_g. They defined the *geometric projection* operator on a subset of attributes $S = \{A_1, \ldots, A_q\}$ as follows:

$$apply_{\Sigma_{Geo}}\ (nest_S(\pi_{S,A_g}(O)),$$

where $\pi_{A_g}(O)$ is the N1NF grouping operation (see Abiteboul & Bidoit, 1986) on S, $nest_S(\pi_{S,A_g}O)$ gives the result with the schema $\{A_1, \ldots, A_q, B\}$ where $B = set(A_g)$, and Σ_{Geo} performs the spatial aggregation function UNION on attribute $B = set(A_g)$. They used the relation cover for representing summary data, in which each descriptive attribute A can be defined on a hierarchical domain. The same operator is redefined as below:

$$apply_{\Sigma_{Geo}}\ (nest_S(gen_{A:A'}(O))),$$

where before applying the nest operator, the abstraction level of hierarchy to which the attribute A belongs is changed. It is indicated by $gen_{A:A'}(O)$. Note, in this case Σ_{Geo} performs the numeric aggregation function SUM.

The model proposed by Rigaux and Scholl (1995) is addressed to generate maps in multiple representations of data using the hierarchy of space partitions and the hierarchy induced by a partial order relationship in the domain of an attribute. In this proposal only one location dimension hierarchy for summary data is considered.

While the above models give a formal definition for the cooperation between spatial and multidimensional environments, some other works consider the architectural aspects of an integration system. For instance, Kouba, Matousek, and Miksovsky (2000) tried to identify some requirements for the correct and consistent functionality of system interconnection. They proposed an integration module that has two different roles: One is the transformation of data from external data sources, and the other refers to the integration of GIS and data warehouse through their common components, that is, location. The GIS under consideration is based on an object-oriented model that identifies the basic GIS elements that are classes and objects. In the GIS system, the structure of the geographical class hierarchy is stored in a metadata object for accessing directly from the integration module. Furthermore, the implementation aspects of the integration of the Microsoft SQL Server 7 Data Warehouse and ArcView GIS System are discussed.

Sindoni G., De Francisci S., Paolucci M., and Tininini L. (2001) have considered the integration of several spatiotemporal data collections of the Italian National Statistics Institute. The integration system is defined mainly by a historical database containing the temporal variation of territorial administrative partitions, a statistical data warehouse providing statistical data from a number of different surveys, and a GIS providing the cartography of the Italian territory up to census tract level. The implemented cooperative system manages the maps of the temporal evolution of a certain number of administrative regions and links to these maps the content of the above-mentioned statistical database of a given year.

A LOGICAL APPROACH

For describing the logical data model, we consider an object-oriented GDB and a cube-based MDDB, the main components of which are indicated in Table 1.

Cooperation by Binding Elements

In order to clarify the need for such cooperation, let us consider the case of a GDB user wishing to display a query result on a map with different combinations of geographic and nongeographic data. Let the query be as follows: "Find all the Italian regions which are adjacent to the Tuscany region, in which the number of cars sold in 1990, in the case of <Corolla>, was greater than 10,000."

For answering this query, it is necessary not only to retrieve the adjacent regions, but also to perform some OLAP operations on the time and product dimensions of the Car_Sales cube shown in Figure 1a. The former analyzes the topological spatial relationship (i.e., adjacency) and can be performed only in GDB. The solution of such queries depends essentially on a data model that enables the component databases to cooperate but remain independent. The main idea of our approach is to explore the additional input that can come to GDB from MDDB.

As we have shown in Table 1 and in Figure 1b, an object-oriented geographic database is characterized by

Table 1. A summary of data models and components of databases

Geographic Database		*Multidimensional Database*	
Geographic Class	$gc = < n, sc, P, M >$, where - n is the name of gc - sc is the class - parent of gc - $P = A_g \cup A_{\bar{g}}$ is set of attributes, where $A_g = \{point, polyline, region\}$ is geometric type $A_{\bar{g}}$ is set of alphanumeric type attributes - M is set of methods go represents an instance of gc	Cube Schema	$C = < C_{name}, L, f_c >$, where - C_{name} denotes the name of the cube - L is a set of levels - $f_c : L_T \to M$ is a function by which a measure (M) is associated with the tuples defined over the combinations of levels (denoted by L_T)
Geographic Object	$go = < id, gc, V, T >$, where - id is the identifier of a go; it is unique - gc the class to which it belongs; it is also unique - V represents the ordered set of the alphanumeric attribute values - T is the ordered set of pairs (latitude, longitude)	Dimension	$D = < L, \preceq >$, where - L is a set of levels - $\preceq$ is a partial order over L's elements such that for each pair of levels l_i and l_j, $l_i \preceq l_j$ if $dom(l_i) \preceq dom(l_j)$ - for any two levels l' and l'', which belong to different dimensions, $l' \cap l'' = 0$
Relationships	- $Full-contains$ represents spatial inclusion relationship among geographic classes and geographic objects, and it is a total and surjective function. It introduces hierarchy over geographic classes/objects. - ISA represents inheritance relationship among class and its subclasses.	Relationship	- *Complete containment* relationship is a total and surjective function between any two levels of a given dimension. This relationship introduces hierarchy over different levels/instance levels of a dimension.
Function λ	l_g and I_g represent, respectively, a geographic level of a cube and its instance - Intensional level : $\lambda : gc \to l_g$ - Extensional level: $\lambda : go \to I_g$	Function λ^{-1}	l_g and I_g represent, respectively, a geographic level of a cube and its instance - Intensional level : $\lambda^{-1} : l_g \to gc$ - Extensional level: $\lambda^{-1} : I_g \to go$
Operators	Topological operators derived from the topological relationships and they are *disjoint, touch, equal, inside, contains, covers, covered by, and overlap* (for a detailed description see Egenhofer & Herring, 1990).	Operators	- $roll-up^{f_{agg}(f_c)}_{l_1 \to l_2}(C)$ allows level changing $l_1 \to l_2$ and yields more aggregate data - $slice^{f_{agg}(f_c)}_{L_2, L_3}(C)$ omits L_1 from the set of levels L_1, L_2, L_3; it is composed by projection and grouped by operators - $dice_\theta(C)$ performs selection according to predicate θ

classes that are organized through hierarchies whose common semantics are space inclusion. The model of MDDB is mainly characterized by the structure of its dimensions, hierarchies over dimensions, and measures. The actual cooperation is provided by linking dynamically the common spatial components of both databases. This link is based on two kinds of correspondences (see Function λ, and Function λ^{-1} in Table 1).

- *Level correspondence* maps a given geographic class of a geographic taxonomy to the corresponding level belonging to the location dimension and vice versa.
- *Level instance correspondence* results from the previous step and maps a generic instance of a geographic class to its corresponding instance of location dimension level and vice versa. In this step, for a correct correspondence any instance changes should be propagated in both environments.

The link extends the geographic class schema through some special attributes, called *binding attributes*, which are added to the set of alphanumeric attributes (see Pourabbas, 2003). A binding attribute of a geographic class is a cube derived from a given cube in a multidimensional database whose geographic level is mapped to the corresponding class name. In Figure 1a, for instance, one of the binding attributes of the geographic class MUNICIPALITY is:

$$< Car\_Sales, \{ model, month\}, f_{Car\_Sales} >,$$

where

$$f_{Car\_Sales} : \{model : Model, month : Month\}_{\mathcal{T}} \rightarrow numeric .$$

The schema of this binding attribute is identical to the schema of the cube *Car_Sales* in MDDB. However, the set of its levels is defined by the levels of the cube except one, Municipality, which is mapped to the class name MUNICIPALITY by λ^{-1}.

Note that the λ, λ^{-1}, full-contains, and complete containment functions indicated in Table 1 play a central role either in linking the heterogeneous environments or in providing data that lack in one environment by knowing data and their relationships in another. For instance, let us suppose only the geographic variable Municipality be defined in MDDB, and let us consider two geographic classes MUNICIPALITY and PROVINCE in GDB (see Figure 1b) between which the full-contains relationship holds. The geographic level Province can be generated by λ, and the complete containment relationship that is equivalent to the full-contains relationship is defined from Municipality to Province.

Generally speaking, a cube can be defined with more geographic levels. The different geographic levels of a cube reflect the cases where each of them is a specialization of a given level of possibly the same location hierarchy. For instance, let us consider the Shoe-Sales cube defined by Province of Distribution, Region of Production, and Month shown in Figure 2.

The geographic levels of such a cube are two specializations of Province and Region in the location dimension hierarchy: Municipality→Province→Region. Consequently, their correspondences obtained by the function λ^{-1} would also have been specializations of the geographic classes PROVINCE and REGION, named PROVINCE OF DISTRIBUTION and REGION OF PRODUCTION. The process of binding attribute generation will yield the subcubes illustrated in Figure 2 in the schema of the above-mentioned geographic classes. The instances of these geographic classes will assume different semantics. A given instance of the first geographic class represents the province of the distribution of shoes which are made in several regions and months, while in the second case an instance represents a region of shoe production in which the items are distributed over several provinces and months.

Query Resolution

Let us recall the query mentioned in the previous subsection. The procedure for solving this query is defined by the following steps.

1) Identify the geographic class and object(s) involved by the geographic operators in the query, and give the results of these operations.
   In our query, they are Region and Tuscany, and the geographic operator is Touch. The result of this

Figure 1. Binding attributes and hierarchies

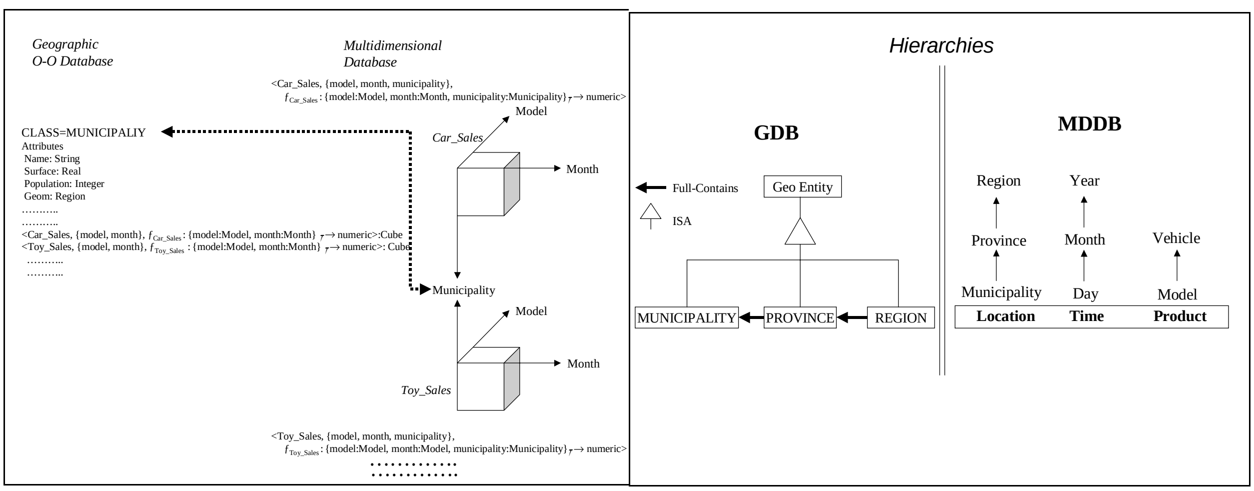


Figure 2. Binding attributes in the case of cube with more geographic levels

Cube

$< Shoe\_Sales, \{Province\ of\ distribution, Region\ of\ production, Month\}, f_{Shoe\_Sales} >$

Specializations of Geographic Classes

CLASS = PROVINCE OF DISTRIBUTION
Attributes
Name : String
....
$< \text{Shoe\_Sales}, \{\text{Region of production, month}\}, f_{Shoe\_Sales} >$: Cube

CLASS = REGION OF PRODUCTION
attributes
name : string
....
$< \text{Shoe\_Sales}, \{\text{Province of distribution, month}\}, f_{Shoe\_Sales} >$: Cube

operator is formed by the instances Liguria, Emilia-Romagna, Umbria, and Lazio.

2) Identify the target of the query that forms the binding attribute in the geographic class.
In our case, it is Car_Sales in regions.
3) Select the cube identified by the binding attribute in MDDB.
In our query, select the cube Car_Sales.
4) Perform aggregations over dimensions of the cube. In our case, it means to obtain Car-Sales by Region, Model, Year from Car-Sales by Municipality, Model, Month. It requires performing two roll-up operations over the latter cube along the Location dimension, which are Municipality → Province and Province → Region. These operations yield Car-Sales by Region, Model, Month. Then, it should be followed by another aggregation over the Time dimension (Month → Year) for obtaining Car-Sales by Region, Model, Year.
5) Select the values of the measure that satisfy the constraints.

In our query, the constraints are Year = 1990, Model = Corolla, and Car_Sales > 10,000. The final measure is obtained by a dice (or selection) operation.

6) Pass the list of the geographic level instances that satisfy the query to GDB.

In our query, it means to compare the list of regions provided by Step 6 with the list of Step 1 in order to identify and visualize the maps of the regions that satisfy the query constraints.

Inferring Binding Attributes

An important aspect related to the concept of binding attribute is *inferring*, with which we mean the maximum number of cubes inferred from a given binding attribute. The inferred cubes can be classified into two categories named covering and deduced.

The covering attributes are based on a process that is able to generate new binding attributes for the parent class based on those of its subclasses, which are related to each other by the full-contains relationship. One important issue related to full-contains, which is based on the well-known "part-whole" relationship, is "value propagation" (Halper, Geller & Perl, 1993; Liu & Halper, 1999). It reflects the interlevel flow of values in the classes of objects organized by hierarchies and allows the "driving" of attributes. Therefore, in order to assimilate the power of value propagation in the context of the full-contains relationship, we introduce the *is-derived* relationship between any two geographic classes or objects. For instance, let us consider the geographic classes PROVINCE and MUNICIPALITY in Figure 1b, which are related by the full-contains relationship. As a consequence of the is-derived relationship, the binding attribute of the PROVINCE (e.g., Milan) can be inferred from the binding attribute $< Car\_Sales, \{model, month\}, f_{Car\_Sales} >$ of the MUNICIPALITY, as follows:

$$slice_{model,month}(dice_{province=Milan}($$

$$(roll-up^{sum(Car\_Sales)}_{Municipality \rightarrow \text{Province}}(Car\_Sales)))$$

The deduced attributes may result from the aggregation over the levels of a given binding attribute in a geographic class through roll-up and slice (or projection) operators. In the case of Car-Sales by Model, Month of the class MUNICIPALITY shown in Figure 1a, they are Car-Sales by Model, Car-Sales by Month, Car-Sales by Vehicle, Month, Car-Sales by Model, Year, Car-Sales by Vehicle, Year, Car-Sales by Vehicle, Car-Sales by Year, and Car-Sales that is the grand total.

Overall, the number of base, covering, and deduced binding attributes of a cube is given by the following formula:

$$A_{Tot} = \left(\prod_{k=1}^{N} (f_k + 1) \right) \left(\sum_{h=1}^{G} \frac{l_h}{l_h + 1} \right),$$

where N represents the total number of levels (geographic and nongeographic), G represents the number of geographic levels, l_h represents the number of base and covering binding attributes generated by a geographic level, and f_k is the number of deduced binding attributes.

The covering and deduced attributes indicate essentially the set of possible answers to a class of queries that can be formulated on the base binding attributes. This fact highlights the usability of binding attributes.

FUTURE TRENDS

We have shown that our approach provides the interoperability of geographic and multidimensional databases through their binding elements, and guarantees the autonomy of data sources. In other words, unlike the models proposed in literature, our approach avoids a different application of a unique set of operators to heterogeneous data sources. This fact enables us to interact with each component database in order to benefit from its own functionalities, and at the same time keep them cooperative through a set of functions.

The discussion up to this point has focused upon the space feature for realizing the interoperability. However, the role of the time feature in such cooperation forms an interesting issue to be considered in future work.

An important assumption made in this work concerns the summarizability of a data cube, which means no cell contains a “not available” value and there is no missing value. In many cases the partial availability of data is realistic and, consequently, the comprehensive aggregation of summary data is compromised. The answering of joint aggregate queries formulated on cubes with missing and/or nonavailable data values can create another open problem.

We discussed the problem of answering spatiosummary queries in the context of GDB. Research that considers this question in the special light of MDDB could raise another interesting issue to be examined.

CONCLUSION

We have presented a logical data model, based on the extension of the geographical data model, which provides the capability of answering queries by the cooperation with MDDB. The cooperation of spatiosummary data is obtained by introducing new attributes into the geographic data structure, called binding attributes. The main idea behind them is to retrieve and manipulate data stored in MDDB and invoked from queries formulated in GDB.

REFERENCES

Abiteboul, S., & Bidoit, N. (1986). Non first normal form relations: An algebra allowing data restructuring. *Journal of Computer and System Sciences*, *33*(3), 361-393.

Agrawal, R., Gupta, A., & Sarawagi, S. (1997). Modelling multidimensional databases. *Thirteenth International Conference on Data Engineering*, 232-243.

Chan, P., & Shoshani, A. (1981). SUBJECT: A directory driven system for organizing and accessing large statistical databases. *Conference on Very Large Data Bases: VLDB 1981*, 553-563.

Egenhofer, M. J., & Herring, J. (1990). A mathematical framework for the definition of topological relationships. *Proceedings of the Fourth International Symposium on Spatial Data Handling*, 803-813.

Gargano, M., Nardelli, E., & Talamo, M. (1991). Abstract data types for the logical modelling of complex data. *Information Systems*, *16*(5), 565-583.

Halper, M., Geller, J., & Perl, Y. (1993). *Value propagation in object-oriented database part hierarchies.* ACM CIKM’93, Washington, DC.

Kouba, Z., Matousek, K., and Miksovsky, P. (2000). On Data Warehouse and GIS Integration. *Lecture Notes in Computer Science*, *N. 1873*, pp. 604 – 613.Springer Verlag, Contance, Germany.

Liu, L. M., & Halper, M. (1999). *Incorporating semantic relationships into an object-oriented database system.* Thirty-Second Annual Hawaii International Conference on System Sciences, HI.

Muller, J. C., Lagrange, J. P., & Weibel, R. (Eds.). (1995). *GISs and generalization.* London: Taylor & Francis.

OLAP Council. (1997). *The OLAP glossary.* Retrieved 1997 from *http://www.olapcouncil.org*

Pourabbas, E. (2003). Cooperation with geographic databases. In *Multidimensional Databases: Problems and Solutions* (chap. 13). Hershey, PA: Idea Group Publishing.

Rafanelli, M., & Shoshani, A. (1990). STORM: A statistical object representation model. *Lecture notes in computer science* (Vol. 420, pp. 14-29). Springer Verlag, Charlotte, NC.

Rigaux, P. and Scholl, M. (1995). Multi-Scale Partitions: Application to Spatial and Statistical Databases. *Lecture Notes in Computer Science* (N. 951), 170-184, Springer-Verlag, Berlin.

Shoshani, A. (1997). OLAP and statistical databases: Similarities and differences. *Proceedings of the 16th ACM SIGACT-SIMOD-SIGART Symposium on Principles of Database Systems*, 185-196.

Sindoni G., De Francisci S., Paolucci M., and Tininini L. (2001). Experiences in Developing a Spatio-Temporal Information Systems. In Procceddings of ETK '2001 – NTTS '2001, Exchange of Technology and Know-how New Techniques and Technologies for Statistics, *18- 22 June 2001, Crete Greece*, 747-756.

KEY TERMS

Cube: It is a group of data cells arranged by the dimensions of the data. Assigning a value to each dimension of a cube, the measure is obtained by a mapping from this assignment.

Data Warehouse: A system for storing, retrieving, and managing large amounts of data using some sophisticated techniques of cleaning, filtering, hashing, and compression.

Map Generalization: This process is used to convert spatial data from one scale-dependent representation into another by calculating the geometry of a more abstract object through the union of the geometries of lower level objects.

Hierarchy: The structures created by an abstraction process by which a selection of a set of attributes, objects, or actions from a much larger set of attributes, objects, or actions according to certain criteria is defined.

OLAP: Online analytical processing is a category of software technology that allows users to transform raw data according to predefined or user-defined functions, and quickly and interactively performs slice, dice, and roll-up operations in various dimensions.

Spatial Partition: The subdivision of the 2D plane into pairs of disjoint regions where each region is associated with an attribute that can have a simple or complex geometric structure.

Topological Eelationships: The intersections of boundary and interior of two regions embedded in R^2.

Credit Card Users' Data Mining

André de Carvalho
USP São Carlos, Brazil

Antonio P. Braga
UFMG, Brazil

Teresa Ludermir
UFPE, Brazil

MINING CREDIT RISK ANALYSIS DATA

The widespread use of databases and the fast increase in the volume of data they store are creating problems and new opportunities for credit companies. These companies are realizing the necessity of making efficient use of the information stored in their databases, extracting useful knowledge to support their decision-making processes.

Nowadays, knowledge is the most valuable asset a company or nation may have. Several companies are investing large sums of money in the development of new computational tools able to extract meaningful knowledge from large volumes of data collected over many years. Among such companies, companies working with credit risk analysis have invested heavily in sophisticated computational tools to perform efficient data mining in their databases.

Credit risk analysis is concerned with the evaluation of the profit and guaranty of a credit application. A typical credit risk analysis database is composed of several thousands of credit applications. These credit applications can be related to either companies or people. Examples of personal credit applications include for student loans, personal loans, credit card concessions, and home mortgages. Examples of company credits are loans, stocks, and bonds (Ross, Westerfield, & Jaffe, 1993).

Usually, the higher the value of the credit asked, the more rigorous is the credit risk analysis. Some large companies have whole departments dedicated to credit risk analysis.

The traditional approach employed by bank managers largely depends on their previous experience and does not follow the procedures defined by their institutions. Besides, several deficiencies in the data set available for credit risk assessment, together with the high volume of data currently available, make the manual analysis almost impossible. The treatment of these large databases overcomes the human capability of understanding and efficiently dealing with them, creating the need for a new generation of computational tools and techniques to perform automatic and intelligent analysis of large databases.

Credit analysis is essentially a classification task that involves the evaluation of the reliability and profitability of a credit application. The application of data mining techniques for credit risk analysis may provide important information that can improve the understanding of the current credit market and support the work of credit analysts (Carvalho, Braga, Rezende, Ludermir, & Martineli, 2002; Eberlein, Breckling, & Kokic, 2000; He, Hawkins, Graco, & Yao, 2000; Horst, Padilha, Rocha, Rezende, & Carvalho, 1998).

The extraction of useful knowledge from large databases is called knowledge discovery in databases (KDD). KDD is a demanding task and requires the use of sophisticated computing techniques (Brachman & Anand, 1996; Fayyad, Piatetsky-Shapiro, Amith, & Smyth, 1996). The recent advances in hardware and software make possible the development of new computing tools to support such tasks. According to Fayyad, Piatetsky-Shapiro, Amith, and Smyth (1996), KDD comprises a sequence of stages:

1. Selection
2. Preprocessing
3. Transformation
4. Data mining
5. Interpretation/evaluation

It is then important to stress the difference between KDD and data mining (DM). While KDD denotes the whole process of knowledge discovery, DM is a component of this process. The DM stage is used as the extraction of patterns or models from observed data. KDD can be understood as a process that contains, at least, the steps of application domain understanding, data selection and preprocessing, DM, knowledge evaluation and consolidation, and use of the knowledge. At the core of the knowledge discovery process, the DM step usually takes only a small part (estimated at 15–25%) of the overall effort (Brachman & Anand, 1996).

The KDD process begins with the understanding of the application domain, considering aspects such as the objectives of the application and the data sources. Next, a representative sample, selected according to statistical techniques, is removed from the database, preprocessed, and submitted to the methods and tools of the DM stage, with the objective of finding patterns/models (knowledge) in the data. This knowledge is then evaluated regarding its quality and usefulness, so that it can be used to support a decision-making process.

Credit analysis databases usually cover a huge number of transactions performed over several years. The analysis of these data may lead to a better understanding of the customer's profile, thus supporting the offer of new products or services. These data usually hold valuable information, e.g., trends and patterns, that can be employed to improve credit assessment. The large amount makes its manual analysis an impossible task. In many cases, several related features need to be simultaneously considered in order to accurately model credit user behavior. This need for automatic extraction of useful knowledge from a large amount of data is widely recognized.

DM techniques are employed to discover strategic information hidden in large databases. Before they are explored, these databases are cleaned. Next, a representative set of samples is selected. Machine-learning techniques are then applied to these selected samples. The use of data mining techniques on a credit risk analysis database allows the extraction of several relevant information regarding credit card transactions.

The data present in a database must be adequately prepared before data mining techniques can be applied to it. The main steps employed for data preparation are as follows:

1. Preprocess the data to the format specified by the algorithm to be used
2. Reduce the number of samples/instances
3. Reduce the number of features/attributes
4. Features construction, which is the combination of one or more attributes in order to transform irrelevant attributes to more significant attributes
5. Elimination of noise and treatment of missing values

Once the data have been prepared, machine-learning (ML) techniques can be employed to discover useful knowledge. The quality of a knowledge extraction technique can be evaluated by different measures, such as accuracy, comprehensibility, and new, useful knowledge.

CRITICAL ISSUES OF CREDIT RISK ANALYSIS DATA MINING

There are, of course, a few problems that can arise from the use of data mining techniques for credit risk analysis. The main advantages and problems associated with this technology are briefly presented in Table 1. Current Internet technology may allow the invasion of company databases, which may lead to the access of confidential data. Besides, the stored data may be made available to other companies or used without the knowledge or authoriza-

Table 1. A summary of critical issues of credit risk analysis data mining

Training of employees Credit analysts should be trained with the new tools available and shown the benefits of the new technology
User ignorance and perceptions Lack of adequate understanding of the data mining and its usefulness
Inadequate information Applicants may have supplied incorrect information,intentionally or not
Maintain integrity of data Maintain up-to-date and accurate information on the databases
Security Maintain secure and safe systems and denying unauthorized users access
Privacy and confidentiality agreements Address individual's right to privacy and the sharing of confidential information

tion of the customer. On the other hand, the extraction of useful information from credit risk analysis databases should result in benefits not only to the companies, but also to the customers, with the offering of better and more convenient credit products.

CONCLUSION

Several companies perform data mining on personal data stored in their databases. This is particularly true for credit risk analysis databases, where customers' personal data can be employed to improve the quality of credit assessment and support the offer of new credit products. Data mining systems usually employ artificial intelligence and statistical techniques to acquire relevant, interesting, and new knowledge for most applications, including credit risk analysis. Although one cannot avoid the use of these sophisticated techniques to support credit assessment, great care should be taken to guarantee privacy and that personal rights are not violated when working with private information.

REFERENCES

Brachman, R., & Anand, T. (1996). *The process of knowledge discovery in databases: A human-centered approach* (pp. 37–57). Cambridge, MA: AAAI Press/The MIT Press.

Carvalho, A., Braga, A., Rezende, S., Ludermir, T., & Martineli, E. (2002). Understanding credit card users behaviour: A data mining approach. In R. A. Sarker, H. A. Abbass, & C. S. Newton (Eds.), *Heuristic and optimization for knowledge discovery* (chap. 14; pp. 240–261). Hershey, PA: Idea Group Publishing.

Eberlein, E., Breckling, J., & Kokic, P. (2000). A new framework for the evaluation of market and credit risk. In B. G. Nakhaeizadeh & K. -H. Vollmer (Eds.), Datamining and computational finance (pp. 51–67). Heidelberg: Physica-Verlag.

Fayyad, U., Piatetsky-Shapiro, G., Amith, S., & Smyth, P. (1996). *From data mining to knowledge discovery: An overview* (pp. 1–34). Cambridge, MA: AAAI Press/The MIT Press.

Giudici, P. (2003). *Applied data mining: Statistical methods for business and industry*. New York: Wiley.

He, H., Hawkins, S., Graco, W., & Yao, X. (2000). Application of genetic algorithm and k-nearest neighbour method in real world medical fraud detection problem. Journal of Advanced Computational Intelligence, 4(2), 130–137.

Horst, P., Padilha, T., Rocha, C., Rezende, S., & de Carvalho, A. (1998, May). Knowledge acquisition using symbolic and connectionist algorithms for credit evaluation. In *Proceedings of the IEEE World Congress on Computational Intelligence, WCCI'98*, Anchorage, USA.

Mendes Filho, E., de Carvalho, A., & Matias, A. (1997). Credit assessment using evolutionary MLP networks. Decision technologies for computational finance. In *Proceedings of the Fifth International Conference Computational Finance, CF'97* (pp. 365–371). London. Advances in Computational Management Science. Dordrecht: Kluwer Academic Publishers.

Ross, S., Westerfield, R., & Jaffe, J. (1993). *Corporate finance*. New York: Richard D. Irwin, Inc.

KEY TERMS

Consumer Credit: A loan to an individual to purchase goods and services for personal, family, or household use.

Credit: Delivery of a value in exchange of a promise that this value will be paid back in the future.

Credit Scoring: A numerical method of determining an applicant's loan suitability based on various credit factors, such as types of established credit, credit ratings, residential and occupational stability, and ability to pay back loan.

Data: The set of samples, facts, or cases in a data repository. As an example of a sample, consider the field values of a particular credit application in a bank database.

Data Mining: The process of extracting meaningful information from very large databases. It is one of the main steps of the KDD process.

KDD: This is the process of knowledge discovery in large databases.

Knowledge: Knowledge is defined according to the domain, considering usefulness, originality, and understanding.

Machine Learning: Subarea of artificial intelligence that includes techniques able to learn new concepts from a set of samples.

Critical Realism as an Underlying Philosophy for IS Research

Philip J. Dobson
Edith Cowan University, Australia

INTRODUCTION

Many recent articles from within the information systems (IS) arena present an old-fashioned view of realism. For example, Iivari, Hirschheim, and Klein (1998) saw classical realism as seeing "data as describing objective facts, information systems as consisting of technological structures ('hardware'), human beings as subject to causal laws (determinism), and organizations as relatively stable structures" (p. 172). Wilson (1999) saw the realist perspective as relying on "the availability of a set of formal constraints which have the characteristics of abstractness, generality, invariance across contexts."

Fitzgerald and Howcroft (1998) presented a realist ontology as one of the foundational elements of positivism in discussing the polarity between hard and soft approaches in IS. Realism is placed alongside positivist, objectivist, etic epistemologies and quantitative, confirmatory, deductive, laboratory-focussed and nomothetic methodologies. Such a traditional view of realism is perhaps justified within the IS arena, as it reflects the historical focus of its use, however, there now needs to be a greater recognition of the newer forms of realism—forms of realism that specifically address all of the positivist leanings emphasised by Fitzgerald and Howcroft (1998). A particular example of this newer form of realism is critical realism. This modern realist approach is primarily founded on the writings of the social sciences philosopher Bhaskar (1978, 1979, 1986, 1989, 1991). The usefulness of such an approach has recently been recognized in the IS arena by Dobson (2001) and Mingers (2002).

BACKGROUND

Bhaskar's brand of realism (referred to by Searle, 1995, as a form of external realism) argues that there exists a reality totally independent of our representations of it; the reality and the "representation of reality" operating in different domains—roughly a transitive epistemological dimension and an intransitive ontological dimension. For the realist, the most important driver for decisions on methodological approach will always be the intransitive dimension—the target being to unearth the real mechanisms and structures underlying perceived events. Critical realism acknowledges that observation is value laden, as Bhaskar pointed out in a recent interview:

...there is no conflict between seeing our scientific views as being about objectively given real worlds, and understanding our beliefs about them as subject to all kinds of historical and other determinations. (Norris, 1999)

The critical realist agrees that our knowledge of reality is a result of social conditioning and thus cannot be understood independently of the social actors involved in the knowledge derivation process. However, it takes issue with the belief that the reality is a product of this knowledge derivation process. The critical realist asserts that "real objects are subject to value laden observation"; the *reality* and the value-laden *observation of reality* operate in two different dimensions, one intransitive and relatively enduring and the other transitive and changing.

An important aspect of a critical realist approach is that it not only provides direction on the characteristics and behaviour of the underlying objects of enquiry, but it also provides direction as to how to examine these objects. The philosophy is presented as an underlabourer to social enquiry in that it can help with "clearing the ground a little...removing some of the rubbish that lies in the way of knowledge" (Locke, 1894, p. 14). This integral and important role for philosophy in the enquiry process can help to avoid many potentially false pathways and avenues.

For example, Bhaskar (1979) presented fundamental difficulties with the way that prediction and falsification have been used in the open systems evident within the social arena. For the critical realist, a major issue with social investigation is the inability to create closure—the aim of "experiment" in the natural sciences. Bhaskar (1979) argued that this inability implies that theory cannot be used in a predictive manner and can only play an explanatory role in social investigations, because:

...in the absence of spontaneously occurring, and given the impossibility of artificially creating, closed systems , the human sciences must confront the problem of the direct scientific study of phenomena that only manifest

themselves in open systems—for which orthodox philosophy of science, with its tacit presupposition of closure, is literally useless. In particular it follows from this condition that criteria for the rational appraisal and development of theories in the social sciences, which are denied (in principle) decisive test situations, cannot be predictive and so must be exclusively explanatory. (p. 27)

As Mingers (2002) suggested, such an argument has specific ramifications with respect to the use of statistical reasoning to predict future results. Bhaskar (1979) argued that the primary measure of the "goodness" of a theory is in its explanatory power. From Bhaskar's perspective, predictive use of theories is not possible in open social systems, and therefore, predictive power cannot be a measure of goodness. From this point of view, theory acts as primarily an explanatory tool for explaining events in hindsight.

Critical realism uses abductive or retroductive reasoning as its main focus. Positivist approaches are associated more with deductive or inductive reasoning. Deductive reasoning is the fundamental reasoning of mathematics, whereby some statement "*p*" leads to implications "*q*"—a movement from the general to the particular. For example, the general claim that "all crows are black" moves to the particular inference that the next one seen will be black. For the crows example, retroductive or abductive reasoning follows from an observation of numerous black crows to a theory as to a mechanism to explain why crows are disposed to be black. As Mingers (2002) described:

We take some unexplained phenomenon and propose hypothetical mechanisms that, if they existed, would generate or cause that which is to be explained. So, we move from experiences in the empirical domain to possible structures in the real domain. This does not of itself prove that the mechanism exists, and we may have competing explanations, so the next step is to work toward eliminating some explanations and supporting others. (p. 300)

Outhwaite (1987) similarly suggested that the critical realist method involves "the postulation of a possible [structure or] mechanism, the attempt to collect evidence for or against its existence and the elimination of possible alternatives" (p. 58). The realist agrees that we have a good explanation when (a) the postulated mechanism is capable of explaining the phenomenon, (b) we have good reason to believe in its existence, and (c) we cannot think of any equally good alternatives. Such an explanatory target suggests that philosophical considerations must play an important role in the critical realist method, because such an approach often requires transcending, or speculating, perhaps nonobservable mechanisms and structures to explain perceived happenings. Such initial proposition is transcendental or metaphysical in its focus, and as such, any explanation or discovery made is seen to be fallible and extendable as knowledge grows. As Wad (2001) argued:

If we take explanation to be the core purpose of science, critical realism seems to emphasise thinking instead of experiencing, and especially the process of abstraction from the domains of the actual and the empirical world to the transfactual mechanisms of the real world. (p. 2).

This type of thinking is called transcendental by Bhaskar, in that it gives an important role to the crossing of the divide between the empirical and speculative activities of scientific work. As Wad pointed out, this is necessary because often the experienced world of events is not explainable in terms of the empirical facts but only by way of incorporating nonexperienced mechanisms incorporated in objects that may be within or outside our domain of investigation.

RESEARCH IMPLICATIONS

Sayer (2000) contended: "Compared to positivism and interpretivism, critical realism endorses or is compatible with a relatively wide range of research methods, but it implies that the particular choices should depend on the nature of the object of study and what one wants to learn about it" (p. 19). As Mingers (2002) suggested, critical realism supports methodological pluralism in that it suggests that an external reality is open to multiple interpretations and understandings.

Yet, critical realism also has important things to say about the objects of enquiry in that it is an ontologically bold philosophy (Outhwaite, 1987, p. 34). It not only encompasses an external realism in its distinction between the world and our experience of it, but it also suggests a stratified ontology and a so-called depth realism in defining the objects that make up such a world. This concept suggests that reality is made up of three ontologically distinct realms: first, the empirical, that is experience; second, the actual, that is events (i.e., the actual objects of experience); and third, the transcendental, nonactual or deep, that is structures, mechanisms, and associated powers. This so-called depth realism proposes that "the world is composed not only of events and our experience or impression of them, but also of (irreducible) structures and mechanisms, powers and tendencies, etc. that, although not directly observable, nevertheless underlie actual events that we experience and govern or produce them" (Lawson, 1997, p. 8). Critical realism, in its use of retroduction, involves a movement from a surface phenomenon to a deeper causal thing; it involves the steady unearthing of deeper levels of structures and mechanisms.

The ontological complexity assumed by critical realism is, however, matched with a conservative epistemology heavily dependent on scientific argument. As Wad (2001) said, critical realism has little to say concerning practical advice:

One may get the feeling that critical realism develops a huge superstructure of ontological and epistemological insights, but when it comes to practical research we are left with the usual methodological suspects, delivered by inductivists, positivist and empirical-analytical scientists. (p. 12)

As Stones (1996) argued, realist methodologies need to be able to account for the underlying ontological richness they implicitly assume. They also need to reflect the belief that any knowledge gains are typically provisional, fallible, incomplete, and extendable. Realist methodologies and writings, thus, must reflect a continual commitment to caution, scepticism, and reflexivity.

FUTURE TRENDS

Critical realism is becoming influential in a range of disciplines, including geography (Pratt, 1995; Yeung, 1997), economics (Lawson, 1997; Fleetwood, 1999), organization theory (Tsang & Kwan, 1999), accounting (Manicas, 1993), human geography (Sayer, 1985), nursing (Ryan & Porter,1996; Wainwright, 1997), logistics and network theory (Aastrup, 2002), and library science (Spasser, 2002). The application of critical realism within the IS field has been limited to date. Mutch (1995, 1999, 2002) has applied critical realist thinking in the examination of organizational use of information. In so doing, he commented on how difficult it is to apply such a wide-ranging and sweeping philosophical position to day-to-day research issues. Mingers (2001, 2002) examined the implications of a critical realist approach, particularly in its support for pluralist research. Dobson (2001, 2002) argued for a closer integration of philosophical matters within IS research and proposed a critical realist approach. Information systems are social systems, and it makes sense to apply a modern social philosophy such as critical realism to their examination.

CONCLUSION

In researching the social context within which information technology (IT) and IS operate, a modern social philosophy such as critical realism has considerable potential. It can provide useful insight into the type of (retroductive) questions that may be asked and also the means by which an examination can progress. The integrated nature of the philosophy encourages a consistency in research in that it recognizes the tripartite connections between ontology, methodology, and practical theory. As Archer (1995) argued:

The social ontology endorsed does play a powerful regulatory role vis-à-vis the explanatory methodology for the basic reason that it conceptualises social reality in certain terms. Thus identifying what there is to be explained and also ruling out explanations in terms of entities or properties which are deemed non-existent. Conversely, regulation is mutual, for what is held to exist cannot remain immune from what is really, actually or factually found to be the case. Such consistency is a general requirement and it usually requires two-way adjustment. (p. 17)

The required consistency between ontological and epistemological matters has led to the critical realist observation that many of the things that traditionally have been done in researching the social arena are actually inconsistent with the underlying nature of the social objects proposed (see, for example, Lawson, 1997, who presented some of the inconsistencies evident in traditional use of economic theory in the social arena). There are, however, few practical examples of the use of critical realism in the IS field and the obvious potential it has is not yet realized. While, as Wilson (1999) observed, the realist argument has shown a remarkable resilience and resourcefulness in the face of the "rise of relativism," and more practical examples of its application need to be developed. Carlsson (2003) examines IS evaluation from a critical realist perspective.

REFERENCES

Aastrup, J. (2002). *Networks producing intermodal transport*. Ph.D. dissertation, Copenhagen Business School.

Archer, M. (1995). *Realist social theory: The morphogenetic approach*. New York; London: Cambridge University Press.

Banville, C., & Landry, M. (1989). Can the field of MIS disciplined? *Communications of the ACM, 32*(1), pp. 48–60.

Bhaskar, R. (1978). *A realist theory of science*. Sussex: Harvester Press.

Bhaskar, R. (1979). *The possibility of naturalism*. Hemel Hempstead: Harvester Wheatsheaf.

Bhaskar, R. (1986). *Scientific realism and human emancipation*. London: Verso.

Bhaskar, R. (1989). *Reclaiming reality: A critical introduction to contemporary philosophy*. London: Verso.

Bhaskar, R. (1991). Philosophy and the idea of freedom. Oxford: Blackwell.

Carlsson, S.A. (2003). Advancing information systems evalution (research): A critical realist approach. *Electronic Journal of Information Systems Evalution*, (6)2, 11-20.

Dobson, P. (2001). The philosophy of critical realism—An opportunity for information systems research. *Information Systems Frontiers, (3)*2, 199-210

Dobson, P. (2002, January). Critical realism and IS research—Why bother with philosophy? *Information Research*. Retrieved from *http://InformationR.net/ir/*

Fitzgerald, B., & Howcroft, D. (1998). Towards dissolution of the IS research debate: From polarization to polarity. *Journal of Information Technology, 13*, 313–326.

Fleetwood, S. (Ed.). (1999). *Critical realism in economics: Development and debate*. London: Routledge.

Iivari, J., Hirschheim, R., & Klein, H. K. (1998). A paradigmatic analysis contrasting information systems development approaches and methodologies. *Information Systems Research, 9*(2), 164–193.

Kuhn, T. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago, IL: The University of Chicago Press.

Lawson, T. (1997). *Economics and reality*. London: Routledge.

Mingers, J. (2001). Combining IS research methods: Towards a pluralist methodology. *Information Systems Research, 12*(3), 240–259.

Mingers, J. (2002). Real-izing information systems: Critical realism as an underpinning philosophy for information systems. In *Proceedings of the 23rd International Conference on Information Systems* (pp. 295–303).

Norris, C. (1999). Bhaskar interview. *The Philosophers' Magazine, 8*, 34

Outhwaite, W. (1987). *New philosophies of social science: Realism, hermeneutics, and critical theory*. New York: St. Martin's Press.

Pratt, A. (1995). Putting critical realism to work: The practical implications for geographical research. *Progress in Human Geography, 19*(1), 61–74.

Ryan, S., & Porter, S. (1996). Breaking the boundaries between nursing and sociology: A critical realist ethnography of the theory–practice gap. *Journal of Advanced Nursing, 24*, 413–420.

Sayer, A. (1985). Realism in geography. In R. J. Johnston (Ed.), *The future of geography* (pp. 159–173). London: Methuen.

Sayer, A. (2000). *Realism and social science*. Thousand Oaks, CA: Sage.

Sayer, R. A. (1992). *Method in social science: A realist approach*. London: Routledge.

Searle, J. R. (1995). *The construction of social reality*. New York: Free Press.

Spasser, M. A. (2002). *Realist activity theory for digital library evaluation: Conceptual framework and case study. Computer Supported Cooperative Work* (Vol. 11, pp. 81–110). Dordrecht: Kluwer Academic Publishers.

Stones, R. (1996). *Sociological reasoning: Towards a past-modern sociology*. New York: Macmillan.

Tsang, E., & Kwan, K. (1999). Replication and theory development in organizational science: A critical realist perspective. *Academy of Management Review, 24*(4), 759–780.

Wad, P. (2001). *Critical realism and comparative sociology*. Draft paper for the IACR conference, 17–19 August.

Wainwright, S. P. (1997). A new paradigm for nursing: The potential of realism. *Journal of Advanced Nursing, 26*, 1262–1271.

Wilson, F. (1999). Flogging a dead horse: The implications of epistemological relativism within information systems methodological practice. *European Journal of Information Systems, 8*(3), 161–169.

KEY TERMS

Author's Note: In philosophy, definitions become a basis for debate—they often reflect the area from which the author derives. Perhaps the following reflect realist origins.

Closed and Open Systems: A *closed system* is one restricted in such a way that laws have uniform effects. An *open system* is one that is not closed. Closed systems do not usually occur spontaneously in nature and generally require human intervention, such as in laboratory experiments (from www.raggedclaws.com/criticalrealism).

Critical Realism: The careful or critical application of the scientific approach to the social sciences.

Epistemology: The study of knowledge or how we come to know.

Ontology: The study of what exists.

Philosophy: "The critical examination of the grounds for fundamental beliefs and an analysis of the basic concepts employed in the expression of such beliefs" (*Encyclopaedia Britannica*, p. 388, *Micropedia*, Vol. 9, 1985) or the "rational, methodical, and systematic consideration of those topics that are of greatest concern to man" (*Macropedia*, Vol. 25, p. 742, 1985 edition).

Realism: The belief that there is a reality independent of our perceptions of it.

Transitive and Intransitive Dimensions: The *intransitive dimension* in the philosophy of science corresponds roughly to ontology, and the *transitive dimension* corresponds roughly to epistemology. Intransitive objects exist and act independently of our knowledge of them (except when we use our knowledge to intervene (see www.raggedclaws.com/criticalrealism).

Critical Realism in IS Research

Sven A. Carlsson
Lund University, Sweden

INTRODUCTION AND BACKGROUND

The Information Systems (IS) field is dominated by research approaches and theories based in positivism (Arnott, Pervan, O'Donnell, & Dodson, 2004; Chen & Hirschheim, 2004; Schultze & Leidner, 2003). IS scholars have pointed out weaknesses in these approaches and theories and in response different strands of post-modern theories and constructivism have gained popularity—see Lee, Liebenau, and DeGross, (1997), Trauth (2001), Whitman and Woszczynski (2004), and Michael Myers' "Qualitative Research in Information Systems" (http://www.qual.auckland.ac.nz). The approaches and theories argued for include interpretivism, ethnography, grounded theory, and theories like Giddens' (1984) structuration theory and Latour's (1987) actor-network theory. (For simplicity, we refer to these different approaches and theories as "post-approaches" and "post-theories" when distinction is not required).

Although these approaches and theories overcome some of the problems noted with "traditional" approaches and theories, they have at least three major weaknesses and limitations. First, their fascination with the voices of those studied leads to IS research as mere reportages and local narratives. Second, their focus on agency leads to them ignoring the structural (systemic) dimension—the agency/structure dimension is collapsed, leading to a flat treatment of the dimension. Third, their rejection of objectivist elements leads to problems when researching artifacts like IT-based IS. For elaborate critiques of post-approaches and post-theories, see Lòpez and Potter (2001) and Archer, Bhaskar, Collier, Lawson, and Norrie (1998).

An alternative to traditional positivistic models of social science as well as an alternative to post-approaches and post-theories is critical realism (Bhaskar, 1978, 1989, 1998; Harré & Secord, 1972). Critical realism (CR) argues that social reality is not simply composed of agents' meanings, but that there exist structural factors influencing agents' lived experiences. CR starts from an ontology that identifies structures and mechanisms through which events and discourses are generated as being fundamental to the constitution of our natural and social reality. This article briefly presents critical realism and exemplifies how it can be used in IS research.

CRITICAL REALISM

Critical realism has primarily been developed by Roy Bhaskar and can be seen as a specific form of realism. Unfortunately, Bhaskar is an opaque writer, but clear summaries of CR are available in Sayer (2000) and Archer et al. (1998). CR's manifesto is to recognize the reality of the natural order and the events and discourses of the social world. It holds that:

… we will only be able to understand—and so change—the social world if we identify the structures at work that generate those events or discourses… These structures are not spontaneously apparent in the observable pattern of events; they can only be identified through the practical and theoretical work of the social sciences. (Bhaskar, 1989, p. 2)

Bhaskar (1978) outlines what he calls three domains: the *real*, the *actual*, and the *empirical*. The real domain consists of underlying structures and mechanisms, and relations; events and behavior; and experiences. The generative mechanisms, residing in the real domain, exist independently of but capable of producing patterns of events. Relations generate behaviors in the social world. The domain of the actual consists of these events and behaviors. Hence, the actual domain is the domain in which observed events or observed patterns of events occur. The domain of the empirical consists of what we experience, hence, it is the domain of experienced events. Bhaskar (1978, p. 13) argues that:

…real structures exist independently of and are often out of phase with the actual patterns of events. Indeed it is only because of the latter we need to perform experiments and only because of the former that we can make sense of our performances of them. Similarly it can be shown to be a condition of the intelligibility of perception that events occur independently of experiences. And experiences are often (epistemically speaking) 'out of phase' with events—e.g., when they are misidentified. It is partly because of this possibility that the scientist needs a scientific education or training. Thus I [Bhaskar] will argue that what I call the domains of the real, the actual and the empirical are distinct.

CR also argues that the real world is ontologically stratified and differentiated. The real world consists of a plurality of structures and mechanisms that generate the events that occur.

USING CRITICAL REALISM IN IS RESEARCH

CR has primarily been occupied with philosophical issues and fairly abstract discussions. In recent years, attention has been paid to how to actually carry out research with CR as a philosophical underpinning—see Layder (1998), Robson (2002), and Kazi (2003). This section briefly presents how CR can be used in IS research by discussing how CR can be used in theory development and how CR can be used in IS evaluation research.

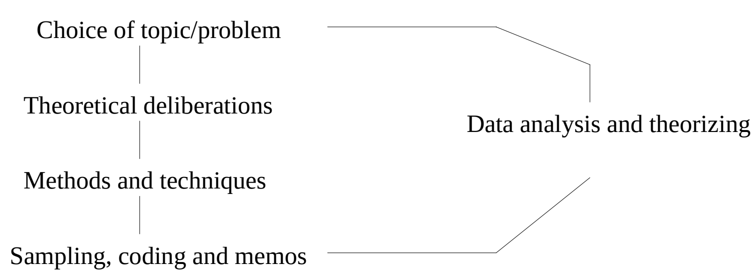
Bhaskar (1998) says that explanations (theories) are accomplished by the RRRE model of explanation comprising a four-phase process: (1) *Resolution* of a complex event into its components (causal analysis); (2) *Redescription* of component causes; (3) *Retrodiction* to possible (antecedent) causes of components via independently validated normic statements; and (4) *Elimination* of alternative possible causes of components. This is a rather abstract description of explanation development, and here we will instead use Layder's (1998) less abstract "adaptive theory." It is an approach for generating theory in conjunction with empirical research. It attempts to combine the use of pre-existing theory and theory generated from empirical data. Figure 1 depicts the different elements of the research process. There is not some necessary or fixed temporal sequence. Layder stresses that theorizing should be a continuous process accompanying the research at all stages. Concerning research methods and research design, CR is supportive of: (1) the use of both quantitative and qualitative methods, (2) the use of extensive and intensive research design, and (3) the use of fixed and flexible research design.

To exemplify how CR and Layder's adaptive theory can be used in IS research, we will use a project on the use of Executive Information Systems (EIS). The project was done together with Dorothy Leidner.1 Here a new discussion of the research is carried out. The overall purpose of the study was to increase our understanding of the development and use of EIS, i.e., develop EIS theory.

Layder's adaptive theory approach has eight overall parameters. One parameter says that adaptive theory "uses both inductive and deductive procedures for developing and elaborating theory" (Layder, 1998). The adaptive theory suggests the use of both forms of theory generation within the same frame of reference and particularly within the same research project. We generated a number of hypotheses (a deductive procedure), based on previous EIS studies and theories as well as Huber's (1990) propositions on the effects of advanced IT on organizational design, intelligence, and decision making. These were empirically tested. From a CR perspective, the purpose of this was to find patterns in the data that would be addressed in the intensive part of the study. [For a discussion of the use of statistics in CR studies, see Mingers (2003).] We also used an inductive procedure. Although, previous theories as well as the results from the extensive part of the project were fed into the intensive part, we primarily used an inductive approach to generate tentative explanations of EIS development and use from the data. The central mode of inference (explanation) in CR research is retroduction. It enables a researcher, using induction and deduction, to investigate the potential causal mechanisms and the conditions under which certain outcomes will or will not be realized. The inductive and deductive procedures led us to formulate explanations in terms of what mechanisms and contexts would lead (or not lead) to certain outcomes—outcomes being types of EIS use with their specific effects.

Another parameter says that adaptive theory "embraces both objectivism and subjectivism in terms of its ontological presuppositions" (Layder, 1998). The adaptive theory conceives the social world as including both subjective and objective aspects and mixtures of the two. In our study, one objective aspect was the IT used in the different EIS and one subjective aspect was perceived effects of EIS use.

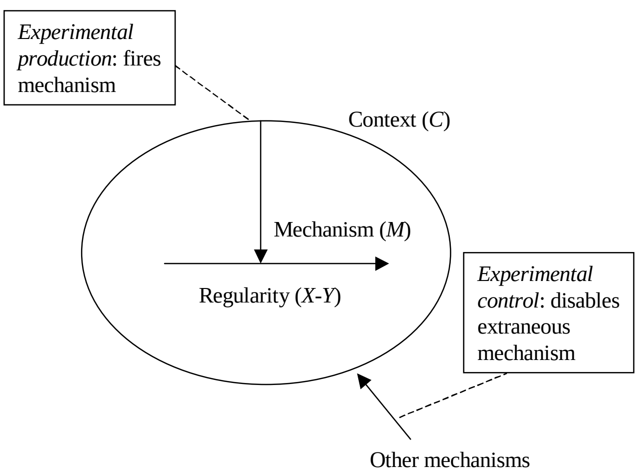
Figure 1. Elements of the research process (Layder, 1998)



Two other parameters say that adaptive theory "assumes that the social world is complex, multi-faceted (layered) and densely compacted" and "focuses on the multifarious interconnections between human agency, social activities, and social organization (structures and systems)" (Layder, 1998). In our study, we focused the "interconnections" between agency and structure. We addressed *self* (e.g., perceptions of EIS), *situated activity* (e.g., use of EIS in day-to-day work), *setting* (e.g., organizational structure and culture), and *context*, (e.g., national culture and economic situation). Based on our data, we can hypothesize that national culture can in certain contexts affect (generate) how EIS are developed and used and how they are perceived. We can also hypothesize that organizational "strategy" and "structure" as well as "economic situation" can in certain contexts affect (generate) how EIS are developed and used and how they are perceived.

Our study and the results (theory) were influenced by, e.g., Huber's propositions, the "theory" saying that EIS are systems for providing top managers with critical information, and Quinn's competing values approach (Quinn, Faerman, Thompson, & McGrath, 1996). The latter theory was brought in to theorize around the data from the intensive (inductive) part of the study. Adaptive theorizing was ever present in the research process. In line with CR, we tried to go beneath the empirical to explain *why* we found what we found through hypothesizing the mechanisms that shape the actual and the events. Our study led us to argue that it is a misconception to think of EIS as systems that just provide top managers with information. EIS are systems that support managerial cognition and behavior—providing information is only one of several means, and it can be one important means in organizational change. Based on our study, we "hypothesize" that "tentative" mechanisms are, for example, national culture, economic development, and organizational strategy and culture. We also hypothesized how the mechanisms together with different actors lead to the development and use of different types of EIS, for example, EIS for personal productivity enhancement respectively EIS for organizational change.

Figure 2. Realistic experiment (Pawson & Tilley, 1997)



IS evaluation (research) is of critical concern to practitioners and academics. Pawson and Tilley (1997) and Kazi (2003) advocate evaluation approaches that draw on the principle of CR and which, according to Bryman (2001, p. 40) see:

> *...the outcome of an intervention [like the implementation of an EIS] as the result of generative mechanisms and the contexts of those mechanisms. A focus of the former element entails examining the causal factors that inhibit or promote change when an intervention occurs.*

Driving realistic IS evaluation research is the goal to produce ever more detailed answers to the question of *why* an IS initiative—e.g., an EIS implementation—works, for *whom*, and in *what* circumstances. This means that evaluation researchers attend to how and why an IS initiative has the potential to cause (desired) changes. Realistic IS evaluation research is applied research, but theory is essential in every aspects of IS evaluation research. The goal is not to develop theory per se, but to develop theories for practitioners, stakeholders, and participants.

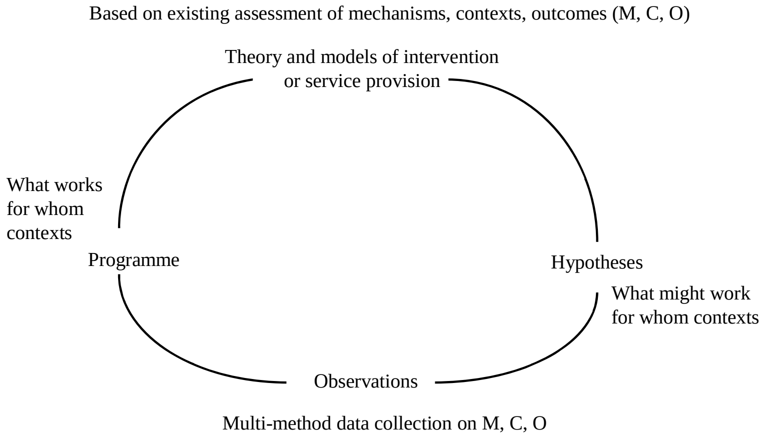
A realistic evaluation researcher works as an experimental scientist, but not according to the logics of the traditional experimental research. Said Bhaskar (1998, p. 53):

> *The experimental scientist must perform two essential functions in an experiment. First, he must trigger the mechanism under study to ensure that it is active; and secondly, he must prevent any interference with the operation of the mechanism. These activities could be designated as 'experimental production' and 'experimental control.'*

Figure 2 depicts the realistic experiment.

Realistic evaluation researchers do not conceive that IS initiatives "work." It is the action of stakeholders that makes them work, and the causal potential of an IS initiative takes the form of providing reasons and resources to enable different stakeholders and participants to "make" changes. This means that a realistic evaluation researcher seeks to understand *why* an IS initiative works through an understanding of the action mechanisms. It also means that a realistic evaluation researcher seeks to understand *for whom* and *in what circumstances (contexts)* an IS initiative works through

Figure 3. Realistic effectiveness cycle (Kazi, 2003; Pawson & Tilley, 1997)



the study of contextual conditioning. Realistic evaluation researchers orient their thinking to context-mechanism-outcome (CMO) configurations. A CMO configuration is a proposition stating what it is about an IS initiative (IS implementation) that works, for whom, and in what circumstances. A refined CMO configuration is the finding of IS evaluation research—the output of a realistic evaluation study.

Realistic IS evaluation based on the above may be implemented through a realistic effectiveness cycle (Kazi, 2003; Pawson & Tilley, 1997) (see Figure 3). The starting point is theory. Theory includes proposition on how the mechanisms introduced by an IS invention into pre-existing contexts can generate desired outcomes. This entails theoretical analysis of mechanisms, contexts, and expected outcomes. This can be done using a logic of analogy and metaphor. The second step consists of generating "hypotheses." Typically, the following questions would be addressed in the hypotheses: (1) what changes or outcomes will be brought about by an IS intervention?; (2) what contexts impinge on this?; and (3) what mechanisms (social, cultural and others) would enable these changes, and which one may disable the intervention? The third step is the selection of appropriate data collection methods—realists are committed to methodological pluralism. In this step it might be possible to provide evidence of the IS intervention's ability to change reality. Based on the result from the third step, we may return to the programme (the IS intervention) to make it more specific as an intervention of practice. Next, but not finally, we return to theory. The theory may be developed, the hypotheses refined, the data collection methods enhanced, etc. This leads to the development of transferable and cumulative lessons from IS evaluation research.

CONCLUSION AND FURTHER RESEARCH

This article argued that CR can be used in IS research. CR overcomes problems associated with positivism, constructivism, and postmodernism.

Although CR has influenced a number of social science fields [2] e.g., organization and management studies (Ackroyd & Fleetwood, 2000; Reed, 2003; Tsang & Kwan, 1999; Tsoukas, 1989) and social research approaches and methods (Byrne, 1998; Robson, 2002), it is almost invisible in the IS field. CR's potential for IS research has been argued by, for example, Carlsson (2003a, 2003b, 2004), Dobson (2001a, 2001b), Mingers (2003, 2004a, b), and Mutch (2002). They argue for the use of CR in IS research and discuss how CR can be used in IS research.

Bhaskar has further elaborated on CR, and in the future, his more recent works, e.g., Bhaskar (2002), could be explored in IS research.

REFERENCES

Ackroyd, S. & Fleetwood, S. (Eds.)(2000a). *Realist perspectives on management and organisations*. London: Routledge.

Archer, M., Bhaskar, R., Collier, A., Lawson, T., & Norrie, A. (Eds.)(1998). *Critical realism: Essential readings*. London: Routledge.

Arnott, D., Pervan, G., O'Donnell, P., & Dodson, G. (2004). An analysis of decision support systems research: Preliminary results. In R. Meredith, G. Shanks, D. Arnott & S. Carlsson (Eds.), *Decision support in an uncertain and* *complex world* (pp. 25-38). Melbourne, AU: Monash University.

Bhaskar, R. (1978). *A realist theory of science*. Sussex, UK: Harvester Press.

Bhaskar, R. (1989). *Reclaiming reality*. London: Verso.

Bhaskar, R. (1998). *The possibility of naturalism* (3rd ed.). London: Routledge.

Bhaskar, R. (2002). *Reflections on meta-reality: Transcendence, enlightenment and everyday life*. London: Sage.

Bryman, A. (2001). *Social research methods*. Oxford, UK: Oxford University Press.

Byrne, D. (1998). *Complexity theory and the social sciences*. London: Routledge.

Carlsson, S.A. (2003a). Advancing information systems evaluation (research): A critical realist approach. *Electronic Journal of Information Systems Evaluation*, 6(2), 11-20.

Carlsson, S.A. (2003b). Critical realism: A way forward in IS research. In *Proceedings of the 11th European Conference on Information Systems*.

Carlsson, S.A. (2004). Using critical realism in IS research. In M.E. Whitman & A.B. Woszczynski (Eds.), *The handbook of information systems research* (pp. 323-338). Hershey, PA: Idea Group Publishing.

Carlsson, S.A., Leidner, D.E., & Elam, J.J. (1996). Individual and organizational effectiveness: Perspectives on the impact of ESS in multinational organizations. In P. Humphreys, L. Bannon, A. McCosh, P. Migliarese, & J.C. Pomerol (Eds.), *Implementing systems for supporting management decisions: Concepts, methods and experiences* (pp. 91-107). London: Chapman & Hall.

Chen, W. & Hirschheim, R. (2004). A paradigmatic and methodological examination of information systems research. *Information Systems Journal*, 14(3), 197-235.

Dobson, P.J. (2001a). The philosophy of critical realism—An opportunity for information systems research. *Information Systems Frontier*, 3(2), 199-201.

Dobson, P.J. (2001b). Longitudinal case research: A critical realist perspective. *Systemic Practice and Action Research*, 14(3), 283-296.

Giddens, A. (1984). *The constitution of society*. Cambridge, UK: Polity Press.

Harré, R., & Secord, P. (1972). *The explanation of social behavior*. Oxford, UK: Blackwell.

Huber, G.P., (1990). A theory of the effects of advanced information technologies on organizational design, intelligence, and decision making. *Academy of Management Review*, 15(1), 47-71.

Kazi, M.A.F. (2003). *Realist evaluation in practice*. London: Sage.

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Hemel Hampstead: Harvester Wheatsheaf.

Layder, D. (1998). *Sociological practice: Linking theory and social research*. London: Sage.

Lee, A.S., Liebenau, J., & DeGross, J. (Eds.)(1997). *Information systems and qualitative research*. London: Chapman & Hall.

Leidner, D.E. & Carlsson, S.A. (1998). Les bénéfices des systèmes d´information pour dirigeants dans trois pays. *Systèmes d´Information et Management*, 3(3), 5-27.

Leidner, D.E. & Elam, J.J. (1995). The impact of executive information systems on organizational design, intelligence, and decision making. *Organization Science,* 6(6), 645-665.

Leidner, D.E., Carlsson, S.A., Elam, J.J., & Corrales, M. (1999). Mexican and Swedish managers´ perceptions of the impact of EIS on organizational intelligence, decision making, and structure. *Decision Sciences*, 30(3), 633-658.

Lòpez, J. & Potter, G. (Eds.)(2001). *After postmodernism: An introduction to critical realism*. London: Athlone.

Mark, M.M., Henry, G.T., & Julnes, G. (2000). *Evaluation*. San Francisco, CA: Jossey-Bass.

Mingers, J. (2002). Real-izing information systems: Critical realism as an underpinning philosophy for information systems. In *Proceedings of the 23rd International Conference on Information Systems* (pp. 295-303).

Mingers, J. (2004b). Re-establishing the real: Critical realism and information systems. In J. Mingers & L. Willcocks (eds.), *Social theory and philosophy for information systems* (pp. 372-406). Chichester, UK: Wiley.

Mingers, J. (2004a). A critique of statistical modelling from a critical realist perspective. In *Proceedings of the 11th European Conference on Information Systems*.

Mutch, A. (2002). Actors and networks or agents and structures: Towards a realist view of information systems. *Organizations*, 9(3), 477-496.

Pawson, R. & Tilley, N. (1997). *Realistic evaluation*. London: Sage.

Quinn, R.E., Faerman, S.R., Thompson, M.P., & McGrath, M.R. (1996). *Becoming a master manager* (2nd ed.). New York: John Wiley & Sons.

Reed, M.I. (2003). The agency/structure dilemma in organization theory. In H. Tsoukas & C. Knudsen (Eds.), *The Oxford handbook of organization theory* (pp. 289-309). Oxford, UK: Oxford University Press.

Robson, C. (2002). *Real world research* (2nd ed.). Oxford, UK: Blackwell.

Sayer, A. (2000). *Realism and social science*. London: Sage.

Schultze, U. & Leidner, D.E. (2003). Studying knowledge management in information systems research: Discourses and theoretical assumptions. *MIS Quarterly*, 26(3), 213-242.

Trauth, E.M. (ed.)(2001). *Qualitative research in IS: Issues and trends*. Hershey, PA: Idea Group Publishing.

Tsang, E.W. & Kwan, K.-M. (1999). Replication and theory development in organizational science: A critical realist perspective. *Academy of Management Review*, 24(4), 759-780.

Tsoukas, H. (1989). The validity of idiographic research explanations. *Academy of Management Review*, 14(4), 551-561.

Whitman, M.E. & Woszczynski, A.B. (Eds.)(2004). *The handbook of information systems research*. Hershey, PA: Idea Group Publishing.

KEY TERMS

Constructivism (or Social Constructivism): Asserts that (social) actors socially construct reality.

Context-Mechanism-Outcome Pattern: Realist evaluation researchers orient their thinking to context-mechanism-outcome (CMO) pattern configurations. A CMO configuration is a proposition stating what it is about an IS initiative that works, for whom, and in what circumstances. A refined CMO configuration is the finding of IS evaluation research.

Critical Realism: Asserts that the study of the social world should be concerned with the identification of the structures and mechanisms through which events and discourses are generated.

Empiricism: Asserts that only knowledge gained through experience and senses is acceptable in studies of reality.

Positivism: Asserts that reality is the sum of sense impression—in large, equating social sciences with natural sciences. Primarily uses deductive logic and quantitative research methods.

Postmodernism: A position critical of realism that rejects the view of social sciences as a search for overarching explanations of the social world. Has a preference for qualitative methods.

Realism: A position acknowledging a reality independent of actors' (including researchers') thoughts and beliefs.

Realist IS Evaluation: Evaluation (research) based on critical realism aiming at producing ever more detailed answers to the question of *why* an IS initiative works (better), for *whom,* and in *what* circumstances (contexts).

Retroduction: The central mode of inference (explanation) in critical realism research. Enables a researcher to investigate the potential causal mechanisms and the conditions under which certain outcomes will or will not be realized.

ENDNOTES

[1] See Carlsson, Leidner, & Elam (1996), Leidner & Carlsson (1998), Leidner & Elam (1995), and Leidner, Carlsson, Elam, & Corrales (1999).

[2] For examples, see *Journal of Critical Realism*.

Critical Strategies for IS Projects

Dane Peterson
Southwest Missouri State University, USA

Chung S. Kim
Southwest Missouri State University, USA

INTRODUCTION

Despite enormous progress in the methodologies and technologies used in the development and implementation of information systems (IS), the failure to achieve IS success on a consistent basis continues to plague the efforts of many companies (Clegg et al., 1997). The success or failure of an IS has been defined in a variety of ways depending on a number of factors, such as the nature of the system and the perspective of the various stakeholders (i.e., IS developers, users, and management). In general, IS success is often subjectively assessed on the basis of how well the system achieves strategic, informational, and transactional benefits for the organization (Mirani & Lederer, 1998).

In order to increase the likelihood of IS success, a variety of approaches and recommendations have been offered. One of the earliest and most popular proposals for increasing IS success was based on the concept of critical success factors (CSF). This proposal assumes that IS failure can be significantly reduced by identifying a small set of factors that are essential for the achievement of IS success (Rockart, 1979). Use of the CSFs approach may make the IS planning process more efficient by providing managers with a decision aid for determining the key areas that are likely to need the most attention and how scarce resources should be allocated. Identification of CSF may also enhance communication about potential problems that may arise due to differences in the perspectives among various stakeholders. Another potential advantage of the CSF approach is that it facilitates the learning process. That is, the CSF approach increases stakeholders' understanding of the IS development process and the ultimate objectives of the system and how the overall development process can be evaluated (Pollalis & Frieze, 1993).

BACKGROUND

Some researchers have attempted to identify the critical factors based on an examination of which factors are statistically related to IS success. In this approach, user satisfaction is often used as measure of IS success. One of the first examples of a study based on this approach used a discriminant analysis procedure to examine which factors best distinguished between successful and unsuccessful projects (Ginzberg, 1981). The results indicated that organizational commitment, commitment to change, and extent of project definition and planning were the best predictors of user satisfaction.

Perhaps because some researchers are not convinced that user satisfaction provides a sufficient surrogate measure of the overall success and benefits of an IS (see Mirani & Lederer, 1998), many investigators have attempted to identify the CSFs based on the views of IS experts. In these studies, IS experts directly assessed the relative importance of potential success factors (e.g., Burn & Szeto, 2000; Jiang, Klein & Balloun, 1996). In general, the results of studies examining the views of IS experts have demonstrated some agreement with respect to the CSFs, although some differences have been observed among the various studies.

Much of the early research on the CSF approach assumed that once the key factors were identified, the same critical factors might apply to the development of all IS. However, as the nature and types of IS have become increasing diverse over the past two decades, much of the recent research has adapted the CSF approach to identifying the key elements that apply to various types of IS and to new areas of IS applications. For example, studies have attempted to identify the CSFs in areas involving executive information systems (Chen & Lee, 2003; Poon & Wagner, 2001), object-oriented analysis and design (Pei & Cutone, 1995), computer-assisted software engineering (Summer & Ryan, 1994), geographical information systems (Nasirin & Birks, 2003), data warehousing (Mukherjee & D'Souza, 2003), emerging information systems in the public sector (Bajjaly, 1999), implementation of integrated services digital networks (Lai & Clark, 1998), enterprise resource planning systems (Akkermans & van Helden, 2002; Al-Mashari, Al-Mudimigh & Zairi, 2003; Umble, Haft & Umble, 2003), information systems requirements gathering (Havelka & Sooun, 2002), the IT alignment planning process (Peak & Guynes, 2003), strategic planning in Eastern cultures (Ang & Teo, 1997), managing decentralized information technology (Birge, 1997), performance of information centers (Guimaraes, Gupta & Rather, 1999), outsourcing implementation (Kim & Young-Soo, 2003; Soliman, Chen & Frolick, 2003), extranet adoption in e-supply chain (Chow, 2004), and facilitating participation of many people in the IS planning process (Peffers, Gengler & Tunnanen, 2003).

Another perspective for remedying the problem of low IS success rates is to identify the causes of IS failure (Williams & Ramaprasad, 1996). According to this viewpoint, there may be inhibiting factors that play a key role in causing IS failure, and therefore it is important to identify both CSFs and critical failure factors (CFFs). Otherwise, a disproportionate amount of attention may be devoted to enhancing factors only to discover that there are inhibiting factors that prevent IS success.

Several studies have surveyed IS experts in attempts to identify the key factors contributing to unsuccessful or abandoned IS projects (e.g., Jiang, Klein & Balloun, 1998; Lyytinen, 1988). For example, Ewusi-Mensah (1997) examined abandoned IS projects and found that poorly defined project goals, lack of project team experience and expertise, inappropriate technology, lack of top management involvement, and escalating project costs were among the reasons for IS failure.

Many of the factors identified as CFFs are the same factors identified as CSFs. Thus, a CFF may simply be the absence of the CSF. However, there appears to be enough differences in the results of studies on CSFs and CFFs to warrant an investigation on the possibility of both enhancing and inhibiting factors (e.g., Lyytinen, 1987).

PURPOSE OF STUDY

Although numerous studies have investigated potential CSFs or CFFs, apparently there has not been any attempt to examine both types of factors in a single study. One objective of this study was to examine if there are any differences in the relative importance of the factors believed to contribute to the success and failure of IS projects.

Consistent with the notion that CSFs may vary depending on the nature and type of IS, a second goal of this study was to examine potential cultural differences in IS developers' views on CSF. A limitation of prior research attempting to identify the CSFs and CFFs is that almost all of the research has focused on the views of IS developers in Western cultures. With an increasing number of corporations developing and implementing IS applications that cross national boundaries and span diverse cultures, there is a need to determine if the factors viewed as most important by IS developers in the Western culture are the same factors viewed as most important by IS developers in other cultures. It may be particularly important to investigate the views of IS developers in Eastern Asia. Several countries in Eastern Asia play a significant role in the area of global information technology (McIntosh, 1999). In addition, the vast differences between Western and Eastern cultures have created a number of obstacles to the successful development of global IS when Asian offices are involved (Burnson, 1989). Thus, this study focused on an investigation of the views of IS developers from Korea.

The present study addressed the following two goals.

1. Which factors do IS developers in Korea view as most important for contributing to IS success and to IS failure?
2. How similar are the views of IS developers in Korea regarding CSFs and CFFs to the results reported in previous studies involving IS developers from Western cultures?

RESEARCH METHODOLOGY

Critical Factors: Based on prior studies examining CSFs and CFFs, this study identified 18 potential factors. These factors are controllable and applicable to IS developers in international environments. The 18 factors are listed in Table 1. The failure factors are expressed as the absence or insufficient condition of the success factors.

Respondents: A systematic random sampling procedure was used to distribute the surveys to IS developer in 10 Korean organizations. The number of IS developers surveyed in each organization ranged from 4 to 29, resulting in a total of 127 useable surveys. Most of the respondents had a least a college degree (83.5%). The average age of the respondents was 32.24 with a standard deviation of 4.42 years. The average number of years of experience in the field was 6.25.

RESULTS AND CONCLUSIONS

The sample means and rank order of importance for the CSFs and CFFs are presented in Table 1. As illustrated in Table 1, user participation, clearly stated objectives, and top management support were viewed as highly critical with regard to both IS success and IS failure. In general, the factors viewed as least important by IS developers from Korea were technical factors and/or tactic-related operational factors (e.g., methodology, prototyping, etc.). Apparently IS developers view organizational factors as more crucial for IS success, possibly because they have

Table 1. Means, ranks, and t-test statistics for CSF's and CFF's

Critical Factors	CSF Mean	CSF Rank	CFF Mean	CFF Rank	t-statistic
(Insufficient) User participation	6.07	1	5.89	1	-2.59**
(Lack of) Clearly stated objectives	6.06	2	5.38	4	1.10
(Lack of) Team member commitment	5.95	3	5.02	11	3.08**
(Lack of) Top management support	5.91	4	5.57	3	-0.08
(Lack of) Project Leader's monitoring/control	5.84	5	5.17	9	0.78
Project leader's (in)experience	5.80	6	5.63	2	1.58
(Ill)Alignment of project and corporate goals	5.71	7	5.05	10	1.12
Use of (in)appropriate technology	5.69	8	4.98	12	0.56
(Lack of) Detailed project plan	5.65	9	5.19	6	-1.44
(No attempt to) Reengineer business process	5.65	9	5.18	7	-0.88
(Im)Proper project scope	5.54	11	5.20	5	-3.17**
(In)Adequate training for team	5.47	12	5.18	7	0.55
(Lack of) Project leader's feedback	5.45	13	4.85	13	1.97*
(Lack of) Team members' self-control	5.41	14	4.67	14	-0.37
Utilizing an (in)effective methodology	5.41	14	4.59	16	0.27
Team member (in)experience	5.36	16	4.65	15	0.62
(Little use of)Utilizing a prototype	5.11	17	4.44	18	-0.07
(Insufficient) Peer review on project progress	5.11	17	4.56	17	-0.05

less control over these factors compared to technical/operational factors.

Generally, the importance ratings for the CSFs were very similar to those for the CFFs. This conclusion is confirmed by the high positive correlation between the means for the CSFs and CFFs ($r = .841$). Since the success factors were rated higher in importance for success than the failure factors were for contributing to the lack of success (e.g., user participation was 6.07 for success and 5.89 for contributing to failure), the responses from each IS developer were standardized in order to compare the relative importance of the 18 success and failure factors. The last column of Table 1 shows the results of related samples t-tests based on the standardized scores. Only 3 of the 18 factors were significant. The two factors viewed as more likely to contribute to IS failure, user participation and project scope, are factors that are generally outside the control of IS developers. The factor viewed as more likely to contribute to IS success, team member commitment, is generally within the domain of the IS developers.

Briefly, the perceptions of critical factors by IS developers from Korea are generally similar to those observed in studies involving IS developers in Western cultures (Ewusi-Mensah, 1997; Ginzberg, 1981; Jiang et al., 1996; Lyytinen, 1988; Pinto & Prescott, 1990). However, Korean IS developers viewed characteristics of the project leaders (monitoring and experience) as more important than what is generally observed in studies involving IS developers in Western cultures. This may be explained in terms of the high power distance culture of Korea. In a high power distance culture, managers are expected to make most of the decisions and take responsibility for project success. Thus, it might seem reasonable to expect that IS developers in Korea would consider project leaders as more important to IS success.

In summary, factors within the domain of the IS developers appear to be more likely to be viewed as CSFs, while factors outside the domain of IS developers are more likely to be viewed as CFFs. However for the most part, the results of this study demonstrate a strong relationship between perceptions of CSFs and CFFs. The views of developers from Korea were generally consistent with results based on surveys of developers from Western cultures, except that Korean IS developers view project leaders as more crucial.

FUTURE TRENDS

Since the introduction of the concept of CSFs, ISs have become increasing diverse. It seems likely that the most crucial factors could vary depending on a number of characteristics of the IS (operational vs. strategic IS (Burn & Szeto, 2000), cultural differences among the IS developers (Kim & Peterson, 2002), and IS lifecycle stage (Pinto & Prescott, 1990)). Thus the list of potential CSFs identified in past research may best serve as a list of potential risk areas that may be encountered in the development of a specific IS (Schmidt et al., 2001). Future research may examine the best methods that IS developers can use to assess the potential risks of each critical factor and procedures to overcome the risks in the development of a specific type of IS or IS application.

REFERENCES

Akkermans, H., & van Helden, K. (2002). Vicious and virtuous cycles in ERP implementation: A case study of interrelations between critical success factors. *European Journal of Information Systems, 11*, 35-47.

Al-Mashari, M., Al-Mudimigh, A., & Zairi, M. (2003). Enterprise resource planning: A taxonomy of critical factors. *European Journal of Operational Research, 146,* 352-365.

Ang, J. & Teo, T.S.H. (1997). CSFs and sources of assistance and expertise in strategic IS planning: A Singapore perspective. *European Journal of Information Systems, 6,* 164-172.

Bajjaly, S.T. (1999). Managing emerging information systems in the public sector. *Public Performance & Management Review, 23,* 40-48.

Birge, E.M. (1997). Managing the decentralized IT organization. *Information Systems Management, 14,* 77-82.

Burn, J.M., & Szeto, C. (2000). A comparison of the views of business and IT management on success factors for strategic alignment. *Information & Management, 37,* 197-216.

Burnson, A. (1989, August 14). The perils of going global. *Infoworld,* 39-40.

Chen, J.Q., & Lee, S.M. (2003). An exploratory cognitive DSS for strategic decision making. *Decision Support Systems, 36,* 147-161.

Chow, W.S. (2004). An exploratory study of the success factors for extranet adoption in e-supply chain. *Journal of Global Information Management, 12,* 60-77.

Clegg, C., Axtell, C., Damadoran, L., Farbey, B., Hull, R., Lloyd-Jones, R., Nicholls, J., Seell, R., & Tomlinson, C. (1998). Information technology: A study of performance and the role of human and organizational factors. *Ergonomics Journal, 40,* 851-871.

Ewusi-Mensah, K. (1997). Critical issues in abandon information systems development. *Communications of the ACM, 40,* 74-80.

Ginzberg, M.J. (1981). Key recurrent issues in the MIS implementation process. *MIS Quarterly, 5,* 47-59.

Guimaraes, T., Gupta, Y.P., & Rather, R.K. (1999). Empirically testing the relationship between end-user computing problems and information center success factors. *Decision Sciences, 30,* 393-414.

Havelka, D., & Sooun, L. (2002). Critical success factors for information requirements gathering. *European Journal of Information Systems, 11*, 35-47.

Jiang, J.J., Klein, G., & Balloun, J. (1996). Ranking of system implementation success factors. *Project Management Journal*, *27*, 49-53.

Jiang, J.J., Klein, G., & Balloun, J. (1998). Perceptions of system development failures. *Information and Software Technology, 39,* 933-937.

Kim, C.S., & Peterson, D.K. (2002). Cultural differences in developers' perceptions of information systems success factors: Japan vs the United States. *Journal of Global Information Management, 10,* 5-13.

Kim, S., & Young-Soo, C. (2003). Critical success factors for IS outsourcing implementation from an interorganizational relationship perspective. *Journal of Computer Information Systems, 43,* 81-91.

Lai, V.S., & Clark, J.G. (1998). Network evolution towards ISDN services: A management perspective. *Journal of Information Technology, 13,* 67-79.

Lyytinen, K. (1988). Expectation failure concept and systems analysts' view of information system failures: Results of an exploratory study. *Information & Management, 14,* 45-56.

McIntosh, J.C. (1999). Strategic uses of global information technology among manufacturing firms. *Journal of Global Information Technology Management, 3,* 24-41.

Mirani, R., & Lederer, A.L. (1998). An instrument for assessing the organizational benefits of IS Projects. *Decision Sciences, 29,* 803-837.

Mukherjee, D., & D'Souza, D. (2003). Think phased implementation for successful data warehousing. *Information Systems Management, 20,* 82-91.

Nasirin, S., & Birks, D. (2003). DSS implementation in the UK retail organizations: A GIS perspective. *Information & Management, 40,* 325-337.

Peak, D., & Guynes, C.S. (2003). The IT alignment planning process. *Journal of Computer Information Systems, 44,* 9-16.

Peffers, K., Gengler, C.E., & Tuunanen, T. (2003). Extending critical success factors methodology to facilitate broadly participative information systems planning. *Journal of Management Information Systems, 20,* 51-87.

Pei, D., & Cutone, C. (1995). Object-oriented analysis and design. *Information Systems Management, 12,* 54-61.

Pinto, J., & Prescott, J.E. (1990). Planning and tactical factors in the project implementation process. *Journal of Management Studies, 27,* 305-327.

Pollalis, Y.A., & Frieze, I.H. (1993). A new look at critical success factors. *Information Strategy, 10,* 24-35.

Poon, P., & Wagner, C. (2001). Critical success factors revisited: Success and failure cases of information systems for senior executives. *Decision Support Systems, 30,* 393-419.

Rockart, J.F. (1979). Chief executives define their needs. *Harvard Business Review, 57,* 81-93.

Schmidt, R., Lyytinen, K., Keil, M., & Cule, P. (2001). Identifying software project risks: An international Delphi study. *Journal of Management Information Systems, 17,* 5-36.

Soliman, K.S., Chen, L., & Frolick, M.N. (2003). APS: Do they work? *Information Systems Management, 20,* 50-59.

Summer, M., & Ryan, T. (1994). The impact of CASE: Can it achieve critical success factors? *Journal of Systems Management, 45,* 16-22.

Umble, E.J., Haft, R.R., & Umble, M.M. (2003). Enterprise resource planning: Implementation procedures and critical success factors. *European Journal of Operational Research, 146,* 241-258.

Williams, J.J., & Ramaprasad, A. (1996). A taxonomy of critical success factor. *European Journal of Information Systems, 5,* 250-260.

KEY TERMS

Business Process Reengineering: Redesign of business processes with the purpose of a dramatic improvement in business performances and productivity.

Critical Failure Factors: CFFs are the limited number of areas which, without careful monitoring of these areas, may lead to a failure of a system.

Critical Success Factors: CSFs are the limited number of areas in which results, if they are satisfactory, will ensure successful competitive performance for the organization.

Detailed Project Plan: A plan(s) that specifies detailed schedules, milestones, humanpower, and equipment requirements necessary for complementing a project.

Peer Review on Project Progress: A structured walk-through is a widely used technique to provide a test of a proposed system design and is often implemented as a peer feedback and review process.

Proper Project Scope: The scope of the project defines the boundary of the project or which aspects of the system will be included in the project.

Prototype: A simplified and smaller model system that still demonstrates the essence of the system.

Critical Success Factors for Distance Education Programs

Ben Martz
University of Colorado at Colorado Springs, USA

Venkat Reddy
University of Colorado at Colorado Springs, USA

INTRODUCTION

Distance education is playing an ever-growing role in the education industry. As such, it is prudent to explore and understand driving conditions that underlie this growth. Understanding these drivers and their corresponding concerns (Table 1) can help educators in the distance education field better prepare for the industry.

BACKGROUND

Distance education's primary driver is that it is the major growth segment in the education industry. In 1999, nearly 80% of the public, four-year institutions and over 60% of the public, two-year institutions offered distance education courses. Over 1.6 million students are enrolled in distance courses today. Over 90% of all colleges are expected to offer some online courses by 2004 (Institute of Higher Education Policy, 2000). Corporations envision online training warehouses saving large amounts of training dollars. Combined, the virtual education market and its sister market, corporate learning, are predicted to grow to over $21 billion by the end of 2003 (Svetcov, 2000).

A second major driver is employer expectations. Fundamental job market expectations are changing. Today, employees are not expected to stay in the same job for long periods of time; 20-plus year careers are not expected. The current modes of careers include multiple careers, combinations of part-time work in multiple jobs, telecommuting, leaving and re-entering into the full-time work force, switching jobs, and so forth, and today's employee easily accepts the need to maintain a level of knowledge current with the career demands (Boyatzis & Kram, 1999). To complement these changes in employer expectations, employees have begun to accept the need for life-long learning.

A third driver is the profit potential. Cost savings may be obtained and if significant enough may drive up demand and costs may be lowered. For example, elective classes that do not have enough students enrolled in them on-campus may pick up enough distance students to make teaching the course more feasible (Creahan & Hoge, 1998). A final driver is the institution's mission. Most educational institutions serve a geographical region, either by charter or mission, and a distance-learning program may be a practical method to help satisfy this strategic mission (Creahan & Hoge, 1998).

However, the "commercialization" of education raises its own concerns about the basic process of learning (Noble, 1999). For example, are there any problems fundamental to the distance environment because of limited social interaction?

Retention may be one such problem. Carr (2000) reports a 50% drop-out rate for online courses. Tinto (1975) compared the learning retention of distance groups with traditional groups and found that the social integration was a key factor in successful retention of traditional

Table 1. Influences on the distance education industry

Table 1. Influences on the distance education industry	
Drivers	Concerns
Growth segment in education industry	Retention
Job market expectations	Fading Back
Life-long learning as an education paradigm	Less social learning
Profit center for educational institutions	Trust & isolation
Possible strategic competence	Impact of technology

groups. Haythornthwaite et al. (2000) think they found another one. They looked at how social cues such as text without voice, voice without body language, class attendance without seating arrangements, and students signing in without attending Internet class impacted students "fading back." They found that the likelihood of students "fading back" is greater in distance-learning classes than in face-to-face classes. From the United Kingdom, Hogan and Kwiatkowski (1998) argue that the emotional aspects of this teaching method have been ignored. Similar concerns are raised from Australia, where technology has been supporting distance- teaching for many years, as Hearn and Scott (1998) suggest that before adopting technology for distance teaching, education must acknowledge the social context of learning. Finally, two other factors, trust and isolation, have been researched by Kirkman et al. (2002), whereby communication helped improve the measures of trust in students using the virtual environment.

By definition, the paradigm of distance education changes the traditional education environment by expanding it to cover geographically dispersed learning. In turn, this means that students will probably respond differently to this environment than they do to the traditional classroom. In addition, academic researchers have always been interested in explaining how people react to the introduction of technology. This body of work can be useful to the distance education environment.

Poole and DeSanctis (1990) suggested a model called adaptive structuration theory (AST). The fundamental premise of the model is that the technology under study is the limiting factor or the constraint for communication. It further proposes that the users of the technology, the senders and the receivers, figure out alternative ways to send information over the channel (technology). A good example here is how a sender of e-mail may use combinations of keyboard characters or emoticons (i.e., :) – sarcastic smile, ;) – wink, :o – exclamation of surprise) to communicate more about their emotion on a subject to the receiver.

Ultimately, the key to realizing the potential of distance education is trading off the benefits and the concerns to produce a quality product. In the new Malcolm Baldridge evaluation criteria, companies are asked to better show a program's effectiveness through customer satisfaction. In turn, Gustafsson et al. (2000) show customer satisfaction linked significantly to quality at Volvo Car Corporation. Finally, in their more broad analysis of well-run companies, Peters and Waterman (1982) deemed customer satisfaction as a key factor contributing to the companies' performance.

Table 2. Questions that correlate significantly to satisfaction

ID	Question Statement	Correlation	
		Coef.	Sign.
16	I was satisfied with the content of the course	.605	.000
17	The tests were fair assessments of my knowledge	.473	.000
18	I would take another distance course with this professor	.755	.000
19	I would take another distance course	.398	.000
20	The course workload was fair	.467	.000
21	The amount of interaction with the professor and other students was what I expected.	.710	.000
22	The course used groups to help with learning	.495	.000
23	I would like to have had more interaction with the professor.	-.508	.000
26	The course content was valuable to me personally	.439	.000
28	Grading was fair	.735	.000
30	Often I felt "lost" in the distance class	-.394	.000
31	The class instructions were explicit	.452	.000
33	Feedback from the instructor was timely	.592	.000
34	I received personalized feedback from the instructor	.499	.000
36	I would have learned more if I had taken this class on-campus (as opposed to online)	-.400	.000
37	This course made me think critically about the issues covered.	.423	.000
38	I think technology (email, web, discussion forums) was utilized effectively in this class	.559	.000
39	I felt that I could customize my learning more in the distance format	.254	.001
42	The course content was valuable to me professionally	.442	.000
43	I missed the interaction of a "live," traditional classroom	-.341	.002
46	Overall, the program is a good value (quality/cost)	.258(1)	.017
LOHITECH	Aggregate of Yes votes in Q6 through Q15	.270(1)	.012
(1) While significant, the low correlation coefficient below .300 should be noted			

With these perspectives in mind, we suggest that these areas interact to identify satisfaction as one important measure of quality for distance education programs. Therefore, one of the key factors to a program's success will be the satisfaction of one of its key stakeholders – its students. If one can identify what helps satisfies students in a distance education environment, one has a better chance to develop a successful program.

THE RESEARCH STUDY

The distance program used in this study is one of the largest, online, AACSB-accredited MBA programs in the world (US News and World Report, 2001). The methodology used a questionnaire with a battery of 49 questions to gather the data. The questions were developed using the concepts and ideas from literature discussed earlier as a guide.

Once the subject identified his or her reference course, that subject's grade was obtained from administrative records and recorded. In addition, four other demographic questions gathered information on gender, number of courses taken, student status, amount of time expected to spend in the reference course, and the amount of time actually spent in the reference course (Martz et al., 2004).

Two sets of questions were used. The first set asked about the student's use of different technologies (i.e., chat, e-mail, streaming video, etc.) in the class and if used, how effective (five-point Likert: 1 = LO …. 5 = HIGH) did they believe the technology to be in helping them with the class. We created a new variable, LOHITECH, for analysis purposes. Using LOHITECH, respondents can be placed in one of two groups: one group that reported using three or less technologies, while the second group reported using four or more technologies in their reference class. The second set of questions asked students to rate (five-point Likert: 1 = Strongly Agree …. 5 = Strongly Disagree) their experience with the reference distance course against statements concerning potential influences for satisfaction. These questions associated a five-point rating scale to statements about the issues identified earlier. The order of the questions was randomly determined and the questionnaire was reviewed for biased or misleading questions by non-authors.

The questionnaire was sent to 341 students enrolled in the distance MBA program. In Fall 2002, the program served 206 students from 39 states and 12 countries. The majority of these students are employed full-time. The program used in this study has been running since Fall 1996 and has over 179 graduates. It offers an AACSB accredited MBA and its curriculum parallels the on-campus curriculum. Close to 33% of the enrolled students are female. The oldest student enrolled is 60 years old and the youngest is 22. The average age of all students enrolled is 35. Over 25 PhD qualified instructors participate in developing and delivering the distance program annually. Recently, the news magazine *US News and World Report* (2001) classified the program as one of the top 26 distance education programs.

There were 131 useable questionnaires returned. The students' final grade for their reference course was obtained and added to the questionnaire record as a variable. These were separated into two groups: 30 that had not yet taken a course and 101 that had completed at least one course. This second group, those students who had completed at least one course, provided the focus for this study.

RESEARCH RESULTS

Question 24, "Overall, I was satisfied with the course," was used as the subject's level of general satisfaction. The data set was loaded into SPSS for analysis. Table 2 shows that 23 variables, including LOHITECH, proved significantly correlated to satisfaction (Q24).

The large number of significant variables leads to the need for a more detailed analysis on how to group them (StatSoft, 2002). Kerlinger (1986, p. 590) suggests the use of factor analysis in this case "to explore variable areas in order to identify the factors presumably underlying the variables". An SPSS factor analysis was performed with a Varimax Extraction on those questions that had proven

Table 3. Correlation of final constructs to satisfaction

Construct (Component: Loading)	Correlation	Significance
Professor Interaction (Q18: .576, Q21: .643, Q33: .794, Q34: .849)	.771	.000
Fairness (Q17: .722, Q20: .738, Q28: .626, Q31: .512)	.695	.000
Course Content (Q16: .596, Q26: .850, Q39: .689, Q42: .825)	.588	.000
Classroom Interaction (Q23: -.354, Q30: -.514, Q36: -.809, Q43: -.770)	-.515	.000
Technology Use & Value (LOHITECH: .508, Q19: .596, Q22: .542, Q37: .494, Q38: .478, Q46: .700)	.624	.000

significantly correlated to satisfaction. All reliability coefficients (Cronbach Alpha) are above .7000 and all Eigenvalues are above 1.00, indicating an acceptable level for a viable factor (Kline, 1993; Nunnally, 1978). Finally, the five components explain 66.932% of the variance.

In summary, 22 variables from the questionnaire proved significantly correlated to satisfaction. A factor analysis of those 22 variables extracted five possible constructs. These constructs were labeled: Interaction with the Professor; Fairness; Content of the Course; Classroom Interaction; and Value, Technology & Learning, based upon the key characteristics of the underlying questions. Table 3 shows the results of combining the ratings for the questions in each construct and correlating each of them to satisfaction. As can be seen from the table, the constructs hold up well as five indicators of satisfaction.

FUTURE TRENDS

As mentioned earlier, the organization, the school in this case, is a key stakeholder in the success of a distance education program. The future success of distance programs depends largely on satisfying these critical success factors. Distance education courses and programs are not only used for providing an alternative delivery method for students but also to generate revenues for the offering unit/college/university. As the number of distance courses and programs increase at an exponential rate, the necessity to enhance quality and revenues also takes prominence. We conclude with a set of operational recommendations that can impact online program success (Table 4).

The data in this study indicate that a timely and personalized feedback by professors results in a higher level of satisfaction by students. The administrators therefore have to work closely with their faculty and offer them ways to enrich the teacher-student relationships. Paradoxically, a faculty member needs to use technologies to add a personal touch to the virtual classroom. For example, faculty should be encouraged to increase the usage of discussion forums, respond to e-mail within 24 to 48 hours, and keep students up-to-date with the latest happenings related to the course.

The data also indicate that good course content and explicit instructions increase student satisfaction in the virtual classroom. It may well be that this basically sets and manages the expectations for the distance student. This result suggests that faculty should have complete Web sites with syllabi and detailed instructions. In turn, this suggests that distance education administrators should focus their attention on providing faculty with support such as good Web site design, instructional designer support, test design, user interaction techniques, and so forth, appropriate for distance learning.

Since distance students' notion of value intertwines learning and technology, it is imperative that distance administrators offer, and faculty use, the available technology in the distance program. Technology in this case not only refers to the actual software and hardware features of the platform but also how well technology is adapted to the best practices of teaching. The results imply that if technology is available but not used, it lowers satisfaction. So, technology options that are not being used in a course should not appear available. For the program administrator, this would suggest adoption of distance platforms that are customizable at the course level with respect to displaying technological options.

CONCLUSION

This study attempts to identify potential indicators for satisfaction with distance education. A body of possible

Table 4. Recommendations to increase online program success

1	Have instructors use a 24-48-hour turnaround for e-mail.
2	Have instructors use a 1-week turnaround for graded assignments.
3	Provide weekly “keeping in touch” communications.
4	Provide clear expectation of workload.
5	Provide explicit grading policies.
6	Explicitly separate technical and pedagogical issues.
7	Have policies in place that deal effectively with technical problems.
8	Provide detailed unambiguous instructions for coursework submission.
9	Provide faculty with instructional design support.
10	Do not force student interaction without good pedagogical rationale.
11	Do not force technological interaction without good pedagogical purpose.
12	Collect regular student and faculty feedback for continuous improvement.

indicators was derived from the literature surrounding the traditional versus virtual classroom debate. A 49-question questionnaire was developed from the indicators and was administered to MBA students in an established distance education program. One hundred and one questionnaires from students with one or more distance classes were analyzed with the result that 22 variables correlated significantly to satisfaction. A factor analysis of the questionnaire data extracted five basic constructs: Professor Interaction, Fairness, Course Content, Classroom Interaction and Technology Use & Value. Several recommendations for implementing and managing a distance program were extracted from these constructs and discussed.

REFERENCES

Boyatzis, R.E., & Kram, K.E. (1999, Autumn). Reconstructing management education as lifelong learning. *Selections, 16*(1), 17-27.

Carr, S. (2000, February 11). As distance education comes of age the challenge is keeping students. *Chronicle of Higher Education.*

Creahan, T.A., & Hoge, B. (1998, September). *Distance learning: Paradigm shift of pedagogical drift?* Presentation at Fifth EDINEB Conference, Cleveland, OH.

Gustafsson, A., Ekdahl, F., Falk, K., & Johnson, M. (2000, January). Linking customer satisfaction to product design: A key to success for Volvo. *Quality Management Journal*, *7*(1), 27-38.

Haythornthwaite, C., Kazmer, M.M., Robins, J., & Showmaker, S. (2000, September). Community development among distance learners. *Journal of Computer-Mediated Communication*, *6*(1).

Hearn, G., & Scott, D. (1998, September). Students staying home. *Futures, 30*(7), 731-737.

Hogan, D., & Kwiatkowksi, R. (1998, November). Emotional aspects of large group teaching. *Human Relations, 51*(11), 1403-1417.

Institute for Higher Education Policy. (2000). *Quality on the line: Benchmarks for success in Internet distance education.* Washington, D.C.

Kerlinger, F.N. (1986). *Foundations of behavioral research* (3rd ed.). Holt, Rinehart & Winston.

Kirkman, B.L., Rosen, B., Gibson, C.B., Etsluk, P.E., & McPherson, S. (2002, August). Five challenges to virtual team success: Lessons from Sabre, Inc. *The Academy of Management Executive*, *16*(3).

Kline, P. (1993). *The handbook of psychological testing.* London: Routledge.

Martz, W.B, Reddy, V., & Sangermano, K. (2004). Assessing the impact of Internet testing: Lower perceived performance. In C. Howard, K. Schenk & R. Discenza (Eds.), *Distance learning and university effectiveness: Changing educational paradigms for online learning.* Hershey, PA: Idea Group Publishing.

Noble, D.F. (1999). Digital diplomas mills. Retrieved November 28, 2002, from *http://www.firstmonday.dk/issues/issue3_1/noble/index.html*

Nunnally, J. (1978). *Psychometric theory.* New York: McGraw-Hill.

Peters, T.J., & Waterman, R.H., Jr. (1982). *In search of excellence.* New York: Harper and Row.

Poole, M.S., & DeSanctis, G. (1990). Understanding the use of group decision support systems: The theory of adaptive structuration. In J. Fulk & C. Steinfeld (Eds.), *Organizations and communication technology* (pp. 173-193). Newbury Park, CA: Sage Publications.

Rockart, J.F. (1979, March-April). Chief executives define their own data needs. *Harvard Business Review.*

Statsoft. (2002). Retrieved November 30, 2002, from *http://www.statsoftinc.com/textbook/stfacan.html*

Svetcov, D. (2000). The virtual classroom vs. the real one. *Forbes,* 50-52.

Tinto, V. (1975). *Leaving college.* University of Chicago Press.

US News and World Report. (2001, October). *Best online graduate programs.*

KEY TERMS

Classroom Interaction: The interaction that can only be achieved face-to- face in a classroom. For example, the real-time feedback of facial expressions is not (yet) available in a distance course and so would be considered "classroom interaction".

Concerns of "Commercialization": The negative factors that the implantation and use of distance education may create.

Course Content: The main themes covered in a course.

Critical Success Factors: The few key areas in which activities must "go right" so that a project of program succeeds (Rockart, 1979).

Exploratory Factor Analysis: A process used to identify statistically significant constructs underlying a set of data.

Fairness: A subjective term defining the level to which a student feels he or she was treated fairly by the professor with respect to the class, including but not limited to test questions, grading, schedule flexibility, and so forth.

Market Drivers for Distance Education: The key elements that seem to be driving the diffusion and usage of distance education in the marketplace.

Professor Interaction: The amount of communication (e-mail, phone calls, video, chat rooms, etc.) that occurs between a student and the professor.

Satisfaction Constructs for Distance Education: Five constructs identified that seem to help identify satisfaction in distance education programs.

Technology Use: The usage of a technology whether it be e-mail, chat rooms, automated tests, software, and so forth.

Technology Value: The user's benefits (perceived and actual) over the costs (perceived and actual) created by the use of technology.

Critical Success Factors of ERP Implementation

Leopoldo E. Colmenares
University Simon Bolivar, Venezuela

Jim O. Otieno
Middlesex University, UK

INTRODUCTION

An enterprise resource planning (ERP) system is an integrated set of programs that provides support for core organizational activities, such as manufacturing and logistics, finance and accounting, sales and marketing, and human resources. An ERP system helps the different parts of an organization share data and knowledge, reduce costs, and improve management of business processes. In spite of their benefits, many ERP systems fail (Stratman & Roth, 1999). Implementing an ERP system is a major undertaking. About 90% of ERP implementations are late or over budget (Martin, 1998), and the success rate of ERP systems implementation is only about 33% (Zhang et al., 2003).

Over the past few years, limited research has been conducted about ERP implementation issues: mainly, case studies in individual organizations have been reported. A major problem with such ERP case studies is that very few implementation failures have been recorded in the literature, and thus, the reasons why implementations fail are not known to practitioners and researchers. That is a motivation toward conducting empirical studies to explore critical factors that affect ERP systems implementation. A recent summary of ERP literature stated that research of critical success factors (CSFs) in ERP implementation is rare and fragmented (Nah, Lau, & Kuang, 2001). The idea of identifying CSFs as a basis for determining the information needs of managers was popularized by Rockart (1979). CSFs are those factors that are critical to the success of any organization, in the sense that, if objectives associated with the factors are not achieved, the organization will fail—perhaps catastrophically (Rockart, 1979). In the context of ERP project implementation, CSFs represent the essential ingredients, without which a project stands little chance of success. This study examines the CSFs for implementing ERP systems in Venezuela. Managers from seven corporations, who were identified as having key roles in ERP systems implementation, were surveyed in order to assess empirically which CSFs are critical in leading a successful implementation of ERP systems in Venezuela. This article is organized into four sections. First, ERP-related literature is reviewed. The next section introduces the research methodology, followed by presentation of the results. The article ends with the conclusions and implications for future research and practice.

BACKGROUND

Implementing an ERP system is not an easy task. It can cause dramatic changes that need to be carefully administered if the potential advantages of an ERP solution (Al-Mudimigh, Zairi, & Al-Mashari, 2001) are to be gained. In some well-documented cases, spectacular results have been achieved (Johnston, 2002). There is, on the other hand, a relatively high failure rate: it was reported that three-quarters of ERP projects were judged to be unsuccessful by the ERP implementing firms (Kyung & Young, 2002). Also, failures are much less extensively documented. As a result, pitfalls to be avoided tend to be less well known. Venezuelan companies are just starting to use ERP systems. They started applying ERP concepts late in the 1990s. Because of the complex and integrated nature of ERP, and the large investment involved, it is imperative for organizations to study the experiences of others and to learn from their practices and success factors (Zairi et al., 2000). Identifying CSFs relevant to local companies is one way to increase the chances of a successful local implementation (Sum, Ang, & Yeo, 1997).

A literature review was conducted to understand the CSFs in successful ERP implementations. The review covered numerous articles (Bingi, Sharma, & Godla, 1999; Esteves & Pastor, 2001; Falkowski et al., 1998; Holland & Light, 1999; Nah, Lau, & Kuang, 2001; Rosario, 2000; Stefanou, 1999; Sumner, 1999; Wee, 2000). The literature varies according to the variables required for implementation success, so there is no general consensus as to the factors that are key to success in ERP implementation. It is probably a combination of factors that is important in explaining ERP implementation success (Zhang et al., 2003). From the review, 20 factors emerged as critical to the successful implementation of ERP systems. They were obtained after careful analysis and grouping of related subfactors:

1. Top management support
2. User training
3. Use of consultants
4. User participation
5. Vendor package selection
6. Use of steering committee
7. Discipline and standardization
8. Minimal customization
9. Use of vendor's development tools
10. Best people full time
11. Technical and business knowledge
12. Implementation approach
13. Clear goals, focus, and scope
14. Business process reengineering
15. Project management
16. Effective communications
17. Presence of a champion
18. Interdepartmental cooperation and communication
19. Management of expectations
20. Vendor/customer partnership

RESEARCH METHODOLOGY

The choice of an appropriate research methodology is critical in guiding researchers on how best to meet research objectives. In this study, the purpose was to discover the perceptions and experiences of companies using ERP systems in Venezuela, and to use that information as the basis of data collection. The analysis has enabled the identification CSFs of ERP systems implementation in Venezuelan companies.

The targets of the study were the organizations that implemented ERP systems successfully. The key informant method was used for collecting information in a social setting by surveying (or interviewing) a selected number of participants. Seven firms were identified from the list provided by ERP vendors. We contacted the ERP project managers in charge of ERP implementation of each company. About 100 questionnaires were sent to the ERP project managers of each firm, who forwarded the questionnaires to the project team members in charge of individual processes. A total of 72 questionnaires were returned, of which 69 were valid.

The questionnaire consisted of two main parts: the company background and the CSFs. The first part was designed to determine characteristics such as size of the company, type of industry, location of company, etc. The second part consisted of 20 statements about the success factors of ERP systems implementation derived from the literature review. The language used in the survey was Spanish. Translation was easy, because Venezuelans used original English terms for many technical and management concepts and especially for information systems and computing concepts.

Participants were requested to rate the importance of each CSF using a five-point Likert scale, where a score of 5 indicated "extremely critical," and a score of 1 indicated "not critical." This method was employed on the grounds that a rating method avoids the problem of having to consider 20 CSFs simultaneously in order to rank them. The data collected were then analyzed by using SPSS. Based on the responses, descriptive statistics, factor analysis (FA), and reliability tests were carried out to identify the CSFs for the successful implementation of ERP systems and data validity, respectively.

RESULTS

Ranking

The importance rating of the 20 CSFs is listed in Table I. The individual mean value of the Likert rating scale is the popular usage indicator for measuring an item's importance, without regard to the other items; therefore, the higher the value, the more important the factor. Most items are rated above the 3.0 scale (midpoint). The three most important factors, in order of declining importance, are top management support, presence of a champion, and project management, with a mean value ranging from 4.80–4.64. Just as the literature argues, these are key items for ERP implementation management (Johnston, 2002). Conversely, use of steering committee, business process reengineering, and use of vendor's development tools, are the three items lowest in the list, with a mean value ranging from 2.95–2.06.

Factor Analysis

In an attempt to reduce the number of items (CSFs), and to understand their underlying structure, a factor analysis (FA) was performed. FA is a data reduction technique that uses correlations between data variables. The underlying assumption of FA is that a number of factors exist to explain the correlations or interrelationships among observed variables (Chatfield & Collins, 1992). For the present study, FA was performed on all 20 variables using principal components extraction (Tabachnick & Fidell, 1989). The goal of this method is to extract maximum variance from the data set within each factor. It is basically used to reduce a large number of variables to a smaller number of components. The measure of sampling adequacy for the 20 items was 0.87, indicating that the items were suitable for factoring (Kaiser, 1974).

Table 1.Ranking of CSFs

Rank	CSF	Mean	Standard Deviation
1	Top management support	4.80	0.62
2	Presence of a champion	4.75	0.85
3	Project management	4.64	0.92
4	Best people full time	4.58	0.60
5	Effective communications	4.51	0.85
6	Interdepartmental cooperation and communication	4.40	0.91
7	Management of expectations	4.36	1.02
8	Technical and business knowledge	4.33	1.21
9	User participation	4.22	0.82
10	Discipline and standardization	4.09	0.85
11	Vendor package selection	4.02	0.61
12	User training	4.01	1.12
13	Implementation approach	4.00	1.20
14	Clear goals, focus, and scope	3.89	1.14
15	Use of consultants	3.75	0.85
16	Minimal customization	3.68	1.52
17	Vendor/customer partnership	3.15	0.52
18	Use of steering committee	2.95	0.63
19	Business process reengineering	2.84	0.55
20	Use of vendor's development tools	2.06	0.42

Notes: N: 69; Scale: 1–5 (5: "extremely critical," 1: "not critical")

A three-stage factor analysis was conducted with an orthogonal (varimax) rotation to obtain a stable factor structure (Rai et al., 1996), resulting in easily interpretable factors. Under this three-round factor analysis, items were omitted according to the following two criteria: (1) no loading greater than 0.35, or (2) loading greater than 0.35 on two or more factors (Kim & Mueller, 1978).

Table 2 shows the results of this analysis. A first-factor analysis was conducted and produced five factors. According to the two criteria, three items were dropped. A second factor analysis on the remaining 17 items resulted in four factors and the dropping of three items. Finally, a three-factor structure was derived that kept a total of 14 items after three iterations. The minimum eigenvalue from a varimax rotation for which a factor was to be retained was set at 1.0 in order to satisfy the minimum eigenvalue criterion (Nunnally, 1978). The three factors are as follows:

1. *Factor 1*, named "ERP implementation management," comprises six items relating to implementation management: top management support, presence of a champion, project management, management of expectations, implementation approach, and clear goals, focus, and scope.
2. *Factor 2*, named "user aptitudes and communication," comprises four items relating to user participation: effective communication, interdepartmental cooperation and communication, user participation, and user training.
3. *Factor 3*, named "technical knowledge," comprises four items relating to knowledge of business and ERP: best people full time, technical and business knowledge, use of consultants, and discipline and standardization.

Cronbach alpha coefficients were calculated to test the reliability of these CSFs, as shown in the last row of Table 2. The reliability of coefficients obtained ranges from 0.56 (Factor 3) to 0.88 (Factor 1). Srinivasan (1985) proposed that a coefficient of 0.7 or higher is acceptable, while a coefficient of 0.5 or higher is considered sufficient when dealing with exploratory research combined with unvalidated data. Thus, the reliability coefficients in this study are deemed acceptable. The strength of factor analysis is that it provides a basis for data reduction. Rather than look at all 20 items, only three factors need be examined. That simplifies the rankings and clarifies the most important items. Rather than focus on individual items, practitioners and researchers can focus on the broad set of items represented by the essential factors.

FUTURE TRENDS

Many companies around the world are following the trend toward making large investments in implementing ERP systems. Several approaches and methodologies of ERP project implementation recognise a series of critical factors that must be carefully considered to ensure successful implementation of an ERP system project. In essence, there are dominant critical factors hypothesised to play an overriding role in the implementation of an ERP project, and they should be ongoing throughout all implementation levels. Clearly, the dominant factors are those that will shape the overall project culture and, subsequently, the organisational cultura, as ERP is far reaching in nature.

Post-ERP activity seems to follow a clear path. A Deloitte Consulting study of 62 companies segments post-ERP activity into three stages. The first stage entails a three- to nine-month productivity decline that is overcome by redefining jobs, establishing new procedures, fine-tuning ERP software, and taking charge of the new streams of information created by the platform. The second stage, which lasts from six to 18 months, involves skills development, structural changes, process integration, and add-on technologies that expand ERP functionality. The third stage, of one to two years in duration, is one of transformation, where the synergies of people, processes, and technology reach a peak. Perhaps most important, ERP forces discipline and organization around processes, making the alignment of IT and business goals more likely in the post-ERP era.

CONCLUSION

Despite the benefits that can be achieved, there is already evidence of failure in projects related to ERP systems implementation (Davenport, 1998). It is, therefore, important to find out what the CSFs are that drive ERP project success.

Table 2. Results of factor analysis

		Factor Loading (Final Factor Structure 3rd FA)		
CSF	**Mean**	**F1**	**F2**	**F3**
Top management support	4.80	0.608		
Presence of a champion	4.75	0.541		
Project management	4.64	0.586		
Best people full time	4.58			0.741
Effective communications	4.51		0.565	
Management of expectations	4.40	0.420		
Interdepartmental cooperation and communication	4.36		0.452	
Technical and business knowledge	4.33			0.584
User participation	4.22		0.562	
Discipline and standardization	4.09			0.387
Vendor package selection	4.02			
User training	4.01		0.510	
Implementation approach	4.00	0.478		
Clear goals, focus, and scope	3.89	0.411		
Use of consultants	3.75			0.573
Minimal customization	3.68			
Vendor/customer partnership	3.15			
Use of steering committee	2.95			
Business process reengineering	2.84			
Use of vendor’s development tools	2.06			
Eigenvalue		7.56	2.54	1.20
Percentage of variance		58.50	14.36	8.65
Cumulative percentage of variance		58.50	72.86	81.51
Cronbach alpha coefficient		0.88	0.74	0.56

According to the respondents in this study, the six top CSFs for ERP systems implementation in Venezuelans firms are top management support, presence of a champion, project management, best people full time, effective communications, and management of expectations.

This research has derived three composite CSFs in ERP systems implementation in Venezuela:

1. ERP implementation management
2. User's aptitudes and communication
3. Technical knowledge

Four of the six top individual items contribute to the composite factor of ERP implementation management, which is, by far, the most important. It is made up of items that are concerned with the management of ERP systems implementation projects.

The results of ranking ERP CSFs in this study are largely consistent with the literature review, though the relative ranking of some factors varied. In the literature, top management support, change management, presence of a champion, and management of expectations are the four most often cited critical factors.

A majority of factors, 12, were rated as critical (rating >4). Only one factor, business process reengineering, which is rated as critical in most articles reviewed, was not rated as critical in this study (rating <3). Hence, the perceptions on CSFs of Venezuelan managers involved in ERP systems implementation projects are largely consistent with the findings reported in the literature.

There are a number of questions still to be determined. For example, although this article establishes the relative importance of CSFs in seven firms, it has not established the reasons. Future studies could look at differences by size of firms, by industry type, by number of locations, by number of customers, etc.

Finally, it should be noted that all CSFs are based on previous research; the success items can be modified when further research is conducted. For instance, discovery-oriented research through comprehensive interviews with several top-level managers in an attempt to identify new CSFs is an option.

REFERENCES

Al-Mudimigh, A., Zairi, M., & Al-Mashari, M. (2001). ERP software implementation: An integrative framework. *European Journal of Information Systems, 10,* 216–226.

Bingi, P., Sharma, M. K., & Godla, J. (1999). Critical issues affecting an ERP implementation. *Information Systems Management, 16*(2), 7–14.

Chatfield, C., & Collins, A. J. (1992). *Introduction to multivariate analysis* (3rd ed.). London: Chapman & Hall.

Davenport, T. H. (1998). Putting the enterprise into the enterprise system. *Harvard Business Review*, July–August, 121–131.

Esteves, J., & Pastor, J. (2001). Analysis of critical success factors relevance along SAP implementation phases. In *Proceedings of the Seventh Americas Conference on Information Systems* (pp. 1019–1025), Boston, MA.

Falkowski, G., Pedigo, P., Smith, B., & Swanson, D. (1998). A recipe for ERP success. *Beyond Computing,* September, 44–45.

Holland, C. P., & Light, B. (1999). A critical success factors model for ERP implementation. *IEEE Software, 16,* 30–36.

Johnston, S. (2002). ERP: Payoffs and pitfalls. *HBS Working Knowledge,* October, 14–21.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika,* (39) 31–62.

Kim, J., & Mueller, C. W. (1978*). Factor analysis: Statistical methods and practical issues*. Thousand Oaks, CA: Sage Publications.

Kyung-Kwon, H., & Young-Gul, K. (2002). The critical success factors for ERP implementation: An organizational fit perspective. *Information and Management, 40*, 25–40.

Martin, M. H. (1998). An ERP strategy. *Fortune*, February, 95–97.

Nah, F. F. -H., Lau, J. L. -S., & Kuang, J. (2001). Critical factors for successful implementation of enterprise systems. *Business Process Management, 7*(3), 285–296.

Nunnally, J. C. (1987*). Psychometric theory*. Englewood Cliffs, NJ: Prentice Hall.

Rai, A., Borah, S., & Ramaprasad, A. (1996). Critical success factors for strategic alliances in the information technology industry: An empirical study. *Decision Sciences, 27*(1), 141–155.

Rockart, J. (1979). Chief executives define their own data needs. *Harvard Business Review, 57*(2), 238–241.

Rosario, J. G. (2000). On the leading edge: Critical success factors in ERP implementation projects. *Business World*, May, 27–32.

Srinivason, A. (1985). Alternative measures of system effectiveness: Associations and implications. *MIS Quarterly, 9*(3), 243–253.

Stefanou, C. J. (1999). Supply chain management (SCM) and organizational key factors for successful implementation of Enterprise Resource Planning (ERP) systems. In *Proceedings of Fifth Americas Conference on Information Systems* (pp. 800–802).

Stratman, J., & Roth, A. (1999). Enterprise resource planning competence: A model, propositions and pre-test, design-stage scale development. In *Proceedings of 30th Decision Science Institute* (pp. 1199–1201).

Sum, C. C., Ang, J. S. K., & Yeo, L. N. (1997). Contextual elements of critical success factors in MRP implementation. *Production and Inventory Management Journal, 3*, 77–83.

Sumner, M. (1999). Critical success factors in enterprise wide information management systems projects. In *Proceedings of Fifth Americas Conference on Information Systems* (pp. 232–234).

Tabachnick, B. G., & Fidell, L. S. (1989). *Using multivariate statistics*. New York: Harper Collins.

Wee, S. (No date.) Juggling toward ERP success: Keep key success factors high. *ERP News*. Retrieved January 22, 2002, from *http://www.erpnews.com/erpnews/ erp904/ 02get.html*

Zairi, M., Al-Mudimigh, A., & Jarrar, Y. (2000). ERP implementation critical success factors—The role and impact of business process management. In *Proceedings of the 2000 IEEE International Conference on Management of Innovation and Technology* (pp. 122–127).

Zhan, L., Lee, M., Zhang, Z., & Banerjee, P. (2003). Critical success factors of enterprise resource planning systems implementation success in China. In *Proceedings of the 36th Hawaii International Conference on System Sciences* (pp. 562–567).

KEY TERMS

Business Process Reengineering (BPR): Any radical change in the way in which an organisation performs its business activities; BPR involves a fundamental rethink of the business processes followed by a redesign of business activities to enhance all or most of its critical measures—costs, quality of service, staff dynamics, etc.

Critical Success Factors (CSFs): CSFs indicate the few key areas of activity in which favorable results are absolutely necessary for the manager to succeed.

Enterprise Resource Planning (ERP) System: Business software package for running every aspect of a company, including managing orders, inventory, accounting, and logistics. Well-known ERP software providers include BAAN, Oracle, PeopleSoft, and SAP, collectively known to industry insiders as the "BOPS."

Factor Analysis: Any of several methods for reducing correlational data to a smaller number of dimensions or factors; beginning with a correlation matrix, a small number of components or factors are extracted that are regarded as the basic variables that account for the interrelations observed in the data.

Project Management: This is the ensemble of activities concerned with successfully achieving a set of goals. This includes planning, scheduling, and maintaining progress of the activities that comprise the project.

Software Package: Written programs or procedures or rules and associated documentation that pertain to the operation of a computer system and that are stored in read/write memory.

Systems Implementation: Customisation or parameterization and adaptation of the software application according to the needs of the organisation.

Critical Trends in Telecommunications

John H. Nugent
University of Dallas, USA

INTRODUCTION

As in all industries, in order to win in a market, it is important to know as much as possible about that market and have at one's disposal tools that will provide insight and competitive advantage when properly, collectively, consistently, and timely applied. This article presents a series of powerful but easy-to-use and understand analytical and operational tools that deliver insight and competitive advantage to the wireless telecommunications professional. It should be stated, moreover, that as with all good tools, the tools and models as presented herein transition across industry lines and are not limited to the wireless telecommunications industry alone.

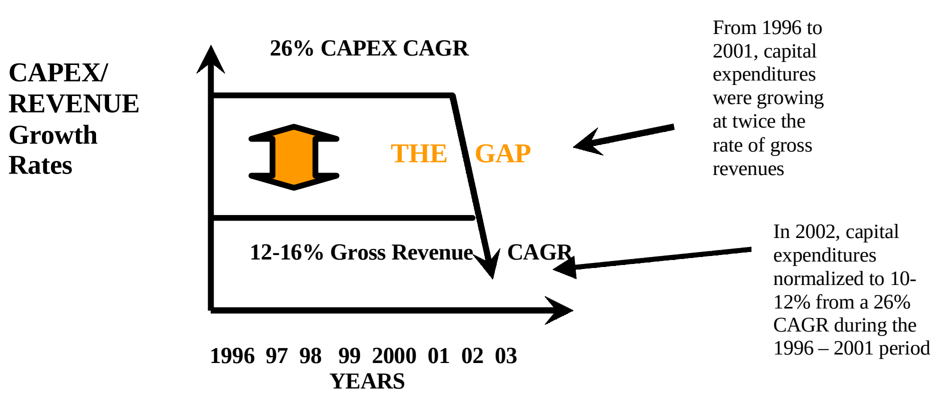
BACKGROUND: ANALYTICAL TOOLS AND MODELS

Because of the turmoil experienced in the telecommunications industry over the past several years, it is useful to view tools that can assist the telecommunications professional with understanding the market(s) and the trends at play. Looking at the telecommunications market from 1996 to 2003, it can be seen that the market exploded in the first half of this period with a 26% cumulative annual capital expenditure growth rate (CAPEX CAGR), and collapsed in the later part of this period (Lehman Brothers, 2000; Hilliard Consulting Group, Inc., 2003).

When capital expenditures so far outstrip gross revenue growth, one knows this situation cannot continue unabated. That is, a return to a more normal state must return. In order to discern approximately when a return to a more normal state will come about, one may examine the underlying market drivers (Nugent, 2001, 2003). Market drivers will often signal the size, breadth and depth of a market.

During the period noted (1996-2003), several large drivers were evident. The first was identified as the Y2K driver. Here, many firms determined it to be better, easier, and less costly and risky to replace versus remediate infrastructure equipment. But. here it was known this driver would be satiated by 2000. A second major driver was The Telecommunications Act of 1996 (www.fcc.gov). This act brought about the creation of many new telecom competitors that raised billions of dollars in the equity and debt markets that went on a spending spree. However, most of these firms had flawed business plans, and through competitive thrusts by the incumbents in the form of administrative delay, regulatory appeal, and litigation, these new entrants were literally bled dry via the consumption of cash in non-revenue producing activities such as regulatory appeals and litigation, and doomed to failure (Nugent, 2001, 2003). Understanding how significant incumbents fight and how they use the most strategic weapons of all – cash position and cash flow – the demise of these new incumbents could be foreseen. Another significant driver was the explosion in the number of

Figure 1. Revenue capital expenditure growth rate comparisons



Source: Hilliard Consulting Group, Inc., 2003 wireless customers brought about by the "Digital One Rate" plan initiated by AT&T. Here, wireless growth exploded from approximately 50 million subscribers to over 120 million in just several years. However, there are models that indicate this type of market satiates at approximately 50% of the overall population or 70% of the adult population (Nugent, 2003). In the U.S., this satiation point is approximately 145 million narrowband voice subscribers – approximately where we are today. So this spending spurt on narrowband voice wireless customer premise equipment (CPE), and infrastructure equipment could have also been estimated to end as the market approached satiation.

Hence, the telecommunications market downturn should not have been a surprise to anyone, as an understanding of the principal market drivers would have permitted an estimate of the market's size, breadth and depth.

FUTURE TRENDS

At a high level, it is also important to understand where a market is today, and where it is going to be tomorrow. To help understand these conditions, a State, Gap, and Trend (SG&T) Analysis tool provides helpful insight (Hilliard Consulting Group, Inc., 2003; Wolford-Ulrich, 2004):

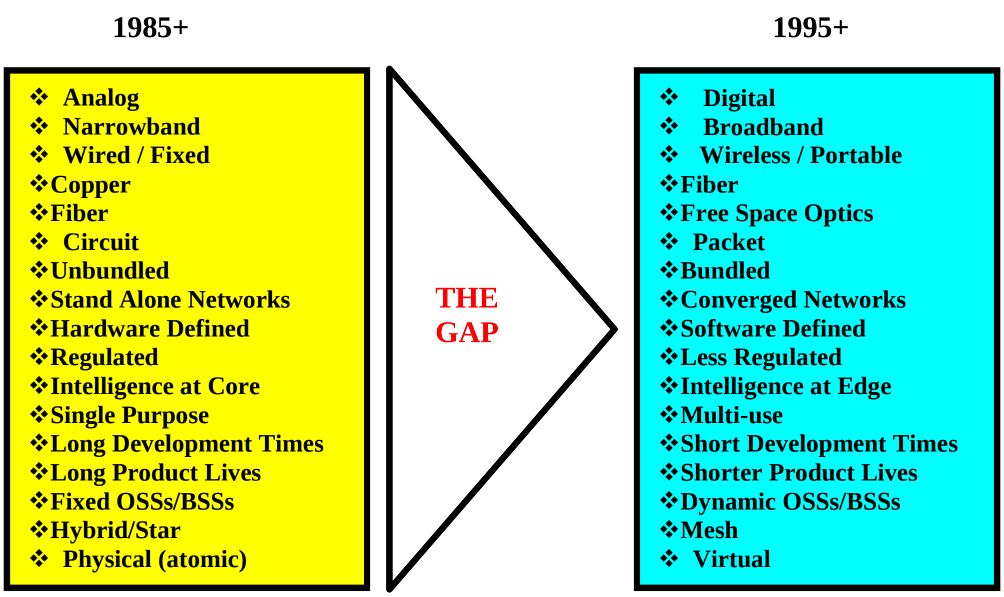
The development of a SG&T tool calls for a "one for one" transition (a "this to that" scenario over a period of time – there can be no ambiguities). Hence, a current and future state can be determined with some clarity.

An examination of this SG&T tool presented below indicates that the telecommunications world is moving from a fixed, tethered, narrowband, analog, circuit-based world, to one principally comprised of mobile, wireless, broadband, digital, packet-based communications. This transition portends significant issues for land-based carriers whose assets principally are in big physical plant (central offices, switching facilities, tethered trunks and circuits, etc.). This model further indicates that land-based carriers' assets are probably depreciating significantly faster than their balance sheets indicate. Supporting this premise is the decline in the number of residential landlines from approximately 168 million lines in 2001 to approximately 152 million landlines in 2003 (FCC/Solomon Smith Barney, 2002).

Moving from a macro model of market trends (SG&T) analysis previously, it can also be seen on a micro level (Product Curve) what attributes successive telecommunications products must follow to win in future markets (Hilliard Consulting Group, Inc., 2003). Here, a Product Curve model is most helpful.

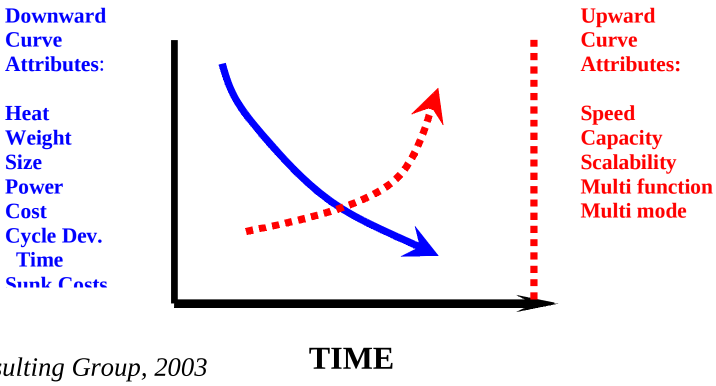
The product curve demonstrates that devices (network and CPE) need to become smaller, consume less power, weigh less, give off less heat, cost less, be developed in faster and faster cycle times, and have less in sunk development costs, while at the same time do more: operate at faster speeds and higher capacities while performing more functions to win in future markets. The product curve also portends troubles for land-line carriers as it can be seen in not too many years, the central office

Figure 2. State gap and trend analysis: Technology transition



Source: Hillard Consulting Group (2003)

Figure 3. Product curve



Source: Hilliard Consulting Group, 2003

of today will be displaced by a laptop wireless broadband tool of tomorrow. The SG&T Analysis and the Product Curve shown above only highlight some important attributes, there are numerous others that may, and should be added for a fuller comprehension of the industry.

To see the product curve in action, a comparison of the original Motorola "Brick" cell phone may be made with the sleek small wireless communication devices we use today. Here, we can see that the devices have become smaller, weigh less, consume less power, cost less, give off less heat, but do more.

Mix Shift Analyses are another way to discern important market changes (Hilliard Consulting Group, Inc., 2003). Here, many consulting firms forecast where we are today, and likely will be tomorrow.

MIX SHIFTS

These mix shift analyses indicate that the telecommunications industry will move from a tethered to a wireless world in the relatively near future, while at the same time the mix of telecommunications traffic will shift from principally voice to principally data. This mix shift does not mean that voice traffic will decline; rather it indicates that data traffic will grow dramatically compared to voice over the period indicated. Moreover, voice will largely become data as Voice over the Internet Protocol becomes more the norm. This trend is highlighted in the SG&A analysis above where we see a shift from circuit to packet communications.

Table 1. Transport shift

Transport \ Year	2003	2008	2015
Landline	80%	50%	10%
Wireless	20%	50%	90%

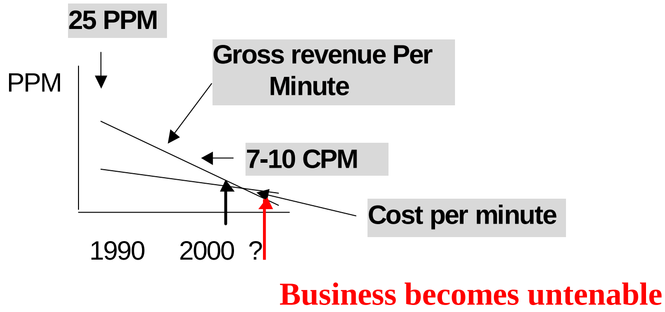
Source: Hilliard Consulting Group, Inc. (2003)

Table 2. Mode shift

Service \ Year	1995	2020
Voice	90%	10%
Data	10%	90%

Source: Hilliard Consulting Group, Inc. (2003)

Figure 4. The minute margin squeeze model for the interexchange carrier (IXC) market



PPM = Price Per Minute, CPM = Cost Per Minute
Source: Hilliard Consulting Group, Inc. (2003)

Operational Tools and Models

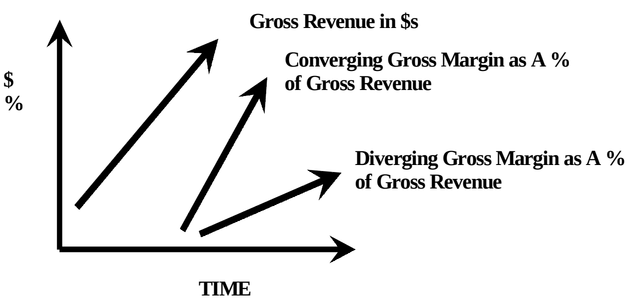
In undertaking a business analysis, there are several tools that work extremely well in examining the performance (present and future) of a telecommunications enterprise.

One set of such tools is referred to as slope analysis tools (Nugent, 2003). Here, a unit price/unit cost model may be used to take current operating parameters and project trends discerned here to future periods in order to foresee likely outcomes (and change strategies today if needed before others discern such trends).

The form of a unit price/unit cost model examined is the Minute Margin Squeeze Model.

This unit price/unit cost model indicates that unit prices are declining significantly faster than unit costs. By plotting actual data for a few years (1990-1993) and then extending these slopes to future periods, one is able to discern likely outcomes. In this case, it would appear the long distance market (IXC) would become quite unattractive after 2002. Viewing the trending in the narrowband voice market, we see a similar trend (58 cents/minute in 1993 to 10 cents or less a minute today).

Figure 5. Converging/diverging gross margin analysis



Source: Hilliard Consulting Group, Inc. (2003)

Another form of slope analysis is the Converging/ Diverging Gross Margin Analysis. Here, actual data from the income statement is plotted for several periods. Converging gross margins indicate increasing operational efficiency while diverging gross margins signify decreasing operational efficiency.

Examining AT&T's Initial Public Offering (IPO) for its wireless unit demonstrates the importance of this tool. The IPO for AT&T's wireless unit was well received and the stock price climbed immediately. However, a reading of the offering document would have shown that the gross margin decreased (DIVERGED) by over 50% in the preceding annual period. This indicated significant operational troubles. By discerning this diverging gross margin and drilling down to determine the reasons, one would have discerned that AT&T's successful "Digital One Rate Offering" caught the company short, far short, of network capacity to support demand. Hence, AT&T had to go off-net and pay other carriers high fees to originate or terminate its traffic. This understanding would have highlighted additional issues facing AT&T Wireless: either take on more debt or dilute current shareholders further by issuing more stock in order to build more infrastructure.

One final operational tool is the Discriminant Function Algorithm (DFA) used to discern changes in corporate health on a prospective basis – inflection points (Amdocs Corp, 2003; Slywotzky, Morrison, Moser, Mundt, & Quella, 1999). This model uses Altman's Z Score algorithm for determining bankruptcy on a prospective basis (Altman, 1983). However, unlike Altman where he uses absolute scores, the DFA model only cares about changes in score - either positive or negative (Nugent, 2003). As an example of what this tools yields on a prospective basis, AT&T is again viewed.

As can be seen, AT&T's Z Score declined from over 5 in 1997 to under 1 by 2001. This is a dramatic negative decline. Yet, during much of this period, Wall Street was in love with this stock. Had one begun plotting this Z score in 1997, one could have discerned by 1999 that things were heading south long before others foresaw this decline.

CONCLUSION

High level analytical and operational tools and models can assist the wireless telecommunications professional in understanding the telecommunications market's characteristics, life cycles, trends, directions, limits, drivers, and likely prospective performance. These tools demonstrate that wireless communications will follow the same life cycle characteristics as wireline communications. Proper utilization of such tools can lead to important and timely insights hopefully leading to competitive advantage based upon an early detection of changes in marketplace dynamics.

REFERENCES

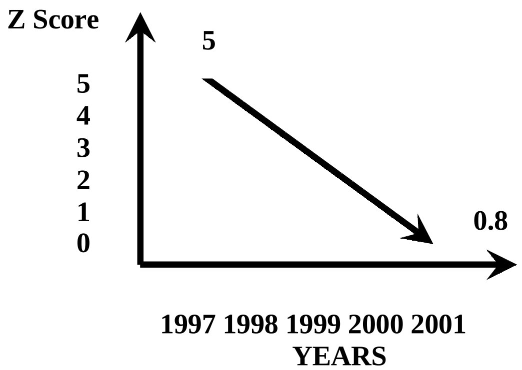
Altman, E.I. (1983). *Corporate financial distress: A complete guide to predicting, avoiding, and dealing with bankruptcy*. John Wiley & Sons.

Amdocs Corporation (2003). State, gap & trend analysis, Available at www.amdocs.com

AT&T Corporation, www.att.com

FCC, Federal Communications Commission, www.fcc.gov

Figure 6. AT&T Z Score, 1997-2001



Source: Hilliard Consulting Group, Inc. (2003)

Hilliard Consulting Group, Inc. (2003). McKinney, TX. Models first developed by Dr. J. Nugent, CPA, CFE, at the Hilliard Consulting Group, Inc. in 1999, and refined in 2001, and 2003.

Lehman Brothers (2000). Report on capital expenditures in the telecommunications industry. Study published for customers of the firm. Not publicly available.

Motorola Corporation.

Nugent, J.H. (2001). Telecom downturn was no surprise. *Dallas Fort Worth TechBiz*, September 10-18, p.22. Available at *www.dfwtechbiz.com*

Nugent, J.H. (2003). *Plan to win: Analytical and operational tools – Gaining competitive advantage* (2nd ed.). New York: McGraw-Hill.

Slywotzky, A.J., Morrison, D.J., Moser, T., Mundt, K.A., & Quella, J.A. (1999). *Profit patterns*. Random House.

Solomon Smith Barney (2002). *The Wall Street Journal*, April 18, p. B5.

The Telecommunications Act of 1996 – *www.fcc.gov*

Wolford-Ulrich (2004). *www.inflectionpoints.com*

KEY TERMS

Capital Equipment Expenditures (CAPEX): Usually measured as a percent of gross revenue.

Converging/Diverging Gross Margin Analysis: A slope analysis tool used to plot actual sales and gross margin data for several periods in order to discern trends and likely outcomes. Measure operational efficiency or inefficiency.

Cumulative Annual Growth Rate (CAGR): The percent of growth from one annual period to the next.

Customer Premise Equipment (CPE): End-user equipment.

Discriminant Function Algorithm: A term developed for Edward Altman in describing his Z Score analytical tool in determining the likelihood of an enterprise going into bankruptcy on a prospective basis. Used as a method for determining inflection points – changes in corporate health versus for bankruptcy prediction.

Inflection Points: Significant changes in corporate performance.

Interexchange Carriers (IXC): Long distance companies that transport inter local access transport area (LATA) traffic.

Local Access Transport Area (LATA): A geographic area defined by the FCC.

Minute Margin Squeeze: Also known as unit price/unit cost model.

Mix Shifts: Shifts in the market between major components usually requiring different technology or solutions.

Product Curve: This tool takes a micro view of transitioning requirements or attributes successive solutions must adhere to in order to win in the future market place.

Slope Analysis: The plotting and visualization of certain operating functions in order to discern trends, often before others see them thereby permitting alteration of strategies in order to gain competitive advantage.

State Gap & Trend Analysis: A tool used to present in a structured format current market or technology states as well as future states. This analysis requires a "one for one" transition – a "this to that" view. This model calls for no ambiguities. The perilous part of this tool is determining how to transition the gap – where one has to be by when with what.

Unit Price/Unit Cost Model: A slope analysis tool used to plot actual unit prices and unit costs for several periods in order to discern future trends and likely outcomes.

Cross-Cultural Research in MIS

Elena Karahanna
University of Georgia, USA

Roberto Evaristo
University of Illinois, Chicago, USA

Mark Srite
University of Wisconsin–Milwaukee, USA

INTRODUCTION AND BACKGROUND

"Globalization of business highlights the need to understand the management of organizations that span different nations and cultures" (Srite et al., 2003, p. 31). In these multinational and transcultural organizations, there is a growing call for utilizing information technology (IT) to achieve efficiencies, coordination, and communication. However, cultural differences between countries may have an impact on the effectiveness and efficiency of IT deployment. Despite its importance, the effect of cultural factors has received limited attention from information systems' (IS) researchers. In a review of cross-cultural research specifically focused on the MIS area (Evaristo, Karahanna, & Srite, 2000), a very limited number of studies were found that could be classified as cross-cultural. Additionally, even though many of the studies found provided useful insights, raised interesting questions, and generally contributed toward the advancement of the state of the art in its field, with few exceptions, no study specifically addressed equivalency issues central to measurement in cross-cultural research. It is this methodological issue of equivalency that is the focus of this article.

METHODOLOGICAL ISSUES

Methodological considerations are of the utmost importance to cross-cultural studies, because valid comparisons require cross-culturally equivalent research instruments, data collection procedures, research sites, and respondents. Ensuring equivalency is an essential element of cross-cultural studies and is necessary to avoid confounds and contaminating effects of various extraneous elements.

Cross-cultural research has some unique methodological idiosyncrasies that are not pertinent to intracultural research. One characteristic that typifies cross-cultural studies is their comparative nature, i.e., they involve a comparison across two or more separate cultures on a focal phenomenon. Any observed differences across cultures give rise to many alternative explanations. Particularly when results are different than expected (e.g., no statistical significance, factor analysis items do not load as expected, or reliability assessment is low), researchers may question whether results are true differences due to culture or merely measurement artifacts (Mullen, 1995).

Methodological considerations in carrying out cross-cultural research attempt to rule out alternative explanations for these differences and enhance the interpretability of results (van de Vijver & Leung, 1997). Clearly, the choice and appropriateness of the methodology can make a difference in any research endeavor. In cross-cultural research, however, one could go to the extreme of classifying this as one of the most critical decisions. In this section, we briefly review such cross-cultural methodological considerations. Specifically, this section will address equivalence (Hui & Triandis, 1985; Poortinga, 1989; Mullen, 1995) and bias (Poortinga & van de Vijver, 1987; van de Vijver & Leung, 1997; van de Vijver & Poortinga, 1997) as key methodological concerns inherent in cross-cultural research. Then, sampling, wording, and translation are discussed as important means of overcoming some identified biases.

Equivalence

Achieving cross-cultural equivalence is an essential prerequisite in ensuring valid cross-cultural comparisons. Equivalence cannot be assumed a priori. Each cross-cultural study needs to establish cross-cultural equivalence. As such, equivalence has been extensively discussed in cross-cultural research, albeit using different terms to describe the phenomenon (Mullen, 1995; Poortinga, 1989).

To alleviate confusion created by the multiplicity of concepts and terms used to describe different but somewhat overlapping aspects of equivalence, Hui and Triandis (1985) integrated prior research into a summary framework that consists of four levels of equivalence: conceptual/ functional equivalence, equivalence in construct operationalization, item equivalence, and scalar equivalence. Even though each level of equivalence is a prerequisite for the subsequent levels, in practice, the distinction between adjacent levels of equivalence often becomes blurry. Nonetheless, the objective in cross-cultural research is to achieve all four types of equivalence. Hui and Triandis' (1985) four levels of equivalence are discussed as follows:

1. *Conceptual/functional equivalence* is the first requirement for cross-cultural comparisons and refers to whether a given construct has similar meaning across cultures. Furthermore, to be functionally equivalent, the construct should be embedded in the same nomological network of antecedents, consequents, and correlates across cultures. For instance, workers from different cultures may rate "supervisor is considerate" as a very important characteristic; however, the meaning of "considerate" may vary considerably across cultures (Hoecklin, 1994).
2. *Equivalence in construct operationalization* refers to whether a construct is manifested and operationalized the same way across cultures. Not only should the construct be operationalized using the same procedure across cultures, but the operationalization should also be equally meaningful.
3. *Item equivalence* refers to whether identical instruments are used to measure the constructs across cultures. This is necessary if the cultures are to be numerically compared.
4. *Scalar equivalence* (or full score comparability; see van de Vijver and Leung, 1997) occurs if the instrument has achieved all prior levels of equivalence, and the construct is measured on the same metric. This implies that "a numerical value on the scale refers to same degree, intensity, or magnitude of the construct regardless of the population of which the respondent is a member" (Hui & Triandis, 1985, p. 135).

Bias: Sources, Detection, and Prevention

To achieve equivalence, one has to first identify and understand factors that may introduce biases in cross-cultural comparisons. Van de Vijner and Poortinga (1997) described three different types of biases: construct bias, method bias, and item bias:

1. *Construct bias* occurs when a construct measured is not equivalent across cultures both at a conceptual level and at an operational level. This can result from different definitions of the construct across cultures, lack of overlap in the behaviors associated with a construct [e.g., behaviors associated with being a good son or daughter (filial piety) vary across cultures], poor sampling of relevant behaviors to be represented by items on instruments, and incomplete coverage of the construct (van de Vijver & Leung, 1997). Construct bias can lead to lack of conceptual/functional equivalence and lack of equivalence in construct operationalization.
2. *Method bias* refers to bias in the scores on an instrument that can arise from characteristics of an instrument or its administration (van de Vijver & Leung, 1997), which results in subjects across cultures not responding to measurement scales in the same manner (Mullen, 1995). Method bias gives rise to concerns about the internal validity of the study. One source of method bias is sample inequivalency in terms of demographics, educational experience, organizational position, etc. Other method bias concerns relate to differential social desirability of responses (Ross & Mirowsky, 1984) and inconsistent scoring across populations (termed "selection-instrumentation effects" by Cook and Campbell, 1979, p. 53). For instance, on Likert scales, Koreans tend to avoid extremes and prefer to respond using the midpoints on the scales (Lee & Green, 1991), while Hispanics tend to choose extremes (Hui & Triandis, 1985). Differential scoring methods may also arise if respondents from a particular culture or country are not familiar with the type of instrument being used.
3. *Item bias* refers to measurement artifacts. These can arise from poor item translation, complex wording of items, or items inappropriate for a cultural context. Consequently, item bias is best prevented through careful attention to these issues. Like method bias, item bias can influence conceptual/functional equivalence, equivalence of operationalization, and item equivalence.

Table 1 presents a summary of how the three types of bias can be prevented or detected. The next section discusses three important methods of bias prevention: sampling, wording, and translation. This article concludes by presenting a set of cross-cultural methodological guidelines derived by a committee of international scholars.

Table 1. Types of bias, prevention, and detection

	Detection	Prevention
Construct bias (focus: constructs)	• Informants describe construct and associated behaviors • Factor analysis • Multidimensional scaling • Simultaneous confirmatory factor analysis in several populations • Comparison of correlation matrices • Nomological network	• Informants describe construct and associated behaviors in each culture
Method bias (focus: administration procedures)	• Repeated administration of instrument • Method triangulation • Monomethod-multitrait matrix	• Sampling (matching, statistical controls) • Identical physical conditions of administering the instrument • Unambiguous communication between interviewer and interviewee • Ensured familiarity with the stimuli used in the study
Item bias (focus: operationalization)	• Analysis of variance • Item response theory • Delta plots • Standardized *p*-difference • Mantel–Haenszel procedure • Alternating least squares optimal scaling • Multiple group LISREL	• Wording • Translation

Sampling

Sampling decisions in cross-cultural studies involve two distinct levels: sampling of cultures and sampling of subjects (van de Vijver & Leung, 1997). Sampling of cultures involves decisions associated with selecting the cultures to be compared in the study. Many studies involve a convenience sample of cultures, typically ones where the researcher has preestablished contacts. Even though this strategy reduces the considerable costs of conducting cross-cultural research, it may hinder interpretability of results, particularly when no differences are observed across cultures (van de Vijver & Leung, 1997). Systematic sampling of cultures, on the other hand, identifies cultures based on theoretical considerations. Typically, this involves selecting cultures that are at different points along a theoretical continuum, such as a cultural dimension. Random sampling of cultures involves selection of a large number of cultures randomly and allows for wider generalizability of results.

Most cross-cultural studies discussing sampling considerations, however, refer to sampling of subjects. Ensuring sample equivalency is an important methodological consideration in cross-cultural research, and it refers to the inclusion of subjects that are similar on demographic, educational, and socioeconomic characteristics. Sample equivalency can be achieved by either matching subjects across groups based on these background variables or statistically controlling for the differences by including such demographic variables as covariates in the cross-cultural comparisons (van de Vijver & Leung, 1997).

Wording and Translation

This is one of the key problems in cross-cultural methods, because in most cases, different cultures also have different languages. Even in cases when subjects from different countries are conversant with English, they may miss the nuances of the intended meanings in questionnaire items (e.g., British, Canadian, and American English all have unique terms). Researchers should ensure that measurement instruments keep the same meanings after translation. Moreover, a given latent construct should be measured by the same questionnaire items in different populations. In fact, researchers such as Irvine and Carrol (1980) made a convincing case for using factor-matching procedures to test for invariance of factor structures across groups before any quantitative analysis is performed.

To translate correctly, there is a need to translate to the target language. This needs to be performed by a native speaker of the target language. Then, the text must be back-translated to the original language, this time by a different native speaker of the original language. Brislin (1986) provided fairly complete guidelines for this process.

IMPLICATIONS

Judging by the issues described above, achieving cross-cultural equivalence is straightforward. However, it is also clear that many precautions can be taken to prevent construct, method, and item bias and thus increase the level of equivalence. These range from sampling, wording, and translation, to careful attention to administration procedures across cultures. A number of guidelines for cross-cultural research have been put forth by an international committee of scholars (Hambleton, 1994; van de Vijver & Hambleton, 1996). Even though the primary focus of these is on research on psychological and educational issues, these guidelines easily generalize to MIS research.

In addition to prevention, various statistical tools can assist in the detection of the various types of biases. In summary, similar patterns of functional relationships among variables need to be shown (Triandis, 1976). Moreover, multimethod measurement can help us to avoid the confound between the interaction of the method and groups studied and is unlikely to share the same statistical underlying assumptions, or even require strong conceptualization ability (Hui & Triandis, 1985). This idea is similar to the notions of multiple operationism and conceptual replication (Campbell & Fiske 1959). Hui and Triandis (1985) claimed that this may not be as difficult as one may think, as long as there is prior planning of research. As an example, Hui and Triandis (1985) mentioned that an instrument may be improved by proper translation techniques, and:

…then establish conceptual/functional equivalence as well as instrument equivalence by the nomological network method and by examination of internal structure congruence. After that, the response pattern method and regressional methods can be used to test item equivalence and scalar equivalence. (p. 149)

The major implication of methodological problems is complications in making valid inferences from cross-cultural data. Clearly, there are many problems with correctly inferring from data in a cross-cultural research project and attributing results to true cross-cultural differences. To do so, alternative explanations need to be ruled out. Establishing (and not merely assuming) the four levels of cross-cultural equivalence previously discussed in this article is a major step in this direction.

FUTURE DIRECTIONS AND CONCLUSION

Initial attempts at reviews of cross-cultural research in MIS (Evaristo, Karahanna, & Srite, 2000) show that, for the most part, MIS studies have refrained from testing theories across cultures, and when comparisons are made, they are often post hoc comparisons utilizing data from prior published studies in other countries. Clearly, this provides some insights into differences across cultures but suffers from a number of methodological shortcomings. In fact, the conclusions of Evaristo, Karahanna, and Srite (2000) were as follows:

In summary, we suggest that there are mainly three points where the MIS cross-cultural research is lacking: lack of theory base (testing or building); inclusion of culture as antecedents of constructs, and general improvement in methodologies used.

All three points are related, although to different extents, to methodological issues. The conclusion is that one critical issue that cross-cultural research in MIS needs to address before reaching the same level of sophistication and quality already attained by mainstream MIS research is to attend to methodological concerns. The current article is a step ahead in this direction and sets the stage for future research.

REFERENCES

Brislin, R. (1986). The wording and translation of research instruments. In *Field methods in cross-cultural research*. Lonner and Berry.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin, 56*, 81–105.

Cook, T., & Campbell, D. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin.

Evaristo, J. R., Karahanna, E., & Srite, M. (2000). *Cross-cultural research in MIS: A review*. Global Information Technology Management Conference, Memphis, TN.

Hambleton, R. K. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment, 10*, 229–244.

Hui, H., & Triandis, H. (1985). Measurement in cross-cultural psychology. *Journal of Cross-Cultural Psychology, 16*(2), 131–152.

Irvine, S., & Carrol, W. (1980). Testing and assessment across cultures: Issues in methodology and theory. In H. C. Triandis & W. Lonner (Eds.), *Handbook of cross-cultural psychology: Methodology* (pp. 181–244). Boston, MA: Allyn and Bacon.

Lee, C., & Green, R. (1991). Cross-cultural examination of the Fishbein Behavioral Intentions Model. *Journal of International Business Studies, 22*(2), 289–305.

Mullen, M. (1995). Diagnosing measurement equivalence in cross-national research. *Journal of International Business Studies,* (Third quarter), 573–596.

Poortinga, Y. H. (1989). Equivalence in cross-cultural data: An overview of basic issues. *International Journal of Psychology, 24*, 737–756.

Poortinga, Y. H., & van de Vijver, F. (1987). Explaining cross-cultural differences: Bias analysis and beyond. *Journal of Cross-Cultural Psychology, 18*(3), 259–282.

Ross, C. E., & Mirowsky, J. (1984). Socially desirable response and acquiescence in cross-cultural survey of mental health. *Journal of Health and Social Behavior, 25*, 189–197.

Srite, M., Straub, D., Loch, K., Evaristo, R., & Karahanna, E. (2003). Inquiry into definitions of culture in IT studies. In F. Tan (Ed.), *Advanced topics in global information management* (Vol. 2). Hershey, PA: Idea Group Publishing.

Triandis, H. (1976). Methodological problems of comparative research. *International Journal of Psychology, 11*(3), 155–159.

van de Vijver, F., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist, 1*, 89–99.

van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage.

van de Vijver, F., & Poortinga, Y. H. (1997). Towards an integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment, 13*(21–29).

KEY TERMS

Conceptual/Functional Equivalence: Refers to whether a given construct has similar meaning across cultures.

Construct Bias: Occurs when a construct measured is not equivalent across cultures both at a conceptual level and at an operational level.

Equivalence in Construct Operationalization: Refers to whether a construct is manifested and operationalized the same way across cultures.

Item Bias: Refers to measurement artifacts.

Item Equivalence: Refers to whether identical instruments are used to measure the constructs across cultures.

Method Bias: Refers to when subjects across cultures do not respond to measurement scales in the same manner. Bias in the scores on an instrument can arise due to characteristics of the instrument or its administration.

Multinational Corporation: A firm that has operations in multiple countries.

Scalar Equivalence: Refers to whether the construct is measured on the same metric across cultures. This occurs if the instrument has achieved all prior levels of equivalence, and the construct is measured on the same metric.

Transcultural Organization: A firm that operates across multiple cultures.

Cross–Culture Communication

Andrew Targowski
Western Michigan University, USA

Ali Metwalli
Western Michigan University, USA

INTRODUCTION

In this millennium, global organizations will increasingly focus on the critical value of the cross-cultural communication process, efficiency, competence and its cost of doing business. In order to successfully communicate cross-culturally, knowledge and understanding of cultural factors such as values, attitudes, beliefs and behaviors should be acquired. Because culture is a powerful force that strongly influences communication behavior, culture and communication are inseparably linked.

Worldwide, in 1983-2003, countries have experienced a phenomenal growth in international trade and foreign direct investment. Similarly, they have discovered the importance of cross-cultural communication. As a result, practitioners and scholars are paying attention to the fact that cultural dimensions influence management practices (Adler, 1983; Child, 1981; Hofstede, 1980; Laurent, 1983; Maruyama, 1984; Triandis, 1982-1983). In recent years, the empirical work in the cross-cultural arena has focused on the role of culture on employee behavior in communicating within business organizations (Tayeh, 1988). But current work on cross-cultural business communication has paid little attention to either (a) how to adapt these seminal works on general communication to the needs of intercultural business or (b) how to create new models more relevant to cross-cultural business exchanges (Limaye, 1991, p. 283). So far there is no adequate model that can explain the cross-cultural communication process and efficiency, let alone estimate the cost of doing business with other cultures worldwide.

The purpose of this article is to define the framework for a cross-cultural asymmetric communication process, efficiency and cost of doing business in the global economy. The research method is based on the architectural design of a cross-cultural communication process and system and their quantitative analysis. Their attributes are estimated in a normative way on a scale from 1 to 5, when 5 is the best value. The attributes for two selected cultures (Western-West and Egyptian) are estimated by expert opinions.

BACKGROUND

A Concept of Culture

A culture is a value-guided, continuous process of developing patterned human behavior within and across cultures and civilizations. Cultures do not satisfy needs; rather, they demand values. In turn, values in their broadest sense define the member of any culture's need for rationality, meaningfulness in emotional experience, richness of imagination and depth of faith (Laszlo, 1972). Human communication, therefore, is a vehicle for cultural dissemination on the one hand, while on the other hand is itself culture-driven.

Cultures are components of a civilization that guide their behavioral patterns. For example, Western civilization currently is composed of the following cultures:

- The Western-West, containing Western Europe and Northern America
- The Western Central, embracing Poland, the Czech Republic, Slovakia, Hungary, Estonia, Latvia, Lithuania, Croatia, and Slovene
- The Western-East, containing Greece and Israel
- The Western-Latin, composed of Latin America's states

There is some opinion that after the end of the Cold War in 1989/1991, that world politics is entering a new phase, and intellectuals have not hesitated to proliferate a vision of what it will be - the end of history (Fukuyama, 1989) or the clash of civilizations (Huntington, 1993). Huntington (1993) predicts that the fundamental source of conflict in this new world will not be primarily ideological or primarily economic. He perceives that the great divisions among humankind and their dominating source of conflict will be cultural.

THE CROSS-CULTURE COMMUNICATIONS PROCESS

We assume that doing business in the global economy depends mostly on the partners' ability to successfully communicate in a cross-cultural environment. People pursue and communicate many common aims, including the values of pure biological survival, social collaboration, creative expression, organizational adaptability or business undertakings. From such common values one can form a hierarchy of human cultural layers:

1. Biological culture layer
2. Personal culture layer
3. Group culture layer
4. Organization culture layer (ex. business enterprise)
5. Regional culture layer
6. National culture layer
7. Global culture layer (including supra-national, regional ones)

Those layers of cultures and the communication climate associated with them filter messages and intentions of business partners and determine the success of business undertakings. A model of the cross-cultural communication process is shown in Figure 1 (Targowski & Bowman, 1988).

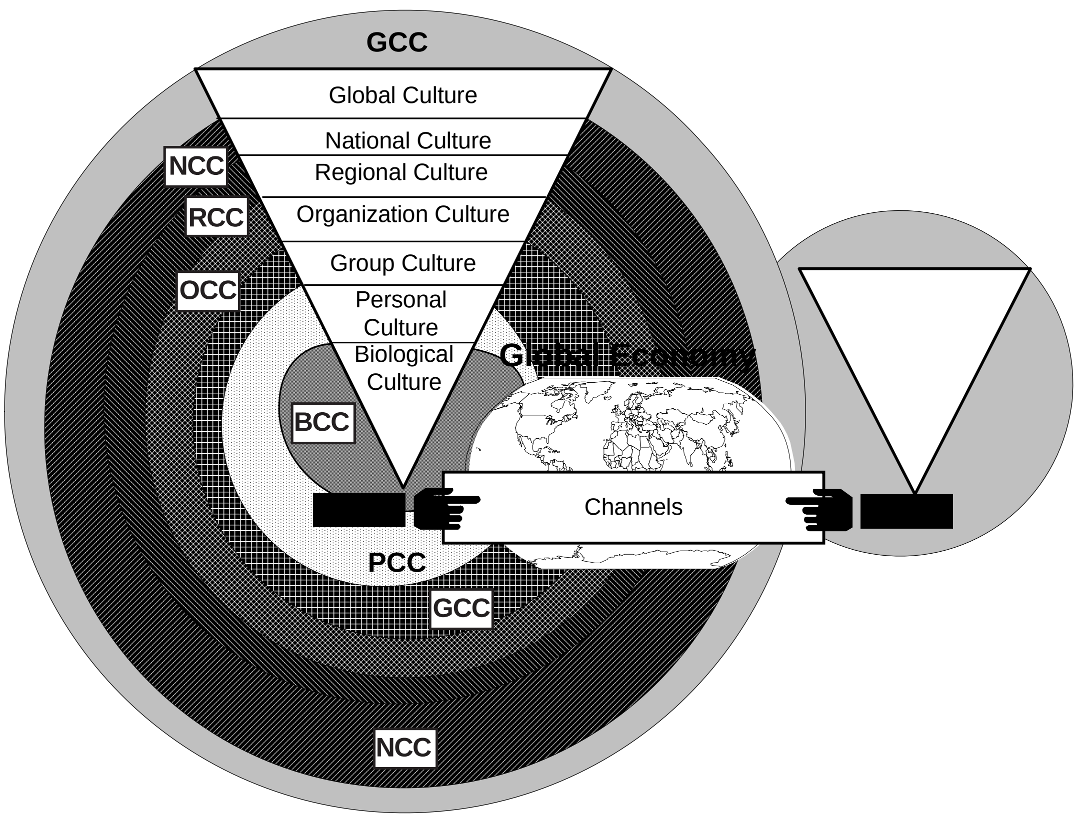
The model of the cross-culture communication process takes place at least between two partners who, in order to communicate, must send both a message and their intentions through several layers of cultures. For example, to communicate in the global economy between two different cultures each partner filters a message and intentions through seven layers of culture (biological, personal, group, organization, regional, national, and global). Of course, to be successful, such cross-culture communication must be based upon a good understanding of rules and practices that govern each layer of culture.

Let's define components of the cross-culture communication process in the global economy.

A). Global Economy

The global economy is largely understood in terms of worldwide economic and political convergence around liberal market principles and the increasing real-time integration of business, technological and financial systems (World Bank, 1997). The globalization process is supported by electronic communication that makes geography, borders, and time zones irrelevant to the way we

Figure 1. The culture layers and communication climates in the cross-culture communication process



conduct our business and personal lives. The "death of distance" will be the single most important economic force shaping all of society over the next half century (Cairncross, 1997). Friends, colleagues, and customers could easily be anywhere – around the corner or around the world – and new ways of communicating will effectively wipe out distance as a cost factor, indeed as a perceptible concept from our lives.

The growth of the global economy triggers the growth of a global information infrastructure (GII), which improves connections among organizations and individuals. Most people on earth will eventually have access to computer networks that are all switched, interactive, and broadband with capacity to receive TV-quality motion pictures. While the Internet will continue to exist in its form, it will also be integrated into other services, such as telephone and television (Targowski, 1996).

Although the Communication Revolution has increased connections among partners through technology, it simultaneously requires a better understanding of rules and practices of cross-cultural communication, regardless of applied media.

B). Culture Layers

The biological culture layer is the basic layer that provides common reactions based on the same physical needs which result from a common biological makeup. This layer is common for humankind in all civilizations and cultures.

The personal culture layer is the means with which the individual survives, operates and develops within the group, organization, region, nation, and globe. The essence of an individual's personal culture is the acceptance of underlying assumptions or "theories-in-use" (Argyris, 1976; Argyris & Schon, 1974). These assumptions are, for the individual, an undebatable understanding of reality, time and space (Schein, 1985). Each individual is additionally a member of various groups, organizations, regional, national, and global cultures.

The group culture layer is a managerial or an employee's tool used to accomplish an organizational task or to protect the interests of group members (formal or informal). A ggroup here may be defined as any collection of individuals united by a common relationship (i.e., work, profession or family).

The organization culture layer is a management tool that uses professional communication to influence organizational performance (Sypher et al., 1985) and that is created (or destroyed) by its leaders (Schein, 1985). A derivative of organization culture is a corporate culture that is a set of broad, tacitly understood rules (policies) informing employees how to behave under a variety of circumstances.

The regional culture layer contains commonalities based on the values of variables that individuals bear within a given region of a nation. Regional variables derive from two sources: 1) environmental influences (Borisoff & Victor, 1989; Farmerer & Richman, 1966; Terpstra & David, 1985) that have particular historical, political, economic, and social characteristics; and 2) traditions whose participants have similar ways of viewing space, time, things, and people (Schein, 1985; Weiss, 1988).

Regional culture may extend beyond national boundaries. In such cases, regional culture is *supranational*. For example, European culture encompasses dozens of nations with some shared values. Supranational culture can also be global (Featherstone, 1990). Regional culture may act as a subset of a particular national culture. In such cases, regional culture is *subnational*. For example, Brittany has a distinct regional culture within France but remains as a part of French national culture. Finally, regional culture may overlap with national culture. For example, Australian culture is at the same time both the culture of a nation state and that of a region (when compared to European or Latin American regions) that happens to be conterminous with a nation state.

The national culture layer is a set of common understandings, traditions, ways of thinking, feeling, behaving, and communication based on the same values of variables that influence communication through the nation. National culture is a learned behavior of its members'

Figure 2. The structure of a culture layer

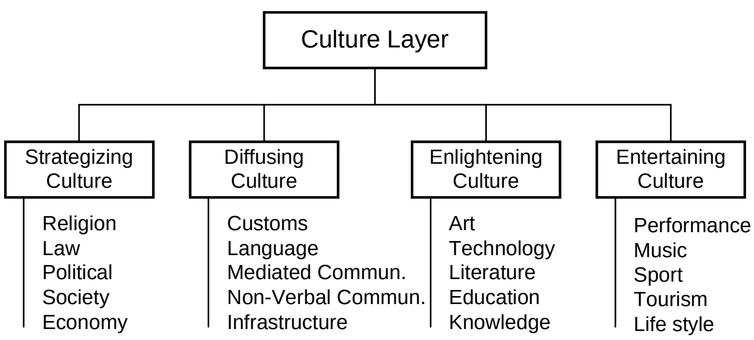


Table 1. Weighted attributes of communication climates (A)

CC Layer	(Territory) Space	Style	Time	Relationship	Frequency (Participation)	Tone (Friendliness)	Quality
Global CC	5-Personal 4-Semi-formal 3-Formal 2-Informal 1-Impersonal	5-Open 4-Semi-open 3-Semi-controlled 2-Controlled 1-Closed	5-Fixed 4-Semi-fixed 3-Inflexible 2-Semi-flexible 1-Flexible	5-V. Friendly 4-Friendly 3-Semi-friendly 2. Unfriendly 1. Hostile	5-High 4-Moderate 3-Low 2-Infrequent 1-Ad-hoc	5-Supporting 4-Guiding 3-Commanding 2-Manipulating 1.Critiquing	5-Transcommunication 4-Pseudocommunication 3-Paracommunication 2-Miscommunication 1-Metacommunication
National CC	5-Personal 4-Semi-formal 3-Formal 2-Informal 1-Impersonal	5-Democratic 4-Authoritarian 3-Dictatorial 2-Totalitarian 1-Chaotic	5-Fixed 4-Semi-fixed 3-Inflexible 2-Semi-flexible 1-Flexible	5-V. Friendly 4-Friendly 3-Semi-friendly 2. Unfriendly 1. Hostile	5-High 4-Moderate 3-Low 2-Infrequent 1-Ad hoc	5-Supporting 4-Guiding 3-Commanding 2-Manipulating 1.Critiquing	5-Transcommunication -Paracommunication -Pseudocommunication -Metacommunication -Miscommunication
Regional CC	5-Personal 4-Semi-formal 3-Formal 2-Informal 1-Impersonal	5-Laisser-faire 4-Coordinated 3-Partnership 2-Dominance 1-Hegemonic	5-Fixed 4-Semi-fixed 3-Inflexible 2-Semi-flexible 1-Flexible	5-V. Friendly 4-Friendly 3-Semi-friendly 2. Unfriendly 1. Hostile	5-High 4-Moderate 3-Low 2-Infrequent 1-Ad hoc	5-Supporting 4-Guiding 3-Commanding 2-Manipulating 1.Critiquing	5-Transcommunication 4-Pseudocommunication 3-Paracommunication 2-Miscommunication 1-Metacommunication
Organization CC	5-Networks 4-System structure 3-Matrix structure 2-Flat hierarchy 1-Tall hierarchy	5-Meritocratic 4-Plutocratic 3-Technocratic 2-Bureaucratic 1-Chaotic	5-Fixed 4-Semi-fixed 3-Inflexible 2-Semi-flexible 1-Flexible	5-V. Friendly 4-Friendly 3-Semi-friendly 2. Unfriendly 1. Hostile	5-High 4-Moderate 3-Low 2-Infrequent 1-Ad hoc	5-Supporting 4-Guiding 3-Commanding 2-Manipulating 1.Critiquing	5-Transcommunication 4-Pseudocommunication 3-Paracommunication 2-Miscommunication 1-Metacommunication
Group CC	5-Sub-networks 4-Networks 3-informal 2-Semi-Formal 1-Formal	5-Meritocratic 4-Plutocratic 3-Technocratic 2-Bureaucratic 1-Chaotic	5-Fixed 4-Semi-fixed 3-Inflexible 2-Semi-flexible 1-Flexible	5-V. Friendly 4-Friendly 3-Semi-friendly 2. Unfriendly 1. Hostile	5-High 4-Moderate 3-Low 2-Infrequent 1-Ad hoc	5-Supporting 4-Guiding 3-Commanding 2-Manipulating 1.Critiquing	5-Transcommunication 4-Pseudocommunication 3-Paracommunication 2-Miscommunication 1-Metacommunication
Personal CC	5-Home 4-Social 3-Semi-social 2-Anti-social 1-Work	5-Meritocratic 4-Plutocratic 3-Technocratic 2-Bureaucratic 1-Chaotic	5-Fixed 4-Semi-fixed 3-Inflexible 2-Semi-flexible 1-Flexible	5-V. Friendly 4-Friendly 3-Semi-friendly 2. Unfriendly 1. Hostile	5-High 4-Moderate 3-Low 2-Infrequent 1-Ad hoc	5-Supporting 4-Guiding 3-Commanding 2-Manipulating 1.Critiquing	5-Transcommunication 4-Pseudocommunication 3-Paracommunication 2-Miscommunication 1-Metacommunication

Source: Two experts'/authors' opinion.

Table 2. The comparison of richness of Western-West and Egyptian cultures

Culture Components	West – West Culture	Islamic (Egyptian) Culture
STRATEGIZING SUB-CULTURE 25	21	20
Religion	3	5
Law	5	5
Political	4	3
Society	4	4
Economy	5	3
DIFFUSING SUB-CULTURE 25	22	21
Customs	4	5
Language	5	5
Mediated communication	5	3
Non-verbal communication	3	5
Infrastructure	5	3
ENLIGHTENING SUB-CULTURE 25	24	16
Art	4	4
Technology	5	2
Literature	5	4
Education	5	3
Knowledge	5	3
ENTERTAINING SUB-CULTURE 25	24	18
Performance	4	4
Music	5	4
Sport	5	3
Tourism	5	4
Life style	5	3
CULTURE RICHNESS (R) $R_{max} = 100$	$R_{WW} = 91$	$R_E = 75$
CULTURE EFFICIENCY (η)	$\eta = 91$ %	$\eta = 75$ %

Source: Two experts'/authors' opinion

historical experience. For example one can recognize an American culture, Chinese culture, Egyptian culture or Polish culture, and so forth.

The global culture layer is the new emerging layer, triggered by the developments of the global economy. At this layer, partners from different cultures and civilizations deliberately apply the same patterns of behavior in order to achieve successful communication in business (political, social, and so forth) endeavors.

Figure 2 provides a structure of a culture layer.

C). Communication Climates

Communication climate (or "atmosphere") can be defined as a set of conditions that transform cultural behavior and information into desired (or undesired) states of a given entity (person, group, organization, region, nation, globe) through the communication process. Communication climate refers to the atmosphere or prevailing condition that exists within a specific entity. For example, within China there are at least three communication climates; the most open communication climate exists in Hong Kong, a less open communication climate is allowed in special economic zones such as in Shenzhen, and a closed communication climate is controlled in the remaining parts of China.

Communication climate consists of seven components: space (territory), style, time, desire for interaction (or relationship), frequency (participation), tone (receptivity or friendliness), and quality. Table 1 provides weighted attributes of communication climates for all layers of culture except the biological communication climate, which more or less is similar for all cultures.

D) Communication Channel

A communication channel is the vehicle or medium in which a message travels. Thus channels range from light waves for nonverbal cues to radio or computers as modes for transmitting sound and visual messages. For example, the effective operation of a highly complex weapon system might hinge on adherence to formal channels of communication, while effective performance in investment banking might rely on informal as well as formal communication channels. The formal channels of communication help create and maintain authority, as well as give authenticity to messages, but they also inhibit communication; indeed, they alienate members. In the global economy, informal channels such as the Internet prevail. They create a horizontal society that exchanges messages which otherwise would not have been created.

THE EFFICIENCY OF CROSS-CULTURAL COMMUNICATION

To apply these attributes in cross-cultural communication, let's take an example of communication in the global economy between Western-West and Egyptian cultures. To assess the influence of each attribute in the communication process, apply weights on the scale from 1 to 5, where 5 is the highest value of the attribute. Table 2 compares those attributes of two cultures, based on the authors' expertise within both cultures.

The cultural difference (D_c) in attribute weights of Western-West culture richness (R_{WW}) and Egyptian culture (R_E) is:

$$D_c = R_{WW} - R_E = 91\text{-}75 = 16 \text{ points} \qquad [1]$$

in favor of Western-West culture. It means that both cultures are not at the same developmental level, yet to succeed in communication between both partners in the global economy both partners must invest in their own abilities to cross-communicate successfully. We will see farther how to evaluate the cost of such investments.

The efficiency of Western-West culture in the global economy is h = 91%, which means that only nine times out of 100 a Western-West businessperson may miscommunicate because of a lack of cultural understanding of a business partner. On the other hand, the Egyptian culture's efficiency in the global economy is h = 75%, which means that an Egyptian businessperson may miscommunicate 25 times out of 100. Almost every fourth transaction will be miscommunicated because of a failure to account for cultural differences.

The role of the communication climate in cross-cultural communication is to facilitate that communication. For example, let us examine a case of communication between the Western-West and Egyptian cultures. Table 3 illustrates that comparison.

The communication climate difference (D_{CC}) between both cultures is:

$$D_{CC} = A_{WW} - A_E = 29\text{-}27 = 2 \text{ points} \qquad [2]$$

in favor of Western-West culture. This means that a Western-West businessperson entering into a transaction with an Egyptian partner should decrease the cultural difference (D_C) = 14 points by D_{CC} = 2, calculated previously. The modified culture difference of Western-West culture ($D_{MC/WW}$) after the adjustment by the communication climate difference is:

$$D_{MC/WW} = D_C - D_{CC} = 16 - 2 = 14 \text{ points} \qquad [3]$$

On the other hand, the Egyptian partner has to increase the Egyptian culture difference ($D_{MC/E}$) by the same coefficient:

$$D_{MC/E} = D_C - D_{CC} = 16 + 2 = 18 \text{ points} \qquad [4]$$

After the adjustments, the modified culture difference (D_{MC}) between Western-West culture and Egyptian culture has increased for Egyptian culture ($D_{MC/E}$) from 14 points to 16 points, and for Western-West culture ($D_{MC/WW}$) it has decreased from 14 to 12 points. The communication climate favors Western-West culture, while it disfavors the Egyptian culture. In other words, a businessperson from a Western-West culture has a communication advantage in the global economy, while a businessperson from the Egyptian culture has to work harder at the communication effort in order to succeed in the global economy.

Table 3. The comparison of communication climate attributes (A) of Western-West (A_{WW}) and Egyptian cultures (A_E) in the global economy settings

Global Comm. Climate	Western-West Culture	Egyptian Culture
Territory (Space)	3-Formal	5-Personal
Style	5-Open	4-Semi-open
Time	5-Fixed	2-Semi-flexible
Relationship	4-Friendly	5-V. Friendly
Frequency (Participation)	5-High	3-High
Tone (Friendliness)	4-Guiding	2-Commanding
Quality	3-Paracommunication	4-Paracommunication
TOTAL POINTS	A_{WW} = 29	A_E = 27

Source: Two experts'/authors' opinion

The ability of a Western-West culture's businessperson (B_{WW}) to deal with a business partner from Egyptian culture is:

$$B_{WW} = R_{WW} : D_{MC/WW} = 91 : 14 = 6.5 \qquad [5]$$

The ability of the Egyptian culture's business partner (B_E) to deal with a business partner from Western-West culture is:

$$B_E = R_E : D_{MC/E} = 75 : 18 = 4.2 \qquad [6]$$

The culture's ability reflects how a given culture's strength can overcome cultural differences. The comparison of both cultures' abilities reflects a partner's cultural strength at the business table.

It also reflects his/her competitive advantage (V). Competitive advantage of a Western-West partner over an Egyptian partner can be computed in the following manner:

$$V_{WW} = B_{WW} : B_E = 6.5 : 4.2 = 1.6 \text{ or } 160\% \qquad [7]$$

In our example, B_{WW} is 1.6 (160%) times stronger than B_E. In common language, this comparison means that in a global economy of bilateral relationships, an American businessperson has almost twice the strength in overcoming cultural differences than does the Egyptian partner. If such knowledge is known to either partner, it can bring competitive advantage to him/her.

THE CULTURE COST FACTOR IN THE GLOBAL ECONOMY

A business entering the global economy is aware that it has to improve the understanding of foreign markets' dynamics and practices. Usually the new entry to the foreign market is associated with two types of cost:

- The explicit cost (C_E) can be anticipated, planned and quantifiable (transportation, building purchase or rental, interpreter's salary and all other overhead costs) in terms of its financial impact on doing business in a specific culture.
- The implicit cost (C_I) of cultural differences is intuitively understandable, but it is usually very difficult to predict or evaluate its structure and range. So, to guarantee success, it is very critical for a global firm to identify, examine and project the implicit costs associated with entering a foreign market.

The following conceptual framework of the implicit cost of the culture factor (C_I) will provide a definition and range of that type of cost:

$$C_I = f(GNP_C, R_C, D_{MC}, N_D, N_W) \qquad [8]$$

Where:

GNP – Gross National Product per capita of a given culture in terms of purchasing power parity (*ppp*);
R_C - Richness of a given culture;
D_{MC} - Adjusted modified culture difference between involved cultures;
N_D - Number of working days in another culture;
N_W - Number of workers in another culture.

Based on the variables' relationships in formula [8] one can define a formula for the culture factor cost of the culture that has a positive culture difference ($+D_{MC}$):

$$C_{I(+D)} = [(\$GNP_{C(+D)} - \$GNP_{C(-D)}) : (R_{C(+D)} + D_{MC})] \times (N_D \times N_W) \qquad [9]$$

In the case of the richer culture (formula [9]), the culture difference (D_{MC}) is in favor of that culture ($R_{C(+D)}$); therefore, both variables are added to decrease the cost of overcoming the culture difference.

This formula provides a balance between economic means, expressed in a GNP level and culture richness (R). In other words, low-rich cultures (measured by a low R) with high GNPs will not communicate in the global economy at the low cost. By analogy, they may remind their associates of the behavior of a *nouveau riche*, whose manners are sometimes questionable. New millionaires from countries that are being transformed from a central planning to a market economy sometimes invade the French Riviera and behave there like an elephant in a china shop. Due to a limited practice in the market economy and old attitudes, very often those *nouveau riche* prefer quick deals rather than long-term business collaboration. They are aware that their low-rich culture does not generate enough confidence in partners from a rich culture to establish long-term business relations. An additional factor plays a role too. A fresh business culture in some of those countries, very often driven by mafias, does not motivate those businesspersons to long-term commitments. As a result of it, they prefer to invest abroad rather than in their own countries. In effect, their business cultures do not develop. They misdevelop.

In the case of the less rich culture (formula [10], the culture difference (D_{MC}) is not in favor of that culture ($R_{C(-D)}$); therefore D_{MC} is subtracted from $R_{C(+D)}$ to increase the cost of overcoming the culture difference.

$$C_{I(-D)} = [(\$GNP_{C(-D)} - \$GNP_{C(-D)}) : (R_{C(-D)} - D_{MC})] \times (N_D \times N_W) \quad [10]$$

The application of these formulas to the example of cross-cultural communication between a Western-West culture [9] and an Egyptian culture [10] for the business duration = 30 days and the involvement of one worker provides the following results [GNP according to World Bank (1998/99)]:

$$C_{WW} = [(\$28{,}740 - \$2{,}940) : (91 + 14)] \times [30 \times 1] = \$7{,}400 \quad [11]$$

$$C_E = [(\$28{,}740 - \$2{,}940) : (75 - 18)] \times [30 \times 1] = \$13{,}600 \quad [12]$$

The richer culture (Western-West), in order to successfully communicate a business plan with the Egyptian culture, must invest only $7,400 to overcome culture differences. However, the less rich culture should spend 1.8 times more than the rich culture to overcome culture differences in 20 culture components (Table 2) and seven culture communication climate attributes (Table 3).

This example only confirms the old truth, that the comprehensive development of a rich culture takes a long time and requires many means. The paradox of this example is that the culture of Egypt is about 6,000 years old, while the Western-West culture is 1,200 years old, but has developed more comprehensively with much, much bigger means and this is reflected in the GNP per capita of both cultures. Apparently 1+ millennium is enough time to enrich the culture of Western civilization.

The cost of overcoming culture differences should focus on learning through education, training, and practice with all culture categories that are disadvantageous. To see the scope of that effort one must analyze the comparison of those two cultures, provided in Table 2 and Table 3.

FUTURE TRENDS

In the global economy cross-cultural communication becomes the basic skills of business communication. The more open the world will be, the more cross-cultural communication will be required not only in business but also in politics, where "dialog among civilizations" is the only solution in transforming from conflicts into peace.

CONCLUSION

This article has outlined and defined a framework for an asymmetric cross-cultural communication process, its efficiency and the cost of doing business in the global economy. A universal system design was developed not only to compare and quantify cultural efficiency and the attributes of communication climates through seven cultural layers (biological, personal, group, organization, regional, national and global), but also to explain the cross-cultural communication process and quantify the cultural cost in the global economy. This culture-specific design will help reduce miscommunication between partners across cultures and raise the awareness of the differences in the level of efficiency and cost in the communication process, behavior and practices between Western and non-Western cultural patterns.

Based on the presented framework, one can state that in order to be successful in cross-cultural communication in the global economy, engaged parties should be aware of the following five efficiency and cost rules:

- Culture Richness Rule I: A party from a less-rich culture (lower R) will more frequently miscommunicate with a party from a richer culture (higher R). (Formula 1).
- Communication Climate Rule II: A party from a "warmer" communication climate (higher A) will be in the advantageous position with a party from a "cooler" communication climate (lower A). (Formulas 3 and 4).
- Communication Ability Rule III: When large differences exist between cultures, the party from the richer culture has the best chance to communicate its own message (higher R). (Formulas 5 and 6).
- Communication Competitive Advantage Rule IV: The difference in communication ability gives a measurable competitive advantage to the more skillful communicator. (Formula 7).
- Communication Cost Rule V: A party with a higher GNP and a richer culture communicates in the cross-culture setting at lower costs than a party with the opposite attributes. (Formulas 8, 9, and 10).

The presented framework of asymmetric cross-cultural communication in the global economy provides pragmatic tools about how to define a communication strategy, train representatives and conduct business talks in order to achieve success.

C

REFERENCES

Adler, N.J. (1983). Cross-cultural management research: The ostrich and the trend. *Academy of Management Review, 8,* 226-232.

Argyris, C. (1976). *Increasing leadership effectiveness.* New York: Wiley-Interscience.

Argyris, C., & Schon, D.A. (1974). *Theory in practice: Increasing professional effectiveness.* San Francisco: Jossey-Bass.

Beamer, L. (1992). Learning intercultural communication competence. *Journal of Business Communication, 29*(3), 285-303.

Berlo, D.K. (1960). *The process of communication.* New York: Holt, Rinehart & Winston.

Borisoff, D., & Victor, D.A. (1989). *Conflict management: A communication skills approach.* Englewood Cliffs, NJ: Prentice Hall.

Cairncross, F. (1997). *The death of distance.* Boston: Harvard Business School Press.

Camerer, C., & Vepsalainen, A. (1988, Fall). The economic efficiency of corporate culture. *Journal of Business Communication, 24*(4), 21-34.

Chaney, L., & Martin, J. (1995). *Intercultural business communication.* Englewood Cliffs, NJ: Prentice Hall.

Charlten, A. (1992). Breaking cultural barriers. *Quality Progress, 25*(9), 47-49.

Child, J. (1981). Culture, contingency and capitalism in the cross-national study of organizations. In L.L. Cummings and B.M. Staw (Eds.), *Research in organizational behavior* (vol. 3, pp. 303-356). Greenwich, CT: JAI Press.

Eiler, M.A., & Victor, D. (1988). Genre and function in the Italian and U.S. business letter. *Proceedings of the Sixth Annual Conference on Languages and Communications for World Business and the Professions,* Ann Arbor.

Farmerer, R.N., & Richman, B.M. (1966). *International business: An operational theory.* Homewood, IL: Richard D. Irwin.

Featherstone, M. (1990). *Global culture, nationalism, globalization and modernity.* Newsbury Park, CA: SAGE Publications, Ltd.

Fisher, G. (1988). *Mindsets: The role of culture and perception in international relations.* Yarmouth, ME: Intercultural Press.

Fukuyama, F. (1989, Summer). The end of history. *The National Interest.*

Granner, B. (1980). Cross-cultural adaptation in international business. *Journal of Contemporary Business, 9*(3), 101-108.

Halpern, J.W. (1983). Business communication in China: A second perspective. *The Journal of Business Communication, 20,* 43-55.

Haworth, D.A., & Savage, G.T. (1989). A channel-ratio model of intercultural communication. *The Journal of Business Communication, 26,* 231-254.

Held, D., Goldblatt, D., McGrew, A., & Perraton, J. (1997). The globalization of economic activity. *New Political Economy, 2*(2), 257-277.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values.* Beverly Hills, CA: Sage.

Howard, E. (1998). Can business cross the cultural divide? *Communication World, 15*(9), 1-7.

Huntington, S. (1993, Summer). The clash of civilizations? *Foreign Affairs.*

Joinson, C. (1995). Cultural sensitivity adds up to good business sense. *HR Magazine,* 82-85.

Laszlo, E. (1972). *The system view of the world.* New York: George Braziller.

Laurent, A. (1983). The culture diversity of Western conceptions of management. *International Studies of Management and Organization, 13*(1-2), 75-96.

Limaye, M., & Victor, D. (1991). Cross-cultural business communication research: State of the art and hypotheses for the 1990's. *Journal of Business Communication, 28*(3), 277-299.

Lindsley, S. (1999). A layered model of problematic intercultural communication in US-owned maquiladoras in Mexico. *Communication Monographs, 66*(6).

Maruyama, M. (1984). Alternative concepts of management: Insights from Asia and Africa. *Asia Pacific Journal of Management, 1*(2), 100-111.

Moran, R., & Richard, D. (1991). Preparing technical professionals for cross-cultural interactions. *Journal of European Industrial Training, 15*(3), 17-21.

Perkins, E.A., &. Stout, J.V. (1987). Group dynamics: Communication responsibilities of managers. *The Proceedings of The Association For Business Communication,* Atlanta, GA.

Richard, L. (1990). *How do you develop Pan-European communication? 7*(8), 1-6.

Rodrik, D. (1997, Summer). Sense and nonsense in the globalization debate. *Foreign Policy*.

Schein, E.H. (1985). *Organizational culture and leadership.* San Francisco: Jossey-Bass.

Scott, J. (1999). Developing cultural fluency: The goal of international business communication instruction in the 21st century. *Journal of Education for Business, 74*(3), 140-143.

Sussman, L., & Johnson, D. (1993). The interpreted executive: Theory, models, and implications. *Journal of Business Communication, 30*(4), 415-434.

Sypher, B.D., Applegate, J.L., & Sypher, H.E. (1985). Culture and communication in organizational context. In W.B. Gudykunst, L.P. Stewart & S. Ting-Toomey (Eds.), *Communication, culture, and organizational process.* Beverly Hills, CA: SAGE Publications.

Targowski, A. (1996). *Global information infrastructure.* Hershey, PA: Idea Group Publishing.

Targowski, A. (2000, June 2-3). The civilization index. *The Proceedings of the East-West Conference*, Western Michigan University, Kalamazoo, MI.

Targowski, A., & Bowman, J. (1988, Winter). The layer-based pragmatic model of the communication process. *Journal of Business Communication, 25*(1), 5-24.

Tayeb, M.H. (1988). *Organizations and national culture: A comparative analysis.* London: Sage.

Terpstra, V., & Dvid, K. (1985). *The cultural environment of business* (2nd ed.) Cincinnati, OH: South-Western Publishing Company.

Triandis, H.C. (1982-83). Dimensions of cultural variations as parameters of organizational theories. *International Studies of Management and Organization, 12*(4), 139-169.

Varner, I., & Beamer, L. (1995). *Intercultural communication in the global workplace.* Chicago: Irwin.

Varner, I.I. (1988). A comparison of American and French business communication. *The Journal of Business Communication, 25*(4), 55-65.

Victor, D.A. (1987). Franco-American business communication practices: A survey. *World Communication, 16*(2), 158-175.

Victor, D.A. (in press). *International business communication.* New York: Harper Collins Press.

Victor, D.A., & Danak, J. (1990). *Genre and function in the U.S. and Indian English-language business letter: A survey.* Paper presented at the Conference on Language and Communication for World Business and the Professions, Ypsilanti, MI.

Weiss, J.W. (1988). *Regional cultures, managerial behavior and entrepreneurship: An international perspective.* Westport, CT: Greenwood Press.

Williams, M. (1991). Will diversity = equality for multicultural communicators? *Communicational World, 8*(3), 27-30.

World Bank. (1997). *The state in changing world: The world development report.* Washington, D.C.

World Bank. (1998-1999). *Knowledge for development: The world development report.* Washington, D.C.

Zong, B., & Hildebrandt, H.W. (1983). Business communication in the People's Republic of China. *The Journal of Business Communication*, *20*, 25-33.

KEY TERMS

Asymmetric Communication: A process in which both communicating parties are at different levels of economic and social development.

Communication Climate: Or "atmosphere," can be defined as a set of conditions that transform cultural behavior and information into desired (or undesired) states of a given entity (person, group, organization, region, nation, globe) through the communication process.

Cross-Culture Communication: A process of communicating among different cultures.

Culture: A value-guided, continuous process of developing patterned human behavior within and across cultures and civilizations.

Global Culture: The new emerging layer of cultures, triggered by the developments of the global economy. At this layer, partners from different cultures and civilizations deliberately apply the same patterns of behavior in order to achieve successful communication in business (political, social, and so forth) endeavors.

Global Economy: Largely understood in terms of worldwide economic and political convergence around liberal market principles and the increasing real-time integration of business, technological and financial systems.

Culture and Anonymity in GSS Meetings

Moez Limayem
City University of Hong Kong, Hong Kong

INTRODUCTION

Group support systems (GSSs) are an increasingly popular means of aiding decision making in a variety of organizational settings, by combining the computer, communication, and decision technologies to improve the decision-making process (Briggs, Nunamaker, & Sprague, 1998). Such technologies make use of anonymity as a key tool to improve the quality of decisions (Nunamaker et al., 1991; Pinsonneault & Heppel, 1997; Postmes & Lea, 2000).

Managers spend a considerable part of their work in meetings and participating in group decisions. Anonymity is generally believed to create an environment that improves group participation, communication, and the objective evaluation of ideas, enhancing the productivity of groups and their decision-making processes. Anonymity, as a distinct aspect of GSS, was expected to increase productivity by reducing the level of social or production blocking, increasing the number of interpersonal exchanges, and reducing the probability of any one member dominating the meeting (Newby, Soutar, & Watson, 2003). Barreto and Ellemers (2002) manipulated two aspects of anonymity separately: visibility of respondents (i.e., participants could or could not see who the other group members were) and visibility of responses (participants could or could not see the responses given by other group members). Results show that when group identification is low, anonymity manipulations affect group members' efforts.

A number of empirical findings have suggested that the use of anonymity and process structure in electronic brainstorming (EBS) generally promote a positive effect on the number of ideas generated (Jessup, Connolly, & Galegher, 1990; Gallupe, Bastianutti, & Cooper, 1991), and quality of ideas achieved in decision making (Zigurs & Buckland, 1998). However, the anonymity function inherent in multiworkstation GSSs has been found to heighten conflict, as members tend to communicate more aggressively because they tend to be more critical (Connolly, Jessup, & Valacich, 1990; Jessup, Connolly, & Tansik, 1990; Valacich et al., 1992); to have no effects on inhibition (Valacich, Dennis, & Connoly, 1994; Valacich et al., 1992); to increase group polarization (Sia, Tan, & Wei, 2002); and to have no effects on group performance (Valacich, Dennis, & Connoly, 1994). Other empirical findings show that, in terms of effectiveness, nominal brainstorming may be equal to (Gallupe, Cooper, & Bastianutti, 1991; Cooper et al., 1998; Barki & Pinsonneault, 2001) or sometimes less than (Valacich, Dennis, & Connoly, 1994; Dennis & Valacich, 1993) electronic brainstorming, indicating that at least as far as laboratory studies are concerned, empirical investigations have proved inconclusive.

BACKGROUND

Ferraro (1998) provided a succinct definition of culture as follows: "Culture is everything that people have, think, and do as members of their society."

Culture has been defined as the collective programming of the mind that distinguishes the members of one group or category of people from another (Hofstede, 1991; Tan, Watson, & Wei, 1995). Culture involves the beliefs, value systems, and norms of a given organization or society, and can exist at national, regional, and corporate levels. In fact, even information systems theories and research are heavily influenced by the culture in which they developed, and a theory grounded in one culture may not be applicable in other countries (Tan, Watson, & Wei, 1995; Triandis, 1987). The theories explaining the effects of GSS have come mainly from a North American perspective and may need to be adjusted for appropriate explanation of the same phenomena in different contexts. Therefore, in order to incorporate a global dimension, theories and models that attempt to explain the effectiveness of technology will need to take into account the cultural background of the group being examined.

Hofstede (1991) identified five dimensions of culture based on his IBM study in 72 different countries:

- *Uncertainty avoidance* is the degree to which a society feels threatened by uncertain and ambiguous situations, which leads members of the society to support beliefs promising certainty and to maintain institutions protecting conformity.
- *Masculinity* refers to a preference for achievement, heroism, assertiveness, and material success; as opposed to *femininity*, which implies a preference for relationships, modesty, caring for the weak, and quality of life.
- *Long-term orientation* stands for the fostering of virtues oriented toward future rewards, in particu-

lar, perseverance and thrift. Its opposite pole, short-term orientation, stands for the fostering of virtues related to the past and present, in particular, respect for tradition, preservation of "face," and fulfillment of social obligation.

- *Power distance* is the extent to which society accepts the fact that power in institutions and organizations is unevenly distributed, and how this fact of inequality is dealt with. It is found that individuals in societies with low power-distance cultures (e.g., United States) may be more inclined to adopt technologies that reduce power distance (Reinig & Mejias, 2003).
- *Individualism* refers to a preference for a loose-knit social framework in society in which individuals are only supposed to take care of themselves and their immediate families. This is opposed to *collectivism*, which implies a preference for a tightly knit social framework in which individuals can expect their relatives and clan to protect them in exchange for loyalty. The people of collectivistic-culture societies (e.g., Hong Kong) may use technologies to sustain group harmony and agreement.

It is interesting that power distance and individualism are found to be inversely related (Hofstede, 1991; Kim et al., 1994; Triandis, 1995). Many Western countries, such as the United States, Great Britain, and Australia, have been described as individualistic, low power-distance cultures, while many Asian countries, such as Hong Kong, Singapore, and China, have been described as collectivistic, high power-distance cultures (Hofstede, 1991).

CULTURE IN GSS STUDIES

Culture was not specifically considered as an important dimension in the early studies of GSS. However, with globalization, it is becoming increasingly important to adapt this tool to the cultural background of the organization or group that intends to use it effectively. These dimensions have been investigated in cross-culture GSS studies (such as those of Robichaux & Cooper, 1998; Tan et al., 1998; Tung & Quaddus, 2002; Watson, Ho, & Raman, 1994). Among the five dimensions, power-distance and individualism have been shown to have impacts on group behavior and group outcomes (Tan et al., 1998; Watson, Ho, & Raman, 1994). This occurs because the anonymity and simultaneous input features of GSS support low power-distance and individualistic cultural norms of desirable group behavior (Watson, Ho, & Raman, 1994).

Watson, Ho, and Raman (1994) later provided empirical support for the inclusion of culture as a dimension of GSS to add to DeSanctis' and Gallupe's (1987) dimensions of group size, member proximity, and task type. Their study examined U.S. and Singaporean cultures using GSS, and the findings suggested that Singaporean groups tended to have a higher pre-meeting consensus and less change in consensus than the U.S. group. This may be explained with reference to the collectivist nature of Singaporean culture, as collectivists have a tendency toward group consensus (Mejias et al., 1997).

Tan, Watson, and Wei (1995) suggested ways that different cultures can be studied with other important variables, such as task type and group size. The study focused on finding a way to examine the robustness of previous and current GSS research across different cultures and to add a cultural perspective to existing GSS knowledge. Hofstede's dimension of power distance was examined in relation to GSS, and the possible impacts of GSS intervention in both high and low power-distance countries were explored.

In studies examining only Singaporean groups (Tan, Watson, & Wei, 1995), the use of GSS resulted in a decreased impact of status and normative influences on decision making. These findings showed that change in consensus was greater in U.S. groups than in Singaporean groups, and influence was more equal in Singaporean groups than in U.S. groups. The higher power-distance of Singaporean groups may explain the differences between these two meeting outcomes, and the study supports the proposition that GSS can overcome the effect of high power-distance on group meetings.

A study comparing North American and Mexican groups participating in GSS sessions showed differences in terms of perception of consensus and satisfaction levels of group members (Mejias et al., 1997). U.S. and Mexican groups were also studied for GSS' effects on participation equity, with Mexican groups reporting higher participation equity levels than U.S. GSS groups (Mejias et al., 1997). It was suggested that high power-distance cultures benefit from GSS, and that these findings indicate that culture has a significant bearing on crucial aspects of GSS meeting outcomes.

Limayem, Khalifa, and Coombes (2003) conducted a study to explain the different effects of anonymity on the behavior of Hong Kong and Canadian groups during GSS sessions. In the Hong Kong Chinese culture, group interactions tend to emphasize harmony, conformance, and reciprocal respect rather than openness and spontaneity. However, the Canadian group's culture, which frequently exhibits openness and spontaneity, will usually allow individuals to deviate from the norm. Anonymity was found to have more significant positive effects for Hong

Kong groups. With anonymity, the performance of the Hong Kong group improved significantly in terms of number of contributions, quality of contributions, and perceived level of participation. No significant differences in the performance measures were found for the Canadian groups, except for the quality of contributions, which deteriorated with anonymity. A qualitative analysis of this negative effect revealed social loafing and lack of accountability as possible causes.

In sum, studies into the use of GSS by different cultures have indicated that there are differences between different cultures using GSS, and that cultural dimensions, such as those proposed by Hofstede (1991), have some relevance in explaining these differences. However, there is still uncertainty as to the specific impacts of culture on the performance of groups in anonymous GSS sessions, and, therefore, more must be done to clearly understand how different cultures respond to anonymity.

FUTURE TRENDS

As we move into the world of virtual organizations and electronic commerce, the use of GSS by groups of different cultures becomes an irreversible trend. For practitioners or users of GSS, groupware applications, or other electronic communication systems, the implications are important. We cannot simply think of anonymity as a concept that is good or bad. The use of anonymity should depend on the culture in which it is applied. For example, it is probably not a good idea to use anonymity in some cases where the culture of the group does not emphasize status hierarchies, conformance, mutual obligation, and reciprocity. In these situations, anonymity could even lead to negative outcomes, such as social loafing due to the reduction in motivation and effort that occurs when individuals work in anonymous groups. Conversely, it may be beneficial to use anonymity for GSS-supported groups with cultures that normally exhibit higher levels of conformance pressure and evaluation apprehension.

In a broader sense, GSS and groupware designers and developers should pay special attention to the implementation of anonymity features. For example, they could make it easier for users to turn these features on and off to accommodate the cultures of the groups using the systems. Finally, facilitators should remember that studies suggest that culture influences participation in the GSS environment (Tung & Quaddus, 2002). Therefore, facilitators should study the culture of the group using the technology before blindly using anonymity to generate or evaluate ideas.

CONCLUSION

Culture is clearly an important factor affecting a group's response to anonymity in the GSS context. Therefore, culture's influence on group structure and evaluation apprehension is important for designers, facilitators, and users of GSS to consider. Considering the lack of research on the effects of culture on GSS, further research in this field would appear to be worthwhile. This may be conducted on other interactions associated with cultures using the anonymous function of GSS. This research also has fascinating implications for technology's effect on other cultures. An interesting line of research for the future will be to isolate the relative impact of anonymity with different cultures engaged in different tasks and situations. The knowledge gained from this and other continuing studies will assist in the effective application of GSS in increasingly diverse and global contexts.

ACKNOWLEDGEMENT

The work described in this article was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project no. 9040564).

REFERENCES

Barki, H., & Pinsonneault, A. (2001). Small group brainstorming and idea quality—Is electronic brainstorming the most effective approach? *Small Group Research, 32*(2), 158–205.

Barreto, M., & Ellemers, N. (2002). The impact of anonymity and group identification on progroup behavior in computer-mediated groups. *Small Group Research, 33*(5), 590–610.

Briggs, R. O., Nunamaker, J. F., & Sprague, R. H. (1998). 1001 unanswered research questions in GSS. *Journal of Management Information Systems, 14*(3), 3–21.

Connolly, T., Jessup, L. M., & Valacich, J. S. (1990). Effects of anonymity and evaluative tone on idea deneration. *Management Science, 36*(6), 689–704.

Cooper, W. H., Gallupe, R. B., Pollard, S., & Cadsby, J. (1998). Some liberating effects of anonymous electronic brainstorming. *Small Group Research, 29*(2), 147–178.

Dennis, A. R., & Valacich, J. S. (1993). Computer brainstorms: More heads are better than one. *Journal of Applied Psychology, 78*(4), 531–538.

DeSanctis, G. L., & Gallupe, R. B. (1987). A foundation for the study of group decision support systems. *Management Science, 33*(5), 589–609.

Gallupe, R. B, Bastianutti, L., & Cooper, W. H. (1991). Unblocking brainstorms. *Journal of Applied Psychology, 76*(1), 137–142.

Ferraro, G. P. (1998). *The cultural dimensions of international business*. Upper Saddle River, NJ: Prentice Hall.

Hofstede, G. (1991). *Cultures and organizations; Software of the mind*. New York: McGraw-Hill.

Jessup, L. M., Connolly, T., & Galegher, J. (1990). The effects of anonymity on GDSS group process with an idea-generating task. *MIS Quarterly, 14*(3), 312–321.

Jessup, L. M., Connolly, T., & Tansik, D. A. (1990). Toward a theory of automated group work: The deindividuating effects of anonymity. *Small Group Research, 21*(3), 333–348.

Kim, U., Triandis, H. C., Kagitcibasi, C., Choi, S. C., & Yoon, G. (1994). *Individualism and collectivism: Theory, methods and applications*. Thousand Oaks, CA: Sage Publications.

Limayem, M., Khalifa, M., & Coombes, J. M. (2003). Culture and anonymity in GSS meetings. In G. Ditsa (Ed.), *Information management: Support systems and multimedia technology* (pp. 150–161). Hershey, PA: Idea Group Publishing.

Mejias, R. J., Shepherd, M. M., Vogel, D. R., & Lasaneo, L. (1997). Consensus and perceived satisfaction levels: A cross cultural comparison of GSS and non-GSS outcomes within and between the United States and Mexico. *Journal of Management Information Systems, 13*(3), 137–161.

Newby, R. Soutar, G., & Watson, J. (2003). Comparing traditional focus groups with a group support systems (GSS) approach for use in SME research. *International Small Business Journal, 21*(4), 421–433.

Nunamaker, J. F., Dennis, A. R., Valacich, J. S., Vogel, D. R., & George, J. F. (1991). Electronic meeting systems to support group work. *Communications of the ACM, 34*(7), 40–61.

Pinsonneault, A., & Heppel, N. (1997). Anonymity in group support systems research: A new conceptualization, measure, and contingency framework. *Journal of Management Information Systems, 14*(3), 89–108.

Postmes, T., & Lea, M. (2000). Social processes and group decision making: Anonymity in group decision support systems. *Ergonomics, 43*(8), 1252–1274.

Reinig, B. A., & Mejias, R. J. (2003). An investigation of the influence of national culture and group support systems on group processes and outcomes. In *Proceedings of the 35th Hawaii International Conference on System Sciences (HICSS'03)* (pp. 6–9), January 2003, Big Island, Hawaii.

Robichaux, B. P., & Cooper, R. B. (1998). GSS participation: A cultural examination. *Information and Management, 33*(6), 287–300.

Sia, C. -L., Tan, B. C. Y., & Wei, K. -K. (2002). Group polarization and computer-mediated communication: Effects of communication cues, social presence, and anonymity. *Information Systems Research, 13*(1), 70–90.

Tan, B. C. Y., Watson, R. T., & Wei, K. K. (1995). National culture and group support systems: Filtering communication to damped power differentials. *European Journal of Information Systems, 4*(2), 82–92.

Tan, B. C. Y., Wei, K. K., Watson, R. T., Clapper, D. L., & McLean, E. R. (1998). Computer-mediated communication and majority influence: Assessing the impact in an individualistic and a collectivistic culture. *Management Science, 44*(9), 1263–1278.

Triandis, H. C. (1987). Individualism and social psychological theory. In C. Kagitcihbasi (Ed.), *Growth and progress in cross-cultural psychology* (pp. 78–83). The Netherlands: Lisse, Swets and Seitlinger.

Triandis, H. C. (1995). *Individualism and collectivism*. Boulder, CO: WestView Press.

Tung, L. L., & Quaddus, M. A. (2002). Cultural differences explaining the differences in results in GSS: Implications for the next decade. *Decision Support Systems, 33*(2), 177–199.

Valacich, J. S., Dennis, A. R., & Connoly. (1994). Idea generation in computer based groups: A new ending to an old story. *Organizational Behavior and Human Decision Processes, 57*(3), 448–468.

Valacich, J. S., Jessup, L. M., Dennis, A. R., & Nunamaker, J. F., Jr. (1992). A conceptual framework of anonymity in group support systems. *Group Decision and Negotiation*, 1(3)f, 219–241.

Watson, R. T., Ho, T. H., & Raman, K. S. (1994). A fourth dimension of group support systems. *Communications of the ACM, 37*(10), 45–55.

Zigurs, I., & Buckland, B. K. (1998). A theory of task/technology fit and group support systems effectiveness. *MIS Quarterly, 22*(3), 313–334.

KEY TERMS

Anonymity in GSS: The situation when participants' names are not made public in a GSS environment.

Culture: The collective programming of the mind, which distinguishes the members of one group or category of people from another.

Group Support Systems (GSS): Any combination of hardware and software that enhances groupwork.

Individualism: Refers to a preference for a loose-knit social framework in society in which individuals are only supposed to take care of themselves and their immediate families. This is opposed to collectivism, which implies a preference for a tightly knit social framework in which individuals can expect their relatives and clan to protect them in exchange for loyalty.

Long-Term Orientation: This stands for the fostering of virtues oriented toward future rewards, in particular, perseverance and thrift. Its opposite pole, short-term orientation, stands for the fostering of virtues related to the past and present, in particular, respect for tradition, preservation of "face" and fulfilling social obligations.

Masculinity: Refers to a preference for achievement, heroism, assertiveness, and material success; as opposed to femininity, which implies a preference for relationships, modesty, caring for the weak, and quality of life.

Power Distance: The extent to which society accepts the fact that power in institutions and organizations is unevenly distributed.

Uncertainty Avoidance: The degree to which a society feels threatened by uncertain and ambiguous situations, which leads them to support beliefs promising certainty and to maintain institutions protecting conformity.

Current Multicast Technology

Gabor Hosszu
Budapest University of Technology and Economics, Hungary

INTRODUCTION

Multicast technology is one-to-many communication, oppositely from the usual one-to-one (unicast) communication, which provides an efficient solution to create multiparty collaborative software by delivering the data flows on an appropriate distribution tree. The root of the distribution tree is the sender and its leaves are the receivers. Many Internet applications, such as distributed simulation, remote education, and videoconference require the underlayer network to support multicast communication.

The main problem of the multicasting is, however, how to construct and maintain the path, called delivery or distribution tree. There are solutions in different levels of the protocol stack, including datalink-layer, network-layer, and application-layer. Only the second and the third level multicast have practical importance; therefore we focus on these in the following.

The network-layer multicast is based on the Internet protocol (IP), which is why this kind of multicasting is called IP multicast. This technology is bandwidth conserving and reduces traffic by simultaneously disseminating a single stream of information to potentially thousands of corporate recipients. Multicast packets are replicated in the network at the point where paths are multiplied by multicast routing protocols, resulting in the most efficient delivery of data to multiple receivers. The main advantage of the multicasting is the smaller bandwidth usage than a normal unicast-based transmission.

The application-layer multicast (ALM) uses a different approach. In this technology, the multiplying nodes are host and the transmission is based on unicast. However, even the application-layer multicast has serious advantages over the pure unicast transmission, since in case of the traditional unicast, the sender has to create as many parallel data streams as receivers joined to the session. However, in case of ALM the sender has to serve a limited number of data streams, since the other group members as nodes also act as multiplication points. In such a way, they create a distribution tree.

BACKGROUND

The traditional unicast require the source to send more than one copy of the data, since there are point-to-point connections between the sender and each of the receivers. Even low-bandwidth applications can benefit from using multicast when there are thousands of receivers. High-bandwidth applications, such as video streaming, may require a large portion of the available network bandwidth for a single stream. In these applications, multicast is the only way to send to more than one receiver simultaneously.

The current Internet applications manage a large and widely distributed set of users, have multiple data streams that vary in content and media type, and make use of multiple unicast and multicast streams in a single session. Examples of these distributed, interactive applications include corporate communications, distance learning, video conferencing, stock quotes, software distribution, network news, collaborative visualization, distributed interactive simulations, and multiplayer games.

Table 1. The IP address ranges

Name	Purpose	Address Range
Class A	Unicast addresses for large networks	1.0.0.0 - 127.255.255.255
Class B	Unicast addresses for medium networks	128.0.0.0 - 191.255.255.255
Class C	Unicast addresses for small networks	192.0.0.0 - 223.255.255.255
Class D	Multicast addresses	224.0.0.0 - 239.255.255.255
Class E	Reserved	240.0.0.0 - 255.255.255.254

THE IP MULTICAST APPROACH

Special IP-addresses are used in the IP-multicast, which do not belong to certain hosts, but define multicast channels. Table 1 presents the address ranges of the IP address space. The Class A, B and C are used for the traditional unicast communication; the Class D addresses are applicable for multicasting.

The first model of the IP-multicasting was proposed by Steve Deering in 1988 in his PhD dissertation, and its first standard was published in the RFC 1112 (Deering, 1989). The most important idea in Deering's model for the multicasting is the so-called multicast group concept. Table 2 summarizes the main factors of this concept.

THE IP MULTICAST ROUTING

The first issue of multicast delivery is the communication between the host and the local router. The host does not have routing functionality; it uses a special signaling protocol, the Internet Group Management Protocol (IGMP), to let the local router know that the host became interested in the traffic of a certain multicast group.

The multicast delivery in the inter-router area is based on the multicast-enable routers, which have multicast routing protocols in their protocol stack. The multicast routing protocols differ from each other according to the underlying unicast routing protocols used. Table 3 presents the most popular multicast routing protocols and their most important properties (Hosszú, 2001).

An important step in creating the delivery tree is to discover all the routers interested in the given multicast session. In case of the DVMRP (Waitzman, 1988), MOSPF (Moy, 1994) and dense mode version of the PIM multicast routing protocols, the routers periodically flood the whole network with a multicast data packet, forwarding it to every router (Adams et al., 2003). Routers not interested in that session send a prune message back to the source of the packet and so the following packets will not be forwarded to these. The flooding step generates a huge amount of unnecessary packets, which is acceptable in dense mode situation, where the majority of the routers are interested in a multicast delivery. This routing method is called flood-and-prune model, and its operation is demonstrated in Figure 1.

In the sparse mode case, where only a small fraction of the routers are interested in a multicast session, the periodically executed flooding means an unacceptable load in the inter-routing area. In such a situation that multicast routing protocols can efficiently operate, rendezvous points (RP) are used to join to a multicast session and in such a way that the periodical flooding phase is avoided. In case of the PIM-SM routing protocol (Fenner et al., 2004), if a host wants to send multicast traffic, it will send the data packets to its local router, and then it forwards the data to the RP, which is now a designated remote router. The other host interested in that multicast traffic will send an IGMP join message to its local router, in order to become a receiver. Then the local router will send a PIM-SM joint message to the RP. Figure 2 shows this mechanism, where two receivers join to the multicast session.

After completing the join phase, the RP continuously forwards the multicast traffic toward the local router of the receiver. The data from the source to the RP is an encapsulated unicast PIM-SM control message. Figure 3 demonstrates this phase.

The root of such a delivery tree is the RP. A more optimized tree can be obtained, if the routers realize that a shorter path can be created between the source and the receiver. In this case the PIM-SM routers can switch over from the RP-rooted tree to the source-rooted tree. Figure 4 shows the created source-rooted tree.

Besides constructing the tree, its stability is also an important problem, since the routing protocol must upgrade at least a part of the tree whenever a member joins or leaves the multicast session (Van Mieghem & Janic, 2002).

Another serious routing problem of the IP-multicast is the inter-AS (autonomous system) multicast routing, since transmitting multicast traffic is not easy among the peering ASs. To overcome this limitation, different wide-area protocols are developed. The current practice is the usage of the protocol set MBGP/PIM-SM/MSDP; however, it has scalability problems, since it uses flooding to inform

Table 2. Components of the multicast group concept

- **IP-style semantics:** Similarly to the traditional (unicast) IP communication, a source can send data at any time; for this it does not have to join to the group of hosts.
- **Open groups:** The source does not have to know the members of the groups for sending data and the receivers (members of the group) do not have to know the source.
- **Variable groups:** The hosts can create, join to or leave any group at any time. The communication does not need any control center to coordinate the activity of the hosts.

Table 3. The widely used multicast routing protocols

Name	Main properties
Distance Vector Multicast Routing Protocol (DVMRP)	Its main problem is that it is not scalable for wide-area usage due to its flooding property.
Multicast Open Shortest Path First (MOSPF)	It uses the underlying OSPF unicast routing tables. Due to the huge control information packets among the routers, it is not scalable for wide-area usage.
Protocol Independent Multicast (PIM)	It can interoperate with any underlying unicast routing protocol. Its sparse mode version called PIM-SM is used for wide-area routing.

the receivers in the remote ASs about a local multicast source (Fenner & Meyer, 2003). The long-term solution may be MASC/MBGP/BGMP stack, which constructs a multicast tree from the ASs and avoids the flooding among them (Rajvaidya & Almeroth, 2003).

From the viewpoint of the market-oriented applications, the Deering's multicast model had to modify in order to be less vulnerable. For instance, a malicious sender can send noisy data to a multicast channel, which is used by other users. In order to avoid such an unwilling situation, the multicast concepts called source-specific multicast (SSM) were developed, where the host specifies not only the multicast group (class D) address sending an IGMP join message to the local router, but also the set of sources, from which it accepts multicast data (Almeroth et al., 2001). In this network the routers can also distinguish data streams having the same multicast address but originating from different sources, if they use SSM routing protocol such as the source-specific version of the PIM-SM, called PIM-SSM (Meyer et al., 2004).

THE CONNECTIONLESS AND CONNECTION ORIENTED TRANSMISSIONS

There are two basic types of communication; one is called connectionless, where the sender starts to send the data packets without any previous negotiation phase with the receiver. Such a method is used in the Internet, where the sender host or router forwards the IP packets without any delivery guarantees. The IP-multicast is similar to the conventional IP unicast communication from the point of view that quality of data transmission is non-reliable, the so-called best-effort. It means that there is not any guarantee for any parameter of the transmission, such as the delay, the error-free delivery, and so forth. Its reason is the connectionless property of the Internet protocol, where there is not any logical connection among the communicating entities in the network level. When a router sends data, it has no any knowledge if the other router receives that data properly or not.

Figure 1. Operation of a flood-and-prune multicast routing algorithm

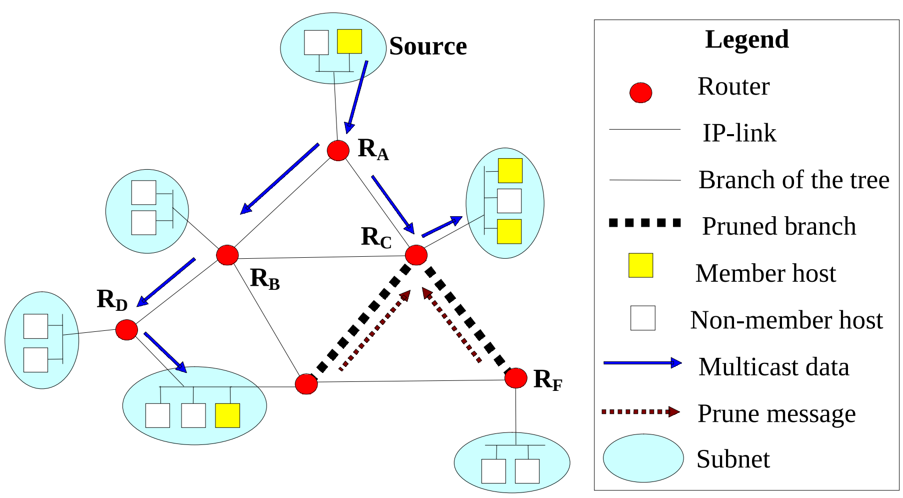
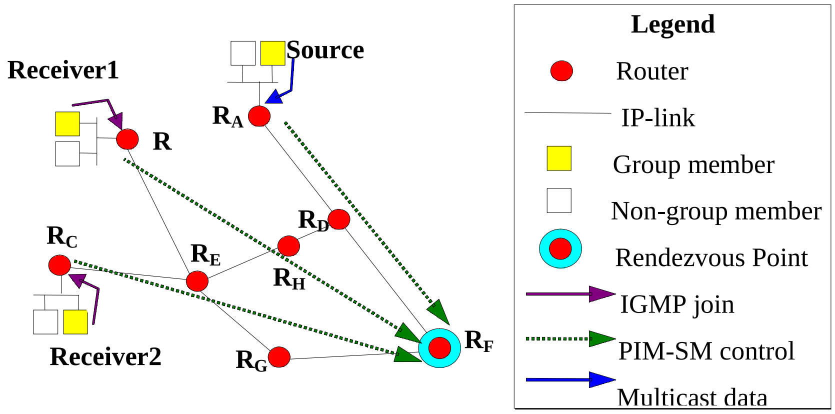


Figure 2. The joining phase in the PIM-SM operation



The other communication form is connection oriented, where the sender and the receiver negotiate with some communication parameters such as the sending rate. An example of the connection oriented data transfer is the Transmission Control Protocol (TCP) that is used in the unicast to get reliable delivery. However, the TCP is not usable for multicasting, since the TCP receiver client acknowledges every packet, and in the case of a large number of receivers, this could cause serious congestion at the network link of the multicast source. That is why the multicast applications use the User Datagram Protocol (UDP) for transmission, which has no reliable mechanism.

METHODS TO INCREASE THE RELIABILITY OF THE MULTICAST

Since most of the multicast applications are media-related software, for example, media conference, voice distribution, shared whiteboard, or various collaborative media tools, they need more reliability than the best-effort delivery. In order to increase the reliability of multicast applications, additional multicast transport protocols are used to achieve the required level (Whetten & Taskale, 2000). Such a protocol is the NORM: NACK-Oriented Reliable Multicast Protocol (Adamson et al., 2004).

The multicast transport protocols have many different property attributes of the data delivery. These attributes can be represented by the protocol parameters. Each protocol parameter describes different reliability mechanisms for the same delivery attribute. Such a protocol parameter is, for instance, the repair method, which can get the values such as "retransmission," "forward-error correction," "interleaving," or different ways of the "local receiver-based repairs". Another parameter is the acknowledgement type, the possible values of which may be "tree-based," "ring-based" or a "simple direct form".

Figure 3. The data transmission among the hosts and the PIM-SM routers

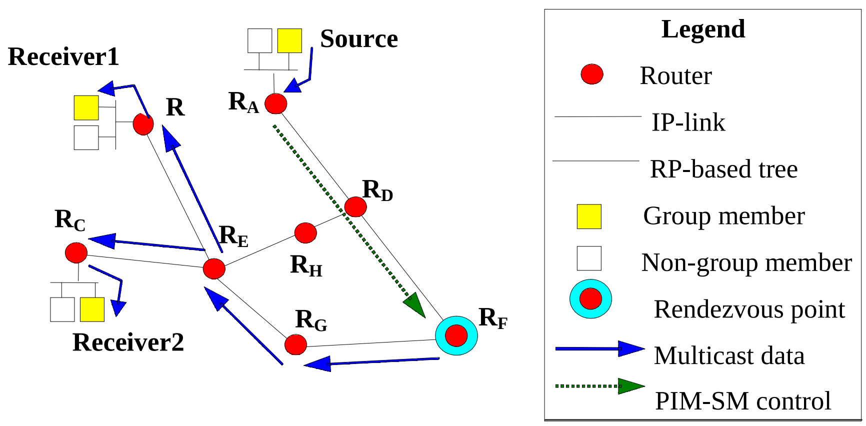
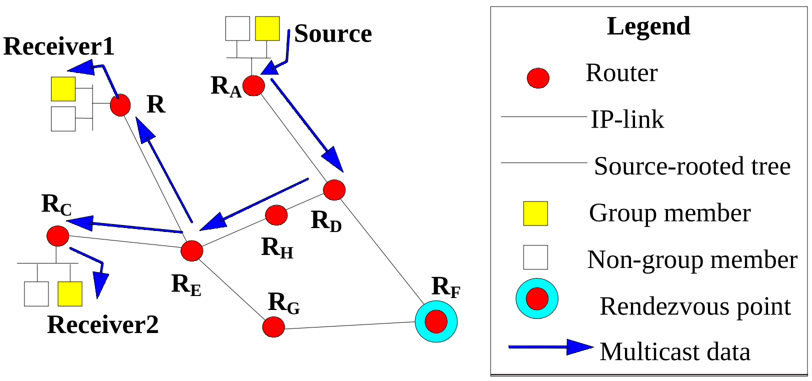


Figure 4. The data transmission through the source-rooted tree



The various applications have different reliability requirements, and therefore these protocol parameters should be optimized in order to create the multicast transport protocol best suited for a given application. However, applying any appropriate mathematical optimization method, at the selection of the protocol parameters mentioned previously, a linearly independent (in other words, orthogonal) set of parameters must be applied. To do this a hyperspace of the parameters is created where all transport protocol corresponds to one point of this space. The optimization procedure means to find the most suitable point in this space to provide the best performances of multicast. The possible values of protocol parameters (which are the types of various mechanisms as the components of these) are the realizations of the protocol functionalities. A quasi-orthogonal subset of the protocol parameters and their possible values are presented in Table 4. These parameters represent the well-known reliable mechanisms of the transport protocols.

To carry out a correct optimization procedure on the appropriately selected protocol parameters, a well usable simulation program should be applied in order to obtain statistically confident results for multicast data transfer. Using an appropriate simulator, the optimized transport protocol can be synthesized for a given media application, satisfying the requirements. That means that by a mathematical method an optimal point in the hyperspace of the protocol parameters can be found.

THE APPLICATION-LAYER MULTICAST TECHNOLOGY

The application-layer multicast (ALM) is also called host-multicast, since the nodes are responsible for the multicast delivery, not the routers. The main difference between the host-multicast and the IP-multicast is that in the case of the ALM, the hosts use unicast between them and the

Table 4. The selected set of the protocol parameters

Protocol parameter	Values
Flow control	Window-based, Rate-based, Multigroup multicast, Receiver give-up, None
Data accuracy	Reliable, Atomic, Non-reliable
Feedback addressee	Original source, Intermediate host, Every member, None
State control	Sender-based, Receiver-based, Shared, None
Feedback control	Structure-based, Timer-based, Representatives-based, Rate-based, None
Way of sending repair	Unicast, Multicast, None
Scope of repair	Global, Global to secondary group, Global to individual members, Local, None
Session membership control	Explicit, Implicit, None

multiplication points of the tree are the hosts themselves, not the routers, as in the IP-multicast (Banerjee et al., 2002; Castro, 2002; Zhang et al., 2002). Table 5 compares the properties of the host-multicast and the IP-multicast.

ALM is a new approach to providing multicast services to group applications. In this peer-to-peer (P2P) architecture, members organize themselves into an overlay topology for data delivery that adapts to the changing network conditions and group dynamics. P2P systems are distributed virtual networks that operate without centralized control. To find a particular piece of data within the network, P2P systems explicitly or implicitly provide a lookup mechanism that matches a given string (called a key) to one or more network nodes responsible for the value associated with that key.

Some of the ALM schemes can take advantage of IP-multicast support where available. However, such additional IP-layer capabilities are not essential for these protocols, and therefore they can be easily deployed in the Internet today. Current Internet P2P applications typically provide lookup functions using time-to-live (TTL) scope controlled flooding procedures. With this approach, the querying node wraps the query in a single message and sends it to all known neighbors. The neighbors check these to see if they are able to reply to the query by matching it to keys in their internal database. If they find a match, they send a response; otherwise, they forward the query to their own neighbors and increase the message's hop count. If the hop count becomes higher than the TTL limit, the forwarding will be cancelled. The TTL value thus determines a boundary for the query that controls its propagation.

THE THEMATIC MULTICAST CONCEPT (TMC)

As an example of the current searching works in the field of ALM, a novel concept of modeling relative density of members is the bunched mode. The bunched mode means a typical multicast scenario where many interested hosts are in certain institutes relatively far from each other. The members of a multicast group are locally in dense mode; however, these spots are far from each other, with their situation globally being similar to the sparse mode. This situation is typical when the collaborative media application has a special topic. That is why this model of communication is called thematic multicast concept (TMC). A typical TMC scenario is shown in Figure 5.

The group-member hosts in a bunch locally elect the designated members (DM). The DM calculates the shortest unicast IP tunnels among them. They exchange their IP addresses and shortest unicast paths among them. In such a way, all of them know all possible shortest unicast paths and calculate the same topology of the inter-bunch IP tunnels. This application-level routing mechanism, called shortest tunnel first (STF), is similar to the MOSPF routing in the network level. The MOSPF routers exchange the link-state information, but in the case of STF, the DMs exchange the path information to every other one, since there are UDP/IP paths to every other DM.

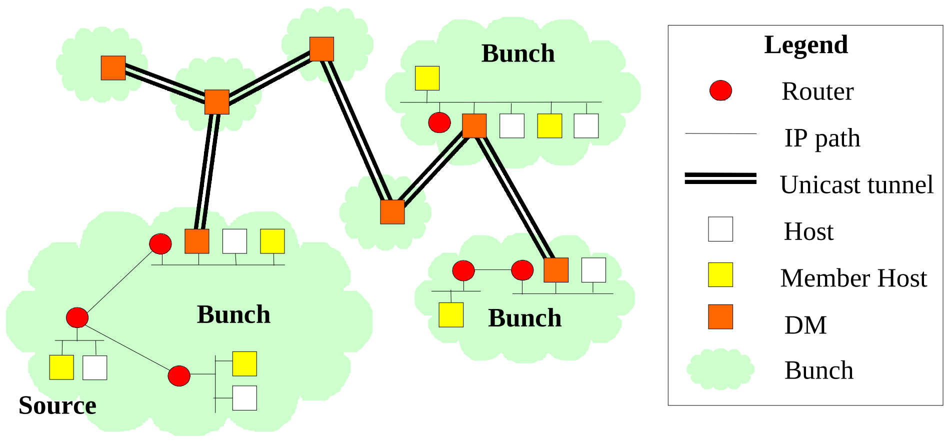
FUTURE TRENDS

Due to the increasing bandwidth, the users send more and more media content through the Internet. The collaborative applications also have strong demands. Therefore the multicast as a communication model will be used in the future. The choice between the IP multicast and the application-layer multicast, however, remains an open question. The most important advantage of the IP multicast is the reliable operation of the intermediate nodes, oppositely to the easy deployment of the ALM. The networking applications, which have a limited number of users, will prefer the ALM as multicasting technology; however, the stable, long-term applications with a number of users (e.g., media streaming) will prefer the IP multicast due to

Table 5. Comparing the properties of the host- and the IP-multicast

- **Deployment cost:** The IP-multicast requires additional knowledge in the routers to the basic unicast communication and extra work from the administrators of the routers; the ALM needs the traditional unicast IP infrastructure, only.
- **Flexibility:** The host- and the IP-multicast are flexible, and it is easy to change the topology of the multicast tree.
- **Speed of delivery:** The ALM theoretically cannot be more optimal than the IP-multicast tree, in such a way that the delivery of the ALM has higher delay than the IP-multicast.
- **Robustness:** In case of the ALM, the hosts are responsible to maintain the tree; in case of the IP-multicast, the routers define the tree. Since the routers operate safely, non-stop, but the user can switch off a host anytime, in such a way the IP-multicast is more robust.

Figure 5. A typical TMC scenario



its more reliable nature. Another side of the problem is that independently from the usage rate of the IP multicast, its sophisticated routing and transporting methods are adapted in the ALM, too.

In the future, new applications as well as methods will emerge in the field of multicasting in order to more efficiently use the existing network.

CONCLUSION

The emerging media applications use multicast technology to efficiently deliver large media data to the receivers. The fundamentals of the network-level IP-multicast and the application-level host-multicast are presented. The main fields of research are the wide-area routing and the reliable multicast transport protocols in the IP-multicast, and a novel application-level routing protocol and communication concept in the host-multicast. The advantages and the drawbacks of the two basic multicast technologies are compared. Selecting the most beneficial technology to apply depends on the requirements of the actual media application.

REFERENCES

Adams, A. et al. (2003). Protocol independent multicast - dense mode (PIM-DM): Protocol specification. *IETF PIM WG, draft-ietf-pim-dm-v2-03.txt*. Work in progress.

Adamson, B. et al. (2004). NACK-oriented reliable multicast protocol (NORM). *IETF, draft-ietf-rmt-pi-norm-09.txt*. Work in progress.

Almeroth, K.C. et al. (2001). Challenges of integrating ASM and SSM IP-multicast protocol architectures. *Int. Workshop on Digital Communications: Evolutionary Trends of the Internet (IWDC'01)*, Taormina, Italy.

Banerjee, S. et al. (2002). Scalable application-layer multicast. *ACM SIGCOMM*.

Castro, M. (2002). Scribe: A large-scale and decentralized application-level multicast infrastructure. *IEEE Journal on Selected Areas in Communications (JSAC), 20*(8).

Deering, S. (1989). Host extensions for IP multicasting. *IETF Network Working Group RFC 1112*.

Fenner, B., & Meyer, D. (2003). Multicast source discovery protocol (MSDP). *IETF Network WG, RFC 3618*.

Fenner, B. et al. (2004). Protocol independent multicast - sparse mode (PIM-SM): Protocol specification. *IETF PIM WG, draft-ietf-pim-sm-v2-new-09.txt*. Work in progress.

Francis, P. (2000). Yoid: Extending the multicast Internet architecture. Technical report. ACIRI. *http://www.aciri.org/yoid*

Hosszú, G. (2001). Introduction to multicast technology. In S.M. Rahman (Ed.), *Multimedia networking: Technology, management & applications* (pp. 369-411). Hershey, PA: Idea Group Publishing.

Meyer, D. et al. (2004). Source-specific protocol independent multicast in 232/8. *IETF Internet draft, draft-ietf-mboned-ssm232-08.txt*. Work in progress.

Moy, J. (1994). Multicast extensions to OSPF. *IETF Network Working Group RFC 1584*.

Moy, J. (1998). OSPF Version 2, STD 54. *IETF Network Working Group RFC 2328.*

Rajvaidya, P., & Almeroth, K. (2003). Analysis of routing characteristics in the multicast infrastructure. *Proceedings of IEEE INFOCOM, 12.*

Van Mieghem, P., & Janic, M. (2002). Stability of a multicast tree. *IEEE INFOCOM*, 1099-1108.

Waitzman, D. et al. (1988). Distance vector multicast routing protocol. *IETF Network Working Group RFC 1075*.

Whetten, B., & Taskale, G. (2000). The overview of reliable multicast transport protocol II. *IEEE Network*.

Zhang, B., Jamin, S., & Zhang, L. (2002). Host multicast: A framework for delivering multicast to end users. *IEEE INFOCOM*.

KEY TERMS

ALM (Application-Layer Multicast): A novel multicast technology that does not require any additional protocol in the network routers, since it uses the traditional unicast IP transmission.

Application-Level Multicast: Alternative name of the application-layer multicast.

AS (Autonomous System): A network where the main routers are in common administration. The Internet is composed of peering ASs.

Host-Multicast: Alternative name of the ALM.

IETF (Internet Engineering Task Force): A voluntary association for developing Internet standards.

IP-Multicast: Network-level multicast technology, which uses the special Class-D IP-address range. It requires multicast routing protocols in the network routers.

Multicast Routing Protocol: In order to forward the multicast packets, the routers have to create multicast routing tables using multicast routing protocols.

Multicast Transport Protocol: To improve the reliability of the multicast delivery, special transport protocols are used in addition to the widely used unreliable UDP.

Reliability: The improved quality of data transmission; different types of reliability exist, including data accuracy or real-time delivery.

RFC (Request for Comments): The IETF publish them as de facto standards of the Internet.

TTL (Time-to-Live): A field in the IP packet header. Its value is the allowed hop-count, the number of routers, which can forward the packet before delivery or dropping out.

Unicast: The one-to-one communication way, where only one host transfers data with another host. In the traditional IP the unicast is applied.

Deployment Cost: The IP-multicast requires additional knowledge in the routers to the basic unicast communication and extra work from the administrators of the routers; the ALM needs the traditional unicast IP infrastructure, only.

Flexibility: The host- and the IP-multicast are flexible, and it is easy to change the topology of the multicast tree.

Speed of Delivery: The ALM theoretically cannot be more optimal than the IP-multicast tree, in such a way that the delivery of the ALM has higher delay than the IP-multicast.

Robustness: In case of the ALM, the hosts are responsible to maintain the tree; in case of the IP-multicast, the routers define the tree. Since the routers operate safely, non-stop, but the user can switch off a host anytime, in such a way the IP-multicast is more robust.

IP-Style Semantics: Similarly to the traditional (unicast) IP communication, a source can send data at any time; for this it does not have to join to the group of hosts.

Open Groups: The source does not have to know the members of the groups for sending data and the receivers (members of the group) do not have to know the source.

Variable Groups: The hosts can create, join to or leave any group at any time. The communication does not need any control center to coordinate the activity of the hosts.

Current Network Security Systems

Göran Pulkkis
Arcada Polytechnic Helsinki, Finland

Kaj Grahn
Arcada Polytechnic Helsinki, Finland

Peik Åström
Arcada Polytechnic Helsinki, Finland

INTRODUCTION

Network security is defined as "a set of procedures, practices and technologies for protecting network servers, network users and their surrounding organizations" (Oppliger, 2000, Preface). The need for network security is caused by the introduction of distributed systems, networks, and facilities for data communication. Improved network security is required because of the rapid development of communication networks. Network security is achieved by using software based tools, that is, network security software (Pulkkis, Grahn & Åström, 2003).

BACKGROUND

This article gives a topical overview of network security software, that is, the topics are not covered in detail, and most topics are briefly introduced and left for further study. The main objective is to present "state-of-the-art" of network security software and to stimulate discussion about related skills and education needed by network users, IT professionals, and network security specialists.

PROTECTION AGAINST MALICIOUS PROGRAMS

Malicious software exploits vulnerabilities in computing systems. Malicious program categories are (Bowles & Pelaez, 1992):

- *Host program needed*
    - **Trap door**
    - **Logic bomb**
    - **Trojan horse**
    - **Virus**
- *Self-contained malicious program*
    - **Bacteria**
    - **Worm**

The ideal protection is prevention, which still must be combined with detection, identification and removal of such malicious programs for which prevention fails. Protection software is usually called antivirus software (Stephenson, 1993):

- **First Generation**
  Simple scanners searching files for known virus "signatures" and checking executable files for length changes.
- **Second Generation**
  Scanners using heuristic rules and integrity checking to find virus infection.
- **Third Generation**
  Memory resident "activity traps" identifying virus actions like opening executable files in write mode, file system scanning, and so forth.
- **Fourth Generation**
  Antivirus software packages using many different antivirus techniques in conjunction.

Examples of recent advanced antivirus techniques are *generic decryption* (*GD*) and *digital immune system (DIS)* technology (Stallings, 2000, Chap. 9).

Protection levels of modern antivirus software are:

- **Gateway level protection** consists of mail server and firewall protection. Viruses are detected and removed before files and scripts reach a local network.
- **File server level protection** consists of server software. Viruses are detected and removed even before network users access their files/scripts.
- **End user level protection** consists of workstation software. Viruses undetected in outer defense lines are detected and removed.

All levels should be combined to achieve depth in antivirus defense. Virus definition databases should be automatically and/or manually updated.

FIREWALL SOFTWARE

Firewalls protect computers and computer networks from external security threats. Firewall types are:

- **Packet-filtering router**, which applies a software and/or hardware implemented filtering rule set to each incoming/outgoing IP packet and then forwards or discards the packet. Most TCP/IP routers support basic user defined filtering rules. A packet-filtering firewall can also be a stand-alone network link device, for example, a computer with two network cards.
- **Application-level gateway (proxy server)**, which acts as an application level traffic relay. A typical application level gateway is a protocol oriented proxy server on a network link, for example, an HTTP proxy, a SMTP proxy, a FTP proxy, and so forth. An HTTP proxy is also a Web page cache for Web usage through the proxy.
- **Circuit-level gateway**, which typically relays TCP packets from one connection to another without examining the contents.

CRYPTOGRAPHIC SOFTWARE

Cryptographic network security software consists of secure network applications and secures network system software.

Software for Secure Network Level Data Communication

Secure network level data communication is based on the Internet Protocol Security (IPSec) protocol. Two computers in the same TCP/IP network implement end-to-end security through the network, when IPSec software is installed and properly configured in both computers. IPSec provides two operation modes:

- **Transport mode**, where original IP headers are used
- **Tunnel mode**, where new IP headers are created and used to represent the IP tunnel endpoint addresses. IPSec is usually embedded in Virtual Private Network (VPN) software. VPN provides secure LAN functionality in geographically distributed network segments and for Internet connected computers. Fundamental VPN types are:
- **Access VPN**, a secure connection to a LAN through a public TCP/IP Network
- **Connection VPN**, a secure remote connection between two logical LAN segments through a public TCP/IP network.

IPSec and VPN functionality is included in Windows 2000/XP. Commercial VPN software products are F-Secure VPN+™, Nokia VPN, Cisco Security VPN Software, and so forth. Open source IPSec and VPN software is also available (Linux, 2003).

Middleware

Middleware is a software layer between the network and the applications for providing services like identification, authentication, authorization, directories, and security (Internet2 Middleware, 2004). Shibboleth (2004) is an example of open source authentication and authorization middleware. Commercial security middleware based on the SHH protocol is presented in SSH Tectia Solution (2004).

Software for Secure Transport Level Data Communication

Many network applications are based on the IETF Transport Layer Security (TLS) standard. The TLS/SSL protocol is based on an established client-server TCP connection. Then both computers execute the SSL Handshake Protocol to agree on the cryptographic algorithms and keys for use in the actual data communication (Stallings, 2000, p. 214). TLS/SSL versions of common application level TCP/IP protocols are available (see Table 1).

VPN solutions can also be implemented using the TLS/SSL protocol and executed on the transport level. This technology, called SSL-VPN provides VPN functionality to geographically distributed network segments and for Internet connected computers using a standard Web browser. A commercial SSL-VPN software product is presented in Symantec™ (2004).

Web Security

Basic Web security features are access level security and transaction level security. Access level security is provided with firewalls which guard against intrusion and unauthorized use. Transaction level security requires protocols for protecting the communication between a

Table1. Secure application level protocols based on TLS/SSL (Oppliger, 2000, p.135)

Secure protocol	Port	Description
HTTPS	443	TLS/SSL protected HTTP
POP3S	995	TLS/SSL protected POP3
IMAPS	993	TLS/SSL protected IMAP4
SMTPS	465	TLS/SSL protected SMTP
NNTPS	563	TLS/SSL protected NNTP
LDAPS	636	TLS/SSL protected LDAP

Web browser and a Web server. Proposed protocols are HTTPS, S-HTTP, and PCT (Pulkkis, Grahn & Åström, 2003, p. 14). HTTPS was originally introduced by Netscape for the Netscape Navigator browser. Presently, HTTPS is an accepted standard supported by practically all Web browsers, while S-HTTP and PCT are seldom used.

E-Mail Security

E-mail traffic to and from e-mail servers is protected, when the SMTPS protocol is used. Sessions with e-mail client programs can be protected

- by using the mailbox access protocols POP3S and IMAPS, and
- by embedding an e-mail client program in a HTTPS Web page.

E-mail content security requires e-mail client program extensions for signing and/or encrypting outgoing messages as well as for decrypting and/or signature checking incoming messages. The most widely used e-mail client security extensions are PGP and S/MIME, see Stallings (2000, Chap. 5).

Secure E-Commerce

Secure electronic transaction (SET), originally introduced by MasterCard and Visa, is an open security specification designed to protect credit card transactions on Internet (Stallings, 2000, Chap. 7). SET provides

- secure communication,
- trust based on X.509v3 certificates, and
- privacy based on strictly controlled access to sensitive information.

Secure Shell (SSH)

The Secure Shell (SSH), a secure remote connectivity protocol in TCP/IP networks, is a de-facto standard being further developed by one of the IETF Security Area Working Group. Two SSH versions have hitherto been developed: SSH1 and SSH2. Commercial as well as open source SSH implementations are available (SSH Communications Security, 2003; OpenSSH, 2003).

Secure Network Management

A protocol for secure network management, SNMPv3, was introduced in 1998 by IETF to address the lack of security in earlier SNMP versions. SNMPv3 incorporates authentication and encryption features to SNMP managers and access control features to SNMP agents (Stallings, 2000, Chap. 8).

Secure DNS (DNSSEC)

The absence of trust in DNS host name resolution is a security hazard in all TCP/IP applications. To address this problem IETF formed a working group to develop the DNSSEC standard. The objective is to provide both authentication and integrity to DNS information. DNSSEC uses public key cryptography to sign DNS information (Securing the Domain Name System DNSSEC, 2003).

Secure Routing Software

Routing protocols and their hardware/software implementations in computer networks are usually open and functionally unprotected. A manifestation of an emerging recognition of routing security in the Internet community is the recently formed IETF Routing Area Working Group "Routing Protocol Security Requirements (rpsec)", which in December 2003 published the Internet Draft "Generic Threats to Routing Protocols" (Barbir, Murphy, & Yang, 2003).

Security Token Access Software

The required randomness in cryptographic key generation may cause security problems, if secret or private keys are used in the main memory of network connected computers. Intruders may identify keys as random areas, since code and data are usually structured. For this reason, a secret key should be changed after some time interval, and a private key should be stored and used in a separate hardware token together with the code of the cryptographic operations using it. Usually, this token is a computer chip on a smartcard. Private keys on a smartcard are protected by pin codes or biometrically by digital fingerprint comparison and/or by digital voice recognition. A numerical smartcard reader dedicated keypad is necessary for pin code security. Software required for accessing cryptographic tokens on smartcards is:

- Device driver for communication with the smartcard through the used smartcard reader.
- PC/SC, a specification set released by an international consortium (PC/SC Workgroup, 2003) for integration with the operating system. In PC/SC, a device manager keeps track of the cards and card readers connected to a computer.
- An Application Programming Interface (API) like PKCS#11, also called CrypTokI, or Microsoft Crypto API.

Wireless Security Software

A radio interface is by nature easy to access. Security threats are either passive or active attacks. Active attacks involve altering data streams and passive attacks, on the other hand, include snooping on transmission. The most important security features are authentication, authorization, confidentiality, integrity and denial of service. The corresponding software is included in the network.

WLAN security is built-up around 802.11i/WPA (Wi-Fi, 2003) or VPN/IPSec. In Bluetooth, there are three security modes handled by the security manager (Grahn, Pulkkis, & Guillard, 2002). A bonding process including pairing and authentication, and encryption based on the SAFER+ algorithm are implemented. Also, a concept of trusted devices is applied.

The security features in a GSM network can be divided into three subparts, subscriber identity authentication, user and signaling data confidentiality, and subscriber identity confidentiality. In 3G systems, security is based on what was implemented in GSM. The encryption algorithm is stronger, the application of authentication algorithms is stricter and subscriber confidentiality is tighter. The security principles are all incorporated into the Authentication and Key Agreement (AKA) procedure (Grahn et al., 2002).

SECURITY ADMINISTRATION SOFTWARE

Security administration uses intrusion detection software, vulnerability checking software, and software for security software management.

An Intrusion Detection System (IDS) monitors traffic in a network and/or user behavior in a host computer to identify possible intruders and/or anomalous behavior and/or misuse (Stallings, 2000, Chap. 9).

Major vulnerabilities are too short, easily guessed or cracked passwords. A potential intruder could run a password cracker on the encrypted passwords stored in a network. System administrators can use cracking to disable usage of bad passwords.

Intrusion prevention requires regular scans for unnecessary open ports and other vulnerabilities like missing security patches.

Network security software in host computers and in other network nodes like routers is often software managed. Management software examples are:

- F-Secure® Policy Manager™for management of “not only antivirus solutions, but all critical network security solutions on all tiers of the network” (F-Secure, 2003).
- Symantec™Gateway Security 5400 Series with firewall, antivirus protection, intrusion prevention/detection, and vulnerability check features managed from a centralized console (Symantec™, 2003).

SECURITY SOFTWARE DEVELOPMENT

Antivirus protection programming skills require knowledge about self-modifying programs/scripts and of virus sensitive operating system features.

Firewall software programming skills are based on detailed knowledge of TCP/IP protocol stack implementation software.

The open source toolkit OpenSSL is available for TLS/SSL application design (The OpenSSL Project, 2003). OpenSSL is installed as a C function library. Also, commercial development tools are available (Certicom, 2003; RSA BSAFE, 2003).

S/MIME e-mail extensions can with special toolkits be added to existing network software and be embedded in network software being developed. Freeware S/MIME v3 toolkits are (S/MIME Freeware Library, 2003) and the Mozilla S/MIME Toolkit. Phaos S/MIME Toolkit is a Java package (Phaos S/MIME, 2003) for secure messaging in Java applications.

IPSec software development is usually VPN software development. IPSec can be integrated in the networking software and/or hardware of a router/a computer node. Commercial IPSec developer toolkits such as SSH QuickSec™and SSH IPSEC Express™are available (SSH Communications Security, 2003).

Toolkits for SSH protocol integration during network software design are also available (SSH Communications Security, 2003).

In smartcard application development usually some development kit is used. Microsoft offers a Smartcard Toolkit to be used together with visual programming tools.

RSAEuro is an open source cryptographic toolkit providing C functions for the use of digital signatures, data encryption and for creating cryptographically safe random numbers. RSAEuro can be downloaded from public FTP servers in European university networks.

DESIGN OF SECURE SOFTWARE

Network security software implements security features. Other network software implements functionality and other features like usability, efficiency, simplicity, safety, dependability, reliability, and so forth. Security requirements for any network software include:

- absence of vulnerabilities and security holes; and
- secure interfaces.

Security should be integrated in the network software life cycle starting from the specifications. The need to assess vulnerability and to react on security incidents should be proactively minimized before network software is used. A recent handbook for secure software design is available (Viega & McGraw, 2002).

FUTURE TRENDS

IPSec is a mandatory part of the new version of the IP protocol, IPv6 (IP version 6 Working Group, 2004). Thus IPSec is automatically included in the IP software in all nodes in future TCP/IP networks. Also DNSSEC and secure routing protocols will be included in the system software of future TCP/IP networks.

New wireless network protocols emerging are among others Wireless USB (WUSB) (Kolic, 2004) and ZigBee™ (2004). WUSB will offer the same functionality as standard wired USB devices. ZigBee is a low-power, short-range, wireless technology. Both technologies will be used in networking solutions for home/industrial automation.

Wi-Fi Protected Access version 2 (WPA2) includes full 802.11i support in a WLAN (WPA Specification Documentation, 2003). WPA2 will replace RC4 with AES. It will also include the CCM protocol. The new standard implementation is hardware accelerated and will require replacement of most access points and some NIC's (Network Interface Cards).

Session key agreements in future wired network, will be based on absolutely secure quantum cryptography protocols (Bennett & Brassard, 1984), which are physically implemented by transmission of randomly polarized laser pulses in optical fibers (Stucki, Gisin, Guinnard, Ribordy, & Zbinden, 2002). Absolutely secure means that verified reception of a session key is also a proof that the same key has not been eavesdropped. Commercial Quantum key distribution technology is already available (Quantum Key Distribution, 2004).

CONCLUSION

Software solutions and tools are network security cornerstones. Today, network security software is a large and complex rapidly expanding area. Network security software skills are a needed by every computer and computer network user. This has profound implications on all education, since use of computer networks is inevitable.

Education for professional network security software skills should include:

- installation, configuration, and test use of all categories of available network security software solutions and products,
- source code inspection exercises of open source network security software solutions,
- programming exercises and projects with TLS/SSL application development environments and cryptographic toolkits.

Network security software development skills are important in upper level network security education.

REFERENCES

Barbir, A., Murphy, S., & Yang, Y. (2003). Generic threats to routing protocols, Internet-Draft, Internet Engineering Task Force, December 17, 2003. Retrieved on March 13, 2004, from http://www.ietf.org/internet-drafts/draft-ietf-rpsec-routing-threats-04.txt

Bennett, Ch., & Brassard, G. (1984). Quantum cryptography: Public key distribution and coin tossing. *International Conference on Computers, Systems and Signal Processing*, Bangalore, India, December 1984 (pp.175-179).

Bowles, J., & Pelaez, C. (1992, August). Bad code. *IEEE Spectrum*.

Certicom (2003). Certicom products. Retrieved on November 9, 2003, from *http://www.certicom.com/products/*

F-Secure (2003). F-Secure products and services. Retrieved on November 9, 2003, from http://www.f-secure.com/products/

Grahn, K., Pulkkis, G., & Guillard, J.-S. (2002). Security of mobile and wireless networks. In *Proceedings of the Informing Science + IT Education Conference*, Cork, Ireland, June 2002, (pp.587-600). Retrieved on April 2, 2004, from *http://ecommerce.lebow.drexel.edu/eli/2002Proceedings/papers/Grahn152Secur.pdf*

IETF. (2003). The Internet Engineering Task Force. Retrieved on November 9, 2003, from http://www.ietf.org

Internet2 Middleware (2004). Internet2 Middleware Initiative (I2-MI) Portal. Retrieved March 27, 2004, from *http://middleware.internet2.edu/*

IP Version 6 Working Group (ipv6). (2004). Retrieved on April 2, 2004, from http://www.ietf.org/html.charters/ipv6-charter.html

Kolic, R. (2004). An introduction to wireless USB (WUSB). Retrieved April 2, 2004, from *http://deviceforge.com/articles/AT9015145687.html*

Linux (2003). Linux FreeS/WAN Project. Retrieved on November 9, 2003, from *http://www.freeswan.org/*

OpenSSH. (2003). Project goals. Retrieved on November 8, 2003, from http://www.openssh.com/goals.html.

Oppliger, R. (2000). *Security technologies for the Word Wide Web*. USA: Artech House.

PC/SC Workgroup. (2003). Retrieved on November 9, 2003, from http://www.pcscworkgroup.com

Phaos S/MIME. (2003). Retrieved on November 9, 2003, from http://www.phaos.com/products/smime/smime.html

Pulkkis, G., Grahn, K., & Åström, P (2003). Network security software. In R. Azari (Ed.), *Current security management & ethical issues of information technology*. Hershey, PA, USA: IRM Press, (pp.1-41).

Quantum Key Distribution (QKD). (2004). id Quantique SA, Geneva, Switzerland. Retrieved April 2, 2004, from *http://www.idquantique.com/qkd.html*

RSA BSAFE. (2003). Security protocols. Retrieved on November 8, 2003, from *http://www.rsasecurity.com/products/bsafe*

S/MIME Freeware Library (SFL). (2003). Retrieved on November 9, 2003, from *http://digitalnet.com/knowledge/sfl_home.htm*

Securing the Domain Name System DNSSEC. (2003). Retrieved on November 9, 2003, from http://www.dnssec.org

Shibboleth (2004). Shibboleth Project Portal. Retrieved on March 27, 2004, from *http://shibboleth.internet2.edu/*

SSH Communications Security. (2003). Retrieved on November 9, 2003, from http://www.ssh.com

SSH Tectia Solution. (2004). Retrieved on March 27, 2004, from *http://www.ssh.com/products/tectia/*

Stallings, W. (2000). *Network security essentials*. USA: Prentice-Hall.

Stephenson, P. (1993, November). Preventive Medicine. *LAN Magazine*.

Stucki, D., Gisin, N., Guinnard, O., Ribordy, G., & Zbinden, H. (2002). Quantum key distribution over 67 km with a plug&play system. *New Journal of Physics*, *4*, 41.1-41.8. Retrieved on April 2, 2004, from *http://www.idquantique.com/files/njp-2002.pdf*

The OpenSSL Project. (2003). Retrieved on November 8, 2003, from *http://www.openssl.org*.

Symantec™. (2004). Symantec™Clientless VPN Gateway 4400 Series. Retrieved on March 27, 2004, from *http://enterprisesecurity.symantec.com/content/displaypdf.cfm?pdfid=653*

Symantec™. (2003). Symantec™products and services. Retrieved on November 9, 2003, from *http://www.symantec.com/product/*

Viega, J., & McGraw, G. (2002). *Building secure software*. USA: Addison Wesley

Wi-Fi (2003). Wi-Fi protected access: Strong, standards-based, interoperable security for today's Wi-Fi networks. (2003). Retrieved on December 13, 2003, from *http://www.wi-fi.org/OpenSection/pdf/Whitepaper_Wi-Fi_Security4-29-03.pdf*

WPA Specification Documentation. (2003). Version 2.0. Link on Wi-Fi Protected Access Portal. Retrieved on December 12, 2003, from *http://www.wi-fi.org/OpenSection/protected_access.asp?*

ZigBee™ (2004). ZigBee™ Alliance. Retrieved on April 2, 2004, from *http://www.zigbee.org*

KEY TERMS

E-Mail Protocols: Simple mail transport protocol (SMTP) is a set of commands for transport of ASCII encoded e-mail messages. Post office protocol (POP3) retrieves new messages from a mailbox to a remote e-mail client. A remote e-mail client can simultaneously access several mailboxes on different mail servers with the Internet message access protocol (IMAP).

Internet Engineering Task Force (IETF): An open international community engaged in Internet architecture evolution (IETF, 2003). Working Groups in several topical areas develop technical drafts and Internet standards.

Internet Protocol Security (IPSec): Developed by an IETF Security Area Working Group. IPSec introduces a new TCP/IP protocol stack layer below IP. IPSec adds authentication and optionally encryption to transmitted data packets. Authentication ensures that packets are from the right sender and have not been altered. Encryption prevents unauthorized reading of packet contents.

Pretty Good Privacy (PGP): An e-mail extension used to encrypt/decrypt and cryptographically sign e-mail as well as to verify e-mail signatures. Verification of a signature is a proof of sender identity and message authenticity.

Secure Multipurpose Internet Mail Extensions (S/MIME): A secure e-mail standard based on MIME. S/MIME, being further developed by an IETF Security Area Working Group, accomplishes privacy and authentication by using encryption/decryption, digital signatures and X.509 certificates

Simple Network Management Protocol (SNMP): An application layer TCP/IP protocol for management information exchange between network devices. SNMP includes two main software entity types: managers and agents.

Virus: Malicious code added to an executable file loaded to a computer and executed without the user's knowledge and consent. Computer viruses often copy and spread themselves to other computers in the same network.

X.509 and PKI: A X.509 certificate is a standardized data structure for proving the ownership of a public cryptographic key with a cryptographic signature of a trusted third party, a certification authority (CA). A public key infrastructure (PKI) is hardware, software, people, policies and procedures needed to issue, manage, store, distribute and revoke X.509 certificates.

Curriculum Development in Web-Based Education

Johanna Lammintakanen
University of Kuopio, Finland

Sari Rissanen
University of Kuopio, Finland

INTRODUCTION

A paradigm shift has taken place in the last decade, with a move from traditional to Web-based education at different educational levels (Harasim, 2000; Karuppan, 2001; Kilby, 2001). Web-based education (WBE) has moved on from the delivery of educational content to Web-based sites with interactive functions (Carty & Philip, 2001). Concurrently, new innovative kinds of pedagogical experiments have shifted the paradigm from teaching to learning (Pahl, 2003). As summarised by Armstrong (2001), what educators have in fact realised is that a good Web-based education theory and good education theory are one and the same; the only difference is that WBE transcends the barriers of space and time. The paradigmatic shift has occurred as part of planned educational policy, while at the same time good international or national experiences have also supported the growth of WBE. In addition, there have been attempts to have more coherent and cohesive educational systems and degrees especially in the European context (The Bologna Declaration, 1999.)

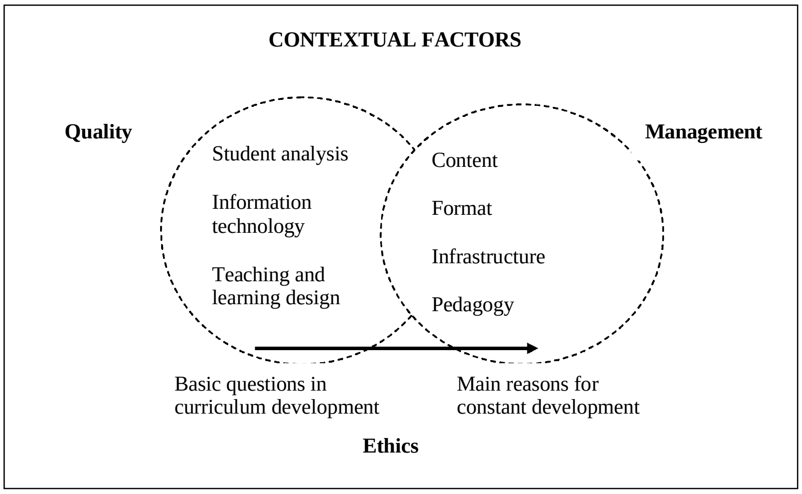
The aim of this article is to pursue the discussion of some essential issues and promoting factors facing Web-based curriculum development (Figure 1). At the beginning, the main concerns in curriculum development are quite often related to students, new technology and pedagogical issues. However, the curriculum development is a process due to constantly evolving information technology and changes in course contents. The second part of this article focuses on this. Additionally, curriculum development does not happen in a vacuum. Therefore, quality, ethics and management are briefly summarized as important contextual concerns in WBE curriculum development.

BACKGROUND

Curriculum Development in Web-Based Education

Basic questions at the first phase of the WBE curriculum development have been summarized in this review into

Figure 1. Curriculum development as a continuous process



three overall themes: 1) student analysis; 2) information technology; and 3) teaching and learning design (adapted from Alexander, 2001).

Student Analysis

One crucial component in curriculum development is the identification of potential users, and analysis of their needs (Karuppan, 2001; Lammintakanen & Rissanen, 2003). Variables such as age, gender, being employed or unemployed are premised as having an effect on computer use (Karuppan, 2001; compare to Lu, Yu, & Liu, 2003). Furthermore, learning materials should support the student's particular learning style in order to facilitate learning (Karuppan, 2001). At best, Web-based education encourages the student to take control over his or her own learning. In turn, curriculum development should support this by promoting a completely new way of thinking in students: from what they hope to acquire from the course to what they themselves contribute to the creation of knowledge (McFadzean & McKenzie, 2001).

Although Web-based learning is said to be a flexible way of learning in terms of availability (anywhere and anytime), it is crucial to take into consideration the place from where the students participate, for example, the home or work place, and also the kind of skills that they have.

The following reasons have been recognized as major obstacles in students' use of information and communication technology (ICT): 1) a lack of student workstations; 2) students' lack of time; 3) students' ICT skills; 4) course overlap; 5) insufficient course hours; and 6) teachers' lack of time (Sinko & Lehtinen, 1999). Careful consideration of these aspects provides an idea of what kinds of learner support systems are needed from the educational institution (e.g., tutoring, technical support; Lammintakanen & Rissanen, 2003). In sum, support from the educational organization and other students, as well as the individual's experience of technology have a major influence on the student's learning experiences (Alexander, 2001).

Information Technology

Similarly, fundamental questions exist concerning the technology used in education. One of the problems concerns the interaction between the equipment used, for example, is technology available to students, and is it accessible especially if interactive text, video and voice are combined? What does the technology cost learner? This is something that should especially be taken into consideration in those countries where tuition fees are not implemented (e.g., Lammintakanen & Rissanen, 2003). Web-based education is supposed to be cost-effective from the organizational side (e.g., Karuppan, 2001), although hardware, software and labour costs are somewhat expensive. Similarly, it would be imperative that the equipment be available for both the students and educators "just on time" because of the rapid development of information technology (IT).

Teaching and Learning Design

Although the choices made during the planning process determine whether the Web-based education is based on constructivism or other learning theories, constructivism is usually closely related to WBE (Jefferies & Hussain, 1998). Web-based education is believed to promote a constructivist approach by allowing all-round interaction, transferring the responsibility of learning to the student, and enhancing the construction of knowledge by interaction. During the curriculum development stage, a careful evaluation is needed on whether or not the chosen technology supports teaching strategies that encourage active involvement, critical thinking, and fosters relationships between learners (Armstrong, 2001). It has, however, been shown that although teachers have adopted the model of constructivist epistemology in principle, they have not always implemented it in the ways they organize the learning situations (Sinko & Lehtinen, 1999).

The changing roles of teachers are obvious in WBE. Previous research has shown that the role of the teacher is not diminished, however, traditional teacher duties, such as instructing the learners and information communication, are. The teacher's new role can be described as a learning catalyst and knowledge navigator, or as tutor acting as a facilitator for learning and group processes (see Volery & Lord, 2000). Moreover, the tutor's duty is to maintain a safe environment for learning, and encourage novel problem solving processes (McFadzean & McKenzie, 2001). Web-based learning forces teachers to become course designers who make decisions based on their understanding of the probable needs, expectations and behaviors of students on their own campuses (Blythe, 2001).

To briefly summarize, previous studies have shown that the technology affects learning in many ways (e.g., McFadzean & McKenzie, 2001; Sinko & Lehtinen, 1999). Curriculum development is a time-intensive requiring adequate financial resources in order to develop tightly organised courses. In addition, it is to be expected that the faculty workload would increase (Armstrong, 2001; Carty & Philip, 2001). Unfortunately, too many WBE applications are still mere tutorials or online books (Kilby, 2001).

C

Table 1. The factors of change promoting curriculum development (Pahl, 2003)

Content	The course subject evolves Changes in content to improve the material
Format	Changes in • Staff • Student body (qualifications, numbers, mode of learning) • Timetable (where and when the course takes place) • Syllabus (the content and organization of the course) • Curriculum (level, extent, prerequisites) • Legal and/or financial environment
Infrastructure	Improvements in hardware technology Systems and language technology face constant minor changes Learning devices are developing
Pedagogy	Knowledge acquisition, modeling of and access to educational knowledge Active learning in terms of engaging the student through interactive systems Collaborative learning supportive systems Autonomous learning Evolving instructional design

Curriculum Development as a Constant Process

The focus of the Web-based education and especially curriculum development has taken a step forward during the last years. However, while a substantial body of research has focused on this new teaching medium, the results have been mixed and with no significant improvement over traditional methods. Additionally, there is a lack of systematic and scientific knowledge, especially with regard to the effects and outcomes of WBE (Karuppan, 2001; Lu et al., 2003; Reisman, Dear, & Edge, 2001). Still, the use of WBE has rapidly increased.

Previously, the challenges of curriculum development focused merely on the lack of skills and suitable equipment for WBE. Often flexibility and maintenance aspects have been neglected in the design and development of new technologies. In fact, while in previous stages, the main concerns were in the planning processes (i.e., how to begin with WBE); nowadays, important questions are how to update the curriculum and develop it further. The curriculum development can be based on students' evaluations, experiences of others as well as previous studies (e.g., Lammintakanen & Rissanen, 2003). But, there exits also factors that "force" us to develop the curriculum. For example, Pahl (2003) has summarized both internal and external reasons why the curriculum needs to be constantly evolved (Table 1). The evolvement of the design of a Web-based course can be due to four dimensions: content, format, infrastructure, and pedagogy.

FUTURE TRENDS

The Context of Curriculum Development

In previous sections, the WBE curriculum development has been described from the pragmatic perspective. However, some existing and concurrently future challenges that have not been mentioned yet have become more important during the curriculum development process. These themes have been summarized into three interrelated categories: 1) ethical and legal issues; 2) quality assurance and accreditation; and 3) managerial and organizational issues. The challenges arose partly from the literature, and partly from practical experiences (e.g., Alexander, 2001; Kilby, 2001; Roffe, 2002), and concern both national and international curriculum development, since a more global perspective in design and courseware provision is expected (Kilby, 2001; see e.g., MIT Open Courseware, Potts, 2003).

Ethical and Legal Issues

Web-based education provides good opportunities to make use of the many information sources available via the Internet (Jefferies & Hussain, 1998). The problem is, however, the quality of knowledge: how to select appropriate information from among the mass, and how to avoid the use of misinformation (Calvert, 1999). At its worst, the Internet is a tool towards fabrication, falsification and plagiarism, which leads to copyright considerations. In addition, from the ethical perspective, students' privacy and confidentiality, and their respectful and dignified treatment on the Web-based environment are imperative (Armstrong, 2001).

Quality Assurance and Accreditation

Quality is a crucial concern in WBE: there are no common quality standards for course design, delivery and evaluation, nor is there is an accreditation system. Some institutions and countries (e.g., the UK) have developed quality assurance protocols that demonstrate that the online programmes are of equal quality to those delivered by traditional methods (Roffe, 2002). From the students' point of view, the important question is how the courses can be accepted as part of the curriculum; how different educational institutes recognise the courses offered from other institutions nationally and internationally.

One quality issue concerns the lack of information bases (portals, registers), which include information on Web-based courses, learning materials or tools for learning that are available on the Internet. In terms of curriculum development, this kind of meta-knowledge would be very useful, and could potentially even reduce the teacher's workload. But, teachers are not very keen to share material, which they themselves have made, and experiences with colleagues across the educational institutions. However, MIT Open Courseware is an exception, and it is open and available all over the world. MIT offers a standardized process for course modelling and encourages extending collaboration and interdisciplinary teaching (Potts, 2003). Registers would also provide students with the information on the available Web-based courses.

Managerial and Organizational Issues

The tradition of individualism in teaching is still part of the organizational culture at different educational institutions. At best, the organizational culture can support both Web-based curriculum development and joint teaching. The allocation of both human and technical resources, and a clear strategic decision from managers are requirements for faculty development, and the incorporation of WBE into the curriculum (Carty & Philip, 2001). As well as time and resources, it is essential that the organization and its managers have a positive attitude to WBE, and that they promote its implementation (e.g., Alexander, 2001). However, the managerial approach to lead teaching and curriculum development work is not yet visible in different educational organizations.

CONCLUSION

In conclusion, there is no consistent paradigm for WBE, rather there are multiple ways of making use of the Web in education, and these will vary depending on the subject being taught and the needs of the learner. Curriculum development requires a great deal of effort from different stakeholders, and therefore, motivation and commitment for long-term WBE development strategies are also needed at organisational level. In addition, Web-based education appears to support cultural cohesion and more rapid information transformation. Standardization and registers are an essential part of future steps towards the improved utilization of WBE requiring international cooperation (see ISO/IEC, which has developed standards for information technology for learning, education and training, http://jtsc36.org).

REFERENCES

Alexander, S. (2001). E-learning developments and experiences. *Education and Training 43*(4/5), 240-248.

Armstrong, M.L. (2001). Distance education: Using technology to learn. In V. Saba & K.A. McCormick (Eds.), *Essentials of computers for nurses – Information for the new millennium* (3rd ed.) (pp.413-436). McGraw-Hill.

Blythe, S. (2001). Designing online courses: User-centred practices. *Computers and Composition, 18*(4), 329-346.

The Bologna Declaration (1999). Retrieved on October 24, 2003 from *http://www.minedu.fi/opm/koulutus/yliopistokoulutus/bolognaprosessi.html*

Calvert, P.J. (1999). Web-based misinformation in the context of higher education. *Asian Libraries, 8*(3), 83-91.

Carty, B., & Philip, E. (2001). The nursing curriculum in the Information Age. In V. Saba & K.A. McCormick (Eds.), *Essentials of computers for nurses – Information for the new millennium* (3rd ed.) (pp.393-412). McGraw-Hill.

Harasim, L. (2000). Shift happens. *Online Education as a New Paradigm in Learning. Internet and Higher Education, 3*(1-2), 41-61.

ISO/IEC JTC1 SG36. Retrieved on March 31, 2004 from *http://jtc1sc36.org/*

Jefferies, P., & Hussain, F. (1998). Using the Internet as a teaching resource. *Education + Training, 40*(8), 359-365.

Karuppan, C.M. (2001). Web-based teaching materials: A user's profile. *Internet Research: Electronic Networking Applications and Policy, 11*(2), 138-148.

Kilby, T. (2001). The direction of Web-based training: A practitioner's view. *The Learning Organization, 8*(5), 194-199.

Lammintakanen, J. & Rissanen, S. (2003). An evaluation of Web-based education at a Finnish University. In A. Aggarwal, (Ed.), *Web-based education. Learning from experience* (pp.440-453). Hershey, PA: Information Science Publishing.

Lu, J., Yu, C-S., & Liu, C. (2003). Learning style, learning patterns and learning performance in a WebCT –based MIS course. *Information & Management, 40*(6), 497-507.

McFadzean , E., & McKenzie, J. (2001). Facilitating virtual learning groups. A practical approach. *Journal of Management Development, 20*(6), 470-494.

Pahl, C. (2003). Managing evolution and change in Web-based teaching and learning environments. *Computers and Education, 40*(2), 99-114.

Potts, J.P. (2003). A new model for open sharing. A presentation in *WCET Annual Conference*, November 5th, 2003. Retrieved on March 31, 2004 from *http://ocw.mit.edu/OcwWeb/index.htm*

Reisman, S., Dear, R.G., & Edge, D. (2001). Evolution of Web-based distance learning strategies. *The International Journal of Educational Management, 15*(5), 245-251.

Roffe, I. (2002). E-learning: Engagement, enhancement and execution. *Quality Assurance in Education, 10*(1), 40-50.

Sinko, M., & Lehtinen, E. (1999). The challenges of ICT in Finnish education. Helsinki: Atena. Electronic Publication available via the Internet: *http://www.sitra.fi/eng/index.asp?DirID=297*

Volery, T., & Lord, D. (2000). Critical success factors in online education. *The International Journal of Educational Management, 14*(5), 216-223.

KEY TERMS

Asynchronous Mode: A non-real-time education where students and teachers interact with each other but no at same time, e.g., by using bulletin board or email.

Computer and Information Literacy: The abilities to perform computer operations at a skill level high enough to meet the demands of the society, and to use the tool of automation in the process of accessing, evaluating and utilizing information (Carty & Philip, 2001).

Learning Style: The way in which individuals acquire and use information, strategies to process information in learning, and problem solving situations (Karuppan, 2001).

Learning Tools: Included in web-based learning environments for managing the course and are geared to facilitating student learning in the environment.

Portal: Acts as a channel between the content creator and end user. It does not typically have a content of its own.

Synchronous Mode: Real-time education where students and teachers can have a dialogue simultaneously, e.g., by using chat.

Web-Based Education (WBE): Differs from traditional classroom teaching in two essential elements: 1) physical distance and 2) time, allowing more flexibility to the learner. The most basic form of WBE is to deliver syllabi, lecture notes, reading materials and assignments via the Internet. The more advanced level includes computer conference facilities, a help desk, linkage of conferencing and web page assignment, testing and course management tools and evaluation (Karuppan, 2001).

Web-Based Learning Environment: A specially developed programme using Internet technology for the design and development of teaching and learning purposes. Trademarks are, for example, WebCT, WebBoard, Top Class, Virtual – U.

Web-Based Misinformation: Used to describe information found in the Internet that does not fit normative patterns of "truth", i.e., it is incomplete, out-of-date, confused, or low consensus "knowledge" (Calvert, 1999).

Data Collection Methodologies for Web-Based Experiments

Stu Westin
University of Rhode Island, USA

INTRODUCTION

Experimental studies involving the use of the World Wide Web (WWW) are becoming increasingly common in disciplines such as management information systems (MIS), marketing, and e-commerce. The focus of these studies is varied and may involve issues of human factors and interface design (Otto et al., 2000; Koufaris, 2002; Liang & Lai, 2002; Palmer, 2002), issues of information processing and search strategies (Spence, 1999; Johnson et al., 2000; Xia & Sudharshan, 2000; Chiang et al., 2004), issues of vendor trustworthiness (Grazioli & Jarvenpaa, 2000; Jarvenpaa et al., 2000; Norberg, 2003), or a myriad of other topics. Regardless of the issue being studied, data collection for online Web research often proves to be a vexing problem, and ideal research designs are frequently sacrificed in the interest of finding a reasonable data capture mechanism. In this article, we discuss some of the methodological complexities that arise when conducting Web-based experiments. We then describe an innovative, software-based methodology that addresses these problems.

BACKGROUND

Despite the difficulties involved, the research community has recognized the importance of Web-based experimental research (Saeed et al., 2003; Gao, 2003).

Server-side data collection approaches based on active server page (ASP) scripts or the like can prove useful in some circumstances. Consider, for example, the situation where one wants to investigate the impact of download time. This has been studied, for example, as a factor affecting user satisfaction (Otto et al., 2000), and Web site success (Palmer, 2002). Using ASP scripts, a delay mechanism can be easily built into a Web page so that the server will delay serving the requested page to the client until some precise, predetermined time has passed. Different experimental treatment levels are accomplished by merely manipulating the delay time that is scripted into the Web page. Here, the experimental subject, using an ordinary browser, will have the perception that the page is slow to download because of the delay between when the page is requested (e.g., by clicking a hyperlink) and when the page is available in the browser.

As another scenario, consider the situation where the researcher wants to study the end user's Web search strategy (e.g., Jansen et al., 2000; Johnson et al., 2000; Chiang et al., 2004) by considering which pages are accessed, along with the sequence of page access. In this case, we can analyze the so-called *click-stream data*. The research community has, in fact, called on investigators to exploit "non-intrusive means of collecting usage and exploration data through click streams" (Gao, 2003, p. 31) in future Web studies. Here again, ASP scripts in the Web pages can provide a simple data collection mechanism by logging each page request (page ID, server time stamp) in a server database. In both of the above research scenarios, standard Web browser software such as Internet Explorer (IE) can be used in the experiment.

In considering these research situations, it is obvious that client-side data collection mechanisms can be constructed just as easily. In both cases, Java applets, Java scripts, or VB scripts can be embedded into the HTML pages to handle the required tasks, and, again, standard browser software can be used. The only difference in this client-side approach is that the data collection is being handled by the client rather than by the server machine. Neither approach provides any obvious benefits over the other, although in the client-side approach, the Web pages could be stored locally, and thus WWW, or even network access, is not required.

One flaw in all of these research scenarios lies in the fact that experimental access must be restricted to a limited set of Web pages that have been appropriately scripted for data collection. If the experimental subject is allowed to "wander" beyond this limited set of pages (an activity that is fundamental to the nature of the Web), then these actions will not be recorded, and the validity of the experiment will be nullified; there will simply be no data collection scripts to execute. Related to this is the fact that all Web pages used in the experiment must be developed and maintained by the investigator—a task that can be labor intensive if a large number of pages are to be made available. Obviously, the experimental pages should be large in number and professional in appearance if external validity is to be maintained.

In some situations, the research data can be collected without the use of client- or server-side scripting. Click-stream data, for example, can often be gleaned through the use of standard network management software, or through *network sniffers* that can be configured to monitor Internet requests and page downloads. In this case, the experimental treatment can involve pages other than those created specifically for the research study, and, again, standard browser software can be used for the experiment. The problem here can be in the precision or in the format of the data, as the software was not designed for this purpose. Pages containing multiple frames, for example, may be logged as individual (frame) downloads in some circumstances and as a single-page download in others. Client requests that are satisfied through the local cache may not be logged at all.

A problem with all of the data collection methodologies discussed thus far is that they suffer from a lack of experimental control. This lack of control comes from the fact that the instrument with which the experimental subject is interacting (a standard Web browser such as IE) was not designed to be used as a research tool.

Consider the situation in which we wish to study WWW use behavior through analyzing click-stream data. There are ways of gathering data on page requests or page downloads, as noted above. However, there is no means, short of direct visual observation, of recording how a particular page was requested. The page request could have come in the form of a click on a hyperlink, but the request could just as likely have been generated automatically through a dynamic action on the page (e.g., *meta refresh*), or through the *Back* or *Forward* buttons in the browser interface. Normal click-stream data will not distinguish between these circumstances, so the precise behavior or intentions of the experimental subject cannot be determined.

Another problem has to do with the occurrence of multiple windows. Many Web sites open hyperlinks in new browser windows. The problem here is that the data collected cannot reflect which of the open windows is active when actions occur, or even that there are multiple windows in use. Again, the data cannot capture or misrepresents the behavior in question; true *streams* cannot be traced.

A CLIENT-SIDE SOLUTION

As noted earlier, the methodological problems, for the most part, stem from a lack of experimental control. Logic and research experience suggest that, for maximum experimental control, the experimental manipulations (treatments) and the data collection mechanisms should be as close to the experimental subject as possible. That is, they should be embedded in the browser. This leads to the development of a custom IE-lookalike browser for use in Web-based experiments. The creation of such a software application is feasible with currently available programming tools and software techniques. The numerous benefits of this approach certainly outweigh the software development costs. The benefits are greatest when research designs are complex and when precision is of prime importance. This particular methodology has been employed in several research studies, including Norberg (2003) and Chiang et al. (2004).

With custom browser software, there is no need to depend on scripts or applets in experiment-specific Web pages to administer experimental treatments or to record user actions. Consequently, there is no need to restrict the experimental domain to a limited set of custom Web pages. With this approach, the experimental domain can include the entire Web. The custom software can be built with the ability to precisely record user activity and to preempt or modify actions that could be harmful or inappropriate in the experimental context. Experimental control and experimental manipulation can be integrated into the browser.

The software that we know as Internet Explorer is essentially a software interface surrounding a set of dynamic link libraries (DLLs) that provide the requisite Internet processing functionality. Microsoft, in its *Visual Studio* suite of software development products, provides a software object called the *WebBrowser Control* (see Microsoft Corporation, 2004a). This control can be used in Visual Basic (VB) or in C++ applications to add Web browsing functionality to software applications. The *WebBrowser Control* can be thought of as an object wrapper around the Internet-processing DLLs noted above. The *WebBrowser* object works with the standard event-based model of Windows computing.

With the *WebBrowser* control, event handlers are provided for all of the major occurrences in an Internet session, such as *request to navigate to a page*, *page download complete*, or *request for a new window*. Key data such as URL, Target Frame, and Page Title are available with the events. In some cases, actions can be preempted through a *Cancel* argument in the event handler. One important example of this is the *BeforeNavigate* event handler. This routine is triggered after a navigation has been requested by the client, but before the request is fulfilled. This allows the custom software to inspect and evaluate the situation and to possibly modify or cancel the request before it is allowed to proceed.

Properties and methods of the *WebBrowser* object can be used to dynamically emulate all of the features of the IE interface, such as the status bar, the browser window caption, and the standard buttons (Back, Forward, Stop,

Table 1. Sample Web browser members

Event/Method	Behavior/Details	Possible Use
BeforeNavigate2 event	Triggered after page request, but before navigation begins; provides target URL and frame information; allows cancellation of navigation	Can analyze URL request and covertly cancel or redirect inappropriate requests; software can contain a list of irrelevant URLs or heuristic rules (e.g., requests containing forbidden protocols or Web domains can be halted) (e.g., no returns to prior pages)
Navigate2 method	Forces browser to navigate to a new location (URL)	Can be used to covertly modify target for (cancelled) navigation request; can be used to covertly effect programmatic delay between user's request and download activity
DocumentComplete event	Triggered when a page has been successfully downloaded and is available for browsing; provides full URL details	Can be used to capture click-stream data (record URL); can be used to time stamp click-stream data and to determine and record page-viewing time
GoForward/*GoBack* methods	Forces navigation to subsequent (or previous) page in system-maintained URL history list	Can emulate full behavior of IE Forward (and Back) buttons, but this behavior can be modified if needed; note that button action can be included in click-stream data record

Refresh, Home, etc.). In short, an emulation of IE can be built with the inclusion of as few or as many features of the IE interface as are needed in the experimental context.

Table 1 describes selected events and methods of the *WebBrowser* object that are pertinent to this discussion. See Microsoft Corporation (2004b) for a full list and description of the members of this object.

Table 2 provides a sample structure of click-stream data that has been generated with this methodology. This represents one of several system-generated data tables analyzed in Chiang et al. (2004). This data table holds one record per subject per URL visited (i.e., one item in a stream). Fields 1 and 2 were generated automatically by the software and provide primary and secondary sort fields for click-stream analysis. Field 3 provides data for frame-based pages and is derived from the *BeforeNavigate2* event. Field 4 (captured in *BeforeNavigate2* event) records the time at which a page was initially requested (e.g., user click). Field 5 records the duration for which the page was available to the user (*DocumentComplete* event for current page → *BeforeNavigate2* event for next page). Field 6 indicates how this page was requested ("Automatic" indicates a redirect, as by a script action or a meta refresh). Field 7 was captured by the *DocumentComplete* event handler. By writing the data record within the *DocumentComplete* event routine, unsuccessful or aborted downloads were culled from the stream. Sequence gaps in Field 2 indicate such.

In this particular study, randomized assignment of subjects to treatment cells, as well as experimental session duration were managed completely by the software application. (After a fixed duration, the browser was programmatically disabled, and the user was shown a "thank you" page.). A second data table holding Subject_ID and experimental treatment information was later merged with the above table for analysis.

Table 2. Data record structure

	Field	Details
1	Subject_ID	Auto generated by application
2	URL_Sequence#	Auto generated by application
3	Target_Frame	If any
4	URL_Start_Time	Since session began (sec./1000)
5	URL_Duration	Sec./1000
6	User_Action	Click, Back-Button, Forward-Button, Home-Button, Automatic
7	Full_URL	Can be parsed if needed

CONCLUSION

By developing software to host a custom browser research instrument, the investigator is free to include (covertly) all of the requisite mechanisms of experimental control and data capture into the software; no external scripting or network monitoring is needed. Timers to control the duration of the experiment or the occurrence of experimental treatments can be embedded into the application. Experimental treatment randomization can also be built in. User activity down to the keystroke or mouse-click level can be monitored and recorded with millisecond accuracy if needed. Certain events can also be blocked or modified, if necessary. For example, an attempt to open a page in a new window can be intercepted and the page redirected to the initial window. No special (i.e., scripted) Web pages are needed, but attempts to "wander" to irrelevant sites or inapposite protocols (e.g., *mailto*, *ftp*) can easily be halted if desired. The cache can be controlled programmatically through calls to the Windows API. Once the basic system is developed, modifications and new features are fairly simple to put into effect.

The above discussion and summary tables touch on the basic advantages and capabilities of this custom software approach. Many of the complexities, such as the need to develop a custom, preemptive keyboard handler for the browser, are ignored here. Also beyond the scope of this discussion are many of the advanced capabilities, such as how to gain access to an additional rich source of experimental data through use of the document object model (DOM). A more thorough presentation can be found in Westin (2003).

REFERENCES

Chiang, K., Dholakia, R. R., & Westin, S. (2004). Needle in a cyberstack: Consumer search for information in the web-based environment. In B. Kahn & M. Luce (Eds.), *Advances in consumer research*. Provo, UT: Association of Consumer Research, 31 (forthcoming).

Gao, Y. (2003). Website interactivity and amusement: Techniques and effects. In S. Gordon (Ed.), *Computing information technology: The human side* (pp. 22-34). Hershey, PA: IRM Press.

Grazioli, S., & Jarvenpaa, S. (2000). Perils of Internet fraud: An empirical investigation of deception and trust with experienced Internet consumers. *IEEE Transactions on Systems, Man, and Cybernetics, 30*(4), 395-410.

Jansen, J. B., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: A study and analysis of user queries on the web. *Information Processing and Management, 36*, 207-227.

Jarvenpaa, S., Tractinsky, N., & Vitale, M. (2000). Consumer trust in an Internet store. *Information and Technology Management*, 1, 45-71.

Johnson, E. J., Moe, W., Fader, P., Bellman, S., & Lohse, J. (2000). On the depth and dynamics of world wide web shopping behavior. Paper presented at the *2nd Marketing Science and the Internet Conference, California, April 28-30*. University of Southern California.

Koufaris, M. (2002). Applying the technology acceptance model and flow theory to online consumer behavior. *Information Systems Research, 13*(2), 205-223.

Liang, T., & Lai, H. (2002). Effect of store design on consumer purchases: An empirical study of on-line bookstores. *Information & Management, 39*(6), 431-444.

Microsoft Corporation. (2004a). *WebBrowser Control*. Available from *http://msdn.microsoft.com/library/default.asp?url=/workshop/browser/webbrowser/browser_control_node_entry.asp*

Microsoft Corporation. (2004b). *WebBrowser Object*. Available from *http://msdn.microsoft.com/library/default.asp?url=/workshop/browser/webbrowser/reference/objects/webbrowser.asp*

Norberg, P. (2003). Managed profiles: The value of personal information in commercial exchange. Unpublished doctoral dissertation, University of Rhode Island, Rhode Island.

Otto, J., Najdawi, M., & Caron, K. (2000). Web-user satisfaction: An exploratory study. *Journal of End User Computing, 12*(4), 4-20.

Palmer, J. W. (2002). Web site usability, design, and performance metrics. *Information Systems Research, 13*(2), 151-167.

Saeed, K. A., Hwang, Y., & Yi, M. Y. (2003). Toward and integrative framework for online consumer behavior research: A meta-analysis approach. *Journal of End User Computing, 15*(4), 1-26.

Spence, R. (1999). A framework of navigation. *International Journal of Human–Computer Studies, 51*, 919-945.

Westin, S. (2003). Building a custom client-side research tool for online web-based experiments. In S. Gordon (Ed.), *Computing information technology: The human side* (253-266). Hershey, PA: IRM Press.

Xia, L., & Sudharshan, D. (2000). An examination of the effects of cognitive interruptions on consumer on-line decision processes. Paper presented at the *2nd Marketing Science and the Internet Conference, California, April 28-30*, University of Southern California.

KEY TERMS

ASP (Active Server Page) Scripting: A simple server-side scripting approach where script code (usually VBScript or Jscript) is mixed with HTML code on a Web page. The script code is processed by a script engine before the page is rendered by the server. This can be used to create dynamic Web pages and to share data within or between Web sessions. This is a predecessor of ASP.NET technology.

Click-Stream: In Web research, the click-stream is the sequence of Web pages that is visited by the experimental subject. A click-stream data record can be as simple as URL and sequence number, or a time stamp can be added. This latter approach allows for analysis of page viewing time.

Client-Side/Server-Side Scripting: In a Web environment, this term relates to the fact that scripted tasks can be handled by the browser software (client side) or by the Web server software (server side). A single Web page may contain both client-side and server-side scripts. The script host is determined by the *RUNAT* attribute of the *SCRIPT* tag.

Event Handler: A procedure (subroutine) that executes in response to an event. The event may represent a specific user action (e.g., a mouse click) or may be a manifestation of a system process (e.g., page has finished loading). Details surrounding the event are provided as arguments of the procedure.

Meta Refresh: A special HTML tag that redirects the visitor to a new page. The result is that the final destination of the navigation is different from the initial target.

Method: In object-oriented programming, methods are the actions or behaviors that an object can perform. At the coding level, a method is created by including a procedure (function or sub) within the class.

Network Sniffer: A hardware or software mechanism that monitors, and possibly records, data traffic on a network.

Object Wrapping: Adding code to convert an existing set of software procedures into an object class. The procedures can then be used as methods in an object-oriented software environment.

Data Communications and E-Learning

Michael W. Dixon
Murdoch University, Australia

Johan M. Karlsson
Lund Institute of Technology, Sweden

Tanya J. McGill
Murdoch University, Australia

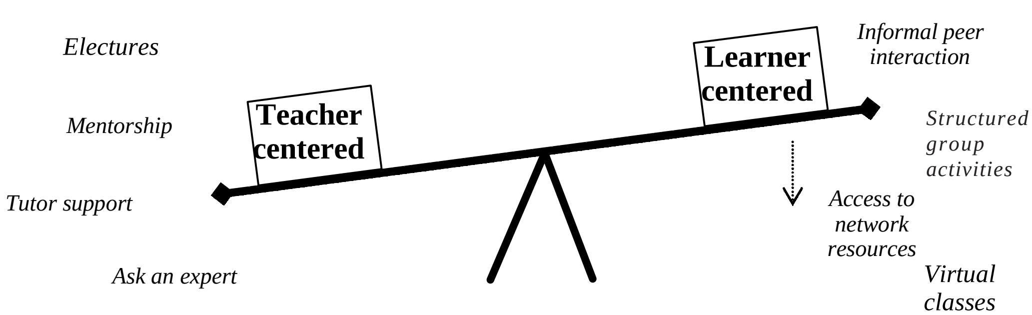
INTRODUCTION

Information and communications technology (ICT) has increasingly influenced higher education. Computer-based packages and other learning objects provide a useful supplement to students studying conventionally by illustrating aspects of the curriculum. Other packages are directed at aspects of course administration such as automated assessment (for example, see Randolph et al. (2002)). Initially such software and materials played only a supplementary role in course offerings, but this has changed rapidly. For example, Coleman et al. (1998) describe a successful early attempt to replace all lecturing with computer-aided learning. Remote delivery of courses also became a viable option because of the advent of the WWW. For example, Petre and Price (1997) report on their experiences conducting electronic tutorials for computing courses. Online education of various sorts is now routinely available to vast numbers of students (Alexander, 2001; Chen & Dwyer, 2003; Peffers & Bloom, 1999). Various terms have been used to label or describe forms of education supported by information technology. These include e-learning (e.g., Alexander, 2001; Campbell, 2004), Web-based learning (e.g. Huerta, Ryan & Igbaria, 2003; Khosrow-Pour, 2002), online learning (e.g., Simon, Brooks & Wilkes, 2003), distributed learning and technology-mediated learning (e.g., Alavi & Leidner, 2001); with e-learning probably the most commonly used term used to describe education and training that networks such as the Internet support.

E-learning has become of increasing importance for various reasons. These include the rise of the information and global economy and the emergence of a consumer culture. Students demand a flexible structure so that they can study, work and participate in family life at the same time (Campbell, 2004). This flexibility is reflected in alternative delivery methods that include online learning and Internet use. We have also become more sensitive to cultural and gender differences, and to the learning needs of the challenged. These needs may be addressed by e-learning (Campbell, 2004).

A number of studies have compared student learning and satisfaction between e-learning and traditional classroom teaching. In an early study, Hiltz and Wellman (1997) found that mastery of course material was equal or superior to that in the traditional classroom and that e-learning students were more satisfied with their learning on a number of dimensions. In particular, they found that the more students perceived that collaborative learning was taking place, the more likely they were to rate their learning outcomes as superior to those achieved in the traditional classroom. They did however identify some disadvantages to e-learning. These included ease of procrastination and information overload. More recently, Piccoli,

Figure 1. Categories of online activities



Ahmad and Ives (2001) found that the academic performance of students in the two environments was similar, but that while e-learning students had higher levels of self-efficacy, they were less satisfied with the learning process. Alexander's comment that "the use of information technology does not of itself improve learning" (Alexander, 2001, p. 241) perhaps highlights the fact that e-learning can be many things and that the intention to introduce e-learning is no guarantee of success.

The different types of teaching and learning activities that are made possible by the Internet are shown in Figure 1. Harasim and Hiltz (1995) divided these activities into two categories: learner or teacher centered. There is, however, no common agreement about which category is the best and many researchers argue for a mixture of learning activities, emphasizing group learning (Bento & Schuster, 2003; Klobas & Renzi, 2003). At the moment there still seems to be an overemphasis on teacher centered approaches, which hopefully will slowly change as a better knowledge of e-learning develops.

BACKGROUND

This article provides an illustration of blended e-learning by describing how we deliver and manage courses in a postgraduate degree in telecommunications management. We aim to foster learner centered education while providing sufficient teacher centered activities to counter some of the known concerns with entirely learner centered education. We use the Internet as the communication infrastructure to deliver teaching material globally and Lotus LearningSpace to provide the learning environment. While the primary aim of our approach is to enhance the student learning process, there are also other incentives that are consistent with this. The university is able to attract a more diverse range of students – those requiring flexibility of study and the other benefits of e-learning. Thus initiatives of this type can benefit the university while meeting the additional needs of students that are discussed in the introduction.

The use of learning and content management systems (LCMS) such as Blackboard, WebCT and Lotus LearningSpace have made e-course development less onerous for faculty. These systems provide a set of tools for publishing, communicating, and tracking student activity. Various guidelines have been suggested for evaluating and choosing software for e-learning (Klobas & Renzi, 2000). After establishing our requirements for a software tool for developing and delivering courses online, we evaluated various alternatives. The requirements that we identified included:

- Instructors should not have to program and debug HTML code;
- All courses should have the same professional look and feel without having to hire computer programmers to write special software, and students should always be presented with the same interface for all their courses;
- The software should be fully integrated (one software package should allow the instructor to do everything required, such as course development and course management);
- Professional support.

After evaluating various alternatives we choose Lotus LearningSpace (LS). Successful use of LS by instructors proved to be significantly less dependent on the technical knowledge of the instructor than was the case with some other popular LCMS. It allows the instructor to focus on the learning of the students rather than on creating and debugging HTML.

LS provides instant feedback to the students and instructor, and enables progress and problems that students encounter as they go through the curriculum to be monitored. Students also have a discussion area where they can ask questions and communicate with the instructor as well as with other students.

LS allows us to create distributed courses that students and instructors can access whether they are online or off-line. Students are able to download material for a course onto their machine so they can go through the curriculum without having to have a direct Internet connection. Using the offline access method makes it easier for students to learn wherever they are located and for instructors to develop and manage course material and reduce critical network bandwidth requirements. Features that facilitate flexible student centered learning include:

- Schedule - provides students with a structured approach to assignments, materials, and assessments. Through the schedule, students can link to everything required to complete their course.
- MediaCenter - allows immediate and searchable access to all materials for the course as the instructor makes them available.
- CourseRoom - provides a discussion group facility, which hosts collaborative interchange between student groups and/or students and instructors.
- Profiles - helps students and instructors get to know their classmates to form productive teams and to network outside the course.

Features that facilitate course management include LS Central for course management and the Assessment

Manager for setting up and tracking of students' progress in their courses.

USING LOTUS LEARNINGSPACE TO ENHANCE LEARNING OF DATA COMMUNICATIONS

Depending on course content and pedagogy the migration to e-learning can be more or less laborious. We have chosen one course in a telecommunication management degree to provide an example of how e-learning was implemented. Appropriately, this course is about networking technology.

The Data Communications Systems course provides an introduction to networking and networking devices focusing on Local Area Networks (LANs) and Wide Area Networks (WANs). Network design, Ethernet, ISDN, Frame Relay, and TCP/IP are introduced, including IP addressing and router protocols. There is a strong practical context that provides students with the opportunity to build, configure and problem solve in a multi-router network environment. This course also includes the first part of the Cisco Certified Network Associate (CCNA) curriculum. The Data Communications Systems course provides a good example of a difficult class to integrate with LS because of the external curriculum, which is contained on a separate Web server, and the external assessment requirement for students.

Students are required to log in to LS once they have selected the course. Students can access LS with a Lotus Notes Client or through a Web browser. The Lotus Notes Client allows students to work off-line by placing an image of what is on the LS server on their laptop or PC at home. This allows students to work while they travel and can reduce the amount of time they are required to be connected to the Internet. When they connect up to the Internet they can resynchronize their copy with the master copy located on a Lotus Notes Server; this flexibility is a major goal of e-learning. The students then use LS schedule to follow the schedule for the course (see Figure 2). Through the schedule, students can access the external curriculum, which resides on a separate Web server.

The course takes a blended learning approach with both face-to-face classes and online learning. This enables us to take advantage of the benefits of technology mediated learning, but does not risk losing the strengths of face-to-face teaching. Theory is presented to students in the form of online topics (Web-based), mini lectures, and laboratories. LS is used to integrate course material from a variety of sources, for example, the Cisco material that contributes to the CCNA. This kind of external mate-

Figure 2. Sample schedule screen

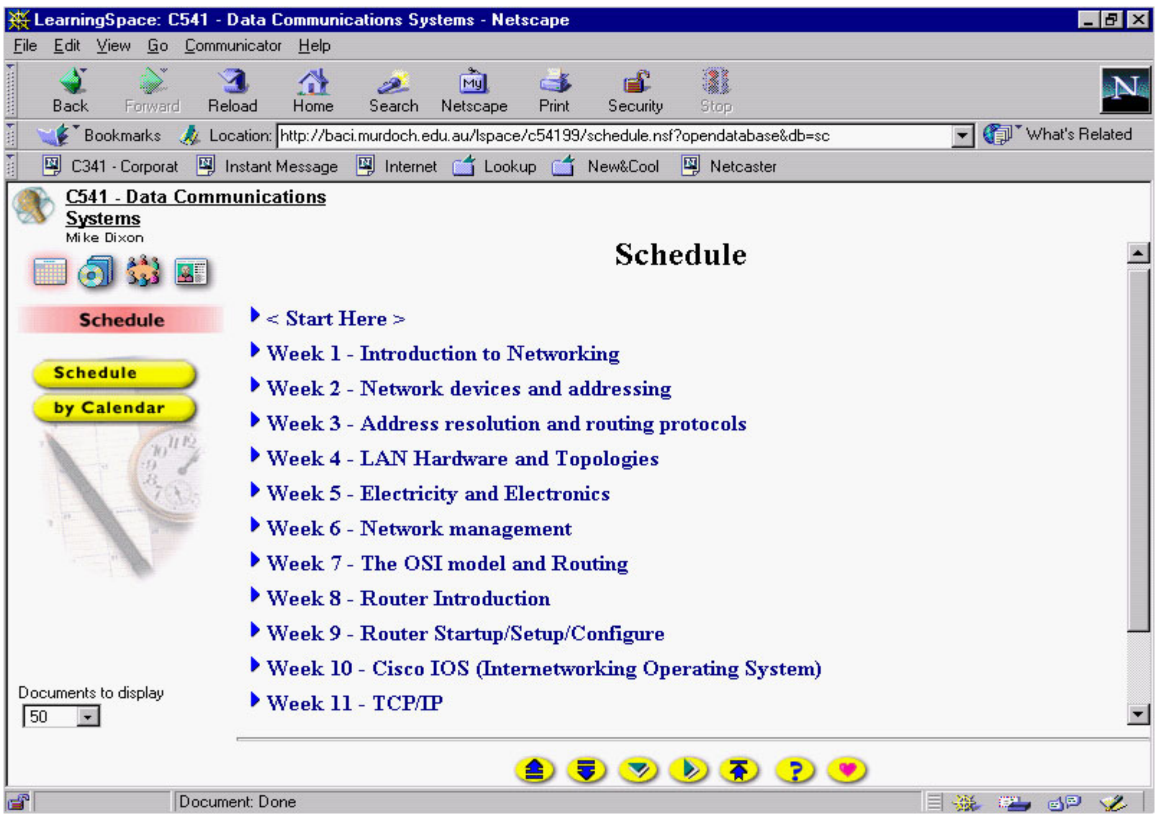
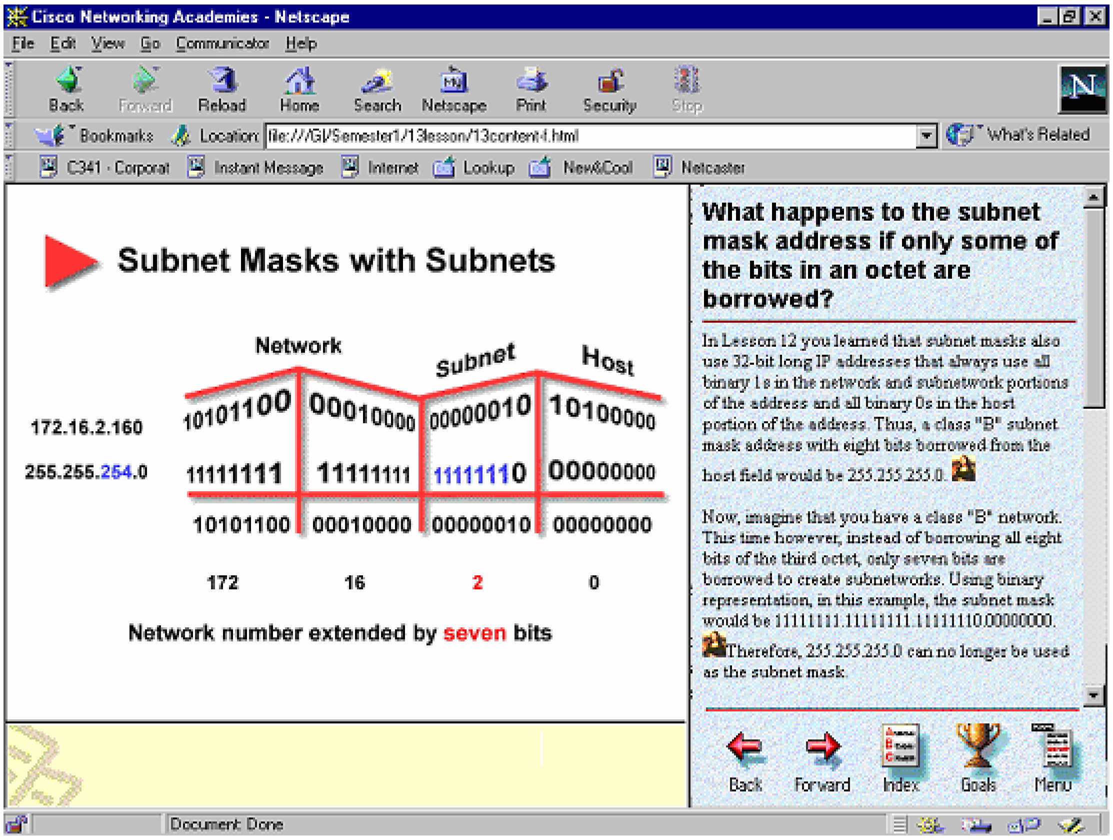


Figure 3. Sample course material screen



rial must be integrated seamlessly with local content so that students see a totally integrated system when it comes to the delivery of each course. The online teaching material combines Web based text and graphics to explain concepts (see Figure 3 for a sample screen). Short movies are also used to illustrate concepts that are difficult to communicate with just static text and graphics. The students are given aims and objectives at the start of each topic. The teaching material covers these topics in detail and at the end of each topic students have optional self-assessment quizzes that allow them to gauge their understanding of the material. The multiple modes of content delivery cater to different cognitive styles of students.

Instructors also use these online quizzes to measure the understanding of the students before they attend classes. The instructor is able to identify students who are not keeping up to date with their work and also areas that students are having problems with. The mini lectures can then focus on the areas where students are having problems. The instructor can also discuss problem areas in the discussion group. Assignments are submitted online, graded locally and returned with the instructor's personal comments and grade.

There are two Cisco online exams for the course and students are required to pass these exams with a minimum score of 80%. Students are allowed to take the exam more than once. All students must also take a separate supervised written final exam to meet University requirements. The online Cisco exams are done on separate assessment servers, as they are part of the certification process. The Cisco Networking Academy results of all the students around the world are maintained in a database in Arizona so that instructors can analyze how their class has answered questions and compare the results of their students with those of students in other institutions around the world. The final exams for each course are taken locally in a more traditional way.

The course also includes on-campus practical work to prepare students to solve real-world networking problems. Students work in groups during the practical lab sessions. Students are given timed individual and group practical exams at the end of each major component. Students are allowed to take these exams as many times as they like but are required to pass all these exams.

These factors all contribute to facilitating student learning. Faster and more frequent feedback on the material keeps students more in touch with their progress. Testing and checkpoints with built in repetition are important for long-term retention and understanding of the material. The facility to continue to work with the teaching material until an 80% pass is achieved enhances performance. Students see important material multiple times so that their learning is reinforced and students are able to study wherever they are and still be part of the student community. The use of a virtual discussion group enhances the sense of community among the students and teachers. Combining a learner centered approach with LS allows us to achieve a quality course online.

FUTURE TRENDS

E-learning will continue to play an increasing role in education and training. Greater broadband access will enable delivery of richer content and greater interactivity. Convergence of information technologies such as notebooks, phones and television and the development of pervasive computing will provide even greater flexibility to students. Educators and students are coming to understand that learning is lifelong and that technology is a valuable tool in supporting it.

CONCLUSION

E-learning is changing the face of university education. This article discussed an approach used to adopt a learner centered environment within an Internet based degree. As the course used to illustrate the issues in the article is about telecommunications, the Internet is a very appropriate medium for instruction. Students learn about telecommunications while accessing information through the Internet. Technology can provide flexibility of learning and hence enrich the learning experience and diversify the student mix.

REFERENCES

Alavi, M., & Leidner, D.E. (2001). Research commentary: Technology-mediated learning — a call for greater depth and breadth of research. *Information Systems Research, 12*(1), 1-10.

Alexander, S. (2001). E-learning developments and experiences. *Education + Training, 43*(4/5), 240-248.

Bento, R., & Schuster, C. (2003). Participation: The online challenge. In A. Aggarwal (Ed.), *Web-based education: Learning from experience* (pp. 156-164). Hershey, PA: Information Science Publishing.

Campbell, K. (2004). *E-ffective writing for e-learning environments.* Hershey, PA: Information Science Publishing.

Chen, W.-F., & Dwyer, F. (2003). Hypermedia research: Present and future. *International Journal of Instructional Media, 30*(2), 143-148.

Coleman, J., Kinniment, D., Burns, F., Butler, T., & Koelmans, A. (1998). Effectiveness of computer-aided learning as a direct replacement for lecturing in degree-level electronics. *IEEE Transactions on Education, 41*, 177-184.

Harasim, L., Hiltz, S.R., Teles, L., & Turoff, M. (1995). *Learning networks - a field guide to teaching and learning on-line.* Cambridge, MA: The MIT Press.

Hitlz, S.R., & Wellman, B. (1997). Asynchronous learning networks as a virtual classroom. *Communications of the ACM, 40*(9), 44-49.

Huerta, E., Ryan, T., & Igbaria, M. (2003). A comprehensive Web-based learning framework: Toward theoretical diversity. In A. Aggarwal (Ed.), *Web-based education: Learning from experience* (pp. 24-35). Hershey, PA: Information Science Publishing.

Khosrow-Pour, M. (Ed.). (2002). *Web-based instructional learning.* Hershey, PA: IRM Press.

Klobas, J., & Renzi, S. (2000). Selecting software and services for Web-based teaching and learning. In A. Aggarwal (Ed.), *Web-based learning and reaching technologies: Opportunities and challenges* (pp. 43-59). Hershey, PA: Idea Group Publishing.

Klobas, J., & Renzi, S. (2003). Integrating online educational activities in traditional courses: University-wide lessons after three years. In A. Aggarwal (Ed.), *Web-based education: Learning from experience* (pp. 415-439). Hershey, PA: Information Science Publishing.

Peffers, K., & Bloom, S. (1999). Internet-based innovations for teaching IS courses: The state of adoption, 1998-2000. *Journal of Information Technology Theory and Application, 1*(1), 1-6.

Petre, M., & Price, B. (1997). Programming practical work and problem sessions via the Internet. *ITiCSE 97 Working Group Reports and Supplemental Proceedings*, 125-128.

Piccoli, G., Ahmad, R., & Ives, B. (2001). Web-based virtual learning environments: A research framework and a preliminary assessment of effectiveness in basic IT skills training. *MIS Quarterly, 25*(4), 401-427.

Randolph, G.B., Swanson, D.A., Owen, D.O., & Griffin, J.A. (2002). Online student practice quizzes and a database application to generate them. In M. Khosrow-Pour (Ed.), *Web-based instructional learning.* Hershey, PA: IRM Press.

Simon, J.C., Brooks, L.D., & Wilkes, R.B. (2003). Empirical study of students' perceptions of online courses. In T. McGill (Ed.), *Current issues in IT education.* Hershey, PA: IRM Press.

KEY TERMS

Blended Learning: E-learning used in conjunction with other teaching and learning methods.

Cisco Certified Network Associate (CCNA): A data communications industry certification.

Distributed Learning: Using a wide range of information technologies to provide learning opportunities beyond the bounds of the traditional classroom.

E-Course: Another term for an online course.

E-Learning: The use of new multimedia technologies and the Internet to improve the quality of learning.

Learning and Content Management Systems (LCMS): These systems provide a set of tools for publishing, communicating, and tracking student activity.

Learning Objects: Small (relative to the size of an entire course) instructional components that can be reused in different learning contexts. Learning objects are generally considered to be digital materials deliverable over the Internet.

Online Learning: An inclusive term for any form of learning supported by computer based training.

Pervasive Computing: Technology that has moved beyond the personal computer to everyday devices with embedded technology and connectivity. The goal of pervasive computing is to create an environment where the connectivity of devices is embedded in such a way that the connectivity is unobtrusive and always available.

Technology-Mediated Learning: Learning where the learner's interactions with learning materials, other students and/or instructors are mediated through information technologies.

Web-Based Learning (WBL): Use of Internet technologies for delivering instruction.

Data Dissemination in Mobile Databases

Agustinus Borgy Waluyo
Monash University, Australia

Bala Srinivasan
Monash University, Australia

David Taniar
Monash University, Australia

INTRODUCTION

The development of wireless technology has led to *mobile computing*, a new era in data communication and processing (Barbara, 1999; Myers & Beigl, 2003). With this technology, people can now access information anytime and anywhere using a portable, wireless computer powered by battery (e.g., PDAs). These portable computers communicate with a central stationary server via a wireless channel. Mobile computing provides *database applications* with useful aspects of wireless technology known as mobile databases.

The main properties of mobile computing include mobility, severe power and storage restriction, frequency of disconnection that is much greater than a traditional network, bandwidth capacity, and asymmetric communications costs. Radio wireless transmission usually requires a greater amount of power as compared with the reception operation (Xu, Zheng, Zhu, & Lee, 2002). Moreover, the life expectancy of a battery (e.g., nickel-cadmium, lithium ion) was estimated to increase time of effective use by only another 15% (Paulson, 2003). Thus, efficient use of energy is definitely one of the main issues.

Data dissemination (can also be called *data broadcasting*) is one way to overcome these limitations. With this mechanism, a mobile client is able to retrieve information without wasting power to transmit a request to the server. Other characteristics of data dissemination include: scalability as it supports a large number of queries; query performance which is not affected by the number of users in a cell as well as the request rate; and effective to a high-degree of overlap in the user's request. In this article, the terms data dissemination and data broadcasting are used interchangeably.

The ultimate challenge in data dissemination is to minimize *the response time* and *tuning time* of retrieving database items. Response time is the total of elapsed time required for the data of interest to arrive in the channel and the download time, while tuning time is the amount of time that a client is required to listen to the channel, which is used to indicate its energy consumption. In some cases, the response time is equal to the tuning time.

This article describes a state-of-the art development in data dissemination strategies in mobile databases. Several strategies for improving the query performance by disseminating data to a population of mobile users will be explained.

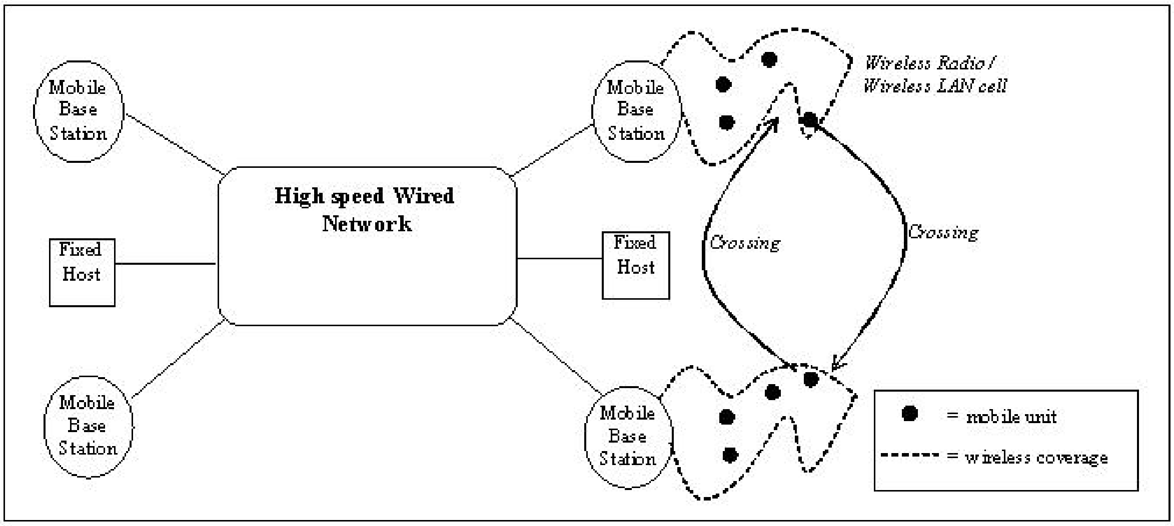
BACKGROUND

In general, each mobile user communicates with a mobile base station (MBS) to carry out any activities such as a transaction and information retrieval. MBS has a wireless interface to establish communication with the mobile client, and it serves a large number of mobile users in a specific region called a "cell". The number of mobile clients in a cell can be infinite. In mobile environment architecture, each MBS is connected to a fixed network as illustrated in Figure 1.

Mobile clients can move between cells while being active and this intercell movement is known as the handoff process (Trivedi, Dharmaraja, & Ma, 2002). Each client in a cell can connect to the fixed network via wireless radio, wireless local area network (LAN), wireless cellular, or satellite. Each of the wireless networks provides a different bandwidth capacity. However, this wireless bandwidth is too small compared with the fixed network such as asynchronous transfer mode (ATM) that can provide a speed of up to 155Mbps (Elmasri & Navathe, 2003).

Data dissemination refers to the periodic broadcasting of database items to mobile clients through one or more wireless channels (or also called broadcast channels), and the clients filter their desired data on the fly. Access to data is sequential. The behavior of the broadcast channel is unidirectional which means the server disseminates a set of data periodically to a multiple number of users. This mechanism is also known as the *push-mechanism* (Malladi & Davis, 2002; Yajima, Hara, Tsukamoto, & Nishio, 2001). It must be noted that data dissemination is different from the data replication mecha-

Figure 1. Mobile environment architecture



nism. Conventional data replication distributes a set of database items to one or more identified clients according to a pre-determined requirement. However, data dissemination broadcasts the database items periodically to an unbounded number of mobile clients, and the clients filter the data on air based on individual interest.

Figure 2 shows the mechanism of data dissemination. In this article, the term data item corresponds to database record or tuples, and data segment contains a set of data items. A complete broadcast file is referred to as a broadcast cycle. The terms mobile client, mobile computer, mobile unit, mobile user and client are used interchangeably.

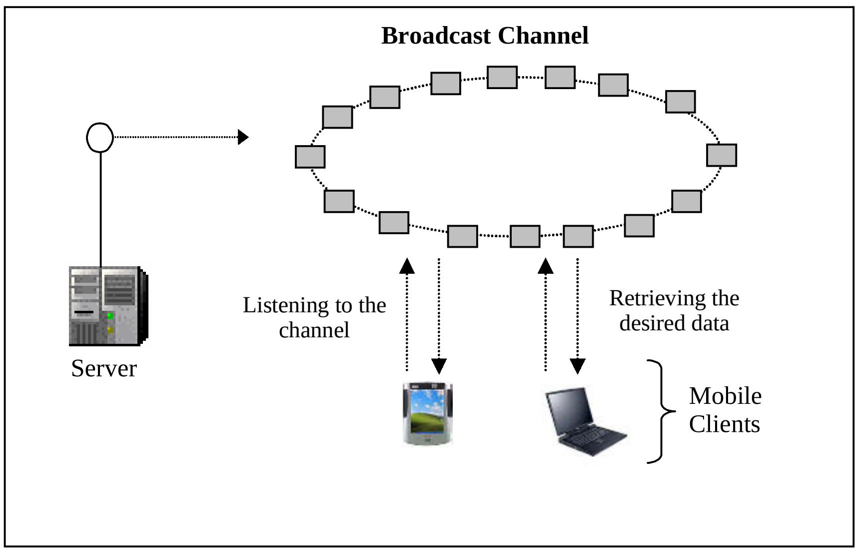
DATA DISSEMINATION

Data dissemination schemes are classified into two categories: one is to minimize query response time, and the other minimizes tuning time.

Minimizing Query Response Time

There are several data dissemination schemes, which include:

Figure 2. Data dissemination mechanism



(i) Selection of Data Items to be Broadcast,
(ii) Non-Uniform Frequency Distribution of Broadcast Data Items,
(iii) Distribution of Data Items over Multiple Channels, and
(iv) Organization of Data Items.

These schemes aim to minimize the query response time by either reducing the waiting time for the desired data to arrive, or, both waiting and download time.

a. Selection of Data Items to be Broadcast

A selection mechanism is designed to reduce the broadcast cycle length, which eventually reduces the query response time. During each broadcast cycle, additional items might be qualified as hot and some previously hot items might cool down, and therefore need to be replaced depending on the size of the cycle. A replacement algorithm is required to replace the cold data items with the new hot items. The server determines a set of appropriate database items to be broadcast, using information from queries received. Since hot items are those accessed by most clients, the access pattern of each client on the database items has to be derived from the clients back to the server. Finally, statistics will be compiled and the desired data items can be broadcast appropriately.

There are at least three replacement algorithms namely *Mean Algorithm*, *Window Algorithm*, and *Exponentially Weighted Moving Average* (*EWMA*) *Algorithm* (Leong & Si, 1997). These algorithms maintain a score for each database item to estimate the access probability of the next broadcast cycle. The scores are defined by measuring the cumulative access frequencies of each database item over the length of the observation period.

However, as the size of the database increases, the number of broadcast items may also increase accordingly. This situation will lead to an increase in response time. The next scheme can be used to improve the response time of the majority requests by manipulating the frequency of hot items to be broadcast.

b. Non-Uniform Frequency Distribution of Broadcast Data Items

The difference in bandwidth capacity between the downstream communication and upstream communication has created a new environment called the *Asymmetric Communication Environment*. In fact, there are two situations that can lead to communication asymmetry (Acharya, Alonso, Franklin, & Zdonik, 1995). One is raised as a result of the capability of physical devices. For example, servers have powerful broadcast transmitters, while mobile clients have little transmission capability. The other is due to the patterns of information flow in the application, such as the situation where the number of servers is far less than the number of clients. It is asymmetric because there is not enough capacity to handle simultaneous requests from multiple clients.

Broadcast Disk is an information system architecture, which utilizes multiple disks of different sizes and speeds on the broadcast medium. This architecture is used to address the above situations (Acharya et al., 1995). The broadcast consists of chunks of data from different disks on the same broadcast channel. The chunks of each disk are evenly scattered. However, the chunks of the fast disks are broadcast more frequently than the chunks of the slow disks. This is the opposite of a flat broadcast where the expected delay before obtaining an item of interest is the same for all broadcast items. With this differing broadcast frequency of different items, hot items can be broadcast more often than others. The server is assumed to have the indication of the clients' access patterns so that it can determine a broadcast strategy that will give priority to the hot items.

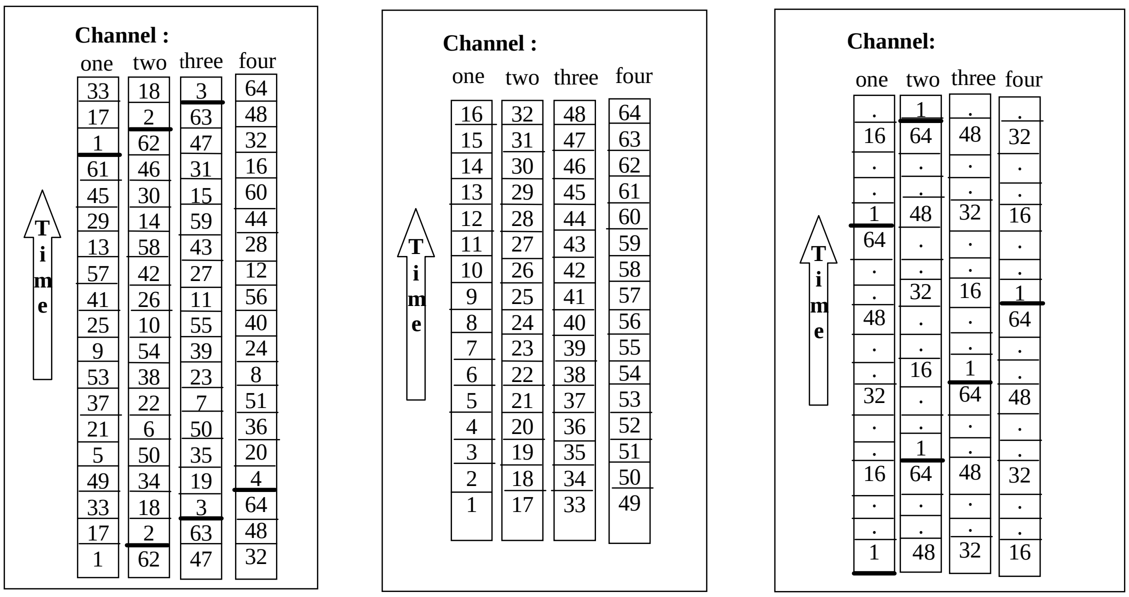
c. Distribution of Data Items over Multiple Channels

An alternative strategy to improve query response time is to distribute the broadcast data over more than one broadcast channel. Moreover, a certain pattern of distribution such as Data Stripping, Data Replication, and Data Replication can be used to handle obstacles like noise and signal distortion that may affect wireless transmission (Leong & Si, 1995). Figure 3 illustrates the three mechanisms.

As shown in Figure 3(a), the data stripping mechanism broadcasts consecutive data items over a multiple channel array and the data items are broadcast at certain intervals to allow the client sufficient time to switch from one channel to another. The data partitioning mechanism in Figure 3(b) allows the database to be partitioned into a number of data segments, and each data segment is placed in a different channel. In Figure 3(c), the database is replicated across all channels, and is therefore called the data replication mechanism.

To avoid the effect of too large data items in a channel, a strategy to determine the optimum number of database items to be broadcast in a channel is needed (Waluyo, Srinivasan, & Taniar, 2003a). In this strategy, the query response time over an on-demand channel is used as a threshold point. Subsequently, the length of the broadcast cycle is split, and broadcast over multiple channels.

Figure 3. Data stripping, partitioning, and replication mechanism



(a) Data Stripping Mechanism

(b) Data Partitioning Mechanism

(c) Data Replication Mechanism

d. Organization of Data Items

The previous schemes are concerned with retrieving a single data item. However, in most cases multiple data items are involved. Organizing database items over the broadcast channel can be applied to reduce waiting time as well as download time. Traditional semantic query optimization is employed to reduce the download time when the application involves multiple entity types, and the query accesses related entities from different entity types (Si & Leong, 1999). The organization of broadcast data is designed to match the query access pattern from mobile clients as closely as possible. As the query initiated by each client varies, this problem is sometimes considered an NP-hard problem.

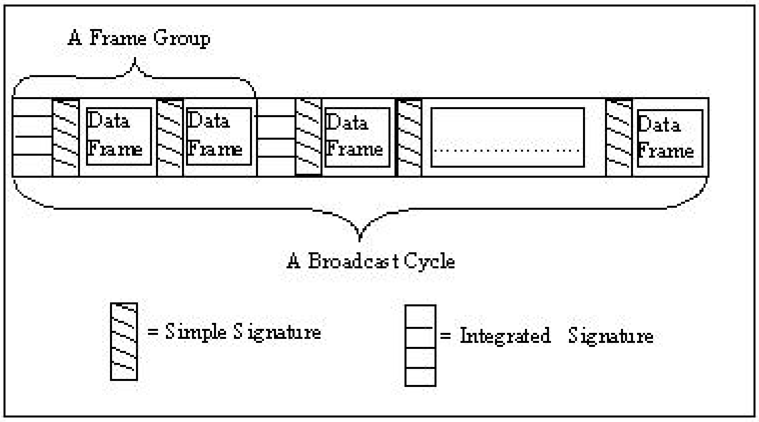
The difficulty is to decide the broadcast order in advance, even without much knowledge of any future query. In general, an access graph is needed to identify the optimal organization of database items over a channel. The access graph is used to represent the dependency of data items. Once the access graph is built, a certain algorithm is utilized to determine the best broadcast program. A cost model called the Semantic Ordering Model (SOM) can be used to determine the most efficient access graph (Si & Leong, 1999). SOM is defined into two models: namely, Simple and Extended SOM. Simple SOM considers only the relationship among entity types while Extended SOM incorporates historical query access patterns. The branch and bound like algorithm (Si & Leong, 1999), Heuristics algorithm (Hurson, Chehadeh, & Hannan, 2000), randomized algorithm (Bar-Noy, Naor, & Schieber, 2000), and Genetic algorithm (Huang & Chen, 2002, 2003) are some algorithms that can be used to identify the most effective organization of broadcast data items. The final broadcast program can be distributed over either a single or multiple channels.

Minimizing Tuning Time

A broadcast indexing scheme is needed to reduce the tuning time by providing accurate information for a client to tune in at the appropriate time for the required data (Lee, Leong, & Si, 2002). In this scheme, some form of directory is broadcast along with the data, and the clients obtain the index directory from the broadcast and use it in subsequent reads. The information generally also contains the exact time of the data to be broadcast. As a result, mobile clients are able to conserve the energy of their unit by switching to "*doze*" mode and back to "*active* mode" when the data is about to be broadcast.

In general, a broadcast indexing technique involves a trade-off between optimizing the client tuning time and the query response time. The consequence of minimizing one of them is the increase of the other. For instance, to minimize the response time is to reduce the length of broadcast cycles. In this case, the index can be broadcast once in each cycle but it will make the tuning time suffer since the client will have to wait for the index to arrive which happens only once in each broadcast cycle. On the other hand, when the index directory is frequently broad-

Figure 4. Multi-level signature



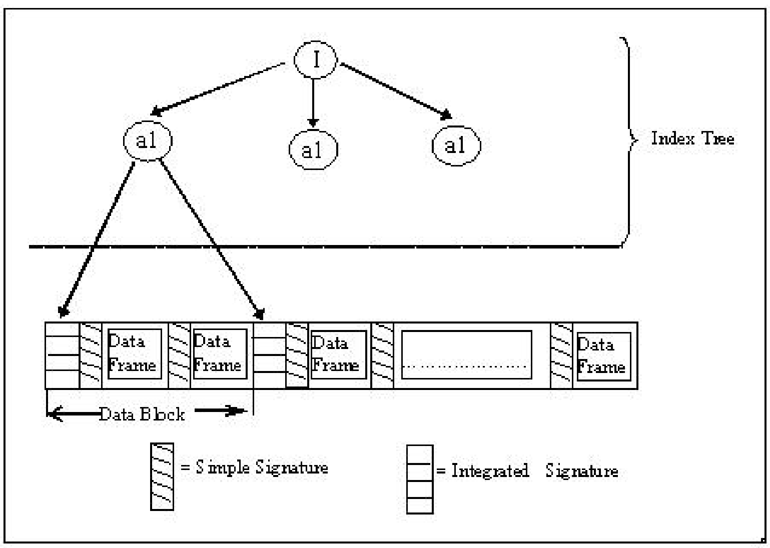
cast in each broadcast cycle to reduce the tuning time, the response time will be greatly affected due to the occupancy of the index in the cycle. Thus, it is necessary to find the optimal balance between these two factors.

Clustering index, non-clustering index, and multiple index methods are used to determine the index distribution over a broadcast channel (Imielinski, Viswanathan, & Badrinath, 1997). Clustering index refers to clustering of a record's attribute whenever all records that belong to the attribute have the same value consecutively. Non-clustering index defines a method for indexing the non-clustering attributes by partitioning each non-clustered attribute in the broadcast cycle into a number of segments called *meta segments*. In multiple indexes, a second attribute is chosen to cluster the data items within the first clustered attribute.

Broadcast indexing can be *tree-indexing* based, *signature* based, or *hybrid indexing* that is the combination thereof. Tree-indexing based such as Global Index structure is utilised to minimise the index response time (Taniar & Rahayu , 2002; Waluyo, Srinivasan, & Taniar, 2003b). The design of the global index model is based on *B*+ tree structure; it incorporates an index channel, and the data items are broadcast separately in data channels.

A signature based index is derived by hashing the attribute values into bit strings followed by combining them together to form a bit vector or signature (Lee & Lee, 1996). The signature is broadcast together with the data on every broadcast cycle. This mechanism is also applied to the query initiated by the client. To process the query, mobile clients need to tune into the broadcast channel and verify the query signature with the data signature by performing a certain mathematical operation such as the "AND" operation. If the signature is not matched, the client can tune to the "doze" mode while waiting for the next signature to arrive. The main issue with this method

Figure 5. Hybrid indexing



is to determine the size of the signature as well as the number of levels of the signature. There are three different signature-based index algorithms, namely *simple signature*, *integrated signature* and *multilevel signature* (Lee & Lee, 1996).

In simple signature, the signature frame is broadcast ahead of each data frame. This makes the total number of signature frames the same as the data frames. An integrated signature is applied for a group of data frames and the signature is calculated accordingly. Figure 4 shows a combination of these two algorithms, forming a multi-level signature. This technique is designed to interleave with data items in a single channel.

A hybrid indexing technique made up of index tree and signature is expected to outperform the single indexing techniques by integrating the advantages of the two techniques into a single operation (Lee, Hu, & Lee, 1998; Hu, Lee, & Lee, 1999). This technique is shown in Figure 5.

CONCLUSION

The inherent limitations and characteristics of mobile computing such as power, storage, asymmetric communication cost, and bandwidth, have become interesting challenges and research opportunities in the field.

This article describes main issues, and several approaches in data dissemination in mobile databases that have been derived from literature. We classify each scheme into two categories: one is to minimize query response time, and the other is to minimize tuning time. Broadcasting schemes that aim to minimize response time include:

(i) Selection of Data Items to be broadcast,
(ii) Non-Uniform Frequency Distribution of Broadcast Data Items,
(iii) Distribution of Data Items over Multiple Channels, and
(iv) Organization of Data Items.

To minimize the tuning time, broadcast indexing of data items is applied. Hybrid schemes that integrate the advantages of these approaches are also of great potential. These techniques can certainly be improved in a number of ways and a vast amount of research is continuing for this purpose.

REFERENCES

Acharya, S., Alonso, R., Franklin, M., & Zdonik, S. (1995). Broadcast disks: Data management for asymmetric communication environments. In *Proceedings of ACM Sigmod* (pp.199-210).

Barbara, D. (1999). Mobile computing and databases: A survey. *IEEE Transactions on Knowledge and Data Engineering*, *11*(1), 108-117.

Bar-Noy, A., Naor, J., & Schieber, B. (2000). Pushing dependent data in clients-providers-servers systems. In *Proceedings of the 6th ACM/IEEE on Mobile Computing and Networking* (pp.222-230).

Elmasri, R., & Navathe, S.B. (2003). *Fundamentals of database systems* (4th ed.). Addison Wesley, U.S.A.

Hu, Q., Lee, W.C., & Lee, D.L. (1999). Indexing techniques for wireless data broadcast under data clustering and scheduling. In *Proceedings of the 8th ACM International Conference on Information and Knowledge Management* (pp.351-358).

Huang, J.-L., & Chen, M.-S. (2002). Dependent data broadcasting for unordered queries in a multiple channel mobile environment. In *Proceedings of the IEEE GLOBECOM* (pp.972-976).

Huang, J.-L., & Chen, M.-S. (2003). Broadcast program generation for unordered queries with data replication. In *Proceedings of the 8th ACM Symposium on Applied Computing* (pp.866-870).

Hurson, A.R., Chehadeh, Y.C., & Hannan, J. (2000). Object organization on parallel broadcast channels in a global information sharing environment. In *Proceedings of the 19th International Performance, Computing and Communications* (pp.347-353).

Imielinski, T., Viswanathan, S., & Badrinath, B.R. (1997). Data on air: Organisation and access. *IEEE Transactions on Knowledge and Data Engineering*, *9*(3), 353-371.

Lee, D.L., Hu, Q., & Lee, W.C. (1998). Indexing techniques for data broadcast on wireless channels. In *Proceedings of the 5th Foundations of Data Organization* (pp.175-182).

Lee, K.C.K., Leong, H.V., & Si, A. (2002). Semantic data access in an asymmetric mobile environment. In *Proceedings of the 3rd Mobile Data Management* (pp.94-101).

Lee, W.C., & Lee, D.L. (1996). Using signature techniques for information filtering in wireless and mobile environments. *Journal on Distributed and Parallel Databases*, *4*(3), 205-227.

Leong, H.V., & Si, A. (1995). Data broadcasting strategies over multiple unreliable wireless channels. In *Proceed-*

ings of the 4th Information and Knowledge Management (pp.96-104).

Leong, H.V., & Si, A. (1997). Database caching over the air-storage. *The Computer Journal*, *40*(7), 401-415.

Malladi, R., & Davis, K.C. (2002). Applying multiple query optimization in mobile databases. In *Proceedings of the 36th Hawaii International Conference on System Sciences* (pp. 294-303).

Myers, B.A., & Beigl, M. (2003). Handheld computing. *IEEE Computer Magazine*, *36*(9), 27-29.

Paulson, L.D. (2003). Will fuel cells replace batteries in mobile devices? *IEEE Computer Magazine*, *36*(11), 10-12.

Si, A., & Leong, H.V. (1999). Query optimization for broadcast database. *Data and Knowledge Engineering*, *29*(3), 351-380.

Taniar, D., & Rahayu, J.W. (2002). A taxonomy of indexing schemes for parallel database systems. *Distributed and Parallel Databases*, *12*, 73-106.

Trivedi, K.S., Dharmaraja, S., & Ma, X. (2002). Analytic modelling of handoffs in wireless cellular networks. *Information Sciences*, *148*, 155-166.

Waluyo, A.B., Srinivasan, B., & Taniar, D. (2003a). Optimal broadcast channel for data dissemination in mobile database environment. In *Proceedings of the 5th Advanced Parallel Processing Technologies*, LNCS, *2834*: 655-664.

Waluyo, A.B., Srinivasan, B., & Taniar, D. (2003b). Global index for multi channels data dissemination in mobile databases. In *Proceedings of the 18th International Symposium on Computer and Information Sciences*, LNCS, *2869*, 210-217.

Xu, J., Zheng, B., Zhu, M., & Lee, D.L. (2002). Research challenges in information access and dissemination in a mobile environment. In *Proceedings of the Pan-Yellow-Sea International Workshop on Information Technologies for Network Era* (pp.1-8).

Yajima, E., Hara, T., Tsukamoto, M., & Nishio, S. (2001) Scheduling and caching strategies for correlated data in push-based information systems. *ACM SIGAPP Applied Computing Review*, *9*(1), 22-28.

KEY TERMS

Broadcast Channel: Unidirectional wireless channel to disseminate a set of database items periodically to multiple numbers of mobile users.

Broadcast Cycle: A complete broadcast file.

Data Dissemination/Broadcasting: Periodical broadcast of database information to mobile clients through one or more wireless channels.

Data Item: Database record or tuples.

Data Segment: A set of data items.

Mobile Base Station (MBS): Fixed host that has wireless interface for communicating with mobile clients.

Mobile Computing: The ability of mobile users to keep connected to the wireless network while traveling, and to access information such as news, weather forecast, email, and query to central database server.

Mobile Database: Mobile users connected to the wireless network and equipped with database application to conduct activity like transaction and information retrieval from central database server.

Query Response Time: The total elapsed time while waiting for the data of interest to arrive in the channel and downloading the data.

Tuning Time: The total time a mobile client must listen to the channel, which is used to indicate its energy consumption.

Data Mining and Mobile Business Data

Richi Nayak
Queensland University of Technology, Australia

INTRODUCTION

Research and practices in mobile (m-) business have seen an exponential growth in the last decade (CNN, 2002; Leisen, 2000; McDonough, 2002; Purba, 2002). M-businesses allow users to access information, perform transactions and other operations from anywhere at anytime via wireless networks. Consequently, m-business applications are generating a large volume of complex data (Magic-sw, 2002). Monitoring and mining of this data can assist m-business operators to make sound financial and organisational decisions.

Data mining (DM) or knowledge discovery in databases is the extraction of interesting, meaningful, implicit, previously unknown, valid and actionable information from a pool of data sources (Dunham, 2003). This valuable and real-time information inferred from the data can be used for decision-making. For example, common use of mobile phones and personal digital assistance (PDAs) has increased the number of service providers. The DM technology can help providers to develop services and sales strategies for future benefits. An example of existing applications of data mining in m-business is MobiMine (Kargupta, Park, Pittie, Liu, Kushraj, & Sarkar, 2002), which enables a user to monitor stock prices from a handheld PDA.

BACKGROUND: PROCESS OF KNOWLEDGE DISCOVERY

Data mining - an interactive, iterative, non-trivial process - is usually divided into many subtasks (Figure 1). Prior to commencing the mining process, businesses should identify and define their goals, objectives and limitations. Accordingly, data is gathered and collated from multiple sources as each source may send data in different formats. The next phase is to ensure quality of the data by removing noise, handling missing information and transforming to an appropriate format. A reduced volume of the data set " representative of the overall processed data " is also derived by applying data reduction techniques.

Once the data is pre-processed, an appropriate data mining technique or a combination of techniques is applied for the type of knowledge to be discovered (Table 1). The discovered knowledge is then evaluated and interpreted, typically involving some visualization techniques. When the mined results are determined insufficient, an iterative process of performing preprocessing and mining begins until adequate and useful information is obtained. Lastly, the information is presented to user to incorporate into the company's business strategies.

Figure 1. The data mining process

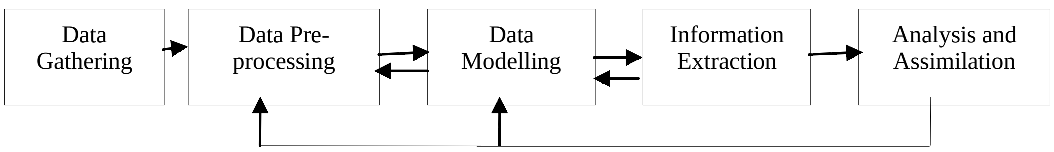


Table 1. Various data mining tasks

Mining Task	Goal	Approaches
Predictive Modelling	To predict future needs based on previous data	Decision tree, Neural networks
Clustering	To partition data into segments	Demographic, Neural networks
Link Analysis	To establish association among items	Counting occurrences of items such as Apriori Algorithms
Deviation Detection	To detect any anomalies, unusual activities	Summarization and Graphical representation

DATA MINING OPPORTUNITIES IN M-BUSINESS DOMAIN

Taking Advantage of Location Information

With the Global Positioning System (GPS) mobile technology, it is possible to identify the location of users (Cousins & Varshney, 2001; Duri, Cole, Munson, & Christensen, 2001). Based on the locations that a person frequents most and the personal information given, it is possible to classify the user in a pre-defined category with data mining techniques. For example, if a person is most often sighted in supermarkets, department stores and at home, and is seen shuttling between sales events, then this person can be classified as a possible homemaker interested in sales events. In terms of a business-to-consumer relation, such information allows businesses to provide the appropriate marketing information to the specific category of users.

In terms of a business-to-business relation, the ability to track the location of the employees is ideal to determine the work efficiency of the employees. Analysis of employee's time spent on the duty will determine the employee who is performing best and most suited for the next pay increment and promotion. Businesses like courier companies are dependent on the information regarding the locations of the transported parcels. Data mining techniques are able to analyse various routes and time spent in receiving parcels over a period. The outcome determines the efficiency of the business processes and factors behind their failure or success.

Personalization of M-Business Applications

Due to the limited screen space provided on mobile devices, it is difficult for mobile users to browse the product or service catalogues on the devices. It is important for vendors to provide only the products or services that match the needs of individual users. Short message service (SMS) is used primarily for simple person-to-person messaging. Information obtained from analysing the user data about previously accessing these services can be used to create personalized advertises to the customer delivered by SMS (Mobilein.com, 2002).

Relevant services can be offered based not only on the personal profile of the device holder, but also on the device holder's location and time factor. For example, m-business applications used in the travel industry can assist users to find attractions, hotels and restaurants of their preference on requested location and time. The *clustering data mining* technique groups the customers with similar preferences. When a new customer mentions his preferences, a recommendation can be made based on the previous similar preferences. *Associative data mining* can be used to indicate which places a person is most likely to visit in a single trip or in two consecutive trips, with having inputs such as location and time of visits to attractions for each user. This provides great convenience for users as these services can be used while driving, for example, a suggestion can be made based on the association rule that if the user is on place A then the user should visit the place B, previous 80% visitors have done so.

Predicting Customer Buying and Usage Patterns

Service providers can analyse the consumer behaviour data (e.g., by analysing gateway log files and content server log files on WAP) and predict the consumers buying and usage patterns, or to understand how mobile subscribers use their wireless services. Using the stored data, companies can apply data mining to identify customer segments using *clustering data mining techniques*, to distinguish customers' consumption patterns using *deviation detection techniques*, and transaction trends using *associative data mining techniques*. This information can then be used to provide better services to the customers or to attract potential customers.

Predicting Future and Better Usage of Mobile Technology

Data, about the number of mobile phones in the market, the number of users subscribing a service, the amount of usage measured in currency, the users' satisfaction and feedback, can be extracted and analysed with data mining. The resulting information can be used to predict the trends and patterns of usage of mobile phones and services. For example, some of the popular services bought through m-commerce technology are mobile ringing tones, logos and screensavers. The most common used interface for these kind of transactions are short message service (SMS) and the standard e-commerce interface, the Internet. An example is Nokia's focus on screensavers, logos and ringing tone availability. This is most likely to be a result of previous research on their users' trends, by capturing the data on the users' demands and needs, and then analysing the users' feedbacks. This information helped Nokia to develop a new market product where the product is no longer just a mobile phone, but also provides extra features like SMS, logos and additional ringing tones and screensavers (Nokia, 2002).

Trend Analysis of Costs versus Benefits

Any m-business constantly analyses whether the profits derived from the business is sustainable. Data mining can assist to do a trend analysis of the business over a period. A possible way to analyse the data collected from the profit derived from the business is the use of linear regression (a *value prediction technique* of the DM). A graph based on average returns versus average investment into the business can be plotted. Analysis of graph indicates whether the amount of investment incurred is greater or smaller than the returns derived. Many other business factors can also be considered during regression analysis.

Optimisation of Delivery Content

A mobile commerce platform should integrate with existing backend databases and businesses applications to deliver data via all the channels such as WAP, VoxML, TruSync, BlueTooth or any wireless protocol. Data mining can also be used to match which channel is best at a time to deliver the information. Data mining techniques can optimise the amount and format of the content for delivery based on the connection speed of the device requesting the information. Data mining techniques help to decide what tasks, activities and transactions are most economical and beneficial to use at the time.

Fraud Detection in M-Business

The analysis of the types of fraudulent activities in telecommunication systems is one of the applications that data mining can assist in a mobile environment. The dynamic nature of different fraudulent activities and the changes of the normal usage can lead in the detection of fraudulent through observing behavioural patterns. A data mining system will have plenty of examples of normal usage and some examples of fraud usage. Based on these previous examples, a *predictive data mining* system establishes facts about fraudulent activities. Whenever a change in the normal usage is detected, the system analyses the change, and is able to predict whether the change is a fraud or not.

DATA MINING CHALLENGES

In order to apply data mining efficiently in m-business, certain requirements have to be met. Ideally, the methods used for mining mobile data should be able to: (1) mine different kinds of knowledge in databases; (2) deal with diverse type of data types such as relational, temporal and spatial types of data; (3) mine information from heterogeneous databases and global information systems; (4) handle noise and incomplete data that is mostly the case in m-business domain; (5) perform the mining tasks efficiently regardless of the size and complexity of the data set; (6) support interactive mining of knowledge at multiple levels of abstraction; (7) support integration of the discovered knowledge with existing knowledge; and (8) deal with the issues related to applications of discovered knowledge and social impacts such as protection of data security, integrity and privacy.

Distributed Environment

In m-business environment, data can reside in many different geographical locations. Most data mining systems are currently based on centrally-located data; data is stored in a single database and the mining techniques are focused on this data set. XML is proving to be an essential way to perform data exchange not only on the Web but also wirelessly between applications or between users and applications. XML has provided the facilities to integrate data and documents to allow for data communication in "a flexible and extensible representation" (Graves, 2002). If every mobile device is able to transmit XML documents that can be read and processed, regardless which platform the mobile device is running on, data integration from multiple sources becomes easier.

But, as a result of convergence between computation and communication, the new data mining approaches have to be concerned with distributed aspects of computation and information storage. A distributed data mining approach typically works by: (1) analysing and compressing local data for minimisation of network traffic; and (2) analysing and generating global data models after combining local data and models (Park & Kargupta, 2002).

Clickstream Data

Users of mobile devices are highly restricted on the Web pages that they can visit, due to small display screen. On a WAP phone, the average number of links it has to other Web site is an average of five links, while a standard Web page has an average of 25 links. If a user is to have three clicks on the Web via a WAP phone, there are only 5^3 (= 125) pages that are accessible to the user, compared with the standard Web page having 25^3 (= 15625) accessible pages (Barnes, 2002). It is quite unlikely that the user will be going to the site that he really wants from the links available. As a result, the usage of data mining to analyse clickstream data collected from users of mobile devices to predict the user's interest is not going to be accurate.

Security and Privacy

With the technology of sending personal messages to mobile users, it has become possible for users to specify the types of information that they prefer, and for businesses to provide those information only. For example, if the user indicates that his preference is a particular brand of product above a particular price, then it can be analysed that the user may also be interested in another similar brand of the same standard. This opens data mining possibilities such as classifying the users based on their reported needs; finding correlations between various needs.

Unfortunately, some users who do not believe in the security of mobile data might inaccurately declare their personal information and preference. This will result in incorrect data mining output. Thus, although data mining results have classified the user as a potential person to send information to, but in reality, it can add expense to incur cost in conducting data mining and including the irrelevant people into the mobile service. A possible solution is use of XML that allows documents to be complex and tagged with unmeaningful names in data transfer. The document is not useful to an unauthorized person without the knowledge of how to transform (decrypt) the document appropriately.

Cost Justification

With the issue that a data mining application is usually computationally expensive, there is always a concern whether the benefits of data mining justify the cost incurred in the process. Also, there is a difficulty to strike a balance between the security and privacy of data transferred versus the computational cost required to process the "encrypted" documents. The more complex the communication document is with the concern of security and privacy, the more computational power is needed to process these documents.

Technological Limitation

Although there are a number of mobile technologies available, there exist several limitations and constraints of the technologies adversely affecting the performance of data mining in m-business domain. Some of the limitations are low bandwidth, limited battery power, unreliable communications that result in frequent disconnections. These factors increase the communication latency, additional cost to retransmit data, time-out delays, error control protocol processing and short disconnections (Madria, Mohania, Bhowmick, & Bhargava, 2002). These limitations pose significant problems in collecting data for mining purposes. For example, the present low bandwidth means that the data transfer is slow. This implies that data mining processes have to be delayed until most data transfers have been completed and received.

Furthermore, the potential of gathering knowledge about a user's location is appealing in terms of m-business and data mining, this potential is not yet realized until present technologies improve to provide adequate and up to standard location.

FUTURE TRENDS AND CONCLUSION

The success of an m-business depends on the ability to deliver attractive products or services that are personalized to the individual user at the right time on the right location. These information intensive services can only be obtained by collecting and analysing combined demographic, geographic, and temporal data. This data can not be transformed into useful information with traditional reporting techniques and tools. Data mining enables the user to seek out facts by identifying patterns within data. Data mining can give businesses the edge over other businesses by offering marketing that is more focused on particular consumer groups or with suggesting the better use of mobile technology. An investment in data mining to m-business data is an extra expense, but can still help m-businesses to provide the right services to the right people at the right time, and that can make a vital difference.

REFERENCES

Barnes, S.J. (2002, April). The mobile commerce value chain: Analysis and future developments. *International Journal of Information Management*, *22*, 91-108.

CNN. (2002). Peek into the future of mobile shopping. Retrieved September 22, 2002, from *http://www.cnn.com/2000/TECH/computing/11/16/m.commerce.future.idg/*

Cousins, K. & Varshney, U. (2001). A product location framework for mobile commerce environment. In *Proceedings of the First International Workshop on Mobile Commerce*, July 2001 (pp.43-48).

Dunham, M.H. (2003). *Data mining: Introductory and advanced topics*. Prentice Hall.

Duri, S., Cole, A., Munson, J., & Christensen, J. (2001). An approach to providing a seamless end-user experience for location-aware applications. In *Proceedings of the First*

International workshop on Mobile Commerce, July 2001 (pp.20-25).

Graves, M. (2002). *Designing XML databases*. Prentice Hall.

Kargupta, H., Park, B., Pittie, S., Liu, L., Kushraj, D., & Sarkar, K. (2002, January). MobiMine: Monitoring the stock market from a PDA. In *ACM SIGKDD Explorations*, *3*(2), 37-46. ACM Press.

Leisen, B. (2000). M-Commerce: Mobile and multiplying. Wall Street Journal, August 18, 2000, p.B1, B4. Retrieved on September 25 2002, from *http://myphlip.pearsoncmg.com/cw/mpviewce.cfm?vceid=1913&vbcid=1714*.

Madria, S.K., Mohania, M., Bhowmick, S.S., & Bhargava, B. (2002, April). Mobile data and transaction management. *Information Sciences*, *141*(3-4), 279-309.

Magic-sw. (2002). Mobile e-business magic white paper. Retrieved September 20, 2002, from *http://www.magic-sw.com/*

McDonough, Jr., D. (2002). Digital river gives m-commerce a kick [online]. *Wireless News*. Accessed September 24, 2002, from *http://www.wirelessnewsfactor.com/perl/story/17377.htm*

Mobilein.com (2002). MobileIN.com, personalization, mobile communications & intelligent networks and applications. Retrieved September 22, 2002, from *http://www.mobilein.com/personalization.htm*

Nokia. (2002). Nokia: History in brief. Retrieved on September 25 2002, from *http://www.nokia.com/aboutnokia/compinfo/history.html*

Park, B., & Kargupta, H. (2002). Distributed data mining: Algorithms, systems, and applications. In N. Ye (Ed.), *Data mining handbook*.

Purba, S. (2002). *New directions in Internet management*. Auerbach Publications.

KEY TERMS

Clustering: This data mining task is to identify items with similar characteristics, and thus creating a hierarchy of classes from the existing set of events. A data set is partitioned into segments of elements (homogeneous) that share a number of properties.

Data Mining: Data mining (DM) or knowledge discovery in databases is the extraction of interesting, meaningful, implicit, previously unknown, valid and actionable information from a pool of data sources.

Link Analysis: This data mining task establishes internal relationship to reveal hidden affinity among items in a given data set. Link analysis exposes samples and trends by predicting correlation of items that are otherwise not obvious.

Mobile Business Data: Data generating from mobile business activities such as accessing information, performing transactions and other operations from anywhere anytime via wireless networks.

Predictive Modelling: This data mining task makes predictions based on essential characteristics about the data. The classification task of data mining builds a model to map (or classify) a data item into one of several predefined classes. The regression task of data mining builds a model to map a data item to a real-valued prediction variable.

D

Data Mining for Combining Forecasts in Inventory Management

Chi Kin Chan
The Hong Kong Polytechnic University, Hong Kong

INTRODUCTION

The traditional approach to forecasting involves choosing the forecasting method judged most appropriate of the available methods and applying it to some specific situations. The choice of a method depends upon the characteristics of the series and the type of application. The rationale behind such an approach is the notion that a "best" method exists and can be identified. Further that the "best" method for the past will continue to be the best for the future. An alternative to the traditional approach is to aggregate information from different forecasting methods by aggregating forecasts. This eliminates the problem of having to select a single method and rely exclusively on its forecasts.

Considerable literature has accumulated over the years regarding the combination of forecasts. The primary conclusion of this line of research is that combining multiple forecasts leads to increased forecast accuracy. This has been the result whether the forecasts are judgmental or statistical, econometric or extrapolation. Furthermore, in many cases one can make dramatic performance improvements by simply averaging the forecasts.

BACKGROUND OF COMBINATION OF FORECASTS

The concept of combining forecasts started with the seminal work 35 years ago of Bates and Granger (1969). Given two individual forecasts of a time series, Bates and Granger (1969) demonstrated that a suitable linear combination of the two forecasts may result in a better forecast than the two original ones, in the sense of a smaller error variance. Table 1 shows an example in which two individual forecasts (1 and 2) and their arithmetic mean (combined forecast) were used to forecast 12 monthly data of a certain time series (actual data).

The forecast errors (i.e., actual value – forecast value) and the variances of errors are shown in Table 2.

From Table 2, it can be seen that the error variance of individual forecast 1, individual forecast 2, and the combined forecast are 196, 188 and 150, respectively. This shows that the error variance of the combined forecast is smaller than any one of the individual forecasts and hence demonstrates an example how combined forecast may work better than its constituent forecasts.

Bates and Granger (1969) also illustrated the theoretical base of combination of forecasts. Let X_{1t} and X_{2t} be two individual forecasts of Y_t at time t with errors:

Table 1. Individual and combined forecasts

Actual Data (Monthly Data)	Individual Forecast 1	Individual Forecast 2	Combined Forecast (Simple Average of Forecast 1 and Forecast 2)
196	195	199	197
196	190	206	198
236	218	212	215
235	217	213	215
229	226	238	232
243	260	265	262.5
264	288	254	271
272	288	270	279
237	249	248	248.5
211	220	221	220.5
180	192	192	192
201	214	208	211

$$e_{jt} = Y_t - X_{jt}, \qquad j = 1,2$$

such that

$$E[e_t] = 0, \qquad E[e_{jt}^2] = \sigma_j^2 \qquad j = 1,2$$

and

$$E[e_{1t}e_{2t}] = \rho\sigma_1\sigma_2$$

where σ_j^2 is the error variance of the j^{th} individual forecast and ρ is the correlation coefficient between the errors in the first set of forecasts and those in the second set.

Consider now a combined forecast, taken to be a weighted average of the two individual forecasts:

$$X_{ct} = kX_{1t} + (1-k)X_{2t}$$

The forecast error is

$$e_{ct} = Y_t - X_{ct} = ke_{1t} + (1-k)e_{2t}$$

Hence the error variance is

$$\sigma_c^2 = k^2\sigma^2 + (1-k)^2\sigma_2^2 + 2k(1-k)\rho\sigma_1\sigma_2 \qquad \textbf{(1)}$$

This expression is minimized for the value of k given by:

$$k_0 = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$

and substitution into equation (1) yields the minimum achievable error variance as:

$$\sigma_{c0}^2 = \frac{\sigma_1^2\sigma_2^2(1-\rho^2)}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$

Note that $\sigma_{c0}^2 < \min(\sigma_1^2, \sigma_2^2)$ unless either ρ is exactly equal to σ_1/σ_2 or σ_2/σ_1. If either equality holds, then the variance of the combined forecast is equal to the smaller of the two error variances. Thus, a priori, it is reasonable to expect in most practical situations that the best available combined forecast will outperform the better individual forecast—it cannot, in any case, do worse.

Newbold and Granger (1974), Makridakis et al. (1982), Makridakis and Winkler (1983), Winkler and Makridakis (1983), and Makridakis and Hibbon (2000) have also reported empirical results that showed that combinations of forecasts outperformed individual methods.

Since Bates and Granger (1969), there have been numerous methods proposed in the literature for combining forecasts. However, the performance of different methods of combining forecasts varies from case to case. There is still neither definitive nor generally accepted conclusion that sophisticated methods work better than simple ones, including simple averages. As Clemen (1989) commented: *In many studies, the average of the individual forecasts has performed the best or almost best.* Others would agree with the comment of Bunn (1985) that the Newbold and Granger (1974) study and that of Winkler and Makridakis (1983) "demonstrated that an overall *policy* of combining forecasts was an efficient one and that if an *automatic* forecasting system were required, for example, for inventory planning, then a linear combina-

Table 2. Forecast errors and variances of errors

Errors of Individual Forecast 1	**Errors of Individual Forecast 2**	**Errors of Combined Forecast**
1	-3	-1
6	-10	-2
18	24	21
18	22	20
3	-9	-3
-17	-22	-19.5
-24	10	-7
-16	2	-7
-12	-11	-11.5
-9	-10	-9.5
-12	-12	-12
-13	-7	-10
Variance of errors = 196	**Variance of errors = 188**	**Variance of errors = 150**

tion using a 'moving-window' estimator would appear to be the best overall".

DATA MINING AND COMBINATION OF FORECASTS IN INVENTORY MANAGEMENT

Data mining is the process of selection, exploration, and modeling of large quantities of data to discover regularities or relations that are, at first, unknown with the aim of obtaining clear and useful results for the owner of the database.

The data mining process is deemed necessary in a forecasting system, and it is particularly important in combining forecasts for inventory demands. Errors in forecasting demand can have a significant impact on the costs of operating and the customer service provided by an inventory management system. It is therefore important to make the errors as small as possible. The usual practice in deciding which system to use is to evaluate alternative forecasting methods over past data and select the best. However, there may have been changes in the process, generating the demand for an item over the past period used in the evaluation analysis. The methods evaluated may differ in their relative performance over sub-periods of the method that was best only part of the time, or in fact never the best method and perhaps only generally second best. Each method evaluated may be modeling a different aspect of the underlying process generating demands. The methods discarded in the selection process may contain some useful independent information. A combined forecast from two or more methods might improve upon the best individual forecasts. Furthermore, the inventory manager typically has to order and stock hundreds or thousands of different items. Given the practical difficulty of finding the best method for every individual item, the general approach is to find the best single compromise method over a sample of items, unless there are obvious simple ways of classifying the items, by item value or average demand per year, and so forth. Even if this is possible, there will still be many items in each distinct category for which the same forecasting method will be used. All of the points made on dealing an individual data series, as previously noted, apply with even more force when dealing with a group of items. If no one individual forecasting method is best for all items, then some system of combining two or more forecasts would seem *a priori* an obvious approach, if the inventory manager is going to use the same forecasting system for all items.

The need for data mining in combining forecasts for inventory demands comes from the selection of sample items on which forecasting strategy can be made for all items, the selection of constituent forecasts to be combined and the selection of weighting method for the combination.

The selection of the sample items is a process of exploratory data analysis. In this process, summary statistics such as mean and coefficient of variation can be investigated so that the sample selected could represent the total set of data series on inventory demands.

The selection of constituent forecasts to be combined is the first stage of model building. The forecasts methods might be selected from popular time series procedures such as exponential smoothing, Box-Jenkins and regression over time. One could include only one method from each procedure in the linear combination as the three groups of methods were different classes of forecasting model and thus might contribute something distinct, while there was likely to be much less extra contribution from different methods within the same class. It might also be useful to attempt to tailor the choice of methods to particular situations. For example, Lewandowski's FORSYS system, in the M-competition (Makridakis et al., 1982), appears to be particularly valuable for long forecast horizons. Thus it might be a prime candidate for inclusion in situations with long horizons but not necessarily in situations with short horizons. It is also important to note that combining forecasts is not confined to combination utilizing time series methods. The desire to consider any and all available information means that forecasts from different types of sources should be considered. For example, one could combine forecasts from time series methods with forecasts from econometric models and with subjective forecasts from experts.

The second stage of model building is to select the weighting method for the combination of forecasts. The weighting method could be simple average or "optimal" weighting estimated by a certain approach, for instance, the constrained OLS method (Chan, Kingsman, & Wong, 1999b). Furthermore, the "optimal" weights obtained could either be fixed for a number of periods (fixed weighting) or re-estimated every period (rolling window weighting) (Chan, Kingsman, & Wong, 1999a, 2004). The selection process can then be done by comparing the different weighting methods with an appropriate performance measure. In the case of inventory management, the carrying of safety stocks is to guard against the uncertainties and variations in demand and the forecasting of demand. These safety stocks are directly related, or made directly proportional, to the standard errors of forecasts. If, as is usually the case, a stock controller is dealing with many items, it is the performance across the group of items that matters. Hence, the sum of the standard errors of the forecasts, measured by the sum of the root mean squared errors over all the items, can be used to compare the results between the different weighting methods. The ultimate aim is to find the one best overall method for the weighting process to use for all the items.

FUTURE TRENDS AND CONCLUSION

The usual approach in practical inventory management is to evaluate alternative forecasting methods over a sample of items and then select the one that gives the lowest errors for a majority of the items in the sample to use for all items being stocked. The methods discarded in the selection process may contain some useful independent information. A combined forecast from two or more methods might improve upon the best individual forecasts. This gives us some insights in the process of data mining. There are a number of well-known methods in data mining such as clustering, classification, decision trees, neural networks, and so forth. Finding a good individual method from our tool kit to handle the data is clearly an important initial step in data mining. Then we should always bear in mind the power in combining the individual methods. There are two kinds of direction to do the combination. The first one is basically a direct combination of the individual methods, such as simple average or "optimal" weighting. The other one is to classify our data first, and then select the weighting method. Classification is always an important aspect of data mining and combination of forecasts sheds some new light on this. Another important message is that if we are dealing with large data sets, then it is not very worthwhile to find the "best" individual method. Obviously, there may not be any best individual at all. A viable alternative is to find several sensible individual methods and then combine them as the final method. This approach will usually relieve much of our effort in finding the best individual method, as justified by the law of diminishing return.

REFERENCES

Bates, J.M., & Granger, C.W.J. (1969). The combination of forecasts. *Operational Research Quarterly, 20*, 451-468.

Bunn, D.W. (1985). Statistical efficiency in the linear combination of forecasts. *International Journal of Forecasting, 1*, 151-163.

Chan, C.K., Kingsman, B.G., & Wong, H. (1999a). The value of combining forecasts in inventory management – A case study in banking. *European Journal of Operational Research, 117*, 199-210.

Chan, C.K., Kingsman, B.G., & Wong, H. (1999b). A comparison of unconstrained and constrained OLS for the combination of demand forecasts: A case study of the ordering and stocking of bank printed forms. *Annals of Operations Research*, *87*, 129-140.

Chan, C.K., Kingsman, B.G., & Wong, H. (2004). Determining when to update the weights in combined forecasts for product demand – An application of the CUSUM technique. *European Journal of Operational Research, 153*, 757-768.

Clemen, R.T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, *5*, 559-583.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., & Winkler, R. (1982), The accuracy of extrapolation (time series) methods: Results of a forecasting competition, *Journal of Forecasting*, *1*, 111-153.

Makridakis, S., & Hibbon, M. (2000), The M3 – competitions: Results, conclusions and implications. *International Journal of Forecasting*, *16*, 451-476.

Makridakis, S., & Winkler, R.L. (1983). Averages of forecasts: Some empirical results. *Management Science*, *29*, 987-996.

Newbold, P., & Granger, C.W.J. (1974). Experience with forecasting univariate time series and the combination of forecasts (with discussion). *Journal of Royal Statistical Society, Series A*, *137*, 131-149.

Winkler, R.L., & Makridakis, S. (1983). The combination of forecasts. *Journal of the Royal Statistical Society, Series A*, *146*, 150-157

KEY TERMS

Combination of Forecasts: Combine two or more individual forecasts to form a composite one.

Constrained OLS Method: A method to estimate the "optimal" weights for combination of forecasts by minimizing the sum of squared errors as in a regression framework, and the weights are constrained to sum to one.

Data Mining: The process of selection, exploration, and modeling of large quantities of data to discover regularities or relations that are at first unknown with the aim of obtaining clear and useful results for the owner of the database.

Fixed Weighting: "Optimal" weights are estimated and are used unchanged to combine forecasts for a number of periods.

Highest Weighting: Use the individual forecast procedure that is given the highest weight in the fixed weighting method. This is not a combination. This method is equivalent to choosing the forecasting technique which is the best on the weight estimation period.

Rolling Window Weighting: "Optimal" weights are estimated in each period by minimizing the errors over the preceding *m* periods, where *m* is the length in periods of the "rolling window". The weights are then used to combine forecasts for the present period.

Simple Average Weighting: A simple linear average of the forecasts, implying equal weights for combination of forecasts.

Data Mining for Supply Chain Management in Complex Networks

Mahesh S. Raisinghani
University of Dallas, USA

Manoj K. Singh
i2 Technologies, USA

INTRODUCTION

Supply chain comprises the flow of products, information, and money. In traditional supply chain management, business processes are disconnected from stock control and, as a result, inventory is the direct output of incomplete information. The focus of contemporary supply chain management is to organize, plan, and implement these flows. First, at the organizational level, products are manufactured, transported, and stored based on the customers' needs. Second, planning and control of component production, storage, and transport are managed using central supply management and replenished through centralized procurement. Third, the implementation of the supply chain involves the entire cycle from the order-entry process to order fulfillment and delivery. Data mining can create a better match between supply and demand, reducing or sometimes even eliminating the stocks.

Data mining thus has become an indispensable tool in understanding needs, preferences, and behaviors of customers. It is also used in pricing, promotion, and product development. Conventionally, data mining techniques have been used in banking, insurance, and retail business. This is largely because of the fact that the implementation of these techniques showed quick returns. Data mining is being used for customer profiling where characteristics of good customers are identified with the goals of predicting new customers and helping marketing departments target new prospects. The effectiveness of sales promotions/ product positioning can be analyzed using market-basket analysis to determine which products are purchased together or by an individual over time, which products to stock in a particular store, and where to place products in each store (Groth 2000; Kopanakis & Theodoulidis, 2003; Weir, 1998). In addition, data mining is used in a variety of other industries such as the financial, healthcare, and telecommunications industry, among others.

There are a lot of opportunities and applications of data mining even beyond the obvious. One of the potential areas is "Supply Chain Management." One of the realities of the demand and supply in the manufacturing industry is that no matter how well balanced a system is, there is an element of uncertainty that creates a mismatch between demand and supply. The objective of this article is to identify those areas in the supply chain where most of the uncertainty exists and to determine suitable data-mining methods to accurately predict uncertainty. The underlying assumption of this paper is that a data warehouse has been implemented before the data-mining techniques can be applied.

A KEY BUSINESS PROBLEM: UNCERTAINTY

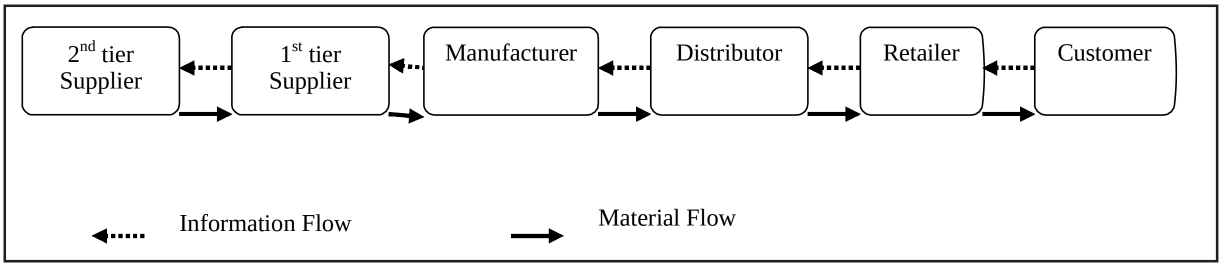
There are two issues that plague supply chain management—variation in the demand and supply and variation in the speed and extent of communication within the supply chain. Variation in demand and supply is due to the inherent uncertainty also present in the processes. Accurately predicting the uncertainties in demand, supply, and processes and then formulating action plans around the prediction is the essence of supply chain management (SCM).

Before we can address the problem of uncertainty in supply chain and explain the use of data-mining techniques, we need to understand the basic process of SCM and where uncertainty exists. In its most simplified form, a supply chain can be depicted as the flow of information from a customer's customer to a supplier's supplier and then the flow of material in the reverse direction, as shown in Figure 1.

The whole supply chain can be conceptually broken down as Supply – Process – Demand. Traditional forecasting planners of supply chains use demand and supply forecasting as a means of controlling uncertainty. However, there are three major drawbacks in those methods, namely:

1. Incorrect forecasting model
2. Incorrect number of parameters
3. Incorrect coefficients values of these parameters.

Figure 1. Information and material flows in the supply chain



Each of these three problems can be solved using data mining. The models are chosen from a finite set of predefined models in data mining. The model can be recreated as many times as needed in order to extract previously unknown patterns and relationships in data. When forecasting using data-mining techniques, the program can detect even minor effects of some parameters.

DATA MINING: METHODS AND PROCESS

Data mining is the process of extracting ideas in data. It can also be defined as "a decision support process that tries to discover patterns and relationships that are beyond the realm of human experience and imagination in the large database and present them to a knowledgeable user for review and examination" or "as the process of extracting previously **unknown, valid,** and **actionable** information from large databases and then using the information to make crucial business decisions" (Groth, 2000).

Data mining not only uses a discovery-based approach in which pattern matching and other algorithms are employed to determine the key relationships in the data but also describes the steps that must be taken to ensure meaningful results.

DATA-MINING METHODS

Data mining is used to build six types of models aimed at solving business problems: classification, regression, time series, clustering, association analysis, and sequence discovery.

Classification

A predictive model is generated based on the historical data. These models are used to assign instances to a group or class by calculating the value of a categorical variable. The value of this categorical variable is generally binary in nature. It can include multiple but discrete values.

Regression

Regression is used to predict values for categorical variables. The values are continuous, real numbers, i.e., it has decimal values, and it has no fixed range in which the values of the variables are fitted.

Time-series forecasting

This method uses a series of existing values and their attributes to forecast future values, except that the values of the categorical variables are dependent on time. Using various data-mining tools, the distinctive features of time can be exploited.

Clustering

Clustering is used to segment a database into clusters, with the members of each cluster sharing a number of interesting properties. These clusters are not predefined and have two basic uses: 1) summarizing the contents of the target databases; and 2) as inputs to the other methods s like supervised learning.

Association

Association is used to describe behavior that is captured in the database. This method relates the occurrences of various events by identifying patterns or groups of items.

Sequencing

Sequencing defines items that are likely to occur together on a sequence basis. This could help marketers in timing their promotions to correlate with the sequential buying order exhibited by their customers.

Data-Mining Process

Robert Grossman, Director of the National Center for Data Mining at the University of Illinois at Chicago, classifies the data-mining process into four phases. Each of these phases is described below:

Phase 1: Data Warehousing

Data warehousing is the foundation for successfully applying data-mining techniques and other analytical and predictive tools. Data warehousing involves the transfer, conversion, and integration of data from legacy systems to a central repository where data is stored and made available to clients.

The downside of data warehousing is the high cost of implementation and the time it takes to complete the process. Data warehouses can cost in excess of $10 million to build and take anywhere from one to three years to complete (Peacock, 1998). This is a very expensive and time-consuming effort.

An alternative is the use of a data mart, which is a functional or departmental data repository. It can constructed as individual components, usually costs between $10,000 and $1 million to build, and can be brought online in less than six months. Data marts can be inconsistent with the data warehouse. To correct this, data may also be enriched with additional attributes. This may be accomplished by adding data extracted from other internal databases or purchased from third-party sources (Asbrand, 1997).

Data-mining provides more meaningful data when it uses large databases extracted into data warehouses. Data-mining technology is more commonly used in large, consumer-oriented businesses such as banking and the retail industry because of the extremely high cost of implementation.

Phase 2: Data Mining Tools

Algorithms are applied to the data in order to produce predictive models. The selection of tools depends on the proper identification of a business problem and analysis to determine the correct technique to use. Some common types of problems and the technique used in data mining are illustrated in Table 1 (META Group, 1997):

The key to using these tools is to understand that it requires a team effort between the analysts, the marketing experts, and information technology experts.

Phase 3: Predictive modeling

During this phase, the predictive models are analyzed and combined to produce a single aggregate model. These techniques may be mixed sequentially or in parallel. Sequentially, the user picks a technique to produce a model and then applies another technique to the results. In parallel, the user chooses different techniques and applies them all to the initial dataset.

Phase 4: Predictive Scoring

Here, the predictive models are applied to score operational data (Grossman, 1998). For instance, a bank could analyze the attributes and habits of its checking account customers for clues that might reveal an acceptable minimum balance in order to retain profitable customers. The bank can use data mining to develop profiles of customer groups inclusive of members consistently hav-

Table 1. Problem types and techniques used in data mining

Example	Problem Type	Technique
What are the top three characteristics of customers who have switched to my competition?	Classification	Neural Networks Decision Tree
What are the largest buckets within my customer base to which I should be marketing a new service?	Clustering	Neural Networks Decision Tree
What is the likelihood a given individual who opens a bank account will also open an IRA within the next three months?	Association and Sequencing	Statistical Techniques Rule Induction
What will the average exchange rate be over the next three months?	Regression and Forecasting	Neural Networks Statistical Techniques

ing trouble maintaining minimum balances. This helps the bank identify profitable customers and predict the minimum balance needed to retain them. As a result, the percentage of profitable customers can rise by a significant percentage (Fabris, 1998).

APPLICATIONS OF DATA MINING IN SUPPLY CHAIN MANAGEMENT

Manufacturers, airlines, banks, insurance companies, credit card companies, and retailers have successfully used data-mining technology. Data mining works best as a supplement to the existing tools and can be used with decision support systems (DSS) to improve the overall result of the system. Zdanowicz (2004) discuss the detection of money laundering and terrorist financing via data mining. Data-mining tools can be used to provide the most accurate picture of the capacity, maintenance, and factory scheduling problems. DSS can take this information as input to provide the planner with an optimal factory scheduling solution.

The supply chain model for any industry will have suppliers, manufacturers, distributors, retailers and customers. Next, we look at each segment of the supply chain to understand the application of data mining.

Retailers

At this point of the supply chain, retailers receive the forecast from primarily two sources, one directly from the individual customer and the other from small- and medium-size organizations. Actual consumption by the individual customers and medium and small organizations is added. The difference between the forecast and the actual consumption is the variation in the demand. The difference between the request and the supply accounts for the variation in the supply.

Data mining can be used at this point in the following ways:

- Market segmentation based on service needs of distinct groups;
- Market basket analysis - retailers can understand the buying behavior of the customers; and
- Target promotion with the use of a computerized approach and an extensive database.

Distributors

At this point in the supply chain, distributors receive the forecast from retailers and large organizations. They pair this data with the actual consumption by the retailers and large organizations. The difference between the forecast and the actual consumption is the variation in the demand. The difference between the promise and the actual supply accounts for the variation in the supply.

Data mining can be used at this point in the following ways:

- Predictions of supply uncertainties to predict the supply uncertainties at the supplier level and at the item level;
- Predictions of process uncertainties can be loss and item obsolescence;
- Predictions of demand uncertainties - market segmentation for retailers and big customers based on factors such as volume of demand, periodicity, variations, etc.;
- Stock out prediction at the warehouse; and
- Strategic implications - logistics becomes a tool to help accomplish corporate strategic objectives.

Manufacturers

At this point, manufacturers receive the demand from distributors or directly from retailers. Manufacturers pass on the demand of retailers to alliance partners and try to fulfill only the demand of distributors. The difference between the forecasted demand from distributors and the actual consumption by the distributors is the variation in the demand. The difference between the promise and the actual supply from component suppliers accounts for the variation in the supply for manufacturers.

Data mining can be used at this point in the following ways:

- Predictions of supply uncertainties - predict supply uncertainties for manufacturers at the supplier level and at the item level;
- Predictions of process uncertainties due to machine breakdown, poor performance, and maintenance schedules;
- Predictions of demand uncertainties based on distributor, item, location, etc.; and
- Predicting future trends in demand - discover trends in the demand of the product.

Mass Customization

Another area where data mining can be used is in identifying the products for "mass customization" at the delivery end. Companies implement mass customization by putting together unique customer orders from a large number of products, while minimizing the inventory of product components. Data mining helps in identifying these permutations and also in identifying the demand pattern for these permutations.

DATA-MINING IMPLEMENTATION ARCHITECTURE

A data-mining application can be an effective tool only when implemented along with some existing decision support system and an enterprise resource planning (ERP) system. The difference between a decision support system (DSS) and a data-mining application is that, unlike data-mining applications, DSSs do not perform queries or analysis of data. Figure 2 illustrates the manner in which this marriage between DSS and data mining affects the implementation architecture.

An ERP system integrates the processes of an organization. Data related to these processes are scattered throughout the system. This large amount of unorganized data cannot be used as it is for data-mining applications. A data warehouse has to be built that organizes ERP data so that it is easily accessible and organized for online analysis.

The smooth flow of information from an ERP to a data-mining tool through a data warehouse will require open metadata integration. Information describing the ERP data, the target data warehouse schema, and the data mappings and transformation rules need to be stored in an open relational metadata repository and be easily accessible to the data- mining tools used in the architecture. This facilitates changes made to the underlying warehouse. It then passes these changes to the analysis tools and makes complete information available to end users, such as the source of the data flows and frequency of updates (Coombs, 1999). Some of the issues in implementing data mining include result interpretation, data selection and representation, and system scalability considerations.

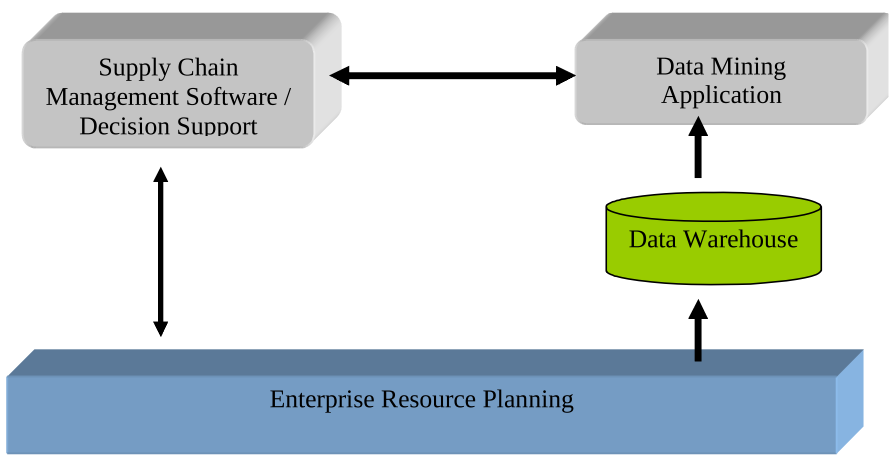
IMPLICATIONS FOR MANAGEMENT

Data mining creates fertile ground for the invention of new tools, analytical methods, and data management to add value to an organization's most valuable asset—its data. Organizations have experienced paybacks of 10 to 70 times their data warehouse investment after data-mining components are added (Chen, Sakaguchi, & Frolick, 2000). But it is not meant to be used for simple querying or reporting purposes, nor complex queries where the parameters are known. It is ideal for cases where only the problem is known, and neither the parameters nor the values of the parameters are known.

It is more important to formulate the business problems and opportunities rather than to start at the technology side. A company should ask the following questions:

1. What are the main business problems and opportunities in the supply chain?
2. What knowledge does a company need to solve these problems or explore opportunities?
3. *Can a company use this knowledge to take appropriate action?*
4. Does a company have the necessary (historical) data available on demand, supply and process behavior to make an analysis potentially successful?
5. *Given a company's knowledge need, what would be the appropriate technique to analyze the available data?*

Figure 2. The data mining implementation architecture



After answering all these questions, a company may decide whether data mining is an appropriate technology.

CONCLUSION

A common characteristic shared by the industry users of data-mining technology is the data intensive. The benefits of data mining have been sufficiently proven in a wide variety of business sectors; it's time to expand the horizon of the application of data mining to other potential areas such as supply chain management. This article makes an initial effort to explore the possibility of using data mining in supply chain management. There is a huge potential for improvement using the techniques of data mining. If small areas can be identified and an incremental approach followed to studying supply chain management, the early adopters of this technology will reap major benefits. Considering the capabilities of today's data-mining software, the power of hidden patterns and relations in supply chain data will lead to important differentiating factors in the competitive global business environment.

REFERENCES

Asbrand, D. (1997). Is data mining ready for the masses? *Datamation*, 11.

Chen, L., Sakaguchi, T., & Frolick, M.N. (2000, Winter). Data mining methods, applications, and tools. *Information Systems Management, 17*(1), 65-70.

Coombs, J. (1999, May). A decision support portal for SAP and other ERP applications. *Data Warehousing, 7*. Retrieved on January 18, 2002 from *http://www.dw-institute.com/research/display.asp?id=5522*

Fabris, P. (1998). Data mining. *CIO Magazine.*

Foley, J. & Russell, J. (1998). Mining your own business. *Information Week*, 18-22.

Grossman, R. (1998). Supporting the data mining process with next generation data mining systems. *Enterprise Systems Journal*, 52-56.

Groth, R. (2000).*Data mining: Building competitive advantage* (pp. 191-209). Englewood Cliffs, NJ: Prentice Hall.

Kopanakis, I. & Theodoulidis, B. (2003). Visual data mining modeling techniques for the visualization of mining outcomes. *Journal of Visual Languages & Computing, 14*, 543-589.

Meta Group Inc. (1997, November). Data mining: Trends, technology, and implementation imperatives.

Peacock, P.R. (1998). Data mining in marketing: Part 1. *Marketing Management, 6*(4), 9-20; Data mining in marketing: Part 2. *Marketing Management, 7*(1), 14-25.

Weir, J. (1998, Fall). Data mining: Exploring the corporate asset. *Information Systems Management*, 68-71.

Zdanowicz, J. S. (2004, May). Detecting money laundering and terrorist financing via data mining. *Communications of the ACM*, 62-75.

KEY TERMS

Data Mining: Process of extracting ideas from data.

Data Warehouse: System for managing data for decision support.

Induction Techniques (Decision Trees and Rule Induction): Assign the largest portion of the discovery process to machine (Nelson, 1998).

Logistics: The art of moving goods within the supply chain.

Neural Networks: Build internal representations of the patterns (Nelson, 1998).

Statistics: High level of user involvement to build models describing the behavior of the data (Nelson, 1998).

Data Mining in Franchise Organizations

Ye-Sho Chen
Louisiana State University, USA

Bin Zhang
Louisiana State University, USA

Bob Justis
Louisiana State University, USA

INTRODUCTION

Franchising has been a popular business approach given the high rate of business failures (Justis & Judd, 2002; Thomas & Seid, 2000), and its popularity continues to increase in today's e-business-centered global economy. For example, *Entrepreneur* magazine¾well known for its Franchise 500 listing¾in early 2001 included a category called Tech Businesses into its Franchise Zone which contains subcategories of Internet Businesses, Tech Training, and Miscellaneous Tech Businesses. At the time of this writing, 35 companies are on the list of Entrepreneur.com. Netchising, combing the power of the Inter*net* for global demand-and-supply processes and international fran*chising* arrangements for local responsiveness, seems to rise as an effective global e-business growth strategy (Beck & Morrison, 2000; Morrison, Beck & Bouquet, 2000). The Netchising business model "offers potentially huge benefits over traditional exporting or foreign direct investment approaches to globalization" and is "a powerful concept with potentially broad applications" (Davenport, 2000, p. 52).

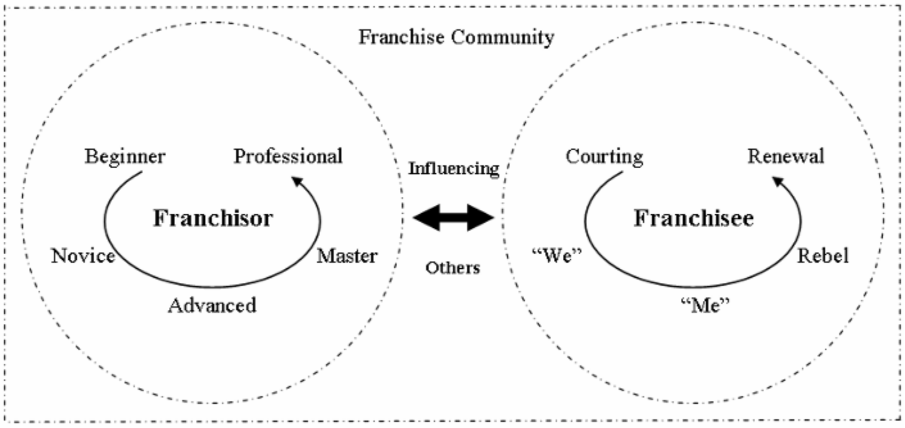
In his best seller, *Business @ the Speed of Thought*, Bill Gates (1999) wrote: "Information technology and business are becoming inextricably interwoven. I don't think anybody can talk meaningfully about one without talking about the other" (p. 6). Gates' point is quite true indeed when one talks about data mining in franchise organizations. Despite its popularity as a global e-business growth strategy, there is no guarantee that the franchising business model will render continuous success in the hypercompetitive environment. This can be evidenced from the constant up-and-down ranking of the Franchise 500 by Entrepreneur.com. Thus, to see how data mining can be "meaningfully" used in franchise organizations, one needs to know how franchising really works. In the next section, we show that (1) building up a good "family" relationship between the franchisor and the franchisee is the real essence of franchising, and (2) proven working knowledge is the foundation of the "family" relationship. We then discuss in the following three sections the process of how to make data mining "meaningful" in franchising. Finally, future trends of data mining in Netchising are briefly described

FRANCHISING: THE FRANCHISOR/ FRANCHISEE RELATIONSHIP

Franchising is "a business opportunity by which the owner (producer or distributor) of a service or a trademarked product grants exclusive rights to an individual for the local distribution and/or sale of the service or product, and in return receives a payment or royalty and conformance to quality standards. The individual or business granting the business rights is called the *franchisor*, and the individual or business granted the right to operate in accordance with the chosen method to produce or sell the product or service is called the *franchisee*" (Justis & Judd, 2002, pp. 1-3). Developing a good "family" relationship between the franchisor and the franchisee is believed to be the most important factor for the success of a franchise (Justis & Judd, 2002). Figure 1 describes how such a "family" relationship is built in the franchise community.

In Figure 1, the franchisor has to learn to deal with many issues to increase the business. The learning process is incrementally developed through the following five stages (Justis & Judd, 2002): (1) Beginner¾learning how to do it; (2) Novice¾practicing doing it; (3) Advanced¾doing it; (4) Master¾teaching others to do it; and (5) Professional¾becoming the best that you can be. At a higher stage of the development, most of the problems in the previous stages have been dealt with. However, more complicated and challenging questions will arise as the franchise continues the expansion. This is especially true when the system reaches the stage of

Figure 1. Understanding how the franchisor/franchisee "family" relationship works



Professional, when many unforeseen and difficult problems could happen all of sudden. Bud Hadfield (1995), the founder of Kwik Kopy franchise and the International Center of Entrepreneurial Development, aptly stated: "The more the company grows, the more it will be tested" (p. 156). To capture the learning process, a counter-clockwise round arrow surrounding the franchisor is used to depict the increasing intensity of learning as the franchisor continues surviving and thriving.

To understand how the "family" relationship is developed, one needs to know the five phases of franchisee life cycle (Schreuder, Krige & Parker, 2000): (1) Courting: both the franchisee and the franchisor are excited with the relationship; (2) "We": the relationship starts to deteriorate, but the franchisee still values the relationship; (3) "Me": the franchisee starts to question the reasons for payments-related issues with the attitude that the success so far is purely of his/her own work; (4) Rebel: the franchisee starts to challenge the restrictions being placed; and (5) Renewal: the franchisee realizes the "win-win" solution is to continue teaming up with the franchisor to grow the system. Similar to the franchisor, a counter-clockwise round arrow surrounding the franchisee is used in Figure 1 to depict the increasing intensity of franchisee life cycle as the franchisee continues learning and growing in the franchise system.

While the franchisee progresses through the life cycle, the "family" relationship is gradually developed through a process of influencing others (Justis & Vincent, 2001), as depicted in Figure 1 with a bi-directional arrow: (1) working knowledge, proven abilities of expanding the franchise system profitably; (2) positive attitude, constructive ways of presenting and sharing the working knowledge; (3) good motivation, providing incentives for learning or teaching the working knowledge; (4) positive individual behavior, understanding and leveraging the strengths of the participants to learn and enhance the working knowledge; and (5) collaborative group behavior, having the team spirit to find the best way to collect, dissimilate, and manage the hard-earned working knowledge. By going through the processes of learning and influencing, both the franchisor and franchisee gain the working knowledge in the franchise community progressively. The franchisor, the franchisee, and the franchise community in Figure 1 are surrounded with dashed lines, meaning there is no limit to the learning process.

MANAGING FRANCHISE ORGANIZATIONAL DATA

There are many contact points within the franchise community where the franchisor and the franchisee can influence each other; in business these contact points are sometimes called "touchpoints." Based on the Customer Service Life Cycle (CSLC) model developed by Ives (2004), Chen, Chong, and Justis (2002) proposed a framework (see Table 1) to harness the Internet to serve the customers for the franchising industry. The 11 sub-stages in Table 1 are based on two well-known franchising books by Justis and Judd (2002) and Thomas and Seid (2000). The model in Table 1 may be used as a comprehensive guide for a franchise system to develop its Web site, especially at the stages of Requirements and Acquisition.

Table 1 also is a comprehensive framework for a franchise organization to model the data needed to serve its customers, that is, franchisees and their customers. A well-designed Internet strategy shall empower the franchisor and the franchisees to collect, use, renew, store, retrieve, transmit, and share the organizational data needed to do the collaborative work in different phases of the CSLC model. Specifically, three types of data are needed:

- *Operational data:* the daily activities at (1) the franchisor headquarters, including six major entity types: employees; business outlets owned by franchisees or companies; prospective franchisees; product development; suppliers (e.g., marketing agents, accountants, insurance providers, attorneys, real estate agents); and government offices, (e.g., taxes and worker compensation); and (2) the franchisee business outlet, including six major entity types: customers, employees, contacts with the headquarters, product inventory, suppliers, and government offices.
- *External data:* the relationship management activities in the franchise community, including three major entity types: the relationship with customers, the relationship with partners and suppliers, and the performance benchmarks in the industry.
- *Legacy data:* the activities that have been working well or gradually adapted since the franchise system came into existence. Examples include (1) rewarding activities to the top performers among the franchisees; (2) efficient procedural activities for the employees at the headquarters supporting the franchisees; and (3) effective and friendly face-to-face activities for the field representatives to serve the franchisees at their outlets.

MANAGING FRANCHISE ORGANIZATIONAL INFORMATION

An architecture, adapted from Inmon (1996), of data mining in franchise organizations with respect to the franchisor/franchisee relationship management depicted in Figure 1 is shown in Figure 2. The architecture consists of four levels: (1) data collection, holding operational, external, and legacy data collected from the franchise

Table 1. The customer-service-life-cycle model in franchising

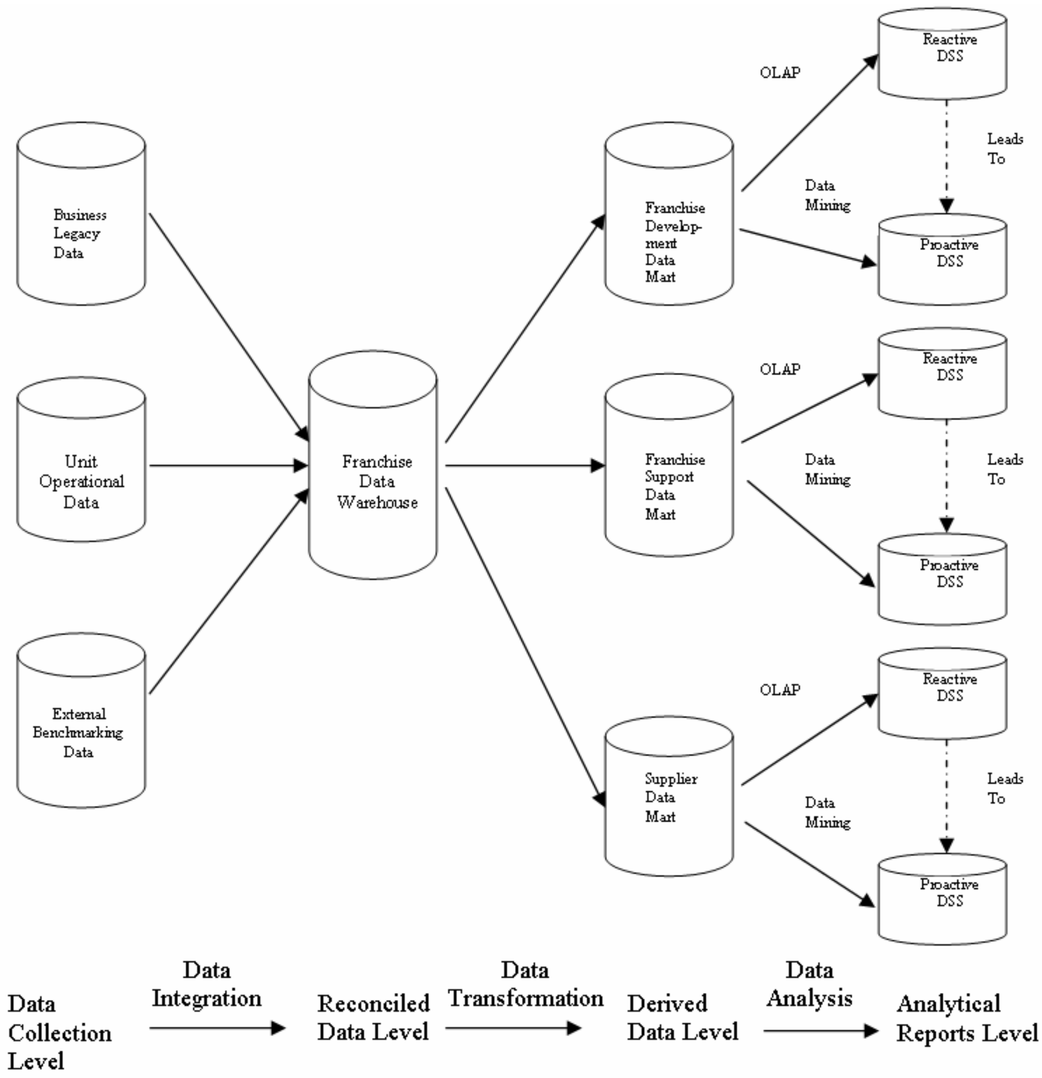
CSLC	Sub-stages	Example: Technology Strategies of WSI Internet (www.wsicorporate.com)
Requirements	Understanding How Franchising Works	Internet
	Investigating Franchise Opportunities	Internet • Global Gateway • Internet Solutions • Portfolio & Technologies • About Us • Franchise Opportunities • Experts Online • Interactive Online
	Obtaining Franchisee Prospectus	Internet • E-mail
	Making the Choice	Internet
Acquisition	Preparing Business Plan	Internet
	Financing the Franchised Business	Internet
	Signing the Contract	Internet
Ownership	Marketing & Promoting the Franchise Products/Services	Internet/Intranet/Extranet • Need a Website? • Live Call • Employment @ WSI • Hot News • WSI ICE Flash • Message from the President
	Managing the Franchise System	Internet Intranet • Serving Franchisees' Customers Extranet
	Building the Relationship between the Franchisor and the Franchisee • The Courting Phase • The "We"-Phase • The "Me"-Phase • The Rebel Phase • The Renewal Phase	Internet Intranet • Knowledge Centre • Training at Headquarters • Newsletter • Meetings • Toll-free Phone Line Extranet • Purchasing Cooperatives
Renewal or Retirement	Becoming a Professional Multi-unit Franchisee or Retiring from the Franchise System	Internet Intranet Extranet

business environment; (2) reconciled data, holding data warehouse data that are subject-oriented, integrated, time-variant, and non-volatile (Inmon, 1996); (3) derived data, containing several data marts (e.g., franchisees, customers, competitors, and suppliers) derived from the data warehouse based on various franchisee/customer-centered segmentations; and (4) the analytical reporting level, producing various relationship performance reports (e.g., business outlet periodical summary, financial, scorecards, and mysterious shopping) for the decision makers using the decision support systems (DSS) for their decision making. To move from the data collection level to the reconciled data level, data integration is needed. It is a very time-consuming process that involves the activities such as cleansing, extracting, filtering, conditioning, scrubbing, and loading. To move from the reconciled data level to the derived data level, data transformation is needed which involves the activities such as replication, propagation, summary, aggregate, and metadata. To move from the derived data level to the analytical reporting level, data analysis is needed which involves two major activities¾online analytical processing (OLAP) and data mining.

A typical OLAP analysis consists of pre-defined multi-dimensional queries. Some examples in franchising are shown below:

- Show the gross margin by product category and by franchise outlets from Thanksgiving to Christmas in the last five years.

Figure 2. An architecture of data mining in franchise organizations



- Which franchise outlets are increasing in sales and which are decreasing?
- Which kinds of customers place the same orders on a regular basis at certain franchise outlets?
- How many franchisees did we lose during the last quarter of 2001, compared to 2000, 1999, and 1998?

Other OLAP activities include spreadsheet analysis, data visualization, and a variety of statistical data modeling methods. Since the query activities are pre-defined, we call the supporting systems reactive DSS.

Data mining, on the other hand, is used to identify hidden relationship patterns of the data residing in the data marts. Typical data mining modeling analysis can be classified into the following three categories:

- Classification and Prediction, using techniques such as RFM (recency, frequency, and monetary), regression, decision tree, and neural network
- Association Rules, using techniques such as market basket analysis, correlation analysis, cross-sell analysis, and link analysis
- Cluster Analysis, using techniques such as partition, hierarchy, outlier, and density analysis

Table 2, adapted from Delmater and Hancock (2001), shows that data mining techniques can be used to help serve franchisees' customers at the different stages of the CSLC model.

Since the data mining queries and related activities are not pre-defined, we call the supporting systems proactive DSS. A major drawback of proactive data mining is that without careful preliminary examination of the data characteristics, the mining activities may end in vain (Delmater & Hancock, 2001). In order to achieve a higher success rate of data mining, we suggest (on the right side of Figure 2) that OLAP-based queries need to be conducted first. For example, one may find, through daily OLAP queries, that certain segments of customers buy certain products frequently. This pattern may lead us to do more in-depth proactive analysis of the customers and products relationships. The result may help the company serve customers better and generate higher profits.

Table 2. Franchisees' customers data mining using the CSLC approach

CSLC	Explanation	Data Mining Activities (and Techniques Used)
Requirements	Finding and reaching the customers	• Lead Generation • Market Analysis & Segmentation (Classification and Prediction) • Mining Web Site Visitors (Association Rules) • Text Mining Usenet Newsgroups (Cluster Analysis)
Acquisition	Selling to the customers	• Customer Acquisition Profiling • Customer Segmentation Strategy (Classification and Prediction) • Online Shopping Tracking (Association Rules) • Pricing Strategy (Association Rules) • Customer-centric Selling (Association Rules) • Text Mining Contact E-Mails (Cluster Analysis) • Scenario Notification (Association Rules)
Ownership	Satisfying the customers after the sales	• Customer Service • Inquiry Routing (Association Rules) • Text Mining E-Mails & Inquiries (Cluster Analysis) • Scenario Notification (Association Rules) • Staffing Level Prediction (Classification and Prediction)
Retirement or Renewing	Retaining the customers so that you can continue coming back	• Customer Retention • Sharper Customer Focus through Loyalty Program (Classification and Prediction) • Detecting Customer Complaints through Text Mining (Cluster Analysis) • Detecting Inappropriate Customer Services (Cluster Analysis) • Individual Customer Profiles (Classification and Prediction) • Scenario Notification (Association Rules)

MANAGING FRANCHISE ORGANIZATIONAL KNOWLEDGE

As was mentioned in the discussions of Figure 1, the key for building the franchisor/franchisee "family" relationship is in the franchise organizational learning. In addition, there are five vital factors for a successful learning program: knowledge, attitude, motivation, individual behavior, and group behavior. Thus, working knowledge is the real foundation of a successful franchise "family" relationship. The working knowledge is structured in many forms of profiles that are embedded in the operational manuals of the franchise business processes. Table 3 gives some examples of those working knowledge profiles with respect to the CSLC business processes associated with the sub-stages in Table 1.

A working knowledge profile is developed when a certain task of the CSLC process is repeated many times with good results. Consider the Site Profile used at the "Marketing & Promoting the Franchise Products/Services" sub-stage in Table 3. The Site Profile is used to help the new franchisee to find a good business site, and typically it is the real estate department at the franchisor headquarters which is responsible in developing the profile. The Site Profile is continuously being tested and enhanced. Various OLAP/Data Mining analytical reports monitoring the performance of the sites are generated at the Analytical Reports Level shown in Figure 2. Based on those reports, the real estate experts and their teams are able to fine-tune the attributes and the parameters within the Site Profile. Most often, the corresponding data collection procedures in the CSLC sub-stage also need to be revised and perfected so that better report scorecards can be generated.

This process of enhancing the working knowledge profile will achieve its high peak when both the franchisor and the franchisees are arriving at the Professional and Renewal stage of growth. A significant phenomenon of being a Professional franchisor and a Renewal franchisee are their ability to leverage the assets of the hard-earned working knowledge profiles into dynamic capabilities and high-business-value-creation completive-advantage strategies. The new products or services coming out of the process of leveraging the working knowledge profiles may transform the franchise business into a more, sometimes surprisingly, profitable enterprise. The capability of leveraging the assets of franchise working knowledge into profitable products or services is at the heart of a successful franchise.

Consider as an example the site selection working knowledge at McDonald's. The Franchise Realty Corporation real estate business, a result of site selection asset leveraging, is the real money-making engine at McDonald's. This can be evidenced from the following speech of Ray Kroc, founder of McDonald's, to the MBA class at the University of Texas at Austin in 1974: "Ladies and gentlemen, I'm not in the hamburger business. My business is real estate" (Kiyosaki, 2000, p. 85), In the book *McDonald's: Behind the Arches* (Love, 1995, p. 152), Kroc commented further: "What converted McDonald's into a money machine had nothing to do with Ray Kroc or the McDonald brothers or even the popularity of McDonald's hamburgers, French fries, and milk shakes. Rather, McDonald's made its money on real estate…." McDonald's makes money out of real estate by leasing properties from landlords and then subleasing the stores to the franchisees. The Professional franchisees, many of them are multi-unit operators, can then focus on expending the business without worrying about finding good locations for the growth. This money-making real estate strategy is what separates McDonald's from other fast-food chains (David, 2003).

Knowledge repository systems, consisting of working knowledge profiles such as the one shown in Figure 3, can be linked into the franchisor headquarters and the franchisee outlets for knowledge sharing and learning. Such a repository has two dimensions. First, there is a working knowledge level for the collaborative team, the franchisee outlet, the franchisor headquarters, and the franchise community. Second, there are user skill levels, including Beginner in the Courting Phase, Novice in the "We"-Phase, Advanced in the "Me"-Phase, Master in the Rebel Phase (since the rebel ones tend to be those who know the system very well and are capable of influencing others to follow them), and Professional in the Renewal Stage of the franchisee life cycle. The foundation of the framework is the working knowledge of the five crucial elements—Knowledge, Attitude, Motivation, Individual Behavior, and Group Behavior—used by the collaborative team, to effectively influence others in building the franchise "family" relationship. The working knowledge profiles at the franchisee outlet, the franchisor headquarters, and the franchise community can be modularized according to the user's level. An intranet-based curriculum of working knowledge modules can then be designed for the users to learn the working knowledge profiles effectively.

FUTURE TRENDS

The third industrial revolution, combining Internet technology with globalization, produces various new data mining opportunities for the growth of franchise organizations. For example, knowledge network applications, using data mining techniques such as social network analysis (Wasserman & Faust, 1994), can be developed to connect Professional franchisees in the world. The goal

Table 3. The CSLC model of franchise working knowledge

CSLC Sub-stages	Examples of Working Knowledge Profiles
Understanding How Franchising Works	• Lead Generation Profile • Website Visitor Profile
Investigating Franchise Opportunities	• Benchmark Profile • Successful Franchisee Profile
Obtaining Franchisee Prospectus	• Prospectus Profile
Making the Choice	• Competitor Profile
Preparing Business Plan	• Business Plan Profile
Financing the Franchised Business	• Financing Institute Profile • Non-traditional Franchising Profile
Signing the Contract	• Franchisee Profile
Marketing & Promoting the Franchise Products/Services	• Site Profile • Customer Profile • Product Profile
Managing the Franchise System	• Support Team Profile • Employee Profile • Supplier Profile
Building the Relationship between the Franchisor and the Franchisee	• Event Management Profile • Best Practices Profile • Crisis Management Profile
Becoming a Professional Franchisee or Retiring from the Franchise System	• Multi-unit Franchisee Profile • Co-branding Profile • Opportunities Profile • Social Network Profile

is to enable the franchise system to venture into new global emerging markets¾ for example, China¾ through international franchising, and develop innovative products/services through asset leveraging. This could be done because franchise capabilities, structured in the working knowledge repository shown in Figure 3, enable the Professional franchisees to work with the franchisor to continuously improve and leverage the current franchise working knowledge. An example of knowledge networks of Professional franchisees can be illustrated in Figure 4. There are six Professional franchisees (A-F) in the figure, with three clusters (A-C, D-E, and F) of knowledge networks. Each Professional franchisee (a dot) has his/her personal knowledge network (arrows pointing out of the dot) tested and built over the years while doing day-to-day problem solving at the franchisee outlet. The knowledge network may include the customers' likes and dislikes, the kind of employees to hire, the competitors' and suppliers' pricing strategies, and the social needs in the local community. Each Professional franchisee is surrounded with a circle with dashed lines, meaning there is no limit to the personal knowledge network. In order to solve the problems more effectively, Professional franchisees may share with each other their approaches. Thus, clusters (connected dots) of knowledge network are formed for solving various problems more effectively.

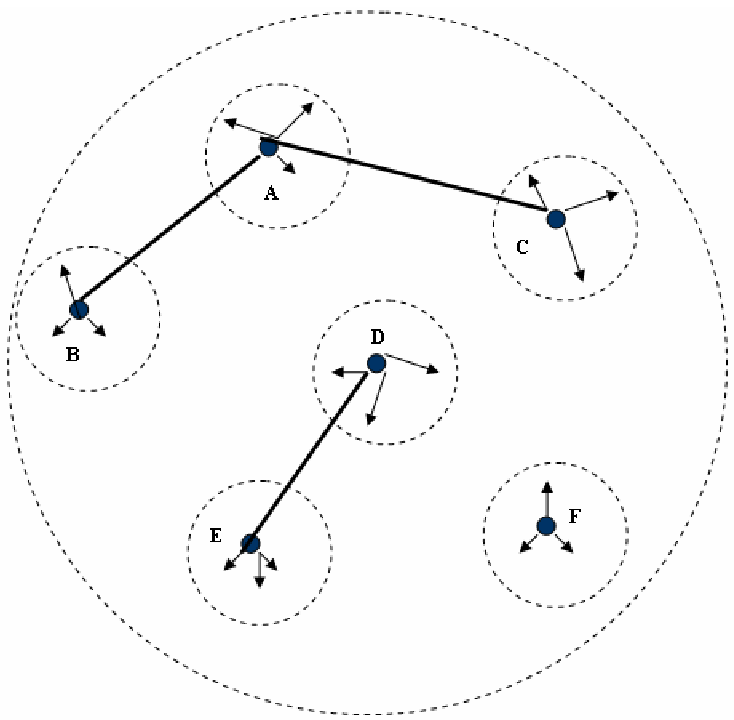
CONCLUSION

Franchising has been popular as a growth strategy for small businesses; it is even more so in today's global and e-commerce world. The essence of franchising lies in managing the "family" relationship between the franchisor and the franchisee. In this article we showed data mining plays an important role in growing and nurturing such a "family" relationship. Specifically, we discussed: (1) how franchise organizational data can be managed effectively using the methodology of Customer Service Life Cycle; (2) how franchise organizational information is deciphered from the customer-centered data using OLAP and data mining analytical techniques; and (3) how the franchise organizational knowledge is leveraged to grow the franchise system. The ability to continue leveraging the

Figure 3. Working knowledge repository in franchise organizations

		User Skill Levels				
		Beginner in the Courting Phase: Beginner Guide	Novice in the "We"-Phase: Practicing	Advanced in the "Me"-Phase: Doing	Master in the Rebel Phase: Teaching Others	Professional in the Renewal Stage: Improving and Leveraging
Working Knowledge Levels	Collaborative Team	Process of Influencing Others for Knowledge Sharing: Knowledge, Attitude, Motivation, Individual Behavior, and Group Behavior				
	Franchisee Outlet	Working Knowledge Profiles for Running the Franchisee Outlet: Customer Profile, Employee Profile, Product Profile				
	Franchisor Headquarters	Working Knowledge Profiles for Running the Franchisor Headquarters: Franchisee Profile, Site Profile, Product Profile, Employee Profile, Event Management Profile				
	Franchise Community	Working Knowledge Profiles for Relationship Management with the Community: Supplier Profiles, Community Profiles				

Figure 4. Knowledge networks of professional franchisees



organizational knowledge assets based on the good "family" relationship is really what a franchise business is about.

REFERENCES

Beck, J.C., & Morrison, A. (2000). *Netchising: Creating a profitable global strategy*. Working Paper, Institute for Strategic Change, Accenture.

Chen, Y., Chong, P.P., & Justis, R.T. (2002, February 8-10). E-business strategy in franchising: A customer-service-life-cycle approach. *Proceedings of the 16th Annual International Society of Franchising Conference*, Orlando, Florida.

Davenport, T. (2000). E-commerce goes global. *CIO Magazine*, *13*(20), 52-54.

David, G. (2003, March 30). Can McDonald's cook again? *Fortune Magazine*.

Delmater, R., & Hancock, M. (2001). *Data mining explained: A manager's guide to customer-centric business intelligence*. Digital Press.

Gates, W. (1999). *Business @ the speed of thought*. Warner Books.

Hadfield, B. (1995). *Wealth within reach*. Cypress Publishing.

Inmon, W.H. (1996). *Building the data warehouse*. John Wiley & Sons.

International Network for Social Network Analysis. (2004). Retrieved February 20, 2004, from *www.sfu.ca/~insna*

Ives, B. (2004). Want to Web empower customer service? Retrieved February 20, 2004, from *isds.bus.lsu.edu/cvoc/projects/cslc/html*

Justis, R.T., & Judd, R.J. (2002). *Franchising*. DAME Publishing.

Justis, R.T., & Vincent, W.S. (2001). *Achieving wealth through franchising*. Adams Media Corporation.

Kiyosaki, R. (2000). *Rich dad, poor dad*. Time Warner.

Love, J. (1995). *McDonald's: Behind the arches*. Bantam Books.

Morrison, A., Beck, A., & Bouquet, C. (2000). Netchising: The next global wave. *eCommerce and global business forum: Building electronic bridges across nations*. Santa Cruz, California.

Schreuder, A.N., Krige, L., & Parker, E. (2000, February 19-20). The franchisee lifecycle concept–a new paradigm in managing the franchisee-franchisor relationship. *Proceedings of the 14th Annual International Society of Franchising Conference*, San Diego, California.

Thomas, D., & Seid, M. (2000). *Franchising for dummies*. IDG Books.

KEY TERMS

Customer Service Life Cycle: Serving customers based on a process of four stages: Requirements, Acquisition, Ownership, and Retirement. Many companies are using the approach to harness the Internet to serve the customers.

Data Mart: A small database with data derived from a data warehouse.

Data Mining: Analytical techniques used to find out the hidden relationships or patterns residing in the organizational data.

Data Warehouse: A database that is subject-oriented, integrated, time-variant, and non-volatile.

Franchisee: The individual or business who receives the business rights and pays the royalties for using the rights.

Franchisee Life Cycle: The stages a franchisee goes through in the franchise system: Courting, 'We,' 'Me,' Rebel, Renewal.

Franchising: A business opportunity based on granting the business rights and collecting royalties in return.

Franchisor: The individual or business who grants the business rights.

Franchisor/Franchisee Learning Process: The stages of learning, including Beginner, Novice, Advanced, Master, and Professional.

Franchisor/Franchisee Relationship Management: The vital factor for the success of a franchise, including: Knowledge, Attitude, Motivation, Individual Behavior, and Group Behavior.

D

Data Mining in Practice

Sherry Y. Chen
Brunel University, UK

Xiaohui Liu
Brunel University, UK

INTRODUCTION

There is an explosion in the amount of data that organizations generate, collect, and store. Organizations are gradually relying more on new technologies to access, analyze, summarize, and interpret information intelligently and automatically. Data mining, therefore, has become a research area with increased importance (Amaratunga & Cabrera, 2004). Data mining is the search for valuable information in large volumes of data (Hand, Mannila, & Smyth, 2001). It can discover hidden relationships, patterns, and interdependencies, and generate rules to predict the correlations, which can help the organizations make critical decisions faster or with a greater degree of confidence (Gargano & Raggad, 1999).

There is a wide range of data mining techniques that has been successfully used in many applications. This paper is an attempt to provide an overview of existing data-mining applications. The paper begins by building a theoretical background to explain key tasks that data mining can achieve. It then moves to discuss applications domains that data mining can support. The paper identifies three common application domains, including bioinformatics data, electronic commerce, and search engines. For each domain, how data mining can enhance the functions will be described. Subsequently, the limitations of current research will be addressed, followed by a discussion of directions for future research.

BACKGROUND

Data mining can be used to achieve many types of tasks. Based on the types of knowledge to be discovered, it can be broadly divided into supervised learning and unsupervised learning. The former requires the data to be preclassified. Each item is associated with a unique label, signifying the class in which the item belongs. In contrast, the latter does not require preclassification of the data and can form groups that share common characteristics (Nolan, 2002). To achieve these two main tasks, four data mining approaches are commonly used: classification, clustering, association rules, and visualization.

Classifications

Classification, which is a process of supervised learning, is an important issue in data mining. It refers to discovering predictive patterns where a predicted attribute is nominal or categorical. The predicted attribute is called the class. Subsequently, a data item is assigned to one of predefined sets of classes by examining its attributes (Changchien & Lu, 2001). One example of classification applications is to analyze the functions of genes on the basis of predefined classes that biologists set (see "Classifying Gene Functions").

Clustering

Clustering is also known as *Exploratory Data Analysis* (EDA; Kohonen, 2000). This approach is used in those situations where a training set of preclassified records is unavailable. Objects are divided into groups based on their similarities. Each group, called a cluster, consists of objects that are similar between themselves and dissimilar to objects of other groups (Roussinov & Zhao, 2003). From a data mining perspective, clustering is unsupervised learning of a hidden data concept. One of the major applications of clustering is the management of customers' relationships, which is described in "Customer Management."

Association Rules

Association rules, which were first proposed by Agrawal and Srikant (1994), are mainly used to find out the meaningful relationships between items or features that occur synchronously in databases (Wang, Chuang, Hsu, & Keh, 2004). This approach is useful when one has an idea of the different associations that are being sought out. This is because one can find all kinds of correlations in a large data set. It has been widely applied to extract knowledge from Web log data (Lee, Kim, Chung, & Kwon, 2002). In particular, it is very popular among marketing managers and retailers in electronic commerce who want to find associative patterns among products (see "Market Basket Analysis").

Visualization

The visualization approach to data mining is based on an assumption that human beings are very good at perceiving structure in visual forms. The basic idea is to present the data in some visual form, allowing the human to gain insight from the data, draw conclusions, and directly interact with the data (Ankerst, 2001). Since the user is directly involved in the exploration process, shifting and adjusting the exploration goals is automatically done if necessary (Keim, 2002). This approach is especially useful when little is known about the data and the exploration goals are vague. One example of using visualization is author co-citation analysis (see "Author Cocitation Analyses").

DATA-MINING APPLICATIONS

As mentioned in the aforementioned discussion, data mining can be used to achieve various types of tasks, such as classification, clustering, association rules, and visualization. These tasks have been implemented in many application domains. The main application domains that data mining can support include bioinformatics data, electronic commerce, and search engines.

Bioinformatics Data

In the past years, bioinformatics has been overwhelmed with increasing floods of data gathered by the Human Genome Project. Consequently, a major challenge in bioinformatics is extracting useful information from these data. To face this challenge, it is necessary to develop an advanced computational approach for data analysis. Data mining provides such potentials. Three application areas, which are commonly presented in the literature, are described below.

Clustering Microarray Data

Unsupervised learning produces clustering algorithms, which are being applied to DNA microarray data sets. The clustering algorithms are often incorporated into the analysis software of microarray images and are therefore frequently used to visualize local and global relationships among hybridization signals captured by the array. Currently, hierarchical clustering is the most popular technique employed for microarray data analysis. Basically, this approach is based on similarity or dissimilarity to proceed successively by either merging smaller clusters into larger ones or by splitting larger clusters (Moller-Levet, Cho, Yin, & Wolkenhauer, 2003). The results of the algorithm are a tree of clusters called a *dendrogram* (Liu & Kellam, 2003). One disadvantage of this approach is that clustering algorithms are very sensitive to noisy data and errors in the data set.

Classifying Gene Functions

Biologists often know a subset of genes involved in a biological pathway of interest and wish to discover other genes that can be assigned to the same pathway (Ng & Tan, 2003). Unlike clustering, which processes genes based on their similarity, classification can learn to classify new genes based on predefined classes, taking advantage of the domain knowledge already possessed by the biologists. Therefore, the classification approach seems more suitable than clustering for the classification of gene functions. To conduct the classification, we need supervised learning to assign pathway memberships that correspond well to the true underlying biological pathways.

Identifying Phenotype Data

In the two aforementioned approaches, the genes are treated as objects while the samples are the attributes. Conversely, the samples can be considered as the objects and the genes as the attributes. In this approach, the samples can be partitioned into homogeneous groups. Each group may correspond to some particular phenotype (Golub, et al., 1999). A phenotype is the observable and physical characteristics of an organism. Over the past decade, growing interest has surfaced in recognizing relationships between the genotypes and phenotypes. Tracing a phenotype over time may provide a longitudinal record for the evolution of a disease and the response to a therapeutic intervention. This approach is analogous to removing components from a machine and then attempting to operate the machine under different conditions to diagnose the role of the missing component. Functions of genes can be determined by removing the gene and observing the resulting effect on the organism's phenotype.

Electronic Commerce

The widespread use of the Web has tremendous impact on the way organizations interact with their partners and customers. Many organizations consider analyzing customers' behavior, developing marketing strategies to create new consuming markets, and discovering hidden loyal customers as the key factors of success. Therefore, new techniques to promote electronic business become essential, and data mining is one of the most popular techniques (Changchien & Lu, 2001). Data mining applications in electronic commerce include customer management and market basket analysis.

Customer Management

For analyzing customers' behaviors, a frequently used approach is to analyze their usage data in order to discover user interests, and then recommendations can be made based on the usage data extracted. Wang et al. (2004) have developed a recommendation system for the cosmetic business. In the system, they segmented the customers by using clustering algorithms to discover different behavior groups so that customers in the same group have similar purchase behavior. For each group's customers, they used the association-rules algorithm to discover their purchase behavior. In addition, they scored each product for each customer who might be interesting in it with the collaborative filtering approach and content-based filtering. They found that this approach could not only recommend the right product to the right person, but also recommend the right product to the right person at the right time.

Market Basket Analysis

Market basket analysis is a typical example among the various applications of association rules, and it aims at identifying associations between consumers' choices of different products (Giudici & Passerone, 2002). The data analyzed in a market basket analysis usually consists of all buying transactions carried out by the consumers in a certain unit of time. The market basket analysis refers to the application of data-analysis techniques to databases that store transactions from consumers buying choices of different products. The aim of the analysis is to understand the association structure between the sales of the different products available. Once the associations are found, they may help in planning marketing policies. For instance, if there is a relationship between two products over time, then retailers can use this information to contact the customer, decreasing the chance that the customer will purchase the product from a competitor. This is the type of data typically analyzed with association rules.

Search Engines

Data mining is of increasing importance for search engines. Traditional search engines offer limited assistance to users in locating the relevant information they need. Data mining can help search engines to provide more advanced features. According to current applications, there are three potential advantages: (a) ranking of pages, (b) improvement of precision, and (c) author co-citation analysis. These advantages are described below.

Ranking of Pages

Data mining identifies the ranking of the Web pages for a particular topic by analyzing the interconnections of a series of related pages. PageRank (Kamvar, Haveliwala, Manning, & Golub, 2003) applies this approach to find pertinent Web pages. In PageRank, the importance of a page is calculated based on the number of pages that point to it. This is actually a measure based on the number of back links to a page. A back link is a link pointing to a page rather than pointing out from a page. This measure is used to prioritize pages returned from a traditional search engine using keyword searching. Google applies this measure to rank the search results. The benefit is that central, important, and authoritative Web pages are given preferences. However, the problem is that it only examines the forward direction. In addition, a much larger set of linked documents is required.

Improvement of Precision

The problem of PageRank is that it is a purely link structure-based computation, ignoring the textual content. Therefore, the precision is low. On a narrowly focused topic, it frequently returns resources for a more general topic. The IBM Almaden Research Centre continues to develop the Clever search engine (Chakrabarti et al., 1999). The main approach is to promote precision by combining content with link information, breaking large hub pages into smaller units, and computing relevance weight for pages. Users' logs are the other source that can be used to improve precision. Zhang and Dong (2002) developed a Chinese image search engine named eeFind by using the matrix analysis on search engine log (MASEL). The basic idea of MASEL is to use the query log to find relations among users, queries, and clicks on results. The relation between pages chosen after a query and the query itself provides valuable information. After a query, a user usually performs a click to view one result page.

Author Co-Citation Analyses

Author co-citation analysis (ACA) has been widely used as a method for analyzing the intellectual structure of science studies. It can be employed to identify authors from the same or similar research fields. Chen and Paul (2001) used visualization to perform such an analysis, and a 3D virtual landscape was applied to represent author co-citation structures. The most influential scientists in the knowledge domain appear near the intellectual structure's center. In contrast, researchers who have unique expertise tend to appear in peripheral areas. The virtual landscape also lets users access further details regarding a particular author in the intellectual structure, such as a list of the author's most cited papers, abstracts, and even the full content of that author's articles.

FUTURE TRENDS

The above three application domains demonstrate that data mining is a very useful technology that opens new opportunities for data analysis. However, there are still many limitations that we need to be aware of and that should be investigated in further work.

Preparation of Data

It is important to know that the effective implementation of data mining methods depends on a large extent on the availability and quality of the data. Therefore, a cleaning and data-transformation step before analysis is usually needed. In addition, we have to understand the characteristics of each application and set clear data mining targets so that truly useful information can be produced with appropriate data mining techniques.

Good Communication

Data mining is an interdisciplinary research area, which involves experts from different domains. One of the significant problems of interdisciplinary research is the wide range and levels of domains of expertise that are present among potential users, so it can be difficult to provide access mechanisms appropriate to all (Kuonen, 2003). Therefore, it is important to have good communications between data mining technologist and users, and among users from different domains.

Multimedia Mining

Basically, Web content consists of a variety of media types such as text, image, audio, video, and so forth. Multimedia mining is engaged to mine the unstructured information and knowledge from these online multimedia sources. However, most current research in the area of data mining still focuses on text data; multimedia mining has received less attention than text mining (Kosala & Blockeel, 2000) and opens a new window for future research to explore.

Evaluation of Effectiveness

As previously described, data mining enhances the functions of the applications in different domains. However, the effectiveness of these applications is still a question. Certainly, there is a need to conduct more empirical studies to verify the effectiveness and evaluate the performance of these applications. In particular, we need more evaluation from user points of view. Better and more detailed evaluation can lead to better understanding of users' needs, and help to develop more effective applications.

CONCLUSION

In summary, data mining can search for interesting relationships and global patterns from various types of resources. These relationships and patterns represent valuable knowledge about the objects that are reflected by many applications in the real world. In this paper, we have given some background to data mining and have provided an overview of three application domains, including bioinformatics data, electronic commerce, and search engines. It should be noted that data mining has also been applied to other application domains, such as digital libraries, Web-based learning, health care, and so forth. It is another direction for future research to investigate what major functions are required for each application domain and to develop concrete criteria for the evaluation of their effectiveness. These works can be integrated together to generate guidelines, which can be used for commercial and research communities to select suitable data mining techniques. The ultimate goal is to enhance the functions and performance of these applications by exploiting the full potential of data mining techniques.

ACKNOWLEDGMENT

This study is in part funded by the Engineering and Physical Sciences Research Council in the UK (grant reference GR/R57737/01) and the Arts and Humanities Research Board in the UK (grant reference MRG/AN9183/APN16300).

REFERENCES

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. *Proceedings of the 20th International Conference on Very Large Databases*, Santiago, Chile.

Amaratunga, D., & Cabrera, J. (2004). Mining data to find subsets of high activity. *Journal of Statistical Planning and Inference*, *122*(1/2), 23-41.

Ankerst, M. (2001). Visual data mining with pixel-oriented visualization techniques. *Proceedings of ACM SIGKDD Workshop on Visual Data Mining*.

Chakrabarti, S., Dom, B. E., Kumar, S. R., Raghavan, P., Rajagopalan, S., Tomkins, A., et al. (1999). Mining the Web's link structure. *Computer, 32*, 60-67.

Changchien, S., & Lu, T. (2001). Mining association rules procedure to support on-line recommendation by customers and products fragmentation. *Expert Systems with Applications*, *20*(4), 325-335.

Chen, C., & Paul, R. J. (2001). Visualising a knowledge domain's intellectual structure. *Computer*, *34*(3), 65-71.

Gargano, M. L., & Ragged, B. G. (1999). Data mining: A powerful information creating tool. *OCLC Systems Services*, *15*(2), 81-90.

Giudici, P., & Passerone, G. (2002). Data mining of association structures to model consumer behaviour. *Computational Statistics & Data Analysis, 38*(4), 533-541.

Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gassenbeck, M., Mesirov, J. P., et al. (1999). Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science*, *286*, 531-537.

Hand, D. J., Mannila, H., & Smyth, P. (2001). *Principles of data mining.* MIT Press.

Kamvar, S., Haveliwala, T., Manning, C., & Golub, G. (2003). Extrapolation methods for accelerating PageRank computations. *Proceedings of the 12th International World Wide Web Conference.*

Keim, D. A. (2002). Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, *8*(1), 1-8.

Kohonen, T. (2000). Self organization of a massive document collection. *IEEE Transactions on Neural Networks*, *11*(3), 574-585.

Kosala, R., & Blockeel, H. (2000). Web mining research: A survey. *ACM SIGKDD Explorations*, *2*(1), 1-15.

Kuonen, D. (2003). Challenges in bioinformatics for statistical data miners. *Bulletin of the Swiss Statistical Society*, *46*, 10-17.

Lee, K. C., Kim, J. S., Chung, & Kwon, S. J. (2002). Fuzzy cognitive map approach to Web-mining inference amplification. *Expert Systems with Applications*, *22*(3), 197-211.

Liu, X., & Kellam, P. (2003). Mining gene expression data. In C. Orengo, D. Jones, & J. Thornton (Eds.), *Bioinformatics: Genes, proteins and computers* (pp. 229-244).

Moller-Levet, C. S., Cho, K.-H., Yin, H., & Wolkenhauer, O. (2003). *Clustering of gene expression time series data* (Tech. Rep.). UK: UMIST.

Ng, S., & Tan, S. (2003). On combining multiple microarray studies for improved functional classification by whole-dataset feature selection. *Genome Informatics*, *14*, 44-53.

Nolan, J. R. (2002). Computer systems that learn: An empirical study of the effect of noise on the performance of three classification methods. *Expert Systems with Applications*, *23*(1), 39-47.

Roussinov, D., & Zhao, J. L. (2003). Automatic discovery of similarity relationships through Web mining. *Decision Support Systems*, *35*(1), 149-166.

Wang, Y., Chuang, Y., Hsu, M., & Keh, H. (2004). A personalized recommender system for the cosmetic business. *Expert Systems with Applications*, *26*(3), 427-434.

Zhang, D., & Dong, Y. (2002). A novel Web usage mining approach for search engine. *Computer Networks*, *39*(3), 303-310.

KEY TERMS

Author Cocitation Analysis: The analysis of how authors are cited together

Bioinformatics: An integration of mathematical, statistical, and computational methods to organize and analyze biological data

Collaborative Filtering: A technique that is used for making recommendations by computing the similarities among users

Content-Based Filtering: A technique that involves a direct comparison between the content or attributes of a user's profile and the document to make recommendations

Dendrogram: A graphical procedure for representing the output of a hierarchical clustering method. It is strictly defined as a binary tree with a distinguished root that has all the data items at its leaves.

Electronic Commerce: Commercial activities that facilitate the buying and selling of goods and services over the Internet

Microarrays: A high-throughput technology that allows the simultaneous determination of mRNA abundance for many thousands of genes in a single experiment

Multimedia Mining: A new field of knowledge discovery in multimedia documents, dealing with nonstructured information such as texts, images, videos, audio, and virtual data

Noisy Data: Errors in the data due to the nature of data collection, measurement, or sensing procedures

Search Engines: Web services that help search through Internet addresses for user-defined terms or topics

Data Warehouse Development

José María Cavero Barca
Universidad Rey Juan Carlos, Spain

Esperanza Marcos Martínez
Universidad Rey Juan Carlos, Spain

Mario G. Piattini
Universidad de Castilla-La Mancha, Spain

Adolfo Sánchez de Miguel
Cronos Ibérica, S.A., Spain

INTRODUCTION

The concept of data warehouse first appeared in Inmon (1993) to describe a "subject oriented, integrated, non-volatile, and time variant collection of data in support of management's decisions" (31). It is a concept related to the OLAP (online analytical processing) technology, first introduced by Codd et al. (1993) to characterize the requirements of aggregation, consolidation, view production, formulae application, and data synthesis in many dimensions. A data warehouse is a repository of information that mainly comes from online transactional processing (OLTP) systems that provide data for analytical processing and decision support.

The development of a data warehouse needs the integration of data that come from different sources, mainly legacy systems. The development of a data warehouse is, like any other task that implies some kind of integration of preexisting resources, complex. This process, according to Srivastava and Chen (1999), is "labor-intensive, error-prone, and generally frustrating, leading a number of warehousing projects to be abandoned mid-way through development" (118). OLTP and OLAP environments are profoundly different. Therefore, the techniques used for operational database design are inappropriate for data warehouse design (Kimball & Ross, 2002; Kimball et al., 1998).

Despite the obvious importance of having a methodological support for the development of OLAP systems, the scientific community and product providers have paid very little attention to the design process. Models usually utilized for operational database design (like the Entity/Relationship-E/R model-should not be used without further ado for analytical environments design. Bearing in mind just technical reasons, databases obtained from E/R models are inappropriate for decision support systems, in which query performance and data loading (including incremental loading) are important (Kimball & Ross, 2002). Multidimensional paradigm should be used not only in database queries but also during its design and maintenance. As stated in Dinter et al. (1999): "To use the multidimensional paradigm during all development phases it is necessary to define dedicated conceptual, logical and physical data models for the paradigm and to develop a sound methodology which gives guidelines how to create and transform these models during the development process." Wu & Buchmann (1997) claimed for data warehouse design methodologies and tools "with the appropriate support for aggregation hierarchies" and "mappings between the multidimensional and the relational models,"(79).

The next section summarizes existing approaches in data warehouse design. Then, our approach for the development of data warehouses is briefly described. Finally, conclusions are presented.

SUMMARY OF EXISTING APPROACHES

There are several proposals for data warehouse design; in this section, we summarize the most relevant ones.

In Kimball and Ross (2002) and Kimball et al. (1998), an approach based on two points is proposed: the data warehouse bus architecture that shows how to construct a series of data marts that, finally, will allow for the creation of a corporate data warehouse, and the business dimensional life cycle (BDL) with the purpose of development of data marts based on dimensional star schemas starting from the business requirements. It is an iterative methodology in which, after a project planning and a business requirements definition task, different activities are developed. These activities can be categorized into three groups: technology activities, data design activities, and specification and development of final user applications activities.

Last, there are two activities related to data warehouse deployment, maintenance, and growth. It is a detailed methodology and, according to the authors, is widely tested. However, in our opinion, it is focused on the relational model from its initial phases.

In Debevoise (1999), an object-oriented methodological approach is proposed, using Unified Modeling Language (UML) to detail the methodology steps. Use case diagrams are used to describe the tasks that the team has to carry out to complete each phase. Use cases will specify what every team member has to do to complete each project cycle part. This methodology is less detailed than the previous one and is a bit difficult to follow.

Cabibbo and Torlone (1998) presented a logical model for multidimensional (MD) database design, and a design methodology to obtain a MD schema from operational databases. As the starting point, they use an ER schema that describes an integrated view of the operational databases. This schema may contain all information valuable for the data warehouse, but the information is in an inappropriate format for this kind of system. The methodology consists of a series of steps for the MD model schema construction and its transformation into relational models and multidimensional matrices. The methodology is incomplete and starts from an ideal assumption; that is, all information is contained in the ER schema. In our opinion, operational schemas should be simply a support, giving more importance to analytical users' requirements.

Golfarelli and colleagues (Golfarelli & Rizzi, 1999; Golfarelli, Maio, & Rizzi, 1998) outlined a methodological framework for data warehouse design based on a conceptual multidimensional model of the same authors, called dimensional fact model (DFM). The methodology is mainly focused on a relational implementation.

Abelló et al. (2001, 2002) reviewed multidimensional data models and proposed a new one, as an extension of UML. Luján-Mora et al. (2002) also extended UML for multidimensional modeling and proposed a methodology also based on UML for the development of data warehouses (Trujillo & Luján-Mora, 2003; Trujillo et al., 2001).

There are many other partial proposals, focused on issues such as model transformation, view materialization, index, etc. For example, Sapia et al. (1999) proposed using data mining techniques in data warehouse design phases (for example, using data mining algorithms for discovering implicit information on data, for conflict resolution in schema integration for recovering lost values and incorrect data, etc.).

The problem with all these works is that they propose to use a new different methodology for data warehouse design, so organizations must use at least two different methodologies: one for OLTP environments and one for OLAP environments. We think that it is better to integrate data warehouse design in the existing methodologies, modifying and adding new activities, so that the training and learning curve for data warehouse design is less difficult.

OUR APPROACH

Our approach is based on applying the experience and knowledge obtained in relational database system development in the last decade (Structed-Query Language or SQL, ER modeling, Computer Aided Software Enginerring or CASE tools, methodologies...) to multidimensional database (MDDB) design. We propose a MDDB development methodology analogous to the traditional ones used in the relational database systems development. Instead of defining a new methodology, we adapt METRICA, an existing traditional methodology (de Miguel et al., 1998), to the development of data warehouses.

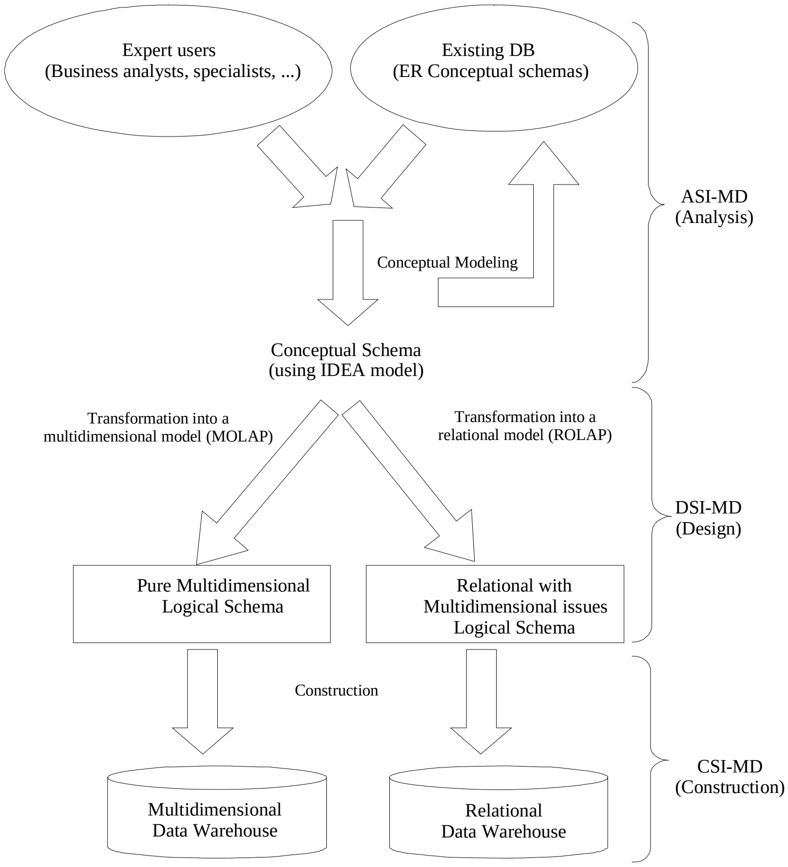
Our methodology (MIDEA) (Cavero et al., 2003) uses as reference framework the Spanish Public Methodology METRICA version 3 proposal (MV3), which is similar to British Structured Systems Analysis Design Method (SSADM) or French Merise. The considered MV3 processes are those that have more influence on the data warehouse development, that is, information system analysis, design, and construction (ASI, DSI, and CSI). The new processes, modified from the original MV3 proposal, have been named as ASI-MD (multidimensional), DSI-MD, and CSI-MD, respectively. Of course, it does not mean that the rest of the processes should not be taken into account on a data warehouse development, but we have considered that the differences should not be significant with respect to any other information system development.

MIDEA uses IDEA, Integrating Data: Elementary-Aggregated, (Sánchez et al., 1999) as a conceptual model. IDEA is a multidimensional conceptual model used to understand and represent analytical users' requirements in a similar manner as the ER model is used to interact with microdata users. Preexisting OLTP system data schema and requirements obtained from analytical data users are the main inputs to the construction of IDEA multidimensional conceptual schema.

This methodology is supported by a CASE tool that incorporates a graphical interface (de Miguel et al., 2000). This tool allows the transformation of a conceptual IDEA schema into a logical schema based on a model supported by some multidimensional or relational products.

Figure 1 presents an overview of the methodology, showing the scope of its three processes: ASI-MD, DSI-MD, and CSI-MD.

Figure 1. Methodology overview



Each process consists of activities, and each activity consists of tasks. The order of the activities does not mean that sequential order is mandatory. The activities can be developed in a different order, or in parallel, overlapping tasks of different activities. However, a process will not be completed until the completion of each of its activities. In every process, a graphic emphasizing its most important activities is included.

One of the activities of the ASI-MD phase is "datawarehouse conceptual modeling." The main purpose of this activity is to obtain, using the IDEA model, the data warehouse multidimensional conceptual schema. It has seven tasks: obtaining preliminary subcell structures, obtaining preliminary dimensions, obtaining preliminary hierarchies, obtaining detailed hierarchies, obtaining detailed subcell structures, obtaining fact schemas, and multidimensional schema verification and validation. For a detailed specification of such tasks, see Cavero et al. (2002).

CONCLUSION

Data warehouse development has turned into a critical success factor for many companies. During this process, some important issues must be taken into account:

1. Integration of OLTP and OLAP database design in a unique methodology
2. Verification and construction of analytical schemas starting from analytical data user requirements, using existing operational schemas as support for the creation of conceptual multidimensional schema

3. Reverse engineering to obtain conceptual schemas from existing specific multidimensional databases
4. Creation or modification of existing operational schemas as a result of analytical users' requirements

ACKNOWLEDGMENTS

This research is carried out in the framework of the following projects: EDAD (07T/0056/2003 1) financed by Autonomous Community of Madrid and DAWIS, financed by the Ministry of Science and Technology of Spain (TIC 2002-04050-C02-01).

REFERENCES

Abelló, A., Samos, J., & Saltor, F. (2001). A framework for the classification and description of multidimensional data models. In *Proceedings of the 12th International Conference on Database and Expert Systems Applications, (DEXA'01),* Springer-Verlag, London, 668-677.

Abelló, A., Samos, J., & Saltor, F. (2002). YAM2 (yet another multidimensional model): An extension of UML. In *International Database Engineering and Applications Symposium, (IDEAS'02),* IEEE Computer Society, Canada, 172-181.

Cabibbo, L., & Torlone, R. (1998). A logical approach to multidimensional databases. *In Lecture Notes in Computer Science. Proceedings of the 6th International Conference on Extending Database Technology: Advances in Database.* Springer-Verlag, London, 187-193.

Cavero, J. M., Costilla, C., Marcos, E., Piattini, M. G., &. Sánchez, A. (2003). A multidimensional data warehouse development methodology. In *Managing data mining technologies in organizations: Techniques and applications*. Hershey, PA: Idea Group Publishing, 188-201.

Cavero, J. M., Marcos, E., Piattini, M. G., &. Sánchez, A. (2002). A methodology for datawarehouse design: Conceptual modeling. In *Data warehousing and Web engineering*. Hershey, PA: IRM Press, 185-197.

Codd, E. F., Codd, S. B., & Salley, C. T. (1993). Providing OLAP (on-line analytical processing) to user–analyst: An IT mandate. *Technical Report*, E. F. Codd and Associates.

de Miguel, A. et al. (1998). METRICA version 3: Planning and development methodology of information systems. Designing a methodology: A practical experience. In *Proceedings of CIICC'98*, Aguascalientes, México. 264-276.

de Miguel, A. et al. (2000). IDEA-DWCASE: Modeling mutidimensional databases. Sotware Demonstration at *EDBT 2000 Software Demonstrations track*. Konstanz, Germany.

Debevoise, N. T. (1999). *The datawarehouse method*. New York: Prentice Hall PTR.

Dinter, B., Sapia, C., Blaschka, M., & Höfling, G. (1999). OLAP market and research: Initiating the cooperation. *Journal of Computer Science and Information Management, 2*(3).

Golfarelli, M., Maio, D., & Rizzi, S. (1998). Conceptual design of datawarehouses from E/R schemas. In *Proc. Hawaii Intern. Conf. on System Sciences,* vol.VII, Kona Hawaii, 334-343.

Golfarelli, M., & Rizzi, S. (1999). Designing the datawarehouse: Key steps and crucial issues. *Journal of Computer Science and Information Management, 2*(3), 1-14.

Inmon, W. H. (1993). *Building the datawarehouse.* New York: John Wiley & Sons.

Kimball, R., Reeves, L., Ross, M., & Thornthwaite, W. (1998). *The datawarehouse lifecycle toolkit: Expert methods for designing, developing, and deploying data warehouses.* New York: John Wiley & Sons.

Kimball, R., & Ross, M. (2002). *The data warehouse toolkit: The complete guide to dimensional modeling* (2nd ed.). New York: John Wiley & Sons.

Luján-Mora, S., Trujillo, J., & Song, I. (2002). Extending UML for multidimensional modeling. In J.-M. Jézéquel, H. Hußmann, & S. Cook (Eds.), *UML 2002 - The Unified Modeling Language, 5th International Conference*, Dresden, Germany, September 30 - October 4, 2002, Proceedings. Lecture Notes in Computer Science 2460, Pages: 290-304. Springer Verlag

Sánchez, A., Cavero, J. M., & de Miguel, A. (1999). IDEA: A conceptual multidimensional data model and some methodological implications. In *Proceedings of the CIICC'99* (pp. 307-318). Cancún, Mexico.

Sapia, C. et al. (1999). On supporting the datawarehouse design by data mining techniques. *GI-Workshop: Data Mining and Data Warehousing*, September 27-28, Magdeburg, Germany.

Srivastava, J., & Chen, P. -Y. (1999). Warehouse creation—A potential roadblock to data warehousing. *IEEE Transactions on Knowledge and Data Engineering, 11*(1), 118-126.

Trujillo, J., & Luján-Mora, S. (2003). A UML based approach for modeling ETL processes in data warehouses. In *Proceedings of the 22nd International Conference on Conceptual Modeling (ER'03)*, 307-320.

Trujillo, J., Palomar, M., Gómez, J., & Song, I. (2001). Designing data warehouses with OO conceptual models. *IEEE Computer*, 34 (12), 66-75.

Wu, M. C., & Buchmann, A. P. (1997). Research issues in data warehousing. In K.R. Dittrich & A. Geppert (Eds.), Datenbanksysteme in Büro, Technik und Wissenschaft (BTW), GI-Fachtagung, Ulm, 5.-7. März 1997, Proceedings. Informatik Aktuell Springer 1997, pp. 61-82.

KEY TERMS

Conceptual Modeling: High-level modeling; modeling at a high level of abstraction.

Data Mart: A data warehouse focused on a particular subject or department. Data warehouses may consist of several data marts.

Data Warehouse: A repository of information coming mainly from online transactional processing systems that provides data for analytical processing and decision support.

Multidimensional Databases: Databases based on multidimensional models.

Multidimensional Modeling: A modeling technique used in data warehouses, multidimensional databases, and OLAP applications, based on defining the facts of interest in terms of their dimension hierarchies.

OLAP (Online Analytical Processing): A category of tools and techniques that allows online analysis of data, usually by means of multidimensional views of data.

OLTP (Online Transaction Processing): In data warehousing context, this refers to the traditional transaction processing of data, as opposed to the analytical processing provided by OLAP.

ENDNOTE

[1] This work is based on Cavero et al. (2002)

Database Integrity

Jorge H. Doorn
INTIA, Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina

INTRODUCTION

One of the factors that stimulated the development of database technology was the need to guarantee the consistency of the stored data. On the other hand, the demand to process more complex data with more semantic content have also lead to a better understanding of the properties of the data and motivated the evolution of the Data Base Management Systems (DBMS) toward the inclusion of facilities to handle such properties. The handling of data properties within the context of a database engine faces two main difficulties: 1) Properties are hard to identify, especially those obvious in the actual world; 2) Data property presentation varies during the process of design of the software.

The first one falls in the field of the Requirement Engineering (Loucopoulos & Karakostas, 1995). The proper requirements elicitation of a software product is a key factor in the success of the whole software development process. However, these requirements are not easy to deal with. They have different intrinsic natures and they may appear showing different faces. In many cases, some requirements are totally or partially hidden in the information collected by the software developers (Jackson, 1995).

The second one falls in the fields of Conceptual modeling and Earlier Logical Design (Doorn & Rivero, 2002). During these activities, issues concerning data modeling and DBMS paradigms promote changes in the way they are considered. Pragmatic issues such as the degree of adhesion of a given DBMS or others related with performance may introduce additional mutations. The understanding of the evolution from actual world data properties to database world integrity restrictions is another a key factor in the success of the software artifact during its life cycle. This understanding must be accompanied with a well defined traceability mechanism, which must clearly identify the actual world data property behind every database integrity restriction (Pinheiro, 2003). Traceability is also a key factor especially during software maintenance activities.

BACKGROUND

Not every data property must be modeled, but each property must be looked at carefully to see if modeling is needed in the context of the scope and in the objective of the software artifact. A more analytical approach may order the data properties by taking into account their importance (Karlsson, 1996).

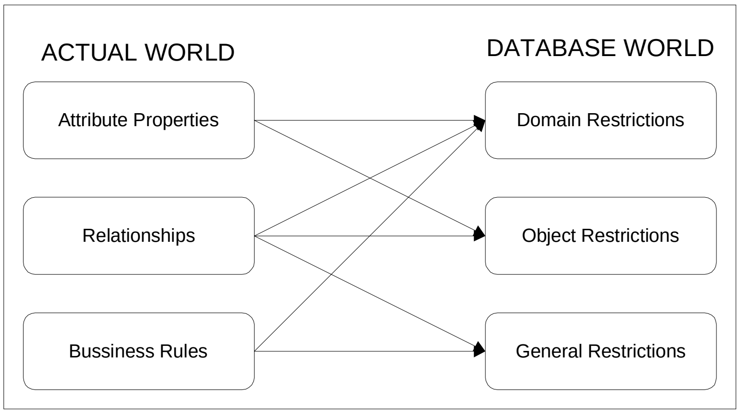
When a data property describes the allowed values for attributes, it is called Attribute Property. Another kind of data property establishes connections among different attributes, known as Relationships. When a data property carries out a semantic that is specific of the Universe of Discourse, not found in other occurrences of the same data, it is called a Business Rule (Ceri et al., 1997; Codd, 1990; Ross, 1997).

Many authors have addressed the restrictions in Databases. In Dey (1999), a profound analysis is done about ternary and higher degree relationships. The authors depict a general framework for the analysis of such constructs and provide generalized rules for their representation in the relational model. Ceri et al. (2000) analyze the order in which triggers, stored procedures, and referential integrity restrictions are executed. Regarding this matter, SQL3 (Cochrane, 1996) has defined a precise execution order and scope of its clauses. A survey of the state of the art of the semantic integrity constraints in some relational and object relational available database systems is done in Rivero et al. (2002).

EVOLUTION OF DATA PROPERTIES

Figure 1 shows all possible mappings from data properties in the actual world to database integrity restrictions. It should be noticed that there are neither arrows from Business Rules to Object Restrictions, nor from Attribute Properties to General Rules. These mappings are extremely hard to find. Also, some other mappings shown in Figure 1 occur very seldom, while others are common. Figure 2 introduces a basic taxonomy of both actual world data properties and database integrity restrictions and gives a more detailed mapping.

Figure 1. Mappings data properties into integrity restrictions



Attribute Properties. Database Engines have a set of built-in data types whose main purpose is to deal with the Attribute Properties of the data to be stored. These data types are useful and have been used for decades, helping users to take care of their data processing needs.

As with all the other Non-Functional Requirements, Attribute Properties are expressed in the Universe of Discourse in declarative ways. An Attribute Property defines the Set of Values that the attribute may have. These sets may be defined by enumeration of the members or by intention. Basically, Attribute Properties defined by enumeration map to Object Restrictions while those defined by intention may map to Domain Restrictions upon the ability of the DMBS to express the restriction (see arrows leaving Attribute Properties in figure 2). When an Attribute Property is defined by enumeration, the list of all possible values for such an attribute should be stored somewhere in the Database. It will become an inclusion dependency when such list is not key of the holding object. Otherwise it will be materialized as a referential integrity constraint.

Relationship Properties. The connection among different attributes is the source of most of the data processing richness and problems. These connections have a scope larger than Attribute Properties since they involve several attributes, at least two, usually belonging to different objects or entities.

When software artifacts are involved, the links among real world things (persons, objects, activities, etc.) are present throughout the whole process of their development. Sooner or later, the links among those things become data Relationships (see arrows leaving Relationships in figure 2). How soon this happens depends upon the software design approach. Some Relationships are binary (cardinalities may be 1:1, 1:N, N:M); however, Relationships may connect three or more real objects. When Relationships must become persistent in the data repository, an obvious issue needs to be analyzed: how are they preserved? Some very old approaches put the linked data together to express the Relationship. These approaches used to have many well-known disadvantages, especially when the Relationship cardinality was not 1:1. This redundancy was also known as the source of a new problem: consistency of data being one of the reasons that pushed towards the creation of the first database models. A Relationship whose cardinality is N:M between entities or objects introduces new problems. Firstly, there may be attributes belonging to the Relationship itself and second, more than one link is needed. The redundancy can be reduced or avoided expressing the links among data in other ways. This implies the inclusion of a special attribute in the data not found in the real world but used to represent the link. This attribute may be either a physical reference telling where the related data is stored, or a logical reference holding a key attribute that permits the related data to be found. In both cases, the redundancy problem is replaced by a referential integrity problem. To sum up, the technique used to store the data in the computer resources will create one or both of the following problems: Data Redundancy or Referential Integrity.

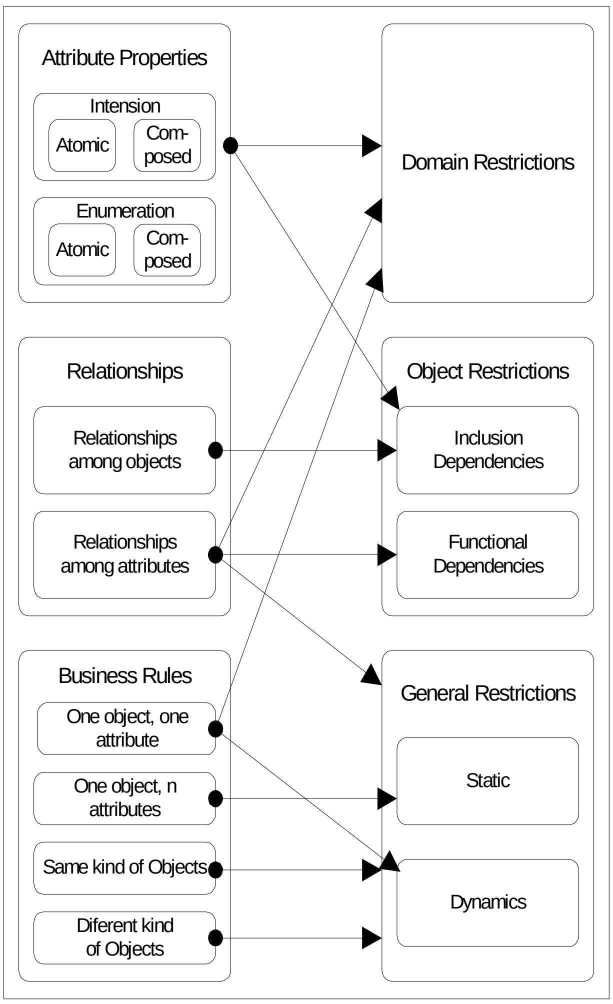
No matter how it is referenced, the referred data should be available every time it is needed. But since the referred data is stored and processed independently from the referring data, the link may become lost. This is called the Referential Integrity problem.

In most cases Relationships maps into Object Restrictions. However, it may happen that Relationships can be reduced to Domain Restrictions when the set of possible values is small. On the other hand, complex Relationships force their implementation by means of General Restriction.

Business Rules. Every organization restrains behavior in some way. This is strongly associated to constraints that define which data may (or may not) be actualized. A Business Rule is a declarative sentence usually describing a correct state of a piece of data.

The process of identifying Business Rules is often iterative and heuristic, where rules begin as general organization statements of policy. Even if the policy is formal and specific, it is typically described in a general and informal fashion, and it often remains for the person responsible to translate it into meaningful specific statements of what to do. These statements are only sometimes originated in a given policy. More often, they arise from the day-to-day operation of the organization. These sentences are sometimes clear, sometimes (perhaps deliberately) ambiguous, and most of the time, contain more than one idea (Ross, 1997).

Figure 2. Detailed mappings



A Business Rule could be confused, due to its appearance, with Attribute Properties and with Relationships. A Business Rule whose scope is only one attribute has in fact the same structure as an Attribute Property sharing all their characteristics. On the other hand, Business Rules involving more than one attribute belonging to the same or different entities or objects may be either similar to or rather different from the Relationships (see arrows leaving Business Rules in figure 2). A General Restriction is dynamic when involve temporal issues. This implies that its evaluation requires some kind of registration of the evolution of the participating data or when the restriction needs the current and the upcoming data.

As mentioned above the need to guarantee the integrity of the stored data was and still is one of the key factors driving the evolution of the Data Base Management Systems (DBMS), creating different Database paradigms. The following paragraphs give a brief summary about how most popular paradigms see the Database integrity problem.

Relational Databases. A rich theory for constraints has emerged in relational databases, mainly based on a fundamental class of constraints called dependencies. Its main motivation is to incorporate more semantics into the relational model.

The term relational model has thus come to refer to the broad class of database models that have relations as the data structure and that incorporate some or all of the query capabilities, update capabilities, and integrity constraints (Abiteboul et al., 1995). Many ideas of other database paradigms already in the scenario (and in their own evolution process) have been borrowed by researchers and vendors, and introduced in current relational products. This fact has created types of incomplete and extended relational engines, which started to be known as postrelational databases. They are incomplete because not all the prescribed issues for the relational model are satisfied, and extended because they go further in some other areas such as active characteristics.

Active Databases. Supporting reactive behavior implies that a database management system may be viewed from a production rule system perspective. These production rules are well known today, in database terminology, as active rules or simply triggers. It is a form of computation, which is motivated by the occurrence of some event, typically a database operation, executing a reaction to that stimulus. Active rules may pose queries to the database to collect information about events and database objects and decide whether events require an action. Then they may execute actions, normally any sequence of database operations (Ceri, 1997). Reactive behavior is seen as an interesting and practical way to check for satisfaction and enforcement of integrity constraints. Nevertheless, integrity constraint maintenance, materialized view maintenance (especially useful in the warehousing area) and implementation of Business Rules, are not the only areas of application of data repositories with reactive behavior. Other interesting applications areas are replication of data for audit purpose, data sampling, workflow processing, scheduling, real-time applications and many others. In fact, practically all products offered today in the marketplace support complex reactive behavior on the client side.

Spatial Databases. Even though there is a very active research area interested in the design of robust and efficient spatial databases, the inability of current GIS regarding the implementation and management of spatial integrity constraints is still evident. This is due to the fact that within the scope of geographic applications, special problems come up due to the locational aspects of data. A modification in a spatial database may cause simultaneous updates in a large number of records in multiple files, making it hard to manage the entire environment. A very sophisticated control is required to avoid redundancy and loss of integrity.

Object-Relational Databases. Object-oriented literature typically uses the term “relationship” to mean, specifically, relationships supported by foreign keys in a relational system. Various levels of support for referential integrity have been implemented in those systems. As in the relational paradigm, triggers are very effective in supporting data integrity in a database, especially for dealing with those restrictions that cannot be expressed declaratively. ORDBMSs demand a system of triggers even more flexible than the relational one.

Object-Oriented Databases. Given that in an OO environment, objects are collected into classes and relationships are established at the class level. A relationship between classes denotes a set of relationships between the objects of the respective classes. Relationships may have attributes. A model is always a compromise to achieve the right amount of expressive power while keeping simplicity and clarity. Domain Restrictions, keys, and referential integrity constraints may all be used straightforwardly in OODBMSs.

Other kinds of constraints are peculiar to OODBMSs, for instance: constraints of the migration of objects between classes (roles); exclusivity constraints between classes; constraints on the definitions of subclasses; and existence dependencies. The last ones are the key to semantic integrity checking since they allow the designers to track and solve inconsistencies in an object-oriented conceptual schema (Snoeck & Dedene, 1998).

Distributed Databases. A distributed database managing system (DDBMS) is a database management system (DBMS) that supports characteristics of a distributed database, that is, the possibility of handling information contained in multiple locations, and preserving data integrity at the same time. A DDB differs from a centralized DB mainly in that data are placed in a number of locations instead of being located in only one site. This characteristic of DDBs causes the control of data integrity to be more complex than for centralized environments. Each transaction can involve more than one location, and it is hard to keep an execution order of the instructions of the transaction to preserve data integrity.

FUTURE TRENDS

The strength of the SQL standards induces one to think that in the future the object relational paradigm will continue to be dominant in the field of DBMS. Improvements to check-and-trigger mechanisms are strongly needed to integrate data properties in a more friendly way with the data itself. Development platforms where data properties become a concern of the engine and the data integrity is fully factorized will form application programs leading to an evolution and the maintenance of less traumatic DBMS-based software artifactss. Naturally, this process will require very active participation on the part of researchers and DBMS producers.

CONCLUSION

Taxonomies of both data properties in the actual world and integrity restriction in the Database world have been given. A mapping of possible evolutions of data properties into the integrity restrictions was also presented. This mapping was always described from the actual world perspective. Its importance can be easily derived from figure 2, where it is shown that merely looking at the integrity restrictions is not enough to know the why’s of their existence. Moreover the actual world to Database world perspective eases traceability, which has been proven to be a key factor in software evolution and maintenance.

REFERENCES

Abiteboul, S., Hull, H., & Vianu V. (1995). *Foundations on databases*. Boston: Addison Wesley.

Ceri, S., Cochrane, R., & Widom, J. (2000). Practical applications of triggers and constraints: Successes and lingering issues. *Proceedings of 26th VLDB Conference*, Egypt, 254-262.

Ceri, S., & Fraternali, P. (1997). *Designing database applications with objects and rules: The IDEA methodology*. New York: Addison Wesley.

Cochrane, R., Pirahesh, H., & Mattos, N. (1996). Integrating triggers and declarative constraints in SQL database systems. *Proceedings of 22th VLDB Conference*, Mumbai India, 123-132.

Codd, E. (1990). *The relational model for database management. Version 2*. New York: Addison Wesley.

Dey, D., Storey, C., & Barron, T. (1999). Improving database design through the analysis of relationships. *ACM TODS*, *24*(4), 453-486.

Doorn, J., & Rivero, L. (Eds.) (2002). *Database integrity: Challenges and solutions.* Hershey, PA: Idea Group.

Jackson, M. (1995). *Software requirements & specifications. A lexicon of practice, principles and prejudices*. Workingham: Addison Wesley Publishers, ACM Press.

Karlsson, J. (1996). Software requirements prioritizing. *Proceedings of ICRE'96 Second International Conference on Requirements Engineering. IEEE CSP*. Los Alamitos, CA. 100-116.

Loucopoulos, P., & Karakostas, V. (1995). *System requirements engineering*. London: McGraw Hill International Series in Software Engineering.

Pinheiro, F. (2004). Requirements traceability. In J. Leite, & J. Doorn (Eds.), *Perspectives on software requirements* (pp. 91-114). Kluwer Academic Publishers.

Rivero, L., Doorn, J., & Ferraggine, V. (2000). Inclusion dependencies. In S. Becker (Ed.), *Developing quality complex database systems* (pp. 261-278). Hershey, PA: Idea Group.

Rivero, L., Doorn, J., & Ferragine V. (2002). Database integrity: Fundamentals and current implementations. In J. Doorn & L. Rivero (Eds.), *Database integrity: Challenges and solutions* (pp. 17-65). Hershey, PA: Idea Group.

Ross, R. (Ed.) (1997). *The business rule book. Classifying, defining and modeling rules*. Boston: Database Research Group.

Snoeck, M., & Dedene, G. (1998). Existence dependency: The key to semantic integrity between structural and behavioral aspects of object. *IEEE Trans. on Software Engineering*, *24* (4), 233-251

KEY TERMS

Conceptual Modeling: A formal or semi-formal description of the actual world elements (objects, persons, organizations) to be included in the future software artifact.

Database Paradigms: The rational and theoretical framework or archetype used to formulate and support a database.

DBMS: Data Base Management Systems. It is a collection of programs that enables users to create and maintain a database.

Logical Design: A formal or semi-formal description produced by the translation of the Conceptual Model into terms of the database structures.

Traceability: The ability to describe the life of a requirement in both backward and forward directions.

Universe of Discourse: The environment in which the software artifact will be used. It includes the macrosystem and any other source of knowledge.

D

Database Support for M-Commerce

Hong Va Leong
The Hong Kong Polytechnic University, Hong Kong

INTRODUCTION

M-commerce applications have evolved out of e-commerce applications, riding on the rapid advancement in mobile communication technologies in the past decade. The diffusion of applications on the Internet into the mobile computing environment has taken an accelerating pace ever since. Virtually all e-commerce and m-commerce applications rely on the provision of information retrieval and processing capability. In this regard, database systems serve as the information source and repository for these applications, backed by efficient indexing mechanism. Bean (2003) gave a good report on supporting Web-based e-commerce with XML, which could be easily extended to m-commerce. An m-commerce framework, based on JINI/XML and a workflow engine, was also defined by Shih and Shim (2002). Customers can receive m-commerce services through the use of mobile devices, such as pocket PCs, PDAs, or even smart phones. These mobile devices together with their users are often modeled as *mobile clients*. Central to supporting m-commerce applications are three types of entities: mobile device, mobile communication, and database. In particular, we are more interested in providing efficient access mechanisms to mobile-client-enabled database servers, which are often called *mobile databases*. Mobile databases contain the core information to support the underlying m-commerce applications, while the use of mobile devices serves for the hardware platform, with mobile communication providing the necessary connection between mobile databases and mobile devices for interfacing with real users or customers.

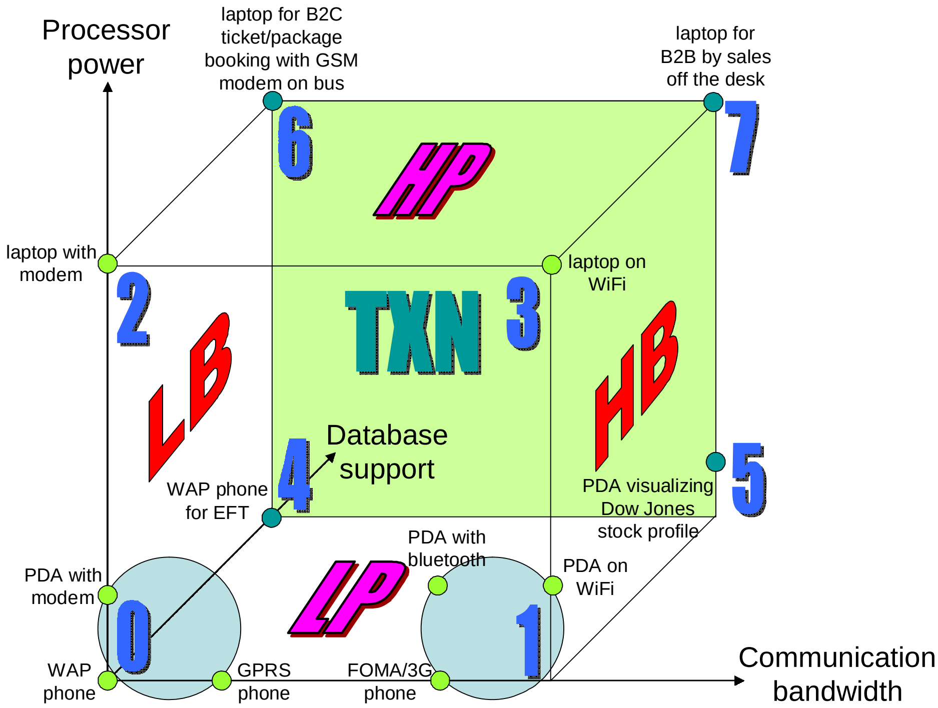
The two major types of data access requirements for a mobile database are data dissemination and dedicated data access. In a mobile environment, data dissemination is preferred, since it can serve a large client population in utilizing the high-bandwidth downlink channel to broadcast information of common interest, such as stock quotations, traffic conditions, special events, or the number of available seats at a performance. On the other hand, dedicated data access is conveyed through uplink channels with limited bandwidth. To disseminate database items effectively, the selected set of hot database items can be scheduled as a broadcast disk (Acharya, Alonso, Franklin & Zdonik, 1995) or the "air storage" (Leong & Si, 1997). The sequence of items to be broadcast is referred to as a broadcast program (Acharya et al., 1995). Proper indexes can be built to facilitate access to database items in the broadcast (Imielinski & Badrinath, 1994). Redundancy can be included to combat the unreliability of wireless communication (Tan & Ooi, 1998).

For dedicated data access, queries and updates to databases are transmitted from the client to the server. Since quite a significant portion of m-commerce services involve location-dependent queries, performance of location-dependent queries should be improved (Madria, Bhargava, Pitoura & Kumar, 2000), dictating effective location maintenance with moving object databases (Wolfson, Sistla, Xu, Zhou & Chamberlain, 1999). This enables efficient execution of common queries such as K-nearest neighbor search. As database systems become more powerful, they can help to process continuous queries (Prabhakar, Xia, Kalashnikov, Aref & Hambrusch, 2002), which are executed continuously, returning possibly location-dependent results.

BACKGROUND

The three fundamental elements in the provision of an m-commerce application, namely, mobile device, mobile communication and database support, can be considered orthogonal. First, the variety of mobile devices differs vastly in computational power, ease of programming, interoperability of operating environments, and support for auxiliary devices. Some mobile clients may be based on a high-end laptop of capacity close to that of a desktop. Other mobile clients may be based on a low-end PDA, or cellular WAP or GPRS phone. Second, mobile communication offers varying bandwidth and reliability, based on low-bandwidth and unreliable GSM connection, medium-bandwidth GPRS or Bluetooth connection, or high-bandwidth 802.11b or 3G/CDMA2000 connection with higher reliability. Third, the database may be as primitive as a file system or simple relational database like MS Access, or as complex as a high-performance Oracle database with transactional support and information retrieval ability. *Transactions* are useful in ensuring a sequence of database operations to be executed consistently without any interference experienced. This leads to a "cube"-like taxonomy as shown in Figure 1.

Figure 1. Taxonomy on m-commerce support



In Figure 1, the taxonomy for m-commerce support is displayed. Planes **LP** and **HP** represent the low computing power equipment and high computing power equipment respectively, whereas planes **LB** and **HB** reflect the availability of low- and high-communication bandwidth. With the availability of transactions in the **TXN** plane, this gives rise to eight different regions, 0 to 7.

Region 0 represents the support of standard file or simple database access from PDA, connecting through low-speed modem or WAP phones. Processing is basically performed at the server, since it is too expensive for clients to support complex mechanism. To combat the low-bandwidth problem, information distillation/extraction (Cowie & Lehnert, 1996) may be performed to reduce the amount of information transmitted. Simple client/server data access paradigm suffices. Region 1 assumes an improved wireless network, with CDMA2000 or WiFi. As a result, data access is more effective, and conventional client/server data processing techniques can be adopted in a rather straightforward manner.

Region 2 corresponds to a mobile client with higher computational power. Information transmitted can be transcoded to reduce the bandwidth consumption. Interactive and intelligent mechanisms such as multi-resolution browsing (Yau, Leong & Si, 2001) can be employed. Data items are cached to combat the low-communication bandwidth, unreliable communication, and frequent disconnection. Research work addressing this issue has drawn much attention in the past, as pioneered by the work on the Coda file system (Mummert, Ebling & Satyanarayanan, 1995). In Coda, frequently accessed files are cached by the clients, and updates made during client disconnection are reintegrated with the file system upon reconnection. Caching in an object-oriented database was studied by Chan, Leong, Si and Wong (1999). Configurations in Region 3 allow easy access to data from the server. With ample bandwidth and processing power, prefetching of data items is done to prepare for potential network disconnection (Jing, Helal & Elmagarmid, 1999). Numerous research works on mobile data access have been conducted with respect to Regions 2 and 3.

Plane **TXN** represents the transactional equivalence of the above-mentioned four regions. Regions 4 and 5 involve the use of PDAs or phones to access databases in a transactional manner. Owing to the low device capability, the only effective mechanism is to execute the transaction at the server. Clients only implement the user interface, supplying the required data and displaying the result sets. Information distillation could be needed for the low-bandwidth configurations in Region 4. The use of a proxy server helps to simplify the client/server design, since the proxy will be responsible for contacting different parties involved in the transaction. Finally, for Regions 6 and 7, there are a lot more research potentials, since the client is more powerful on which more complex algorithms can improve overall system performance. For instance, it would be appropriate to implement on the client a variant of the optimistic concurrency control protocol (Bernstein, Hadzilacos & Goodman, 1987) for Region 6, and two-phase locking with lock caching (Franklin, Carey & Livny, 1997) for Region 7 configurations.

In practical m-commerce applications, clients are moving around and are often far away from their home locations. Research and application focus should therefore be on the low-communication bandwidth, as well as on client mobility. This corresponds to plane **LB**. Under such configurations, a large client population communicates with database servers over wireless communication channels (Alonso & Korth, 1993). Efficient access mechanisms to these mobile databases must be devised after addressing client movement and client location management (Wolfson et al., 1999). This gives rise to a new class of database queries, known as *location-dependent queries* (Madria et al., 2000). A *geographical information system* component enables these queries to be resolved through geo-referencing (Choy, Kwan & Leong, 2000). *Moving object databases* can keep track of the whereabouts of clients (Wolfson et al., 1999).

STRONG DATABASE SUPPORT

M-commerce applications involve access to one or more databases, which are often being accessed concurrently by multiple clients. Transactions are executed to ensure the database consistency despite concurrent access. The correctness criterion of *serializability* on concurrent transactions can be enforced through concurrency control protocols. Concurrency control protocols in a client/server or a mobile environment can be classified according to their nature (Franklin et al., 1997; Jing et al., 1999). In m-commerce applications, simultaneous access to multiple databases, which are managed by different organizations, is a norm rather than an exception. A collection of database systems organized in a coherent way, and functioning in a cohesive and collaborative manner, is referred to as a *federated database system* (Fang, Ghandeharizadeh, McLeod & Si, 1993). Often, it is necessary to provide consistent accesses to those multiple databases with a transaction-like behavior, known as *global serializability* (Breitbart, Garcia-Molina & Silberschatz, 1992). The distributed activity accessing the multiple databases is called a *global transaction*. Tesch and Wäsch (1997) presented an implementation of global transactions on the ODMG-compliant multi-database systems. However, creating a federated database system could involve a lot of coordination efforts, both at the system level and the enterprise managerial level. The execution cost of the global transactions, in terms of concurrency control and atomic commitment, can be high. Thus, global transactions have not been widely adopted, despite their usefulness and convenience. Rather, multiple sub-transactions were commonly executed on individual databases without enforcing global serializability. *Compensating transactions* (Chrysanthis & Ramamritham, 1994) are executed explicitly to undo the effect of failed sub-transactions.

In a mobile environment, it is appropriate to relax serializability, which could be imposing overly restrictive constraints on the execution of transactions. It is often acceptable for a mobile client not to see the updates made by concurrently executing mobile clients, in exchange for a faster execution and a higher probability of committing its transaction. Isolated-only transactions (Lu & Satyanarayanan, 1995) were proposed to reduce the impact of client disconnection. N-ignorance (Krishnakumar & Bernstein, 1994), bounded inconsistency (Wong, Agrawal & Mak, 1997), and update consistency (Shanmugasundaram, Nithrakashyap, Sivasankaran & Ramamritham, 1999) were some of the common weaker forms of correctness criteria. In these approaches, they try to ignore some of the operations in a transaction or allow them to be executed out-of-order in some controlled manner.

Regardless of whether serializability or its weaker form is adopted, transaction processing throughput in a mobile environment can be improved by utilizing the broadcast bandwidth effectively. The database can be broadcast and transactions can be processed against the *broadcast database*. This is very useful for read-only transactions (Pitoura & Chrysanthis, 1999) simply by tuning for a consistent set of database items over the broadcast. To enable update transaction processing, the hybrid protocol by Mok, Leong, and Si (1999) ensures serializability by performing validation for update transactions, and utilizing the uplink channel to request for additional database items not available over the broadcast. Consistency across database items is ensured through the use of timestamps. In update consistency (Shanmugasundaram et al., 1999), a mobile client is only required to see updates made at a server consistent with the values it reads, without having to follow the same serialization order as those observed by other mobile clients; it can be enforced by the cycle-based algorithm.

With the embracement of the Internet computing paradigm, more and more enterprises are willing to publicize their databases as part of their drive towards B2B or B2C e-commerce. Under most cases, these databases can be accessed from outside the enterprise via a Web interface. The ability to access consistent information using global transactions in a federated database system becomes more practical and manageable. Although updates to databases are normally restricted in a non-federated database environment, more and more databases become enabled for the execution of read-only transactions (enquiries) by external parties. The use of global transactions on "loosely" federated databases could become more popular. Furthermore, the presence of a high proportion of read-only transactions even renders the concurrency control for global transactions far more efficient.

Under certain B2B settings, it would be advantageous to automate the workflow process across enterprises (Vonk & Grefen, 2003). It is important to ensure the transactional execution of a low-level workflow process, albeit the more relaxed consistency requirement on higher level processes. Naturally, the overall process can be modeled as a nested transaction, which is destined to be long-lived that global serializability could be too restrictive in delivering good performance. Instead, the adoption of compensating transactions to roll back unintended effects is more appropriate (Tesch & Wäsch, 1997). With respect to the B2C settings, mobile clients would normally only initiate local transactions or simple global transactions spanning across a small number of databases, rather than a long-lived workflow process. The limitation of communication bandwidth and occasional network disconnection implies a longer transaction execution cycle. As a result, existing research works for m-commerce point to supporting global transactions with relaxed correctness criteria, allowing partial transaction commitment and controlled degree of inconsistency, whenever necessary.

FUTURE TRENDS

Owing to the complexity of global transaction processing and the resource limitations of mobile clients, it is sensible to migrate the coordination effort to the proxy server and the Web server, thereby relieving the mobile clients from the complex processing. This is especially true in the context of Regions 4 and 5. As a result, there would be the decoupling of the transaction processing mechanism from the application logic, with basic transactional support at the proxy. The application logic can further be delivered conveniently through the adoption of mobile agents (Yau et al., 2001), preferably intelligent agents that can make sensible decision on behalf of the client, only reporting back the outcome and obtaining confirmation from the client. For instance, mobile agents can act on behalf of the client for event-driven transactions like stock-selling transactions or auctioning. Support for global transactions can be provided primarily on the wired network through the agents. For Region 6 and 7 configurations, more complicated control can be established at the clients, whose higher computing power can also be dedicated to filter for information more effectively, with reference to its local cache, and to provide value-added operations on the data. Now, it is mature to leverage on the computing power of these mobile clients for m-commerce, by engaging in the popular paradigm of peer-to-peer computing (Avancha, D'Souza, Perich, Joshi & Yesha, 2003). Effective visualization of the result set and navigation through the result sequence to decide on the next processing step is also important, since low-end mobile devices normally are only equipped with a relatively small display. This user interface constraint would not change until the penetration of foldable display.

CONCLUSION

Database support is an important and yet fundamental issue in m-commerce applications. With respect to the vast difference in the power and capacity of mobile devices, the varying wireless communication bandwidth, and the new dimension of client mobility and network disconnection, adjustments need to be made to render appropriate database support to these applications. The consistency requirement of accessing multiple databases demands transactional support from database systems. In this article, we gave a generic classification along the three major characteristics for m-commerce environments. The different issues of m-commerce database support were surveyed and discussed. In the future, there should be a division of research efforts, in providing effective transactional support at the server or proxy through agents, while leveraging on the capability of the clients in information organization and presentation, and also in the quest for better value-added computing support and solutions.

REFERENCES

Acharya, S., Alonso, R., Franklin, M. & Zdonik, S. (1995). Broadcast disks: Data management for asymmetric communication environments. *Proceedings of the ACM SIGMOD International Conference on Management of Data* (pp. 199-210). ACM.

Alonso, R. & Korth, H. (1993). Database system issues in nomadic computing. *Proceedings of the ACM SIGMOD International Conference on Management of Data* (pp. 388-392). ACM.

Avancha, S., D'Souza, P., Perich, F., Joshi, A. & Yesha, Y. (2003). P2P m-commerce in pervasive environments. *ACM SIGecom Exchanges*, *3*(4), 1-9.

Bean, J. (2003). *Engineering global e-commerce sites: A guide to data capture, content and transactions*. San Francisco: Morgan Kaufmann.

Bernstein, P.A., Hadzilacos, V. & Goodman, N. (1987). *Concurrency control and recovery in database systems*. Reading, MA: Addison-Wesley.

Breitbart, Y., Garcia-Molina, H. & Silberschatz, A. (1992). Overview of multidatabase transaction management. *VLDB Journal*, *1*(2), 181-239.

Chan, B.Y.L., Leong, H.V., Si, A. & Wong, K.F. (1999). MODEC: A multi-granularity mobile object-oriented database caching mechanism, prototype and performance. *Journal of Distributed and Parallel Databases*, *7*(3), 343-372.

Choy, M., Kwan, M. & Leong, H.V. (2000). Distributed database design for mobile geographical applications. *Journal of Database Management*, *11*(1), 3-15.

Chrysanthis, P.K. & Ramamritham, K. (1994). Synthesis of extended transaction models using ACTA. *ACM Transactions on Database Systems*, *19*(3), 450-491.

Cowie, J. & Lehnert, W. (1996). Information extraction. *Communications of the ACM*, *39*(1), 80-91.

Fang, D., Ghandeharizadeh, S., McLeod, D. & Si, A. (1993). The design, implementation and evaluation of an object-based sharing mechanism for federated database systems. *Proceedings of the International Conference of Data Engineering* (pp. 467-475). IEEE.

Franklin, M.J., Carey, M.J. & Livny, M. (1997). Transactional client-server cache consistency: Alternatives and performance. *ACM Transactions on Database Systems*, *22*(3), 315-363.

Imielinski, T. & Badrinath, B.R. (1994). Mobile wireless computing: Challenges in data management. *Communications of the ACM*, *37*(10), 18-28.

Jing, J., Helal, A.S. & Elmagarmid, A. (1999). Client-server computing in mobile environments. *ACM Computing Surveys*, *31*(2), 117-157.

Krishnakumar, N. & Bernstein, A.J. (1994). Bounded ignorance: A technique for increasing concurrency in a replicated system. *ACM Transactions on Database Systems*, *19*(4), 586-625.

Leong, H.V. & Si, A. (1997). Database caching over the air-storage. *The Computer Journal*, *40*(7), 401-415.

Lu, Q. & Satyanarayanan, M. (1995). Improving data consistency in mobile computing using isolation-only transactions. *Proceedings of the 5th Workshop on Hot Topics in Operating Systems*.

Madria, S.K., Bhargava, B.K., Pitoura, E. & Kumar, V. (2000). Data organization issues for location-dependent queries in mobile computing. *Proceedings of International Conference on Database Systems for Advanced Applications* (pp. 142-156).

Mok, E., Leong, H.V. & Si, A. (1999). Transaction processing in an asymmetric mobile environment. *Proceedings of the 1st International Conference on Mobile Data Access* (pp. 71-81). Springer-Verlag.

Mummert, L.B., Ebling, M. & Satyanarayanan, M. (1995). Exploiting weak connectivity for mobile file access. *Proceedings of the 15th ACM Symposium on Operating System Principles* (pp. 143-155). ACM.

Pitoura, E. & Chrysanthis, P.K. (1999). Scalable processing of read-only transactions in broadcast push. *Proceedings of the 19th International Conference on Distributed Computing Systems* (pp. 432-439). IEEE.

Prabhakar, S., Xia, Y., Kalashnikov, D.V., Aref, W.G. & Hambrusch, S.E. (2002). Query indexing and velocity constrained indexing: Scalable techniques for continuous queries on moving objects. *IEEE Transactions on Computers*, *51*(10), 1124 -1140.

Shanmugasundaram, J., Nithrakashyap, A., Sivasankaran, R. & Ramamritham, K. (1999). Efficient concurrency control for broadcast environments. *Proceedings of the ACM SIGMOD International Conference on Management of Data* (pp. 85-96). ACM.

Shih, G. & Shim, S.S.Y. (2002). A service management framework for m-commerce applications. *Mobile Networks and Applications*, *7*(3), 199-212.

Tan, K.L. & Ooi, B.C. (1998). On selective tuning in unreliable wireless channels. *Data and Knowledge Engineering*, *28*(2), 209-231.

Tesch, T. & Wäsch, J. (1997). Global nested transaction management for ODMG-compliant multi-database systems. *Proceedings of the International Conference on Information and Knowledge Management* (pp. 67-74). ACM.

Vonk, J. & Grefen, P. (2003). Cross-organizational transaction support for e-services in virtual enterprises. *Journal of Distributed and Parallel Databases*, *14*(2), 137-172.

Wolfson, O., Sistla, A.P., Xu, B., Zhou, J. & Chamberlain, S. (1999). DOMINO: Databases fOr MovINg Objects tracking. *Proceedings of the ACM SIGMOD International Conference on Management of Data* (pp. 547-549). ACM.

Wong, M.H., Agrawal, D. & Mak, H.K. (1997). Bounded inconsistency for type-specific concurrency control. *Journal of Distributed and Parallel Databases*, *5*(1), 31-75.

Yau, S.M.T., Leong, H.V. & Si, A. (2001). Multi-resolution Web document browsing in a distributed agent environment. *Proceedings of the International Conference on Mobile Data Management* (pp. 279-281). Springer-Verlag.

KEY TERMS

Broadcast Database: A mobile database whose contents are being broadcast, fully or partially, to a population of mobile clients.

Compensating Transaction: A transaction that is executed to undo the effect of another committed transaction. Unlike ordinary transaction rollback or abort, both original and compensating transactions are visible in the committed projection of the execution history.

Continuous Query: A query that is re-evaluated continuously. For example, the query "*give me the most updated temperature*" will return different readings depending on the current moment. Some continuous queries are also location dependent. For instance, the query "*show me the nearest gas station*" will continually execute a location-dependent query. Advanced query processing technique is needed, in conjunction with moving object databases.

Federated Database System: A network of independent and autonomous databases cooperating in a loosely coupled manner to share and exchange information.

Geographical Information System: An information system that stores and manipulates data for geographical entities such as streets, road junctions, railway, land-use, or even terrain. The data is associated with the location of the entities to allow fast geo-referencing.

Location-Dependent Query: A query whose results depend on the current location of the query issuer. For example, the query "*which is the nearest gas station*?" will return different gas stations depending on the current location of a driver.

Mobile Database: A database accessible to mobile clients. There are appropriate mechanisms to take into account the limitation of the wireless bandwidth, the use of downlink broadcast channel, and the effect of client mobility.

Moving Object Database: A database that maintains efficiently the location information about moving objects, with proper indexing on the object location.

Serializability/Global Serializability: Serializability is the generally accepted correctness criterion for concurrent execution of transactions. The concurrent execution should produce the same effect and lead to the same database state as one possible sequential execution of the same set of transactions. Global serializability is the correctness criterion for concurrent execution of global transactions over many database systems. It is a stronger correctness criterion than serializability.

Transaction/Global Transaction: A transaction is a sequence of operations on a database that should appear as if it were executed non-interfered, even in the presence of other concurrent transactions. A transaction should satisfy the ACID properties, namely, **A**tomicity, **C**onsistency, **I**solation, and **D**urability. A global transaction is a distributed transaction that is executed on two or more database systems.

D

Database Technologies on the Web

J.F. Aldana Montes
Universidad de Málaga, Spain

A.C. Gómez Lora
Universidad de Málaga, Spain

N. Moreno Vergara
Universidad de Málaga, Spain

I. Navas Delgado
Universidad de Málaga, Spain

M.M. Roldán García
Universidad de Málaga, Spain

INTRODUCTION

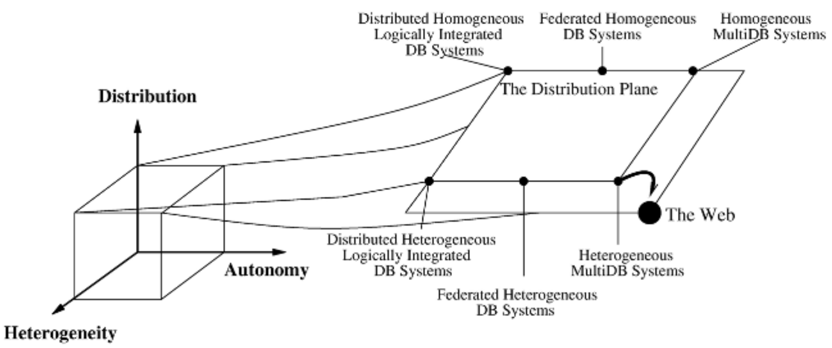
Database community has been seriously disturbed with the Web technologies expansion. Particularly, two reports have produced a special commotion in database field. The first one, the Asilomar report (Bernstein et al., 1998), postulates the new directives in databases tendencies, previewing the Web impact in this field. The second one, Breaking out the Box (Silberschatz & Zdonik, 1996), proposes how database community must transfer its technology to be introduced into Web technology. In this sense, the database box must be broken out into its autonomous functional components, and they must be used to reach a solution for the problem of heterogeneous data sources integration.

BACKGROUND

Thinking about the Web as a huge, highly distributed database, we may consider different dimensions to conceptually describe it. Tamer Özsu and Valduriez (1999) define a classification of database systems with respect to: 1) their distribution; 2) the autonomy of local systems; and 3) the heterogeneity of database systems. The autonomy concept is considered as the distribution of control, not of data. This indicates the degree to which individual DBMSs can operate independently. Whereas autonomy refers to the distribution of control, the distribution dimension deals with the physical distribution of data over multiple sites. With respect to heterogeneity, this can range from hardware heterogeneity, differences in networking protocols, variations in DBMSs, and so forth, to the data model or the policy for managing integrity on the database.

Obviously, the Web is in the distribution plane, and, as shown in Figure 1, we think that "it falls out" of the cube because it presents the highest degree of distribution, heterogeneity, and autonomy. Therefore, traditional distributed database techniques must be further extended to deal with this new environment in order to face Web environment intrinsic problems, such as the management of semi-structured information.

Figure 1. Extending the cube



A lot of work has still to be carried out in the database community to resolve all of the issues related to such a kind of distributed and heterogeneous database, which is what the Web actually is.

Due to space limitations, we only review some related and open issues, such as the integrity problem, query optimisation problem, integration issues in both Web and Semantic Web. But the reader must note that this coverage is not complete. There are some interesting issues uncovered, such as data models, security, transaction processing, and so forth.

THE INTEGRITY PROBLEM IN THE WEB

Some important factors related to this issue include the structure of the underlying database upon which the constraints are imposed, the nature of the imposed constraints, and the method adopted for their evaluation. Traditionally, two restrictions are focused in databases: domain and structural restrictions. So they must be considered in a Web data model.

From a database perspective, XML Schema (2001) provides a new technological standard which enables us to represent data semantics like a database does. We can find a discussion about schema languages in Lee and Chu (2000).

A domain restriction defines the set of values that an attribute may have. XML-Schema provides enhanced data types and user-defined data types. New data types can be created from base data types specifying values for one or more facets for the base data type. Moreover, on XML-Schema we can define subclasses and superclasses of types. We can also disable the restriction or extension of a particular type.

With respect to structural restrictions, in this schema language, we can represent:

1. Uniqueness for attribute. Furthermore, XML Schemas specify uniqueness not only for attributes but also for arbitrary elements or even composite objects.
2. Key for attribute.
3. Foreign key for attribute; using the *key/keyref* mechanism, which complements the *id/idref* one of previous versions of XML Schema and solves its associated problems.

QUERY OPTIMISATION IN THE WEB

Optimisation on regular path queries (Grahne & Thomo, 2000, 2001) and indexing techniques over semi-structured information (McHugh & Widom, 1999) have already been studied. One of the most important studies on XML and semi-structured data optimisation (McHugh & Widom, 1999) has been developed for the LOREL system (Abiteboul, Quass, McHugh, Widom & Wiener, 1997). LOREL not only supports the traditional value index, but also label, edge, and path indexes.

However, other relevant aspects, such as composition reordering, restriction propagations, and more complex and sophisticated techniques, magic rewriting for example (Bancilhon, Maier, Sagiv & Ullman, 1986), are well known in database systems and have yet to be tested in the XML context.

From the XML perspective, constraint propagation cannot be used exclusively for query optimisation, but it can and must be used during new type derivation or inference (Fan, Kuper & Siméon 2001). This is generically called subtyping. One of the main characteristics of XML is that it allows new type extensions and restrictions through inheritance. The computation of this new set of constraints is fundamental in object-oriented systems, and very important in preventing integrity constraints violations in our schema.

Constraint derivation in subtyping is useful with domain and column constraints, but it is also useful with entity and referential constraints (Fan, Kuper, & Siméon, 2001). The existence of physical index associated with keys and references in order to accelerate relationship composition is common.

INTEGRATION ISSUES IN THE WEB

Wrappers were the first building block on Web integration. They act as interfaces to each data source, providing (semi-) structure to non-structured sources or mapping the original data source structure to a common one. The knowledge about evaluating a query over multiple wrappers is encapsulated by mediators. The wrapper-mediator approach provides an interface to a group of (semi-) structured data sources, combining their local schemas in a global one and integrating the information of local sources. So the views of the data that mediators offer are coherent, performing semantic reconciliation of the common data model representations carried out by the wrappers. Table 1 summarizes the most popular wrapper-mediator systems. The next level of abstraction on Web integration corresponds to ontology-based systems. From the data perspective, ontologies enrich the semantic of the schema, resolving synonymy and polysemy problems (Heflin, 2001). The reader can find an excellent review in ontology engineering in (Corcho, Fernández-López & Gómez-Pérez, 2001). Table 2 summarizes the most popular ontology-based systems.

Table 1. A summary of wrapper-mediator systems

Not XML based	• DISCO • GARKIC
XML based	• AMOS • TSIMMIS • MIX • MOCHA

Table 2. A summary of ontology-based systems

For building ontologies	• WebODE • WebOnto
For integration purpose	• Ariadne • OBSERVER

INTEGRATION ISSUES IN THE SEMANTIC WEB

Tim Berners-Lee (2000) outlined his vision for the Semantic Web: "In the context of the Semantic Web, the word semantic means machine processable. For data, the semantics convey what a machine can do with that data". He also underlined that the Semantic Web is a declarative environment, where you say what you mean by some data, and not what you want to do with it. The development of the Semantic Web involves the adoption of different technologies (Figure 2), which allow adding meaning to the structure of the XML documents and describing the necessary semantic information in order to be processed by the machines. For a long time, this problem has been dealt with by means of "ontologies" described as documents or files "that definitely define the relationships between the terms". According to the information sources, the ontologies describe the content of the data repositories regardless of their syntactic representation, enabling their semantic integration (Mena, Illarramendi, Kashyap, & Sheth, 2000).

As a means to materialize a conceptualization or ontology, a language to represent that knowledge is required. Traditionally, these languages have been developed in the area of artificial intelligence and concentrate their formal base on paradigms such as: the 1st and 2nd order predicate calculus, description logic or the object-oriented ones. Nowadays, languages based on XML are ideal candidates for the Semantic Web. Thus, first proposals were SHOE, XOL, OML and RDFS, which inherit the labels of XML and incorporate new characteristics that improve the expressiveness of the initial data model. Latter, other languages that extend RDF and RDFS were proposed, such as OIL, DAML+OIL and OWL. This latter are going to become the standard for representing ontologies, due to the efforts of the W3C Semantic Web Group.

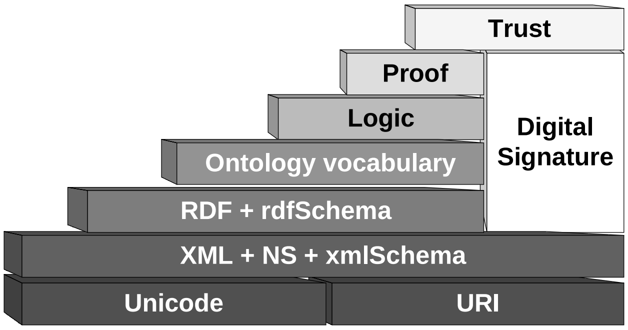
- The most naive setting for the Semantic Web would be the one in which all the data sources commit with a very limited set of well known universally accepted ontologies. It is evident that this scenario is not very realistic because everybody is not expected to commit with a limited group of ontologies. It is easier to imagine that there will be many ontologies and many systems in charge of evaluating specific ontologies. The evaluation should be made in a cooperative way thanks to the interaction of agents or systems, which is essential to respect the interoperability and autonomy of the systems.

CONCLUSION: A DATABASE PERSPECTIVE

Since Codd formally defined the relational model in the early 1970s, it has proved its expressiveness and efficiency, but also has presented limitations. This has motivated the definition of many extensions. Among these, two have shown an unusually high degree of success. Deductive and object-oriented database paradigms are attempts to introduce recursion and object-orientation in databases. XML is a standard based on a semi-structured model that allows structural recursion. Its algebra has functional language characteristics that support both static and dynamic type inference. This means that XML includes and extends the problems of distribution, recursion, and object-orientation in relational models.

It has been shown that there exist several ways to specify integrity constraints in XML, using DTDs or XML Schema, among others. To avoid multiple fetching of constraints expressed in different formats during data management, it would be desirable to choose a unique

Figure 2. Semantic Web (figure by Tim Berners-Lee, http://www.w3c.org/2000/Talks/1206-xml2k-tbl/slide10-0.html)



format of constraints specification. The XML Schema seems to be the best candidate due to its expressiveness, as is shown in recent studies (Lee & Chu, 2000). Nevertheless, other standards, like RDF and RDF Schemas, are complementary and can be used together in a higher abstraction level, as proposed in the Semantic Web. Almost all of the aspects related to maintenance and query optimisation via integrity constraints are open in XML.

Many frontiers are open to research and development. Moreover, we still cannot ensure that the W3C XML query language and algebra recommendations, in its current status, would be valid as a practical query language for data intensive processing. Alternative proposals exist, see the comparative analysis of Bonifati and Ceri (2000), although many of them conform to the W3C's one. A good example is the proposal (Beech, Malhotra & Rys, 1999) developed by three important W3C members: Oracle, IBM, and Microsoft. Together, they have developed a query language and algebra very close to SQL and relational algebra whose results are compatible with XQuery's although these are especially oriented to data intensive processing.

An update semantic model is still undefined revealing the amount of work yet to do. For a complete definition of the data manipulation language, it will be necessary to define new recommendations including the given update commands. Having finished the process of complete formalization of the language, the development of a transactional system for XML would be necessary. It would maintain data integrity under multiple concurrent accesses. This work is mainly related to the logical data schema. Improvements in the physical model will begin later on.

As a result, we may note how many aspects of XML query processing, including integrity constraints management, would have to receive serious attention from the database community. New fields of research are open and in-depth research on all aspects related to this new data model in the Web are of vital interest regarding its application to industry.

The Semantic Web will continue to be a model of growth similar to the Web and its success depends on us being able to elaborate realistic techniques that make this development model feasible. The trade-off adopted among the concepts of expressive power, correctness, completeness and efficiency, for each one of the different mechanisms of inference, will open a wide range of study regarding new evaluation techniques – based on ontologies – for distributed logical programmes within the context of the Semantic Web. On the other hand, the query and efficient management of huge distributed knowledge bases still have many unsolved aspects, related – among other things – with the efficient integration of information and the development of distributed mechanisms of inference.

REFERENCES

Abiteboul, S., Quass, D., McHugh, J., Widom, J., & Wiener, J. (1997). The Lorel query language for semistructured data. *International Journal on Digital Libraries*, *1*(1), 68-88.

Bancilhon, F., Maier, D., Sagiv, Y., & Ullman, J.D. (1986). Magic sets and other strange ways to implement logic programs. *Proceedings of the ACM SIGMOD-SIACT Symposium on Principles of Database Systems* (pp.1-16).

Beech, D., Malhotra, A., & Rys, M. (1999). A formal data model and algebra for XML. Communication to the W3C. Available online at *http://www-db.stanford.edu/dbseminar/Archive/FallY99/malhotra-slides/malhotra.pdf*

Berners-Lee, T. (2000). Semantic Web - XML2000. At XML2000. Available online at: *http://www.w3.org/2000/Talks/1206-xml2k-tbl/Overview.html*

Bernstein, P.A., Brodie, M.L., Ceri, S., DeWitt, D.J., Franklin, M.J., García-Molina, H., Gray, J., Held, G., Hellerstein, J.M., Jagadish, H.V., Lesk, M., Maier, D., Naughton, J.F., Pirahesh, H., Stonebraker, M., & Ullman, J.D. (1998). The Asilomar report on database research. *SIGMOD Record, 27*(4), 74-80.

Bonifati, A., & Ceri, S. (2000). Comparative analysis of five XML query languages. *SIGMOD Record*, *29*(1), 68-79.

Corcho, O., Fernández-López, M., & Gómez-Pérez, A. (2001). Technical Roadmad v1.0. IST Proyect IST-2001-29243 OntoWeb.

Fan, W., Kuper, G.M., & Siméon, J. (2001). A unified constraint model for XML. *The 10th International World Wide Web Conference (WWW'10).*

Grahne, G., & Thomo, A. (2000). An optimization technique for answering regular path queries. *International Workshop on the Web and Databases (WebDB 2000 Informal Proceedings)* (pp.99-104).

Grahne, G., & Thomo, A. (2001). Algebraic rewritings for optimising regular path queries. *International Conference on Database Theory (ICDT 2001)* (pp.301-315).

Lee, D., & Chu, W.W. (2000). Comparative analysis of six XML schema languages. *SIGMOD Record, 29*(3), 76-87.

Heflin, J.D. (2001). Towards the semantic Web: Knowledge representation in a dynamic, distributed environment, PhD Thesis.

McHugh, J., & Widom, J. (1999). Query optimization for XML. *Proceedings of the 25th Very Large Data Bases Conference (VLDB 1999)* (pp.315-326).

Mena, E., Illarramendi, A., Kashyap, V., & Sheth, A.P. (2000). OBSERVER: An approach for query processing in global information systems based on interoperation across pre-existing ontologies. *International of Journal Distributed and Parallel Databases (DAPD), 8*(2), 223-271. ISSN 0926-8782.

Silberschatz, A., & Zdonik, S. B. (1996). Strategic directions in database systems - Breaking out of the box. *Computing Surveys, 28*(4), 764-778.

Tamer Özsu, M., & Valduriez, P. (1999). *Principles of distributed database systems* (2nd ed.).

XML Schema W3C Recommendation 2 May 2001. (2001). Available online at: *http://www.w3.org/XML/Schema*

KEY TERMS

Data Integration: The problem of combining data from multiple heterogeneous data sources and providing a unified view of these sources to the user. Such unified view is structured according to a global schema. Issues addressed by a data integration system include specifying the mapping between the global schema and the sources and processing queries expressed on the global schema.

Domain Restriction: A condition, usually formulated in first order logic, that defines the set of values that an attribute or variable may have. In database terminology, a domain restriction is a kind of integrity constraint.

Integrity Constraint: Properties that the data of a database are required to satisfy and they are expected to be satisfied after each transaction performed on the database. Integrity constraints provide a way of ensuring that changes made to the database do not result in a loss of data consistency. Integrity constraints include key references and cardinality restrictions.

Mediator: Systems that filter information from one or more data sources that are usually accessed using wrappers. The main goal of these systems is to allow users to make complex queries over heterogeneous sources, as if it was a single one, using an integration schema. Mediators offer user interfaces for querying the system, based on the integration schema. They transform user queries into a set of sub-queries that other software components (the wrappers), which encapsulate data sources' capabilities, will solve.

Ontology: A logical theory accounting for the intended meaning of a formal vocabulary , i.e., its ontological commitment to a particular conceptualization of the world. The intended models of a logical language using such a vocabulary are constrained by its ontological commitment. An ontology indirectly reflects this commitment (and the underlying conceptualization) by approximating these intended models.

Query Optimization: A problem that can be stated as follows: Given a query, the set of all execution plans that compute it and a cost function that is defined on the before set, find a execution plan that is of minimum cost using some search strategy.

Semantic Query Optimization: A query optimization process based on knowledge. The knowledge used is usually represented as, but is not limited to, integrity constraints.

Semantic Web: An extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation. Berners-Lee said that in the context of the semantic Web, the word "semantic" meant "machine processable". He explicitly ruled out the sense of natural language semantics. For data, the semantics convey what a machine can do with that data.

Structural Restriction: Conditions that relate one attribute or variable value with respect to other possible values of the same attribute or variable or to the value of others' attributes. In database terminology, a structural restriction is a kind of integrity constraint. Some examples of structural restrictions are: uniqueness of an attribute value, key for attribute and foreign key for attribute.

Decision Support Systems Concept

Daniel J. Power
University of Northern Iowa and DSSResources.com, USA

INTRODUCTION

Since the late 1960s, researchers have been developing and implementing computerized systems to support management decision makers. A number of decision support systems (DSS) typologies were proposed in the early 1980s (cf., Alter, 1980; Sprague & Carlson, 1982), but technology developments and new applications led to an expanded DSS framework (cf., Power, 2000a, 2000b, 2001). The expanded DSS framework developed in detail in Power (2002a) helps decision makers and DSS developers explain and categorize potential decision support projects as well as existing DSS.

Many terms are used for specific types of DSS. For example, some vendors and managers use the terms "business intelligence," "collaborative systems," "computationally oriented DSS," "data warehousing," "model-based DSS," and "online analytical processing (OLAP)" software to label decision support software. Software vendors use these more specialized terms for both descriptive and marketing purposes. The terms used to describe decision support capabilities are important in making sense about what technologies have been deployed or are needed. Some DSS are subsystems of other information systems, and this common structural design adds to the complexity of categorizing and identifying DSS. In general, DSS are a broad class of information systems used to assist people in decision-making activities (cf., Power, 2004).

According to Alter (1980), DSS can "take on many different forms and can be used in many different ways" (p. 71). DSS differ in terms of capabilities and targeted users of a specific system, and in terms of how the DSS is implemented and what it is called. Some DSS focus on data, some on models, and some on facilitating communications and collaboration. DSS also differ in scope. Some DSS are intended for one "primary" user and "stand alone" for analysis, and other DSS are intended for many users in an organization.

BACKGROUND

Traditionally, academics and practitioners have discussed building DSS in terms of four major components: the user interface, the database, the models and analytical tools, and the DSS architecture and network (cf., Sprague & Carlson, 1982). This traditional list of components identifies similarities and differences between categories or types of DSS. The expanded DSS framework is primarily based on the differential emphasis placed on the DSS components when systems are actually constructed. The importance of the components in providing decision support functionality is the major differentiating factor among various types of DSS.

The expanded DSS framework focuses on one major dimension, with five categories and three secondary dimensions. The major characteristic in the framework is the dominant technology component that drives or provides the decision support functionality. Five generic categories based on the dominant component are discussed in this section: communications-driven, data-driven, document-driven, knowledge-driven, and model-driven DSS. The following categories (with explanations based on Power, 2001, 2002a, 2004) can classify DSS currently in use:

- Communications-driven DSS include systems built using communication, collaboration, and decision support technologies. These systems were developed first in the late 1980s and are called groupware. Group DSS also involve communications, but many GDSS derive their functionality by providing a group of users with access to quantitative decision models.
- Data-driven DSS include file drawer and management reporting systems, data warehousing and analysis systems, executive information systems (EIS), and some spatial DSS (SDSS). Business intelligence systems are also examples of data-driven DSS. Data-driven DSS emphasize access to and manipulation of large databases of structured data and, especially, a time-series of internal company data and sometimes external data. Simple file systems accessed by query and retrieval tools provide the most elementary level of functionality. Data warehouse systems that allow the manipulation of data by computerized tools tailored to a specific task and setting or by more general tools and operators provide additional functionality. Data-driven DSS with OLAP, drill-down, and scorecards provide the highest level of functionality and decision support

that are linked to the analysis of a large collection of historical data.

- Document-driven DSS integrate a variety of storage and processing technologies to provide complete document retrieval, summarization, and analysis. A search tool that creates text summaries and rates document relevance provides decision support functionality, but the dominant component is the document base. Examples of documents that might be included in a document database include policies and procedures, product specifications, catalogs, and corporate historical information, including minutes of meetings, corporate records, and important correspondence.
- Knowledge-driven DSS can suggest or recommend actions to managers. These DSS contain specialized problem-solving expertise based upon artificial intelligence and statistics technologies. The "expertise" consists of knowledge about a particular domain, understanding of problems within that domain, and "skill" at solving some of these problems.
- Model-driven DSS include systems that use accounting and financial models, representational simulation models, and optimization models. Model-driven DSS emphasize access to and manipulation of a model. A simple algebraic model with "what if?" analysis provides the most elementary level of model-driven DSS functionality.

Table 1 summarizes the five types of DSS and the expanded DSS framework. Data-driven, document-driven, and knowledge-driven DSS need different and specialized database components to provide decision support. A data-driven DSS needs a relational or a multidimensional database of structured data. A document-driven DSS need a specialized document repository and, in some instances, a relational database to assist in document searching. A knowledge-driven DSS needs to store knowledge, including rules, frames, or likelihood data. A model-driven DSS derives functionality from the model component. Finally, the communications and networking component is the key driver for a communications-driven DSS.

FUTURE TRENDS

The number of DSS of each generic type is expanding rapidly. In general, each generic type of DSS can be targeted to the same user group. Also, a given decision process may benefit from implementation of multiple DSS. Each DSS can have a narrow, specific purpose or a more general purpose. DSS of any type can serve multiple, overlapping purposes. For example, to provide business intelligence to managers, more than one DSS may be needed, including both a data-driven DSS and a document-driven DSS. Finally, each category of DSS can be deployed using a Web-based architecture. Today, Web technologies (Linux, Apache server, MySQL, PHP) provide a powerful DSS development environment. DSS can, should, and will be categorized in terms of these three secondary dimensions—user groups, purpose, and the enabling technology.

One can and should use all four dimensions in Table 1 to categorize a specific DSS. Some specific questions for identifying the type of DSS include the following:

Table 1. Summary of D. Power's expanded DSS framework

Dominant Component	User Group	Purpose	Enabling Technology
Communications **Communications-Driven DSS**	Intra- and interorganization users	Conduct a meeting, post on a Bulletin Board	Web-based or LAN
Database **Data-Driven DSS**	Managers, staff, intra- and interorganization	Query a data warehouse, ad hoc analysis	Mainframe, LAN, Web-based
Document-base **Document-Driven DSS**	Specialists, managers	Search Web pages, find documents	Web-based
Knowledge-base **Knowledge-Driven DSS**	Internal users, customers	Management advice, choose products	LAN or Web-based
Models **Model-Driven DSS**	Managers and staff, customers	Crew scheduling, decision analysis	Stand-alone PC, Web-based

Source: Based on Power (2002a)

1. What is the dominant component and driver of decision support?
2. Who makes up the targeted user group?
3. What is the purpose of the DSS?
4. What is the enabling technology used for implementing the system?

The answers to these questions should help classify the proposed DSS or the DSS product a vendor is trying to sell or an existing system that was previously implemented in a company. For example, a manager may want to build a model-driven, interorganizational, product design, Web-based DSS. Or, a company might currently have a data-driven, intraorganizational, ad hoc query, client/server-based DSS. In the future, a checklist of questions will help DSS analysts and managers categorize and describe DSS.

CONCLUSION

This article explained DSS and summarized a broad conceptual framework that should be used for categorizing DSS. Discussion and debate and efforts to use the framework can improve our understanding of DSS and make the framework more useful for both research and development of DSS. One hopes that the expanded framework improves our overall understanding of computerized systems intended to support decision making. The DSS concept has been reinterpreted and broadened over the past 25 years, and technologies have changed and improved. There remain, however, conceptual overlap problems related to terms associated with computerized decision support, and vendors still use too many marketing terms for decision support development software. In general, the basic DSS goal of supporting decision makers using computing technology remains crucial and important for any organization that is in a complex, rapidly changing environment.

REFERENCES

Alter, S. L. (1980). *Decision support systems: Current practice and continuing challenge*. Reading, MA: Addison-Wesley.

Power, D. J. (2000a). *Decision support systems hyperbook*. Cedar Falls, IA: DSSResources.COM, HTML version.

Power, D. J. (2000b). Web-based and model-driven decision support systems: Concepts and issues. In *Proceedings of the 2000 Americas Conference on Information Systems*, Long Beach, California, August 10-13.

Power, D. J. (2001). Supporting decision-makers: An expanded framework. In A. Harriger (Ed.), *e-Proceedings 2001 Informing Science Conference*, June 19-22, Krakow, Poland.

Power, D. J. (2002a). *Decision support systems: Concepts and resources for managers*. Westport, CT: Greenwood/Quorum Books.

Power, D. J. (2002b). Categorizing decision support systems: A multidimensional approach. In M. Mora, G. Forgionne, & J. N. D. Gupta (Eds.), *Decision making support systems: Achievements and challenges for the new decade* (pp. 20-27). Hershey, PA: Idea Group Publishing.

Power, D. J. (2004). Specifying an expanded framework for classifying and describing decision support systems. *Communications of the Association for Information Systems, 13*(Article 13), 158-166.

Power, D. J., & S. Kaparthi. (1998). The changing technological context of decision support systems. In D. Berkeley, G. Widmeyer, P. Brezillion, & V. Rajkovic (Eds.), *Context-sensitive decision support systems* (pp. 42-54). London: Chapman & Hall.

Sprague, R. H., & Carlson, E. D. (1982). *Building effective decision support systems*. Englewood Cliffs, NJ: Prentice Hall.

KEY TERMS[1]

Business Intelligence (BI): A popularized umbrella term introduced by Howard Dresner of the Gartner Group in 1989 to describe a set of concepts and methods to improve business decision-making by using fact-based support systems. The decision support purpose is to provide managers with information or business intelligence. The term is sometimes used interchangeably with briefing books and executive information systems. A business intelligence system is a data-driven DSS.

Data: Binary (digital) representations of atomic facts, especially from financial transactions. Data may also be text, graphics, bit-mapped images, sound, or analog or digital live-video segments. Structured data are the raw material for analysis using a data-driven DSS. The data are supplied by data producers and are used by information consumers to create information.

Decision Support Systems (DSS): A specific class of computerized information systems that support decision-making activities. DSS are interactive computer-based systems and subsystems intended to help decision makers use communications technologies, data, documents, knowledge, and models to identify and solve problems and make decisions. Five more specific DSS types include communications-driven DSS, data-driven DSS, document-driven DSS, knowledge-driven DSS, and model-driven DSS.

Executive Information Systems (EIS): A computerized system intended to provide current and appropriate information to support decision making for senior managers using a networked workstation. The emphasis is on graphical displays, and there is an easy-to-use interface.

Online Analytical Processing (OLAP): Software for manipulating multidimensional data from a variety of sources that were stored in a data warehouse. The software can create various views and representations of the data. OLAP software provides fast, consistent, interactive access to shared, multidimensional data.

Web-Based DSS: A computerized system that delivers decision support information or decision support tools to a manager or business analyst using a "thin-client" Web browser like Internet Explorer. The computer server that is hosting the DSS application is linked to the user's computer by a network with the transmission control protocol/Internet protocol (TCP/IP). Web-based DSS can be communications-driven, data-driven, document-driven, knowledge-driven, or model-driven.

What If? Analysis: The capability of "asking" the software package what the effect will be of changing input data, decision variables, or independent variables. In a model-driven DSS, a decision variable is a changing factor in the model that is determined by a decision maker. The presence of this capability helps identify a model-driven DSS.

ENDNOTE

1 These Key Term descriptions are based on terms at DSSResources.COM (www.dssresources.com).

Decision Support Systems in Small Businesses

Yanqing Duan
University of Luton, UK

Mark Xu
University of Portsmouth, UK

INTRODUCTION

Decision support systems (DSSs) are widely used in many organisations (Arslan et al., 2004; Belecheanu et al., 2003; Dey, 2001; Gopalakrishnan et al., 2004; Lau et al., 2001; Puente et al., 2002). However, there is a common tendency to apply experience and techniques gained from large organisations directly to small businesses, without recognising the different decision support needs of the small business. This article aims to address the issues related to the development and the implementation of DSSs in small business firms. Our arguments are based on evidence drawn from a large body of DSS literature and an empirical study conducted by the authors in the UK manufacturing sector.

BACKGROUND

Early DSS were developed in parallel with management information system (MIS) in the 1970s. MIS is developed to primarily generate management information from operational systems, whilst DSS as defined by Gorry and Scott Morton (1971) is information systems that focus on supporting people in the unstructured and semi-structured decision-making process. A typical DSS consists of four main components: the database, the model base, the user interface and the users. Central to the DSS are the models and analytical tools that assist managers in decision making and problem solving. Concomitant with advances in the technology of computing, most DSS provide easy access to data and flexible control models with a friendly user interface design; some DSS also incorporate a variety of analytical tools and report/graphic generators. The main purpose of DSS is not to replace managers' ability to make decisions, but to improve the effectiveness of managers' decision making.

DSS in practice can hardly be separated from other types of computer-based systems, as it is often integrated with those systems, for example operational databases, spreadsheets, report generators, and executive support systems. Thus the boundary of DSS has now been extended, and DSS broadly refers to any computer-based information system that affects or potentially affects how managers make decisions. This includes data and model oriented systems, reporting systems, executive support systems, expert systems and group decision support systems.

The success and continued growth of small and medium sized enterprises (SMEs) are critically important to local and national prosperity, but their problems are not always accorded the same importance as those of larger organisations. Compared to the research devoted to large organisations on the use of information systems, SMEs have attracted much less attention. It is also the case that the problems inherent in providing support for small business management are more commonly studied from a social or economic viewpoint. Very few studies indeed have addressed decision support needs in the context of the use of information technology.

Managers of small businesses have often been disappointed with software packages because of the inability of these to adapt well to their needs (Heikkila et al., 1991). There are dangers in seeing small businesses as miniature versions of large businesses; many problems differ, and even similar problems require different solutions. Small enterprises normally have limited resources and less skilled managerial staff. They have higher failure risks and commonly do not have suitable access to the information they need.

DSS IMPLEMENTATIONS

Small business may represent a productive domain for attempts to introduce greater levels of computer-based decision support. Ray (1994) suggests that small business managers and their staff have positive attitudes towards the use of computers in business. Cragg and King (1993) report that many companies have plans to increase their use of computer applications, and found that the wish for better information was the motivating force in all case studies conducted. In the majority of the firms studied by Khan and Khan (1992), managers believed that a computerised system improved their performance in selected areas, but that there is still room for significant further development.

Gordon and Key (1987) point out that if small business managers' problem-solving skills are deficient in any of the critical areas of management decision-making, then they must improve those skills through the use of appropriate educational programmes, consultants, decision support tools, or some combination of these. Unfortunately, the owner-manager (because of involvement in the day-to-day operation of the firm) has not the time, resource or expertise needed to evolve an appropriately analytical approach (Raymond et al., 1989, cited in Naylor & Williams, 1994). There would seem to be as strong a case for the potential benefits of DSS to the smaller business as for its larger counterpart, provided suitable software is available, and it is effectively used by the managers concerned.

Limited research has investigated the success factors for the use of information technology (including DSS) in small businesses (Delone, 1988; Lai, 1994; Raymond & Bergeron, 1992) and the design and development of specific DSSs for SMEs (Chaudhry et al., 1996; Houben et al., 1999). Some work has been done specifically to identify those areas that have not been adapted to DSS, but show potential for its introduction for the small business (Duan et al., 2002). Most research (Levy, 1999) indicates that computer use is still confined to operational activities, although a few studies (Naylor & Williams, 1994) found that some SMEs have realised the value of their information systems as decision support tools and had begun to use them for more complex activities. Other researchers suggest that there are many areas in which DSS can be better developed and utilised to help managers in critical decision-making processes, such as marketing, sales promotion, cash-flow management and customer services. It has been argued that small businesses can improve their organisational performance and increase their competitiveness with appropriate information systems (Levy et al., 1999). The increasing emphasis on competitiveness in small business has led to a new focus on the competitive advantage promised by appropriate use of information technology (Levy et al., 1999; Lin et al., 1993).

A study conducted within the UK manufacturing SMEs by Duan et al. (2002) shows that the extent of DSS use is generally limited and the use of DSS varies considerably among the firms surveyed. However, even where there was a reported low level of DSS use, managers' satisfaction was relatively high. The applications with which managers were most satisfied were: cash management, budget preparation and materials requirements planning. Despite the relatively low usage of DSS generally, the majority of SME managers indicated that they use computers personally to aid business decisions; this suggests that there is, at least, widespread use of desktop computing in managers' offices.

Regarding the inhibitors to the greater use of DSS, lack of staff time to analyse needs and identify solutions is the most significant factor identified. Lack of finance for systems purchase or development, lack of experience of systems development, lack of information on available DSS packages, and unavailability of appropriate software were other factors commonly cited (Duan et al., 2002).

DSS DEVELOPMENT METHODS

DSS for small businesses can be developed and implemented in different ways. Four routes were identified, such as:

- Off-the-peg - purchase of a commercially developed package;
- Bespoke - designed by a software house for the specific application;
- In-house - developed by the firm's own specialist staff;
- User - developed by managers as users.

Research (Duan et al., 2002) shows that the majority of DSS were purchased as commercially developed packages; other systems were developed by managers as users, developed by in-house specialists or developed as bespoke systems by software houses. In view of the normally limited resource base for IT development (Heikkila et al., 1991), it is not surprising that most small firms choose to purchase commercially developed, ready-to-use DSS software. By breaking down the development methods into three decision-making levels, it shows that commercial packages are more commonly used at the operational level (60%) than at the strategic level. In contrast, user-developed DSS are more commonly used at the strategic level than at the operational level.

Research on *in-house* and *user* development methods in small firms is scarce. The evidence from the Duan et al. (2002) survey suggests that small business managers are capable of developing their own DSS, and that a certain proportion do so. Research in Canada by Raymond and Bergeron (1992) found that user-developed DSS in small businesses are more successful than any developed by other means. A study by Lai (1994) in the USA, however, revealed no link between the method of system development and DSS success.

By far the most commonly used DSS in small manufacturing firms are commercial packages purchased off the shelf for operational decision making. The readiness of small business managers to purchase commercial packages, coupled with their recognition that DSS vendors provide valuable support, suggest that small businesses represent a market for software developers that will continue to grow. That many managers are developing their own systems to support strategic decisions might also suggest there to be a market opportunity here.

THE FUTURE NEEDS FOR DSS IN SMES

The research by Duan et al. (2002) attempts to identify the gaps between the current provision of DSS and small business managers' desired levels of DSS support. The findings reveal that:

- the current level of DSS usage is low;
- although DSS usage is limited, managers are generally satisfied with DSS they are using;
- the desired level of support is much higher than the current provision;
- the high standard deviations for current DSS use and desired levels of support indicate high variations among responses. The standard deviation of levels of satisfaction is lower than the other two variables; this suggests that there is less disagreement on this issue.

The study supports the argument that current DSS in small businesses are geared to operational rather than strategic decision making. It is evident that the low-level use of DSS found by Raymond in 1982 has not changed significantly. The desired level of support at the operational level is also much higher than that at the strategic level. Users appear to expect that DSS will provide the most benefit for operational decisions. This is perhaps as well, given the nature of the decision-making tasks at strategic level, involving complex and changing environments, high level of uncertainty and the need to include decision makers' personal intuition and judgement. The lower-level use of DSS and desired support for strategic decision making does not mean that there is no space for further improvement, however. Indeed, the fact that many managers are "going it alone" could mean that professional support will enhance strategic planning. Levy et al. (1998) report that one of their case study firms had been successful in integrating information systems into its business strategy and gained competitive advantages. However, computer support for strategic decisions is still a challenging area for future research and much effort is being expended to overcome the difficulties (Duan & Burrell, 1997; Li et al., 2000), yet again, in the context of the larger business.

CONCLUSION

Decision support systems are designed to assist managers to make more effective decisions. Potentially, they could provide great benefits to SME managers and enhance managers' decision-making capability. However, the extent of DSS implementation in small business is still low in general, although there is significant variation between different firms. The literature review of previous studies indicates that the situation has not changed significantly since Raymond's investigation in 1982, and the present study confirms this. Lack of staff time to analyse needs and identify solutions is the most significant factor holding firms back from adopting, or making further use of DSS. Use of DSS at the operational decision-making level is higher than at the strategic level. Small business managers are generally satisfied with the DSS they are using and are hoping for much better DSS support in the future. DSS development, particularly DSS for strategic decisions in small business, still represents both a challenge and an opportunity for DSS professionals and researchers.

DSS in SMEs are most commonly implemented by purchase of a commercial package, and only rarely by bespoke development. Most DSS are used for operational rather than strategic decision making. Those firms that do use DSS to support strategic decisions rely upon user-developed models.

Although DSS applications in SMEs are still relatively few in number, most DSS users report satisfaction with their systems. To reduce the gaps between current DSS provision and the managers' indicated needs, a greater focus on small business by DSS researchers and practitioners is required. Systems most likely to appeal to small business managers will have to be appropriate to their sector's needs, and capable of implementation with minimal user training. In conclusion, it can be said that the current situation in relation to DSS in small business is full of potential but requiring further professional support.

REFERENCES

Arslan, M., Catay, B., & Budak, E. (2004). A decision support system for machine tool selection. *Journal of Manufacturing Technology Management, 15*(1), 101-109.

Belecheanu, R., Pawar, K.S., Barson, R.J., Bredehorst, B., & Weber, F. (2003). The application of case based reasoning to decision support in new product development. *Integrated Manufacturing Systems, 14*(1), 36-45.

Chaudhry, S.S., Salchenberger, L., & Beheshtian, M. (1996). A small business inventory DSS: Design, development, and implementation issue. *Computers & Operations Research, 23*(1), 63-72.

Cragg, P.B., & King, M. (1993). Small-firm computing: Motivators and inhibitors. *MIS Quarterly, 17*(2), 47-59.

Delone, W.H. (1988). Determinants of success for computer usage in small business. *MIS Quarterly, 12*(1), 51-61.

Dey, P.K. (2001). Decision support system for risk management: A case study. *Management Decision, 39*(8), 634-649.

Duan, Y., & Burrell, P. (1997). Some issues in developing expert marketing systems. *Journal of Business and Industrial Marketing, 12*(2), 149-162.

Duan, Y., Kinman, R., & Xu, M. (2002). The use of decision support systems in SMEs. In S.S. Burgess (Ed.), *Managing information technology in small businesses: Challenges and solutions* (pp. 140-155). Hershey, PA: Idea Group Publishing.

Gopalakrishnan, B., Yoshii, T., & Dappili, S.M. (2004). Decision support system for machining centre selection. *Journal of Manufacturing Technology Management, 15*(2), 144-154.

Gordon, W.L., & Key, J.R. (1987). Artificial intelligence in support of small business information needs. *Journal of Systems Management, 38*(1), 24-28.

Gorry, G., & Scott Morton, M. (1971). A framework for management information systems. *Sloan Management Review, 13*(1), 55-70.

Heikkila, J., Saarinen, T., & Saaksjarvi, M. (1991). Success of software packages in small business: An exploratory study. *European Journal of Information Systems, 1*(3), 159-169.

Houben, G., Lenie, K., & Vanhoof, K. (1999). A knowledge-based SWOT analysis system as an instrument for strategic planning in small and medium sized enterprises. *Decision Support Systems, 26*(2), 125-135.

Khan, E.H., & Khan, G.M. (1992). Microcomputers and small business in Bahrain. *Industrial Management & Data Systems, 92*(6), 24-28.

Lai, V.S. (1994). A survey of rural small business computer use: Success factors and decision support. *Information & Management, 26*(6), 297-304.

Lau, H.C.W., Lee, W.B., & Lau, P.K.H. (2001). Development of an intelligent decision support system for benchmarking assessment of business partners. *Benchmarking: An International Journal, 8*(5), 376-395.

Levy, M., Powell, P., & Galliers, R. (1999). Assessing information systems strategy development frameworks in SMEs. *Information & Management, 36*(5), 247-261.

Levy, M., Powell, P., & Yetton, P. (1998, December). SMEs and the gains from IS: From cost reduction to value added. *Proceedings of IFIP WG8.2 Working Conference, Information Systems: Current Issues and Future Changes*, Helsinki (pp. 377-392).

Li, S., Kinamn, R., Duan, Y., & Edwards, J. (2000). Computer-based support for marketing strategy development. *European Journal of Marketing, 34*(5/6), 551-575.

Lin, B., Vassar, J.A., & Clark, L.S. (1993). Information technology strategies for small businesses. *Journal of Applied Business Research, 9*(2), 25-29.

Naylor, J.B., & Williams, J. (1994). The successful use of IT in SMEs on Merseyside. *European Journal of Information Systems, 3*(1), 48-56.

Puente, J., Pino, R., Priore, P., & de la Fuente, D. (2002). A decision support system for applying failure mode and effects analysis. *International Journal of Quality & Reliability Management, 19*(2), 137-150.

Ray, C.M. (1994). Small business attitudes toward computers. *Journal of End User Computing, 6*(1), 16-25.

Raymond, L. (1982). Information systems in small business: Are they used in managerial decisions? *American Journal of Small Business, 5*(4), 20-26.

Raymond, L., & Bergeron, F. (1992). Personal DSS success in small enterprises. *Information & Management, 22*(5), 301-308.

KEY TERMS

Database: A collection of related information. The information held in the database is stored in an organised way so that specific items can be selected and retrieved quickly.

Decision Support System (DSS): An interactive computer-based system, which helps decision makers utilise data and models to solve semi-structured to unstructured problems.

Executive Information System (EIS): A computer-based information delivery and communication system designed to support the needs of senior managers and executives.

Expert Systems: A computer-based system that performs functions similar to those normally performed by a human expert. It has a knowledge base, an inference engine and a user interface.

Group Decision Support Systems (GDSS): Information systems that support the work of groups (communication, decision making) generally working on unstructured or semi-structured problems.

Management Information System (MIS): A business information system designed to provide past, present, and future information appropriate for planning, organising and controlling the operations of an organisation.

Small and Medium Sized Enterprises (SMEs): The definition of SMEs varies in different countries. It is normally defined as having between 10 and 249 employees in the UK and Europe.

D

Decision-Making Support Systems

Guisseppi Forgionne
University of Maryland Baltimore County, USA

Manuel Mora
University of Aguascalientes, Mexico

Jatinder N. D. Gupta
University of Alabama in Huntsville, USA

Ovsei Gelman
National Autonomous University of Mexico, Mexico

INTRODUCTION

Decision-making support systems (DMSS) are specialized computer-based information systems designed to support some, several or all phases of the decision-making process (Forgionne et al., 2000). They have the stand-alone or integrated capabilities of decision support systems (DSS), executive information systems (EIS) and expert systems/knowledge based systems (ES/KBS). Individual EIS, DSS, and ES/KBS, or pair-integrated combinations of these systems, have yielded substantial benefits for decision makers in real applications.

The evolution of the systems has presented unique challenges and opportunities for DMSS professionals. To gain further insights about DMSS achievements, challenges, and opportunities, we asked recognized leaders in the field for their views. This article overviews the DMSS concepts, presents the expert views regarding achievements, challenges, and opportunities, and examines the implications for DMSS research and practice.

DMSS ARCHITECTURES

Decision-making support systems involve various creative, behavioral, and analytic foundations that draw on a variety of disciplines (Sage, 1981). These foundations give rise to various architectures that deliver the fundamental support concepts to individual and group users. The architectures, which are summarized in Table 1, include: (a) the classic systems (Alter, 1996), which include decision support systems (DSS), expert and knowledge based systems (ES/KBS), executive information systems (EIS), group support systems (GSS), and spatial decision support systems (SDSS) and (b) new systems (Forgionne, 1991; Forgionne et al., 2002; Gray & Watson, 1996; Mora et al., 2003; Power, 2002; Turban & Aronson, 1998), which include management support systems (MSS), decision technology systems (DTS), Integrated DMSS, data warehouse (DW)-based and data mining (DM)-based DMSS (DW&DM-DMSS), intelligent DMSS (i-DMSS), and Web-based DMSS or knowledge management DMSS.

The architectures have been applied to a variety of public and private enterprise problems and opportunities, including the planning of large-scale housing demand (Forgionne, 1997), strategic planning (Savolainen & Shuhua, 1995), urban transportation policy formulation (Rinaldi & Bain, 2002), health care management (Friedman & Pliskin, 2002), decision making in the pharmaceutical industry (Gibson, 2002), banking management (Hope & Wild, 2002), entertainment industry management (Watson & Volovino, 2002) and the military arena (Findler, 2002). The reported, and other applications, draw on advanced information technologies (IT), including intelligent agents (Chi & Turban, 1995), knowledge-based (Grove, 2002) and knowledge-management procedures (Alavi, 1997), synthetic characters (Pistolesi, 2002), and spatial decision support systems (Silva et al., 2002) among others.

DMSS ACHIEVEMENTS AND FUTURE TRENDS

Once created, DMSS must be evaluated and managed. A variety of approaches have been suggested to measure DMSS effectiveness. There are economic theory-based methodologies, quantitative and qualitative process and outcome measures, and the dashboard approach. These approaches suggest various organizational structures and practices for managing the design, development, and implementation effort. Most of these approaches suggest much more user involvement than had heretofore been

Table 1. Decision-making support systems architectures

Classic DMSS Architectures	Description	Main Decision-Making Phase Supported					DMSS' SUPPORT CHARACTERISTICS
		INTELLIGENCE	DESIGN	CHOICE	IMPLEMENTATION	LEARNING	
DSS	A DSS is an interactive computer-based system composed of a user-dialog system, a model processor and a data management system, which helps decision makers utilize data and quantitative models to solve semi-structured problems.			**A**			(A) What-if, goal-seeking & sensitivity analysis.
ES & KBS	An ES/KBS is a computer-based system composed of a user-dialog system, an inference engine, one or several intelligent modules, a knowledge base and a work memory, which emulates the problem-solving capabilities of a human expert in a specific domain of knowledge.	**A**		**B**			(A&B) Symbolic pattern-based recognition; fuzzy data; how and why explanation facilities.
EIS	An EIS is a computer-based system composed of a user-dialog system, a graph system, a multidimensional database query system and an external communication system, which enables decision makers to access a common core of data covering key internal and external business variables by a variety of dimensions (such as time and business unit).	**A**			**B**		(A&B) Key performance indicators (KPIs) in graphs and text tables; data exploring and searching through drill-down, roll-up, slice and dice and pivoting operations; networking communications to internal and external bulletin boards.
GSS	A GSS an integrated computer-based system composed of a communication sub-system and model-driven DMSS (DSS), to support problem formulation and potential solution of unstructured decision problems in a group meeting.		**A**	**B**			(A) Idea generation through brainstorming facilities; pooling and display of ideas; generation of alternatives and criteria. (B) Preference models; voting schemes; conflict negotiation support.
SDSS	A SDSS a computer-based system composed of a user-dialog sub-system, a geographic/spatial database sub-system, a decision model sub-system and a set of analytical tools, which enables decision makers to examine situations based strongly on spatial data.	**A**		**B**			(A) Spatial data searching support; visualization tools for maps, satellite images and digital terrains. (B) What-if analysis of scenarios, goal-seeking analysis, sensitivity analysis of decision variables upon spatial data.

Table 1. Decision-making support system architectures (continued)

Modern DMSS Architectures	Description	Main Decision-Making Phase Supported: INTELLIGENCE	DESIGN	CHOICE	IMPLEMENTATION	LEARNING	DMSS' SUPPORT CHARACTERISTICS
MSS, DTS or I-DMSS	These systems are the result of the triple-based integration (i.e., DSS, EIS and ES/KBS) and have the aim to offer a full support to decision makers in all phases of the DMP.	**A**	**B**	**C**	**D**		(A&D) Visual data exploring through graphs; color codes and tables; data exploration with drill-drown, roll-up, slice and dice, pivoting operations. (B) Intelligent advice through AI-based capabilities to support the models selection task. (C) Numerical modeling through available numerical-based models; what-if, goal-seeking and sensitivity analysis.
DW & DM DMSS	DW&DM-DMSS are computer-based systems composed of a user-dialog sub-system, a multidimensional database subsystem and an online analytical processing (OLAP) component enhanced with knowledge discovery algorithms to identify associations, clusters and classifications rules intrinsic in the data warehouse.	**A**					(A) OLAP capabilities of aggregation, slice and dice; drill-down; pivoting; trend analysis; multidimensional query; graphics and tabular data support. Knowledge discovery patterns using statistical-based, tree-decision or neural networks.
Web-DMSS & KM-DMSS	Web-DMSS & KM-DMSS are computer-based systems composed of an user-dialog sub-system, a text &multimedia document storage subsystem and publishing/retrieval subsystem to preserve and distribute knowledge in the organization using intranets.	**A**				**B**	(A&B) Document publishing and retrieval facilities.
i-DMSS	Computer-based systems composed of a user-dialog sub-system, a multidimensional database and knowledge base subsystem and a quantitative & qualitative processing sub-system, all enhanced with AI-based techniques, designed to support all phases of the DMP.	**A**	**B**	**C**	**D**	**E**	(A&D) Visual data exploring through graphs; color codes and tables; data exploration with drill-drown, roll-up, slice and dice, pivoting operations. (B) Intelligent advice through AI-based capabilities to support the models selection task. (C) Numerical and qualitative modeling through numerical-based or symbolic models; what-if, goal-seeking and sensitivity analysis. (E) Symbolic reasoning through knowledge-based models for explanations about how and why the solution was reached.

practiced, and they also suggest a larger role for specialists outside of traditional information systems practitioners to carry out the technical design, development, and implementation tasks.

The suggested changes and developments present unique challenges and opportunities for DMSS professionals. To gain further insights about DMSS achievements, challenges, and opportunities, a recent study compiled opinions from recognized leaders in the field (Forgionne, Gupta & Mora, 2002). As detailed elsewhere (idem, 2002), the professionals were selected on the basis of their demonstrated accomplishments and reputation within the DMSS community. All experts are long-time professors, consultants, and practitioners with substantial influential publications, leadership positions, and project successes in the DMSS field. The expert verbatim views are summarized in Table 2.

DMSS Achievements

The expert opinion implicitly suggests that DMSS have been recognized as a unique type of information system. Further, the opinions seem to focus on the deployment of new and advanced information technology (IT) to improve the quality of the overall DMSS design. This collective opinion suggests that advanced IT progress can be used to improve DMSS design, development, and implementation. The expert views on core DMSS concepts indicate that distributed DMSS, through Web-based, mobile computing and accessibility concepts, will be a critical factor in maintaining DMSS achievements.

DMSS Challenges

The experts also offered opinions about the next generation of DMSS. In their collective opinion, the next generation of DMSS will involve: (a) the use of portals, (b) the incorporation of previously unused forms of artificial intelligence through agents, (c) better integration of data warehousing and data mining tools within DMSS architectures, (d) creation of knowledge and model warehouses, (e) the integration of creativity within DMSS architectures, (f) the use of integrated DMSS as a virtual team of experts, (g) exploitation of the World Wide Web, (h) the exploitation of mobile IT, and (i) the incorporation of advanced IT to improve the user interface through video, audio, complex graphics, and other approaches. A recent survey on artificial intelligence and DMSS literature (Mora et al., 2003) suggests similar research issues: (a) increasing the processing capabilities of DMSS through intelligent agents, fuzzy systems and neural networks, (b) improving the user-interface capabilities through multimedia and virtual environments, and (c) deploying the potential of distributed knowledge-bases using the Web. In short, the experts and literature on AI and DMSS implicitly recognize the relevance of improving the DMSS user interface, information and knowledge representations schemes and intelligent processing capabilities through the deployment of advanced IT.

Table 2. DMSS achievements, challenges, and opportunities

DMSS Issue	Expert Collective Opinion
Key Achievements	the evolution of DMSS software and hardware; the implementation of DMSS in a variety of organizations; the creation of DMSS tailored design and development strategies
Research Issues and Practical Problems	providing quality data for decision support; managing and creating large decision support databases; model management and model reuse; building knowledge driven DMSS; improving communication technologies; developing a uniform and comprehensive DMSS scheme; developing an effective toolkit; developing and evaluating a synergistic integrated DMSS; collecting insights about the neurobiology of decision support for managers' less structured work; the application of agent and object-oriented methodologies; developing DMSS though well-established methodologies
Core DMSS Architectural Concepts and Opportunities	Web technology; accessibility; security; effective data, idea, and knowledge management, possibly through the use of smart agents; effective model management; effective dialog management; EIS-like features; incorporation of basic and common DMSS functionalities; mobile computing; user-centric design

DMSS Opportunities and Future Trends

Future opportunities, trends and challenges discerned by the experts include: (a) availability of DMSS packages for specific organizational functions, such as customer relationship management, (b) system functional and technical integration, (c) reducing software tool costs, (d) consolidation and innovation, (e) the creation of a technology role for the decision maker through the DMSS application, (f) the integration of the decision maker into the design and development process for DMSS, (g) developing effective design and development tools for user-controlled development, (h) accommodating the structural changes in the organization and job duties created by DMSS use, (i) developing measures of DMSS effectiveness, (j) recognizing the group dimension of organizational decision making, (k) incorporating the cognitive dimension of human decision making, (l) utilization of smart agents and features, (m) distribution of DMSS expertise through collaborative technologies, (n) incorporating rich data, information and knowledge representation modes into DMSS, and (o) focusing user attention on decisions rather than technical issues. Common themes suggested by this disparate expert opinion are: (a) the DMSS should focus decision makers on the decision process rather than technical issues, and (b) DMSS development may require specialized and new IT professionals. However, DMSS success, and in particular large-scale DMSS projects, will be realized in organizations through a systematic and well-managed implementation approach.

CONCLUSION

In some ways, the DMSS field has not progressed very much from its early days. There is still significant disagreement about definitions, methodologies, and focus. Experts vary on the breadth and depth of the definitions. Some favor analytical methodologies, while others promote qualitative approaches. Some experts focus on the technology, while others concentrate on managerial and organizational issues. There does not seem to be a unified theory of decision making, decision support for the process or DMSS evaluation. Furthermore, the achieving of successful implementation of large-scale DMSS is still a complex and open research problem (Mora et al., 2002).

In spite of the diversity, opinions are consistent regarding some key DMSS elements. Most experts recognize the need for problem pertinent data, the role of the Internet in providing some of the necessary data, the need for system integration within DMSS architectures and between DMSS and other information systems, and the importance of artificial intelligence within DMSS processing. The DMSS concept also continues to be successfully applied across a variety of public and private organizations and entities. These applications continue to involve the user more directly in the design, development, and implementation process.

The trends will create DMSS that are technologically more integrated, offer broader and deeper support for decision making, and provide a much wider array of applications. In the process, new roles for artificial intelligence will emerge within DMSS architectures, new forms of decision technology and methodology will emerge, and new roles will be found for existing technologies and methodologies.

As the evolution continues, many tasks that had been assigned to human experts can be delegated to virtual expertise within the DMSS. With such consultation readily available through the system, the decision maker can devote more effort to the creative aspects of management. Support for these tasks can also be found within DMSS. In the process, the decision maker can become an artist, scientist, and technologist of decision making. The DMSS-delivered virtual expertise can reduce the need for large support staff and corresponding organizational structures. The organization can become flatter and more project-oriented. In this setting, the decision maker can participate more directly in DMSS design, development, implementation, and management. Such changes will not occur without displacements of old technologies and job activities, radical changes in physical organizations, and considerable costs. As the reported applications indicate, however, the resulting benefits are likely to far outweigh the costs.

REFERENCES

Alavi, M. (1997). KPMG Peat Marwick U.S.: One giant brain. Creating a system to manage knowledge. *Harvard Business Review* (reprinted collection), 75-95.

Alter, S. (1996). *Information systems: A management perspective*. Menlo Park, CA: Benjamin/Cummings.

Chi, R., & Turban, E. (1995). Distributed intelligent executive information systems. *Decision Support Systems, 14*(2), 117-130.

Findler, N. (2002). Innovative features in a distributed decision support system based on intelligent agent technology. In M. Mora, G. Forgionne & J. Gupta (Eds.), *Decision making support systems: Achievements, challenges and trends* (pp. 174-192). Hershey, PA: Idea Group Publishing.

Forgionne, G. (1991). Decision technology systems: A vehicle to consolidate decision making support. *Information Processing and Management, 27*(6), 679-797.

Forgionne, G. (1997). HADTS: A decision technology system to support army housing management. *European Journal of Operational Research, 97*(2), 363-379.

Forgionne, G., Gupta, J., & Mora, M. (2002). Decision making support systems: Achievements, challenges and opportunities. In M. Mora, G. Forgionne & J. Gupta (Eds.), *Decision making support systems: Achievements, challenges and trends* (pp. 392-402). Hershey, PA: Idea Group Publishing.

Forgionne, G., Mora, M., Cervantes, F., & Gelman, O. (2002, July 3-8). I-DMSS: A conceptual architecture for next generation of DMSS in the Internet age. In F. Adam, P. Brezillon, P. Humpreys & J. Pomerol (Eds.), *Proceedings of the International Conference on Decision Making and Decision Support in the Internet Age (DSIAge02)*, Cork, Ireland (pp. 154-165).

Forgionne, G., Mora, M., Cervantes, F., & Kohli, R. (2000, August 10-13). Development of integrated decision making support systems: A practical approach. In M. Chung (Ed.), *Proceedings of the AMCIS 2000 Conference*, Long Beach, CA (pp. 2132-2134).

Friedman, N., & Pliskin, N. (2002). Demonstrating value-added utilization of existing databases for organizational decision-support. *Information Resources Management Journal, 15*(4), 1-15.

Gibson, R. (2002). Knowledge management support for decision making in the pharmaceutical industry. In M. Mora, G. Forgionne & J. Gupta (Eds.), *Decision making support systems: Achievements, challenges and trends* (pp. 143-156). Hershey, PA: Idea Group Publishing.

Gray, P., & Watson, H. (1996, August 16-18). The new DSS: Data warehouses, OLAP, MDD and KDD. *Proceedings of the AMCIS Conference 1996*, Phoenix, AZ.

Grove, R. (2000). Internet-based expert systems. *Expert Systems, 17*(3), 129-135.

Hope, B., & Wild, R. (2002). Procedural cuing using expert support system. In M. Mora, G. Forgionne & J. Gupta (Eds.), *Decision making support systems: Achievements, challenges and trends* (pp. 101-119). Hershey, PA: Idea Group Publishing.

Mora, M., Cervantes, F., Gelman, O., Forgionne, G., Mejia, M., & Weitzenfeld, A. (2002). DMSS implementation research: A conceptual analysis of the contributions and limitations of the factor-based and stage-based streams. In M. Mora, G. Forgionne & J. Gupta (Eds.), *Decision making support systems: Achievements, challenges and trends* (pp. 331-356). Hershey, PA: Idea Group Publishing.

Mora, M., Forgionne, G., Gupta, J., Cervantes, F., & Gelman, O. (2003, September 4-7). A framework to assess intelligent decision-making support systems. In V. Palade, R. Howlett & L. Jain (Eds.), *Proceedings of the 7th KES2003 Conference*, Oxford, UK (pp. 59-65). Heiderberg, FRG: Springer-Verlag.

Pistolesi, G. (2002). How synthetic characters can help decision making. In M. Mora, G. Forgionne & J. Gupta (Eds.), *Decision making support systems: Achievements, challenges and trends* (pp. 239-256). Hershey, PA: Idea Group Publishing.

Power, D. (2002). Categorizing decision support systems: A multidimensional approach. In M. Mora, G. Forgionne & J. Gupta (Eds.), *Decision making support systems: Achievements, challenges and trends* (pp. 20-27). Hershey, PA: Idea Group Publishing.

Rinaldi, F., & Bain, D. (2002). Using decision support systems to help policy makers cope with urban transport problems. In M. Mora, G. Forgionne & J. Gupta (Eds.), *Decision making support systems: Achievements, challenges and trends* (pp. 86-100). Hershey, PA: Idea Group Publishing.

Sage, A. (1981). Behavioral and organizational considerations in the design of information systems and process for planning and decision support. *IEEE Transactions on Systems, Man and Cybernetics, 11*(9), 640-678.

Savolainen, V., & Shuhua, L. (1995). Strategic decision making and intelligent executive support system. *Proceedings of the 12th International Conference on Systems Science*, Wroclaw, Poland (pp. 285-295).

Silva, F., Eglese, R., & Pidd, M. (2002). Evacuation planning and spatial decision making: Designing effective spatial decision support systems through integration of technologies. In M. Mora, G. Forgionne & J. Gupta (Eds.), *Decision making support systems: Achievements, challenges and trends* (pp. 358-373). Hershey, PA: Idea Group Publishing.

Turban, E., & Aronson, J. (1998). *Decision support systems and intelligent systems.* Upper Saddle River: Prentice-Hall.

Watson, H., & Volonino, L. (2002). Customer relationship management at Harrah's Entertainment. In M. Mora, G. Forgionne & J. Gupta (Eds.), *Decision making support systems: Achievements, challenges and trends* (pp. 157-172). Hershey, PA: Idea Group Publishing.

D

KEY TERMS

Data Warehousing–Data Mining (DW-DM) DMSS: Computer-based system composed of a user-dialog sub-system, a multidimensional database subsystem and an online analytical processing (OLAP) component enhanced with knowledge discovery algorithms to identify associations, clusters and classifications rules intrinsic in a data warehouse.

Decision Making Support System (DMSS): An information system designed to support some, several or all phases of the decision making process.

Decision Support System (DSS): An interactive computer-based system composed of a user-dialog system, a model processor and a data management system, which helps decision makers utilize data and quantitative models to solve semi-structured problems.

Executive Information System (EIS): A computer-based system composed of a user-dialog system, a graph system, a multidimensional database query system and an external communication system, which enables decision makers to access a common core of data covering key internal and external business variables by a variety of dimensions (such as time and business unit).

Expert System/Knowledge Based System (ES/KBS): A computer-based system composed of a user-dialog system, an inference engine, one or several intelligent modules, a knowledge base and a work memory, which emulates the problem-solving capabilities of a human expert in a specific domain of knowledge.

Group Support System (GSS): An integrated computer-based system composed of a communication sub-system and model-driven DMSS (DSS), to support problem formulation and potential solution of unstructured decision problems in a group meeting.

Intelligent Decision Making Support Systems (i-DMSS): Computer-based system composed of a user-dialog sub-system, a multidimensional database and knowledge base subsystem, and a quantitative and qualitative processing sub-system enhanced with AI-based techniques, designed to support all phases of the decision making process.

Management Support Systems (MSS), Decision Technology Systems (DTS), or Integrated Decision Making Support Systems (I-DMSS): Systems that integrate DSS, EIS and ES/KBS to offer full support to the decision maker in all phases of the decision-making process.

Spatial Decision Support System (SDSS): A computer-based system composed of a user-dialog sub-system, a geographic/spatial database sub-system, a decision model sub-system and a set of analytical tools, which enables decision makers to analyze situations involving spatial (geographic) data.

Web-DMSS & Knowledge Management (KM)-DMSS: Computer-based system composed of a user-dialog sub-system, a text and multimedia document storage sub-system, and publishing/retrieval subsystem to preserve and distribute knowledge in intranet-supported organizations.

Decision-Making Support Systems and Representation Levels

Jean-Charles Pomerol
Université Pierre et Marie Curie, France

Frédéric Adam
University College Cork, Ireland

INTRODUCTION

The concept of Decision Support System (DSS), which was first coined by Gorry and Scott Morton (1971), was proposed in an attempt to focus the attention of IS researchers and practitioners more closely on the decision-making processes of managers. It sought to acknowledge the importance of decision-making as the key activity that managers must perform in organizations (Huber, 1982).

Even though there are good functional definitions of what DSS should do, a readily understandable definition that takes into account the specificities of human reasoning (e.g., studies by Herbert Simon; Pomerol & Adam, 2004) is still lacking. In this chapter, we try to bridge the gap between human reasoning and the understanding and design of DSSs. We begin with a description of the human process of decision-making, then we give a semi-formal definition of Decision Making Support Systems (DMSS) and conclude with a few words about the architecture of such systems.

BACKGROUND

Human decision-making has its origin in a dissatisfaction commonly referred to as a decision problem. The dissatisfaction arises from the difference between the current state of affairs and another, not yet existing, more desirable state of affair. The notion of state of affair or *state of the world* refers to the seminal work of Savage (1954). As described by Simon, managerial decision-making is characterized by a number of key factors: (1) the personal dimension, which is at the core of the decision-making process, in that what one person wants may not be desirable for another; (2) the issue of uncertainty, whereby the current state of the world may not be known with certainty; and (3) the difficulties inherent in evaluating the desirable state of the world when it includes many different attributes that are not fully compatible (e.g., increasing market share and reducing costs). The process of human decision-making is represented in Figure 1.

In Figure 1, we have sketched what may be regarded as a realistic human decision process, tracking the main components of decision reasoning. For the sake of simplicity, we have divided the process into two main parts: diagnosis and look-ahead. It is, of course, not always easy to separate these two, but from an engineer's point of view, it facilitates the design of systems aimed at supporting the process of decision-making.

DECISION MACHINES, WHAT-IF ANALYSIS, AND LOOK-AHEAD

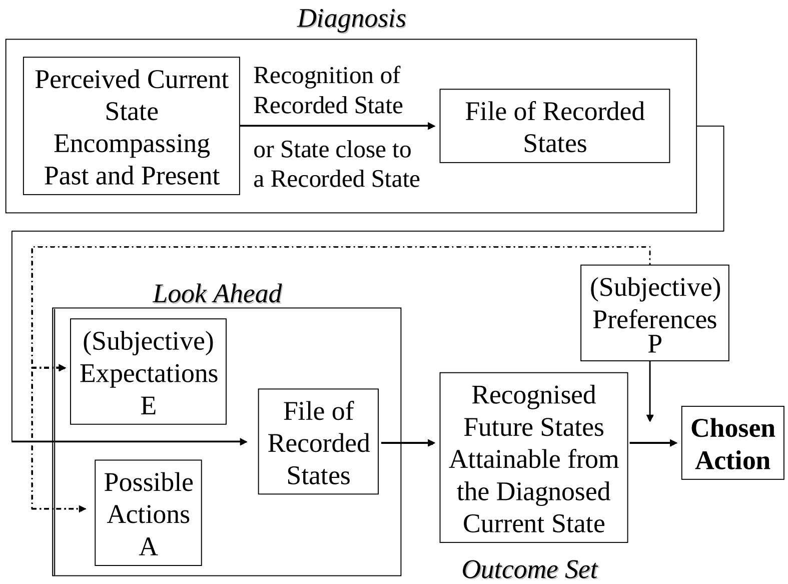
The model presented in Figure 1 can be used to predict what may be the most appropriate systems able to support the decision process. A number of corresponding DMSS designs can be proposed—diagnosis machine (or decision machine), "what-if" analysis machines, and look-ahead machines—which seek to tackle the most difficult and abstract level in human decision-making.

Decision Machines

A decision machine is an automaton adhering to one-to-one correspondence between the diagnosed current state and a proposed action. As said before, the word "decision" is, in this case, improper because the decision has already been made by the designer of the system. However, when people are unaware of the program or when it is so sophisticated that it is impossible to look through its operations, one can refer to these as decision machines. As such, most decision machines are mere classifiers linking a recognised state to an action. Numerous of these machines exist in the context of almost continuous decision (i.e., control of industrial processes, underground train driving, etc.).

With a programmed decision machine relating the current state to an action, one does not capture the full complexity of human decision-making. In addition, in many concrete situations, the set of all the possible current states cannot be described either extensionally or intentionally. Thus, the human decision maker is always

Figure 1. The decision process (adapted from Pomerol, 1997)



indispensable, working in an interactive way with the machine, mainly because unexpected (not programmed) states might occur. Many accidents have resulted from the bad recognition of the current state (wrong diagnosis) by a subject (Boy, 1991). Thus, the designers of decision support systems are confronted with the paradoxical problem of developing systems capable of helping people in situations that neither the user nor the program can foresee. This is one of the most difficult challenges in the development of decision support systems.

"What-If" Analysis

Although various frameworks have been proposed to cope with uncertainty, many decision makers discover that, in real situations, events are either very interdependent or the probabilities remain unknown (e.g., what is the probability that the price of oil will be higher in three months than today?), such that rational decision-making is impossible. Predicting or even identifying all the possible reactions of other agents and competitors is another key difficulty. The ability to envision the future and to anticipate events exists only in most advanced animals. Key components of the intelligent behavior of human beings are the capacity for anticipation, the ability to decide against immediate short-term advantage, and the desire to pursue future gains.. This type of multi-criteria choice can be regarded as a basic expression of rationality. This is consistent with Damasio's view that "will power is just another name for the idea of choosing according to long-term outcomes rather than short-term ones" (Damasio, 1994, p. 175).

In any case, "what-if" analysis or, more accurately, "scenario reasoning," should produce two outputs: all possible outcomes at a given horizon and the probability or plausibility of each outcome. Decision makers exercise their preferences on probabilistic outcomes (preferably multi-attribute), and then make their decisions and implement the resulting actions in accordance with the chosen scenario. Unfortunately for non-aided decision makers, scenario reasoning may lead to a combinatorial explosion such that it is often impossible to handle long, precise, and diverse scenarios (Pomerol, 2001). This is the very reason why support from machines is necessary.

Look-Ahead Machines

Two capabilities appear to be necessary in a look-ahead machine: (1) the ability to combine many actions and events (with their probabilities); and (2) the ability to imagine the possible actions and to anticipate all possible reactions of other agents and/or nature. According to Figure 1, this "imagination" ability is simply provided by the file of recorded states, such that for a given subject, all possible events and reactions of the other agents are drawn from a set of memorised items. However, forecasts never predict what is really new (Hogarth & Makridakis, 1981; Makridakis, 1990). It is, therefore, unlikely that look-ahead machines can escape this weakness. Another reason for using recorded states is that human forecasts are often too optimistic because human beings remember success more easily than failures (Kahneman & Lovallo, 1993). The intrinsic difficulty in forecasting is the main weakness of many formalised planning processes. This is particularly true for long-term or strategic planning (Mintzberg, 1994). Bearing in mind that the basis for scenario building is often restricted to the set of recorded actions, events, and situations (i.e., states), there are two types of look-ahead machines: simulators and decision-making support systems (DMSS).

A simulator is a machine in which a real industrial or social process has been modelled on a reduced scale. Then, some initial data and parameters are fed into the simulator and the user observes the evolution of the variables in which he or she is interested.

When it is possible to develop at least a partial model of the given process, it is possible to exploit it by developing a DMSS. Research and development conducted by one of the authors in the DMSS area over a period of many years led to the conclusion that DMSSs are look-ahead machines designed to perform an exploration at several cognitive levels. At the data level, it is called "what-if" or sensitivity analysis. Roy (1998) has also described another type of "what-if" analysis called robustness analysis. In robustness analysis, the question raised by the decision maker is, "What input modifications can a decision endure before really becoming bad?" It differs from sensitivity analysis, which looks for the largest modification of some input that leaves a decision unchanged.

DECISION-MAKING SUPPORT SYSTEMS

A Functional Definition for DMSS

There are many ways to conceive of and technically describe a DMSS, but most typical "textbook" definitions are not really helpful for the purpose of teaching students what DMSSs are about or for the purpose of developing better systems. Keen and Morton (1978) present a comprehensive sample of the definitions and ideas relying upon functional views put forward by IS researchers in the 1970s. These ideas are suitably summarised in the following quote.

Decision support implies the use of computers to: (1) assist managers in their decision process for semi-structured tasks, (2) support rather than replace managerial judgement and (3) improve the effectiveness of decision making rather than its efficiency.(Keen & Morton, 1978, p. 1)

The obvious drawback of such definitions is that, while they clearly describe what a DMSS is meant to do, they are weak at describing how it should do it and do not provide any clear guidelines for DMSS development or usage. According to Naylor (1982), this lack of formal theorisation of what a DMSS should do greatly contributed to the relative lack of success of the DMSS concept. During the 1980s, many authors attempted to go beyond the limitations inherent in such definitions and to put forward new ideas about the design and development of DMSS applications (Bonczek et al., 1981; Lévine & Pomerol, 1989; Sprague & Carlson, 1982). These new definitions were more or less inspired by Artificial Intelligence (AI), while the functional perspective was changed by the introduction of the concept of EIS. Since then, other concepts have been proposed either on the technological (i.e., OLAP, data warehouse) or conceptual side (I.e., Groupware, drill down, exception reporting, etc.), but this does not fundamentally change the nature of decision-making support nor does it change the difficulties inherent in the design of executive systems with clear decision-making relevance.

DMSS as Information Processing System

Bonczeck et al. (1981) were the first authors to introduce the idea of AI in DMSS theory and to regard a DMSS as a special kind of Information Processing System (IPS) as described by Newell and Simon (1972). This novel vision of DMSS amounts to viewing DMSS applications as essentially problem solving applications. Their vision of DMSS, as illustrated in Figure 2, shows the typical structure of a DMSS application.

In conformity with Newell and Simon's (1972) model, the IPS features a processor (the PPS), memories (the KS) and an input/output device (the LS). Sprague and Carlson (1982) extended this vision of a DMSS based on representation, operations, memory aids, and control mechanisms. The main contribution of Sprague and Carlson was to emphasise the importance of two additional factors: (1) the role of the representation—a certain structure applied to the data; and (2) the methods available to the decision maker to use and control the DMSS (i.e., how the DMSS supports the decision-making process).

In a simplified diagram, Sprague (1987) illustrates his vision of DMSS as a system with memories (a database), an operation performed according to programmed models, and a dialogue device (see Figure 3).

It is worth noting that the all-important concept of control has disappeared from this diagram, despite the acknowledgement of its importance for the successful use of the DMSS application. Lévine and Pomerol reintroduced the control aspect of DMSS in their vision of DMSS (1989, 1995). They proposed that using a DMSS amounts to performing a heuristic search up to the moment where the decision maker feels that he or she has gathered sufficient knowledge to make a decision.

Figure 2. Structure of a DSS according to Bonczek et al. (1981)

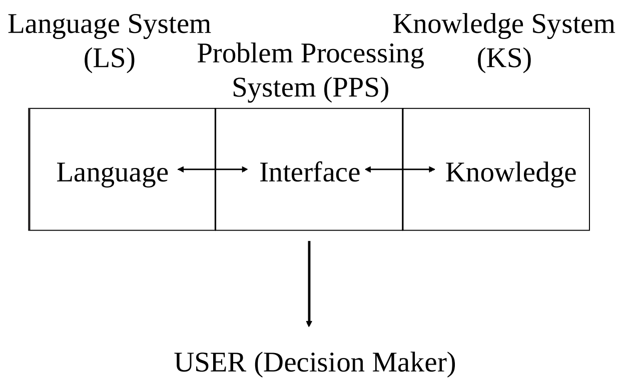
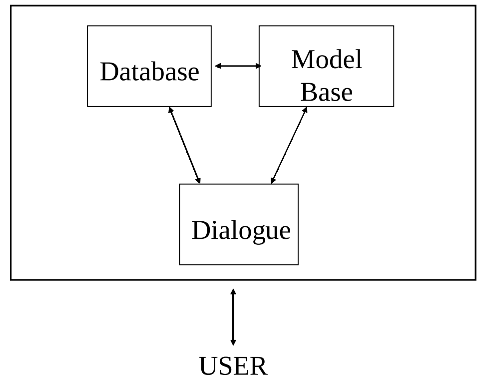


Figure 3. A DSS according to Sprague (1987)



Levine and Pomerol (1995) suggested that this heuristic search occurs at different levels. These representation levels initially described by Humphreys and Bekerley (1985) are characterised by their level of abstraction and a specific way of thinking that corresponds to the degree of formalisation of the problems. Humphreys (1989) later described how problem solving involves passing through the five levels from the highest to the lowest level of abstraction as described below.

(1) At the highest level, representations are mainly cultural and psychological; managers are more or less aware of what a problem involves, but its expression is mostly beyond language. It is at this level that the problem is shaped.
(2) At this level, representations become explicit, and the problem can be broken down into a number of sub-problems, some of which can be formalised. The structuration of the problems is still partial rather than detailed, and managers will refer to the "marketing function" or the "marketing process" without being able to formalise processes in greater details.
(3) At this level, decision makers are able to define the structure of the problems they must solve. They are able to put forward models that can be used for the investigation of the alternatives they will pursue.
(4) At this level, the decision makers will perform sensitivity analysis with the models they have defined in the previous stage in order to determine which input values are the most suitable.
(5) At the lowest level, managers decide upon the most suitable values, and the representation of the problem they must solve is stable and fully operational.

The process described by Humphreys is a top-down process whereby the structuration of the concepts investigated is refined from one level to the next until decisions can be made. Levels 1 and 2 are generally considered strategic levels of reflection handled by top executives, whereas the remaining three levels correspond to more operational and tactical levels. The top level is obviously a very important stage of the decision-making process, as the only factors limiting the horizon of the decision maker are either psychological (unconscious) or cultural (e.g., one's educational background). In human decision making, this step of shaping the framework for further studying the problem appears under the name "setting the agenda" (Simon, 1997) or "problem setting" (Checkland, 1981). The role of the person who shapes the problem, the "problem owner" in Checkland's words, is very important for the follow-up of the decision and action. This stage is all the more important in that it conditions the outcome of the decision-making process. Avenues that are not envisaged at that stage are unlikely to be explored at a later stage. The manager's vision of the problem is then considerably refined at the second level to prepare for the formalisation of the third level, where the manager enters the realm of "hard system thinking" and where representations and computational models must be fully stabilised. At the remaining levels, the manager will define the processing to be applied to the data and identify the most appropriate input values of the model. The refining process from top cognitive level to lower ones also can be interpreted as a decision process in which high-level decisions shape the decision framework, and within which smaller decisions are then made leading to actions, which is the ultimate level.

Based on this view of DMSS, we can conclude that existing DMSS applications often cover levels 3, 4, and 5 and neglect the two initial levels that, admittedly, are much more problematic to handle.

FUTURE TRENDS – TOWARDS A CONTINGENCY VIEW OF DMSS

The definition of DMSS we have proposed addresses four key problems:

(1) It takes into account the problem of representation levels.
(2) It accounts for the continued success of spreadsheet-based DMSS applications, which are still the most used DMSSs.
(3) It helps researchers understand the classical "what-if" analysis as a heuristic search for better insights (generally of a look-ahead type).
(4) It provides some insights into the reasons why the human decision maker is so important in the decision-making process (i.e., to express his or her preferences and to evaluate the outcomes generated by the DMSS).

This definition envisions the DMSS as an Information Processing System (IPS) designed so that the decision maker can perform a heuristic search at any representation level. At the lower levels (3,4, and 5), the model in the DMSS is virtually complete, and the system may be able to deliver a complete solution ready to be implemented. By contrast, at the higher levels (1 and 2), the model, only present in the manager's mind and in an incomplete form, is not available for modeling, and the decision-making process must be enacted as a dialogue between the DMSS and the decision maker. An important consequence is that the interaction that takes place between the DMSS and the manager is fundamentally different at the lower levels than at the higher levels. At the lower levels of abstraction, the exchanges between the IPS and the manager are limited to data entry and output of results; at the higher levels, they must enable the control of often abstract, heuristic searches.

Naturally, the development of DMSS applications aimed at the first and second levels is very ambitious. The support that decision makers get from their DMSS comes mainly from the ability of the underlying models to accurately represent and simulate reality. However, models are, at best, incomplete and can only provide partial support to the decision maker. In the future, developers of DMSS must address the challenges inherent in designing applications able to handle interaction with their users at the five different levels of abstraction.

CONCLUSION

In this chapter, we provide a technical definition of DMSS, taking into account the specificities of human decision making, particularly the look-ahead process often illustrated by scenario thinking and multi-attribute choice. The arbitration process required between short-term desires and long-term advantages can be supported by tools allowing a heuristic search at different cognitive levels. The search stops when a satisfying trade-off has been attained. This introduces the DMSS as a look-ahead machine and paves the way to a better understanding and a better design of future DMSS.

REFERENCES

Bonczek, R.H., Holsapple, C.W., & Whinston, A.B. (1981). *Foundations of decision support systems*. New York: Academic Press.

Boy, G. (1991). *Intelligent assistant systems*. New York: Academic Press.

Checkland, P.D. (1981). *Systems thinking, systems practice*. Chichester: John Wiley and Sons.

Damasio, A.R. (1994). *Descartes' error*. New York: Pitman and Sons.

Gorry, A., & Morton, Scott M. (1971). A framework for management information systems. *Sloan Management Review, 13*(1), 55-70.

Hogarth, R.M., & Makridakis, S. (1981). Forecasting and planing: An evaluation. *Management Science, 27*, 115-138.

Huber, G. (1984). The nature and design of post-industrial organisations. *Management Science, 30*(8), 928-951.

Humphreys, P. (1989). Intelligence in decision making - A process model. In G. Doukidis, F. Land, & E. Miller (Eds.), *Knowledge-based management systems*. Chichester: Hellis Hovwood.

Humphrey, P., & Bekerley, D. (1985). Handling uncertainty: Levels of analysis of decision problems. In G. Wright (Ed.), *Behavioural decision making*. London: Plenum Press.

Kahneman, D., & Lovallo, D. (1993). Timid choice and bold forecast: A cognitive perspective on risk taking. *Management Science, 39*, pp. 17-31.

Keen, P.G., & Morton, Scott M. (1978). *Decision support systems: An organisational perspective*. Reading, MA: Addison-Wesley.

Lévine P., & Pomerol, J.-Ch. (1989). *Systèmes interactifs d'aide à la décision et systèmes experts*. Paris: Hermès.

Lévine, P., & Pomerol, J.-Ch. (1995). The role of the decision maker in DSSs and representation levels. *Proceedings of the 29th Hawaiï International Conference on System Sciences*, *3*.

Makridakis, S. (1990). *Forecasting, planning and the strategy for the 21st century*. New York: Free Press.

Mintzberg, H. (1994). *The rise and fall of strategic planning*. New York: Free Press.

Naylor, T.H. (1982). Decision support systems or whatever happened to MIS? *Interfaces*, *12*(4), 92-94.

Newell, A., & Simon, H. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Pomerol, J.-Ch. (1997). Artificial intelligence and human decision making. *European Journal of Operational Research*, *99*, 3-25.

Pomerol, J.-Ch. (2001). Scenario development and practical decision making under uncertainty. *Decision Support Systems*, *31*, 197-204.

Pomerol, J.-Ch., & Adam, F. (2004, July). Practical decision making – From the legacy of Herbert Simon to decision support systems. *IFIP WG 8.3 Conference*, Prato, Italy.

Roy, B. (1998). A missing link in OR-DA: Robustness analysis. *Foundations of Computing and Decision Science*, *23*, 141-160.

Savage, L.J. (1954). *The foundations of statistics*. New York: John Wiley and Sons.

Simon, H.-A. (1977). *The new science of management decision*, revised edition (first edition 1963). Englewood Cliffs, NJ: Prentice-Hall.

Simon, H.-A. (1997). *Administrative behavior*, 4ième édition (first edition 1947). New York: Free Press.

Sprague, R. (1987). DSS in context. *Decision Support Systems*, *3*(2), 197-202.

Sprague, R., & Carlson, E.D. (1982). *Building effective decision support systems*. Englewood Cliffs, NJ: Prentice Hall.

KEY TERMS

Decision-Making Process: Process whereby managers make decisions, including the stages described by Herbert Simon: intelligence, design, choice, and review.

Decision Support System: A computer system that enables managers to solve a given problem in their own personalised way.

Look-Ahead: Ability of human decision makers to anticipate all potential future scenarios and the likelihood that they will come to pass.

Representation Level: Degree to which the problems faced by managers are well formalized in their own minds and can be represented for the purpose of DSS development.

Robustness Analysis: A form of "what-if" analysis that seeks the maximum variation of a parameter that keeps a decision within acceptable values.

Sensitivity Analysis: A form of "what-if" analysis that seeks the largest variation of an input value into a model input that leaves the outcome of the decision unchanged.

Structure: Characteristic of a decision situation that determines whether the problem it involves can be programmed or described in specific terms (i.e., for the purpose of developing a relevant DSS).

Defining and Understanding ERP Systems

David Sammon
University College Cork, Ireland

Frédéric Adam
University College Cork, Ireland

INTRODUCTION

The last 15 years have seen the emergence on the software market of a category of software called Enterprise Resource Planning systems or ERP, which has become the focus of both researchers and practitioners in the information systems area. At this time, the ERP software market is one of the fastest growing markets in the software industry with long-term growth rates of 36-40%. Some estimates put the eventual size of the market by the year 2010 at US$1 trillion (Bingi et al., 1999). Since these estimates have been put forward, the ERP market has slowed down, but the overall growth of the enterprise-wide application market is still quite strong, thanks to a number of additional segments, such as Customer Relationship Management (CRM) and Supply Chain Management (SCM). Also, more recently, a new trend is emerging in the market: the re-implementation and extension of ERP, referred to as ERP II (Humphries and Jimenez, 2003).
Fundamentally, ERPs are all integrated "mega packages" (Gable et al., 1997) which provide support for several or all functional areas of the firm depending upon the configuration purchased by the client. Their complexity is reflected in the complexity of their implementation and deployment in organisations where they have been observed to have a substantial impact on everyday activities in both the short term and the long term. This has led to many reports of unsuccessful implementation, which are however matched by many reports of substantial benefits accruing to implementing firms. Thus, managers look upon ERP software as necessary evils and much research has been carried out in order to increase the success rate of ERP implementations and to ensure that benefits materialize.

THE EMERGENCE OF ERP

The historical origin of ERP is in inventory management and control software packages that dictated system design during the 1960s (Kalakota and Robinson, 2001). The 1970s saw the emergence of Material Requirements Planning (MRP) and Distribution Resource Planning (DRP), which focused on automating all aspects of production master scheduling and centralised inventory planning, respectively (Kalakota and Robinson, 2001). During the 1980s, the misnamed MRPII (Manufacturing Resource Planning) systems emerged to extend MRP's traditional focus on production processes to other business functions, including order processing, manufacturing, and distribution (Kalakota and Robinson, 2001). In the early 1990s, MRPII was further extended to cover areas of Engineering, Finance, Human Resources, Project Management, etc. MRPII is a misnomer, as it provided automated solutions to a wide range of business processes, not just those found within a company's manufacturing and distribution functions (Kalakota and Robinson, 2001). However, although MRP II systems overcame some of the drawbacks of MRP systems they became less relevant because:

- Manufacturing is moving away from a "make to stock" situation and towards a "make to order" ethos where customisation is replacing standardisation. This has lead to a far more complex planning process.
- Quality and cost are only minimum requirements for organisations wishing to compete in the marketplace. Competition has moved to a basis of aggressive delivery, lead-times, flexibility and greater integration with suppliers and customers with greater levels of product differentiation.

As a result, MRPII was further extended and renamed ERP (Kalakota and Robinson, 2001). An ERP system differs from the MRPII system, not only in system requirements, but also in technical requirements, as it addresses technology aspects such as graphical user interface, relational database, use of fourth generation language, and computer-aided software engineering tools in development, client/server architecture, and open-systems portability (Russell and Taylor, 1998; Watson and Schneider, 1999). Also, while "MRP II has traditionally focused on the planning and scheduling on internal resources, ERP strives to plan and schedule supplier resources as well, based on the dynamic customer demands and schedules" (Chen, 2001). This brief evolutionary definition of ERP is depicted in Figure 1.

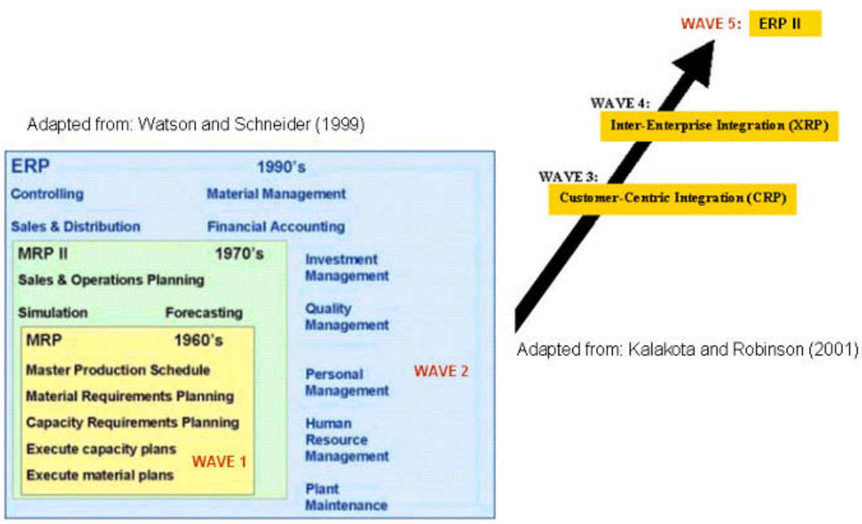
Kalakota and Robinson (2001) position ERP as the second phase in the "technology" and "enterprises internal and external constituencies" integration process, as illustrated in Figure 1. According to Kalakota and Robinson (2001), *Wave 1* of the evolution of ERP addresses the emergence of Manufacturing Integration (MRP), while *Wave 2* relates to Enterprise Integration (ERP). The combined impact of "key business drivers" (replacing legacy systems, gaining greater control, managing globalisation, handling regulatory change, and improving integration of functions across the enterprise) forced the "structural migration" from MRP to ERP (Kalakota and Robinson, 2001). Another significant factor in the second wave of ERP development was Y2K preparation, which was often cited as the major reason for ERP adoption (Brown et al., 2000; Kalakota and Robinson, 2001; Themistocleous et al., 2001). A new wave, *Wave 5*, now exists and positions ERP II as the new approach to enterprise integration.

ERP DEFINED

Although there is no agreed-upon definition for ERP systems, their characteristics position these systems as integrated, all-encompassing (Markus and Tanis, 2000; Pallatto, 2002), complex mega packages (Gable et al., 1997) designed to support the key functional areas of an organisation. The American Production and Inventory Control Society (APICS) defines ERP as, "an accounting-oriented information system for identifying and planning the enterprise-wide resources needed to take, make, ship, and account for customer orders" (Watson and Schneider, 1999). As a result, by definition, ERP is an operational-level system. Therefore, an Enterprise Resource Planning (ERP) system is a generic term for an integrated enterprise-wide standard information system (Watson and Schneider, 1999) that impounds deep knowledge of business practices accumulated from vendor implementations throughout organisations (Shang and Seddon, 2000).

ERP can be further defined as a strategic business solution that integrates all business functions, including manufacturing, financial, and distribution (Watson and Schneider, 1999). ERP systems are also being referred to as "enterprise systems" (Davenport, 1998; Chen, 2001) and "enterprise-wide Information Systems" (Al-Mashari, 2000; Milford and Stewart, 2000). It is a customised, packaged, software-based system that handles the majority of an enterprise's information systems' requirements (Watson and Schneider, 1999). It is a software architecture that facilitates the flow of information among all functions within an enterprise (Watson and Schneider, 1999). As a result, ERP systems are traditionally thought of as transaction-oriented processing systems (Davenport, 1998; Chen, 2001) or transactional backbones (Kalakota and Robinson, 2001). However, they are continually redefined based on the growing needs of organisations. Therefore, various definitions point to ERP systems as being enterprise-wide information systems that accommodate many features of an organisation's business processes. They are highly complex, integrated systems, which require careful consideration before selection, implementation, and use. Neglect of any of these areas can lead a company down the path to failure already

Figure 1. Evolution of ERP Systems



worn by FoxMeyer, Unisource Worldwide Inc., etc. (Adam and Sammon, 2004).

INSIDE ERP SYSTEMS

ERP systems use a modular structure (i.e., multi-module) to support a broad spectrum of key operational areas of the organisation. According to Kalakota and Robinson (2001), the multiple core applications comprising an ERP system are "themselves built from smaller software modules that perform specific business processes within a given functional area. For example, a manufacturing application normally includes modules that permit sales and inventory tracking, forecasting raw-material requirements, and planning plant maintenance." Typically, an ERP system is integrated across the enterprise with a common relational database, storing data on every function. ERP are widely acknowledged as having the potential to radically change existing businesses by bringing improvements in efficiency, effectiveness, and the implementation of optimised business processes (Rowe, 1999). One of the key reasons why managers have sought to proceed with difficult ERP projects is to end the fragmentation of current systems, to allow a process of standardisation, to give more visibility on data across the entire corporation and, in some cases, to obtain competitive advantage (Adam and Sammon, 2004).

ERP systems have expanded to include "back office" functions, such as operations, logistics, financials or human resources and "non-transaction-based systems" (Davenport, 1998) or "front office" functions, such as sales, marketing, customer services as an integral component (Davenport, 1998; Chen, 2001). This inclusion is a result of the emergence of Supply Chain Management (SCM) (Chen, 2001; Turban, et al., 2001) and Customer Relationship Management (CRM) strategies and systems (Chen, 2001), as illustrated in Figure 2.

While the names and numbers of modules of the ERP systems available on the market may differ, a typical system integrates all its functions by allowing modules to share and transfer information freely and by centralizing all information in a single database (Chen, 2001). Figure 3 provides an overview of such an ERP system.

ERP packages force an organisation to implement a proven set of business processes, which means that there is no need for the organisation to "reinvent the wheel." ERP packages encapsulate reusable "best practice" business processes. As state of the art technology and processes move forward, purchasers of packaged software move with them (Weston, 1998; Krumbholz et al., 2001; Stefanou, 2000). ERP packages give the foundation to the business, thus the management can concentrate on "grabbing market share" (Weston, 1998). Kalakota and Robinson (2001) stress that the popularity of ERP systems stems from the fact that they appear to solve the challenges posed by portfolios of "disconnected, uncoordinated applications that have outlived their usefulness."

These "legacy systems" provide one of the biggest drags on business productivity and performance because maintaining many different computer systems leads to

Figure 2. ERP Extended

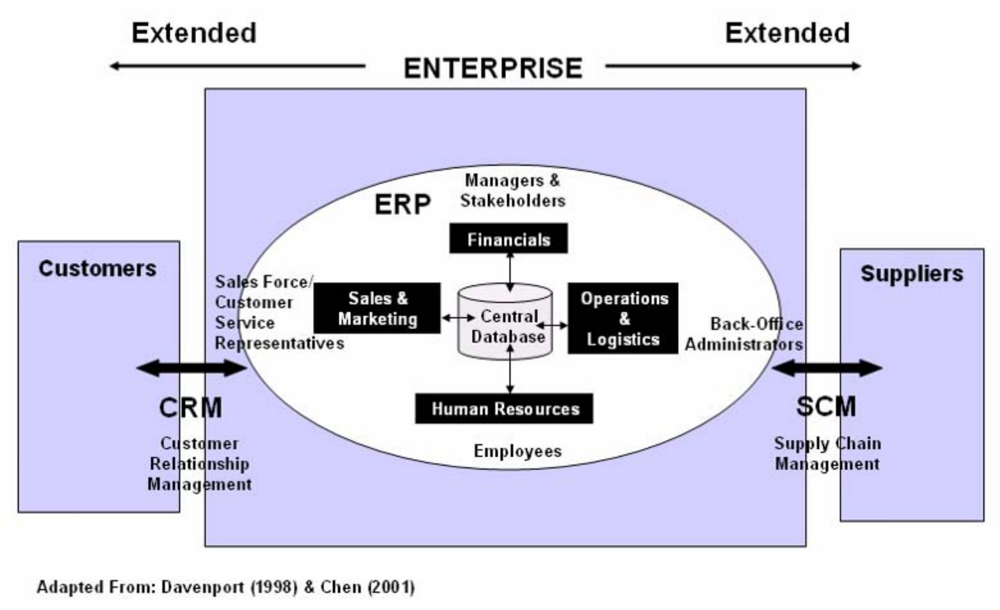
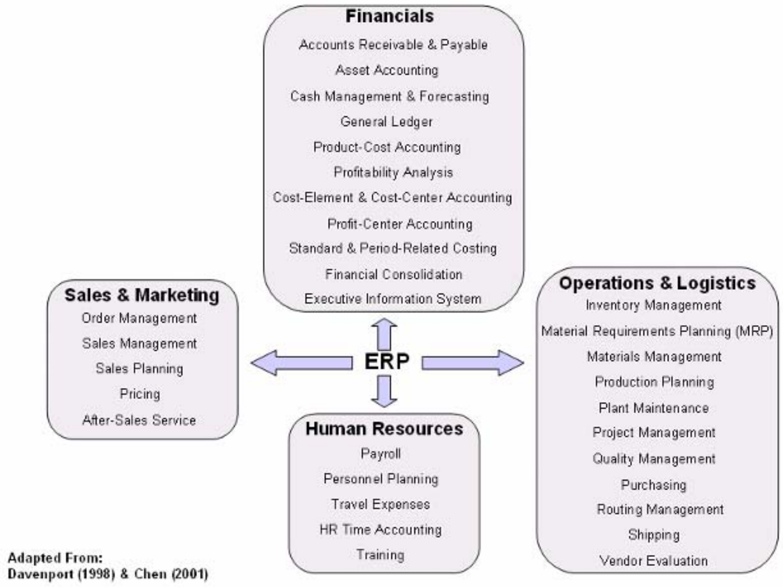


Figure 3. Module Overview of an ERP System



enormous costs. These include: direct costs such as rationalisation, redundancy, re-keying, reformatting, updating, debugging, deleting, etc., but more importantly, include indirect costs such as a company's purchasing and sales system which cannot communicate with its production/scheduling systems, leading manufacturing productivity and customer service to suffer. Crucially, management may be left to make vital decisions based on information from incompatible systems, thereby relying on instinct rather than sound business rationale.

MacVittie (2001) identified three goals behind the implementation of an ERP system:

1. Integration of Financial Data: When managers depend on their function's or unit's perspective of financial data, conflicting interpretations will arise (e.g., Finance will have one set of sales figures while Marketing will have another). Using an ERP system provides a single version of sales.
2. Standardisation of Processes: A manufacturing company that has grown through acquisitions is likely to find that different units use different methods to build the same product. Standardising processes and using an integrated computer system can save time, increase productivity, and reduce head count. It may also enable collaboration and scheduling of production across different units and sites.
3. Standardisation of Human Resource Information: This is especially useful in a multi-site company. A unified method for tracking employee time and communicating benefits is extremely beneficial because it promotes a sense of fairness among the workforce as well as streamlining the company as a whole.

BENEFITS OF ERP

Clearly, a properly implemented ERP system can achieve unprecedented benefits for business computing (Watson and Schneider, 1999). However, some companies have difficulty identifying any measurable benefits or business process improvements (James and Wolf, 2000; Donovan, 2001). For example, Pallatto (2002) observed that some vendors and consultants are presently "soft-peddling" the term ERP due to bad experiences and management frustration, when original business goals and benefits were not achieved, with their ERP implementations.

Rutherford (2001) observed that only around 10% to 15% of ERP implementations deliver the anticipated benefits. According to James and Wolf (2000) companies that were able to identify benefits thought that they could have been realized without the implemented ERP system. They felt that most the benefit they got from their ERP came from changes, such as inventory optimization, which could have been achieved without investing in ERP (James and Wolf, 2000). Therefore, ERP systems can be considered as a catalyst for radical business change that results in significant performance improvement (Watson and Schneider, 1999). According to James and Wolf (2000), reporting on an instance of an ERP implementation, "many of the benefits that we are able to achieve today could not have been predicted at the time that we started work on ERP. In fact, in hindsight it appears that much of the value of these large systems lay in the infrastructure foundation they created for future growth based on Information Technology."

Shang and Seddon (2000) presented a comprehensive framework of business benefits that organisations might be able to achieve from their use of ERP systems. They present 21 ERP benefits consolidated across five benefit dimensions (operational, managerial, strategic, IT infrastructure, organizational). Shang and Seddon (2000) analyzed the features of ERP systems, literature on IT benefits, Web-based data on 233 ERP-vendor success stories, and interviews with 34 ERP cases to provide a comprehensive foundation for planning, justifying and managing the ERP system. The focus and goal of the Shang and Seddon (2000) framework is "to develop a benefits classification that considers benefits from the point of view of an organisation's senior management" and that can be used as a "communication tool and checklist for consensus-building in within-firm discussions on benefits realisation and development" (Shang and Seddon 2000).

Shang and Seddon (2000) also commented that there were few details of ERP-specific benefits in academic literature and that "trade-press articles" and "vendor-published success stories" were the major sources of data. However, they warned that, "cases provided by vendors may exaggerate product strength and business benefits, and omit shortcomings of the products." This was also observed in Adam and Sammon's (2004) study of the "ERP Community" and the *sales* and *needs* discourse that characterise it. It is therefore extremely important to be able to assess the suitability of the ERP system for any organization. A study conducted by Sammon and Lawlor (2004) reiterates this argument, highlighting that failure to carry out an analysis of the mandatory and desirable features required in a system with an open mind will lead to the blind acceptance of the models underlying the ERP packages currently on sale on the market, with detrimental effects on the organization and its operations.

The justification for adopting ERPs centres around their business benefits. However, Donovan (1998) believes that to receive benefit from implementing ERP there must be no misunderstanding of what it is about, or underestimation of what is involved in implementing it effectively, and even more importantly, organisational decision makers must have the background and temperament for this type of decision making (Donovan, 2001).

CONCLUSION

Kalakota and Robinson (1999) put forward four reasons why firms are prepared to spend considerable amounts on ERP systems:

1. ERP systems create a foundation upon which all the applications a firm may need in the future can be developed.
2. ERP systems integrate a broad range of disparate technologies into a common denominator of overall functionality.
3. ERP systems create a framework that will improve customer order processing systems, which have been neglected in recent years.
4. ERP systems consolidate and unify business functions such as manufacturing, finance, distribution and human resources.

Thus, despite the slowdown of ERP package sales from 1999 on (Remy, 2003), there is evidence that the trend towards ERP and extended ERP systems is well established (Stefanou, 2000). Some segments of the ERP market maybe saturated, but powerful driving forces including e-commerce (Bhattacherjee, 2000), Enterprise Application Integration (EAI) (Markus, 2001), Data Warehousing (DW) (Inmon, 2000; Markus, 2001; Sammon et al., 2003) and ERP II (Pallatto, 2002) will further push this market to new heights in years to come. This will mean that considerable proportions of the IT investment of many firms will converge towards ERP-type projects and that IS researchers need to pursue their efforts at developing better frameworks and better methodologies for the successful deployment of Enterprise Systems in firms.

REFERENCES

Adam, F., & Sammon, D. (2004). *The enterprise resource planning decade: Lessons learned and issues for the future*. Hershey, PA: Idea Publishing Group.

Al-Mashari, M. (2000). Constructs of process change management in ERP context: A focus on SAP R/3. In *Proceedings of the 6th Americas Conference on Information Systems*, August 10-13, Long Beach, California, USA.

Bhattacherjee, A. (2000, August). Beginning SAP R/3 implementation at Geneva Pharmaceuticals. *Communications of the AIS*, *4*(2), 1-39.

Bingi, P., Sharma, M., & Godla, J. (1999, Summer). Critical issues affecting an ERP implementation. *Information Systems Management*, 7-14.

Brown, C.V., Vessey, I., & Powell, A. (2000). The ERP purchase decision: Influential business and IT factors. In *Proceedings of the 6th Americas Conference on Information Systems*, August 10-13, Long Beach, California, USA (pp. 1029-1032).

Chen, I.J. (2001). Planning for ERP systems: Analysis and future trend. *Business Process Management Journal*, *7*(5), 374-386.

Davenport, T. (1998, July-August). Putting the enterprise into the enterprise system. *Harvard Business Review*, 131.

Donovan, M. (1998, September). There is no magic in ERP software: It's in preparation of the process and people. *Midrange ERP*, 8.

Donovan, M. (2001). Successful ERP implementation the first time. Retrieved July 25, 2001 from the World Wide Web at: *www.mdonovan.com/pdf/perfor8.pdf*

Gable, G., van den Heever, R., Scott, J., & Erlank, S. (1997). Large packaged software: The need for research. In *Proceedings of the 3rd Pacific Asia Conference on Information Systems (PACIS)*.

Humphries, T., & Jimenez, M. (2003). ERP scenario. In *Proceedings of Software Infrastructure track at Gartner Symposium ITXPO*, Florence, Italy, March 10-12.

Inmon, W.H. (2000). ERP and data warehouse: Reading the tea leaves. Retrieved October 18, 2001 from the World Wide Web at: *www.billinmon.com/library/articles/arterpfu.asp*

James, D., & Wolf, M.L. (2000). A second wind for ERP. *The McKinsey Quarterly*, 2, 100-107. Retrieved October 1, 2002 from the World Wide Web at: *www.mckinsely quarterly.com*

Kalakota, R., & Robinson, M. (2001). *E-business 2.0: Roadmap to success*. Reading, MA: Addison-Wesley.

Krumbholz, M., Galliers, J., Coulianos, N., & Maiden, N.A.M. (2000). Implementing enterprise resource planning packages in different corporate and national cultures. *Journal of Information Technology*, 15, 267-279.

McVittie, L. (2001, March). Buckle up: Implementing ERP takes time and patience. Retrieved from the World Wide Web at: *www.networkcomputing.com*

Markus, M.L. (2001). Viewpoint. *Business Process Management Journal*, 7(3), 171-180.

Markus, M.L., & Tanis, C. (2000). The enterprise systems experience – From adoption to success. In R.W. Zmud (Ed.), *Framing the domains of IT management: Projecting the future through the past*. Retrieved September 10, 2001 from the World Wide Web at: *Pinnaflex.com*

Milford, M., & Stewart, G. (2000). Are ERP implementations qualitatively different from other large systems implementations? In *Proceedings of the 6th Americas Conference on Information Systems*, August 10-13, Long Beach, California, USA (pp. 966-971).

Pallatto, J. (2002). Get the most out of ERP. IT toolbox – ERP knowledge base. Retrieved October 15, 2002 from the World Wide Web at: *www.erpassist.com/documents/document.asp?i=1334*

Remy, C. (2003, June 6). Les PGI nouvelle vague tenus d'apporter intégration et RSI. *Le Monde Informatique*, 985.

Rowe, F. (1999). Cohérence, Intégration informationnelle et changement: Esquisse d'un programme de recherche à partir des Progiciels Intégrés de Gestion. *Systèmes d'Information et Management*, *4*(4), 3-20.

Russell, R.S., & Taylor, B.W. (1995). *Production and operations management: Focusing on quality and competitiveness*. Englewood Cliffs, NJ: Prentice Hall.

Rutherford, E. (2001). ERP's ends justify its means. CIO Online Quick Poll Report. Retrieved July 18, 2001 from the World Wide Web at: *http://www.cio.com/poll/042401_erp.html*

Sammon, D., Adam, F., & Carton, F. (2003). The realities of benefit realisation in the monolithic enterprise systems era: Considerations for the future. In *Proceedings of the 10th European Conference on Information Technology Evaluation*, Madrid, Spain, September 25-26.

Sammon, D., & Lawlor, D. (2004). An examination of an ERP software selection process: An Irish case study. In F. Adam & D. Sammon (Eds.), *The enterprise resource planning decade: Lessons learned and issues for the future*. Hershey, PA: Idea Publishing Group.

Shang, S., & Seddon, P. (2000). A comprehensive framework for classifying the benefits of ERP systems. In *Proceedings of the 6th Americas Conference on Information Systems*, August 10-13, Long Beach, California, USA (pp. 1005-1014).

Stefanou, C. (2000). The selection process of enterprise resource planning, ERP, systems. In *Proceedings of the 6th Americas Conference on Information Systems*, August 10-13, Long Beach, California (pp. 988-991).

Turban, E., McLean, E., & Wetherbe, J. (2001). *Information technology for management: Making connections for strategic advantage* (2nd ed.). New York: Wiley.

Watson, E., & Schneider, H. (1999, February). Using ERP systems in education. *Communications of the AIS*, 1(9), 1-47.

Weston, R. (1998, January). ERP users find competitive advantages. *Computerworld*, 19.

KEY TERMS

APICS: The American Production and Inventory Control Society was founded in 1957 and is a global leading provider of information and services in production and inventory management.

ERP Community: A model defining the collective relationships and interactions between the three *de facto* actors (the ERP vendor, the ERP consultant, the implementing organization) within the ERP market.

ERP II: ERP II is understood to mean the re-implementation and expansion of ERP. It is an extended, open, vertical, and global approach to systems integration and can be understood as an application and deployment strategy for collaborative, operational, and financial processes within the enterprise and between the enterprise and key external partners and markets, in an effort to provide deep, vertical-specific functionality coupled with external connectivity.

Extended Enterprise: Understood as the seamless Internet-based integration of a group or network of trading partners along their supply chains.

Legacy Systems: Mission-critical "aging" systems that supported business functions for many years. However, they are no longer considered state-of-the-art technology and have limitations in design and use. Within organisations, the vast majority were replaced pre-Y2K with ERP systems.

Modular System Design: An ERP system can be broken down by functional area (e.g., sales, accounting, inventory, purchasing, etc.). It increases flexibility in terms of implementing mutli-module (from one or many vendors), or single-module ERP functionality.

Return On Investment (ROI): A technique used for measuring the return on an investment and is often used in the justification of new ERP systems, or to measure how well an ERP system has been implemented, or for implementing new or additional functionality.

D

Delineating Knowledge Flows for Enterprise Agility

Mark E. Nissen
Naval Postgraduate School, USA

INTRODUCTION

The practice of knowledge management (KM) purports to take the power of knowledge to the group, organization and even enterprise level (Davenport & Prusak, 1998). Although this potential benefit of KM is not viewed universally (Gore & Gore, 1999; McDermott, 1999), many scholars (e.g., Drucker, 1995) assert that knowledge represents one of the very few sustainable sources of competitive advantage. Hence, the knowledge-based organization—one that competes on the basis of its differential knowledge (e.g., see Grant, 1996, for discussion of the knowledge-based view of the firm)—appears to offer great promise in terms of performance and capability.

Drawing from Nissen (2004), the knowledge-based organization must be able to apply substantial knowledge, when and where it's needed, to affect organizational goals. However, knowledge is not distributed evenly through the organization, so rapid and efficient knowledge flow is critical to enterprise performance. The larger, more geographically dispersed, and time-critical an enterprise (e.g., global manufacturing firms, telecommunication and software companies, military forces), the more important knowledge flow becomes in terms of efficacy. Unfortunately, our collective knowledge of how knowledge flows is quite primitive (Alavi & Leidner, 2001; Nissen, 2002). Lacking knowledge-flow theory and application for guidance, even enterprises with multimillion-dollar KM projects have difficulty seeing past *information* technologies such as intranets and Web portals. Further, Nissen, Kamel, and Sengupta (2000) note such KM projects rely principally upon trial and error, one of the least effective approaches known.

BACKGROUND

This section draws heavily from Nissen (2004) to summarize key background pertaining to knowledge flow. It focuses in particular on important concepts from the emerging literature on knowledge management and augments current work on knowledge flows (e.g., Baumard, 2002; Echeveria-Carroll, 1999; Fang, Lin, Hsiao, Huang, & Fang, 2002; Foss & Pedersen, 2002; Gupta & Govindarajan, 2000; Schulz & Jobe, 2001; Zhuge, 2002). For the purposes of this article, three important concepts from the KM literature are summarized: 1) knowledge hierarchy, 2) knowledge management life cycle, and 3) current knowledge-flow theory.

Knowledge Hierarchy

Many scholars (e.g., Davenport & Prusak, 1998; Nissen et al., 2000; von Krogh, Ichijo, & Nonaka, 2000) conceptualize a hierarchy of knowledge, information, and data. As illustrated in Figure 1, each level of the hierarchy builds on the one below. For example, data are required to produce information, but information involves more than just data (e.g., need to have the data in context). Similarly, information is required to produce knowledge, but knowledge involves more than just information (e.g., it enables action). We operationalize the triangular shape of this hierarchy using two dimensions—abundance and actionability—to differentiate among the three constructs.

Briefly, data lie at the bottom level, with information in the middle and knowledge at the top. The broad base of the triangle reflects the abundance of data, with exponentially less information available than data, and even fewer chunks of knowledge in any particular domain. Thus, the width of the triangle at each level reflects decreasing abundance in the progress from data to knowledge. The height of the triangle at each level reflects actionability (i.e., the ability to take appropriate action, such as a good decision or effective behavior). Converse to their abundance, data are not particularly powerful for supporting action, and information is more powerful than data. But knowledge supports action directly, hence its position near the top of the triangle. Curiously, there is current speculation as to one or more additional levels "above" *knowledge* in such hierarchies (e.g., *wisdom*; cf. Spiegler, 2000). The present article does not attempt to address "wisdom flow."

Knowledge Management Life Cycle

Nissen et al. (2000) observe a sense of process flow or a life cycle associated with knowledge management. Integrating their survey of the literature (e.g., Davenport &

Figure 1. Knowledge hierarchy (adapted from Nissen, 2002)

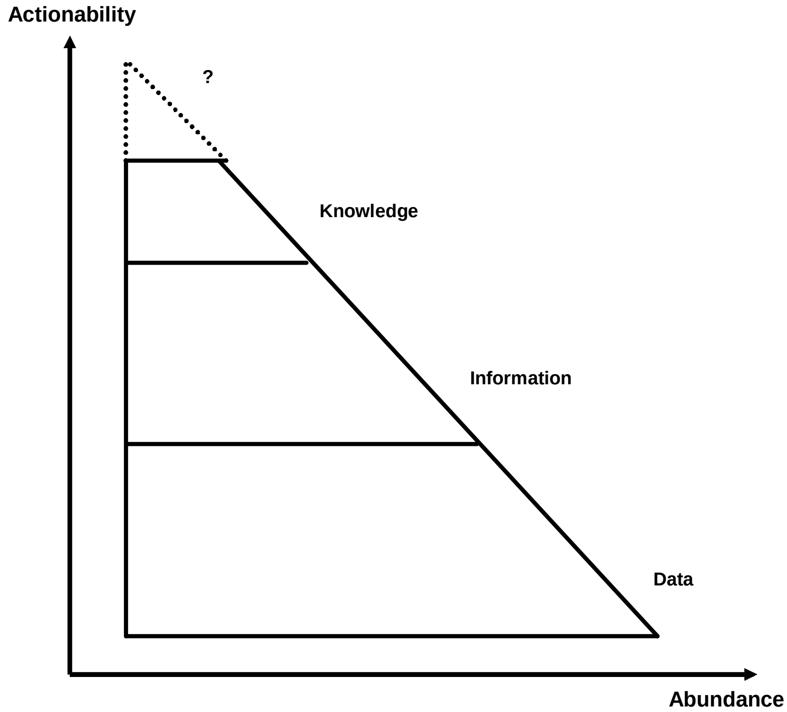


Table 1. Knowledge management life cycle models (Adapted from Nissen et al., 2000)

Model	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6
Despres and Chauvel	Create	Map/ bundle	Store	Share/ transfer	Reuse	Evolve
Gartner Group	Create	Organize	Capture	Access	Use	
Davenport & Prusak	Generate		Codify	Transfer		
Nissen	Capture	Organize	Formalize	Distribute	Apply	
Amalgamated	Create	Organize	Formalize	Distribute	Apply	Evolve

Prusak, 1998; Despres & Chauvel, 1999; Gartner Group, 1999; Nissen, 1999), they synthesize an amalgamated KM life cycle model as outlined at the bottom of Table 1.

Briefly, the creation phase begins the life cycle, as new knowledge is generated within an enterprise; similar terms from other models include *capture* and *acquire*. The second phase pertains to the organization, mapping, or bundling of knowledge, often employing systems such as taxonomies, ontologies, and repositories. Phase 3 addresses mechanisms for making knowledge formal or explicit; similar terms from other models include *store* and *codify*. The fourth phase concerns the ability to share or distribute knowledge in the enterprise; this also includes terms such as *transfer* and *access*. Knowledge use and application for problem solving or decision making in the organization constitutes Phase 5, and a sixth phase is included to cover knowledge refinement and evolution, which reflects organizational learning—and thus a return to knowledge creation—through time. It is important to note, as in the familiar life cycle models used in IS design (e.g., System Development Life Cycle), progression through the various phases of this Life Cycle Model is generally iterative and involves feedback loops between stages; that is, all steps need not be taken in order, and the flow through this life cycle is not necessarily unidirectional.

Current Knowledge-Flow Theory

This section summarizes the dynamic model developed by Nissen (2002). It begins by building upon Nonaka's (1994) work to conceptualize an extended model of knowledge-flow dynamics. This extended model is intended to help managers to understand better how enhancing knowledge flow can increase enterprise intelligence. The first step is to augment Nonaka's two-dimensional framework by incorporating a third dimension, the *KM life cycle*. We operationalize the construct using the life cycle stages from the Amalgamated Model presented in Table 1.

The second step requires relabeling the epistemological and ontological dimensions. Because the terms *epistemological* and *ontological* can be confusing when used in the present context, we relabel them as *explicitness* and *reach*, respectively, to target directly the principal focus of each dimension (i.e., differentiating explicit vs. tacit knowledge; indicating how broadly knowledge reaches through an enterprise).

In Figure 2, we note a few notional vectors for illustrating and classifying various dynamic patterns of knowledge as it flows through the enterprise. For example, the simple linear flow labeled "P&P" (i.e., Policies & Procedures) depicts the manner in which most enterprises inform, train, and attempt to acculturate employees: explicit documents and guidelines that individuals in the organization are expected to memorize, refer to, and observe. As another example, the cyclical flow of knowledge labeled "KMLC" (i.e., KM life cycle) reflects a more complex dynamic than its simple linear counterpart. This flow describes a cycle of knowledge creation, distribution, and evolution within a workgroup, for example.

Further, Nonaka's dynamic model of knowledge flow can also be described in this space by the curvilinear vector sequence corresponding to the processes labeled "SECI" (i.e., socialize, externalize, combine, internalize), "externalize", "combine", and "internalize". Thus, our model subsumes the one proposed by Nonaka and shows a complex dynamic as knowledge flows along the life cycle. Moreover, examination of this space suggests also including (and connecting) the "create" and "refine" vectors, which are not part of Nonaka's theory but represent key elements of the empirically derived Life Cycle Model (e.g., the key to knowledge generation and evolution). Clearly, a great many other flows and patterns can be shown in this manner.

APPLICATION EXAMPLE – NEW PRODUCT DEVELOPMENT

Drawing concepts from Nissen (2004) in this section, we know the business environment now changes very rapidly, and the frequency of such changes is only expected to increase through time. Business processes must change even more rapidly for the knowledge-based organization to anticipate—not simply react to—new competitive requirements and market opportunities. Hence, the days of fixed business rules (e.g., embedded in static workflow or ERP systems) are long gone, as enterprises must be able to reconfigure processes quickly to capitalize on often-ephemeral advantages. To accomplish such agile reconfigurability, enterprise knowledge must flow quickly and efficiently across space, time and organizational units as process changes dictate. Again, enterprise knowledge flow is critical to performance.

To help ground in practice our knowledge-flow model previously described, we illustrate its application through an example of new product development in 21st century.

Figure 2. Notional knowledge-flow vectors (adapted from Nissen, 2002)

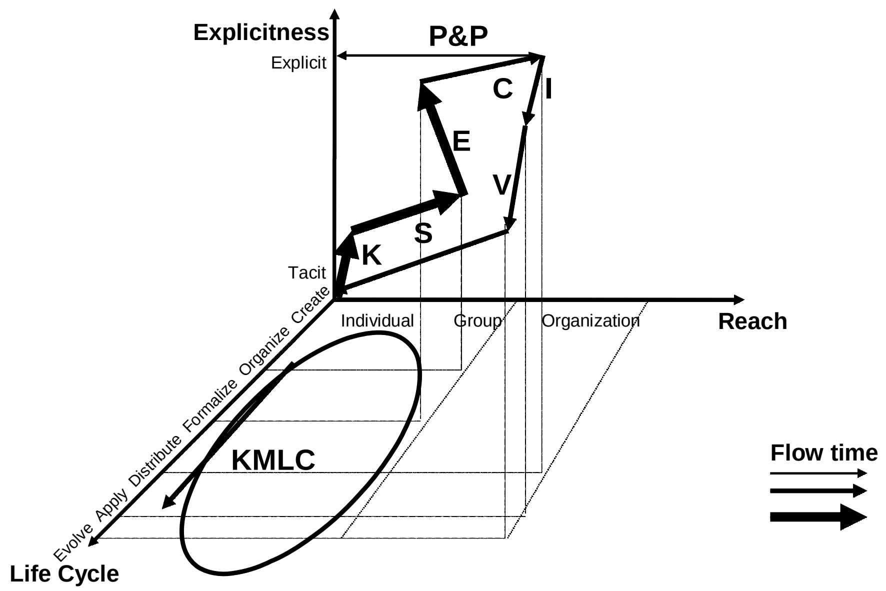


Figure 3. Current enterprise knowledge (adapted from Nissen, 2004)

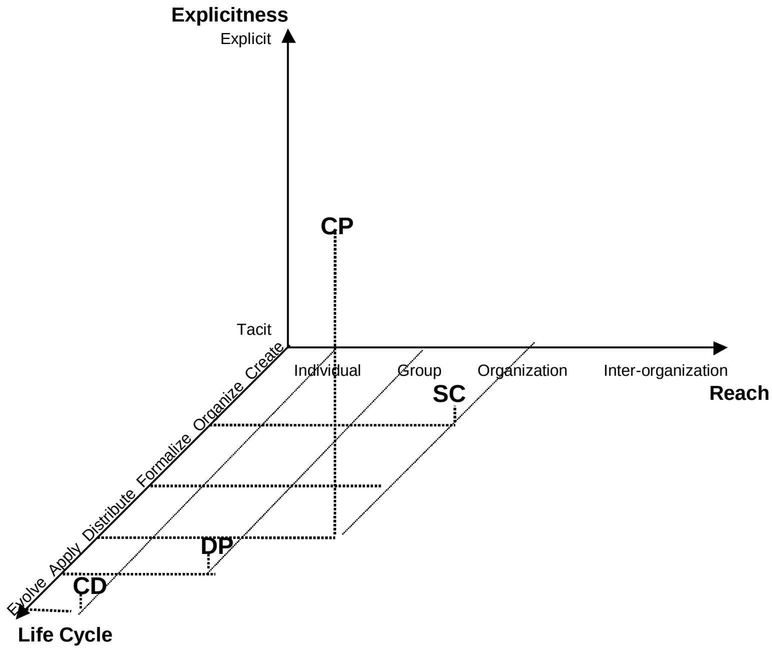
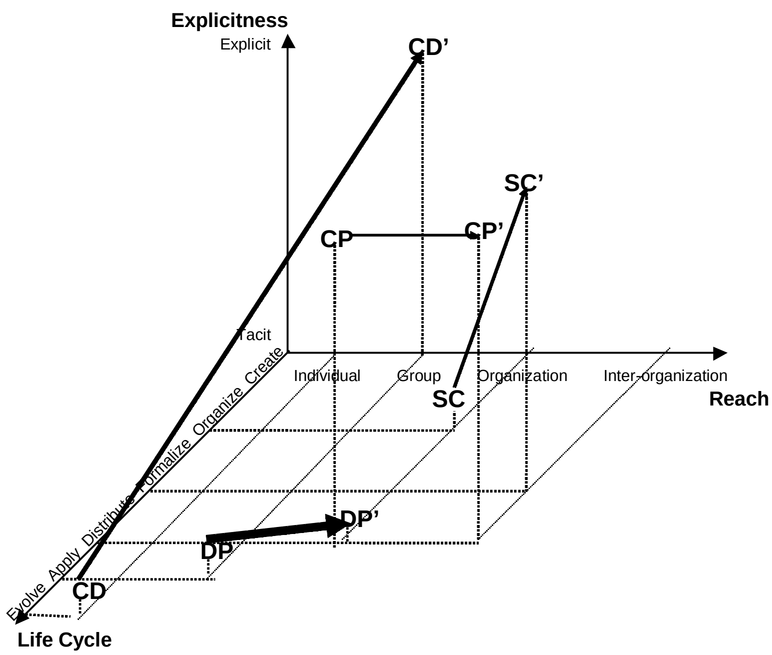


Figure 4. Required knowledge flows (adapted from Nissen, 2004)



We can similarly ground the knowledge-flow model in other application domains (e.g., banking, consulting, manufacturing, telecommunications)—in which some of the specifics vary, but the principles hold constant. However, this example should convey clearly the key ideas. Typically heavy investments in plant and equipment for new product development confine most physical capital to specific geographical locations, and processes based upon such plant and equipment are very resistant to short-term changes. Yet global customer markets and supplier bases demand just such short-term changes. This represents a challenging context in terms of knowledge flow. Further, increasingly global markets and operations cause the demand for specific products to be unpredictable and to shift continually. New product design cycles can barely keep up with such demand shifts, and dynamic supplier networks change with each new product offering.

Here we discuss four areas of knowledge that appear to be critical: 1) customer demands, 2) competitor products, 3) design processes, and 4) supplier capabilities. In Figure 3, we use our multidimensional model to depict and examine such critical knowledge. Beginning with customer demands, the three-dimensional plot point corresponding to such knowledge is labeled "CD" in the figure. Notice it is plotted at the tacit end of the explicitness axis (e.g., future customer wants are generally not explicitly stated), the individual level in terms of the reach dimension (e.g., customer preferences are developed by individuals), and evolve stage of the life cycle (i.e., customer tastes evolve through time).

Even with this single step of classifying and plotting customer-demand knowledge in terms of our multidimensional model, we have identified a critical knowledge element, and can compare and contrast it visually with other key elements of knowledge. In similar fashion, knowledge of competitor products (i.e., labeled "CP" in the figure), internal design processes (i.e., labeled "DP") and supplier capabilities (i.e., labeled "SC") are also plotted in this figure. This set of points provides us with a view of the firm's critical knowledge as it exists today. We explain below how such knowledge is *not flowing* to where (and when) a knowledge-based organization needs it to be, even though it is deemed critical. This motivates the need for inducing knowledge flow.

In Figure 4, we plot these same four points along with their counterparts depicting how knowledge in each area needs to flow in terms of our three model dimensions. For instance, customer-demands knowledge needs to be made explicit, and its reach must extend at least to the level of a marketing group within the firm. The question of where along the life cycle depends upon how agile and proactive the firm wishes to be; as depicted in the figure (i.e., point CD'), ideally the firm would want to *create* customer demand through new-product innovation and brand loyalty. A vector connecting the current and desired knowledge points (i.e., CD-CD') is shown to delineate the corresponding knowledge flow requirement. Here we understand now what knowledge is critical, the state of such knowledge, and how it must flow to enable new product innovation.

Points and vectors corresponding to necessary knowledge flows in terms of competitor products (i.e., CP-CP'), supplier capabilities (i.e., SC-SC'), and design processes (i.e., DP-DP') are classified, identified, and plotted in similar fashion in Figure 4. For instance, unlike (tacit) customer demands, knowledge of competitor products is already explicit and distributed, but the requisite knowledge flow shows such knowledge must also be made explicit *across* organizations (i.e., within our firm as well as the competitors'); this requires business intelligence to identify what competitors are doing in relevant marketspaces.

The knowledge flow corresponding to supplier capabilities is similar (e.g., required to flow beyond the supplier and into our firm), except such capabilities (e.g., in terms of new product development) must also be made explicit (i.e., not left as unarticulated potential) through formalization; this requires competitive sourcing to identify and help develop promising suppliers. Finally, design-process knowledge is required to flow beyond a specific design group, which suggests organization-wide distribution, however, distributing tacit knowledge as such remains a difficult undertaking, as it is often referred to as "sticky" in the product-development context (Szulanski, 1996; von Hippel, 1994).

FUTURE TRENDS

What kinds of future trends are suggested by inducing knowledge flows in the analysis described previously? For one, this particular analysis provides insight into the reality that one size does not fit all in terms of KM and IT. Rather, the key lies in focus on how the requisite knowledge associated with each vector can be induced to flow in situations where it is currently "clumped" or "constricted" (e.g., the marketing flow described previously). For another, to induce knowledge flows, the extant tools and processes employed by most successful firms can be integrated and targeted to focus specifically on the corresponding knowledge. Using knowledge-flow induction as outlined previously, the manager can ask what is preventing this specific knowledge from flowing today, and reconfigure the enterprise's tools and processes to obviate the problem.

The analysis described previously also points to future trends in terms of new IT. For instance, we now know better than to focus IT on information and data, and such knowledge can be used to focus research and development specifically upon classes of technologies that can address knowledge directly. Some such technologies such as those of data mining and artificial intelligence are known relatively well, whereas others such as "knowledge amplifiers" and "knowledge engines" have yet to be invented. Finally, inducing knowledge flows as described previously has implications in terms of research. Despite the progress reflected in this article, our level of ignorance pertaining to knowledge-flow dynamics remains high. Nonetheless, having a multidimensional model and representational method to depict knowledge flows can help guide such research.

CONCLUSION

The knowledge-based organization appears to offer great promise in terms of performance and capability. This article summarizes the current state of the art pertaining to knowledge flow, and provides an overview of a model to help induce the flow of knowledge through an organization. Using a global manufacturing firm as an example to illustrate how the knowledge-flow model provides practical guidance, we identify knowledge elements that are critical to effective performance in an unpredictable, dynamic business environment, and we use the multidimensional model to illustrate how to identify specific knowledge flows required for success. This enables managers to analyze the knowledge flows of their organizations, and provides insights both in terms of future trends and directions for research and development.

REFERENCES

Alavi, M., & Leidner, D.E. (2001). Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly, 25*(1), 107-136.

Baumard, P. (2002). Tacit knowledge in professional firms: The teachings of firms in very puzzling situations. *Journal of Knowledge Management, 6*(2), 135-151.

Davenport, T.H., & Prusak, L. (1998). *Working knowledge: How organizations manage what they know*. Boston, MA: Harvard Business School Press.

Despres, C., & Chauvel, D. (8 March 1999). Mastering information management: Part six – knowledge management. *Financial Times*, 4-6.

Drucker, P.F. (1995). *Managing in a time of great change*. New York, NY: Truman Talley.

Echeverria-Carroll, E.L. (1999). Knowledge flows in innovation networks: a comparative analysis of Japanese and US high-technology firms. *Journal of Knowledge Management, 3*(4), 296-303.

Fang, S-C., Lin, J.L., Hsiao, L.Y.C., Huang, C-M., & Fang, S-R. (2002). The relationship of foreign R&D units in Taiwan and the Taiwanese knowledge-flow system. *Technovation, 22*(6), 371-383.

Foss, N.J., & Pedersen, T. (2002). Transferring knowledge in MNCs: The role of sources of subsidiary knowledge and organizational context. *Journal of International Management, 8*(1), 49–67.

Gartner Group. (1998). Knowledge management scenario. Conference presentation, Stamford, CN, presentation label SYM8KnowMan1098Kharris.

Gore, C., & Gore, E. (1999). Knowledge management: The way forward. *Total Quality Management, 10*(4/5), S554-S560.

Grant, R.M. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal, 17*, 109-122.

Gupta, A.K., & Govindarajan, V. (2000). Knowledge flows within multinational corporations. *Strategic Management Journal, 21*(4), 473–496.

McDermott, R. (1999). Why information technology inspired but cannot deliver knowledge management. *California Management Review, 41*(4), 103-117.

Nissen, M.E. (1999). Knowledge-based knowledge management in the re-engineering domain. *Decision Support Systems, 27*, 47-65.

Nissen, M.E. (2002). An extended model of knowledge-flow dynamics. *Communications of the Association for Information Systems, 8*, 251-266.

Nissen, M.E. (2004). Inducing enterprise knowledge flows. In J. Gupta & S. Sharma (Eds.), *Creating knowledge based organizations*. (Chapter 9, pp.185-202). Hershey, PA: Idea Group Publishing.

Nissen, M.E., Kamel, M.N., & Sengupta, K.C. (2000). Integrated analysis and design of knowledge systems and processes. *Information Resources Management Journal, 13*(1), 24-43.

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science, 5*(1), 14-37.

Schulz, M., & Jobe, L.A. (2001). Codification and tacitness as knowledge management strategies: an empirical exploration. *Journal of High Technology Management Research, 12*(1), 139–165.

Spiegler, I. (2000). Knowledge management: a new idea or a recycled concept? *Communications of the Association for Information Systems, 3*(14), 1-24.

Szulanski, G. (1996). Exploring internal stickiness: impediments to the transfer of best practice within the firm. *Strategic Management Journal, 17*, 27-43.

von Hippel, E. (1994). "Sticky information" and the locus of problem solving: implications for innovation. *Management Science, 40*(4), 429-439.

von Krogh, G., Ichijo, K., & Nonaka, I. (2000). *Enabling knowledge creation: How to unlock the mystery of tacit knowledge and release the power of innovation*. New York, NY: Oxford University Press.

Zhuge H. (2002). Knowledge flow management for distributed team software development. *Knowledge-Based Systems, 15*(8), 465–471.

KEY TERMS

Competitive Advantage: Employing organizational resources in an advantageous manner that cannot be imitated readily by competitors.

Extended Model of Knowledge-flow Dynamics: A four-dimensional model used to classify and visualize flows of knowledge in the organization.

Knowledge-based Organization: An organization that competes on the basis of its differential knowledge.

Knowledge Flow: The movement or propagation of knowledge across space, time, people and organizations.

Knowledge Hierarchy: The hierarchical relationship between knowledge, information and data, with each level building upon the one below.

Knowledge Management: Capitalizing on intellectual assets in the organization.

Knowledge Management Life Cycle: The cycle of activities associated with managing knowledge.

Delivering Web-Based Education

Kathryn A. Marold
Metropolitan State College of Denver, USA

INTRODUCTION

A decade of hindsight allows us to examine the phenomenon of Web-based course delivery and evaluate its successes and failures. When Web-delivered courses mushroomed from campuses in the 1990s, they were embraced by students, faculty, and administrators alike. The prospect of "electronic tutelage" (Marold, 2002), which allowed students through Web interface to take college courses for credit any time, any place (ATAP), was immediately popular with students. The interruptions of job and schedule changes, relocation, childbirth, failed transportation to campus, and so forth no longer necessitated an interruption in progress toward a degree. Likewise, faculty saw online teaching as an opportunity to disseminate knowledge and assess student progress according to their personal preferences, and to communicate personally with their students, albeit virtually. Administrators saw the revenue without physical classroom allocations as an immediate cash cow. In the beginning, there was satisfaction all around. Although this state of affairs was not necessarily universal, generally it could be concluded that Web-based education was a very good thing.

The Evolution Of Web-Based Course Delivery

Web-based education is a variation of distance learning: the content (college courses from an accredited North American institution, for purposes of this chapter) is delivered via the World Wide Web. The Web course content covers a quarter or semester of curriculum that the student must complete and prove a level of mastery within a given timeline. For the most part, Web-based courses use existing college curriculum and timelines. Web-based education is currently the most popular form of distance education. As educators are inclined to do, it was not long before they wanted to stand back and evaluate what they had created and determine the success of Web-delivered courses as a form of distance education. With McLuhanesque procedures, a glance in the "rear view mirror" was in order (McLuhan, 1964.) The results of many measures of success show that for *some* of the students, *some* of the time, in *some* situations, Web-based education is quite successful. Likewise, for many persons in many situations and in many phases of their formal education, Web-delivered education is *not* the answer.

BACKGROUND

The advent of the World Wide Web in the early 1990s promised a more effective, user-friendly form of Internet distance education. The graphical hypertext and, indeed, the hypermedia nature of the Web could enhance course delivery. Almost immediately, Web courses began to flourish. A new mode of delivery was firmly established.

Web-Based Education's Successes and Failures

Numerous publications have exposed problems associated with the Web-based form of distance education. The population taking the courses was sometimes the problem (Haga, 2001). The attrition and failure rate of Web-delivered courses was higher than the classroom arena (Terry, 2001). The content of the course could be problematic (Haga, 2002). The credibility of course credit achieved online was sometimes suspect (Moreno, 2000). The level of courses offered online was sometimes suspect (Marold, 2003). Research findings suggest the following conclusions concerning Web-based education (see Table 1).

There are almost as many reports of success with Web-based education as there are reports of failures. Students who are successful with Web courses tend to take more of them, sometimes as many as 90 hours of the 120 hours required for a bachelor's degree. There are now entire degrees offered online. The earliest research on Web-based education reported no statistical difference in final grades between Web-based groups and classroom groups (Mawhinney, 1998; Schulman, 1999).

The conclusion that Web-delivered education, like all distance education, is only appropriate for some students cannot be denied. It is obvious that Web courses are not going to go away. It is also undeniable that regardless of how enrollments in Web-based courses are screened, there will be students who enroll in Web courses that should not be in them. It has been shown time and again that some Web students enroll for all of the wrong reasons

Table 1. Successes and failures of Web-based education

Positive	Negative
Survey level courses are the most successful.	The attrition and failure rates for upper level, analytical Web-based courses often reach 50%.
Courses at the 1-2 level of Bloom's taxonomy (Bloom, 1956) of learning immersion are more successful than those at the 3-5 level.	Students at the B and C level (the vast majority of students in any institution) are the most at risk for not completing and not passing Web-delivered courses.
Students with GPA of 3.5 or better are the most successful at completing and excelling in Web-delivered courses.	Web-delivered courses are a disaster for the passive learner without time management and independent study skills.
Graduate level Web-delivered courses are more successful.	Both students and faculty alike indicate time spent on an Internet delivered course is more than it would be on its classroom equivalent.
Courses delivered via a 3rd party distributor or a portal (such as *WebCt* or *Blackboard*) are more successful than self-hosted Web courses	
Internet students generally do better than their classroom counterparts on exams. Internet students generally do worse on projects than their counterparts on assigned projects.	
Analytical and problem solving courses are least successful.	
Web-delivered courses are a godsend for the highly motivated, independent learner.	
Final grades on Web-based education courses generally do not differ significantly from those earned in the classroom. Web-based courses are here to stay. They are an accepted, credible method of course delivery	

(Haga, Marold, & Helms, 2001). It is equally obvious that Web courses fill an enormous need for many students and are, therefore, very successful in many instances.

While the students who are at risk for failure in Web-based courses that are analytical and require problem solving are those students who are generally classified as mid-level achievers, taking prerequisite courses online seems to alleviate the risk slightly (Pence, 2003). The student group at the greatest risk is the mid-level achieving group, which in a normal distribution is the largest number of students in the class (Marold & Haga, 2004). Pence suggests some alleviating factors, such as taking the prerequisite course online from the same institution. This suggests that as students become more accustomed to the requirements and idiosyncrasies of online learning, the risk decreases. Experience makes a difference. In addition, the majority of students taking online courses indicate that they would take another online course, even though they perceive them to be more work than an equivalent classroom course. Despite attrition and failure rates that sometimes reach 50%, Web-based education is clearly a student favorite.

Tables 2 and 3 show some of the research results of a decade of Web-based education.

In the above research of two separate Web-based required computer information systems junior level courses in the same department of a large urban state school, student tests were higher in the Internet version, but their project scores (application of learning) were lower.

In Table 2, there were three different courses at freshman, sophomore, and junior levels, offered online as well as in the classroom, from the same department of a large urban institution. All three courses were survey-type

Table 2. Comparing Internet scores with classroom scores (Haga, 2001)

Telecommunications	**Internet**	**Classroom**
Projects average	60.8	83.5
Exams average	64.2	62.7
Visual Basic Programming		
Projects average	69	73.3
Exams average	72	71.3

Table 3. Comparing Internet scores with classroom scores (Marold, Larsen, & Moreno, 2002)

Introduction to Computers 1010	**Internet**	**Classroom**
Projects average	92.4	92.4
Exams average	78.2	73.5
Computer Applications for Business 2010	**Internet**	**Classroom**
Projects average	90.3	91.8
Exams average	77.5	70.01
Micro-based Software 3270	**Internet**	**Classroom**
Projects average	87.5	93.8
Exams average	77.7	67.8

** Revised classroom exams average for 2010*

Table 4. Successful Web-course characteristics

• More is not better. Simplicity, organization, and clarity are paramount.
• Frequent communication is important, both synchronous (chat, messaging) and asynchronous (e-mail, forum, discussion boards).
• Smaller modules and units work better.
• More frequent, less intricate assignments are better than complex comprehensive assignments due at the end of the course.
• Student uploaded projects and assignments work better than hard copy submissions.
• Individualized assignments are more likely to be submitted.
• The more choices in assignments and projects, the better.
• Proctored exams and secure quizzes are essential for course acceptance in the community and the workplace.
• Courses shorter in duration are more successful than ones that span semesters or quarters.
• Electronic grade books and regular feedback to student assignments submitted are essential.
• Samples of student work communicate expectations for assignments.
• Profiles or student Web pages for class members build community.
• Web links to helpful course content and enrichment material are appreciated, although not used extensively.
• More than one route to the same location in a course individualizes the course.
• Calendars of due dates and course announcements provide quick reference.

courses with a heavy skills component. This research was done earlier than the study shown in Table 2. The test scores were also higher for the Internet sections, and the projects lower.

Although Web-based grades differ from classroom grades on individual projects and exams, generally over a decade of looking back, we see that final grades on courses completed on the Internet and in the traditional classroom do not differ significantly (Haga, 2002; Marold, 2002; Moreno, 2000; Presby, 2001).

DESIGNING AND DELIVERING WEB-BASED EDUCATION

It is heartening to know that the success of a Web-delivered course is not directly related to the faculty's skill in instructional design and Web publication (Marold, 2002). Templates and Web design software suffice for course design. Simple courses are just as successful as complex ones. Studies that count visits to various pages in Web courses show that many of the available pages are infrequently (and sometimes never) used by a majority of online students (Haga, 2001). The same is true for extensive instructions on how to take an online course, or lists of Web links available for extra information supplementing the course material. Factors that contribute to successful Web-delivered course are shown in Table 4. These are general conclusions based upon a decade of studies of Web-based education.

For every individual characteristic listed here, one undoubtedly can find research contradicting it. From their inception, Web delivered courses have been examined and reexamined. However, in general, these are factors that have been found to make a positive difference for faculty and designers of Web-based courses.

FUTURE TRENDS

As Web courses become more common and educators become more proficient at designing and deploying them, undoubtedly their success rate will improve. As students become more accustomed to taking Web courses, they will become more proficient at completing them. As McLuhan (1964) noted, it takes a period of adjusting to a new medium before we are entirely comfortable with it. Once Web courses become part of the "every day" of our existence, Web-based education may take its legitimate place alongside other modes of delivery.

CONCLUSION

In conclusion, Web-based education is working—for some. Entire degrees are obtainable online; students, faculty, administrators, and organizations accept credits earned online. For better or for worse, Web-based education is firmly entrenched in all areas of higher education. While Web-based education has not yet achieved the same level of success as classroom delivered instruction, it is part of most institutions' programs. The fact that Web-based education is truly working in a cost-effective manner for some of the student population some of the time, assures its continuance. Like the printing press of the late 1400s for mass distribution of knowledge, Web-based education provides mass distribution of knowledge in a new, effective way.

REFERENCES

Bloom, B.S. et al. (1956). *Taxonomy of educational objectives. Handbook I.: The cognitive domain*. New York: David McKay.

Haga, W., & Marold, K. (2002). Is the computer the medium *and* the message? A comparison of VB programming performance in three delivery modes. *International Business and Economic Research Journal, 1*(7), 97-104.

Haga, W., Marold, K., & Helms, S. (2001). Round 2 of online learning: Are Web courses *really* working? *Proceedings of the Twenty-Ninth Annual Conference of IBSCA*. Providence, RI.

Marold, K. (2002). The twenty-first century learning model: Electronic tutelage realized. *Journal of Information Technology Education, (1)*2, 113-123.

Marold, K., & Haga, W. (2003). The emerging profile of the on-line learner: Relating course performance with pretests, GPA, and other measures of achievement. *Proceedings of the Information Resources Management Association*. Philadelphia, PA.

Marold, K., & Haga, W. (2004). E-learners at risk: Midlevel achievers and online courses. In A. Appicello (Ed.), *Instructional Technologies: Cognitive Aspects of Online Programs*. Hershey, PA: Idea Group Publishing.

Marold, K., Larsen, G., & Moreno, A. (2002). Web-based learning: Is it working? In M. Khosrowpour (Ed.), *Web-based instructional technologies*. Hershey, PA: Idea Group Publishing.

Mawhinney, C. et al. (1998, October). Issues in putting the business curriculum online. *Proceedings of the Western Decision Sciences Institute*. Puerto Vallarta, MX.

McLuhan, M. (1964). *Understanding media*. New York: McGraw Hill.

Moreno, A., Larsen, G., & Marold, K. (2000). The credibility of online learning: A statistical analysis of IS course delivery at three levels. *Proceedings of the Western Decision Sciences Institute*. Maui, HI.

Pence, N.K. et al. (2003). An exploration of the impact of online delivery in prerequisite courses on CIS majors' course sequence. *Proceedings of the International Business and Economic Research Conference*. Las Vegas, NV.

Presby, L. (2001). Increasing productivity in course delivery. *T.H.E. Journal, 28*(7), 52-58.

Schulman, A., & Sims, R.L. (1999). Learning in an online format versus an in-class format: An experimental study. *T.H.E. Journal, 26*(11), 54-56.

Terry, N. (2001). Assessing enrollment and attrition rates for the online MBA. *T.H.E. Journal, 28*(7), 64-68.

KEY TERMS

ATAP: Any time, any place learning. A basic characteristic of Web-based education courses in that they are available to the student on a 24-7 basis.

Bloom's Taxonomy of Learning: A scale that represents an organization of learning levels (five levels) that are characterized by the student's immersion into the theory and application of principles of a course content.

Chat Sessions: Live discussions online with a variable number of participants in a Web-based class. They can be formal and led by the instructor, or they can be leaderless informal conversations. Chat sessions are synchronous.

Electronic Gradebooks: Maintaining a record of a student's progress in a Web-based education class by posting grades on the course Web pages. General gradebooks show all enrollees; personalized gradebooks can only be viewed by the individual student.

Electronic Tutelage: Learning of new complex concepts (sometimes called *scientific* concepts), not with the intervention of a physical tutor, but via electronically delivered materials; the basic theory behind Web-based education.

Synchronous and Asynchronous Communication: Synchronous communication via the Web is immediate communication, such as in chat or instant messaging. Asynchronous communication is delayed communication via the Web, such as threaded discussions, forums, or e-mail messages, where each participant does not have to be online at the same time.

Web-Based Education: A variation of distance learning; the content (college courses from an accredited North American institution, for purposes of this chapter) is delivered via the World Wide Web. The Web course content covers a quarter or semester of curriculum that the student must complete within a given timeline in return for course credit.

Web Links: Hyperlinks (hot links) to other Web sites that are embedded in the active pages of a Web-based education course.

Web Profiles: Short biographies of students enrolled in a Web-based education course. The profiles may contain Web links to students' own home pages and digitized photos of the students.

Democratic E-Governance

Ari-Veikko Anttiroiko
University of Tampere, Finland

INTRODUCTION

The changing role of the state and a managerialist view of the operations of public-sector organizations gave rise to the idea of new public governance. Gradually, more citizen-centered views of governance also emerged, reflecting a need to strengthen the role of citizens and communities in governance processes at different institutional levels. This development, especially since the mid-1990s, has been affected by new technologies, leading to a kind of coevolution of institutional arrangements and technological solutions that have paved the way for a better understanding of the potentials of democratic e-governance.

BACKGROUND

Discussion about governance has acquired new dimensions since the early 1990s due to the gradual erosion of the hierarchical, mainly state-centric, bases of political power. Among the core topics has been the decline of the nation state and the rise of the regions and local governments as the new key players in coping with external challenges and imposing a political will within territorial communities. Also, after the Second World War, and the 1980s in particular, international organizations and regional institutions started to gain more power in the international arena (Pierre, 2000, p. 1). Another widely discussed aspect of public governance relates to the functioning and ways of working of public-sector organizations. In fact, this theme has dominated much of the recent discourse around governance. In all, the entire institutional landscape and the overall understanding of the role of public-sector organizations has gradually changed practically everywhere in the world, thus fueling the discussion about governance. One important governance agenda-setter was the Organisation for Economic Cooperation and Development (OECD) Public Management Committee (PUMA), which carried out work on this topic during the first half of the 1990s, and as a synthesis, published a policy paper entitled *Governance in Transition* in 1995 (OECD, 1995).

OECD's policy lines have been more or less neoliberal, which means that governance issues were discussed and still are, to a large extent, within the framework of New Public Management (NPM). In essence, its message is that the approach to the management of public organizations and services needs to be based on managerialism and market-based coordination (Walsh, 1995). After a gradual widening of these NPM perspectives, some people have referred to this area as new governance or new public governance. Yet, it is important to keep in mind that new governance attempts to widen the horizons inherited from the original market-oriented NPM doctrine.

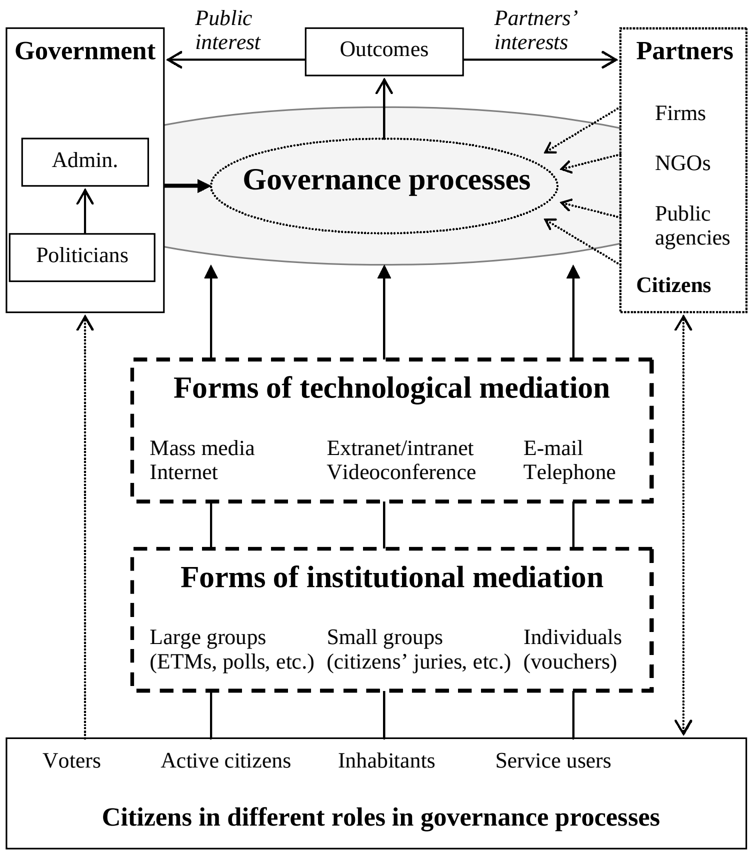
As presented above, contemporary understanding and use of the concept of governance have their roots in the changing role of the state and in a managerialist view of the operations of public organizations. These two discourses have been challenged by another approach that could be called democratic governance. It emphasizes the interactions between citizens, political representatives, and administrative machinery, providing a special view of citizens' opportunities to influence and participate in governance processes.

DEFINITION OF GOVERNANCE

One of the reasons behind the revival of the concept of governance was the need to distinguish between the traditional, institutionally oriented conception of "government" and more dynamic and network-based ways of thinking and working in policy processes. *Government* refers to the institutions and agents that perform the governmental functions, i.e., to formal institutions of the state or those of decentralized territorial governments and their ability to make decisions and take care of their implementation, whereas *governance* as a process concept refers to the new modes and manners of governing within policy networks and partnership-based relations (Stoker, 1998, p. 17; Jessop, 1998, pp. 30–31; see also Kooiman, 1993). In spite of the common root of these two terms, they should not be used as synonyms. The role of government in public governance may vary considerably, with two basic models being state-centric and society-centric models of governance (Pierre & Peters, 2000, p. 29).

The way the concept of governance is used here can be specified as "public governance," which aims to pursue collective interest in the context of intersectoral stakeholder relations. In this sense, governance refers to the coordination and the use of various forms of formal or

Figure 1. Aspects of democratic e-governance (cf. Anttiroiko, 2004, p. 40).



informal types of interaction and institutional arrangements in the policy-making, development, and service processes to pursue collective interest (Anttiroiko, 2004, p. 28).

One essential aspect of recent discussions about governance is to identify how to maintain the "steering" role of political–administrative institutions despite the internal and external challenges to them. The other element causing one of the core dilemmas of governance is the tension between public and private interests, which, in turn, pose a challenge to the democratic control of public governance.

E-TRANSFORMATION IN DEMOCRATIC GOVERNANCE

E-Transformation in governance processes refers to the fact that the information society development profoundly affects the relationships of different actors, forms and channels of communication and interaction, and the entire fabric of network and partnership relations. The transformational aspect of governance leads us back to the issue of the information society, which forms a background for understanding a wider transformative nature of information and communication technologies (ICTs).

The introduction of ICTs in the public sector in the 1960s in most of the advanced countries started to reshape their data-processing activities, such as record keeping and financial administration. Electronic systems started to replace old manual systems. This picture started to change dramatically in the 1990s. At the core of this revolution was the Internet (Seneviratne, 1999, pp. 44–45). Along with this new wave, the democratic potential of ICTs became a global issue.

Since the 1990s, a need for reconstruction of technology along more democratic lines has become apparent. New ICTs have the potential to restructure government and to strengthen democracy, and to create a closer relationship between public administration and citizens, in particular (Pardo, 2002, p. 95; see also Becker & Slaton, 2000). It has even been said that new ICTs applied by government contribute to the emergence of a different type of governance, i.e., more "direct" government, as concluded by Pardo (2002, pp. 90–91).

Figure 2. Continuum of citizen influence (applied from Bishop & Davis, 2002, pp. 20–21).

Democratic e-governance combines three conceptual elements: governance as a core activity area and manner of governing, democracy as a kind of basic principle applied in governance, and ICTs as a tool. As to a formal definition, we may conclude that democratic e-governance is a technologically mediated interaction in transparent policy-making, development, and service processes in which political institutions can exercise effective democratic control and, more importantly, in which citizens have a chance to participate and effectively influence relevant issues through various institutionally organized and legitimate modes of participation (Anttiroiko, 2004, pp. 40–41). At a practical level, democratic e-governance requires both institutional and technological mediation of civic and community interests in formal governance processes, as illustrated in Figure 1.

METHODS OF DEMOCRATIC E-GOVERNANCE

There is nothing inherently "democratic" in governance. It can be, and historically has been, performed in various ways that cannot be called democratic in the true sense of the word. In the history of the institution of community governance, the early modern times represent the era of elite control that since the 19th century began to transform into a conventional democratic mold having expression in the form of civic rights and a representative system of government. This was followed by the rise of professionalism and managerialism in the 20th century. The new phase emerging has, in some descriptions, been called the era of citizen governance (Box, 1998, p. 30; see also Hirst, 2000). Democratic governance requires that political institutions be capable of steering governance processes and affecting their outcomes, that citizens be given a chance to influence and participate in these processes in principle whenever they see fit, that governance processes be made transparent, and that key actors be held accountable for their actions to political institutions and ultimately to society as a whole.

Institutional mediation tools of democratic governance are needed to facilitate civic involvement. A fundamental aim is to supplement the representative system that is considered by many to be too hierarchical, inflexible, and distant from the point of view of ordinary citizens. Gross (2002, p. 250) summarized the basic requirements of e-democracy in the following way: citizens need to be able to access information, to discuss political issues, and to make decisions by voting electronically.

Conventional ways of political influence and participation in modern democracies include voting and campaigning in elections and being an active member of a political party or pressure group. Another category includes memberships of advisory committees or other bodies with a stake in policy processes and also various forms of client or customer involvement in implementa-

Table 1. Methods and forms of democratic e-governance.

1. Facilitating information processes
• *Presenting, disseminating, and sharing information (Web sites, e-BBS, etc.)*
• *Collecting and processing data (e.g., database management tools and e-document management)*
• *Facilitating communicative or two-way information processes (e-mails and e-feedback systems)*
2. Supporting communication and negotiation
• *Facilitating discussion and interaction (electronic discussion forums, e-mails, etc.)*
• *Generating understanding and awareness (idea-generating tools, simulations, etc.)*
• *Facilitating citizen–expert interaction (e.g., consensus conferences)*
3. Citizen consultation and involvement in preparation and planning
• *Consultative referendum*
• *Moderated deliberative polling (scientific deliberative polling, electronic town meeting, etc.)*
• *Other forms of citizen consultation (e-citizens' juries, standing e-panels, etc.)*
• *Participatory planning*
• *Modeling decisions and advising on possible consequences (expert systems, decision support systems, etc.)*
4. Community-based deliberation and participation
• *Virtual communities and cyberassociations*
• *Community networks and local associations*
• *Local and neighborhood governments*
5. Political transactions and decision making
• *Making proposals and initiatives (e-initiatives, e-petitions, etc.)*
• *Participating and voting in elections (e-electioneering, e-voting, etc.)*
• *Making decisions (initiative and referendum processes, including e-referendums)*
6. Implementation and service processes
• *Various forms of user democracy (e-feedback systems, e-vouchers, etc.)*

tion of public policies. More direct forms include voting in referendums and participating in the consultative or advisory bodies set up on an ad hoc basis. Last, there are various types of community group actions as well as political demonstrations that aim at changing public policy, and even various forms of civil disobedience (Birch, 1996, p. 81). One way to systematize these forms on the basis of the degree of citizen influence is presented in Figure 2.

Figure 2 systematizes citizen influence within a continuum: the two extreme ends being a chance to obtain information and a direct political decision making. The idea behind this is originally an old concept of eight rungs on a ladder of citizen participation that proceeds from nonparticipation through a certain degree of tokenism to the exercise of direct citizen control (Bishop & Davis, 2002, p. 17).

Methods of democratic e-governance are based on the functions they serve in policy processes. On the basis of the functions of institutional and technological mediation tools, the methods applicable in e-governance can be presented as in Table 1 (Anttiroiko, 2004; see also Grönlund, 2002; Gross, 2002; 6, 2001; Becker & Slaton, 2000; Barber, 1984).

Since the 1970s, small groups of academics, community organizers, activists, government officials, and media professionals have been experimenting with electronic media and ICTs. Since then, new democratic practices have emerged, and many experiments have been conducted in different parts of the world. Almost all of these have shown how interested ordinary people have been in the opportunity to participate, and how much they have appreciated being consulted and included in these processes (Becker & Slaton, 2000, p. 95; see also Tsagarousianou et al., 1998).

FUTURE TRENDS

Democratization of governance is conditioned by such pervasive changes as globalization, technological development, new forms of social organization, and increased individualism. It may be that a hybrid model of democratic governance is in the making, in which the new technology employed is evolving along with the societal and governmental structures. Becker and Slaton (2000, p. 210) have claimed that despite some setbacks, genuine democracy will increase in degree and scope. This development is supported by such factors as greater influence of social movements, new methods of direct democracy, wider use of consensus-building mechanisms, and new forms of e-enabled democratic political organization. New technological mediation tools and the Internet, in particular, may prove vital in rethinking conceptions of democratic governance. However, only modest democratic gains can be achieved through electronic means, unless a radical redesign of democratic institutions is accomplished, which is not foreseeable in the immediate future (Anttiroiko, 2003).

CONCLUSION

A new discourse about democratization of public governance reflects a gradual transition from both the state-centric model of governance and managerial and market-based views of new public management to the politically oriented coalition-building and stakeholder-involving new governance model that is rooted in the values of authentic democracy (cf. Reddel, 2002; Becker & Slaton, 2000; Barber, 1984). In this sense, democratic e-governance is a technologically mediated interaction in governance processes in which special attention is paid to citizens' chances to participate and influence relevant public policies. New forms of democratic e-governance are expected to pave the way for a greater citizen self-governance that matches the requirements of a changing society.

REFERENCES

6, P. (2001). E-governance. Do digital aids make a difference in policy making? In J. E. J. Prins (Ed.), *Designing e-government. On the crossroads of technological innovation and institutional change* (pp. 7-27). The Hague: Kluwer Law International.

Anttiroiko, A.-V. (2003). Building strong e-democracy. The role of technology in developing democracy for the information age. *Communications of the ACM, 46*(9), 121-128.

Anttiroiko, A.-V. (2004). Introduction to democratic e-governance. In M. Malkia, A.-V. Anttiroiko, & R. Savolainen (Eds.), *e-Transformation in* governance (pp. 22-49). Hershey, PA: Idea Group Publishing.

Barber, B. (1984). *Strong democracy: Participatory politics for a new age*. Berkeley: University of California Press.

Becker, T., & Slaton, C. D. (2000). *The future of teledemocracy*. Westport, CT: Praeger.

Birch, A. H. (1996). *The concepts and theories of modern democracy*. First published 1993. Reprinted 1995 and 1996. London; New York: Routledge.

Bishop, P., & Davis, G. (2002). Mapping public participation in policy choices. *Australian Journal of Public Administration, 61*(1), 14-29.

Box, R. C. (1998). *Citizen governance. Leading American communities into the 21st century*. Thousand Oaks, CA: Sage.

Gronlund, A. (Ed.). (2002). *Electronic government: Design, applications and management*. Hershey, PA: Idea Group Publishing.

Gross, T. (2002). e-Democracy and community networks: Political visions, technological opportunities and social reality. In A. Gronlund (Ed.), *Electronic government: Design, applications and management* (pp. 249-266). Hershey, PA: Idea Group Publishing.

Hirst, P. (2000). Democracy and governance. In J. Pierre (Ed.), *Debating governance. Authority, steering, and democracy* (pp. 13-35). Oxford: Oxford University Press.

Jessop, B. (1998). The rise of governance and the risks of failure: The case of economic development. *International Social Science Journal—Governance, 155, L*(1, March), 29-45.

Kooiman, J. (Ed.). (1993). *Modern governance. New government–society interactions*. Thousand Oaks, CA: Sage.

OECD. (1995). *Governance in transition. Public management reforms in OECD countries*. Organisation for Economic Cooperation and Development. Paris: OECD.

Pardo, M. d. C. (2002). New information and management technologies for the 21st century public administration. *Workshop Report of Twenty-Fifth International Congress of Administrative Sciences: Governance and public administration in the 21st century: New trends and new techniques* (pp. 83-99), Athens, July 2001. Brussels: IIAS, Proceedings, 2002.

Pierre, J. (2000). Introduction: Understanding governance. In J. Pierre (Ed.), *Debating governance. Authority, steering, and democracy* (pp. 1-10). Oxford: Oxford University Press.

Pierre, J., & Peters, B. G. (2000). *Governance, politics and the state*. Houndmills: Macmillan.

Reddel, T. (2002). Beyond participation, hierarchies, management and markets: "New" governance and place policies. *Australian Journal of Public Administration, 61*(1), 50-63.

Seneviratne, S. J. (1999). Information technology and organizational change in the public sector. In G. D. Garson (Ed.), *Information technology and computer applications in public administration: Issues and trends* (pp. 41-61). Hershey, PA: Idea Group Publishing.

Stoker, G. (1998). Governance as theory: Five propositions. *International Social Science Journal—Governance, 155, L*(1, March), 17-28.

Tsagarousianou, R., Tambini, D., & Bryan, C. (Eds.). (1998). *Cyberdemocracy: Technology, cities and civic networks*. London: New York: Routledge.

Walsh, K. (1995). *Public services and market mechanisms. Competition, contracting and the new public management*. Houndmills: Macmillan.

KEY TERMS

E-Democracy: Electronic democracy (e-democracy) as a tool-oriented conception of democracy refers to new democratic practices in which ICTs and innovative institutional arrangements are utilized (cf. teledemocracy).

E-Governance: Most commonly, the concept of governance is associated with "public governance," which refers to coordination, interaction, and institutional arrangements that are needed to pursue collective interest in policy-making, development, and service processes in the context of intersectoral stakeholder relations. Electronic governance (e-governance) is technologically mediated communication, coordination, and interaction in governance processes.

E-Government: Electronic government (e-government) refers to those aspects of government in which ICTs are or can be utilized, the basic functions being to increase efficiency in administrative processes (e-administration), to guarantee easy access to information for all, to provide quality e-services, and to enhance democracy with the help of new technological mediation tools (e-democracy).

Multilevel Governance: A governance system comprising various institutional levels: local, regional, national, macroregional, and international or global levels. For example, the European Union forms a multilevel governance system with municipalities, regions, member states, and the EU-level institutions, as does any federal system of government, like those in the United States or Germany.

Network: A new form of organization. Networks are loose sets of actors who work together in order to promote their interests within a common operational framework that is held together by some shared interests, reciprocity, and trust. In their most characteristic form, networks are flexible ways of organizing activities that require the competences of several independent actors.

New Public Management (NPM): Neoliberally oriented public management doctrine based on a market-oriented view stating that, instead of direct political control and hierarchies, public organizations should rely on indirect control—that is, market-based coordination—in the interaction between public organizations and their environments. It emphasizes the efficiency and effectiveness of public organizations, customer focus in provision of public services, and market-based conditioning frameworks, such as privatization, competition, and contracting out.

Teledemocracy: A normative theory on democracy that is dedicated to greater, improved, and more direct citizen participation and influence in all aspects of government. It is based on the notion of transformational politics that emphasizes the necessity to fine-tune a democratic system to meet the requirements of an increasingly complex information society. This is evident in how it favors the utilization of new ICTs in democratic processes, in such forms as electronic town meeting, scientific deliberative polling, and new democratic use of the Internet. The most prominent academic developer and advocate of teledemocracy is Professor Ted Becker, Auburn University, Alabama.

D

Departure of the Expert Systems Project Champion

Janice C. Sipior
Villanova University, USA

INTRODUCTION

This article discusses the expert system (ES) project champion by examining the experiences of Ciba-Geigy Corporation with an ES project, impeded by the departure of the project champion. The OpBright Expert System, developed to support the identification of appropriate optical brightener products by sales representatives, was intended to provide a competitive advantage through superior customer service. With the promotion and transfer of the vital force committed to the project's success, the ES encountered a stalemate. The difficulties in maintaining momentum for the ES without a project champion are discussed. Finally, suggestions are presented to guide organizations away from the same fate.

BACKGROUND

The role of project champion has been recognized as vital to successful project development since the time of Schon's (1963) seminal work. A project champion for information systems is defined as "a key *individual*, whose *personal* efforts in support of the system are critical to its successful adoption" (Curley & Gremillion, 1983, p. 206). The project champion, for ES projects in particular, is recognized as critical to the successful application of this technology (Hayes-Roth & Jacobstein, 1994; Sipior, 2000: Wong, 1996). Champions of ES projects differ from those of other projects due to the necessity to identify, document, and distribute knowledge and expertise, facilitating knowledge-sharing. The characteristics of project champions are discussed in the next section.

Formal Position

A project champion is frequently an executive from the area of application (Willcocks & Sykes, 2000), but may come from external organizations, such as a consultants or vendors (Thomas, 1999). Champions may be managers (Beath, 1991); or hold other formal positions (Mayhew, 1999; Pinto & Slevin, 1989; Thomas, 1999). Surprisingly, champions rarely come from formal IT functions (Martinsons, 1993; Willcocks & Sykes, 2000) and may even view IT managers as too conservative, adversaries to technological innovations, and even inept (Beath & Ives, 1988). Rather than being assigned to the role, interest and personal conviction to a project compel the champion to emerge (Pinto & Slevin, 1989; Schon, 1963). Formally appointing an individual could actually lead to his demise (Howell & Higgins, 1990). Once convinced, the champion exhibits an entrepreneurial spirit (Bolton & Thompson, 2000; Pinto & Slevin, 1989; Schon, 1963).

Leadership Qualities

The champion tends to go well beyond job responsibilities, and may even go against management directives (Beath, 1991; Curley & Gremillion, 1983). Champions are characterized as more than ordinary leaders. They exhibit transformational leadership behaviors (Howell & Higgins, 1990). Such leadership is particularly valuable for implementing systems intended to bring about organizational change (Beath, 1991; Landers, 1999), such as redefining responsibilities, realigning lines of authority, shifting power centers, and adjusting reward schemes. As knowledge repositories, ES certainly has the potential to invoke change of this nature.

Base of Power

Some level of power is held by champions (Mayhew, 1999; Pinto & Slevin, 1989), attributable to formal position or personal relationships. Diminished power can result in project failure (Scott & Vessey, 2002). Champions are perceived as influential or prestigious by organizational members (Curley & Gremillion, 1983). This perception by others may be the result of a planned influence strategy to attract followers (Schon, 1963). Such influence strategies include impression building, rational justification, assertion, or persuasive communication (Howell & Higgins, 1990). Although activities of champions may be intentionally fostered, their influence tactics are not always regarded in a positive light (Beath, 1991).

Visionary Perspective for Change

The champion is willing to put himself on the line, risking his reputation, to complete the project. The champion serves as a visionary and directs his energies to bring about change to achieve that vision (Landers, 1999; Willcocks & Sykes, 2000). Primary among the influence strategies is persuasive communication (Sumner, 2000). The vision must be clearly communicated in order that others understand and support the vision (Kotter, 1995). An unrealistic or misunderstood vision can result in failure well after the project is underway (Royer, 2003).

A CASE STUDY OF THE PROJECT CHAMPION AT CIBA-GEIGY

Ciba-Geigy Corporation, an international chemical manufacturing firm headquartered in Basel, Switzerland, continually strives to gain market position by fostering their progressive image. Ciba-Geigy emphasizes customer service, especially important to the Dyestuffs and Chemicals Division. This division produces over 2,000 products including fabric dyes, optical brighteners, and industrial chemicals, representing approximately 20% of corporate sales. The OpBright Expert System, developed to support the identification of appropriate optical brightener products, was championed by the vice president (VP) of the division as providing benefits realizable from managing internal knowledge. The VP was convinced, as was found in previous research, that effective knowledge management can impact business performance (Alavi & Leidner, 1999; Hansen, Nohria, & Tierney, 1999; Zack, 1999). Included among the anticipated benefits of OpBright are gaining competitive advantage, faster response to customers, consistent quality customer service, training new salespeople, and managing product expertise, as discussed in the following sections.

Gain Competitive Advantage

As a leading dyestuffs and chemicals producer, Ciba-Geigy recognizes the value of IT as an important means for gaining competitive advantage. Continually striving to gain market position by fostering their progressive image, Ciba-Geigy has emphasized the need to utilize IT in direct marketing. The use of laptops by the sales force, championed by the VP, provides a highly visible means for projecting this image as well as enhancing sales force performance.

Respond More Quickly to Customers

Improved communication between the sales force and the division office, in terms of such factors as speed, receipt and response, and content completeness, was realized through the use of laptops by the sales force. For example, access to the online order processing inventory and sales service system enables sales representatives to complete a sales transaction more quickly, increasing employee productivity and providing more responsive and effective customer service. This taste of success led the VP to seek further improvement. In informal meetings with the corporation's computer vendor, IBM, the VP became convinced that customer support could be enhanced through the implementation of an ES, a recognized benefit of expert system applications (Mattei, 2001).

Provide Consistent Quality Customer Service

The VP had the insight to identify the importance of offering fast, expert advice regarding the appropriate use of optical brighteners for individual customer's applications at the time of on-site sales calls. Optical brighteners are used for a wide variety of end-products. For textiles, paper products, and detergents, optical brighteners are applied to enhance coloring, that is, to make "whites whiter and brights brighter." Non-textile applications include testing for leaks, such as those in automotive parts. Salespeople are thus challenged to make appropriate and specific recommendations concerning a wide range of applications, wherein the factors to consider can vary widely. The inability of a salesperson to answer customers' questions can result in delayed or lost sales. Recognizing this impact, the VP championed the expert system as a means of increasing sales profitability. Individual customer questions could be addressed on the spot while maintaining consistency and quality in responses.

Train the Sales Force

By managing and distributing knowledge about optical brightener product features, areas of application, and troubleshooting solutions, the sales force is able to develop a greater understanding of the optical brightener product line and technical characteristics. New salespeople benefit by having unconstrained access to a "technical expert". Sales force training is thereby enhanced during formal training sessions and while on the job.

Manage Critical Product Expertise

The VP envisioned an entire family of ES, for all optical brightener applications, would be developed in the future. The domain for the first ES was appropriately narrowed to include the application of optical brighteners to fabrics only. An expert in this area of application was identified. This individual has extensive experience with the optical brightener product category, having served as a customer support technician and troubleshooter for over 15 years. His knowledge about customer requirements, properties of fabrics, and application processes enabled him to recommend appropriate optical brightener products based on features of those products. This critical product expertise was captured, documented, and can be distributed through OpBright. Such astute management can preclude the loss of valuable corporate knowledge and expertise resulting from normal turnover or absence (Mattei, 2001), promotion (Prietula & Simon, 1989), or retirement.

Momentum without a Project Champion

The VP classified the development of the ES as a research and development (R&D) effort. No formal cost/benefit analysis was performed. Even for R&D projects, a formal analysis is recommended (Brenner & Tao, 2001). However, the cost of this first-time project was viewed as an investment in experience to be applied to future areas of application, an argument commonly made by champions in securing funding (Sharpe & Keelin, 1998). The VP was convinced that the impact of this technology would far outweigh the initial investment.

OpBright encountered a stalemate shortly after its completion. The VP was promoted and transferred to another division. The incoming VP reviewed the OpBright expert system project and restated the objective to develop a *prototype* ES. The objective was met as the prototype had been developed. Without a replacement project champion, OpBright remains at a standstill.

FUTURE TRENDS

Lessons learned, from the Ciba-Geigy case study, to avoid pitfalls attributable to the departure of the champion are discussed in the following sections. These insights may serve to guide future research; focusing on identifying strategies organizations may employ to nurture the continuation of the role of project champion.

Incorporate Expertise Management in Strategic Planning

An assessment of areas of expertise critical to organizational processes and the potential for applying ES technology, should be included within the strategic planning process. Expertise management thereby becomes a formalized managerial activity (Sipior & Garrity, 1990). Incorporating expertise management in strategic planning is clearly unique to ES in particular, differentiating ES project champions from champions of technological innovations in general.

Secure the Support of Top Management

It is well recognized that securing the support of top management is a prerequisite to the long-term success of ES projects (Sipior & Volonino, 1991; Wong, 1996). Broad support for technological innovation from the top individual alone is insufficient unless it is translated into support for specific applications of ES technology by management levels below him.

Top management support is more likely if the ES fits with corporate strategy. By incorporating expertise management within strategic planning, this fit becomes more likely (Zack, 1999). In turn, this fit will garner broader organizational commitment (Meador & Mahler, 1990). To gain the necessary support, a results-oriented approach is preferable since it enables management to buy into the impact ES can have, rather than focusing on the technology itself (Sipior & Volonino, 1991). For management support to be on-going, the results should have continued benefit.

Secure the Support of the Next Generation of Top Management

Top management support is not sufficient, unless the support of the next generation of top management is secured (Kotter, 1995). If successors are ambivalent about the ES project, as was the case for Ciba-Geigy, they certainly will not take the initiative to understand this technology and maintain project momentum.

Recognize, Support, and Nurture a Project Champion

As discussed, a project champion tends to emerge, rather than be assigned to the role. Evidence suggests both the inability to nurture such individuals as well as instances wherein this individual has been successfully developed

(Schon, 1963). When a champion does not emerge naturally, a company may be able to "*find* or *make* a champion for the system" (Curley & Gremillion, 1983, p. 207).

Formalize Project Measurement, Monitoring, and Follow-up

Formalization of expert system project measurement, monitoring, and follow-up evaluation can improve the probability of project success (Brenner & Tao, 2001). Project measurement should include an evaluation at each phase of the project: (1) the validity of the technology to be employed in terms of delivering its claimed capabilities, (2) the benefits to be derived by its application, and (3) the products and services to be produced by employing the technology (Brenner & Tao, 2001). For each phase, project characteristics should be evaluated. The type of idea that generated the project could be rated from high of three to low of zero for a technological advance, new or novel technology, new twist on a known technology, or no technological advance. Additional characteristics to consider include expertise and capabilities gained, time saved, R&D dollars saved, intellectual property expected from the project, importance of the project objective, commercial impact such as increased sales revenue, technical leverage, and internal and external relationship building. Changes in the project should be monitored as the project advances. At project implementation, the estimates from the initial project evaluation should be compared to actual project performance data. Formalizing measurement, monitoring, and follow-up evaluation forces project accountability in terms of delivering the intended results.

CONCLUSIONS

The experiences of Ciba-Geigy underscore the critical importance of the role of project champion in ES development. Conversely, the loss of this individual can threaten the very existence of a project. Organizations should thus take heed and devote attention to harnessing the enthusiasm and drive of ES project champions for advantage, before it is too late to preclude the loss of momentum generated by these individuals. Future research may be directed toward exploring the reasons for failure of projects supported by champions, and identifying strategies organizations may employ to nurture the continuation of the role of project champion for ES.

REFERENCES

Alavi, M., & Leidner, D.E. (1999). Knowledge management systems: Issues, challenges, and benefits. *Communications of the Association for Information Systems*, *1*(2), Article 1.

Beath, C.M. (1991). Supporting the information technology champion. *MIS Quarterly*, *15*(3), 355-372.

Beath, C.M., & Ives, B. (1988). The information technology champion: Aiding and abetting, care and feeding. *Proceedings of the 21st Annual Hawaii International Conference on Systems Sciences IV* (pp. 115-123), Kailua-Kona, Hawaii, USA.

Bolton, B., & Thompson, J. (2000). A breed apart. *Director*, *53*(10), 54-57.

Brenner, M.S., & Tao, J.C. (2001). *Research Technology Management*, *44*(3), 14-17.

Curley, K.F., & Gremillion, L.L. (1983). The role of the champion in DSS implementation. *Information and Management*, *6*, 203-209.

Hansen, M.T., Nohria, N., & Tierney, T. (1999). What's your strategy for managing knowledge? *Harvard Business Review*, *77*(2), 106-119.

Hayes-Roth, F., & Jacobstein, N. (1994). The state of knowledge-based systems. *Communications of the ACM*, *37*(3), 27-35.

Howell, J.M., & Higgins, C.A. (1990). Champions of technological innovation. *Administrative Science Quarterly*, *35*(2), 317-341.

Kotter, J.P. (1995). Leading change: Why transformation efforts fail. *Harvard Business Review*, *73*(2), 59-67.

Landers, T.L. (1999). Are you a project champion? *Modern Materials Handling*, *54*(2), 33.

Martinsons, M.G. (1993). Cultivating the champions for strategic information systems. *Journal of Systems Management*, 31-34.

Mattei, M.D. (2001). Using "expert systems" for competitive advantage. *Business and Economic Review*, *47*(3), 17-20.

Mayhew, D.J. (1999). Business: Strategic development of the usability engineering function. *Interactions*, *6*(5), 27-34.

Meador, C.L., & Mahler, E.G. (1990). Choosing an expert systems game plan. *Datamation,* (August), 64-69.

Pinto, J.K., & Slevin, D.P. (1989). The project champion: Key to implementation success. *Project Management Journal, 20*(4), 15-20.

Prietula, M.J., & Simon, H.A. (1989). The experts in your midst. *Harvard Business Review, 67*(1), 120-124.

Royer, I. (2003). Why bad projects are so hard to kill. *Harvard Business Review, 81*(2), 48-56.

Schon, D.A. (1963). Champions for radical new inventions. *Harvard Business Review, 41*(2), 77-86.

Scott, J.E., & Vessey, I. (2002). Managing risks in enterprise systems implementations. *Communications of the ACM, 45*(4), 74-81.

Sharpe, P., & Keelin, T. (1998). How SmithKline Beecham makes better resource-allocation decisions. *Harvard Business Review, 76*(2), 45-57.

Sipior, J.C. (2000). Expert system stalemate: A case of project champion departure. *Information Resources Management Journal, 13*(4), 16-24.

Sipior, J.C., & Garrity, E.J. (1990). The potential of expert systems for competitive advantage. *Proceedings of the Decision Sciences Institute*, San Diego, CA USA.

Sipior, J.C., & Volonino, L. (1991). Changing the image of expert systems from sensational to organizational tools. *Proceedings of the 24th Annual Hawaii International Conference on System Sciences*, Kauai, Hawaii, USA.

Sumner, M. (2000). Critical success factors in enterprise wide information management systems projects. *Proceedings of the 2000 ACM SIGCPR Conference on Computer Personnel Research*, Chicago, IL, USA.

Thomas, J. (1999). Share and share alike. *Logistics Management and Distribution Report, 38*(3), 44-48.

Willcocks, L.P., & Sykes, R. (2000). The role of CIO and IT function in ERP. *Communications of the ACM, 43*(4), 33-38.

Wong, B.K. (1996). The role of top management in the development of expert systems. *Journal of Systems Management*, 36-40.

Zack, M.H. (1999). Managing codified knowledge. *Sloan Management Review, 40*(4), 45-58.

KEY TERMS

Assertion: Authoritatively convince others of the project's potential benefits so they dedicate their efforts to the project.

Charismatic Behavior: Captivate others into believing in the project as the champion himself does.

Impression Building: Portray outcomes of endeavors as highly positive achievements to promote an image of competence and success.

Inspirational Behavior: Influence others by using emotional appeals, and vivid and persuasive images, to elevate their performance.

Intellectual Stimulation: Challenge others to aspire to imaginative use of their individual skills.

Persuasive Communication: Rely more on persistence, rather than clear and concise arguments, to attain agreement.

Rational Justification: Analyze how the project advances goals and objectives and upholds values of the organization.

Transformational Leaders: Inspire others, through charisma, inspiration, intellectual stimulation, and individualized consideration, to transcend their own self-interests for a higher collective purpose.

Design Levels for Distance and Online Learning

Judith V. Boettcher
Designing for Learning and the University of Florida, USA

INTRODUCTION

The importance of design for instructional programs — whether on campus or online or at a distance — increases with the possible combinations of students, content, skills to be acquired, and the teaching and learning environments.

Instructional design —as a profession and a process— has been quietly developing over the last 50 years. It is a multidisciplinary profession combining knowledge of the learning process, humans as learners, and the characteristics of the environments for teaching and learning. The theorists providing the philosophical bases for this knowledge include Dewey (1933), Bruner (1963), and Pinker (1997). The theorists providing the educational and research bases include Vygotsky (1962), Knowles (1998), Schank (1996), and Bransford, Brown, and Cocking (1999).

Instructional design offers a structured approach to analyzing an instructional problem and creating a design for meeting the instructional content and skill needs of a population of learners usually within a specific period of time. An instructional design theory is a "theory that offers explicit guidance on how to better help people learn and develop" (Reigeluth, 1999).

BACKGROUND

This entry describes a multi-level design process for online and distance learning programs that builds on a philosophical base grounded in learning theory, instructional design, and the principles of the process of change as reflected in the writings of the theorists listed above. This design model builds on traditional instructional design principles, as described by Gagne (1965), Dick & Carey (1989), and Moore & Kearsley (1996). It integrates the strategic planning principles and the structure of the institutional context as described in Kaufman (1992) and Boettcher & Kumar (1999), and also integrates the principles of technological innovation and the processes of change as described by E. M. Rogers (1995) and R. S. Rosenbloom (1998).

This entry describes a six-level design process promoting congruency and consistency at the institution, infrastructure, program, course, activity, and assessment level. It also suggests a set of principles and questions derived from that framework to guide the instructional design process.

Figure 1. Six levels of design for learning

Six Levels of Design	Design Responsibility	Sponsor/Leader	Design and Review Cycle
Institution	Entire campus leadership and community	Provost, CIO and Vice-presidents	3-5 Years
Infrastructure	Campus and Technology Staff	Provost, CIO and Vice-presidents	2-3 Years
Degree, Program	College/Deans/Faculty	Dean and Chairs	1-3 Years
Course	Faculty	Dept Chair	1-2 Years
Unit/Learning Activity	Faculty	Faculty and or Faculty team	1-2 Years
Student Assessment	Faculty	Faculty and or Faculty team	1-2 Years

SIX LEVELS OF DESIGN

Effective instructional design for online and distance learning benefits from instructional planning at six levels. Figure 1 summarizes these six levels of design, and identifies the group or individuals usually responsible for the design at that level and the length of the design cycle at each level. Ideally, the design at each of these six levels reflects philosophies of teaching and learning that are consistent with the institutional mission and consistent with the expectations of the students and society being served.

Level One: Institutional Design

The design work to be done at an institutional level is similar to the strategic planning and positioning of an institution. Institutional planning generally begins with an institution's current vision and mission statements and then proceeds through a data collection and input process that addresses a set of questions such as the following:

Institutional Questions:

- What programs and services comprise our primary mission? For whom?
- To what societal needs and goals is our institution attempting to respond?
- What life goals are most of our students working to achieve?
- What type of learning experiences are our students searching for?
- What changes in our infrastructure are needed to match our desired services, programs, and students?
- Does our institution have any special core competencies, resources, or missions that are unique regionally or nationally that might form the basis for specialized online and distance programs? What are the strengths of our mature faculty? Of our young faculty?

Level Two: Infrastructure Design

People often think that buildings, classrooms, Web applications, communication services, and servers are neutral as far as having an effect on teaching and learning. Nothing could be more misleading. Design of the infrastructure includes design of all the elements of the environment that impact the teaching and learning experiences of faculty and students and the staff supporting these experiences. It includes design of the following:

- Student services, faculty services, and learning resources.
- Design of administrative services, including admission processes, financial processes, and institutional community life events.
- Design of physical spaces for program launching events, hands-on, lab, or network gathering events, as well as celebratory graduation events.

Physical and Digital Plants

Infrastructure design for online and distance teaching and learning programs focuses on the design of the network and Web infrastructure. Infrastructures for online learning have offices, classrooms, libraries, and gathering spaces for the delivery and management of learning and teaching. However, these offices and classrooms are accessed through Web services, rather than through physical buildings. The good news about online infrastructures is that they support an unparalleled new responsiveness, feedback, and access for learning activities.

After almost ten years of building online campuses, we now know that a "digital plant" infrastructure is needed to support the new flexible online and distance environments. We know that this new digital plant needs to be designed, built, planned, maintained, and staffed. The infrastructure to support the new programs cannot be done with what some have called "budget dust" (McCredie, 2000). It is not nearly as easy or inexpensive as we all first thought. Some experts suggest that, a "full implementation of a plan for technology support on campus costs about the same as support of a library — approximately 5% of the education and general budget" (Brown, 2000).

Components of a Digital Infrastructure

What exactly is a digital plant infrastructure? One way of describing this infrastructure is to think of it in four major categories of personal communication tools, networks, hardware for servers, and software applications. A key component of the digital infrastructure is the group of individuals who make the systems work. This digital plant is shown in Figure 2 (Boettcher and Kumar, 2000).

Some of the questions that might be used to guide the development of the digital infrastructure follow.

Personal communication tools and applications:

- Will all students have their own computer? Their own laptop?
- Do we expect students all to be proficient with word processing applications, mail, Web applications, researching on the Internet? With collaborative

tools and with one or more course management systems?

Networks that provide access to Web applications and resources and to remote, national, and global networks:

- What physical wired or wireless networks are needed to support Web applications, such as e-mail servers, directory servers, and Web application services?
- How often will higher bandwidths be needed for video conferencing for programs? For meetings? For downloading large files? For streaming video?

Dedicated servers and software applications that manage campus services:

- What types of interactive Web services will be provided? What hardware and software will be required?•What type of administrative systems and course management system will we use?
- What do we need to do to assure student, faculty, and staff accessibility from anywhere at anytime?

Software applications and services from external providers, such as research and library services that are licensed to the institutional community, Internet services, and out-sourced services, such as network services:

- What licensed services are required and desired?
- What budget is required to support these services currently and into the future?

Technology decisions for students have always been part of the instructional design process for distance learning. A comforting way of thinking about the technology for the infrastructure design level is in terms of the generations of technologies used in distance learning. (Sherron and Boettcher, 1997). Distance learning was made possible with the widespread availability of technologies, such as the mail, radio, telephone, television, and audio and videocassettes. In the 21st century we simply have more technology and more choices.

Now let's look at the design of programs and courses. Design issues at these levels are principally the responsibility of the institutional academic leadership.

Level Three: Program Design

At the program level of design, instructional planners answer questions about the type of program to be offered, to whom, and over what period of time and at what cost. When venturing into new business areas, the following two guidelines are useful: (1) focus on programs that can leverage institutional core competencies and strengths, (2) plan a phased approach, gaining experience in delivering programs in one or two areas before launching others, and (3) recognize that online and distance learners generally are interested in achieving or completing an instructional goal that can assist in their current or future career path.

It is in the next four levels of design that the principles of Vygotsky are most applied, building on Vygotsky's (1962, 1978) view of the learner as a goal-oriented learner within a specific learning context using specific resources as directed by a teacher. These four core elements of all learning experiences provide a framework for the design process:

- The person doing the learning — the learner
- The person guiding and managing the learning — the faculty/teacher/mentor
- The content /knowledge/skill to be acquired/or problem to be solved
- The environment or context within which the learning experience occurs

Program Level Planning Questions:

Program planning design has four categories of planning — curriculum, design/development process, faculty, and student.

Curriculum questions:

- What is the degree or certificate program to be offered online? Will it be a full master's degree (10 to 16 courses), an undergraduate minor (four to six courses) or a certificate program (two to four courses)?
- What types of courses are envisioned? Will these courses be a fully developed "course in a box" with a minimal amount of interaction or a highly interactive and collaborative course requiring or using many online resources and applications?

Design and development questions:

- Who are the faculty who will design, develop and deliver the courses in the program?
- Who will lead the effort to develop the degree or certificate program for online or distance delivery? Which organization will be marketing the program?
- What course management system or similar Web tool will be used for the content management? What tools and resources will be available and

supported for the interaction and collaboration activities?
- What is the schedule for design and development and delivery of courses and program? For the marketing and recruiting of the students?

Faculty questions:

- What training will be available to faculty as they transition to online teaching and learning programs?
- What tools and resources and support will be available to faculty?
- Will faculty have any released time or budget for teaching and learning resources in the new online or distance environment?
- What type of access to the network is recommended and available? Will dial-up be sufficient, or will DSL or cable access be recommended or required?

Student questions:

- Who are the students who will enroll in this course of study? How will we find them and market the program to them?
- What will our students bring to the program experience?
- What tools and resources will the student in this program or certificate program require or be likely to use?
- Where will the students be doing their learning and with what types of content resources and applications? What level of network access is required or recommended?

The question of technology access for students was particularly important in the mid-1990s, when technology access was relatively scarce. However, the latest data from the Campus Computing Study of 2002 suggests that more than 75% of all students own their own desktop or notebook computer (Green, 2002). If all students have their own computers and access to the Internet, this access greatly impacts the design of communication activities and course experiences.

Course Design — Level Four

Design at the course level is usually considered to be the responsibility of the faculty member. In online and distance courses, however, the stand-alone course is the exception rather than the rule. Most online and distance courses are part of a curriculum, certificate, or degree program. This means course-level design occurs within the context of the larger program and that many of the design decisions are made in collaboration with other faculty within the academic program or department. Faculty at the course level are primarily responsible for design decisions on content, objectives, student goals, learning experiences, and assessment for a particular course. Many of these questions for this design level parallel questions at the program level design. The following questions are more specific to a single course:

Course questions:

- Where does this course fit within the context of the degree or certificate program to be offered online? Is it an early course focused on core discipline concepts, peer discussions, and standard problems or a later course focusing on applications and complex scenarios?
- What is the core set of knowledge, skills, and attitudes/values to be acquired by the students?
- What is the set of content resources required and recommended? What content resources will students use to customize the learning experience for their needs and state of knowledge and personal interests?
- Will students be a cohesive cohort group?

Design and development questions:

- What types of instructional activities and experiences will support student learning of the knowledge, skills, and attitudes of the students?
- What course management system or similar Web tool will be used for the content management? For the interaction and collaboration activities?
- What is the schedule for design and development and delivery of this course?

Faculty questions:

- What training is needed for a faculty member to transition to online teaching and learning programs?
- What tools and resources and support will be needed to support the delivery of this course?

Student questions:

- Who are the students taking this course? What are their hopes and expectations? What future courses will depend on the knowledge, skills, and attitudes acquired in this course?
- What knowledge and expertise do the students bring to the course? What is their zone of proximal development (Vygotsky, 1978)?

- What types of teaching and learning strategies best suit the students in this course? What are the life style and learning styles of the students?
- When and where will the students be likely to gather for their collaborative work? When and where will they do their more self-study activities?
- Where will the students be doing their learning? What level of network access is required or recommended?

The next two design levels are within the course parameters and generally are the responsibility of the faculty designing the course.

Level Five: Unit/Learning Activity

Many of the design questions for the unit/learning activity level and the student assessment level are derived from a design model that focuses on integrating student life style and learning styles into instructional planning (Boettcher, 2003). Examples of cognitive learning style design questions include: "How do students process information?"; "How do students respond in their minds when challenged with new concepts and rich content structures?", "What knowledge do students bring to the learning experience?"

The life style of the learner is also addressed in these questions. Life style includes all the elements in a learner's current life situation. Where will the learner be working? Will they have a personal space where they can control sound, temperature, disturbances, and network access? Will they have to "ask" their family if they can access the network? A life-style focus encourages analysis of the where, when, with whom, and with what resources the learner is going to be doing their learning work.

Learning work consists of constructing new knowledge, applying and integrating knowledge, and solving problems with that new knowledge. Mobile, wireless technologies enable learners to study anywhere at any time. Initially, the ability to study anywhere seemed to hold the promise of solving many problems associated with access to learning. However, we have not addressed the question of just when and just where this "anytime" is likely to occur.

Learning activity questions:

- What is the knowledge, skill or attitude that is the desired outcome of this learning activity?
- What kinds of problems can students solve now? What kinds of problems do we want students to be able to solve at the conclusion of the experience?
- What instructional strategy or experience will support the learner learning the desired knowledge, skill, or attitude?
- When, where, with whom, and with what resources is the instructional activity envisioned to occur?
- What role will the teacher/mentor/faculty play during this activity? Will the teacher be present physically or at a distance, synchronously or asynchronously?
- When will a learner know that he/she knows? What feedback or result from a problem being solved will make the learner's knowledge evident?

Level Six: Assessment of Student Learning

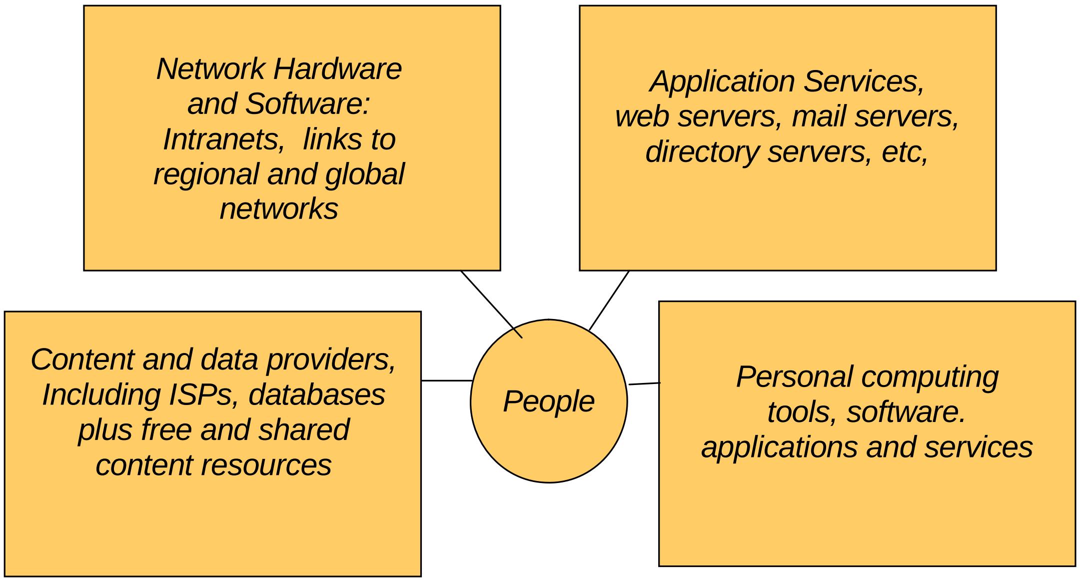
Assessment is fundamental to the design process. Assessment planning helps to balance the goals of learning effectiveness and teaching and learning efficiency. In productive, customized, enriched learning experiences, each learner begins with an existing, personal store of knowledge representation. Learning experiences expand and enrich that knowledge base in individual and personal ways. Productive learning experiences mean that all students complete a learning experience with an expanded, yet different, store of knowledge. The goal is that learners learn core principles so that the core principles can be effectively applied, but the particular way the knowledge base is constructed in individuals is unique.

What we can design into instructional planning is that all learners share some of the same experiences and that assessment focuses on the common learning that is achieved. Assessment can also provide for demonstration of knowledge and skills in more complex environments and for some elements of customized knowledge acquisition. Here are some selected questions for assessing student learning:

Assessment Questions

- How will the learners know the goals and objectives for the learning? It is good to plan for core concepts, practice of core concepts, and customized applications of concepts.
- Will learners be generating a set of their own goals for learning? How will the faculty mentor and learner communicate and agree on goals for learning, particularly for customized applications of concepts?
- In what ways and where will the students be evaluated and graded?
- How will students demonstrate their competency in concept formation? In solving problems?
- If we don't see the students on a regular basis, can we design ways to "see" their minds virtually through conversation and experiences?

Figure 2. Teaching and learning infrastructure - "digital plant"



The portfolio project within the National Learning Infrastructure Initiative (NLII) is also useful as an assessment of complex learning (www.educause.edu/nlii/keythemes/eportfolios.asp).

FUTURE TRENDS

The process of instructional design is a professional task requiring knowledge of educational research, learning processes, and, increasingly, a respect for context, and a comfort level with innovation and instructional technologies. It is a labor-intensive task requiring interaction with content experts and administrators and institutional representatives. Much of the current instructional design processes focus on the analyses of needs and contexts for learning for learners, their tools and their resources — as a group. Future work in instructional design will focus more on the learner as an individual with unique knowledge structures and thus promote the needs for a rich contextual learning environment that is multi-leveled and customizable.

The design of instructional planning will become more of a priority as the demand for effective and efficient learning grows as a result of time pressures, budget pressures, and increasing demands for accountability in education.

CONCLUSION

These instructional design principles reaffirm the iterative nature of design work and the sharing of design work among the hierarchical groups of an institution. Instructional design, when done well, results in delighted and productive learners and faculty pleased with their roles and their work. Consistently applied, instructional design principles keep teaching and learning focused on the who, when, where, how, and why of teaching and learning and help to ensure that the money and time invested in learning programs provide an appropriate return for individuals and for society. Instructional design is a powerful tool that moves teaching and learning into the science of learning and knowing.

REFERENCES

Boettcher, J.V. (2000). How much does it cost to put a course online? It all depends. In M.J. Finkelstein, C. Francis, F. Jewett, & B. Scholz (Eds.), *Dollars, distance, and online education: The new economics of college teaching and learning* (pp. 172-197). Phoenix, AZ: American Council on Education/Oryx Press.

Boettcher, J.V. (2003). Design levels for distance and online learning. In R. Discenza, C. Howard, & K. Schenk

(Eds.), *Distance learning and university effectiveness: Changing educational paradigms for online learning.* Hershey, PA: Idea Group.

Boettcher, J.V., & Kumar, V.M.S. (2000, June). The other infrastructure: Distance education's digital plant. *Syllabus*, (13), 14-22.

Bransford, J.D., Brown, A.L., & Cocking, R.R. (1999). *How people learn. Brain, mind, experience, and school.* Washington, DC: National Academy Press. Retrieved from the World Wide Web at: *www.nap.edu/books/0309070368/html/*

Brown, D.G. (2000). Academic planning and technology. In J.V. Boettcher, M.M. Doyle, & R.W. Jensen (Eds.), *Technology-driven planning: Principles to practice* (pp. 61-68). Ann Arbor, MI: Society for College and University Planning.

Bruner, J.S. (1963). *The process of education.* New York: Vintage Books.

Business-Higher Ed Forum.(2003). *Building a nation of learners: The need for changes in teaching and learning to meet global challenges.* American Council on Education. Retrieved from the World Wide Web at: *www.acenet.edu/bookstore/pubInfo.cfm?pubID=285*

Dewey, J. (1933). *How we think* (1998 ed.). Boston, MA: Houghton-Mifflin.

Dick, W., & Carey, L. (1989). *The systemic design of instruction.* New York, Harper Collins.

Gagne, R. M. (1965). *The conditions of learning.* New York: Holt, Rinehart & Winston.

Green, K. C. (2002). *Campus computing, 2002.* Encino, CA: The Campus Computing Project. Retrieved from the World Wide Web at: *www.campuscomputing.net*

Kaufman, R. (1992). *Strategic planning plus : An organizational guide.* Thousand Oaks, CA: Sage Publications.

Knowles, M. (1998). *The adult learner: A neglected species*. Houston, TX: Gulf.

Moore, M.G., & Kearsley. G. (1996). *Distance education: A systems view.* Belmont, CA: Wadsworth.

Newell, H.A. (1996). *Sciences of the artificial.* Boston, MIT Press.

Pinker, S. (1997). *How the mind works.* New York: W.W. Norton.

Reigeluth, C.M. (1999). What is instructional-design theory and how is it changing? In C.M. Reigeluth (Ed.), *Instructional-design theories and models, volume II: A new paradigm of instructional theory* (pp. 5-29). Mahwah, NJ: Lawrence Erlbaum.

Rogers, E.M. (1995). *Diffusion of innovations.* New York, Free Press.

Rosenbloom, R.S. (1998). *Sustaining American innovation: Where will technology come from?* Forum on Harnessing Science and Technology for American's Economic Future, National Academy of Sciences Building, Washington, DC, National Academy of Science.

Schank, R.C. (1996). Goal-based scenarios: Case-based reasoning meets learning by doing. In D. Leake (Ed.), *Case-based reasoning: Experiences, lessons & future directions* (pp. 295-347). AAAI Press/The MIT Press.

Schrum, L. & Benson, A. (2002). Establishing successful online distance learning environments: Distinguishing Factors that contribute to online courses and programs. In R. Discenza, C. Howard, & K. Schenk (Eds.), *The design and management of effective distance learning programs* (pp. 190-204). Hershey, PA: Idea Group Publishing.

Sherron, G. T., & Boettcher, J. V. (1997). *Distance learning: The shift to interactivity.* CAUSE Professional Paper Series #17. Retrieved September 22, 2004 from the World Wide Web at: *www.educause.edu/asp/doclib/abstract.asp?ID=pub3017*

Vygotsky, L.S. (1962). *Thought and language.* (E. Hanfmann & G. Vakar, trans.) Cambridge: MIT Press.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge: Harvard University Press.

KEY TERMS

Instructional Design: The process of analyzing the students, content, and intended context of an instructional program to provide detailed specifications for an instructional program or curriculum to achieve effective and efficient student learning within an affordable and accessible delivery format.

Instructional Design Theory: A "theory that offers explicit guidance on how to better help people learn and develop" (Reigeluth, 1999).

Instructional Strategy: An instructional strategy is a communication activity used to engage the learner in an educational experience and to assist the learner in acquiring the planned knowledge, skill, or attitude. Instructional strategies include lectures, discussions, reading assignments, panel presentations, study and media projects, problem analysis and solutions, field trips and assessment activities.

Learning Infrastructure: The set of physical and digital buildings, applications, services, and people that provide and support the environments for learning.

Learning Theory: A set of hypotheses or beliefs that explain the process of learning or acquiring knowledge and skill.

Online Course: A set of instructional experiences using the digital network for interaction, learning and dialogue. An online course does not require any face-to-face meetings in a physical location. Similar courses such as web-centric courses (also called hybrid or blended courses) are similar to online courses, but require regular scheduled face-to-face classes or meetings.

Zone of Proximal Development: This is a key concept in Lev Vygotsky's theory of learning. The Zone of Proximal Development (ZPD) is the "distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under the adult guidance or in collaboration with more capable peers" (Vygotsky, 1986).

ENDNOTE

Note: This article is an adaptation of the following book chapter. Boettcher, J.V. (2003). Design levels for distance and online learning. In R. Discenza, C. Howard, & K. Schenk (Eds.), *Distance learning and university effectiveness: Changing educational paradigms for online learning*. Hershey, PA: Idea Group.

Designing Agents with Negotiation Capabilities

Jana Polgar
Monash University, Australia

SOFTWARE AGENTS TODAY

Agents are viewed as the next significant software abstraction, and it is expected they will become as ubiquitous as graphical user interfaces are today. Agents are specialized programs designed to provide services to their users. Multiagent systems have a key capability to reallocate tasks among the members, which may result in significant savings and improvements in many domains, such as resource allocation, scheduling, e-commerce, and so forth. In the near future, agents will roam the Internet, selling and buying information and services. These agents will evolve from their present day form - simple carriers of transactions - to efficient decision makers. It is envisaged that the decision-making processes and interactions between agents will be very fast (Kephart, 1998).

The importance of *automated negotiation systems* is increasing with the emergence of new technologies supporting faster *reasoning engines* and mobile code. A central part of agent systems is a sophisticated reasoning engine that enables the agents to reallocate their tasks, optimize outcomes, and negotiate with other agents. The *negotiation strategy* used by the reasoning engine also requires high-level inter-agent communication protocols, and suitable collaboration strategies. Both of these subsystems – a *reasoning engine* and a *negotiation strategy* - typically result in complicated agent designs and implementations that are difficult to maintain.

Activities of a set of *autonomous agents have to be coordinated*. Some could be mobile agents, while others are static intelligent agents. We usually aim at decentralized coordination, which produces the desired outcomes with minimal communication. Many different types of *contract protocols* (cluster, swaps, and multiagent, as examples) and *negotiation strategies* are used. The evaluation of outcomes is often based on marginal cost (Sandholm, 1993) or game theory payoffs (Mass-Colell, 1995). Agents based on constraint technology use complex search algorithms to solve optimization problems arising from the agents' interaction. In particular, coordination and negotiation strategies in the presence of incomplete knowledge are good candidates for constraint-based implementations.

SELECTED NEGOTIATION AND REASONING TECHNIQUES

Negotiation space is determined by two components: *negotiation protocol* and *negotiation strategy*. The *negotiation protocol* defines the rules of behavior between the participants in terms of interactions, deals, bidding rules, temporal constraints and offers, as components of the protocol. Two agents must first agree on the negotiation protocol before any interaction starts.

The *negotiation strategy* is a specification of the sequence of actions the agent intends to make during the negotiation. Strategies should be compatible with the negotiation protocol. The focus of any negotiation strategy is to maximize outcomes within the rational boundaries of the environment. The classification of negotiation strategies is not an easy task since the negotiation strategy can be realized by any algorithm capable of evaluating outcomes, computing appropriate actions, and following the information exchange protocol.

The *negotiation mechanism* is the actual implementation of negotiation strategy and negotiation protocol. This field is evolving fast, with emergence of new agent platforms, wireless encounters and extended mobility.

Negotiation is a search process. The participants jointly search a multi-dimensional space (e.g., quantity, price, and delivery) in an attempt to find a single point in the space at which they reach mutual agreement and meet their objectives. The *market mechanism* is used for many-to-many coupling or interactions between participants. *Auctions* are more appropriate for one-to-many negotiations. The market mechanism often suffers from inability to efficiently scale down (Osborne, 1990) to smaller numbers of participants. On the other hand, one-to-many interactions are influenced by strategic considerations and involve integrative bargaining, where agents search for *Paretto efficient* agreements (tradeoffs).

NEGOTIATION STRATEGIES

Analytical Approach (Game Theory)

The principles of bargaining and negotiation strategies in multiagent systems have attracted economists. Early foundations and mathematical models were investigated by Nash (1950), and the field is still very active. The *game theory* is a collection of analytical tools designed to understand and describe bargaining and interaction between decision makers. Game theory uses mathematical models to formally express real-life strategies (Fudenberg, 1991; Osborne, 1994).

The high-level abstraction allows the model to be applied to a variety of situations. The model places no restrictions on the set of actions available to the player. With regard to mathematical models, there already exist many sophisticated and elaborated strategies for specific negotiation problems. The Contract Net Protocol (CNP) (Sandholm, 1993; Smith, 1980) represents the model of decentralized task allocation where agents locally calculate their marginal costs for performing sets of tasks. The pricing mechanism in Sandholm (1993) generalizes the CNP to work for both cooperative and competitive agents. In Zeng (1996), bilateral negotiation based on the Bayesian method is presented. It demonstrates the static nature of the model. The learning effect is achieved by using dynamic updates of a knowledge base, which is consulted during the negotiation process.

Most of the studies assume perfect rationality (flawless deduction, marginal costs are computed exactly, immediately and without computational cost), and the infinite horizon of strategic bargaining. These are not realistic assumptions. More advanced studies deal with coalition formation and negotiation strategies in the environment of multiple self-interested or cooperative agents with bounded rationality (Sandholm, 1993) and bargaining with deadlines.

Analytical approach has the advantage of stable and reliable behavior. The main disadvantage is the static nature of the model, resulting in potential predictability of the outcomes. The other problems are associated with the notion of perfect rationality.

Contracts in automated negotiations consisting of self-interested agents are typically designed as binding (impossible to breach). In cooperative distributed problem solving, commitments are often allowed to be broken based on some local reasoning. Frequently, the protocols use continuous levels of commitment based on a monetary penalty method (Sandholm, 1993). Unfortunately, the inflexible nature of these protocols restricts an agent's actions when the situation becomes unfavorable. The models that incorporate the possibility of decommitting from a contract with or without reprisals (Sen, 1994; Smith, 1980) can accommodate some changes in the environment and improve an agent's status. However, all of these protocols are somewhat restricting with respect to evolving, dynamic situations.

Evolutionary Strategies

With *evolutionary strategies*, the data used as the basis for negotiation, as well as the algorithm operating on the data, evolve. This approach provides more efficient learning, supports the dynamics of the environment, and is adaptable. However, only a few implementations have been attempted, and these have been of only simple negotiation strategies (Aridor, 1998). *Genetic algorithms* are probably the most common techniques inspired by evolution, in particular by the concepts of natural selection and variation. The basic genetic algorithm is derived from the hypothesis that the candidate solutions to the problem are encoded into "chromosomes". Chromosomes represent a solution or instance of the problem hand encoded into a binary string. The algorithm then operates on this binary string. It begins with a randomly generated set of candidate solutions. The set of candidate solutions is generated as a random string of ones and zeroes. Each chromosome is evaluated and the fitness of the chromosome could be the value of the objective function (or the utility if we want to maximize the outcome). A new population is created by selecting individuals to become parents. A thorough description of the genetic algorithm approach can be found in Goldberg (1989).

A very large amount of research has been carried out in the application of evolutionary algorithms to situations that require decisions. Examples include coalition games, exchange economies, and double auctions. This approach was inspired by the concept of variation and natural selection. The intelligent agents are modeled using classifier systems to select decisions. Although the recent research shows that multiagent systems of classifiers are capable of learning how to play *Nash-Markov equilibrium*, the current limitations of computational resources and the instability of "home-grown" implementations significantly constrain the nature of the strategies. The important question is what design and implementation techniques should be used to ease this conflict and to provide the resources required for genetic learning to operate in an unrestricted way. It is believed that the ability of agents to learn simple games would be beneficial to electronic commerce.

Constraint Agents

The potential of constraint-based agents is still to be fully realized and appreciated. One of the possible frameworks for constraint-based agents is outlined in Nareyek (1998). This framework considers agents as a means for simplifying distributed problem solving. An agent's behavior and the quality of solutions depend on the underlying action-task planning system. The recent results with some constraint planners and constraint satisfaction problems (CSP) indicate the potential advantages of this approach.

Agents operating with only partial knowledge of the surrounding environment are prime candidates for the implementation of reasoning using constraint solvers. Agents can share and exchange knowledge in a distributed fashion. The suggestion-correction process used in negotiation strategies corresponds to mapping constraint satisfaction problems to search algorithms and heuristics (Tsang, 1993). The planning and scheduling systems can be treated as CSPs that allow the use of constraint solvers to support the planning activity of an agent. However, similar to analytical tools, constraint solvers represent static models of CSP. In order to achieve flexibility in the decision-making process, the solver has to be able to incorporate an agent's essential characteristics:

Reactive behavior is characterized by the agent's ability to absorb new information and restrict or relax its actions. When constraint solvers face the relaxation of constraints, they recompute the entire problem. The relaxation in constraint planners and schedulers nearly always results in reduced stability and possibly reduced quality of solutions. Despite the recent implementation of constraint solvers in graphical applications (Borning, 1995; Sadeh, 1995), with real-time computation and stability, the constraint relaxation and adaptive behavior still represent some difficulties for practical applications.

An agent's *rational behavior* and fast reaction to the changes in its environment is difficult to support with constraint solvers. Some scheduling systems incorporate this idea and extend or replace deliberative planning with behavior rules (ILOG, 1996). The majority of existing constraint solvers compute the search for solutions offline. An approach to eliminate this problem is the development of an anytime algorithm (Zilberstein, 1995) and constraint hierarchies (Zanden, 1996). Some CSPs are suitable for iterative local search techniques, such as annealing, taboo search or genetic algorithms (Kautz, 1996). Additional information on suitability can be found in Tsang (1993).

Representation of time in traditional planning systems is based on Allen's model of temporal relations (Allen, 1985). This representation does not seem to be suitable for multi-agent interaction with complex temporal relationships and commitment protocols. Problems arise with respect to concurrency and events that last over a period of time.

Social abilities mean interaction, negotiation, coordination, and/or cooperation among agents. A language or interaction protocol typically supports these social skills. The ultimate goal of cooperation and coordination is to reach a globally optimal solution independent of the language or protocol used. If we map the cooperation goals into distributed problem-solving strategies and let each agent play the role of a cooperating computational unit instead of an autonomous negotiator, it is then possible to deploy distributed constraint satisfaction problem-solving strategies.

Multiagent solutions to the distributed constraint problems require new types of design frameworks based on replaceable components that can accommodate the local autonomy of agents, several negotiation strategies, cooperative problem solving, online search for solution, and support for an agent's rational behavior.

Agent design must accommodate static agents as well as trends in mobile agents. The lightweight architecture required for mobile agents, and an efficient and flexible negotiation strategy, are not mutually exclusive. The main issue is to provide a framework that guarantees that the negotiating agent is not overloaded with the complex intelligence that may not be used. We see an agent as a lightweight core with the ability to "borrow" the intelligence it requires from the hosting environment.

The major problem with this component-based framework in which one or more constraint solvers are used as plug-in components arises from the difficulties associated with the description of the constraint satisfaction problem. The representation of the problem can be either under-constrained or over-constrained. In either case, the agent is dealing with incomplete knowledge and it cannot determine which of the generated solutions is optimal and suitable to use in further negotiation.

One of the most important criteria used to judge constraint solvers is their performance. In order to provide reactive and pro-active behavior, it is important for the solver to generate the solution quickly enough to maintain responsiveness. This becomes more difficult to achieve as the number of constraints and variables becomes larger.

A pre-built constraint solver may be able to maintain constraints efficiently. However, declarative definitions of constraints are difficult to use in high-level agent building tools.

There has been very little work in incorporating constraint solvers into reasoning and negotiation strategies of agent systems. The slow acceptance of constraint solvers seems to have been caused by four reasons:

1. Many constraint solvers only support a limited range of constraints.
2. It is difficult to understand and control constraint solvers. When the programmer sets up a set of constraints, it may not be obvious how the constraint solver will maintain them.

3. If the search space is large then online performance of constraint solvers may not be satisfactory for some application domains (e.g., Eaton, 1997).
4. Multi-agent systems are represented by highly distributed search space (distributed constraint optimization problems (DCOP)).

DCOP have been considered an important subject of research for multi-agent systems and multi-agent cooperative mediation (Mailler & Lesser, 2004). DCOP aim at finding optimal assignment to a set of variables dispersed over a number of agents that have some interdependencies. So far, the descriptions of the problem proposed include distributed partial constraint satisfaction problem (Hirayama & Yokoo, 1997), distributed valued constraint satisfaction problem (Lemaitre & Verfaillie, 1997), and DCOP (Yokoo & Durfee, 1991). Typical problem with these asynchronous protocols is the exponential increase in communication. The protocol suggested in Mailler and Lesser (2004) allows the agents to overlap the context used for their local decision making. During the mediation session, the mediator agent computes a partial solution and recommends the value assignments to other agents.

CONCLUSION AND FUTURE TRENDS

Agents are fairly complex and ambitious software systems. They will be entrusted with advanced and critical applications, such as network administration, database and application monitoring, workflow/business process support, and enterprise integration. As such, agent-based systems must be engineered with valid software engineering principles and not constructed in an ad hoc fashion.

Analytical strategies are tools based on a static mathematical model to evaluate outcomes and generate appropriate action. With *evolutionary* or *genetic approaches*, the learning process is more effective and models are adaptable. The advances in *constraint technology* enable the design and implementation of planning and scheduling tasks to be treated as constraint satisfaction problems. The agent concept then can be used to support dynamic adaptation and collaboration of local or distributed problem-solving modules.

Design and implementation of any reasoning strategies and collaboration protocols can lead to complex systems, and these issues can also lead to computer code that is difficult to maintain. The protocols of interaction are complicated and they typically have many levels or states. If a reasoning engine is very large and complex, it restricts and slows an agent's mobility, rationality, and real-time responsiveness. In recent years, patterns of agent collaboration and negotiation, and some useful components and architectures have emerged. Reusable components and generative patterns can greatly improve the quality of agent designs and implementations.

Any agent, and mobile agents in particular, should have the capability to adapt their negotiation strategy and protocol according to the tasks at hand. In order to match tasks with the appropriate negotiation and collaboration strategies, the agent should have the ability and the means to select a negotiation strategy as a component and then "plug it in" for use.

Nwana (1998) discusses the contradiction between the research in agent coordination and reasoning, and the reality of implementing real applications. He suggests that we should continue with "borrowing and consolidation" using already established AI work. This evolutionary path requires adaptable design techniques to support the trends in AI. Instead of building specific negotiation strategies and protocol cores for each problem domain, we build agents with robust, adaptable core capabilities. The agents' negotiation strategy and protocols are then components or pattern-based building blocks that are "borrowed" and matched with the task at hand.

Future trends in the design of negotiating agents will undoubtedly track those in software engineering in general. Interesting and valuable future developments are expected to appear in the areas of separation of concerns and aspect-oriented programming (AOP). Object-oriented techniques, components and patterns have revolutionized software engineering. However, as software applications get more ambitious and complicated, object technology reaches its limits, especially due to its inability to address behavior and functionality that crosscuts an object hierarchy. Research completed by Kendall (1999) has considered how the separation of concerns and AOP can be used to improve the quality of agent designs. Role modeling is another key area for future research, and this area is also detailed in Kendall (1999).

REFERENCES

Allen, J.F., & Hayes, P.J. (1985). A common-sense theory of time. *Proceedings International Joint Conference on Artificial Intelligence* (pp. 528-531).

Aridor, Y., & Lange, D. (1998). Agent design patterns: Elements of agent application design. *Proceedings of the 2nd International Conference of Autonomous Agents* (on CD). IEEE.

Borning, A. (1995). The OTI constraint solver: A constraint library for constructing interactive graphical user interfaces. *Proceedings of the 1st International Conference on Principles and Practice of Constraint Programming* (pp. 624-628).

Eaton, P.S., Freuder, E.C., & Wallace, R.J. (1997). *Constraint-based agents: Assistance, cooperation, compromise.* Technical Report. Computer Science Department, University of New Hampshire.

Fudenberg, D., & Tirole, J. (1991). *Game theory.* Cambridge: MIT Press.

Goldberg, D.E. (1989). *Genetic algorithms in search, optimisation and machine learning.* Reading, MA: Addison-Wesley.

Hirayama, K., & Yokoo, M. (1997). Distributed partial constraint satisfaction problem. In G. Smolka (Ed.), *Principles and practice of constraint programming (CP-97), Lecture Notes in Computer Science, 1330,* 222-236. Springer-Verlag.

ILOG. (1996). *ILOG SCHEDULER user's manual.*

Kautz, H., & Selman, B. (1996). Pushing the envelope: Planning, propositional logic, and stochastic search. *Proceedings of the 13th National Conference on Artificial Intelligence (AAAI-96)*, 1194-1201.

Kendall, E. (1999). Role model designs and implementations with aspect oriented programming. *Proceedings of the 1999 Conference on Object- Oriented Programming Systems, Languages, and Applications (OOPSLA'99)* (pp. 132-145). ACM Press.

Kephart, J.O., Hanson, J.E., Levine, D.W., Grosof, B.N., Sairamesh, J., Segal, R.B., & White, S.R. (1998, July 4-7). Dynamics of an information-filtering economy. *Proceedings of 2nd International Workshop on Cooperative Information Agents (CIA-98)*, Paris. Retrieved October 6, 2003, from *http://citeseer.ist.psu.edu/70606.html*

Lemaitre, M., & Verfaillie. (1997). An incomplete method for solving distributed valued constraint satisfaction problems. *Proceedings of the AAAI Workshop on Constraint and Agents* (on CD).

Mailler, R., & Lesser, V. (2004). Solving distributed constraint optimization problems using cooperative mediation. Retrieved October 4, 2003, from *ftp://mas.cs.umas.edu/pub/mailler-569.pdf*

Mass-Colell, A., Whinston, R., & Green, J.R. (1995). *Microeconomic theory.* Oxford University Press.

Nareyek, A. (1998). *Constraint-based agents.* Technical Report. German National Research Center for Information Technology, Berlin.

Nash, J. (1950). The bargaining problem. *Econometrica, 18,* 155-162.

Nwana, H.S., & Ndumu, D.T. (1998). A perspective on software agents research. *ZEUS Methodology Documentation.* British Telecom Laboratories.

Osborne, M.J., & Rubinstein, A. (1990). *Bargaining and markets.* Academic Press.

Osborne, M.J., & Rubinstein, A. (1994). *A course in game theory.* Cambridge: The MIT Press.

Sadeh, N.M., & Fox, M.S. (1995). Variable and value ordering heuristics for the job shop scheduling constraint satisfaction problem. *Technical Report CMU-RI-TR-95-39.* Carnegie Melon University.

Sandholm, T. (1993). An implementation of the contract net protocol based on marginal cost calculations. *Proceedings of the 12th International Workshop on Distributed Artificial Intelligenc*e (pp.256-262). Retrieved March 4, 2000, from *http://citeseer.ist.psu.edu/sandholm93 implementation.html*

Sen, S., & Durfee, E. (1994). The role of commitment in cooperative negotiation. *International Journal of Intelligent Cooperative Information Systems, 3*(1), 67-81.

Smith, R.G. (1980). The contract net protocol: High-level communication and control in a distributed problem solver. *IEEE Transactions on Computers, C-29*(12),1104-1113.

Tsang, E., & Borrett, P.K. (1993). *Foundation of constraint satisfaction.* Academic Press.

Yokoo, M., & Durfee, E.H. (1991). Distributed constraint optimization as formal model of partially adversarial cooperation. *Technical Report CSE-TR-101-91.* University of Michigan, Ann Arbor.

Zanden, B. (1996, January). An incremental algorithm for satisfying hierarchies of multi-way dataflow constraints. *ACM Transactions on Programming Languages and Systems, 18*(1), 30-72.

Zeng, D., & Sycara, K. (1996, August). How can an agent learn to negotiate? In J.P. Muller, M.J. Wooldgridge & N.R. Jennings (Eds.), *Intelligent agents III: Agent theories, architectures, and languages. Proceedings of European Conference on Artifical Intelligence'96.* LNCS. Springer.

Zilberstein, S., & Russell, S. (1995). Approximate reasoning using anytime algorithms. In S. Natarajan (Ed.), *Imprecise and approximate computation.* Kluwer Academic Publishers.

KEY TERMS

Agent: A program designed to provide specialized and well-defined services. An agent can be *static* – executing on the computer where it was installed, or *mobile* – executing on computer nodes in a network.

Bayesian Method: Means of quantifying uncertainty, based on the probability theory. The method defines a rule for refining a hypothesis by factoring in additional evidence and background information. It uses results of previous events to predict results of future events.

Evolutionary Game Theory: Study of equilibria of games played by population of players where the “fitness” of the players derives from the success each player has in playing the game. It provides tools for describing situations where a number of agents interact. Evolutionary game theory improves upon traditional game theory by providing dynamics describing how the population will change over time.

Genetic Algorithm: Class of algorithms used to find approximate solutions to difficult-to-solve problems, inspired and named after biological processes of inheritance, mutation, natural selection, and generic crossover. Genetic algorithms are a particular class of evolutionary algorithms.

Negotiation Mechanism: The actual implementation of negotiation strategy and negotiation protocol.

Negotiation Strategy: Specification of the sequence of actions the agent intends to make during the negotiation.

Paretto Efficient Agreement: State of things in which no player (in game theory) is worse off than the others. It typically refers to the distribution of resources. The agreement could lead to cooperation of players.

Reasoning Engine: A computer program that enables the agents to negotiate with other agents and that involves negotiation strategy.

Designing High Performance Virtual Teams

Ian K. Wong
Queen's University, Canada

D. Sandy Staples
Queen's University, Canada

INTRODUCTION

In the past several decades, we have seen tremendous advancements in the development of communication technology. Since the invention of the Internet in 1969, there has been rapid development of Internet-based communication tools and technologies. This technology has revolutionized business practices by offering another important and effective channel for communication (Foo & Lim, 1997), and has allowed people to work on projects irrespective of their physical location. One resulting business practice that has been adopted in recent years is virtual teamwork. Virtual teams are groups of individuals who work at interdependent tasks, who share responsibility for outcomes, and who work together from different locations. Recently, the use of teams as fundamental building blocks in organizations is increasing, as is the use of virtual teamwork (Furst, Blackburn & Rosen, 1999). This article identifies the characteristics of high performing virtual teams.

BACKGROUND

In addition to the basic definition of a virtual team, all virtual teams have important characteristics that contribute to their overall success. To analyze the characteristics of a team's situation, Cohen's (1994) model of team effectiveness can be used as an organizing framework. The model identifies strengths and weaknesses that readers can use to inform their own design and operations of effective virtual teams. According to Cohen, there are several broad characteristics that all potentially affect how successful the team will be at meeting its task, and are therefore worthy of examination. These characteristics are listed in Table 1 and will be examined in detail in the following paragraphs. Although Cohen's team effectiveness model is based on traditional teams (i.e., collocated), these characteristics have been found to be very important in empirical research on virtual teams (Staples & Cameron, 2004; Wong & Staples, 2004).

TASK DESIGN

Appropriate task design can be a powerful motivator (Cohen, 1994). Both job characteristics theory (e.g., Hackman & Oldman, 1976, 1980) and sociotechnical theory (e.g., Cummings, 1978) suggest that group task design is critical for employee motivation, satisfaction, and performance. Both theories suggest that to positively impact performance and attitudes, the task should be designed according to the criteria specified in Table 2.

Job characteristics theory, which has fairly strong empirical support, suggests that task attributes influence effectiveness through their impact on critical psychological states such as motivation and satisfaction with the work. For example, in a case study of one particular business development virtual team, team members commented that high satisfaction and motivation levels reflected the high perceived significance of the project (Wong & Staples, 2004). Positive motivation and satisfaction levels have a positive effect on the quality of the work and overall productivity of the team (i.e., an indirect effect exists between task design and productivity and quality) (Cohen, 1994). Also, the team must have autonomy in determining how their work will be done, because autonomy enhances worker attitudes, behaviors, and performance (Cohen & Bailey, 1997). Finally, when a remote worker receives managerial feedback in the form of advice and help, the worker's effectiveness increases (Staples, 2001). This would result in an increase in virtual team performance.

Table 1. Characteristics of virtual teams affecting team effectiveness

- Design of the team's task
- The characteristics of the members of the team
- The processes used by the team
- The organizational context of the team

CHARACTERISTICS OF THE TEAM AND ITS MEMBERS

Team member characteristics that influence the success of a virtual team are listed in Table 3 and are described in more detail below (Cohen, 1994).

The size of the team can affect the ability of the team to do its task (Cohen, 1994). If the team is too big, higher coordination costs result. If the team size is too small, it will not have the resources needed to complete its work, and team members will be less likely to be committed to the team. The size of the team should also correspond to the stage of the project. For example, a virtual team developing a new product may need more human resources as the product moves from the design stage into the manufacturing stage.

Stability of team membership is necessary for team effectiveness. If turnover is high, time and effort will be spent orientating new members, performance norms will not develop, and performance will suffer. However, some turnover can be beneficial, in that it could revitalize a stagnant team and enhance creativity (Cohen, 1994).

The collective knowledge and skills of a team will greatly impact the team's ability to carry out its task. Such skills include technical skills, information systems (IS) skills, and interpersonal skills. Information systems skills are needed to use the information technology tools and systems that are available to communicate virtually and share information virtually, which is the norm given the lack of face-to-face interaction in virtual teams. Effective communication skills among team members are also vital to the effectiveness of a virtual team (Grenier & Metes, 1995).

The degree of virtuality (degree of team geographic distribution) could contribute to team effectiveness. There has been limited empirical evidence to suggest that greater geographic distribution (high degree of virtuality) of a team leads to lower performance (Cramton, 2002). This is presumably due to reduced face-to-face contact, reduced opportunities to build social relationships, and the difficulties of communicating and coordinating virtually using communication technology rather than communicating face-to-face. This implies that higher virtuality could be negatively related to team performance and satisfaction with the work and the team.

Team performance beliefs have been found to be a strong predictor of group effectiveness in previous research (Cohen, 1994). For example, team beliefs, assessed via a concept called group potency, were found to be positively related to the commitment to the team, satisfaction with being part of the team, and motivation with the team's tasks (Staples & Cameron, 2004). Group potency captures efficacy beliefs at the group level. Group potency (sometimes referred to as group efficacy) is "a collective belief in the capability of the group to meet a task objective" (Gibson, Randel & Earley, 2000, p. 71).

TEAM PROCESSES

There are several behavioral characteristics pertaining to team process that positively affect team effectiveness.

Table 2. Task design criteria necessary to positively impact performance and attitudes

The task should be designed such that:

- A variety of skills are required (leadership, communication, different technical skills, etc.) such that a team of people are needed to work together to complete the overall task.
- A whole and identifiable piece of work exists so that members can see the outcome of their efforts.
- It is perceived to have significant impact on the lives of other people so that team members feel their work is important and are motivated to complete the task.
- The team has considerable autonomy and independence in determining how the work will be done so that team members feel empowered and responsible for their actions.
- The team is provided with regular and accurate feedback such that the team can understand how it is performing and make adjustments as needed.

Table 3. Characteristics of the team members that affect team effectiveness

- The size of the team
- The stability of the team, in terms of turnover
- The skills of the members of the team
- The relative locations of the team members (i.e., their virtualness)
- The team members' beliefs about their team's capabilities

These are coordination, caring (i.e., team spirit), sharing of expertise, and effectiveness of communications. According to Cohen and Bailey (1997), how team members coordinate is an important characteristic of a team. Since team members are interdependent on each other to get their work done, coordinating interdependent tasks is critical. Good coordination among team members leads to working together without duplication and wasted efforts. Caring about each other implies working together with energy and team spirit. This can motivate team members and foster commitment to the team goals, resulting in higher performance. Sharing and benefiting from others' knowledge and expertise is also important because it supports effective cross-training and decision making to fulfill interdependencies. Good communication processes are required in order to make this possible. Effective communication processes were important in building a successful virtual team according to 84% of team members interviewed in six case studies of virtual teams (Staples, Wong & Cameron, 2004). Similarly, Cohen (1994) found that most high-performing teams are teams in which members are comfortable sharing information and communicate well. These team process variables are part of most models of team effectiveness and have been found to be associated with group effectiveness in previous research (Cohen, 1994).

ORGANIZATIONAL CONTEXT

Lastly, the organizational context that a team works in can create the conditions for a team to be successful or for it to fail (Cohen, 1994). The team with the best internal processes may still perform poorly if it lacks the resources or information needed to do its task. A team will not be able to make good decisions without proper information, without sufficient training, and without adequate resources. The key variables that potentially interact to create an environment where the employee wants to be involved and can participate to complete their tasks effectively are listed in Table 4 and explained below.

According to Duarte and Snyder (2001), human resource policies must be designed and integrated in such a way that virtual team members are recognized, supported, and rewarded for their work. Cohen, Ledford, and Spreitzer (1996) found that management recognition was positively associated with team ratings of performance, trust in management, organizational commitment, and satisfaction for both self-directed and traditionally managed groups in organizations. As such, it is important that an effective reward system with performance measures is in place to reward results. Grenier and Metes (1995) and Lurey and Raisinghani (2001) also suggest that it is important for the organization to reward high levels of team performance. If rewards are solely at the individual level, this does not stimulate the completion of interdependent tasks and the sharing of information and expertise.

Next, it is important that team members have access to continual online training and technical support (Duarte & Snyder, 2001; O'Hara-Devereaux & Johansen, 1994). Training and team building is important because it ensures that employees develop the knowledge required to contribute to organizational performance (Cohen, 1994). It should be available to enable employees to develop the necessary skills and knowledge required to complete their tasks at hand. In addition, providing training to team members shows management's commitment to continual growth and development for members.

Management support and the power structure are also key variables in attaining virtual team effectiveness. Teams not only need to be well budgeted and resourced, but team members also want encouragement and symbolic gestures of appreciation. Symbolic gestures—such as a pat on the back or having management verbally communicate that the team is doing a good job¾demonstrate respect for the team members; this can positively motivate the members. With respect to power structure, it has been shown that team members must have independence and decision-making capability to be successful in virtual groups (Lipnack & Stamps, 1997). In addition, according to Cohen and Bailey (1997), an organization needs to provide team

Table 4. Organizational context factors that affect team effectiveness

- Human resource policies that emphasize and reward teamwork
- Human resource policies that do not penalize (i.e., overlook) virtual workers
- The availability of training to work virtually, work effectively in teams, and carry out task responsibilities
- The amount of support management provides to the team
- The degree of autonomy the team is given to make decisions and take responsibility for their own actions
- The ability of the organization to share information that is needed to carry out tasks and make effective decisions

members autonomy in their work. Worker autonomy has been shown to have clear benefits because it enhances worker attitudes, behaviors, and performance (whether measured objectively or rated subjectively by team members).

Finally, access to information is necessary for team members to effectively complete their tasks. A practical study of a project-based virtual team suggested that most of the information team members needed was held either by themselves or other team members (Wong & Staples, 2004). Therefore, good communication between team members is essential for information to be easily accessed. In addition, resources need to be available for virtual communication to be possible and effective. These resources include information technology infrastructure and information technology tools that are needed to communicate and share information electronically in the virtual setting.

FUTURE TRENDS

The need to compete in a rapidly changing, hypercompetitive, and global marketplace is prompting many organizations to transform their organizational structures from large, hierarchical structures to agile, flexible, new structures (Morris, Marshall & Rainer, 2001). Business is increasingly being done at an international level, and organizations are responding by creating cross-border teams. Consequently, we will see a continuing trend of virtual teams emerging that have team members less geographically concentrated and more geographically dispersed. It remains to be seen how organizations will deal with the increase in geographic dispersion. As shown in research (Cramton & Webber, 1999), higher team member geographic dispersion may lead to lower team performance. Therefore, future research may focus on how to maintain team performance at a high level for teams that have a high degree of virtualness.

Furthermore, management will be continually trying to adopt new communication tools and technologies that have improvements in quality and capabilities. As such, training to use new technology will become especially important; how well team members can communicate and interact with each other depends very much on their ability to properly use the communication tools. Future research may deal with the effectiveness of new, emerging technologies and how they may positively or negatively affect virtual team performance.

CONCLUSION

The success of a virtual team depends heavily on its characteristics. Specifically, the main characteristics are the team's task design, the characteristics of the team members, the processes used by the team, and the organizational context. Analyzing these characteristics can reveal a great deal about how well a team is functioning, and what potential it has to improve. An organization's management can also use these characteristics to effectively design its virtual teams and thereby give it the best chance possible for success. Therefore, it is in management's best interests when putting together a virtual team to carefully consider the characteristics of the team. This will increase the team's chance of success if the variables that yield positive outcomes are maximized.

REFERENCES

Cohen, S.G. (1994). Designing effective self-managing work teams. In M.M. Beyerlein, D.A. Johnson & S.T. Beyerlein (Eds.), *Advances in interdisciplinary studies of work teams, volume 1, series of self-managed work teams*. Greenwich, CT: JAI Press.

Cohen, S.G., & Bailey, D.E. (1997). What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management, 23*(3), 239-290.

Cohen, S.G., Ledford, G.E., & Spreitzer, G.M. (1996). A predictive model of self-managing work team effectiveness. *Human Relations*, *49*(5), 643-676.

Cramton, C.D. (2002). Finding common ground in dispersed collaboration. *Organizational Dynamics, 30*(4), 356-367.

Cummings, T.G. (1978). Self-regulating work groups: A sociotechnical synthesis. *Academy of Management Review*, *3*(3), 625-634.

Duarte, D.L., & Snyder, N.T. (2001). *Mastering virtual teams: Strategies, tools, and techniques that succeed*. San Francisco, CA: Jossey-Bass.

Foo, S., & Lim, E. (1997). A hypermedia database to manage World Wide Web documents. *Information and Management, 31*(5), 235-249.

Furst, S., Blackburn, R., & Rosen, B. (1999, August). Virtual teams: A proposed research agenda. *Proceedings of the Academy of Management Conference*, Chicago, IL.

Gibson, B.G., Randel, A.E., & Earley, P.C. (2000). Understanding group efficacy: An empirical test of multiple assessment methods. *Group & Organizational Management, 25*(1), 67-97.

Grenier, R., & Metes, M. (1995). *Going virtual*. Upper Saddle River, NJ: Prentice-Hall.

Hackman, J.R., & Oldman, G.R. (1976). Motivation through the design of work: Test of a theory. *Organizational Behavior and Human Performance, 16,* 250-279.

Hackman, J.R., & Oldman, G.R. (1980). *Work redesign*. Reading, MA: Addison-Wesley.

Lipnack, J., & Stamps, J. (1997). *Virtual teams: Reaching across space, time, and organizations with technology*. New York: John Wiley & Sons.

Lurey, J., & Raisinghani, M. (2001). An empirical study of best practices in virtual teams. *Information & Management, 38*(8), 523-544.

Marks, M.A., Mathieu, J.E., & Zaccaro, S.J. (2001). A temporally based framework and taxonomy of team processes. *The Academy of Management Review*, *26*(3), 356-376.

Morris, S., Marshall, T., & Rainer, R. (2001). Impact of user satisfaction and trust on virtual team members. *Information Resources Management Journal*, *15*(2), 22-30.

O'Hara-Devereaux, M., & Johansen, R. (1994). *Global work: Bridging distance, culture & time*. San Francisco, CA: Jossey-Bass.

Staples, D.S. (2001). Making remote workers effective. In N.J. Johnson (Ed.), *Telecommuting and virtual offices: Issues & opportunities* (ch. 11, pp. 186-212). Hershey, PA: Idea Group Publishing.

Staples, D.S., & Cameron, A.F. (2004). Creating positive attitudes in virtual team members. In S. Godar & P. Ferris (Eds.), *Virtual & collaborative teams: Process, technologies, & practice* (pp. 76-98). Hershey, PA: Idea Group Publishing.

Staples, D.S., Wong, I.K., & Cameron, A.F. (2004). Best practices for virtual team effectiveness. In D. Pauleen (Ed.), *Virtual teams: Projects, protocols and processes* (pp. 160-185). Hershey, PA: Idea Group Publishing.

Wong, I., & Staples, D.S. (2004). A virtual team in action—An illustration of a business development virtual team. In D. Pauleen (Ed.), *Virtual teams: Projects, protocols and processes* (pp. 91-114). Hershey, PA: Idea Group Publishing.

KEY TERMS

Group Potency: A collective belief in the capability of the group to meet a task objective.

Job Characteristics Theory: Task attributes influence effectiveness through their impact on critical psychological states such as motivation and satisfaction with the work.

Organizational Context: The conditions within the organization that a team works in that influences the successfulness of the team's activities and the involvement of the team members. Reward systems, level of management support, resources provided, and organizational culture all are important organizational factors that potentially affect a team's ability to succeed.

Task Design: The way key attributes of the task are arranged, in terms of the influence of these attributes on the effectiveness of a team in performing the task. Research has found key design attributes include the need for a variety of skills, the perceived importance of the task, the independence and autonomy given to people to determine how the task will be done, and the way task feedback is provided.

Team Characteristics: The composition of the team and the shared beliefs held within the team about the team. Team composition includes such things as the number of team members, the skills the team members collectively possess, and the stability of team membership.

Team Effectiveness: The ability of a team to perform its tasks on time, on budget, and with acceptable quality, as well as the satisfaction, motivation, and commitment of the team members.

Team Processes: "Members' interdependent acts that convert inputs to outcomes through cognitive, verbal, and behavioral activities directed toward organizing taskwork to achieve collective goals" (Marks, Mathieu & Zaccaro, 2001, p. 357).

Virtual Team: A group of individuals who work at interdependent tasks, who share responsibility for outcomes, and who work together from different locations.

Virtuality/Virtualness: The degree to which team members are geographically distributed such that opportunities for meeting informally and/or face-to-face are reduced.

ENDNOTE

1 As one reviewer pointed out, geographic dispersion is just one type of variety that may affect team performance. Other types of diversity include team member values (reflecting different organizational and/or national cultures), backgrounds, and mental models. Research on the optimal mixes of different types of diversity in virtual teams would be very valuable.

Designing Hypertext and the Web

Yin-Leng Theng
Nanyang Technological University, Singapore

INTRODUCTION

Hypertext and Web technology still continue to excite many, ever since it became popular in the late 1980s. Early visionary thinkers such as Bush (1945), Engelbart (1963), and Nelson (1987) in their own ways believed that hypertext and its underlying philosophy would enable man to locate, retrieve, and easily use the store of human knowledge that lies in books, journals, and associated materials in libraries all over the world.

In hypertext, end users not only benefit from the information they read, but also from the richness of associations supported by the network of nodes and links. Hypertext has affected us directly or indirectly in almost every facet of our lives, ranging from scientific work to business and education needs, to our general way of life. For example, the most widely used hypertext system since 1994 is the World Wide Web (or simply the Web) on the Internet.

Though hypertext systems promise a whole new paradigm and freedom of reading and information retrieval, they are not without problems. Ironically, it is precisely this freedom and power with which hypertext systems equip end users that give rise to many of these problems.

BACKGROUND: BRIEF HISTORY OF HYPERTEXT AND THE WEB

When Vannevar Bush (1945) envisioned his hypertext "memex," he dreamed of a personal microfiche-based system that would help him tackle the problem of information overload at that time. His vision of the "memex" heralded the beginning of a search for a system that mimics the human mind to access information quickly and intuitively by reference. In 1965, Nelson (1987) coined the term "hypertext" and presented it as a radical new way of structuring textual information into non-sequential format, a computer-based incarnation of Bush's dream "memex" (Berk & Devlin, 1991). Even though the technology in the "early" years of the hypertext history (for example, Conklin, 1987; Nielsen, 1995) was not sophisticated enough for many of the ideas to be realized, hypertext pioneers staunchly believed that hypertext technology had something special to offer.

Not surprisingly, as the years roll by, hypertext systems grow more sophisticated and computer is the technology that has enabled the concept of hypertext to be seen and not just heard. The growing popularity of the Internet and advancements in networking saw the birth of networked hypertext systems such as the Web. The Web project initiated in 1990 was originally created as an online information tool for high-energy physics research at CERN (the European Center for Nuclear Physics Research in Geneva, Switzerland). Berners-Lee and colleagues, the originators of the Web, built it based on the hypertext paradigm. Based on its likeness to a spider's web, this world of hypertext links is also called the Web.

Although the Web was first made available in 1991, it was only after the release of Mosaic by the National Center for Supercomputing Applications (NCSA) in January 1993 that it really gained prominence. Mosaic, NCSA's Web client, made the Web accessible to a wide and diverse user community because of its easy-to-use, graphical interface. Mosaic and the Web succeeded in establishing a universal hypertext. With the release of Netscape Navigator in 1994 by a commercial company co-founded by the original author of Mosaic, the number of end users on the Internet escalated to a phenomenal figure. Today, the Web is used by millions across the world. It has changed the Internet to the extent that it has become almost synonymous with the modern use of the Internet.

Since the same usability issues arise in hypertext, hypermedia, multimedia, and the World Wide Web, we will call them all "Web" for conciseness.

CONTINUING DESIGN AND USABILITY PROBLEMS

Since Halasz's seminal paper at the Hypertext'87 Conference, seven design issues in hypertext systems were identified; these issues formed the nucleus of multiple research agendas in: (a) search and query; (b) composites; (c) virtual structures; (d) computation in/over hypertext network; (e) versioning; (f) collaborative work; (g) extensibility; and (h) tailorability.

Although these issues were revisited several times over the intervening 15 years, and many research systems have attempted to address the original seven issues (Whitehead et al, 2002), we are still *not* producing better, usable hypertext applications.

With the exponential growth of the Web in 1994 and 1995, these issues became particularly acute. Nielsen (1996) listed 10 top mistakes in Web design, including: (1) use of frames; (2) bleeding edge technology; (3) scrolling text, marquees, constantly running animations; (4) complex URLs (Uniform Resource Locators); (5) orphan pages; (6) long scrolling pages; (7) lack of navigation support; (8) non-standard link colors; (9) outdated information; and (10) overly long download times. When Nielsen (1999) revisited these mistakes three years later, he concluded that all 10 mistakes were still mistakes. Apart from scrolling long pages, which is causing fewer navigation problems, the other nine mistakes still cause significant usability problems and should be avoided in modern Web sites.

A survey of current work (e.g., Wu, Meng, Yu & Li, 2001) seems to suggest one or more of the following problems still persist in hypertext/Web applications: (a) structured search for documents is not sufficiently supported; (b) proving properties of Web sites is difficult or impossible; (c) maintenance is not supported by formal mechanisms; and (d) personalizing of information, or adaptation to user groups, is difficult or impossible.

The question we want to ask is: Could it be possible that these well-intentioned efforts are not achieving their aims because wrong or inappropriate solutions are being sought based on incorrect or incomplete assumptions?

If it is a *psychological* problem, then it may be entirely due to end users' inability to exploit computer screens, complex information structures, and that nothing in the design is going to ameliorate this. Thus, as a psychological problem, it can be alleviated but not solved by better design (Theng, 1997).

If it is an *engineering* problem, it can be attributable to bad system design, and poor design causes psychological problems. The "blame," therefore, should not rest on end users alone!

NEW DESIGN AND USABILITY PROBLEMS

The world is becoming a global marketplace with end users from across the world. However, according to Galdo and Nielsen (1996), this heightened interest in internationalization and localization has not yet been translated into increased usability for international end users.

Globalization brings with it new issues relating to culture and ethics. With respect to design, standard human-computer interaction (HCI) guidelines may not be adequate. They may need to be modified, extended, or restricted.

Culture and Design

Culture is defined as "learned behavior consisting of thoughts and feelings" (Galdo & Nielsen, 1996). There is little provision in hypertext/Web applications to cater to end users' browsing and inter-cultural needs (Theng, Duncker, Mohd-Nasir, Buchanan & Thimbleby, 1999). One reason for the neglect of cultural aspects may be that usability failure is rather commonplace, and cultural usability issues are hard to recognize as such, more so since designers cannot help but see the world from their particular cultural point of view and may be defensive about their work. Thus, cultural usability issues for system designers may come disguised as illiteracy problems or simply as "user faults," rather than as cultural differences.

Apart from colors (Duncker, Theng & Mohd-Nasir, 2000), cultural factors that can influence the design of interactive interfaces for international end users include the following: spoken and written languages; the reading/writing direction; meanings and preferences of colors; and interpretations of signs, pictures, symbols used (Galdo & Nielsen, 1996).

In order to produce usable and useful interactive systems, designers need to ensure that good design features are incorporated into the systems, taking into consideration end users' cultural preferences.

Ethics and Design

According to the Oxford Dictionary, ethics is the "study of right and wrong in human behavior." The notion of the code of ethics and professional conduct is not new. The general moral imperatives in the ACM (Association for Computing Machinery) code of Ethics and Professional Conduct (adopted by the ACM Council, October 16, 1992) include one's contribution to society and human well-being, avoidance of harm to others, honesty and trustworthiness, fairness and non-discriminate actions, respect for property rights, privacy, and confidentiality. A specific professional responsibility in the ACM Code of Ethics is to "give comprehensive and thorough evaluations of computer systems and their impacts, including analysis of possible risks." "Ensuring that users have their needs clearly articulated during assessment and design of requirements" is one of the organizational leadership moral imperatives.

Apart from ensuring that the technical aspects of the Internet are taken care of, designers need also to consider the ethical and legal issues that may arise, owing to the increase in the use of computer-mediated communication devices (Gringras, 1997).

In recent years, Cyberethics, a growing area of research, explores the application of classical ethics to the latest information communication technologies, including the Internet. The fundamental principles of Cyberethics are privacy, property, access, and accuracy (Barroso, 2001). They imply a new set of responsibilities for both professionals and users of the Internet, and have implications for the design of safe, usable, and useful systems.

Duquenoy and Thimbleby (1999) explore the idea of justice to aid improved design. John Rawls' classic *A Theory of Justice* is used to explain why design needs to consider these two principles: liberty and equality (Rawls, 1972). Applying the Principle of Liberty, designers should ask themselves whether their designs persecute, discriminate, and oppress users (Duquenoy & Thimbleby, 1999).

Berners-Lee (1996) at the British Computer Society, in which he was awarded the Distinguished Fellowship, ended the presentation with a somber thought for designers/developers that they were wrong to ignore ethics "because when you design those protocols, you are designing the space in which society will evolve. You are designing, constraining the society which can exist for the next 10-20 years."

NEW DESIGN APPROACHES

Past solutions to address hypertext problems failed to take into consideration the context of use of the systems. Recent promise of the "Semantic Web" has also raised new expectations in that the Web should be (Marshall & Shipman, 2003): (1) a universal library readily accessed and used by humans in a variety of information use contexts; (2) the backdrop for the work of computational agents completing sophisticated activities on behalf of their human counterparts; and (3) a method for federating particular knowledge bases and databases to perform anticipated tasks for humans and their agents.

We propose a shift in mindset to embrace both a conscientious attempt designing with the *heart* to put end users at the center of design incorporating new issues arising from cultural and ethical perspectives, as well as an engineering approach involving the *mind* working with end users.

DESIGNING WITH THE HEART

To provide multi-cultural interfaces to hypertext/Web applications, we envisage the development of "boundary objects" between different cultures accessing shared information resources. "Boundary objects," from a social science perspective, are used by different communities without presupposing a fully shared definition of an object (Star & Griesemer, 1989). To achieve the emergence of inter-cultural boundary objects in hypertext/Web applications, cooperative and communicative features need to be introduced that allow negotiation and articulation across sites. Some ideas for implementation of boundary objects in three areas to create culturally sensitive user interfaces are (Theng, 2003; Theng et al., 1999): (1) creation of boundary objects as part of the hypertext/Web interface; (2) creation of a learning environment; and (3) creation of opportunities to create learning objects by users.

To enable equal access/opportunity/use, we suggest some ideas for implementation to create ethically sensitive user interfaces (Theng, 2003): (i) provision of adaptive and adaptable interfaces for users; (ii) provision of workspace and equal opportunity for search results; and (iii) integration of technology into the social environment.

To capture explicit and tacit knowledge occurring in a continuous, spiral manner, knowledge management proponents such as Nonaka and Konno (1998) suggest management of both explicit knowledge (information) as well as tacit knowledge. For richer exchange of user interactions in hypertext/Web applications, perhaps hypertext/Web applications should: (1) create opportunities to share experiences; (2) articulate tacit knowledge in some "externalization" form; and (3) create ways to integrate different types of knowledge and re-package them into different forms.

DESIGNING WITH THE MIND

To address limitations in current research in hypertext/Web applications, a systematic, engineering approach addressing fundamental issues that concern the *mind* should be considered (Theng, 2003, 1997):

- **For What Purpose?** It is important that designers identify the goals of hypertext/Web applications from the perspectives of the author/developer as well as from the reader/end user. Author/developer goals refer to authoring strategies to consistently apply basic Web document design principles on every single page in the hypertext/Web applications designers create. Reader/end user goals refer to the tasks readers/end users commonly perform when using hypertext/Web applications.
- **How To?** To design user-centered hypertext/Web applications, we need to involve end users and their needs throughout the different stages in the well-accepted iterative development lifecycle: (i) feasibility study; (ii) conceptual design; (iii) building; (iv) implementation; (v) integration; and (vi) maintenance.
- **With What?** This refers to the resources available for the development and delivery environments. They include: (i) data; (ii) rules; and (iii) authoring

tools. Data refers to digitized objects (e.g., text, images, sounds, videos) used to build hypertext/Web applications. They can also refer to existing collections of related information found, for example, in archives or on the Web. Rules are defined as ways in determining data structures of the digitized objects, that is, how data are to be captured and presented. Authoring tools refer to the authoring help and environment provided for designers/developers of hypertext/Web applications. These tools should make authoring as well as maintenance of hypertext/Web applications as easy as possible.

- **Level of Abstraction?** The essence of hypertext/Web applications is the representation of ideas through chunks of text or other media interconnected by relations, typically navigational links similar to knowledge representations such as frames and semantic nets. Instead of viewing knowledge expression as an all at once event, Shipman, Moore, Maloor, Hsieh, and Akkapeddi (2002) advocate viewing it as a constructive process in knowledge building. Perhaps the description of hypertext/Web applications can be at three levels: (i) *instance* level, which addresses specific application and domain needs; (ii) *modeling* level, which addresses the issues of how models for specific domain are created; and (iii) *meta-modeling* level, which addresses the issues of how to go about creating a model.

FUTURE TRENDS AND CONCLUSION

Many hypertext/Web applications are still poorly designed and built in terms of how information is structured and displayed. We argue for a move from treatment to prevention, from treating the end users' symptoms to avoiding bad design. This article makes several theoretical and practical contributions to address design of hypertext/Web applications with the *heart* and *mind*.

In the conventional iterative design process, there is great emphasis to encompass all activities relating to gathering and analyzing end users' needs, specifying end users' requirements, and building and testing prototypes. This is uncontentious wisdom, but so far, doing this is not easy since using these established usability techniques requires a competent level of "craft skills," and applying them is usually laborious and time consuming.

With more informed and sophisticated end users, there is greater awareness and concern that systems should be user focused. Future developments would see more emphasis on Web usability engineering that focuses on integration of creative and practical-to-use design and usability evaluation aids throughout the traditional iterative design process to help designers and developers.

ACKNOWLEDGEMENTS

The author would like to thank colleagues and friends who had given invaluable feedback and support: Harold Thimbleby, Cecile Rigny, Matthew Jones, Norliza Nasir, Elke Duncker, and Paola Ameldi.

REFERENCES

Barroso, P. (2001). Cyberspace: Ethical problems with new technology. In H. Krawczyk, B. Wiszniewski, S. Szejko, S. Rogerson & T. Bynum (Eds.), *Ethicomp2001*. Retrieved January 31, 2002, from *www.ccsr.cse.dmu.ac.uk/conferences/ccsrconf/ethicomp2001/index.html*.

Berk, B. & Devlin, J. (1991), Why hypertext? In B. Berk & J. Devlin (Eds.), *Hypertext/hypermedia handbook* (pp. 9-11). Armadillo Associates.

Berners-Lee, T. (1996). The World Wide Web—past, present and future. *Journal of Digital Information*, *1*(1).

Bush, V. (1945). As we may think. *Atlantic Monthly*, *7*, 101-108.

Carroll, J. (2000). *Making use: Scenario-based design of human-computer interactions*. Boston: MIT Press.

Conklin, J. (1987). Hypertext: An introduction and survey. *IEEE Computer*, 17-41.

Duncker, E., Theng, Y.L. & Mohd-Nasir, N. (2000). Cultural usability in digital libraries. *Bulletin of the American Society for Information Science*, *26*(4).

Duquenoy, P. & Thimbleby, H. (1999). Justice and design. In M. Sasse & C. Johnson (Eds.), *Interact'99* (pp. 281-286).

Galdo, E.M. & Nielsen, J. (1996). *International user interfaces*. New York: John Wiley & Sons.

Gringras, C. (1997). *The laws of the Internet*. Boston: Butterworth Heinemann.

Landauer, T. (1995). *The trouble with computers: Usefulness, usability and productivity*. Boston: MIT Press.

Marshall, C. & Shipman, F. (2003). Which Semantic Web? *Proceedings of the 14th ACM Conference on Hypertext and Hypermedia* (pp. 57-66). ACM Press.

Nelson, T.H. (1987). *Literary machines. (Ver. 87.1)*. South Bend, IN: The Distributors.

Nielsen, J. (1995). *Multimedia and hypertext: The Internet and beyond*. AP Professional.

Nielsen, J. (1996). *Ten top mistakes in Web design*. Retrieved October 7, 2001, from *www.useit.com/alertbox/9605.html*

Nielsen, J. (1999). *'Top ten mistakes' revisited three years later*. Retrieved October 7, 2001, from *www.useit.com/alertbox/990502.html*

Nonaka, I. & Konno, N. (1998). The concept of "ba": Building a foundation for knowledge creation. In J.W. Cortada & J.A. Woods (Eds.), *The knowledge management yearbook 1999-2000* (pp. 37-51). Boston: Butterworth Heinemann.

Rawls, J. (1972). *A theory of justice*. Oxford University Press.

Shipman, F., Moore, J., Maloor, P., Hsieh, H. & Akkapeddi, R. (2002). Semantics happen: Knowledge building in spatial hypertext. *Proceedings of the 13th ACM Conference on Hypertext and Hypermedia* (pp. 25-34). ACM Press.

Star, S.L. & Griesemer, J.R. (1989). Institutional ecology, translations and boundary objects: Amateurs and professionals in the Berkeley's Museum of Vertebrate Zoo. *Social Science, 19*, 387-420.

Theng, Y.L. (1997). *Addressing the 'lost in hyperspace' problem in hypertext*. PhD Thesis: Middlesex University.

Theng, Y.L. (2003). Designing hypertext and the Web with the heart and the mind. In P. Bommel (Ed.), *Information modeling for Internet applications* (Chapter XIV, pp. 299-319). Hershey, PA: Idea Group Publishing.

Theng, Y.L., Duncker, E., Mohd-Nasir, N., Buchanan, G. & Thimbleby, H. (1999). Design guidelines and user-centred digital libraries. In S. Abiteboul & A. Vercoustre (Eds.), *Lecture Notes in Computer Science: Research and Advanced Technology for Digital Libraries, Third European Conference ECDL 1999* (pp. 167-183). Berlin: Springer-Verlag.

Whitehead, J., De Bra, P., Grønbæk, K., Larsen, D., Leggett, J. & Schraefel, M. (2002). Seven issues, revisited. Panel discussion. *Proceedings of the 13th ACM Conference on Hypertext and Hypermedia* (pp. 171-171). ACM Press.

Wu, Z., Meng, W., Yu, C. & Li, Z. (2001). Towards a highly scalable and effective metasearch engine. *Proceedings of the 10th International World Wide Web Conference* (pp. 286-395). ACM Press.

KEY TERMS

Hypermedia: A term used to describe the interlinking of textual information and other forms of media, such as audio, video, and photographic images.

Hypertext: A term conceived and coined by Ted Nelson, who described it as interactive branching of text information structured into non-sequential format into nodes and links. The non-linear nature of hypertext provides freedom to readers who enjoy associational thinking and reading.

Information: Commonly known as a collection of facts or data. In Computer Science, it refers to processed, stored, or transmitted data. In Knowledge Management, it refers to codified knowledge.

Knowledge: Includes but is not limited to those descriptions, hypotheses, concepts, theories, principles, and procedures which to a reasonable degree of certainty are either true or useful. In the growing discipline in Knowledge Management, "knowledge" is divided into explicit (information), implicit, and tacit knowledge.

Multimedia: A term used to reflect diverse platforms for communicating ideas and meaning through a mix of media information that may include text, audio, video, and still pictures.

Usability: ISO 9241-11 defines usability as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use." Usability of hypertext/Web is commonly measured using established usability dimensions covering these categories of usability defects such as screen design, terminology and system information, system capabilities and user control, navigation, and completing tasks.

Usefulness: This is debatable. Some make the distinction between usability and usefulness. Although it is impossible to quantify the usefulness of a system, attempts have been made to measure its attainment in reference to system specifications and the extent of coverage of end users' tasks supported by the system, but not on end user performance testing.

World Wide Web: Originally created as an online information tool for high-energy physics research at CERN (the European Center for Nuclear Physics Research in Geneva, Switzerland), information is organized as a series of documents referring to each other with links of search and retrieval of text, images, sound, and video. Based on its likeness to a spider's web, this world of hypertext links is also called the Web.

Designing OMIS-Based Collaboration for Learning Organizations

Kam Hou Vat
University of Macau, Macau

INTRODUCTION

Today, the view that knowledge is a valuable organizational resource has become widely recognized and accepted in the business community. This is largely due to the emergence of the knowledge-based economy (OECD, 1996), characterized by a highly competitive and turbulent business environment. One consequence is the increase in organizations' efforts to deliberately manage knowledge. Organizations are realizing that their competitive edge is mostly the intellectual capital (brainpower) (Stewart, 1997) of their employees, and they are particularly interested in harnessing their human resources in order to stay ahead of the pack, through their soaring attention on specific aspects of knowledge management (De Hoog, van Heijst, van der Spek et al., 1999), which deals with the conceptualization, review, consolidation, and action phrases of creating, securing, combining, coordinating, and retrieving knowledge. Undeniably, with Web-based and intranet technologies (Dunn & Varano, 1999), the connectivity and possible sharing of organizational knowledge (bits and pieces of individual know-how scattered throughout the organization) are greatly enabled to cultivate the knowledge culture of the organization. In a knowledge-creating organization (Nonaka & Takeuchi, 1995), employees are expected to continually improvise and invent new methods to deal with unexpected difficulties, and to solve immediate problems and share these innovations with other employees through some effective communication channels or knowledge-transfer mechanisms. In fact, complete organizational knowledge is created only when individuals keep modifying their knowledge through interactions with other organizational members. The challenge that organizations now face is how to devise suitable information system (IS) support (Vat, 2000, 2002a, 2002b) to turn the scattered, diverse knowledge of their people into well-documented knowledge assets ready for deposit and reuse to benefit the whole organization. This article presents some learning organization perspectives of employee-based collaboration through the design of a specific IS support called the *organizational memory information system*—hence, the term OMIS.

THE BACKGROUND OF A LEARNING ORGANIZATION

The concept of the learning organization took seed several decades ago and gained major recognition with the incredible success of Peter Senge's 1990 book, *The Fifth Discipline*. Senge describes a learning organization as a place where people continually expand their capacity to create the results they truly desire, where new and expansive patterns of thinking are nurtured, where collective aspiration is set free, and where people are continually learning how to learn together. At the core of the learning organization are five essential learning disciplines—personal mastery, mental models, shared vision, team learning, and systems thinking—that may be briefly described as follows. Personal mastery has to do with individual learning and can be seen as the basic building block through the actualization of which the learning organization is typically constructed. Mental models are about how individuals reflect on their own knowledge, using such models to improve the internal understanding of an organization's functions and processes. Shared vision implies a sense of group commitment to a matrix of organizational goals, while team learning describes a sharing and utilization of knowledge involving collective thinking skills. The purpose of systems thinking is to understand relationships and interrelationships, as well as the context and the forces that affect the behavior of a system or organization. For the early half of the 1990s, the idea of learning organization had been criticized as the mere re-incarnation of earlier ideologies, such as organization development and total quality management (Rasmussen, 1997). Nonetheless, as more entities adopt the practices underlying the learning organization, it appears that the learning organization concept is passing from buzzword status to a meaningful expression of best organizational practices. Nowadays, many organizations that are engaged in constantly revamping and retooling themselves may be seen as reaching for that ideal goal of learning organizations. In fact, in this modern age of information technology and swift change, learning has become an integral part of the work of an organization run along principles intended to encourage constant reshaping and change. More importantly, learning organizations can be characterized as the organizations that continuously transform themselves by developing the skills of all their people and by achieving what Argyris (1992) has called *double-loop learning*, which helps transfer learning from individuals to a group, provide for organizational renewal, keep an open attitude towards the outside world, and support a commitment to knowledge. And this is often facilitated by the provision of some organizational knowledge transfer mechanisms, an example of which is the organizational memory information system (OMIS) to bring about the fundamental shifts in thinking and interacting, and the new capabilities needed in the learning organizations.

OMIS: AN ORGANIZATIONAL LEARNING EXPERIENCE

Lately, an organization's ability to learn is often considered a process of development to organizational memory. By organizational memory (Walsh & Ungson 1991), we are referring to various structures within an organization that hold knowledge in one form or another, such as databases and other information stores, work processes, procedures, and product or service architecture. As a result, an organizational memory (OM) must be nurtured to assimilate new ideas and transform those ideas into action and knowledge, which could benefit the rest of the organization (Ulrich, Von Glinlow & Jick 1993). Through understanding the important components of the OM (Vat, 2001), an organization can better appreciate how it is currently learning from its key experiences, to ensure that relevant knowledge becomes embedded within the future operations and practices of the organization. In practice, creating and using an OM is a cooperative activity necessarily involving many members of an organization. If those individuals are not adequately motivated in contributing to the OM initiative, and the organizational culture does not support knowledge sharing (Orlikowski, 1992), it is not likely to turn the scattered, diverse knowledge present in various forms, into well-structured knowledge assets ready for deposit and reuse in the OM. Consequently, it is important to distinguish between the organizational memory (OM encompassing people) and the OMIS that captures in a computational form only part of the knowledge of the organization. The OM captures the knowledge of the organization. The associated OMIS makes part of this knowledge available either by providing direct access to it (for example, codified knowledge assets such as experience reports), or indirectly by providing knowledge maps (for example, tacit knowledge assets such as personnel with specific expertise). Managing the OM deals first of all with the question of "Which knowledge should go into the OMIS?" Answering this question requires determining what knowledge is owned by the members of the organization, what knowledge is needed now, what is going to be needed in the future, and for what purposes. This helps the organization to define not only a strategy for acquiring the needed knowledge, but also to establish validation criteria in relation to the defined goals. Besides, we also need to deal with "who needs the knowledge, when, and why," as well as the policies for accessing and using the OMIS. This contextualization of the OMIS with respect to the organization's ability to learn is essential to implement the mechanisms of organizational knowledge transfer.

FUTURE TRENDS OF DESIGNING OMIS

When designing an OMIS to support an organization to learn (Vat, 2001, 2002a), we consider the following modes of learning: 1) individual, 2) group, and 3) repository. Individual learning is characterized by knowledge being developed, and possibly the result of combining an insight with know-how from other sources in the organization, but it is often not distributed and is not secured for reuse. Group learning is centered around the concept of communication in two possible modes: supply-driven or demand-driven. The former is characterized by an individual who has found a way to improve the work process and communicates this to one's co-workers. The latter refers to a worker who has recognized a problem in the current process and asks fellow workers whether they have a solution for this problem. In each case, knowledge is developed, distributed, and possibly combined with knowledge from other parts of the organization, but it is seldom secured. In repository learning, the communication element is replaced by collection, storage, and retrieval of knowledge items. Namely, it is typified by storing lessons learned in some information repository so that they can be retrieved and used when needed. Overall, in repository learning, knowledge is developed, secured, distributed, and is possibly the result of knowledge combination. It is convinced that the requirements of an OMIS design should be formulated in terms of the following usage scenarios. Namely, an OMIS should facilitate individual workers to access the knowledge required by combination, to submit a lesson learned, and to decide which of the co-workers would be interested in a lesson learned. Also, there should be criteria to determine if something is a lesson learned, how it should be formulated and where it should be stored, and how to distribute some newly asserted knowledge piece to the workers in need. The perceived technical issues, nevertheless, could include the following: How do we organize and index the

OM to enhance its diffusion? How do we retrieve relevant elements of the OM to answer a user request or proactively push relevant elements towards users? How do we adapt the answer to users, in particular to their tasks, according to the knowledge contexts? These problems are largely related to information retrieval, and they are bound to the OM framework for knowledge distribution, whose goal is to improve organizational learning, with the aid of the OMIS support.

CONCLUSION

Much of today's literature (Badaracco, 1991; Hamel & Prahalad, 1994; Quinn, 1992; Pinchot & Pinchot, 1994) supports the supposition that intellectual material in the form of information, knowledge, and any other form of intellectual property, is a valued organizational asset, and organizations are increasingly dependent on information technology (IT) for the transfer of knowledge and information. Conspicuously missing, however, is often a discussion of collaboration (Schrage, 1990) as a regenerative source of ideas that will advance organizations to learn, change, and excel (Menon, 1993; Stewart, 1994). In other words, simply transferring information at accelerated (IT) speeds contributes little added value to knowledge. Organizations must go beyond simple information transfer processes to survive and prosper. Garvin (1993) characterizes organizational learning as a continual search for new ideas. To collaborate is to work in a joint intellectual effort, to partition problem solving to produce a synergy such that the performance of the whole exceeds that of any individual contributor. The central issue in organizational learning is how individual learning is transferred to the organizational level. Here, we are assuming an organization of learners who take ownership for their own development and learning on a self-directed basis. Yet, only with a clear understanding of the transfer process can we manage learning processes consistent with organizational goals, issues, and values. If this transfer process is indeed actualized in the design of the OMIS, we could well have a learning organization which has the capability of capturing learning in its different paths and incorporating that learning into the running of its daily operations.

REFERENCES

Argyris, C. (1992). *On organizational learning*. Cambridge, MA: Blackwell Business.

Badaracco, J. (1991). *The knowledge link*. Boston: Harvard Business School Press.

De Hoog, R., van Heijst, G., van der Spek, R. et al. (1999). Investigating a theoretical framework for knowledge management: A gaming approach. In J. Liebowitz (Ed.), *Knowledge management handbook*. Springer-Verlag.

Dunn, J.R. & Varano, M.W. (1999). Leveraging Web-based information systems. *Information Systems Management*, *16*(4), 60-69.

Garvin, D.A. (1993). Building a learning organization. *Harvard Business Review*, *71*(4), 78-91.

Hamel, G. & Prahalad, C. (1994). *Competing for the future*. Boston: Harvard Business School Press.

Menon, A. (1993). Are we squandering our intellectual capital? *Marketing Research: A Magazine of Management*, *5*(3), 18-22.

Nonaka, I. & Takeuchi, H. (1995). *The knowledge creating company: How Japanese companies create the dynamics of innovation*. Oxford: Oxford University Press.

OECD. (1996) *The knowledge-based economy*. Paris: Organization for Economic Co-operation and Development (GD(96)102).

Orlikowski, W.J. (1992). *Learning from notes: Organizational issues in groupware implementation*. Working Paper #3428. Boston: MIT Sloan School of Management.

Pinchot, G. & Pinchot, E. (1994). *The end of bureaucracy and the rise of intelligent organization*. San Francisco: Berrett Koehler.

Quinn, J.B. (1992). *Intelligent enterprise*. New York: The Free Press.

Rasmussen, R.V. (1997). *Learning organization links*. Retrieved August 13, 1998, from courses.bus.ual-berta.ca/org-a652/learninglinks.htm.

Schrage, M. (1990). *Shared minds*. New York: Random House.

Senge, P.M. (1990). *The fifth discipline: The art and practice of the learning organization*. London: Currency Doubleday.

Stewart, T. (1994). Measuring company I.Q. *Fortune*, *129*(2), 24.

Stewart, T. (1997). *Intellectual capital: The New Wealth of organizations*. New York: Doubleday.

Ulrich, D., Von Glinlow, M. & Jick, T. (1993). High-impact learning: Building and diffusing a learning capability. *Organizational Dynamics*, *22*, 52-66.

Vat, K.H. (2000, November 1-2). Designing knowledge infrastructure for virtual enterprises in organizational learning. *Proceedings of the 10th Annual Business Information Technology Conference* (BIT2000), Manchester, England (CD-ROM Paper No. 45).

Vat, K.H. (2001, November 1-4). Towards a learning organization model for knowledge synthesis: An IS perspective. *CD-Proceedings of the 2001 Information Systems Education Conference* (ISECON2001), Cincinnati, Ohio, USA (Association of Information Technology Professionals).

Vat, K.H. (2002a). Designing organizational memory for knowledge management support in collaborative learning. In D. White (Ed.), *Knowledge mapping and management* (pp. 233-243). Hershey, PA: IRM Press.

Vat, K.H. (2002b, July 14-18). Engineering component-based knowledge applications for e-learning organizations: The software architects' challenge in organizational transformation. *Proceedings of the 6th World Multi-Conference on Systemics, Cybernetics and Informatics* (SCI2002) (Volume 1, pp. 262-267), Orlando, FL, USA (International Institute of Informatics and Systemics).

Walsh, J.P. & Ungson, G.R. (1991). Organizational memory. *Academy of Management Review*, *16*(1), 57-91.

KEY TERMS

Collaboration: To facilitate the process of shared creation involving two or more individuals interacting to create shared understanding where none had existed or could have existed on its own.

Double-Loop Learning: Together with single-loop learning, describes the way in which organizations may learn to respond appropriately to change. Single-loop learning requires adjustments to procedures and operations within the framework of customary, accepted assumptions, but fails to recognize or deal effectively with problems that may challenge fundamental aspects of organizational culture, norms, or objectives. Double-loop learning questions those assumptions from the vantage point of higher order, shared views, in order to solve problems.

Knowledge Management: The broad process of locating, organizing, transferring, and using the information and expertise within the organization, typically by using advanced information technologies.

Learning Organization: An organization that focuses on developing and using its information and knowledge capabilities in order to create higher-value information and knowledge, to modify behaviors to reflect new knowledge and insights, and to improve bottom-line results.

OMIS: An information system supporting the development of organizational memory, whose design philosophy is often organization specific. An example philosophy is to consider the OMIS as a meaning attribution system in which people select certain resource items out of the mass potentially available and get them processed to make them meaningful in a particular context in order to support their purposeful actions.

Organizational Learning: A process of leveraging the collective individual learning of an organization to produce a higher-level organization-wide intellectual asset. It is a continuous process of creating, acquiring, and transferring knowledge, accompanied by a modification of behavior to reflect new knowledge and insight, and produce a higher-level asset.

Organizational Memory: A learning history that tells an organization its own story, which should help generate reflective conversations among organizational members. Operationally, an organizational memory has come to be a close partner of knowledge management, denoting the actual content that a knowledge management system purports to manage.

Designing Web Applications

Gustavo Rossi
UNLP, La Plata, Argentina

Daniel Schwabe
Departamento de Informática, PUC-Rio, RJ, Brazil

INTRODUCTION

The World Wide Web (WWW) has become the most widely used platform for application development and information delivery. Web applications have evolved from static, read-only Web sites to current, collaborative, mobile, and pervasive information systems. Most companies are automating their core work flows using Web technologies; new businesses supported by the provision of complex Web services appear every day.

These application fields impose new modeling, design, and implementation requirements. Applications must have good performance, but they must also be usable and often adaptable to the individual user, his or her location, preferred interface device, and so forth. The time frame of the development life cycle must be short, which means we must improve design and implementation reuse. Applications evolve constantly and in unforeseen ways, so modularity is a must.

Web applications are different from "conventional" applications mainly because they are based on the hypermedia metaphor and because they allow users to access information by navigating through multimedia nodes that are connected by links. More complex structures, such as hierarchical indexes and landmarks, are often necessary to help the user find his or her way through the ocean of information. Successful Web applications provide good navigation topologies, helping the user to complete tasks without experiencing the "lost in hyperspace" syndrome.

In this context, conventional software engineering approaches fail to fulfill the needs of this application domain, if only because they neglect the navigational dimension of Web applications. They simply consider them just as a particular case of interactive applications; for example, they do not provide meaningful abstractions to model the unique features of this kind of software. In particular, treating a Web application just as any other transactional software, in which queries over data are posed using a browser as an interface, ignores the richness of the hypermedia paradigm for describing information systems. The same is true when considering a Web application just as a kind interactive application, in which the standard MVC (model-view-controller) paradigm (Knight & Dai, 2002) can be applied. While database- and programming-centric approaches can be used in the implementation step, conceptual modeling of Web applications requires a more careful approach.

Fortunately, the Web engineering community has come forth with a brand new set of methods and data models to address these issues. In the following section we survey the most interesting aspects of those methods.

BACKGROUND

The leading methodologies in this field, such as Web Modeling Language or WebML (Ceri, Fraternali, & Bongio, 2000; WebML, 2003), Object-Oriented Hypermedia Design Method, also called OOHDM (Schwabe & Rossi, 1998), UML-based Web Engineering or UWE (Hennicker & Koch, 2000), Object-Oriented Hypermedia or OO-H (Cachero, Gómez, & Pastor, 2000), or W2000 (Baresi, Garzotto, & Paolini, 2001), consider the Web-application-development life cycle as an iterative, incremental process, in which different models are specified before the implementation.

The design space is thus partitioned in basically three data models: conceptual, hypertext, and presentation models. Though each method provides its own primitives for describing each model, there is a general consensus on considering these three aspects separately.

The application or conceptual model describes the underlying application domain using well known design primitives: WebML uses data-modeling languages (such as the entity relationships), while OOHDM, UWE, and OO-H use object-oriented concepts in the syntax of UML (unified modeling language).

The conceptual model can be viewed as the data model in traditional applications. The navigation model, meanwhile, is built as a hypertext view over the conceptual model. There may be different views for different user profiles, providing distinct navigation paths or even navigation objects. Generally speaking, a navigation or hypertext schema describes the items that the user will navigate and the links he or she will follow to traverse those items. The navigation schema is described using nodes and links, which abstract the usual components of a Web site.

To model these aspects, almost all methods use their own proprietary notations; UWE, however, has chosen to use a lightweight UML extension by defining stereotypes for each typical hypertext component such as navigation object, link, and index. In this way, it can benefit from all existing concepts and tools developed in the UML world.

In OOHDM, the high-level navigation structure of the application is described by a navigational contexts schema. A navigational context is a set of nodes sharing some properties (products of a brand, results of a query, etc.), and the navigational context schema describes all meaningful contexts together with indexes for accessing the nodes in a context. Contexts allow specifying the navigation alternatives for all its members at the same time, without having to detail it for each individual item. For example, it is possible to state, "From any given product, one can navigate to any of its related products," instead of having to say, "From product CD-ROM, it is possible to navigate to product CD Blanks," from, "From product Shaver, it is possible to navigate to product Shaving Blades," and so on.

All modeling methods provide a design armoury for describing abstract aspects of the user interface in terms of interface objects. These objects show how pages will look like though avoiding delving into the implementation aspects.

Once the different Web application models have been completely specified, a running Web application can be built. Different methods provide different strategies for mapping models into running applications. In general, implementing the conceptual model is straightforward using either conventional databases or object-oriented tools. The navigational views may be materialized or not, but they can be easily described using current server page technologies; interfaces can be built by using HTML (hypertext markup language) or XML (extensible markup language) and Extensible Stylesheet Language (XSL) tools.

Some methods go further. For example, WebML provides a set of tools for automatically generating the running application by performing a set of transformations on the previously described models (WebRatio, 2003).

The main contribution of these methods to the field of Web engineering has been threefold. First, they posited the idea that modeling Web applications involve more than defining conceptual models, that is, the introduction of the navigational modeling activity is essential to understand the dynamics of these applications (Rossi & Schwabe, 1999).

Second, their notations are based on well-known modeling practices (object orientation and entity relationship), which improve abstraction and reuse capabilities, (see, for example, Koch & Kraus, 2002), and provide a good framework for extending the models. In this sense, these methods have been recently used to specify customized applications (Abrahão, Fons, González, & Pastor, 2002; Matera, Ceri, Dolog, & Nejdl, in press; Rossi, Schwabe, & Guimaraes, 2001) and Web-based business processes (Brambilla, 2003; Schmid, & Rossi, 2004).

Third, they have introduced abstract interface specification (or presentation) as an important activity in the engineering process, factoring out interface technology from the design itself, a crucial aspect in times of continuous, fast technological and standards evolution.

The Web engineering community has also discovered a large set of Web design patterns that record and convey design experience (Garzotto, Paolini, Bolchini, & Valenti, 1999; Lyardet, Rossi, & Schwabe, 2000; Molina, Meliá, & Pastor, 2002; Rossi, Schwabe, & Garrido, 1997). These patterns allow the designer to reuse smart solutions that we (and other people) have found in successful applications while solving recurrent problems. These patterns have generated the architecture of Web frameworks, semicomplete designs in a particular application domain that can be refined and customized to obtain running applications (Fayad, Schmidt, & Johnson, 1999; Schwabe, Emeraldo, & Rossi, 2001). More comprehensive work toward obtaining a complete set of design guidelines from a system of patterns can be read in Van Duyne, Landay, and Hong (2003).

FUTURE TRENDS

Related Work

Since their inception, modeling and design methodologies have evolved quickly, and now we can say that the Web engineering field is entering adolescence. There are many other methods deserving mention, such as Web Site Design Method or WSDM (De Troyer & Leune, 1998), focusing on audience-driven modeling and design, and Web Unified Modeling Language, also called WUML (Kappel, 2001), devised specifically for ubiquitous Web applications.

Coming from a complementary perspective, there are proposals aimed at integrating conventional data models and database management systems with the Web platform. One of the first such approaches is Strudel (Strudel, 2003), which allows building Web sites as queries to a database through a specially designed language. Another example is the Araneus Web-base management system (Mecca, Atzeni, Masci, Merialdo, & Sindoni, 1998). Araneus combines a data model for Web documents, several languages for wrapping, querying, creating, and updating Web sites, and a set of methods for site design and implementation. Araneus has also adapted the three-tier idea of separating the application data model from hypertext (navigation) and presentation specification.

Trends

The emergence of the concept of Web services as a way to deploy and use distributed functionality in the Web and to achieve universal interoperability has led to a new trend: the specification of business processes as a combination of Web services. Business processes can be described by modeling the behavior of an actual participant in the process (the so-called *executable* business processes) or by describing the business protocols, that is, the mutually visible message-exchange behavior of the parties (the *abstract* processes). Business Process Execution Language for Web Services, aslo called BPEL4WS (IBM, 2003) provides a language for the formal specification of business processes and business interaction protocols. Meanwhile, the proposers of WebML have extended their modeling approach to allow specifying Web applications that compose Web services to support arbitrary business processes and work flows (Brambilla, Ceri, Comai, Fraternali, & Manolescu, 2002). In this regard, existing work in the Workflow Management Coalition (WfMC, 2003) can also be mentioned.

Current trends in design methods include the extension of existing approaches to support the construction of semantic Web applications, and the use of modern design principles to build pervasive Web software (Kappel, Pröll, Retschitzegger, & Schwinger, 2003).

The emergence of the semantic Web (Berners-Lee, Hendler, & Lassila,, 2001) envisages the next generation of the WWW as a vast pool of semantically annotated data. Such annotations follow semantic models, which are in many ways equivalent to the design models used by the methodologies previously mentioned. Several initiatives are investigating the design, development, and implementation of applications for the semantic Web, such as Corcho, Gomez-Pérez, López-Cima, López-García, and Suárez-Figueroa (2003), Golbeck, Alford, and Hendler (2003), Jin, Decker, and Wiederhold (2002) and Lima and Schwabe (2003a, 2003b).

CONCLUSION

Web modeling and design approaches are evolving rapidly. A healthy community with people coming from diverse fields such as hypermedia, databases, and conceptual and object-oriented modeling has established the basis of a well-grounded discipline. In the meantime, as implementation approaches for Web applications have also evolved, we can build modular and usable Web applications using solid engineering approaches. Modern architectures such as Java2 Enterprise Edition (J2EE) or the variants of the model-view-controller paradigm for the Web (Knight & Dai, 2002) make the mapping of Web design schemata into running applications feasible and straightforward. Tools for assisting the designer in this process already exist as previously mentioned.

The time in which building Web software was a handcrafted, amateur process is over. The Web community is now adopting good engineering practices by seamlessly combining the wisdom from traditional software engineering with the creativity of hypermedia and interface designers.

REFERENCES

Abrahão, S., Fons, J., González, M., & Pastor, O. (2002). Conceptual modeling of personalized Web applications. *Proceedings of the International Conference on Adaptive Hypermedia*, (pp. 358-362).

Baresi, L., Garzotto, F., & Paolini, P. (2001). Extending UML for modeling Web applications. *Proceedings of 34th Annual Hawaii International Conference on System Sciences (HICSS'34)*.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic Web. *Scientific American*.

Bieber, M. (2000). Hypertext [Electronic version]. In A. Ralston, E. Reilly, & D. Hemmendinger (Eds.), *Encyclopedia of Computer Science* (4th ed., pp. 799-805). Nature Publishing Group. Retrieved from *www.web.njit.edu/~bieber/pub/cs-encyclopedia/csencyclopedia00.pdf*

Brambilla, M. (2003). Extending hypertext conceptual models with process-oriented primitives. *Proceedings of ER 2003 (International Conference on Conceptual Modeling)*, 246-262.

Brambilla, M., Ceri, S., Comai, S., Fraternali, P., & Manolescu, I. (2002, December). Model-driven specification of Web services composition and integration with data-intensive Web applications. *IEEE Bulletin of Data Engineering*.

Cachero, C., Gómez, J., & Pastor, O. (2000). Object-oriented conceptual modeling of Web application interfaces: The OO-HMethod presentation abstract model. *ECWEB'00*, 206-215.

Ceri, S., Fraternali, P., & Bongio, A. (2000). Web modeling language (WebML): A modeling language for designing Web sites. Proceedings of WWW9 Conference, Amsterdam, Netherlands. (Also in (2000). *Computer Networks, 33*, 137-157).

Corcho, O., Gomez-Pérez, A., López-Cima, A., López-García, V., & Suárez-Figueroa, M.-d.-C. (2003). ODESeW: Automatic generation of knowledge portals for intranets and extranets. *Proceedings of ISWC 2003*, 802-817.

De Troyer, O., & Leune, C. (1998). WSDM: A user-centered design method for Web sites, computer networks and ISDN systems. *Proceedings of the 7th International World Wide Web Conference*, 85-94.

Fayad, M., Schmidt, D., & Johnson, R. (Eds.). (1999). *Building application frameworks.* Wiley.

Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1995). *Design patterns: Elements of reusable object-oriented software.* Addison Wesley.

Garzotto, F., Paolini, P., Bolchini, D., & Valenti, S. (1999). Modeling-by-patterns of Web applications. *Proceedings of ER (Workshops) 1999*, 293-306.

Golbeck, J., Alford, A., & Hendler, J. (2003). Handbook of human factors in Web design. In R. Proctor & K. P. Vu (Eds.), *Organization and structure of information using semantic Web technologies.* Retrieved from *www.mindswap.org/papers/Handbook.pdf*

Halasz, F. G. (1991). *Seven issues reconsidered.* Keynote address of Hypertext'91. Retrieved from *http://www2.parc.com/spl/projects/halasz-keynote/* (See also Slide 4 at *http://www2.parc.com/spl/projects/halasz-keynote/recon/sld004.htm*)

Hennicker, R., & Koch, N. (2000). A UML-based methodology for hypermedia design. In A. Evans, S. Stuart, & B. Selic (Eds.), *Lecture notes in computer science: Vol. 1939. UML 2000: The unified modeling language—Advancing the standard.* York, UK: Springer Verlag.

IBM. (2003). *Specification: Business process execution language for Web services.* Retrieved *from http://www-106.ibm.com/developerworks/library/ws-bpel/*

Jin, Y., Decker, S., & Wiederhold, G. (2002). OntoWebber: Building Web sites using semantic Web technologies. Retrieved from *www.db.stanford.edu/~yhjin/docs/owedbt.pdf*

Kappel, G., Pröll, B., Retschitzegger, W., & Schwinger, W. (2001). Modeling ubiquitous Web applications: The WUML approach. *Proceedings of ER 2001 Workshops*, 183-197.

Knight, A., & Dai, N. (2002). Objects and the Web. *IEEE Software, 19*(2), 51-59.

Koch, N., & Kraus, A. (2002). The expressive power of UML-based Web engineering. *Proceedings of 2nd International Workshop on Web Oriented Software Technology (IWWOST'02) at ECOOP'02*, 105-119.

Lima, F., & Schwabe, D. (2003a). Application modeling for the semantic Web. *Proceedings of LA-Web 2003*, 93-102. (Also retrieved from *http://www.la-web.org*)

Lima, F., & Schwabe, D. (2003b). Modeling applications for the semantic Web. In *Lecture notes in computer science: Vol. 2722. Proceedings of the third international conference on Web engineering (ICWE 2003)*. Heidelberg, Germany: Springer Verlag.

Lyardet, F., Rossi, G., & Schwabe, D. (2000). *Patterns for e-commerce applications.* Proceedings of Europlop 2000, Kloster Irsee, Germany.

Matera, M., Ceri, S., Dolog, P., & Nejdl, W. (in press). *Integrating WebML and UML for modelling multi-channel context-aware Web applications.* Proceedings of the International Conference on Web Engineering, Munich, Germany.

Mecca, G., Atzeni, P., Masci, A., Merialdo, P., & Sindoni, G. (1998). The Araneus Web-base management system. *Proceedings of ACM SIGMOD International Conference on Management of Data 1998*, 544-546.

Molina, P., Meliá, S., & Pastor, O. (2002). User interface conceptual patterns. *Proceedings of Interactive Systems Design, Specification and Validation*, 159-172.

RDF. (1999). *W3C recommendation.* Retrieved from *http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/*

Rossi, G., & Schwabe, D. (1999). Web application models are more than conceptual models. In *Lecture Notes in Computer Science, Vol. 1727. Proceedings of the World Wild Web and conceptual modeling '99 Workshop, ER'99 Conference*. Paris: Springer.

Rossi, G., Schwabe, D., & Garrido, A. (1997). *Design reuse in hypermedia applications development.* Proceedings of ACM International Conference on Hypertext (Hypertext'97), Southampton, UK.

Rossi, G., Schwabe, D., & Guimaraes, R. (2001). *Designing personalized Web applications*. Proceedings of the WWW10, Hong Kong, China.

Schmid, H., & Rossi, G. (2004, January). Modeling and designing business processes in Web applications. *IEEE Internet Computing*, 19-27.

Schwabe, D., Emeraldo, L., & Rossi, G. (2001, Spring). Engineering Web applications for reuse. *IEEE Multimedia,* 2-12.

Schwabe, D., & Rossi, G. (1998). An object oriented approach to Web-based application design. *Theory and Practice of Object Systems, 4*(4).

Strudel. (2003) Retrieved from *http://www.research.att.com/~mff/strudel/doc/*

UML. (2003). *UML resource center.* Retrieved from *http://www.rational.com/uml/*

UWA Project. (2003). Retrieved from *http://www.uwaproject.org*

Van Duyne, D., Landay, J., & Hong, J. (2003). *The design of sites.* Addison Wesley.

W3C Semantic Web Activity. (2001). Retrieved from *http://www.w3.org/2001/sw/*

WebML. (2003). *The WebML Web site*. Retrieved from *http://www.webml.org*

WebRatio. (2003). *The WebRatio development tool.* Retrieved from *www.webratio.com*

WfMC. (2003). *The Workflow Management Coalition.* Retrieved from *http://www.wfmc.org*

KEY TERMS

Customization: Customizable applications are those applications that can adapt themselves to the context in which they execute; applications (in particular, Web applications) can be adapted to the user profile, the location, the network connection, and so forth. Building applications that can adapt themselves to the current context involve many different software and information representation problems. In the context of Web applications, we can adapt the application behavior to the user profile, we can change the hypermedia topology for different users' tasks, or we can adapt the user interface for different appliances.

Hypermedia: Hypermedia is a metaphor for representing and accessing multimedia information. Information is represented as a structured collection of multimedia information items, called nodes. The collection's structure can be represented explicitly or implicitly and is typically seen as a graph, whose nodes are connected by links which represent meaningful relationships among them. The graph is traversed by navigating following links by selecting anchors in nodes. The WWW uses the hypermedia paradigm extensively; nodes are usually Web pages and links indicate the URLs that the user can reach from the current page (Halasz, 1991).

Model-View-Controller: The model-view-controller (MVC) is the most widely used software metaphor for building interactive applications. Originally conceived in the context of the Smalltalk programming environment, it is now currently used in Web applications. The MVC partitions the software space into three modules: the model that contains application objects and behaviors, the view that represents how these objects are presented to the user, and the controller that handles the interaction with the user. Using MVC allows a clear decoupling of concerns that helps building applications in which changes in the look, feel, and the interaction style may be changed without affecting the application core (Knight, & Dai, 2002).

Navigation: Navigation is the usual way of traversing a hypermedia graph. While the user explores a particular node, he or she can click an anchor to follow a link to explore another different node. While navigation is a good metaphor to freely traverse a graph of information when the graph is huge or very dense, the user might get lost. Different tools such as maps, landmarks, and navigation histories may help to overcome this problem (Bieber, 2000).

Patterns: Software patterns represent proven solutions to recurrent design problems. A pattern names and describes a problem that occurs once and again, and presents a good solution (a solution used by many expert designers) in an abstract way so it can be used in different occurrences of the same problem. The core of the solution is usually expressed using UML diagrams showing class and object relationships that serve to solve the recurrent problem. In Gamma (1995), the authors present a good catalogue of design patterns for object-oriented software. The same concept applies to the hypermedia and Web domain when designers recognize that they constantly face similar problems. Web navigation and interface patterns record and convey problems related to Web applications' navigational and interface structures.

Semantic Web: According to its original proponents (Berners-Lee et al., 2001, p. 35), "the semantic Web is an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation." In addition to the currently available data meant for human consumption, the semantic Web provides metadata describing this data, allowing it to be processed by programs, allowing for more effective discovery, automation, integration, and reuse across various applications. The metadata are described using the resource description framework (RDF, 1999). The initiative is led by the World Wide Web consortium (W3C, 2001).

UML: The unified modeling language (UML) is the standard modeling and design notation to build object-oriented applications. UML comprises a set of submodels that allow a designer to easily express both structural and dynamic aspects of object societies. UML and its accompanying process model, the rational unified process (RUP), covers all steps of the software development process by providing different diagrammatic tools such as use-case diagrams, class diagrams, sequence diagrams, state models, and so forth. UML is itself described by a metamodel and possesses its own extension mechanisms to adapt it to specific application domains.

Web-Base: According to Mecca et al. (1998), a Web-base is a collection of data of heterogeneous nature, specifically, highly structured data as in relational or object-oriented databases and semistructured data in the Web style. It incorporates both databases and Web sites. A Web-base management system (WBMS) is a system for managing such Web-bases.

Developing Dynamic Balanced Scorecards

Pallab Saha
National University of Singapore, Singapore

INTRODUCTION

With the advent of scientific management in the late 19th century came attempts to increase worker efficiency by setting standards for various factors in an industrial system. While measuring financial factors of a business was prevalent even during late 19th century, the scientific management ideology heralded the start of an era of formally measuring non-financial factors. This ignited several performance measurement frameworks that were used by businesses.

The most widely used among them included Tableau de Bord, developed in early 20th century in France, which focused on improving production processes. Management by Objectives (MBO) was markedly different and radical as compared to early "command & control" structures advocated by Frederick Taylor. Performance Matrix, first developed in 1986, is a list of performance metrics that are priority weighted. Balanced Scorecard (BSC), first published in the *Harvard Business Review* in 1992 by Kaplan and Norton, is a comprehensive framework that imbibed the good ideas of all the earlier frameworks plus presented some new attributes in a concise and understandable manner that managers found easy to comprehend (Epstein & Manzoni, 1998).

BSC focuses on the need to derive performance measures from strategic goals and objectives (Epstein & Manzoni, 1998; Kaplan & Norton, 1992, 1996a, 1996b; Niven, 2002). The impetus for proposing the BSC was the increasing disillusionment of solely using financial metrics to plan, monitor, control and manage organizational performance (Niven, 2002). The BSC categorizes performance measures into four perspectives, namely: financial, customer, internal business process and learning and growth, with the assumption that these capture an organizational performance in a holistic manner. BSC also recognizes the fact that having isolated and unlinked measures constituting the scorecard does not provide a mechanism for managers to "see" the impact of changes and test business hypotheses (Kaplan & Norton, 2004). A well-architected BSC therefore must describe organizational strategy through a system objectives and measures that are linked (Niven, 2002). These measures are linked together in a chain of cause-effect relationships from the performance drivers in the learning and growth perspective all the way through to financial perspective (Kaplan & Norton, 2004; Niven, 2002). However intuitive cause-effect chains may seem, creating them is perhaps the most challenging aspect in scorecard development (Niven, 2002). The concept of balance central to the BSC performance measurement system specifically relates to (Niven, 2002):

- Balance between financial and non-financial measures
- Balance between internal and external constituents of the organizations
- Balance between lag and lead indicators

BACKGROUND

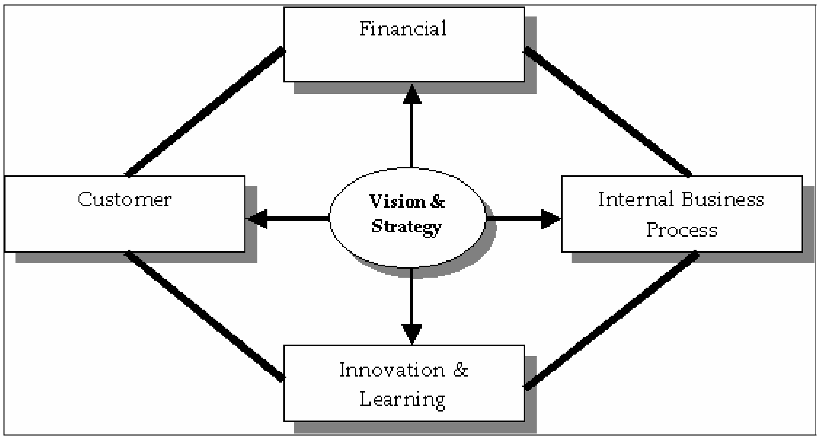
The balanced scorecard framework was selected by the *Harvard Business Review* as one of the most influential management theories of the last 75 years. Driven by its simplicity and power, the BSC framework has been adopted by nearly 50% of Fortune 1000 companies in a Bain & Company Survey in 1999. Notwithstanding its prevalence and industry acceptance, the BSC framework in its classical form as proposed by its founders has certain key deficiencies that often come as obstacles to its effective implementation (Fowler, 2003). These are:

- *Deficiency 1:* A cause-effect diagram used to depict the strategy map expresses the causality in a unidirectional manner.
- *Deficiency 2:* A cause-effect diagram, while showing the cause-effect linkages, does not take into consideration the time varying impact of these influences.
- *Deficiency 3:* A BSC provides no mechanism for validating the performance measures specified.

These key deficiencies have far-reaching implications on the effectiveness of BSC as a strategy implementation framework (Takikonda, 1998; Veen-Dirks & Martin, 2002).

- Cause-effect linkages are viewed as one-way in nature, thus emphasizing one-way thinking.

Figure 1. Basic structure of a BSC performance measurement system



- Unidirectional approach leads to difficulty in reliable simulations.
- While linkages from non-financial to financial measures are shown, feedback loops depicting the impact of financial on non-financial measures is absent.
- It is also assumed that both cause and effect occur in the same place and time, thus not provisioning for delays in causality, thereby missing the temporal/dynamic complexity.
- A static cause-effect diagram makes it difficult to identify "good" predictor metrics.

These five implications necessitate enhancement of the BSC with systems thinking/system dynamics approach to develop the dynamic balanced scorecard (DBSC).

DYNAMIC BALANCED SCORECARD

A broad methodology that combines systems thinking/system dynamics with balanced scorecard theory is shown next. Listed as follows are suggested steps or activities involved that culminate into development of a DBSC.

1. Assessing strategic and organizational landscape.
2. Identifying critical success factors (CSF).
3. Analyzing performance over time for identified CSFs.
4. Developing critical success loops (CSL) for each CSF.
5. Synthesizing individual CSLs into one single global CSL (GCSL).
6. Identifying and defining strategic resources.
7. Specifying strategies and strategic objectives.
8. Building strategy map(s).
9. Identifying performance measures for strategic resources.
10. Specifying targets for performance measures.
11. Identifying initiatives to address strategic objectives and CSFs.
12. Developing Dynamic BSC.
13. Validating DBSC through dynamic simulation.
14. Deploying DBSC as the performance management framework.

While these specific steps can vary from organization to organization, the approach remains largely similar. As the objective of this article is to highlight the use of systems thinking/system dynamics in order to overcome (most) of the deficiencies of the traditional BSC, steps that are specifically different are highlighted and described as follows.

Step 3: Analyzing Performance Over Time for CSFs

Following identification of CSFs that the organization "thinks" is key to its success, analyzing their behavior over time is a good starting point to uncover the underlying systemic structures among interrelated variables. The key here is to select a time horizon that is appropriate for the specific success factor. The "performance over time" graph is a simple graph showing CSFs on one axis and time on the other. Strategic assessment performed earlier identifies the number of other significant variables that might impact specific CSFs. Analysis of CSF performance over time often provides critical input to identify initial linkages between the variables and their impacts on CSFs. This helps in developing testable hypotheses about how variables' behavior might be interrelated (Kim, 1997).

Step 4: Developing Critical Success Loops (CSL) for Each CSF

Representing the systemic structure of each CSF is accomplished using causal loop diagrams (CLD) to develop CSL (Kim, 1997). Drawing CSLs allows the graphical depiction of the structure that exists between the CSF and its impacting variables. The trick here is to identify at least one reinforcing loop for each CSF that would be self-sustaining (Sterman, 2000). Causal loop diagrams used to develop CSLs should clearly identify:

- Linkages between the variables as either reinforcing or balancing (link polarity)
- Feedback loops as either reinforcing or balancing (loop polarity)
- Delays between variables depicting the temporal perspective

Figure 2 shows an illustration of a CSL as a causal loop diagram where *process quality* is the CSF. The diagram clearly identifies link polarities, loop polarities and delays.

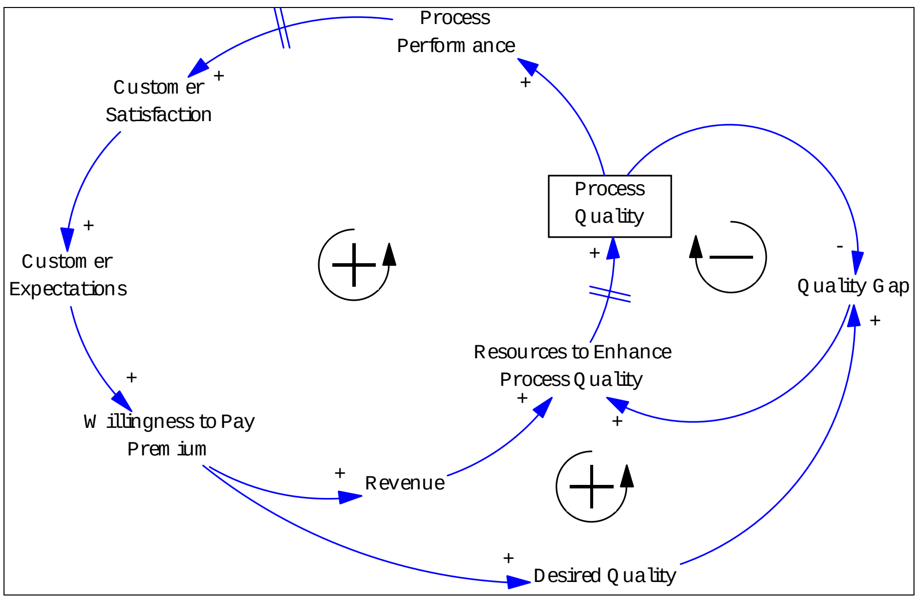
Step 5: Synthesizing Individual CSLs into One Single Global CSL

The next step is to uncover the interplay between the CSFs. This is accomplished by identifying the linkages between the separate CSLs. This involves developing a global critical success loop that logically links all identified CSFs and their impacting variables as a causal loop diagram (Kim, 1997). The global view of the success factors allows depiction of the "big picture" of various variables and their interrelationships. A global CSL is often a starting point to identify policy interventions required to impact CSFs and steer their performance in a manner that an organization desires.

Step 6: Identifying and Defining Strategic Resources

As the primary aim is to identify policy interventions required to impact CSFs and steer their performance in a manner that an organization desires, a critical activity involves identifying the strategic resources that an organization has access to (Warren, 2002). Identifying strategic resources are critical, as the organization "uses" these resources to impact critical success factors. Care must be taken to clearly identify both tangible and non-tangible resources.

Figure 2. Process quality critical success loop



Step 7: Specifying Strategies and Strategic Objectives

Strategic resources are elements that an organization has access to. As resources accumulate or deplete over time, all policy level interventions fundamentally seek to *add*, *increase*, *raise*, *enhance*, *build*, *grow*, *augment*, *boost*, *extend*, *expand* or *develop* either an existing resource or a new resource (Warren, 2002). However in either case, an organization has to work with what is available to it or has access to, which are the strategic resources. In specifying strategic objectives it is imperative to ensure that objectives (and goals) seek to impact resources over time.

Step 8: Building Strategy Map(s)

Figure 3 shows a partial strategy map depicting strategic objectives with linkages. Strategy map is a tool that helps organizations both to communicate their strategies and processes and systems that are needed to implement those strategies (Kaplan & Norton, 2004). In short, strategy maps depict the transformation of various organizational resources into desired outcomes.

Step 9: Identifying Performance Measures for Strategic Resources

It is critical to measure resources in order for them to be controlled, managed and steered. Table 1 lists *resources*, their *typical measures*, *units of resources*, *inflows & outflows*, *units of inflows* and *outflows* and *impacting factors* (Warren, 2002). It is to be noted at this point that a firm is also expected to specify targets for each performance measure (Step 10). Impacting factors are drivers that drive inflows and outflows that accumulate or deplete each resource. Hence when identifying initiatives to impact strategic resources, alignment with impacting factors is essential (McGarvey & Hannon, 2004). The presumption is that strategic initiatives affect impacting factors, which in turn impact the strategic resources.

Step 12: Develop Dynamic Balanced Scorecard

Causal loop diagrams provide an excellent method to capture the dynamics of the problem at a high level, but are not really amenable to simulation in order to fully understand the underlying dynamics of the variables (Sterman, 2000). Hence stock and flow diagrams are used to develop the dynamic balanced scorecard. Strategic resources identified earlier become the "stocks" that accumulate or deplete with time.

Figure 4 shows the partial stock and flow diagram that can be used to validate resources and their performance measures through the use of dynamic simulations. Following validation, measures can then be part of the Dynamic BSC amenable to *what – if* analysis through the development of business dashboards.

FUTURE TRENDS

Organizations are moving toward a service-oriented approach to business operation, which is driving the adop-

Figure 3. Partial strategy map depicting strategic objectives with linkages

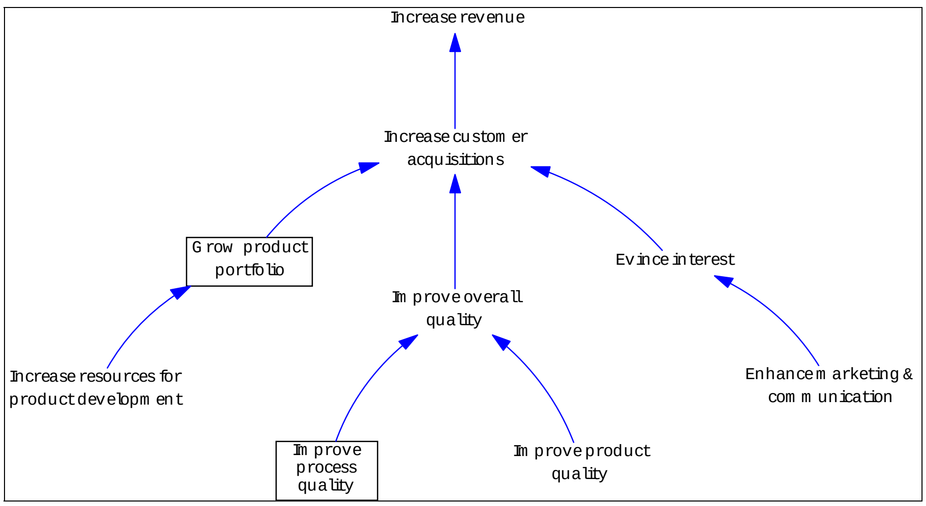


Table 1. Resources, measures, flows and impacting factors

Resources	Performance Measures	Inflows & Outflows	Units of Inflows & Outflows	Impacting Factors (Illustrative)
Revenue	Total revenue (Dollars)	Revenue added Revenue lost	Dollars / quarter	Sales, Competitors performance, New products, Price
Customer base	Individuals (Count)	Customers acquired Customers lost	Individuals / quarter	Sales effort, Price, Demand, Quality, Word of mouth
Products	Number of products (Count)	Products added Products retired	Products / quarter	Development effort, Staff skills, Demand
Overall quality	Defects (Defects Per Million Opportunities)	Increase in defects Reduction in defects	Defects Per Million Opportunities / quarter	Product quality, Process quality, Quality improvement effort
Staff	Number of staffs (Count)	Staffs added Staffs depleted	People / quarter	Salaries, Growth opportunities, Firm performance
Market reputation	Index	Increase in reputation Reduction in reputation	Index value / quarter	Word of mouth, Competitors perception, Customer performance

tion of continuous business process management (Kalakota & Robinson, 2003). However for business process management to be successful, it is imperative for organizations to link process and business performance in an integrated manner to develop an enterprise performance management system (Smith & Fingar, 2003). Development of such management systems depends to some extent on the availability of generic scorecards that organizations can configure based on their specific business needs. Hence future research could focus on two areas:

- Development of dynamic scorecards for industry-specific initiatives like straight-through processing and collaborative planning, forecasting and replenishment.
- Development of dynamic scorecards to monitor and manage compliance to regulatory requirements mandated by law (e.g., the Sarbanes-Oxley Act).

CONCLUSION

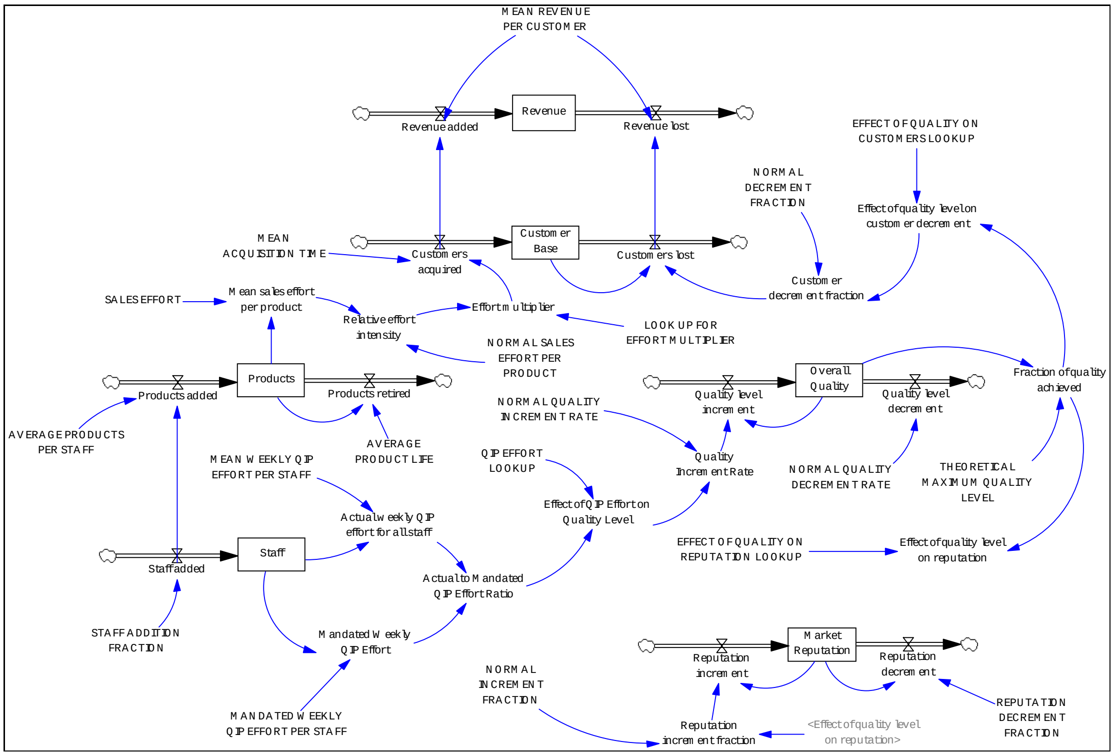
The problems of understanding complex system behavior and the challenge of developing models to capture such behavior are apparent in the field of performance measurement. Organizations are faced with conflicting goals of maximizing financial and non-financial returns. Existence of multiple measures in a scorecard addressing such conflicting goals increases the structural complexity of the measurement system and makes it difficult to predict its behavior. Enhanced with use of systems thinking, balanced scorecard practice can be made richer with the incorporation of the dynamic perspective of strategies and performance measures. While BSCs do assume the systemic nature of businesses, they fail to explicitly address the systemic thinking in its original form. This brings forth serious shortcomings in implementation. Synergizing system thinking and balanced scorecards provides a practical way to address these shortcomings.

REFERENCES

Epstein, M., & Manzoni, J-F. (1998). Implementing corporate strategy: From tableaux de bord to balanced scorecards. *European Management Journal, 16*(2), 190-203.

Fowler, A. (2003). Systems modeling, simulation and the dynamics of strategy. *Journal of Business Research, 56,* 135-144.

Figure 4. Partial stock and flow diagram used for validating resources and measures



Kalakota, R., & Robinson, M. (2003). *Services blueprint: Roadmap for execution.* Boston: Addison-Wesley.

Kaplan, R., & Norton, D. (1992). The balanced scorecard – Measures that drive performance. *Harvard Business Review, 70*(1), 71-79.

Kaplan, R., & Norton, D. (1996a). *The balanced scorecard: Translating strategy into action.* Boston: Harvard Business School Press.

Kaplan, R., & Norton, D. (1996b). Linking the balanced scorecard to strategy. *California Management Review, 39*(1), 53-79.

Kaplan, R., & Norton, D. (2000, September/October). Having trouble with your strategy? Then map it. *Harvard Business Review*, 167-176.

Kaplan, R., & Norton, D. (2004). *Strategy maps: Converting intangible assets into tangible outcomes.* Boston: Harvard Business School Press.

Kim, D.H. (1997). From key success factors to key success loops. *The Systems Thinker, 8*(5), 6-7.

Lyneis, J.M. (1999). System dynamics for business strategy: A phased approach. *System Dynamics Review, 15*(1), 37-70.

McGarvey, B., & Hannon, B. (2004). *Dynamic modeling for business management.* New York: Springer-Verlag.

Niven, R.P. (2002). *Balanced scorecard step-by-step: Maximizing performance and maintaining results.* New York: John Wiley & Sons.

Smith, H., & Fingar, P. (2003). *Business process management: The third wave.* Tampa, FL: Meghan-Kiffer Press.

Sterman, J. (2000). *Business dynamics: Systems thinking and modeling for a complex world.* Boston: Irwin McGraw-Hill.

Takikonda, L.U., & Takikonda, R.J. (1998). We need dynamic performance measures. *Management Accounting, 80*(3), 49-51.

Veen-Dirks, P.V., & Wijn, M. (2002). Strategic control: Meshing critical success factors with the balanced scorecard. *Long Range Planning, 35,* 407-427.

Warren, K. (2002). *Competitive strategy dynamics*. Chichester, UK: John Wiley & Sons.

KEY TERMS

Balanced Scorecard: A selected collection of measures (both financial and non–financial) derived from an organization's strategies. Usually balanced scorecard is considered to be the most prevalent form of performance measurement system.

Causal Loop Diagram: Consists of two or more variables connected by links, which usually take the form of arrows. These diagrams depict three major components: feedback loops, cause-effect relationships and delays.

Collaborative Planning, Forecasting and Replenishment: An industry-wide initiative that involves collaboration between trading (supply chain) partners in the retail industry in order to achieve lower costs, higher efficiency and better customer satisfaction.

Performance Measurement: A management practice that involves specifying and managing a set of metrics addressing business performance, usually in the form of a scorecard. A good performance measurement system is targeted to implement a coherent set of strategies.

Sarbanes-Oxley Act: Requires compliance with a comprehensive reform of accounting procedures for publicly held corporations to promote and improve the quality and transparency of financial reporting by both internal and external independent auditors.

Stock & Flow Diagram: Consists of a set of stocks and flows, stocks representing resources that either accumulate or deplete over time through inflows and outflows. Stock and flow diagrams are largely used in system dynamics and rarely in systems thinking approaches.

Straight-Through Processing: An industry-wide initiative to streamline, transform and automate processes in the financial securities industry in order to achieve lower settlement times and quicker trading.

Strategic Resources: Anything to which an organization has access to that may be strategically useful to it.

Strategy Map: A depiction of strategic objectives with linkages with the aim of showing transformation of various organizational resources into desired outcomes.

System Dynamics: A methodology for modeling and managing complex feedback systems. This includes dynamic simulation models that allow policy formulation and testing.

Systems Thinking: Systems thinking, closely related to system dynamics, is also a methodology to model and analyze dynamic, complex and interdependent nature of systems. However, systems thinking seldom includes developing dynamic simulation models as in system dynamics.

Developing Trust in Virtual Teams

Niki Panteli
University of Bath, UK

INTRODUCTION

During the last few years, there has been an increasing acknowledgment of the importance of trust in business interactions within the management and organizational literatures (e.g., Kramer & Tyler, 1996; Mayer, Davis, & Schorman, 1995; Rousseau, Sitkin, Burt, & Camerer, 1999). Trust, as a positive and confident expectation in the behavior of another party (Cook & Wall, 1980; Currall & Judge, 1995), enables cooperation and becomes the means for complexity reduction, even in situations where individuals must act under uncertainty with ambiguous and incomplete information. Therefore, it is not surprising that in the current age of global and digital economy and virtuality (Shepherd, 2004), there has been an overwhelming interest in trust. Motivated by the need to better understand trust in the digital era, this paper views the case of global virtual teams in commercial business organizations.

BACKGROUND

Trust has received significant recognition as a phenomenon worthy of detailed study in organizational and management studies (Dirks & Ferrin, 2001). In organizations, individuals must often act under uncertainty with ambiguous and incomplete information. This lack of explicit knowledge introduces risk and thus the requirement for trust. Accordingly, trust is defined as the willingness of a party to be vulnerable to the actions of another party (Mayer et al., 1995) based on a state of a positive, confident, though subjective, expectation regarding the behavior of somebody or something in a situation that entails risk to the trusting party (Baba, 1999; Cook & Wall, 1980; Currall & Judge, 1995).

Numerous scholars agree that trust is highly beneficial for the functioning of organizations. Trust "is at the heart of knowledge exchange" (Davenport & Prusak, 1998, p.35). High levels of trust are also key to effective communication (Dodgson, 1993) as they "improve the quality of dialogue and discussions … [that] facilitate the sharing of … knowledge" (Ichijo, von Krogh, & Nonaka, 2000, p.200), and committed relationships (ibid). The centrality of trust is further accentuated by its absence: "mistrust ... makes success harder to attain" (Kanter, 1994, p.105) as it weakens relationships, increases dependence on less information, compromises rational and unprejudiced analysis and exploration, and undermines learning (Luhmann, 1979). Furthermore, it has been recognized that if trust is not prominent, this may lead to dissatisfaction, absenteeism, and even intention to quit (Cunningham & MacGregor, 2000). At the inter-organizational level, trust also plays a vital role since it is found to affect the degree of cooperation among participating parties (Grabowski & Roberts, 1998; Newell & Swan, 2000). This is particularly important for virtual organizations. The business motivation for virtual arrangements is the potential for increased value-added and competitive advantage from the enhanced knowledge stock and core competencies, which are deemed to accrue to such networks (Alavi &Leidner, 2001).

Clearly, there is little dispute over the significance of trust in the organizational literature. However, there seems to be little agreement on how trust is developed and maintained in both the traditional and the virtual organizational literature.

In the traditional literature on trust where face-to-face communication is the norm, trust develops as the degree of familiarity with other people increases; i.e., the more we get to know others, the more likely it is that we trust them (Lewicki & Bunker, 1995,1996). Lewicki and Bunker (1996) take the view that trust varies over time and takes on a different character at the various stages (early, developing, and mature stages) of a relationship, as we not only begin to feel more comfortable with other people as we spend more time with them, but also as our knowledge of their integrity and competence improves. Based on this view, Lewicki and Bunker (1996) suggest three categories of trust, each corresponding to a different stage of the relationship:

- Calculus-Based Trust, the type of trust that is grounded in the rewards to be derived from pursuing and preserving the relationship or in the fear of punishment for violating trust within the relationship;
- Knowledge-Based Trust that assumes that the more information one has about others, the more able one is to predict their actions; and
- Identification-Based Trust, the type of trust that is characterized by mutual understanding among all

parties to the point that each can effectively act for the other.

These types of trust are "linked in a sequential iteration in which the achievements of trust at one level enables the development of trust at the next level" (p. 119).

Familiarity with other people has also been identified as an important antecedent of trust development in virtual teams. According to Handy (1995), for trust to develop in virtual environments there is a need for constant face-to-face communication. As he puts it: "paradoxically, the more virtual an organization becomes, the more its people need to meet in person" (Handy, 1995, p.46). This view has also been reinforced by Lipnack and Stamps (1997, p.226): "if you can drop by someone's office, see first-hand examples of prior work, and talk with other colleagues, you can more easily evaluate their proficiency." Researchers have already argued that the lack of proximity impersonalizes trust (Nandhakumar, 1999), while the virtual context of a geographically dispersed workforce may constrain or even impede rich information exchange[1] since communication becomes highly computer-mediated (Davenport & Pearlson, 1998). It follows, therefore, that trust based on familiarity with other individuals could not be easily developed in virtual settings.

In the following section, the challenges of developing trust in a virtual team setting are discussed by drawing upon the findings of existing empirical research.

Trust and Virtual Teams: Empirical Findings

While trust has been identified as a key feature for the success of virtual interactions, empirical research in this area has remained limited. Jarvenpaa and Leidner (1999) have conducted one of the most detailed research projects into studies on trust and virtual teams thus far. Their eight-week study of 75 teams of university students, each consisting of four to six members, highlighted significant differences in the behaviors and strategies between high- and low-trust teams and supported the existence of swift trust; this type of trust presumes that roles are clear and that each team member has a good understanding of others' roles and responsibilities (Meyerson, Weick, & Kramer, 1996).

However, trust is not always swift. Tucker and Panteli (2003) have illustrated the significance of shared goals and power in influencing trust development; these factors were not identified in the context of university settings as the tasks are often well-articulated in advance while power differentials, which could influence the degree of inter-dependence among members, are not significant in the case of university students. In business environments, however, power differentials prevail. Power, defined as the capability of one party to exert an influence on another to act in a prescribed manner, is often a function of both dependence and the use of that dependence as leverage (Rassingham, 1999). Indeed, power is an important contextual factor that affects trust (Hart & Saunders, 1997) in that it suggests the existence of a unilateral dependency or an imbalanced relationship (Allen, Colligan, Finnie, & Kern, 2000).

Accordingly, within a business environment where conflict and power differentials prevail, building trust is not always a swift process. Instead, it is found that the process of jointly constructing team goals holds significant value as it may provide the "glue" to hold team members together long enough to enable trust development.

Shared goals are and should be a key characteristic of virtual teams. They could provide a means to developing a common sense of identity for team members that can be of particular benefit to those global virtual teams who meet infrequently or perhaps not at all. These benefits include the establishment of a foundation upon which to build trust and minimize the use of coercive power in pursuit of a collaborative and productive relationship. However, the study finds that even though shared goals are important for the success of virtual teams, these should not be taken for granted. Indeed, goals may not be shared either because they do not exist at all, or because team members have not become aware of them, have their own priorities, or share different interpretations of the team's role. Furthermore, this study has also shown that the construction of shared goals is often not a one-off activity, but rather it is a process that requires the ongoing participation of all parties involved. Though this could be a time-consuming, iterative, and difficult process, these findings allow us to argue that it is far better to invest in it and as up front in the project as possible than deal with the vicious, destructive, downward spirals that result from team members with conflicting goals and poor levels of trust.

In considering power within virtual teams, there is an increasing recognition in the literature that knowledge is indeed power and that teams are often formed to create knowledge through combination and exchange. Within these teams, the team member with power at any given time is the one with the most relevant knowledge at that time. Tucker and Panteli (2003) found that in high-trust teams power differentials do not disappear; rather, power shifts from one member to another throughout the life cycle of a project depending on the stage and requirement of each stage.

Further to the issues of shared goals and power, Tucker and Panteli (2003) found support for the need for face-to-face interaction. However, the opportunities to meet face-to-face have been severely limited by economic pressures and, more recently, terrorist attacks. Under these circumstances, those virtual teams that work well tend to undertake regular communications via synchronous, "live" computer-mediated communication (CMC) such as the telephone and videoconferencing systems. Participants confirmed that synchronous media offered more feedback and therefore facilitated understanding more effectively than asynchronous technologies such as voicemail and e-mail. The use of asynchronous technologies was, however, regularly used for documenting and recording agreements and providing brief, simple updates to work progress. The teams that worked well also included a social and fun element in their interactions that appeared to help in creating a stronger shared social context.

Table 1 details the common features and behaviors observed within the global virtual teams studied in Tucker and Panteli (2003, p.91).

FUTURE TRENDS

The increasing pressure to perform in a short period of time with unfamiliar people and within a computer-mediated environment makes it imperative to study not only the type of trust but also how trust is formed and developed in a virtual team context.

It is readily acknowledged that what has been attempted here is only an exploration of contingencies to provide a better understanding of trust within the virtual team environment. There is no doubt that this is only the beginning of our understanding of trust in a virtual context.

Virtual interactions, however, exist at multiple levels—between individuals, individuals and organizations, and between organizations. Nandhakumar, Paneli, Powell, and Vidgen (2004, p.79) have introduced a digital era interaction (DEI) matrix to support exploration of trust relationships at these different levels.

The digital era interaction (DEI) matrix in Figure 1 indicates areas where digital era developments might be expected with consequent implications for trust. It views trust in two dimensions allowing an exploration of its nature in organization-to-organization settings, organization-to-individual and individual-to-organization interactions, and at an individual-to-individual level. In this paper, trust at the individual-to-individual level (and, more specifically, employee-to-employee) was explored. Thus, although the individual-to-organization (I2O) quadrant is currently the least developed of the four, it may prove to be an interesting sector in the longer term. Consumer-to-Business (C2B) is another growing area that allows consumers to organize themselves and use their stronger bargaining power to obtain a better price. As it was put: "The matrix is not exhaustive (for example, it does not include E2C, employee to consumer), but it does cater for the principal electronic relationships that currently exist and highlights ones that will be significant in the future" (p.79).

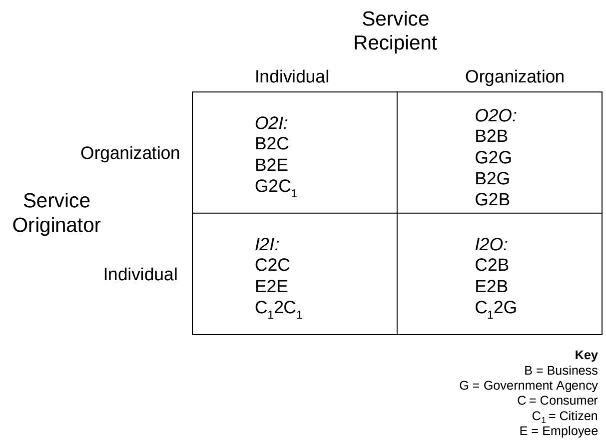
CONCLUSION

This paper reinforces arguments in the existing literature on the significance and complexity of trust dynamics in building effective virtual teams. It defined trust and outlined its central role in virtual interactions while also

Table 1. Differences between high-trust and low-trust global virtual teams

High-Trust Global Virtual Teams	***Low-Trust Global Virtual Teams***
Factors related to Shared Goals:	**Factors related to Shared Goals:**
*Awareness of shared goals	*Lack of awareness of shared goals
*Take time to build shared goals	*Lack of shared goals
*Open debate for shared goals up front	*Opinions of others not considered
*Team-based goals have primacy	*Individual goals take primacy
Factors related to Power:	**Factors related to Power:**
*Availability of facilitators	*Power battles
*Facilitators focus on win-win	*Coercion
*Recognition of knowledge as power	*Misunderstandings and conflict of interests
*Recognition that power moves; power in many places	*Use of positional power
*Power differentials are minimized	*Perception of 'I have Power'
Communication:	**Communication:**
*Face-to-face where possible	*Asynchronous CMC
*Regular synchronous CMC	* Time difference matters
*Social interaction	*Little or no social interest

Figure 1. Digital era interactions (DEI)



identifying some of the challenges involved in developing trust in the digital era, arguing that trust is necessary but not sufficient for promoting effective, collaborative virtual interactions using empirical findings. The paper illustrates the significance of shared goals and power in influencing trust development. It has also become apparent that while the agreement of shared goals provides a mobilizing force for the members of global virtual teams, the process of developing these goals holds significant value in terms of the exchange of information, learning, improving understanding, and an opportunity to demonstrate trustworthiness. Repositioning trust in this way increases our chances of making sense of complex virtual, computer-mediated situations, and puts us in closer touch with the challenges of developing trust. In doing so, the paper carries forward an important debate in the digital era.

REFERENCES

Alavi, M. & Leidner, D.E. (2001). Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly*, *25*(1), 107-136.

Allen, D., Colligan, D., Finnie, A., & Kern, T. (2000). Trust, power and inter-organisational information systems: The case of the electronic trading community TransLease. *Information Systems Journal*, *10*, 21-40.

Baba, M. (1999). Dangerous liaisons: Trust, distrust, and information technology in American work organizations. *Human Organization*, *58*(3), 331-346.

Cook, J. & Wall, T. (1980). New work attitudes measures of trust: Organizational commitment and personal need fulfilment. *Journal of Occupational Psychology, 53*(1), 39-52.

Cunningham, J.B. & MacGregor, J. (2000). Research note: Trust and the design of work: Complementary constructs in satisfaction and performance. *Human Relations*, *53*(12), 1575-1591.

Currall, S. & Judge, T. (1995). Measuring trust between organization boundary role persons. *Organization Behaviour and Human Decision Processes*, *64*(2), 151-170.

Daft, R.L. & Lengel, R.H. (1986). Organizational information requirements, media richness and structural design. *Management Science*, *32*(5), 554-571.

Davenport, T.H.& Pearlson, K. (1998). Two cheers for the virtual office. *Sloan Management Review*, Summer, 51-65.

Davenport, T.H. & Prusak, L., (1998). *Working knowledge: How organizations manage what they know*. Cambridge, MA: Harvard Business School Press.

Dirks, K.T. & Ferrin, D.L. (2001). The role of trust in organizational settings. *Organization Science*, *12*(4), 450-467.

Dodgson, M., (1993). Learning, trust and technological collaboration. *Human Relations*, *46*(1), 77-95.

Grabowski, M. & Roberts, K.H. (1998). Risk mitigation in virtual organizations. *Journal of Computer-Mediated Communication*, *3*(4), June.

Handy, C. (1995). Trust and the virtual organization. *Harvard Business Review*. May-June, 40-50.

Hart, P. & Saunders, C. (1997). Power and trust: Critical factors in the adoption and use of electronic data interchange. *Organization Science, 8*(1), 23-42.

Ichijo, K., von Krogh, G., & Nonaka, I., (2000). Knowledge enablers. In G. von Krogh, J. Roos, & D. Kleine (Eds.), *Knowing in firms: Understanding, managing and measuring knowledge*, pp. 173-203. London: Sage Publications.

Jarvenpaa, S.L. & Leidner, D.E. (1999). Communication and trust in global virtual teams. *Organization Science*, *10*, 791-815.

Kanter, R.M., (1994). Collaborative advantage: Successful partnerships manage the relationships, not just the deal. *Harvard Business Review*, July-August, 98-108.

Kramer, R.M. & Tyler, T.R. (1996). Whither trust. In R.M. Kramer & T.R. Tyler (Eds.), *Trust in organizations:*

Frontiers of theory and research. Thousand Oaks, CA: Sage Publications.

Lewicki, R.J. & Bunker, B.B. (1995). Trust relationships: A model of trust development and decline. In B. Bunker & J. Z. Rubin (Eds.), *Conflict, cooperation and justice*.San-Francisco: Jossey-Bass.

Lewicki, R.J. & Bunker, B.B. (1996). Developing and maintaining trust in working relationships. In R.M. Kramer & T.R. Tyler (Eds.), *Trust in organizations: Frontiers of theory and research*. Thousand Oaks, CA: Sage Publications.

Lipnack, J. & Stamps, J. (1997). *Virtual teams: Reaching across space, time, and organizations with technology*. New York: John Wiley & Sons.

Luhmann, N. (1979). *Trust and power*. London: John Wiley and Sons.

Mayer, R.C., Davis, J.H., & Schorman, F.D., (1995). An integrative model of organizational trust. *Academy of Management Journal, 20*(3), 709-734.

Meyerson, S., Weick, K.E., & Kramer, R.M. (1996). Swift trust and temporary groups. In R.M. Kramer & T.R. Tyler (Eds.), *Trust in organizations: Frontiers of theory and research*. Thousand Oaks, CA: Sage Publications.

Nandhakumar, J. (1999). Virtual teams and lost proximity: Consequences on trust relationships. In P. Jackson (Ed.), *Virtual working – Social and organizational dynamics*. London: Routledge.

Nandhakumar, J., Panteli, N., Powell, P., & Vidgen, R., (2004). Trust in the digital era. In N. Mylonopoulos, N. Pouloudi, & G. Doukidis, *Social and economic transformation in the digital era*. Hershey, PA: Idea Group Publishing.

Newell, S. & Swan, J. (2000). Trust and inter-organizational networking. *Human Relations, 53* (10), 1287-1328.

Rassingham, P. (1999). Risks in low trust among trading partners in electronic commerce. *Internet Research: Electronic Networking Applications and Policy*, *10*(1), 56-62.

Rousseau, D., Sitkin, S., Burt, R., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review, 23*(3), 393-404.

Shepherd, J. (2004). Why the digital era? In N. Mylonopoulos, N. Pouloudi, & G. Doukidis, *Social and economic transformation in the digital era*. Hershey, PA: Idea Group Publishing.

Tucker, R. & Panteli, N. (2003). Back to basics: Sharing goals and developing trust in global virtual teams. In N. Korpela, R. Montealegre, & A. Poulymenakou (Eds.), *Organizational information systems in the context of globalization.* Boston, MA: Kluwer Academic Publishers.

KEY TERMS

Computer-Mediated Communication: Communication that is facilitated using information technologies such as email, videoconferencing, teleconferencing.

Power: The ability to influence others.

Power Differentials: The existence of imbalanced power relationships.

Shared Goals: Goals that articulate what the teams stand for and their shared vision.

Social Interactions: A chain of interrelated messages that include a social and fun element and contribute to increasing familiarity among participants.

Trust: A state of a positive, confident though subjective expectation regarding the behavior of somebody or something in a situation that entails risk to the trusting party.

Virtual Teams: A group of geographically dispersed individuals who work on a joint project or common task and communicate electronically.

ENDNOTE

[1] This view corresponds to the media richness theory that argues that electronic media are lean (Daft & Lengel, 1986).

Diffusion of E-Learning as an Educational Innovation

Petek Askar
Hacettepe University, Turkey

Ugur Halici
Middle East Technical University, Turkey

INTRODUCTION

Most of the discussions related to education are about technological innovations. Indeed as Rogers (1995) stated, we often use the word "innovation" and "technology" as synonyms. Technology is regarded as an agent of change in educational settings, and a quick analysis of the educational projects all over the world shows us that it is not possible to define a future vision of education without technology, especially e-learning, which brings two important concepts together: technology and learning. Therefore as a form of distance learning, e-learning has become a major instructional force in the world.

Besides the technological developments, the last two decades have brought a tremendous increase in knowledge in education, particularly in learning. The emerging views of learning which should be taken into consideration for every learning environment could be stated as follows: personalized, flexible, and coherent (learning is connected to real-life issues); not bounded by physical, geographic, or temporal space; rich in information and learning experiences for all learners; committed to increasing different intelligences and learning styles; interconnected and collaborative; fostering interorganizational linkages; engaged in dialogue with community members; accountable to the learner to provide adaptive instructional environments (Marshall, 1997).

WWW is an environment that fits the new paradigm of learning and facilitates "e-learning" which faces a challenge of diffusion. Diffusion is defined by Rogers (1995) as the process by which an innovation is communicated through certain channels over time among the members of a social system. Therefore the adoption of WWW as a learning environment is influenced by the following set of factors: 1) the individuals' perception of the attributes of e-learning, 2) the nature of the communication channels, 3) the nature of the social system, and 4) the extent of the change agents' efforts in the e-learning. These are the variables that affect the diffusion of e-learning in the schools and countries.

E-LEARNING AND INSTRUCTIONAL DESIGN

E-learning not only opens up new ways of learning and teaching, but also leads to a new way of thinking and organizing learning content. Collaborations among different stakeholders cause new standards for design of knowledge on the Internet. In traditional computer-based instruction, content comes in units called courses. However a new paradigm for designing instruction, grounded in the object-oriented notion of computer science, is called "learning objects."

Learning object is defined by the Learning Technology Standards Committee (2002) of the Institute of Electrical and Electronics (IEEE) as any entity, digital or non-digital, that can be used, re-used, or referenced during technology-supported learning. The features of learning objects are self-contained, interactive, reusable, and tagged with metadata. By the use of learning objects, one can learn just enough, just in time, and just for them. Learning objects can be considered a movement within the field of e-learning, one aimed at the componentization of learning resources, with a view to reusability (Duchastel, 2004).

The idea of educational software as a package is becoming outdated and making way for learning objects as a new way of designing instructional materials. In designing learning objects, the studies on multiple representation of knowledge become important since people have different learning styles and strategies. The associations between these two constructs are the main focus of the new instructional design principles. Therefore, the development of learning objects and the way of creating teaching units are well suited for what we call the Information Age.

A representation of knowledge could be decomposed into its parts, where the parts are far from arbitrary. Then they can be used and reused in a great variety of combinations, like a child's set of building blocks. Every combination is meaningful and serves as an instructional whole. Holland (1995) compares building blocks to the features of the human face. The common building blocks are: hair, forehead, eyebrows, eyes, and so on. Any combination is different and may never appear twice. This analogy could be true of e-learning platforms, where learning objects are put together to make up a meaningful whole, which we call instructional materials.

The five fundamental components of instructional design process are learners, content, objectives, methods, and assessment. Hence, for a systematic instructional design of a subject matter, the basic steps are: learner characteristic identification, task analysis, objectives, content sequencing, instructional strategies, message design, and instructional delivery and evaluation.

The awareness of learner differences with respect to entry competencies, learning styles and strategies, motivation, and interest are critical. However it is difficult to accomplish this task by using ongoing approaches. Indeed, new technologies, if used properly, enable us to make the lessons more individualized. The Learning Objects Metadata Working Group (IEEE, 2002) stated its goal as: to enable computer agents to automatically and dynamically compose personalized lessons for an individual learner. This leads a paradigm shift to approaches to instructional design. As Wiley (2001) stated, a problem arose when people began to actually consider what it meant for a computer to "automatically and dynamically compose personalized lessons." It seems that the idea of learning objects is challenging, but opens to new concepts, strategies, and research areas in the instructional design process.

E-LEARNING AND SCHOOL MANAGEMENT

For most of the last two decades, technology has been implemented in schools, and its potential to change the educational systems has been argued for. There are tremendous efforts to encourage the integration of computers and Internet into schools. However, in one of the diffusion studies conducted by Askar and Usluel (2001), two paths to the adoption of computers are presented. One path is related to the use of technology in the school management system; the other one is related to the use of technology in the teaching and learning process. For many reasons the rate of adoption of computers in management applications is quicker than the learning-teaching applications. Indeed the concerns related to use of computers in the teaching-learning process are still at the awareness stage. On the other hand, the need for using computers and the Internet for management purposes is more relevant and seems more convenient for the current school system.

Educators assert that the central purpose of school management systems should be to improve instructional program quality. In light of this idea, a typical configuration of a Web-based school management system designed—taking this idea into consideration—includes administration, assessment, and communication. The features are: student enrollment, attendance, registration, test scores, grades and other record-keeping tasks, formative and summative evaluation, and feedback to parents and teachers about student learning and performance. In addition, new online management systems include item- banking capability for adaptive testing and online learning modules.

E-LEARNING AND THE COMMUNITY

The modern world requires individuals and communities to be able to continually develop and utilize different skills and knowledge. There is growing consensus among OECD countries that modern economies cannot afford a significant number of uneducated people (OECD, 2000). However, education systems throughout the world are ill equipped to address individual and community learning needs. The existing school system is not flexible for those who for some reason left school early.

Distance education is a recognized solution all over the world for bridging the learning and education divide between the educated and poorly educated. It gives people the opportunity to continue their formal education. Despite the initial concerns that distance education might be lower in quality than traditional method of schooling, many forms of distance education are gaining acceptance (Belanger & Jordan, 2000). Therefore distance education is receiving positive attention from governments as a solution to the educational problem mentioned above.

Also, the trend towards lifelong learning is universal. The transformations taking place in all societies require an increasing participation of individuals, an ability to innovate and solve problems, and a capacity to learn and go on learning (Mayor, 1994). Moreover, the term "open learning" is used to lower barriers that stand in the way of individuals and communities wishing to engage in different learning opportunities.

One of the solutions for the above mentioned problems is learning centers, which are flexible learning organizations and which serve the learning needs of the individuals and communities. A school that is well equipped and organized could be opened during non-traditional school hours. Therefore, schools as learning centers can be critical resources to meet the growing need for distance education students and other community members. However, in highly centralized education systems, it is very difficult to organize schools for those other than the registered students. The rules and regulations for conventional school become real barriers for open learning environments.

FUTURE TRENDS IN TECHNOLOGY

While adopting the current technology for enhancing teaching and learning, advances in micro- and nanotechnology push the limits of miniaturization, and of minimizing the costs and power of microelectronic components and micro-systems. Explorations of alternative materials are expected to allow organic flexible materials for displays, sensors, and actuators so that they can be placed anywhere and can take any shape. Furthermore it is expected that not only PCs, but also all our surroundings, will be interfaced. Instead of only " writing and reading" in human-computer interaction, all senses are to be used intuitively. Information search will be context-based instead of "word " based. Mobile and wireless devices will be used not only for voice transfer, but also for full multimedia (IST WP, 2002)

As information and communication technologies change, open systems and services are to be developed in support of ubiquitous, experiential and contextualized learning, and virtual collaborative learning communities improving the efficiency and cost-effectiveness of learning for individuals and organizations, independent of time, place, and pace. Next-generation learning solutions are expected to combine cognitive and knowledge-based approaches, with new media having intelligence, virtual and augmented reality, virtual presence, and simulation (ISTC, 2001).

CONCLUSION

Our current educational system is highly resistant to change. Making a client-based change rather than a technological based one will be the most important innovation to accomplish for the educational change. While technology is pushing the limits of e-learning environments, special care should be taken in educational and organizational frameworks. The stakeholders of the systems are students, teachers, principals, learners, and community. Their attitudes, needs, and expectations from e-learning are important issues for the change process. The innovation adoption variables of relative advantage, compatibility, visibility, ease of use, results demonstrability, and triability should be considered by school administrators seeking to increase the rate of adoption of e-learning within their organization (Jebeile & Reeve, 2003).

Complexity is another issue to be considered. Fullan (1991) defines complexity as the difficulty and extent of change required of the individuals responsible for implementation. Therefore, there should be an emphasis on simplifying the process of e-learning, while moving from approaches based on knowledge transfer to systems based on the dynamic construction and user-friendly exchange of knowledge between learners, teachers, and learning communities.

The educational community, as the end user of e-learning systems, should be given the opportunity of observing and trying the e-learning systems. Awareness or being informed about the innovation is the key factor for changing the negative attitudes or beliefs. It is known that if people see the implementation and results of innovation, they are more likely to adopt them for their usage. Unfortunately, the benefits of e-learning are not well known and well recognized by all relevant stakeholders. Therefore, a comprehensive and systematic awareness campaign is needed to speed up the rate of adoption.

REFERENCES

Askar, P. & Usluel, Y. (2001, March 5-10). Concerns of administrators and teachers in the diffusion of IT in schools: A case study from Turkey. *Proceedings of the 12th International Conference of Society for Information Technology and the Teacher Education*, Orlando, Florida. Retrieved from *www.aace.org/dl/index.cfm/fuseaction/View/papered/3970*

Belanger, F. & Jordan, D.H. (2000). *Evaluation and implementation of distance learning: Technologies, tools and techniques.* Hershey, PA: Idea Group Publishing.

Duchastel, P. (2004) Learning objects and instructional design. *Interactive Technology and Smart Education*, *1*(1), 67-70.

Holland, J.H. (1995). *Hidden order: How adaptation builds complexity*. Perseus Books.

IEEE Learning Technology Standards Committee. (2002). *Draft standard for learning object metadata.* Retrieved November 22, 2002, from *ltcs.ieee.org/wg12/LOM_1484_12_1_V1_Final_Draft.pdf*

IST WP. (2002, December 17). *IST priority workprogramme 2003-2004, information society technologies.* Retrieved from *fp6.cordis.lu/fp6/call_details.cfm?CALL_ID=1*

ISTC. (2001, October 17). *Technology supported learning, ISTC–Information Society Technologies Committee.* Final Report from the Working Party on Education and Training, Luxembourg. Retrieved from *www.proacte.com/downloads/eandt/ISTC-CONsolidated-report-et.DOC*

Jebeile, S. & Reeve, R. (2003) The diffusion of e-learning innovations in an Australian secondary college: Strategies and tactics for educational leaders. *The Innovation Journal*, *8*(4), 1-21.

Marshall, S.P. (1997). Creating sustainable learning communities for the twenty-first century. In F. Hesselbein et al. (Eds.), *The organization of the future* (pp. 177-188). San Francisco: Jossey-Bass.

Mayor, F. (1994). Lifelong learning for the 21st century. *Proceedings of the 1st Global Conference on Lifelong Learning*, Rome, UNESCO, DG/94/39. Retrieved November 17, 2002, from *unesdoc.unesco.org/ulis/dgspeech other.html*

OECD/National Centre for Adult Literacy. (2000). *Learning to bridge the digital divide (schooling for tomorrow)*. France: OECD Publications.

Rogers, E.M. (1995). *Diffusion of innovations*. New York: The Free Press.

Shepherd, C. (2000). Objects of interest. Retrieved November, 18, 2002, from *www.fastrak- onsulting.co.uk/tactix/Features/perfect_etutor.htm*

Wiley, D.A. (2001). Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. In D.A. Wiley (Ed.), *The instructional use of learning objects*. Bloomington, IN: Association for Educational Communications and Technology. Retrieved February, 12, 2004, from *wiley.ed.usu.edu/articles.html*.

KEY TERMS

Diffusion: The process by which an innovation is communicated through certain channels over time among the members of a social system.

Distance Education: A type of formal education in which the majority of the instruction, which is transmitted through technology, occurs when student and instructor are not in the same place.

Innovation: An idea, practice, or object that is perceived as new by an individual or other unit of adoption.

Instructional Design: The systematic process of translating general principles of learning and instruction into plans for instruction and learning.

Learning Object: Any entity, digital or non-digital, that can be used, re-used, or referenced during technology-supported learning.

Learning Styles: The ways in which a person takes in and processes information and experiences.

Rate of Adoption: The relative speed with which an innovation is adopted by members of a social system.

Diffusion of Innovations in Organisations

Davood Askarany
University of South Australia, Australia

INTRODUCTION

During the past two decades, the world has witnessed a significant change in the nature of the technological and administrative practices and processes faced by organisations in different areas of their operations, such as manufacturing processes, operation technologies, and information systems (Shields, 1997). To keep pace with other competitors in the global market, organisations are keen to be able to use the best and the latest ideas, techniques, practices, and processes in different aspects of their activities. This has placed a greater emphasis on the diffusion of innovation as a solution to cope with the requirements of such changes. This view is consistent with systems approach theory that suggests that all parts of a system are related to each other, and any change in one part of a system may require the consideration of appropriate change(s) in other parts of the organisation, otherwise, the system may not work properly (Kellett & Sweeting, 1991).

An overview of the diffusion of advanced techniques and the recognition of factors influencing the diffusion of such innovations is expected to facilitate the implementation of recently developed, advanced, and up-to-date techniques and practices in organisations. Given this, the current chapter explains a diffusion model applicable to studies investigating the diffusion of advanced techniques (both technological and administrative innovations). The model incorporates most of the innovation factors addressed in the diffusion and advanced techniques literature. An "advanced technique" is referred to as an innovation in this overview.

BACKGROUND

Rogers (1995) defined an innovation as "an idea, practice, or object that is perceived as new by an individual or other unit of adoption." Further, he suggested that if the individual has no perceived knowledge about an idea and sees it as new, it is an innovation. Likewise, Damanpour and Gopalakrishnan (1998) defined innovation as "the adoption of an idea or behaviour new to the organisation." The common criterion in any definition of innovation is newness. According to Rogers (1995), newness in an innovation might be expressed not only in terms of new knowledge, but also in terms of first persuasion, or a decision to adopt. The second element that needs some clarification is diffusion: Wolfe (1994) explained diffusion of an innovation as a way new ideas are accepted (or not) by those to whom they are relevant. Rogers (1995) extended this definition to consider diffusion as a process by which an innovation is communicated through certain channels over time among the members of a social system.

A clear understanding of the complexities of the innovation process and of alternative diffusion methods is central to any innovation diffusion study. Depending on the source of innovation, the diffusion of the innovation might follow different stages, so that alternative approaches and perspectives might be applicable under different innovation diffusion processes.

According to Damanpour and Gopalakrishnan (1998), diffusion of innovations in organisations takes place in two ways: generation and adoption. In the case of generation, innovations are generated by organisations for their own use or for export to other organisations. In the case of adoption, innovations are imported into the organisation for adoption. The process of adoption of an innovation is a very long and difficult process, especially because many innovations need a long period of time to become widely adopted (Rogers, 1995). Rogers further emphasised that increasing the diffusion rate of an innovation is a common problem for potential adopters of that innovation.

The process of innovation diffusion is different when the innovation is generated by the organisation; in this case, the main stages include the stages of idea generation, project definition, design, development, and marketing and commercialisation (Cooper & Kleinchmidt, 1990). In the case of adoption of an innovation, which has been developed outside the organisation, the stages will be awareness of innovation, attitude formation, evaluation, decision to adopt, trial implementation, and sustained implementation (Zaltman, Duncan, & Holbek, 1973). Furthermore, Wolfe (1994, p. 411) added that when innovation is generated in the organisation, the stages "tend to be mulled and overlapping," while in the case of adoption, the stages "tend to occur in the expected order." Depending on whether the innovation is generated within or adopted by an organisation, two alternative general models can be formulated to describe the diffusion process.

Contributing to the diffusion of innovation literature, Rogers (1995) suggested that there are six phases for the diffusion of an innovation: recognition of a problem or need, basic and applied research, development, commercialisation, diffusion and adoption, and consequences. Given this explanation, Rogers emphasised that these six phases are somehow arbitrary, as they might not always occur in order, and some of them might be skipped in the case of particular innovations. An innovation development consists of all decisions and activities and their impacts that occur during these phases. These suggested stages for innovation development are largely consistent with the generation approach of Damanpour and Gopalakrishnan (1998).

However, Zahra and Covin (1994) adopted a different perspective, suggesting that there are three major sources of innovation: imitative, acquisitive, and incubative. They defined these three major sources of innovation as follows. Imitative sources are those innovations that are first introduced by other firms and then copied by organisations. Acquisitive sources also include those innovations that have been developed by other firms but are acquired through purchase, licensing, acquisition, or merger. Finally, incubative sources are those innovations that have been developed in organisations for their own use. This categorisation is compatible with the generation and adoption approach of Damanpour and Gopalakrishnan (1998) in that "imitative innovations" and "acquisitive innovations" can be classified as "adopted" innovations and "incubative innovations" as "generated" ones.

From a process point of view, Rogers (1995) divided the innovation process in organisations into two subprocesses: an initiation process and an implementation process. The initiation process includes two stages: agenda setting and matching. These two stages involve all activities such as information gathering, conceptualising, and planning for the adoption of an innovation. The implementation process includes three stages: redefining/restructuring, clarifying, and routinizing. These three stages contain all of the actions, events, and decisions involved in implementing an innovation. This classification is again consistent with the adoption method explained by Damanpour and Gopalakrishnan (1998). However, regardless of types, phases, or sources of innovations, there are a variety of factors influencing the diffusion of advanced techniques in organisations. Reviewing the diffusion literature, Askarany (2003) summarised a variety of influencing factors and established a diffusion model applicable to studies investigating the diffusion of advanced techniques.

A DIFFUSION MODEL

Addressing the diffusion of advanced techniques, Askarany (2003) classified all influencing factors into three main categories: factors related to attributes of innovations, to adopters of innovations, and to the social system.

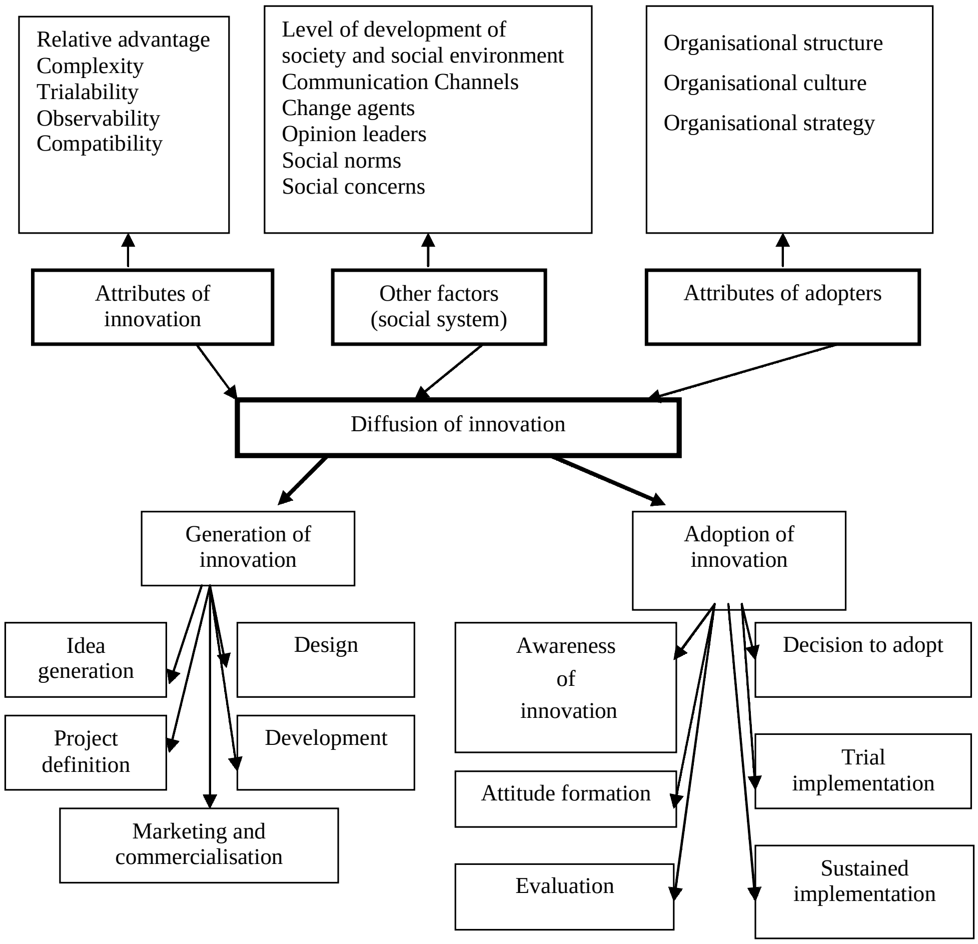
Following Rogers (1995), attributes of innovations include five categories: relative advantage, compatibility, complexity, trialability, and observability. These five categories include a variety of influencing factors, such as the degree of uncertainty associated with the innovation; the amount of investment required to adopt the innovation; the extent of economic advantage of an innovation; continuity of the innovation progress; overall benefit of an innovation (including economic and noneconomic advantage of an innovation); reinventing and dynamics aspects of innovations; profitability, flexibility, and capability of modification of an innovation; availability of an innovation and the information about it for potential adopters; and the type of innovation.

Factors related to the adopters of innovations include three categories: organisational strategy, organisational structure, and organisational culture. In other words, most characteristics of organisations can be explained by these three categories. These categories might include factors such as size of organisations, aggressiveness and innovativeness of their managers, level of information of organisations about the innovation, learning perspectives of organisations, resistance to change, technical skills of the users of an innovation in organisations, competition, and awareness of an innovation as a possible solution or as an available technique for progress.

Factors related to social system include the level of development of a society, communication channels in a society, social concerns, change agents, opinion leaders, and social norms. It might also be possible to include all of the influential factors that could not be related to the innovation category or the adopter's category under a social system category.

Given the above classification, Askarany (2003) suggested that the following general diffusion model (Figure 1) can be developed. This diffusion model is highly likely to be applicable to any diffusion study with minor modifications. Under this model, in general, what makes a diffusion research different from other diffusion research is the type or group of influencing factors and the kind or the number of advanced techniques. So, depending on the type of influencing factors and kind of advanced techniques, a more detailed model can be adopted from the

Figure 1. General diffusion model



suggested general model and modified to tailor a specific diffusion research. Also, depending on the research approaches (generation of innovation or adoption of innovation), the central focus of the research might change.

One of the main advantages of the presented model is that it is not only a comprehensive model that distinguishes between generation and the diffusion of innovations, but it is also a simplified model that classifies most influencing factors addressed in the literature into three main categories.

The above diffusion model does not give priority to any influencing groups. However, according to Rogers (1995), the accumulated body of literature on diffusion indicates that much effort has been devoted to studying the innovativeness and determining the characteristics of adopters, while relatively little effort has been spent on investigating how the "properties" of innovations affect their rates of adoption. He suggested that between 51% and 87% of variance in the adoption rate of innovations can be explained by the *characteristics of innovations*, suggesting that it is these that have the most significant influence on their diffusion. Highlighting this view, recent studies investigating the impact of other influencing factors (e.g., characteristics of innovators and social systems) on the diffusion of administrative innovations have failed to explain the majority of variances in the diffusion of such advanced techniques (e.g., Askarany & Smith, 2003; Beng, Schoch, & Yap, 1994; Bork & Morgan, 1993; Chenhall, 2003; Gurd, Smith, & Swaffer, 2002; Jazayeri & Hopper, 1999; Krumwiede, 1998; Lukka & Shields, 1999; Rogers, 1983, 1995; Williams & Seaman, 2001).

FUTURE TRENDS

The suggested diffusion model in this chapter is highly likely to be applicable to any future diffusion study with minor modifications. By classifying all influencing factors into three main categories, the model proposes that factors related to the characteristics of innovations are among the first influencing factors that determine whether there is a need for change, innovation, or diffusion of innovation. If an advanced technique is expected to replace an applied technique, the first question in considering such a technique for implementation by an organisation would be as follows: what are the advantages or disadvantages of the new technique compared with the existing technique? Before thinking about the other influencing factors, such as organisational culture, strategy, structures, and so on, potential adopters of an advanced technique would ask about the benefits that such an advanced technique is going to offer and that might not be achievable with their current techniques. Even when no replacement is involved, the characteristics of that advanced technique are still among the first factors that potential adopters would consider when adopting such an innovation. Further research is needed to determine the level of the impact of characteristics of innovations on their diffusion.

CONCLUSION

The technological and administrative changes of the last two decades have placed a greater emphasis on the diffusion of advanced techniques as a solution to cope with the requirements of such changes. To facilitate the implementation of advanced and up-to-date techniques and practices in organisations, this chapter explains a diffusion model clarifying different diffusion processes and influencing factors. The developed model suggests that characteristics of innovations are among the most important factors to organisations, managers, consultants, and other decision makers in making decisions regarding implementing such innovations.

REFERENCES

Askarany, D. (2003). An overview of the diffusion of advanced techniques. In B. F. Tan (Ed.), *Advanced topics in global information* (Vol. 2, pp. 225-250). Hershey, PA: Idea Group Publishing.

Askarany, D., & Smith, M. (2003). The relationship between technological innovation, activity-based costing and business size. Paper presented at the *Information Science + Information Technology Education Joint Conference*, Pori, Finland.

Beng, A. K., Schoch, H., & Yap, T. (1994). Activity based costing in the electronics industry: The Singapore experience. *Small Business and Enterpreneurship, 11*(2), 28-37.

Bork, H. P., & Morgan, M. J. (1993). Is ABC really a need not an option? *Management Accounting Research, 71*(8), 26-27.

Chenhall, R. H. (2003). Management control systems design within its organisational context: Findings from contingency-based research and direction for the future. *Accounting, Organizations and Society, 28*, 127-168.

Cooper, R. G., & Kleinchmidt, E. J. (1990). New product success factors: A comparison of "kills" versus success and failures. *R & D Management, 20*, 47-63.

Damanpour, F., & Gopalakrishnan, S. (1998). Theories of organizational structure and innovation adoption: The role of environmental change. *Journal of Engineering and Technology Management, 15*, 1-24.

Gurd, B., Smith, M., & Swaffer, A. (2002). Factors impacting on accounting lag: An exploratory study of responding to TQM. *British Accounting Review, 34*, 205-221.

Jazayeri, M., & Hopper, T. (1999). Management accounting within world class manufacturing: A case study. *Management Accounting Research, 10*, 263-301.

Kellett, B. M., & Sweeting, R. C. (1991). Accounting innovation and adoptions: A U.K. case. *Management Accounting Research, 2*(1), 15-26.

Krumwiede, K. R. (1998). The implementation stages of activity-based costing and the impact of contextual and organizational factors. *Journal of Management Accounting Research, 10*, 239-277.

Lukka, K., & Shields, M. (1999). Innovations in management accounting focus. *Management Accounting, London, 77*(3), 33-38.

Rogers, E. M. (1983). *Diffusion of innovations*. New York: Free Press.

Rogers, E. M. (1995). *Diffusion of innovation* (4th ed.). New York: Free Press.

Shields, M. D. (1997). Research in management accounting by North Americans in the 1990s. *Journal of Management Accounting Research, 9*, 3-61.

Williams, J. J., & Seaman, A. E. (2001). Management accounting systems change and departmental performance: The influence of managerial information and task uncertainty. *Management Accounting Research, 13*, 419-445.

Wolfe, R. A. (1994). Organizational innovation: Review, critique and suggested research directions. *Journal of Management Studies, 31*(3), 405-431.

Zahra, S. A., & Covin, J. G. (1994). The financial implications of fit between competitive strategy and innovation types and sources. *Journal of High Technology and Management, 5*, 183-211.

Zaltman, G., Duncan, R., & Holbek, J. (1973). *Innovations and organizations*. New York: Wiley.

KEY TERMS

Compatibility of Innovation: The degree of consistency of an innovation with the needs, expected values, and norms of potential adopters and their social systems.

Complexity of Innovation: The degree to which an innovation seems difficult to understand and use.

Diffusion: A process by which an idea, product, practice, behaviour, or object is communicated and circulated to those to whom it is relevant.

Innovation: Any idea, product, practice, behaviour, or object that is apparent as new.

Observability of Innovation: The degree to which the results of an innovation can be observed or demonstrated.

Relative Advantage of Innovation: The degree to which an innovation seems to be better than the idea, object, practice, or process, which is replacing.

Trialability of Innovation: The degree to which an innovation can be tried on a limited basis before full implementation.

Diffusion Patterns of the Internet Technology Cluster in Irish SMEs

James Griffin
Tipperary Institute, Ireland

INTRODUCTION

Electronic commerce has been recognised as a source of fundamental, pan-sectoral change to the conduct of business; Chan and Swatman (2000) use the term: "A new paradigm for doing business." Other authors have gone further, viewing modern IT developments as the latter part of a period starting in the mid-1970s that represents a transition to nothing less than a new phase of capitalist development (Amin, 1994). Benjamin, Rockhart, Scott Morton, and Wyman (1983) also suggest that the world economy has been fundamentally altered by the globalisation of competition which has largely been caused by the declining cost and consequent increasing spread of IT developments.

The resulting shift in business practices as businesses attempt to exploit these new opportunities will necessitate wide-scale adoption of new processes and technologies. Elliot and Loebbecke (2000) suggest that this requires new thinking on how organisations adopt innovations and the revision of theoretical models of adoption. Bamfield (1994) identifies innovation theory as an appropriate framework for understanding IT adoption processes. La Rovere (1998) concurs, stating that the diffusion of innovations in information technology (IT) is becoming an increasingly important area of study.

Furthermore, any overview of recent Internet-related literature will identify that the issue of mapping diffusion patterns is being increasingly affected by the range and variety of technologies that are drawn into the e-business platform. In terms of understanding the nature of Internet usage and diffusion within SMEs, it is necessary to individually measure the extent to which different elements of what essentially comprise an Internet Technology Cluster are used amongst adopting firms.

This necessitates a definition of the different elements of the Internet Technology Cluster. This can be accomplished through the analysis of past surveys and technical articles written in the field of Internet research. Three basic elements are identifiable. Firstly, several studies have identified e-mail as the most common Internet application used in business (Howe, 2001; Everett, 1998; Feher & Towell, 1997).

Secondly, many of the most common Internet technologies and applications centre around the Internet browser. Graphics, audio, HTML, and HTTP technologies are all involved in the presentation of Web sites to the viewer via the browser, whilst research and communications applications such as search engines, newsgroups and discussion groups, and online journals are viewed via the browser. These elements can be combined together under the banner of Internet browser applications.

Finally, more complex technologies based around back-end activities and remote access to Internet services (through FTP, WAP, and Telnet) can be grouped together to give an indication of the extent of usage of more advanced Internet applications.

BACKGROUND

During the summer of 2001, the author conducted a detailed survey of SMEs in six distinct geographical regions of the Republic of Ireland (three urban and three rural). The survey examined current Internet usage levels, the factors influencing the adoption decision process, and the actual benefits achieved by SMEs that have adopted the Internet.

A stratified sample frame of 700 companies was chosen from the population of 3,500 SMEs. A total usable response rate of 153 responses was achieved. Following readjustments to the sample frame size, the overall response rate for the survey stands at 23.4%.

As highlighted in Table 1 below, the current level of Internet connectivity within Irish SMEs is very high (over 90% of firms). This is no doubt a function of the high level of awareness that has been created over recent years. However, problems arise when the nature of this connectivity is examined in more detail, through the application of the cluster-based definition of the Internet.

The 90.8% of respondents with Internet access were asked to rate the extent to which they used the three separate elements of the Internet Technology Cluster. Rating of each element was on a five-point Likert scale.

In accordance with previous literature, *e-mail* was the most used Internet application, with an average usage

Table 1. Key findings related to the extent of Internet usage in Irish SMEs

	Yes		**No**	
Does your firm currently have Internet access? (n=153)	90.8% (n=139)		9.2% (n=14)	
Does your firm currently have a Web site? (n=139)	52.5% (n=73)		47.5% (n=66)	
	ISDN	**LAN**	**Dial-up**	**Unsure**
Type of connection (n=139)	23% (n=32)	8% (n=11)	62% (n=86)	7% (n=10)

rating of 3.43. *Internet browsers* received an average rating of 2.82, whilst the *Advanced Applications* category received an average usage rating of 1.43.

Only 52.5% of firms with Internet connections have a Web site of their own; this is lower than would have been expected for a 90% rate of connectivity. The type of physical Internet connection is generally a simple dial-up (62%), with few firms engaging in higher speed connections. A worrying 8% of owner managers could not even describe their connection type, suggesting a lack of understanding of, or degree of separation from, Internet applications in their organisation.

Of those respondents without a Web site, 67% did express an intention to set one up within the next 12 months. This fact, coupled with the steadily increasing Web site adoption rate depicted in Figure 1, indicates a pattern that is strongly reminiscent of the types of mimetic and bandwagon diffusion patterns described by Abrahamson and Rosenkopf (1993), DiMaggio and Powell (1983), and O'Neill, Pouder, and Buchholtz (1998). Mimetic and bandwagon diffusion patterns can best be defined as self-reinforcing diffusion patterns that effectively dismiss the technical attributes or properties of an innovation, with its level of adoption instead being a factor of the number of adoptions that have already taken place. Given the coverage and attention that the Internet has received in media, commercial, and academic circles over the past five years in Ireland, it is unsurprising that a mimetic form of diffusion pattern is evident amongst SMEs in Ireland.

CRITICAL FINDINGS

To explore the supposition that Internet technology diffusion in SMEs in Ireland may be following a mimetic pattern, it is helpful to review the main characteristics of mimetic adoption patterns. A review of relevant diffusion theory highlights two key characteristics.

Firstly, many authors, through the examination of the EDI adoption experience in SMEs, have identified a key characteristic of mimetic adoptions to be a reduction in benefit amongst organisations partaking in reactionary and later adoptions vis-à-vis firms engaging in earlier, more strategic adoptions (see Cash, 1985; Friedman, 1998; Raymond, Julien, Carrier & Lachance 1996; Swatman & Swatman, 1992).

Figure 1. Diffusion pattern of Web site development

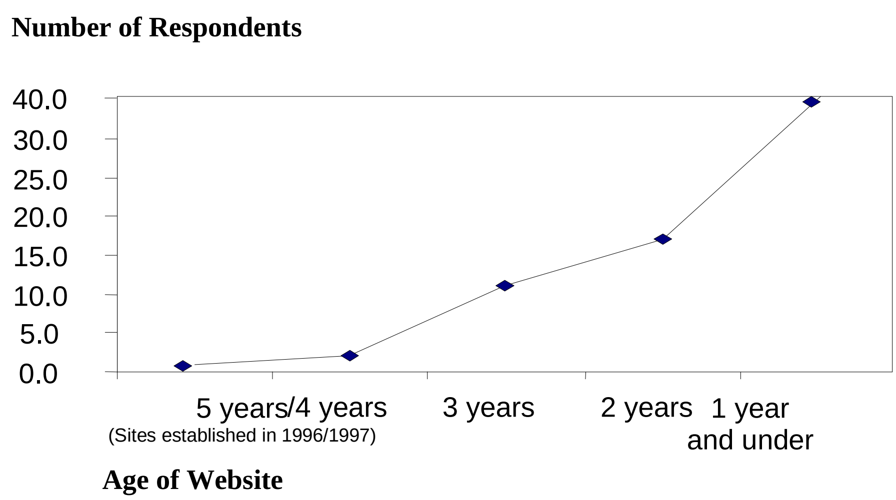
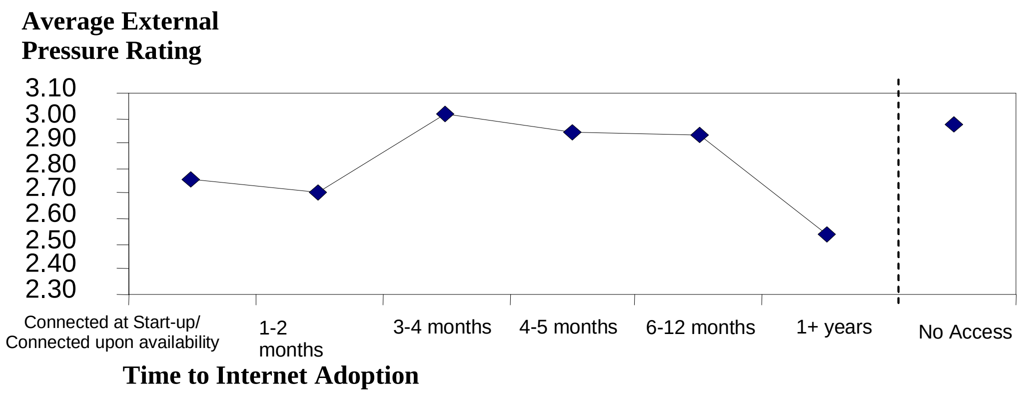


Table 2. Business benefits received compared to Web site age

	Website usefulness factor					
	A	**B**	**C**	**D**	**E**	**F**
	Direct selling	Inquiries for offline sales	Advertising the firm	Information to customers	Informing suppliers/ retailers	Providing after-sales service
Ratings for sites over 1.5 years in existence	1.5	2.7	3.1	3.0	2.2	1.7
Ratings for sites 1.5 years and under in existence	1.4	2.1	2.7	3.1	2.2	1.4
Overall Average Ratings	**1.42**	**2.42**	**2.88**	**3.08**	**2.17**	**1.62**

Figure 2. External pressure felt by respondents (grouped by time taken to adopt)



To identify any significant differences between early and late adopters the average age of respondents' Web sites was calculated. The average age of Web sites in the sample was 1.5 years. Table 2 facilitates comparison between respondents with older and newer Web sites, and average Web site usefulness ratings.

Differences between early and late adopters in terms of benefit accrued through Web site adoption are surprisingly minimal, with both sets of ratings tending towards the lower end of the scales.

This finding highlights a major issue for the diffusion of Web technologies, namely that with low benefits accruing to many early adopters in a majority of business categories, the subsequent poor results of adoptions in these firms will represent a significant weakening of the vital '*Reputation Effect*' that sustains mimetic or bandwagon patterns. The result of this reduction in the reputation effect may well be that the sharp increase in uptake that normally gives rise to the early and late majorities in other mimetic adoption patterns may not occur in the case of the Internet and specifically in the case of Web site technologies.

Secondly, a feature of mimetic adoption patterns is that many firms, during the adoption decision process, feel increasing levels of pressure from their external environment. Authors such as DiMaggio and Powell (1983) and Pierce and Delbecq (1977) identified pressure from many elements of the multi-participant environment as creating a considerable impetus amongst later adopters to acquire the IT innovation in question. Amongst the research sample, one would therefore expect to find the level of external pressure increasing the longer a respondent takes to decide to adopt (indicating a reduction in voluntariness over time as the reputation of the innovation grows). It is evident from Figure 2, however, that the level of external pressure felt by respondents does not closely follow such a pattern.

Specifically, two interesting findings arise from Table 2 that merit deeper examination. Firstly, considerably high external pressure to adopt the Internet was reported by respondents with no Internet connectivity at all, suggesting that despite such pressure, others factors are strong enough to overcome the pressure.

Secondly, 'slow adopters' (i.e., respondents who considered their adoption for over a year) exhibited a major drop in perceived external pressure. These results suggest that the influence of peer networks and the pressures they bring to bear in relation to the Internet adoption decision process are decreasing over time, most likely due to low levels of benefits accruing to the adopting firms. This represents a significant shortening of what Thorelli (1986) termed 'innovation poles', described as the influence of peer networks (or 'macro-cultural' pressure) elsewhere in e-business literature. This is indicative of the reduction in the reputation effect of Internet technologies posited earlier, and strongly suggests that this effect has essentially reached its peak and is now beginning to decline, rather than increase over time.

FUTURE TRENDS

Consideration of these issues suggests an Internet diffusion pattern displaying an initially mimetic pattern. Difficulties arise however when we consider how that diffusion pattern is likely to develop in the future.

With the evident reduction in the reputation effect of the Internet, it is unlikely that a high rate of mimicry will continue. With this in mind, different potential diffusion patterns become apparent. These potential patterns centre on the questions: When will the late majority arrive? and Will there be a late majority?'

One conclusion from these findings may be that we may have arrived (or are arriving) at the diffusion 'plateau' described by authors such as Cooper and Zmud (1990) and Kautz and Larsen (2000) in traditional IT/IS innovation diffusions.

Figure 3. The Elongated Diffusion Pattern

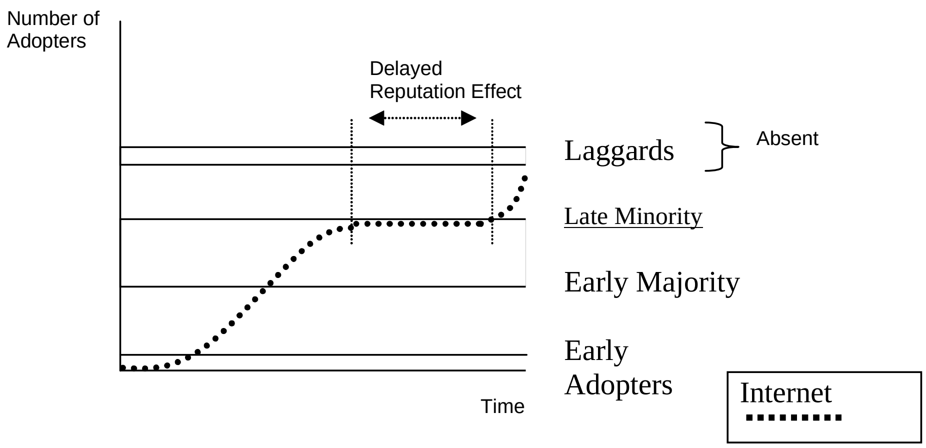
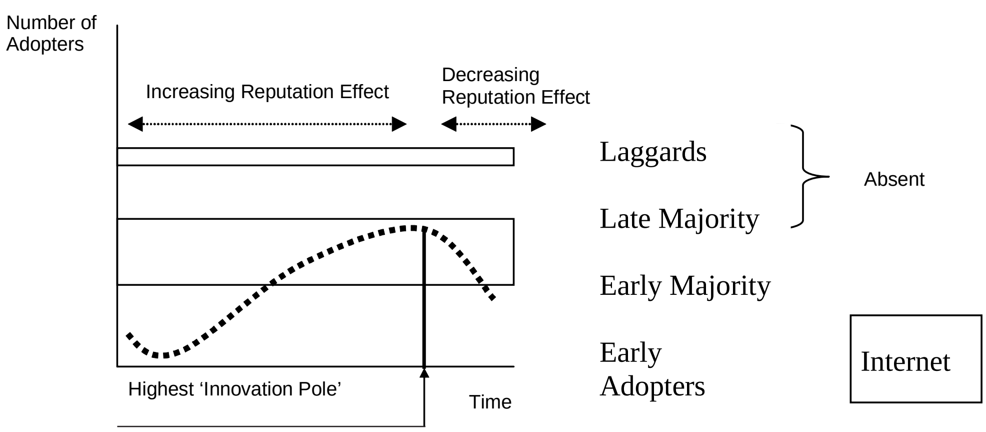


Figure 4. Dipping Diffusion Pattern



If so, the reduction in the reputation effect of the Internet and the low levels of benefits accruing to firms currently may well create a long hiatus before the 'late majority' of adopters arrive. This long hiatus represents a period where policy makers and development agencies will struggle to improve the strategic nature of current Internet adoptions within SMEs.

This could be termed a *'Delayed Reputation Effect'* and will give rise to an *'Elongated Diffusion Pattern'*. Furthermore, if we take into consideration the extremely high connectivity rates that currently exist at this stage of the diffusion process, this makes the potential numbers for a late majority much smaller than that of the early majority.

This fact, combined with the fact that a declining reputation effect will further reduce the number of later adopters, suggests that it may be more accurate to apply the term *'Late Minority'* to any further adopters of Internet technologies. This possible diffusion pattern is presented in Figure 3 below.

Another conclusion that can be reached from these findings is that we may be about to experience what can best be termed a *'Dipping Diffusion Model'* (shown in Figure 4).

If policy makers and development agencies fail to improve the strategic nature of Internet adoptions amongst SMEs, then the benefits of Web site ownership actually experienced will remain low amongst adopters. If this remains the case, it is likely that firms will become disillusioned with the entire cluster of Internet technologies. This will result in an increase in *'Internet Disconnections'*.

In such a pattern, the late majority of adopters will never arrive and disconnections from the Internet actually begin to result in a reduction in overall numbers of adopters over time.

It is important to note that Figures 3 and 4 are indicative models only and are an attempt, based on the analysis of these research findings, to identify the most likely Internet diffusion pattern that will be experienced amongst SMEs in Ireland in the future. Much will depend on the behaviour of policy makers and development agencies, and their ability to repair the damaged reputation of the Internet as a useful facilitation tool for small businesses.

When considering the advantages that this research gives us in understanding future diffusion patterns, a few points need to be clarified. The reality of which of the two potential Internet diffusion patterns discussed will actually be experienced by SMEs in Ireland can only be tested over time. As such, these findings should be viewed as creating the basis for further extensive longitudinal research in this area. Furthermore, whilst it is a necessarily accepted limitation of this research that only SMEs in Ireland have been studied, the conceptual model for this research is drawn from a wide range of established, international IS/IT literature, and in this way is suitable for application to a wider, international sample. If this approach were to be applied by other researchers in this way, more generalisations and conclusions could be made for the global audience.

CONCLUSION

A significant feature of this article is the presentation of the concept of the Internet as a cluster of technologies. The absence of this approach to date has led to much confusion in the interpretation of empirical findings. The application of the concept of an *'Internet Technology Cluster'* in further research, where technologies such as e-mail, browser, and more advanced uses are examined individually, should ensure greater accuracy in the reporting and analysis of Internet adoption and usage activity in the future.

With this issue in mind, the presentation of the Internet usage findings of this empirical research amongst Irish SMEs provides a strong basis for comparative studies in this field.

Whilst uptake is high in terms of connectivity, the nature of usage is not advanced. It is this *low-end-only* use of Internet technologies in the SME sector that has limited the benefits being derived by these firms. Consequently, firms are becoming disenchanted with their Internet experience, leading to a potentially high and damaging 'disconnection rate', and a tendency for later adopters to postpone their adoption process.

This finding has obvious policy and infrastructural implications for the development of Internet usage in Ireland. SMEs in all sectors of the economy must have access to the *sector-specific* training and support necessary for them to climb the learning curve involved in moving from basic Internet usage to the use of more advanced Internet technologies.

This research highlights the substantial training needs, telecommunications access needs, and financial supports that SMEs have regarding advanced Internet usage. If these supports are not made readily available to SMEs, usage will remain basic and benefits limited, leading to the types of problematic diffusion models predicted above.

REFERENCES

Abrahamson, E. & Rosenkopf, L. (1993). Institutional and competitive bandwagons: Using mathematical modeling as a tool to explore innovation diffusion. *Academy of Management Review*, *18*, 487-517.

Amin, A. (1994). *Post-Fordism, a reader*. Oxford: Blackwell.

Bamfield, J. (1994). Technology management learning: The adoption of EDI by retailers. *International Journal of Retail and Distribution Management*, *22*(2), 3-11.

Benjamin, R.I., Rockhart, J.F., Scott Morton, M.C. & Wyman, J. (1983). Information technology: A strategic opportunity. *Sloan Management Review*, (Winter).

Cash, J.I. (1985). Interorganizational systems: An information society opportunity or threat? *The Information Society*, *3*(3), 199-228.

Chan, C. & Swatman, P.M.C. (2000). From EDI to Internet commerce: The BHP steel experience. *Internet Research*, *10*(1), 72-82.

Cooper, R.B. & Zmud, R.W. (1990). Information technology implementation research: A technological diffusion approach. *Management Science*, *36*(2), 123-139.

DiMaggio, P. & Powell, W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*, 147-160.

Elliot, S. & Loebbecke, C. (2000). Interactive, inter-organizational innovations in electronic commerce. *Information Technology and People*, *13*(1), 46-66.

Everett, J. (1998). Internet security. *Employee Benefits Journal*, (September), 14-18.

Feher, A. & Towell, E. (1997). Business use of the Internet. *Internet Research*, *7*(3), 195-200.

Friedman, L. (1998). Technology acquisition decision making revisited: Lessons learned in an age of environmental uncertainty. *International Journal of Technology Management*, *15*(5), 222-237.

Howe, W. (n.d.). *A brief history of the Net*. Retrieved November 16, 2000, from *www.walthowe.com/navnet/history.html*.

Kautz, K. & Larsen, E.A. (2000). Diffusion theory and practice: Disseminating quality management and software process improvement innovations. *Information Technology and People*, *13*(1), 11-26.

La Rovere, R.L. (1998). Diffusion of information technologies and changes in the telecommunications sector: The case of Brazilian small- and medium-sized enterprises. *Information Technology and People*, *11*(3), 194-206.

O'Neill, M.H., Pouder, R.W. & Buchholtz, A.K. (1998). Patterns in the diffusion of strategies across organizations: Insights from the innovation diffusion literature. *The Academy of Management Review*, (January), 17-24.

Pierce, J.L. & Delbecq, A.L. (1977). Organizational structure, individual attitudes and innovation. *Academy of Management Journal*, *2*(1), 27-37.

Raymond, I., Julien P-A., Carrier, J-B. & Lachance, R. (1996). Managing technological change in manufacturing SMEs: A multiple case analysis. *International Journal of Technology Management*, *11*(3), 270.

Swatman, P.M.C. & Swatman, P.A. (1992). EDI system integration: A definition and literature survey. *The Information Society*, *8*, 169-205.

Thorelli, H.B. (1986). Networks: Between markets and hierarchies. *Strategic Management Journal*, *7*(1), 32-45.

KEY TERMS

Advanced Internet Applications: The set of more complex Internet technologies based around design activities, back-end data management activities, and remote access to Internet services.

Dipping Diffusion Pattern: A pattern of IT diffusion in which the late majority/minority of adopters never arrives and disconnections from the Internet result in a reduction in overall numbers of adopters over time.

Elongated Diffusion Pattern: A pattern of IT diffusion in which, due to a reduction in the reputation effect, there is a hiatus prior to the arrival of a late minority of adopters.

Internet Browser Usages: Internet technologies and applications that centre on the Internet browser and that are generally involved in the presentation of Web sites to the viewer, in the undertaking of research and in open-forum communications.

Internet Disconnections: The event of firms becoming disenchanted with Internet applications and removing all Internet-based applications from their organisation.

Internet Technology Cluster: The range of identifiably separate families of technologies involved in the e-business platform comprising: e-mail, browser usages, and advanced Internet applications.

Mimetic Adoption Pattern: A pattern of IT adoption amongst a population of firms in which further adoptions are a function of previous adoptions rather than strategic considerations.

Reputation Effect: The extent of association of a specific information technology with perceived business benefits amongst a population of firms over a period of time, fuelling increased visibility and sustaining a mimetic diffusion pattern.

Digital Asset Management Concepts

Ramesh Subramanian
Quinnipiac University, USA

INTRODUCTION TO DIGITAL ASSET MANAGEMENT

"DAM. Looks like something you might say if you couldn't find a photograph you needed for a front-page story. But DAM—digital asset management—is actually designed to preempt such frustrated outbursts. In an age when oodles of media, including print, images, video and audio, are stored in computers rather than file cabinets, newspapers and other groups need a way to organize, manipulate and share those media quickly and easily." (Grimes, 1998)

Dramatic changes have occurred on the corporate front in the last few years, as more and more businesses have started to conduct commerce on the Internet. New business concepts and products are being developed on a daily basis. The accent is on speed, and changes occur quickly – daily, hourly or even minute-to-minute. Two major facets of these changes are:

1. Large amounts of data are created and stored in digitized forms in organizations, and
2. New "digital products" are created.

As more and more information is created in electronic form, organizations are faced with the following problems:

- The volume of digital data has become cumbersome to manage and reuse (Sharples, 1999).
- Organizations have struggled to reduce cycle time, maintain brand consistency, and coordinate cross-media publishing as well as one-to-one marketing efforts.
- The number of digital assets that an organization may manage has exploded.
- Gistics, a California-based research firm that has studied media asset management for several years, estimates that approximately 30% of all media assets in organizations are misplaced, and then reworked or duplicated.

A 2001 Frost and Sullivan market indicator report by Subha Vivek forecasts tremendous future growth in the U.S. digital media management market (Vivek, 2001). The three market segments that will be affected represent the capture, storage and access, and distribution of digital media, respectively.

The promise of digital asset management has attracted a lot of commercial enterprises and software research laboratories, and several products have been introduced commercially in the last few years. However, due to the "newness" of the field, there is not much academic research literature in the field. A good source of academic thought in this field can be found in the online proceedings of the Annenberg DAM Conference, held at the Annenberg School of Communication, University of Southern California in 1998 (Annenberg DAM Conference, 1998).

BACKGROUND: DIGITAL ASSET MANAGEMENT (DAM) CONCEPTS

This section is adapted from our earlier paper on the subject (Subramanian & Yen, 2002).

A. Definition

A *digital asset* is any asset that exists in a digitized form, and is of intrinsic or commercial value to an organization. *Digital asset management* can be defined as a set of processes that facilitate the search, retrieval, and storage of digital assets from an archive.

B. Basic Features of DAM

The basic features of any DAM system include: storage, search and retrieval, and "thumbnail browsing" (Rosenblatt, 1998). A good DAM system will also include the ability to perform object check-in and check-out.

Other desirable features include:

- Integration of the DAM system with content creation applications on the desktop.
- Enterprise features, that is, features that are necessary for a digital media management system to be useful in a large-scale deployment at a large media company (i.e., an industrial strength, scalable database).
- The ability of a DAM system to have a user interface that can function in a cross-platform environment

(e.g., the Java language from Sun Microsystems, and the development of XML technology).
- The ability to extend the functionality of the DAM system through programming interfaces.

C. What are the basic differences between DAMs and standard data management systems?

One might argue that DAMs are not much different from currently available database management systems that facilitate the search, retrieval, and storage of data. However, DAMs are different in their potential to address four key problems that pertain to the creation, storage, search and dissemination of multi-media data. According to Hilton (Hilton, 2003), those four issues are:

1. Asset mining: New and sophisticated methods for mining multidimensional, multi-media data stores, which can result in the creation of dynamic, "on-demand" digital products
2. Automation: Automated classification and retrieval systems
3. Managing intellectual property (and associated security issues)
4. Engagement: New GUIs and other data manipulation methods as well as collaboration tools

Somani, Choi and Kleewein distinguish traditional data management systems from "content management systems" that handle digital assets, communications and content such as documents, intellectual property, rich media, e-mail and Web data, and discuss the differences between the two types of systems in the following areas (Somani et al., 2002):

1. Data federation to provide in-place access to existing data
2. An expressive data model that accommodates data from very disparate sources
3. Search over metadata and data

Somani et al. then propose an architecture for integrating the two, that is, data and content management systems. A detailed discussion of the architecture is beyond the scope of this article.

DAM SYSTEM ARCHITECTURE

Figure 1 uses a three-tiered architecture of the generic DAM system architecture to show the process followed during an asset creator's session and a client's query session.

In the *asset creation flow*, the Asset Creator creates an asset, which could be in any digital format, and provides the asset and its associated information to the Asset manager. The Asset manager converts the information associated with the asset into an XML metadata format, builds the appropriate data type definitions, and passes the information and the asset to the Metadata manager. The Metadata manager manages the organization of the Metadata Store, which is a database containing meta information on the assets. Appropriate links between the metadata and the actual assets are created and maintained here. The Metadata Store contains information about the digital assets, typically in an XML DTD format. The assets are passed by the Metadata manager to the File manager, which is responsible for the check-in and check-out of the assets into and out of the Digital Asset Store, which is a combination of file systems and databases.

In the *query processing flow*, the Client asks a query. The Query processor parses the query and sends the user information as well as the parsed query to the Metadata manager, which maintains the metadata for the assets. The metadata include not only information about the asset but also information on who is allowed to access the asset. After this information is retrieved from the Metadata store, a message is sent back to the Query processor by the Metadata manager. The message passed may either be a refusal to access the asset, or an affirmation that the requested asset is being retrieved.

The metadata component acts as a store, search, retrieve and security tool, managed by the Metadata manager.

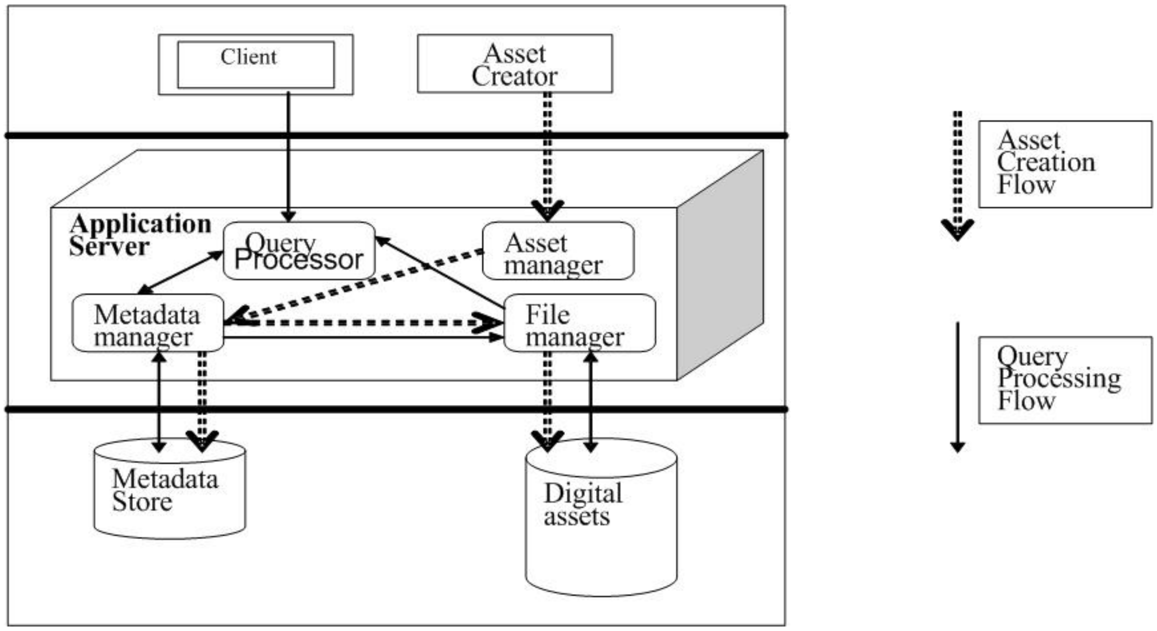
ADDITIONAL ARCHITECTURAL DETAILS

A. An Open and Distributed Architecture

The key to any digital asset management system is to create an open and distributed architecture. A well designed DAM system first should provide the ability for people to take an asset, repository or archive and be able to customize it into their environment and extend it to their existing system and other new systems. The architecture should allow for the following features:

1. Scaling
2. User Interface and Custom Interface

Figure 1. Architecture for DAM system



3. File Management and Asset Association
4. Platform Independence and Transportability

B. Representation and Identification of Digital Assets

1. Representation Issues and Addressable Unit: The three categories of digital asset representation issues are: Production Workflow, Creative Workflow, and the Addressable Unit (Romer, 1998). Production workflow is the ability for metadata to be added throughout the life cycle of digital asset handling (captioning, or cataloging), with appropriate knowledge management capabilities to establish standardization and consistency. Production workflows often deal with known items. Creative workflows are more discovery-oriented, hence more volatile and browse intensive. This workflow is characterized by the need to do many interactive searches and temporarily store candidate assets until a final decision can be made. A good DAM system must allow the designer or user to leverage this valuable production knowledge to help search for what one is looking for, by defining appropriate addressable units within a particular digital asset such as a digital video movie.
2. Identification Issues and Metadata Creation: Organizations are becoming increasingly aware that library science techniques need to be a very integral part of the process of identifying digital assets. DAM systems need to have a very sophisticated metadata model that will store and classify property information, details, names and descriptions, and anything else that the user defines and can be put on top of his/her digital assets. One promising method for classifying digital assets is through the use of "controlled vocabulary". This technique addresses how the hierarchical key words work, so that one can take a system, like the key word structure from the Library of Congress, and apply it to assets inside one's database.

C. Searching Digital Assets

1. "Top Down" Approach: Usually provides specialized searching routines that recognize the properties of the objects being searched. It is also referred to as "Content-based analysis". Image understanding and pattern recognition are all technologies that automatically process images based upon certain inherent properties of the image itself. Some examples are the color, texture and shape recognition systems from Virage (*http://www.virage.com*) or IBM's QBIC application (Faloutsos et al., 1994; Flickner et al., 1995). "Search extenders" facilitate content-based searching of digital files (e.g., content base searching of textural images). Bulldog's Search Extender™ technology allows the search mechanism to recognize the properties of various file types and implement specialized searching routines. They allow users to do context searching,

such as finding the name "Bulldog" very close to information about Sun Microsystems inside the database. Equivalent search extender technology exists for images, video and audio.
2. "Bottom Up" Approach: A method is to design image representation schemes that will match the goals of the search process used for retrieval. The search sub-system was designed with decision support in mind.

TAXONOMY OF DAM SYSTEMS

We have categorized the DAM systems based on two criteria. The first criterion is based on how many people DAM systems need to serve, the price range and their complexity level. In this criterion, DAM can be divided into three categories (Rosenblatt, 1998):

1. The Enterprise category, to give an organization the capabilities to be scalable and with industrial strength. It typically costs in the millions of dollars.
2. The Mid-range category, where the system is purpose-built for a specific asset management purpose. They typically cost in the one hundred to five hundred thousand dollar range.
3. The Low-end category, for basic, low-end solutions that are useful for small installations. These systems cost typically cost around $50K. These are systems that just have basic features, and they run on work group class servers.

The second criterion we use is based on several representative functions of the DAM.

1. The Image Cataloging Feature: With image cataloging feature, systems capture low-resolution thumbnails.
2. The Object Repository Feature: With object repository feature, systems can capture not only thumbnails and metadata, but also high-resolution media files.
3. Content Management Back End for Web Sites Feature: This provides service/editorial environment with content management for the Web. The content can be in different file formats, such as text, graphics, multimedia, XML, and SGML documents.
4. Intelligent Cataloguing and Indexing Feature: This intelligent cataloguing and indexing feature has Web-based interface with intelligent navigation tools, allowing the user to search and organize images, video, audio and text documents easier and more efficiently.

PROTECTION AND DISTRIBUTION ISSUES OF DIGITAL ASSETS

Digital Assets Protection

The rapid growth of multimedia manipulation tools and the wide availability of network access lead to the convenience of digital data processing, delivery and storage. Powerful digital facilities can produce a large amount of perfect digital asset copies in a short period of time. The advanced compression technology also contributes to make the digital content compacted into a small data stream to facilitate transmission. These advantages benefit content users, definitely to raise concerns from content creators and content owners if the intellectual property right cannot be enforced successfully (Su et al., 1999). This is especially true in the present time, which has seen the rapid proliferation of peer-to-peer online file sharing systems such as Napster and Kazaa (Subramanian & Goodman, 2003). Yet, it has been noted by several researchers that senior corporate executives rarely pay a whole lot of attention to computer security and the security of digital assets. Austin and Darby discuss the types of threats that a company is apt to face and propose eight processes, ranging from deciding how much protection each asset receives, to insisting on secure software, to rehearsing a response to a security breach (Austin & Darby, 2003). We discuss the subject of digital assets protection from three different viewpoints, that is, legal, technical and non-technical means to ensure protection of digital assets:

1. Legal Methods: Patents, Copyrights, and Trademarks: A patent is a grant of exclusive right that is given by the government to anybody who files an application for patent protection. It confers upon the applicant the right to exclude everyone else from making, using or selling that patented invention. The term of protection offered by a patent is shorter than that of a copyright by a good deal. The term of protection for a patent is generally 20 years from filing throughout the world. Under the U.S. patent system, that protection is available for any invention. Anything that can be conceived of by humans and put into practice in a useful way that is new and novel can be subject to patent protection. This includes any creation of a human's mind, such as drugs, pharmaceuticals, computer software, inventions, processes, methods of manufacture, chemical compounds, electronic devices, and so forth.
A copyright protects an original work of authorship. The term of protection for a copyright under

the U.S. law is "life of the author and fifty years for the case of a work made in the course of employment by an employee working for an employer". Under certain conditions, the protection period for the latter case could be 75 years from publication or 100 years from creation, whichever expires first. There is a proposal currently in the U.S. Congress to extend the term of copyright protection by an additional 20 years, to bring it up to the level that is guaranteed in the European Union.

A trademark is a right that attaches to a marking used to identify goods and commerce. It helps serve as a guarantee of the source of those goods for consumers, so you know where they come from and you may have expectation as to the quality associated with that mark. Trademarks may be registered with the USPTO; they may also be registered with state authorities. Many states have their own trademark registries. Trademarks can also arise by the operation of common wall.

Copyright laws can be separated into the Anglo-American and the European copyright laws. The Anglo-American laws are followed by America, Great Britain and other British colonies. They adopt a pragmatic, market-oriented approach towards copyrights. On the other hand, under European laws, the copyright is seen as a basic fundamental human right. The economic incentives that it provides are secondary to insuring the protection of someone's personality.

2. Technical Methods for Intellectual Property Protection: Encryption and watermarking are the two most important digital assets content protection techniques. Encryption protects the content from anyone without the proper decryption key. There are two types of encryption existing: symmetric (secret-key) mechanism and asymmetric (public-key) mechanism. Symmetric mechanism uses the same security key to "lock" and scramble a digital file and to recover a bit-exact copy of the original content at the destination. Asymmetric encryption employs dual keys. The sender encrypts the digital with the recipient's public key, and the recipient decrypts it with his or her private key (Cravotta, 1999). Watermarking, on the other hand, further protects the content by embedding an imperceptible signal directly into the content for ownership declaration, play control or authentication (Su et al., 1999).
3. Non-technical methods for intellectual property protection: "This has to do with our providing products at a value above and beyond that provided by those who would pirate our products" (Griffin, 1998).

Digital Assets Distribution

In order to manage the asset owner and the user, the owner of a digital asset must be able to determine the usage scenarios that are appropriate for the asset. The scenarios are summarized by the notion of placement, size, duration, and extent. License management refers to the legal terms and conditions that apply to the specific usage scenario.

CONCLUSION AND FUTURE DIRECTIONS

In the era of e-commerce, digital asset management is emerging as a very hot topic of study for both practitioners as well as researchers. In this article we discuss different concepts and issues of DAM, which include but are not limited to the components and architecture of DAM systems, its basic and desirable features, the taxonomy of DAM systems, and protection and distribution sub-systems of DAM. There are several open issues for research in DAM. A list of these issues includes but is not limited to modeling the storage and organization of digital assets, the digital assets valuation, pricing, rights and licensing models, and methodologies for optimal search and retrieval of digital assets. In a future study, we plan to address some of these open issues.

REFERENCES

Annenberg DAM Conference. (1998). *http://www.ec2.edu/dccenter/dam/*

Austin, R.D., & Darby, C.A.R. (2003, June). The myth of secure computing. *Harvard Business Review, 81*(6).

Cravotta, N. (1999, March). Encryption: More than just complex algorithms. *EDNMagazine*. *http://www.reed-electronics.com/ednmag/index.asp?layout=article&articleid=CA56697&rid=0&rme=0&cfd=1*

Faloutsos et al. (1994). Efficient and effective querying by image content. *Journal of Intelligent Information Systems, 3*, 231-262.

Flickner et al. (1995, September). Query by image and video content: The QBIC system. *Computer, 28*(9), 23-31.

Griffin, J. (1998). Annenberg DAM Conference 1998. *Transcripts of the session "Auditing & Valuation in Different Industries: Licensing & Delivery to Consumers". http://dd.ec2.edu/1998/dam98_2a_transcript.html*

Grimes, B. (1998, November/December). Digital asset management 101. *TechNews, 4*(6). *http://www.naa.org/technews/tn981112/editorial.html*

Hilton, J.L. (2003, March/April). Digital asset management systems. *EDUCAUSE Review, 3*(2).

Romer, D. (1998). Annenberg DAM Conference 1998. *Transcripts of the session "Asset Representation in the Creative Process: Bringing Media to Life". http://dd.ec2.edu/1998/dam98_2b_transcript.html*

Rosenblatt, W. (1998). Annenberg DAM Conference 1998. *Transcripts of the Session "Storage / Retrieval / Presentation". http://dd.ec2.edu/1998/dam98_1b_transcript.html*

Sharples, H. (1999, November). Sights set on site. *Graphic Arts Monthly*, 52-54.

Somani, A., Choy, D., & Kleewein, J.C. (2002). Bringing together content and data management systems: Challenges and opportunities. *IBM Systems Journal*, *41*(4), 686.

Su, Po-Chyi et al. (1999, April 25-28). Digital image watermarking in regions of interest. *Proceedings of the IS&T Image Processing/Image Quality/Image Capture Systems (PICS)*, Savannah, Georgia. *http://biron.usc.edu/~pochyisu/pochyi_files/main.htm*

Subramanian, R., & Goodman, B. (2003). Peer-to-peer corporate resource sharing and distribution with mesh. *Proceedings of the 14th IRMA International Conference.*

Subramanian, R., & Yen, M. (2002). Digital asset management: Concepts and issues. In A. Gangopadhyay (Ed.), *Managing business with electronic commerce: Issues and trends*. Hershey, PA: Idea Group Publishing.

Vivek, S. (2001). DAM: It's consolidation time. Frost and Sullivan "Market Indicator" Report. *https://www.frost.com/prod/servlet/market-insight.pag?docid=RCOL-4ZEMMP&ctxht= FcmCtx4&ctxhl=FcmCtx5&ctxixpLink=FcmCtx5 &ctxixpLabel= FcmCtx6* (requires account).

KEY TERMS

Addressable Unit: A specific unit within a particular digital asset such as a digital video movie.

Asset Creator: Anyone who creates an asset, which could be in any digital format, and provides the asset and its associated information to the asset manager.

Asset Manager: The asset manager converts the information associated with the asset into an XML metadata format, builds the appropriate data type definitions, and passes the information and the asset to the metadata manager.

Creative Workflow: These are more discovery-oriented, hence more volatile and browse intensive. This workflow is characterized by the need to do many interactive searches and temporarily store candidate assets until a final decision can be made.

Digital Asset: A digital asset is any asset that exists in a digitized form, and is of intrinsic or commercial value to an organization.

Digital Asset Management (DAM): Digital asset management can be defined as a set of processes that facilitate the search, retrieval, and storage of digital assets from an archive.

Digital Asset Store: A combination of file systems and databases.

Metadata Manager: The metadata manager manages the organization of the metadata store, which is a database containing meta information on the assets.

Metadata Store: The metadata store contains information about the digital assets, typically in an XML DTD format.

Production Workflow: It is the ability for metadata to be added throughout the life cycle of digital asset handling (captioning, or cataloging), with appropriate knowledge management capabilities to establish standardization and consistency. Production workflows often deal with known items.

Search Extenders: Methods that facilitate content-based searching of digital files.

Digital Government and Individual Privacy

Patrick R. Mullen
U.S. Government Accountability Office, Washington, DC, USA[1]

INTRODUCTION

Individual privacy is freedom from excessive intrusion by those seeking personal information about the individual. This allows the individual to choose the extent and circumstances under which personal information will be shared with others. A related concept, confidentiality, is a status accorded to information based on a decision, agreement, obligation, or duty. This status requires that the recipient of personal information must control disclosure. While privacy and confidentiality are concepts that can be applied to people in all societies, this article discusses them in relation to their treatment by the United States government, in particular with the advent of digital government. The concepts associated with digital government can also apply to non-Americans as well, but the discussion in this article is based on U.S. laws, documents, and relevant literature.

BACKGROUND

The growth of the Internet and digital government—that is, e-government—has dramatically increased the federal government's ability to collect, analyze, and disclose personal information about many private aspects of people's lives. Such information, once available only on paper to a limited number of people, is now instantly retrievable anywhere in the world by anyone with a computer and an Internet connection, including by hackers and firms specializing in selling information on individuals. At the same time as this dramatic increase in digital government, the level of trust in government has declined (Council for Excellence in Government, 2001); currently, many Americans perceive the government as a potential threat to individual privacy. Given these forces at work in American society, one should not be surprised to read the results of surveys that show privacy as a top concern of Americans in the 21st century. If Americans do not believe that the government is adequately protecting individual privacy, they may be less willing to provide the government with information. For example, most surveys by statistical agencies are voluntary, and even others that are mandatory, such as the decennial census and the American Community Survey, can suffer from underreporting by respondents. Such reluctance could compromise the ability of government to collect information necessary to develop, administer, and evaluate the impact of various policies and programs (Mullen, 2004).

FUTURE TRENDS

Generally speaking, e-government refers to the use of technology, particularly Web-based Internet applications, to enhance a specific agency's Web site, for access to and delivery of government information and services to individuals, businesses, and other organizations and government agencies. E-government has been seen as promising a wide range of benefits based largely on harnessing the power of the Internet to facilitate interconnections and information exchange between citizens and their government. A variety of actions have been taken in recent years to enhance the government's ability to realize the potential of e-government, culminating in the enactment of the E-Government Act of 2002 (Public Law 107-347), which includes provisions addressing everything from funding of e-government initiatives to measures for ensuring security and privacy. In addition to the E-Government Act, President George W. Bush designated e-government as one of five priorities in his management agenda for making the federal government more focused on citizens and results. The goals of President Bush's e-government initiative are summarized in Table 1.

Schelin (2003) discusses the rapid growth of e-government and provides an overview of the historical premises, theoretical constructs, and associated typologies of e-government that are a framework for understanding e-government and its potential benefits and related challenges. While the Internet opens new opportunities for streamlining processes and enhancing delivery of services, federal executives and managers must also be cognizant of the responsibilities and challenges that accompany these opportunities (Garson, 2003; Pavlichev & Garson, 2004; U.S. General Accounting Office, 2001a). Some of the responsibilities and challenges associated with managing e-government are summarized in Table 2.

Table 1. Goals of President Bush's e-government initiative

According to the President's management agenda, e-government is expected to:

- provide high-quality customer services regardless of whether the citizen contacts the agency by phone, in person, or on the Web;
- reduce the expense and difficulty of doing business with the government;
- cut government operating costs;
- provide citizens with readier access to government services;
- increase access for persons with disabilities to agency Web sites and e-government applications; and
- make government more transparent and accountable.

Table 2. Summary of e-government responsibilities and challenges

- Sustaining committed executive leadership
- Building effective e-government business cases
- Maintaining a focus on the individual
- Protecting individual privacy and confidentiality
- Implementing appropriate security controls
- Maintaining electronic records
- Maintaining a robust technical infrastructure
- Addressing IT human capital concerns
- Ensuring uniform service to the public
- Empowering citizens in democratic processes

PERSPECTIVES ON PRIVACY

In American society, there is an inherent tension between the desire for the free flow of information versus the concern for maintaining individual privacy. This tension is captured in various congressional statements, included in legislation as well as in executive branch guidance to agencies, which explain how to carry out their seemingly conflicting responsibilities under the law. For example, in passing the Paperwork Reduction Act of 1995, the Senate Governmental Affairs Committee (1995) noted that "information obtained by government is a valuable and useful resource to government and society, if managed in a coordinated and systematic manner." The committee also noted the importance of the free flow of information:

"The advent of the electronic information age presents new opportunities and obligations of the Federal government as it strives to fulfill its continuing responsibility to make government information accessible to the American public. The legislation meets this need by providing for improved dissemination of government information to the public, particularly in electronic formats."

However, these same technological trends also raise concerns about information privacy. As Congress stated in passing the Privacy Act (1974):

"…the increasing use of computers and sophisticated information technology, while essential to the efficient operations of the Government, has greatly magnified the harm to individual privacy that can occur for any collection, maintenance, use, or dissemination of personal information."

While the Privacy Act is the primary law regulating the federal government's collection and maintenance of personal information, the legislative intent has been clarified with subsequent Office of Management and Budget (OMB) guidance to federal agencies. OMB's circular on "The Management of Federal Information Resources" (Circular A-130) also captures the balance between the free flow of information versus individual privacy (U.S. Office of Management and Budget, 2000), addressing (1) the need for agency Web sites to post clear and easily accessed privacy policies and (2) federal agency use of Internet cookies—short strings of text sent from a Web server to a Web browser when the browser accesses a Web page—which poses privacy risks because the data contained in persistent cookies may be linked to individuals after the fact (U.S. General Accounting Office, 2001b). In addition, the Federal Trade Commission (FTC) has issued four Fair Information Principles governing online privacy at commercial Web sites. These principles can be used as criteria to assess federal agency Web sites (U.S. General Accounting Office, 2000; Center for Democracy and Technology, 1999). The principles are included in Table 3.

PROTECTING INDIVIDUAL PRIVACY

Individual privacy is closely related to data confidentiality and security (Boruch & Cecil, 1979; Duncan, Jabine &

Table 3. FTC's fair information principles

1. Notice. Data collectors must disclose their information practices before collecting personal information from consumers.
2. Choice. Consumers must be given options with respect to whether and how personal information collected from them may be used for purposes beyond those for which the information was provided.
3. Access. Consumers should be able to view and contest the accuracy and completeness of data collected about them.
4. Security. Data collectors must take reasonable steps to ensure that information collected from consumers is accurate and secure from unauthorized use.

de Wolf, 1993; Lowrance, 1997; U.S. General Accounting Office, 2001c). If a breach in data confidentiality were to occur, individual privacy might be affected. And a breach in security might affect both confidentiality and individual privacy. By the same token, protecting confidentiality—through various security measures—helps protect individual privacy. The concepts and practices of good data stewardship are important for federal agencies to understand in their role as protectors of individual privacy. A loss of confidence in federal agencies' ability to protect individual privacy would have serious consequences for individuals' trust and participation in government programs. With respect to e-government issues, two key factors are:

1. **Privacy Status:** Whether or not the information about an individual, including his or her personal attitudes or experiences, is known to another.
2. **Privacy Rights:** Whether the individual has control over whether information is shared with anyone else.

Many definitions of individual privacy emphasize privacy rights. Logically, however, an individual's privacy may be preserved, enhanced, or reduced by the choices that he or she makes, as well as by the actions of others, specifically excessive intrusion.

In ensuring individual privacy, the federal government faces challenges while implementing e-government. For example, record linkage is sometimes seen as reducing privacy because of the additional information provided about the individual (the whole may be greater than the sum of its parts). Conversely, record linkage has sometimes been viewed as enhancing individual privacy because the collection of new data may be avoided. Because agency public use data files and Web sites now make the results of research and statistical programs widely available to the public, there is an increased need to make sure that the confidentiality of personal data is strongly protected and that re-identification risks are guarded against.

The Privacy Act can be viewed as striking a balance between government needs for information and individual privacy interests. To strike this balance, the Privacy Act contains requirements that implement the principle of openness, which is similar to the FTC's fair information principle of notice. This principle dictates that an agency of the federal government must be open about its record-keeping policies, practices, and systems that relate to individual privacy. The requirements implementing the openness principle are intended to achieve two general goals:

1. facilitate public scrutiny of federal agency record-keeping policies, practices, and systems by interested and knowledgeable parties; and
2. make the individual aware of systems in which a record on him or her is likely to exist.

Specifically, the Privacy Act applies these goals only to information maintained in a system of records and allows exceptions, under various circumstances, to the disclosure and use of information without the consent of the individual. The advent of e-government has brought new challenges for federal agencies in complying with the principles of the Privacy Act. In addition, various rules and regulations to implement the Privacy Act have been updated in attempts to make these requirements relevant in the era of the Internet and Web-based applications.

The government cannot realize the full potential of the Internet until individuals are confident that the government will protect their privacy when they visit its Web sites. To ensure that individuals understand how their personal information is handled when they visit federal Web sites, OMB issues circulars and memorandums, in accordance with the openness principle. For example, OMB's M-99-18 memorandum requires federal agencies to post privacy policies on their Internet Web sites, providing guidance for doing so. Specifically, the memorandum requires each agency to post privacy policies to: (1) its principal Web sites; (2) any other known, major entry points to the Web sites; and (3) any Web page where substantial personal information is collected from the public. The memorandum also requires agencies to post privacy policies that clearly and concisely inform visitors to the Web sites what information the agency collects about individuals, why the agency collects it, and how the agency will use it. Finally, the memorandum requires that privacy policies are clearly labeled and easily accessed when an individual visits a Web site.

CONCLUSION

New information technologies offer many possibilities for the government to improve the quality and efficiency of its service to individual Americans. Web sites are powerful tools for conveying information on federal activities, objectives, policies, and programs. Web sites also provide a simple and quick way for accessing information about the government and what it is doing on the individual's behalf. Congressional interest in the benefits of e-government and Internet-based operations has resulted in the passage of laws designed to encourage the deployment of e-government functions, while also protecting individuals' privacy. Increasingly, federal agencies are using the World Wide Web and other Internet-based applications to provide online public access to information and services, as well as to improve internal business operations. These developments have naturally resulted in concerns about protecting the privacy of individuals for whom (1) information is electronically available and (2) access to information on federal agency Web sites is intended. Therefore, for digital government to succeed, it is critical that agency officials and staff, such as privacy officers and advocates, disclosure officials, and survey managers, exercise good stewardship of the data for which they are responsible. They can exercise good stewardship by guarding data subjects' personal privacy, as well as data confidentiality and security, through use of cybersecurity practices, such as public key infrastructure (PKI) technology.

REFERENCES

Boruch, R.F. & Cecil, J.S. (1979). *Assuring the confidentiality of social research data*. Philadelphia, PA: University of Pennsylvania Press.

Center for Democracy and Technology. (1999). *Policy vs. practice: A progress report on federal government privacy notices on the World Wide Web*.

Council for Excellence in Government. (2001). *The next American revolution* (Hart-Teeter poll reported in e-government).

Duncan, G.T., Jabine, T.B. & de Wolf, V.A. (1993). *Private lives and public policies: Confidentiality and accessibility of government statistics*. Washington, DC: National Academy Press.

Garson, G.D. (2003). *Public information technology: Policy and management issues*. Hershey, PA: Idea Group Publishing.

Lowrance, W.W. (1997). *Privacy and health research: A report to the U.S. Secretary of Health and Human Services*. Washington, DC.

Mullen, P.R. (2004). Digital government and individual privacy. In A. Pavlichev & G.D. Garson (Eds.), *Digital government: Principles and best practices*. Hershey, PA: Idea Group Publishing.

Pavlichev, A. & Garson, G.D. (2004). *Digital government: Principles and best practices*. Hershey, PA: Idea Group Publishing.

Schelin, S.H. (2003). E-government: An overview. In G.D. Garson (Ed.), *Public information technology: Policy and management issues*. Hershey, PA: Idea Group Publishing.

U.S. Congress. (1974). *The Privacy Act of 1974*. Public Law 93-579.

U.S. Congress. (2002). *E-Government Act of 2002*. Public Law 107-347.

U.S. General Accounting Office. (2000, October 11). *Internet privacy: Comparison of federal agency practices with FTC's Fair Information Principles* (GAO-01-113T).

U.S. General Accounting Office. (2001a, July 11). *Electronic government: Challenges must be addressed with effective leadership and management* (GAO-01-959T).

U.S. General Accounting Office. (2001b, April). *Internet privacy: Implementation of federal guidance for agency use of "cookies"* (GAO-01-424).

U.S. General Accounting Office. (2001c, April). *Record linkage and privacy: Issues in creating new federal research and statistical information* (GAO-01-126SP).

U.S. Office of Management and Budget. (1999, June). *Privacy policies on federal Web sites* (M-99-18).

U.S. Office of Management and Budget. (2000, November). *Management of federal information resources* (Circular No. A-130).

U.S. Senate. (1995). *Senate Report 104-8 on the Paperwork Reduction Act of 1995*. Public Law 104-13.

KEY TERMS

Confidentiality: A status accorded to information based on a decision, agreement, obligation, or duty. This status requires that the recipient of personal information must control disclosure.

Cookies: A short string of text that is sent from a Web server to a Web browser when the browser accesses a Web page. The information stored in a cookie includes the cookie name, the unique identification number, the expiration date, and the domain.

Individual Privacy: Freedom from excessive intrusion by those seeking personal information about the individual. Allows for the individual to choose the extent and circumstances under which personal information will be shared with others.

Public Key Infrastructure (PKI): A system of computers, software, and data that relies on certain sophisticated cryptographic techniques to secure online messages or transactions.

Public Use Data Files: Electronic computer files containing the results of federal research and statistical programs that are made available for public use, often on agency Web pages.

Record Linkage: A computer-based process that combines sources of existing personal information with additional information that refers to the same person, or to his or her family, friends, schools, employer, area of residence, or geographic environment.

Re-Identification: The risk that publicly released data could allow public identification of a survey respondent's information or could be used in a way harmful to them. Federal agency data releases must comply with statistical disclosure limitation methods to guard against this risk.

Stewardship: Management of another's property, affairs, or, in this case, data. For agencies, stewardship includes functions of officials and staff, such as privacy officers and advocates, disclosure officials, and survey managers. Stewardship carries responsibility for data subjects' personal privacy, as well as data confidentiality and security.

ENDNOTE

1 The opinions in this article are solely the author's and do not represent those of the U.S. Government Accountability.

Digital Literacy and the Position of the End-User

Steven Utsi
K.U.Leuven, Belgium

Joost Lowyck
K.U. Leuven, Belgium

INTRODUCTION

As an educational setting, the traditional classroom fails to meet the learner's need for suitable skills to learn with educational software. The development of digital learning skills in school curricula challenges designers of educational software. A useful starting point of research in this domain is the study of literacy, both in its traditional and new forms (Tyner, 1998). It is a powerful background for research on the interaction of learners with educational software platforms. A "platfom" is a particular software package, designed for educational use.

BACKGROUND

Both in school and society, the skill to comprehend and handle printed course materials is essential. Literacy has since long been a vital skill for functioning adequately in an industrial society (see e.g. Marvin, 1984).

An Emerging Plural Notion of Literacy

The International Adult Literacy Survey (IALS) describes literacy as a broad range of information processing skills in relation to written or printed language. Traditional literacy is defined as follows (OECD, 1997, p. 2):

"Using printed and written information to function in society,to achieve one's goal and to develop one's knowledge and potential."

However, traditional literacy is increasingly evolving into a new, plural literacy that refers to making sense of meaningful content in more complex and technological environments (Erstad, 1998). The growing importance of images and of communication technologies has a cultural backlash that even transforms the nature of literacy. Gee (1990) opened up so-called "New Literacy Studies" (NLS). He defends a socio-cultural approach of literacy (p. 153):

"Literacy is the mastery of, or fluent control over, secondary discourse."

While primary discourse pertains to infant face-to-face interaction of children with trusted figures (parents, family, and others), secondary discourse develops through contact with public life and its social and cultural conventions. Secondary literacy is in itself a plural concept: a multitude of social institutions and commitments to public life invade an adult's life and are as many "literacies" to master. As Walter (1999, p. 34) points out:

"The existence of multiple literacies, none more valid than the next, but each specific to a culturally-defined community."

According to this plural notion of literacy, literacy can be neither neutral nor universal, since all literacy includes social and cultural conventions that shape a particular type of "literacy". Visual literacy, for instance, complements traditional literacy and claims a unique position in today's school curriculum. Debes (1969) first mentioned "visual literacy". According to visual literacy, a specific "image" language supports communication. In traditional language, words support verbal communication. Visual literacy may not only be a means of communication, but also a way of thinking (Hortin, 1983). Thinking visually, then, means the ability to think and learn in terms of images. And children's acquisition of skills to work effectively and efficiently with educational software has to underpin this recent position of a new and full interpretation of literacy.

Undoubtedly, it is of prime importance to analyse the nature of skills necessary to take full advantage of today's learning opportunities. In a visual oriented culture the acquisition of new reading and writing skills is indispensable, e.g. the analysis and composition of images. Indeed, literacy supposes an active intervention in a social and cultural context. Avgernou and Ericson (1997) define visual literacy as a group of skills that make it possible for an individual to understand and use visuals for intentional communication with others. This concerns different target groups, for instance primary school pupils or even impaired children.

During the last decade, a wide array of "literacies" relating to information and communication technologies (ICT) surfaced: media literacy (Hobbs, 1998; Potter, 1998), electronic literacy (Maylath, 1993), multimedia literacy (Kellner, 1998), computer literacy (Guthrie & Richardson, 1995; Peha, 1995), and digital literacy (Gilster, 1997). This evolution accompanies the expansion of IT to ICT. Indeed, communication is now a central feature of technological environments, clearly depending on both "traditional" and "new" literacies (Plowman & Stephen, 2003):

"(…) the flexible and sustainable mastery of a repertoire of practices with the texts of traditional and new communication technologies via spoken language, print and multimedia."

The overarching notion "information literacy" denotes the ability to access, retrieve, manage, and use information relevant to an identified need for information (Kuhltau, cit. in Campbell, 1994). Originally, information literacy was limited in scope to computer information. The progress of computer sciences and, more generally, the use of ICT in a wide array of domains broadened its meaning into library skills, computer skills, thinking skills, and critical reading skills.

Media literacy pertains to communication through and critical analysis of a diversity of media; it is the end user's ability to navigate both effectively and efficiently and to keep track of position in electronic media, while "criss-crossing the landscape" (Spiro, R. J., Feltovich, R. L., Jacobson, M. J., & Coulson, R. L., 1991). Gilster (1997, p. 1) defines digital literacy as follows:

"(…) the ability to understand and use information in multiple format from a wide range of sources when it is presented via computers."

Computer literacy is the ability to integrate information and build a personal knowledge base. Both electronic literacy (e-mail reading skills) and multimedia literacy (technical multimedia skills) are building blocks of more general "computer" literacy. Electronic and multimedia literacy explain, for instance, the comprehension of hypertext.

When comparing different "literacies", two observations are important. First, critical analysis, interpretation, and processing of information are attributed to media literacy and digital literacy. The processing and integration of information (computer literacy) and technical skills (electronic and multimedia literacy) have to be critically evaluated by computer users. Secondly, without the notion of traditional and visual literacy, none of the newer forms of literacy can be understood. Indeed, media and digital literacy acquire meaning for users through similar basic mechanisms as traditional and visual literacy. Literacy education elucidates implicit messages, ideological content or even idiosyncratic intentions designers may embed in software packages. On the other hand, the study of ICT related literacies informs software designers of problems encountered by learners with educational software platforms. Traditional issues are accessibility of information and user interface design.

Current Research Questions

The "literacies" debate is a theoretical starting point. Empirically, the detection of specific skills that explain interaction with educational software -digital literacy- is a first research path. These skills have to be integrated in the school curriculum and are treated as abilities underlying new "literacies". Before any application of theoretical insights, a primary research question has to pertain to the relationship between "operational skills" (searching, clicking, and/or dragging screen and user interface objects) and content comprehension in educational software. Is retrieval of information influenced by the mastery of operational skills?

Moreover, information can be represented through text, visualization, or talk. Does the integration of these different symbol systems in educational software alter the typical linear end-user interaction with the computer screen interface? The most common pattern of software use is sequencing interface screen after interface screen in a so-called linear fashion. Clicking hotspots and exploring additional in-depth layers of screens, providing e.g. background information, are seldom spontaneous actions. This type of research question addresses conditions that facilitate "switching content" in -for instance- an educational software package fitted out with hotspots and hyperlinks. The content of an educational platform can for example be organized in an adventure game with hyperlinks, combined with an illustrated encyclopaedia supporting the game with textual and verbal background information. A related question points to the relationship between switching content and retrieving or remembering information afterwards. Is switching detrimental to retrieval of information or does it on the contrary support memory?

Research with 3rd and 4th graders using a multimedia comic strip about World War II (see Utsi & Lowyck, 2002) revealed end-users to anticipate crucial events: they look for objects in the interface screens that most probably will play a crucial role in the next few screens. Mere reactions to audio-visual events in interface screens steadily fade, while searching, clicking, and/or dragging objects become increasingly well-considered throughout the user-interface interaction. Throughout the process, visual literacy gradually changes from superficial use of visual cues to deeper comprehension of educational content. Thus, visual literacy is an essential condition for meeting the educational goals. Multimedia literacy skills are effortlessly acquired on the spot: clicking and dragging objects pose no problem. When first confronted with a new educational software platform, visual literacy seems narrowed to multimedia skills, like clicking and dragging, but gradually visual literacy opens up again to more thoughtful, content driven interaction with all the platform's trimmings.

FUTURE TRENDS

Future studies in this research field can make use of the cognitive load model of Kalyuga, Chandler, and Sweller (2000), who suggest a trade-off effect between text and visual cues. Cognitive load can be defined as the amount of mental resources necessary for information processing. High cognitive load requires the end-user to spend extra memory resources to deal with new incoming information. Accordingly, text and visual cues compete for attention of users since the cognitive span of learners is limited, while effort needed for both task completion -clicking one's way through the platform's interface screens- and content comprehension requires maximum investment of cognitive resources and attention (see also Kirschner, 2002). The trade-off of text and visual cues (Utsi & Lowyck, 2002) yields a clear distinction between low performing pupils clinging to visual cues and the ones who use the full range of novel experiences and skills offered in the educational platform.

Processing and integrating information from different sources is a digital and "computer" (Guthrie & Richardson, 1995; Peha, 1995) literacy skill that may acquire its place in school curricula in due time. Recent and future studies can provide input for implementing research findings with regard to digital literacy in educational software. Like any other type of ICT, educational multimedia runs through a cycle of design, development, and implementation. The shift, in the strip story mentioned above, from the user's shallow reactions to visual stimuli toward anticipation of events is a leading theme. This anticipation can be triggered by visual cues or by story line comprehension. Pupils seem to encounter difficulties to cope with the different layers of multimedia information: text, pictures, motion, music, and sound. They hardly break up the linear nature of the "storyline": background information is scarcely accessed. A core concern while designing, developing, and implementing educational media is the trade-off of textual and visual information in educational multimedia, hampering the break up of linear content material. Textual and visual symbol systems compete for the attention of learners, involved in learning with educational software.

Designers of educational software need to try and ease the integration of text and visuals. Throughout the realization of educational software, care should be taken that the precise role of hyperlinks and hotspots is clear to the user. Indeed, available cognitive span is highly occupied by an educational platform using different symbol systems, such as textual and visual information. The end user's decision to access an additional content layer in the platform needs to be an informed decision. If the learner does not know where a hyperlink or a hotspot will lead to, he may neglect to select the hyperlink, because he is not aware of the relevance of this information for the task at hand. The type and relevance of information hyperlinks and hotspots can be realized through embedded, formal features (for instance glowing or blinking) that signal the information value of a hyperlink or hotspot.

Furthermore, educational software platforms presenting information and exercises in a rousing setting, for instance in an adventure game context, need to ensure that a balanced arch of task tension is built up throughout the platform. Task tension is the "track" of the learner to reach intermediary goals that eventually lead to a successful mastery of an assignment. Regularly, the learner has to be attended to the end goal, for example retrieving a tool for repairing a space ship. This tool has to turn up visually from time to time or nearing it may be signalled by a characteristic sound. Ideally, this arch of task tension implies a series of pauses: moments for the learner to relax and explore course content from another point of view, for instance through the access of background information. Ultimately, tension is discharged in a culminating accomplishment of the assignment or end goal. Pauses can be created through opening up rigid linear progress from one content element or exercise to another and/or and through filling in the "gaps" with interesting, but distracting information or modest tasks. In an educational adventure game, for instance a journey through outer space, a simple, but distracting lottery game can be integrated, without any relevance for the end goal of the adventure game: retrieving a tool for repairing a space ship. Educational material that is presented in a serial, linear fashion without consideration of the limited processing capacities of young learners is doomed to be incompletely and thus inadequately processed.

Information sticks better when the end user's attention is triggered by repeating screen elements that suggest an optimal route across specific content elements offered at a learner's initial level. The design, development, and implementation of educational ICT need to balance the signalling function of visual cues, routing pupils' clicking behaviour, and the effort of learners to look for content, relevant for reaching goals. In educational software, this can for instance be achieved via a "road map", lending learners a hand in orienting themselves on the way to their goal. Small screen elements cue the learner's attention and break up the linear character of traditional course material, if presented in a low task tension context.

CONCLUSION

Digital literacy is a baseline set of skills for successfully coping with a complex, often technological world, holding multiple media messages. In line with traditional reading and writing training, digital literate learners may also cope successfully with new ways of communicating. Digital literacy comprises technical computer skills (searching, clicking, and/or dragging) and visual cue awareness that triggers reading and deepened understanding of information. Support of the user-platform interaction needs to be embedded in the design, development, and implementation of educational software platforms. This interaction is more than merely user-friendly, but it is challenging the in-depth understanding of the information at hand. Cognitive load is an important constraint for unconcerned use of educational software. Apart from hampering processing and transferring information to memory, cognitive load undermines spontaneous exploration across content layers. The linear nature of most actual course material can be broken up by the users' inquisitiveness, but only if they decide to do so and if actions to switch to parallel content layers in an educational software platform do not imply a cognitive burden. The integration of visuals and other media types, text and sound, has to be well-considered. Consistent appearance and use of hyperlinks and hotspots, an arch of tension with some moments to pause, and repeating screen elements are important in designing an appealing and effective educational software platform.

REFERENCES

Avgerinou, M. & Ericson, J. (1997). A review of the concept of visual literacy. *British Journal of Educational Technology*, *28*, 280-291.

Campbell, B. S. (1994). *High school principal roles and implementation themes for mainstreaming information literacy instruction.* Unpublished doctoral dissertation, University of Connecticut.

Debes, J. (1969). The loom of visual literacy. *Audiovisual Instruction*, *14* (8), 25-27.

Erstad, O. (1998). Media literacy among young people: Integrating culture, communication and cognition. In B. Höijer & A. Werner (Eds.), *Cultural cognition: New perspectives in audience theory* (pp. 85-101). Göteborg: Nordicom.

Gee, J. P. (1990). *Social linguistics and literacies: Ideology in discourses*. New York: The Falmer Press.

Gilster, P. (1997). *Digital literacy*. New York: Wiley.

Guthrie, L. F., & Richardson, S. (1995). Language arts: Computer literacy in the primary grades. *Educational Leadership*, *53*(2), 14-17.

Hobbs, R. (1998). Literacy in the information age. In J. Flood, S. B. Heath, & D. Lapp (Eds.), *Handbook of research on teaching literacy through the communicative and visual arts* (pp. 7-14). New York: Simon and Schuster Macmillan.

Hortin, J. (1983). Visual literacy and visual thinking. In L. Burbank & D. Pett (Eds.), *Contributions to the study of visual literacy* (pp. 92-106). Blacksburg, Virginia: IVLA Inc.

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology*. *92*(1), 126-136.

Kellner, D. (1998). Multiple literacies and the critical pedagogy in a multicultural society. *Educational Technology*, *48*(1), 103-122.

Kirschner, P. (2002). Cognitive load theory: implications of cognitive load theory on the design of learning. *Learning and Instruction*, *12*, 1-10.

Marvin, C. (1984). Constructed and reconstructed discourse: Inscription and talk in the history of literacy. *Communication Research*, *11*(4), 563-594.

Maylath, B. (1993). Electronic literacy: What's in store for writing and its instruction? Paper presented at the *Annual Meeting of the Conference on College Composition and Communication* (44th, March 31-April 3), San Diego, CA,

OECD (1997). *Literacy skills for the knowledge society: Further results from the International Adult Literacy Survey*. Paris: OECD/HRD Canada.

Peha, J. M. (1995). How K-12 teachers are using networks. *Educational Leadership*, *53*(2), 18-25.

Plowman, L., & Stephen, C. (2003). A "benign addition"? Research on ICT and pre-school children. *Journal of Computer Assisted Learning*, *19*, 149-164.

Potter, W. J. (1998). *Media literacy*. London: Sage.

Spiro, R. J., Feltovich, R. L., Jacobson, M. J., & Coulson, R. L. (1991). Cognitive flexibility, constructivism, and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domains. *Educational Technology*, *31*, 24-33.

Tyner, K. (1998). *Literacy in a digital world*. Wahwah, N.J.: Lawrence Erlbaum

Utsi, S., & Lowyck, J. (2002). Empirical validation of the concept "multimedia literacy". Paper presented at the *SIG Instructional Design of the European Association for Research on Learning and Instruction* (EARLI) (Erfurt, 27th - 29th June).

Walter, P. (1999). Defining literacy and its consequences in the developing world. *International Journal of Lifelong Education*, *18*(1), 31-48.

KEY TERMS

Children: Tutees, enrolled in primary school.

Cognitive Load: Amount of mental resources necessary for information processing.

Educational Software: Software packages, supporting specific goals in the education of target groups, e.g. primary school tutees or impaired children.

End-User: Tutee, working with dedicated educational software packages.

Impaired Learners: learners, hampered by physical of psychological deficiencies.

Instructional Design: Lay-out of an optimal integration of educational content and interface lay out of end-user software.

Learning: Cognitive processing and integration of new educational content, if possible induced through exercises or games.

Literacy: Operational and cognitive skills, necessary to work effectively and efficiently with educational software.

Digitization of Library Information and Its Accessibility for People with Disabilities

Axel Schmetzke
University of Wisconsin-Stevens Point, USA

INTRODUCTION

After 20 years of digitization efforts, hardly a single type of library information resource remains that has not shifted, at least to some extent, to an electronic, Web-based format: information about the library itself, catalogs, indexes, dictionaries and encyclopedias, books and journals, tutorials, reserve materials and reference services. The online migration of these resources has opened unprecedented opportunities to people with "print disabilities", who cannot independently access printed works because of lack of sight, dyslexia or insufficient motor control (Coombs, 2000), but who are able to access electronic text with the help of assistive input and output technology, such as modified computer keyboards and screen readers with speech or Braille output (Mates, 2000; Lazzaro, 2001).

The extent to which these new opportunities become realized depends on the design of the Web environment. From the perspective of accessibility, design in the online world matters as much as it does in the physical world. This article seeks to determine the extent to which the library profession addresses the need of people with disabilities for accessibly designed online resources—by reviewing the professional library literature for coverage of this issue, by summarizing empirical accessibility studies, and by analyzing pertinent policies adapted by libraries and their professional organizations.

COVERAGE OF ONLINE ACCESSIBILITY IN THE LIBRARY LITERATURE

In 1996, accessible Web design began to emerge as an issue in the professional library literature. Since 1999, there has been a noticeable increase in library-related journal publications that investigate the accessibility of Web-based library information, seek to raise awareness concerning the need for accessible Web design, and provide practical tips (for a detailed overview, see Schmetzke, 2003, pp.153-156). Since 2001, two library journals, *Computers in Libraries*, "Making Sure", 2001), and *Library Hi Tech* (Schmetzke, 2002a, 2002b) have devoted special-theme issues to online accessibility; *Information Technology and Disability* reports regularly on the subject. In 1999, the American Library Association began publishing monographs that addressed accessible Web design (Lazzaro, 2001; Mates, 2000; McNulty, 1999). Gradually, accessibility is acknowledged as an important aspect even in more general works: While some authors, such as Pace (2003) in his book *The Ultimate Digital Library*, continue to ignore the issue, others deal with it, at least briefly, in connection with topics such as Web page design (Garlock & Piontek, 1999), Web site usability testing (Norlin & Winter, 2002), digital resources selection and digital video (Hanson & Lubotsky Levin, 2003), Web-based instruction (Sharpless Smith, 2001) and virtual reference service (Coffman, 2003).

EMPIRICAL RESEARCH FINDINGS

Of the online resources provided by libraries, Web pages have been studied the most. The vast majority of studies employed Bobby, a software-based accessibility checker, to investigate conformance to the 1999 Web Accessibility Guidelines (WCAG), developed by the World Wide Web Consortium's Web Accessibility Initiative. Recently, researchers also began looking at compliance with the "Access Board" standards, a similar set of accessible design criteria developed under Section 508 of the U.S. Rehabilitation Act Amendments of 1998 (Architectural and Transportation Barriers Compliance Board, 2000).

At the library Web sites evaluated between 1999 and 2002, 19% to 75% of the Web pages were found to be free of major accessibility problems (Blake, 2000; Kester, 1999; Lilly & Van Fleet, 1999, 2000; Schmetzke, 2001a, 2003; Spindler, 2002; Yu, 2002); the average number of errors per page varied between 1.3 and 6.1 (Schmetzke, 2002c). Web accessibility tends to be higher at academic libraries than at public libraries. While Web accessibility has improved in some pockets, such as the libraries within the University of Wisconsin system (Schmetzke, 2004), a comparison of 2000 and 2002 data that were collected from the same nationwide sample consisting of the Web sites of 24 mostly larger academic libraries shows that, on the average, the percentage of barrier-free pages has actually slightly declined—from 59% to 53% (Schmetzke, 2003). A break-down of the Web sites into those that had undergone a major redesign during the period in question, and those that did not, revealed that accessibility at the former had drastically declined (from 47% to 24%), whereas accessibility at the latter had improved considerably (from 68% to 81%). Apparently, Web designers involved in the complete re-design of their Web sites tended not to pay attention to accessibility. That the occasion of a redesign can be an opportunity for improving accessibility was shown by a British study involving Web accessibility audits at 11 higher-education Web sites, at least some of which included "gateways" to information provided by libraries. Of the six sites that underwent a major redesign, all showed significant improvements with regard to the problems revealed during the audit (Sloan, Gregor, Booth, & Gibson, 2002).

Perhaps most remarkable is the fact that the Web pages of accredited schools of library and information science (SLIS)—those institutions that train the next generation of librarians—tended to be particularly inaccessible. In 2002, only 30% of the SLIS pages (at U.S. campuses) were free of barriers. Accessibility was barely higher, at 36%, at Canadian schools (Schmetzke, 2003). It is reasonable to assume that these figures not only reflect wide-spread unawareness about the need for accessible design among the SLIS Web designers but also among the library school faculty and staff, who hire the designers and give them direction. Similar lack of awareness among the leadership was also reported for the area of distance education (Schmetzke, 2001b) and in connection with several high-profile technology-promoting initiatives in higher education (Blair, Goldmann, & Relton, 2004).

Information about the accessibility of Web-based library resources other than library Web pages is comparatively scarce. Prior to 2002, little had been published in this area. Then, in 2002, *Library Hi Tech* (Schmetzke, 2002a, 2002b) published two special-theme issues that included accessibility studies on selected online catalogs, online indexes and databases, e-journals, online reference works and courseware. While few of the online resources reviewed were found to be absolutely inaccessible, most contained at least some accessibility problems (for an overview, see Schmetzke, 2002c). Several authors pointed out that lack of usability, rather than accessibility, was often the problem (Axtell & Dixon, 2002; Byerley & Chambers, 2002). Stewart (2003), who currently provides the most comprehensive and up-to-date information on the accessibility and usability of selected online indexes, databases and electronic journals, provides a similar picture. Comparing his current findings with unpublished data collected in 1999, Stewart observes a "complete reversal in the inaccessibility of the online library databases from approximately 95% inaccessible to 95% accessible". However, he cautions that accessibility, defined in terms of conformance to certain accessible design standards, does not automatically result in usability. As an example, he points to SilverPlatter's database interface which, for the most part, conforms to the Access Board Standards (Section 508), but which is so poorly designed that it is extremely difficult to use with an audio browser.

Until a few years ago, anecdotal evidence suggested that vendors showed little, if any, concern for the accessibility of their products and that their sales representatives were typically ill prepared to discuss the issue. A recent survey by Byerley and Chambers (2003) reveals that the situation has changed significantly: Vendors have become more aware of accessibility and started to remove access barriers from their products. In line with previous findings that usability is often more a problem than accessibility (see previous paragraph), the authors discovered that vendors' efforts are largely focused on conformance to Section 508 standards. Only three of the 11 companies that responded to the survey reported that they conduct usability tests involving people with disabilities. The survey also revealed that most database companies do not provide accessibility information on their corporate Web sites, which makes it difficult for accessibility-conscientious customers to make informed purchasing decisions.

ACCESSIBILITY POLICIES

Under the pressure of the Americans with Disabilities Act of 1990 (ADA Hankbook, 1995) and the widening influence of Section 508, many U.S. colleges and universities have adopted campus-wide accessible Web policies during the past years. Typically, these policies either recommend or require compliance with WCAG, the Access Board Standards issued under Section 508, or some combination or variation thereof (Bohman, 2004).

Some, mostly larger, academic libraries have picked up the campus-wide mandate for accessible Web pages and addressed it in their own policies. Among the first to do so was Yale University Library (2000) which, in its *Library Services for Persons with Disabilities Policy Statement*, requires compliance with WCAG's priority levels one and two.

Very few libraries have adopted policies that address the issue of accessibility in connection with the selection and procurement of online information products. An extensive Web search conducted by this author in November 2003, along with an inquiry posted to axslib-l, an email-based discussion group dedicated to accessibility issues in libraries, yielded all but a handful of such policies: The University of Washington Libraries' (2001) *Selection Guidelines for Internet Resources* direct librarians to consider accessibility when selecting online resources and to weigh the "value of the resource … against the access difficulties presented"; the University of Vermont's (2000) *Electronic Resources Collection Development Policy* simply includes being "accessible to people with a variety of disabilities" as one criterion among the many that collection developers must consider. Neither policy requires that only accessible resources must be selected; they merely emphasize the need to consider accessibility in the selection process. Considerably stronger in tone is the policy adopted by California State University (CSU) (2004). Its *Principles for CSU Acquisition of Electronic Information Resources* stipulates that "[i]nformation providers should offer interfaces that comply with basic standards for accessibility by users with disabilities." In a similar vein, the University of Wisconsin Libraries' *Strategic Directions for 2003-2005* call for accessible "[W]eb-based information resources (such as journal databases), [W]eb-based instructional applications (such as courseware and simulations) and online services". These more stringent policies echo the procurement guidelines of the International Coalition of Library Consortia (1999), an international group comprising over 160 library consortia worldwide, which considers its members responsible for "ensur[ing] that vendor platforms are … ADA compliant" and which urges participating libraries to discuss with vendors the development of accessible products**.**

Recognition of accessibility as an important issue varies among professional library organizations. In 2001, the American Library Association (ALA) (2001) approved a policy drafted by one of its branch organizations, the Association of Specialized and Cooperative Library Agencies (ASCLA), that calls for "equal access to library resources" and urges libraries to use "strategies based upon the principles of universal design to ensure that library policy, resources and services meet the needs of all people". However, the need to create a barrier-free electronic environment is not mentioned at all in the policy's section that addresses, in some detail, the accessibility of library services and collections.

While the call for a barrier-free Web environment is present in the ALA-ASCLA policy at least in the form of a general principle, it is not mentioned at all in the recent *Guidelines for Distance Learning Library Services* issued by one of ALA's major branches, the Association of College and Research Libraries (2000) —guidelines that not only reflect the professional views of the broader library and higher-education community, but that also have been influential in shaping individual libraries' policies and practices.

FUTURE TRENDS

Despite considerable coverage of online accessibility issues in the library literature, much remains to be done. Librarians need to further educate themselves and their leadership, reshape their institutional policies, and develop effective implementation strategies that ensure 100% accessibility of their home-grown Web resources and that challenge vendors to make their Web-based products more accessible. Unless this happens, the online library environment is unlikely to transform into what Berners-Lee (n.d.), WWW inventor and the current director of the World Wide Web Consortium (1999), had envisioned: "The power of the Web is its universality. Access by everyone regardless of disability is an essential aspect."

CONCLUSION

To a large extent, disability is a social construct. Whether individuals with "disabilities" can pursue independent and fulfilling lives is not merely a matter of their particular internal conditions but also a question of enabling or disabling external factors put in place by society and its institutions. Libraries clearly are part of this nexus. By neglecting to remove all barriers from their Web pages and by not adopting policies that seek to realize the opportunities new information technology provides, many libraries currently fail to create the conditions that would enable all people, including those with "disabilities", to participate fully in the evolving information society.

REFERENCES

ADA Handbook: Disability discrimination: Statutes, regulations and related materials. (1995). Cincinnati: Anderson Publishing Co.

American Library Association. (2001, January 16). Association of Specialized and Cooperative Library Agencies. *Library Services for People with Disabilities Policy*. Retrieved May 5, 2004, from *http://www.ala.org/Content/NavigationMenu/ASCLA/Issues/Library_Services_for_People_with_Disabilities_Policy_Passes.htm*

Architectural and Transportation Barriers Compliance Board. (2000). *Electronic and information technology accessibility standards*. 36 CFR Part 1194. Retrieved May 5, 2004, from *http://www.access-board.gov/sec508/508standards.htm*

Association of College and Research Libraries. Distance Learning Section. Guidelines Committee (2000). Guidelines for distance learning library services. *College and Research Libraries News, 61*(11), 1023-1029.

Axtell, R., & Dixon, J.M. (2002). Voyager 2000: A review of accessibility for persons with visual disabilities. *Library Hi Tech*, *20*(2), 141-147.

Berners-Lee, T. (no date). Cited from the Web Accessibility Initiative Web site. Retrieved May 5, 2004, from *http://www.w3.org/WAI/*

Blair, M.E., Goldmann, H., & Relton, J. (2004). Access to electronically-mediated education for students with disabilities: Policy issues. *National Center on Disability and Access to Education*. Retrieved May 5, 2004. *http://ncdae.org/papers/policy.htm*

Blake, S. (2000). Universal access, the ADA, and your library Web page. *Arkansas Libraries, 57*(1), 19-24.

Bohman, P.R. (2004). University Web accessibility policies: a bridge not quite far enough. *WebAim*. Retrieved May 5, 2004, from *http://www.webaim.org/coordination/articles/policies-pilot*

Byerley, S.L., & Chambers, M.B. (2002). Accessibility and usability of Web-based library databases for non-visual users. *Library Hi Tech*, *20*(2), 169-178.

Byerley, S.L., & Chambers, M.B. (2003). Accessibility of Web-based library databases: The vendors' perspectives. *Library Hi Tech, 21*(3), 347-357.

California State University. (2004, January 21). Principles for CSU acquisition of electronic information resources. May 5, 2004, from *http://seir.calstate.edu/acom/ear/docs/principles.shtml*

Coffman, S. (2003). *Going live. Starting & running a virtual reference service*. Chicago: American Library Association.

Coombs, N. (2000). Enabling technologies. Untangling your Web. *Library Hi Tech, 18* (1), 93-96.

Garlock, K.L., & Piontek, S. (1999). *Designing Web interfaces to library services and resources*. Chicago: American Library Association.

Hanson, A., & Lubotsky Levin, B. (2003). *Building a virtual library*. Hershey, PA: Information Science Publishing.

International Coalition of Library Consortia. (1999, January). *Guidelines for technical issues in Request for Proposal (RFP) requirements and contract negotiations*. Retrieved May 5, 2004, from *http://www.library.yale.edu/consortia/techreq.html*

Kester, D. (1999). Measuring the sight of your Web site. *North Carolina Libraries*, *57*(3), 114-117.

Lazzaro, J.J. (2001). *Adaptive technologies for learning & work environments* (2nd ed.). Chicago: American Library Association.

Lilly, E.B., & Van Fleet, C. (1999). Wired but not connected: Accessibility of academic library home pages. *The Reference Librarian*, No. 67/68, 5-28.

Lilly, E.B., & Van Fleet, C. (2000). Measuring the accessibility of public library home pages. *Reference & User Services Quarterly, 40*(2), 156-163.

Making Sure Everyone Has Access to Information (special issue). (2001). *Computers in Libraries, 21*(9).

Mates, B.T. (2000). *Adaptive technology for the Internet: Making electronic resources accessible to all.* Chicago: American Library Association. [Free online version at *http://www.ala.org/ala/products/books/editions/adaptivetechnology.htm*]

McNulty, T. (1999). *Accessible libraries on campus. A practical guide for the creation of disability-friendly libraries.* Chicago: Association of College and Research Libraries, American Library Association.

Norlin, E., & Winters, CM! (2002). *Usability testing for library web sites. A hands-on guide.* Chicago: American Library Association.

Pace, A.K. (2003). *Ultimate digital library: Where the new information players meet.* Chicago: American Library Association.

Schmetzke, A. (2001a). Web accessibility at university libraries and library schools. *Library Hi Tech, 19*(1), 35-49.

Schmetzke, A. (2001b). Online Distance Education—"anytime, anywhere" but not for everyone, *Information Technology and Disabilities*, *7*(2). Retrieved May 5, 2004, from *http://www.rit.edu/~easi/itd/itdv07n2/contents.htm*

Schmetzke, A. (Ed.). (2002a). Accessibility of Web-based information resources for people with disabilities (special issue). *Library Hi Tech, 20*(2).

Schmetzke, A. (Ed.). (2002b). Accessibility of Web-based information resoureces for people with disabilities (part two) (special issue). *Library Hi Tech*, *20*(4).

Schmetzke, A. (2002c). The accessibility of online library resources for people with print disabilities: Research and strategies for change. *Computers Helping People with Special Needs. Proceedings of the 8th International ICCHP Conference*, July 15-20, 2002, Linz, Austria. Berlin: Springer Verlag: 390-397.

Schmetzke, A. (2003). Web accessibility at university libraries and library schools: 2002 follow-up study. In M. Hricko (Ed.), *Design and implementation of Web-enabled teaching tools* (pp.145-189). Hershey, PA: Information Science Publishing.

Schmetzke, A. (2004). Web page accessibility on University of Wisconsin campuses: 2004 survey and six-year trend data. Retrieved May 5, 2004, from *http://library.uwsp.edu/aschmetz/Accessible/UW-Campuses/Survey2004/contents2004.htm*

Sharpless Smith, S. (2001). *Web-based instruction. A guide for libraries*. Chicago: American Library Association.

Sloan, D., Gregor, P., Booth, P., & Gibson, L. (2002). Auditing accessibility of UK higher education Web sites. *Interacting with Computers*, *12*, 313-325.

Spindler, T. (2002). The accessibility of Web pages for mid-sized college and university libraries. *Reference & User Services Quarterly*, *42*(2), 149-154.

Stewart, R. (2002, November 15). *Accessibility of online databases. A usability study of research databases.* Technology Access Program. Oregon University. Retrieved May, 2004, from *http://tap.oregonstate.edu/research/ahg.htm*

University of Vermont Libraries. (2000, June 16). Electronic Resources Coordinating Council. *Electronic resources collection development policy*. Retrieved May 5, 2004, from *http://bailey.uvm.edu/ercc/appendix6.html*

University of Washington Libraries. (2001, November 29). *Selection guide for Internet resources*. Retrieved May 5, 2004, from *http://www.lib.washington.edu/msd/internetselguide.html*

World Wide Web Consortium (W3C) (1999). *Web content accessibility guidelines 1.0*. Retrieved May 5, 2004, from *http://www.w3.org/TR/WAI-WEBCONTENT/*

Yale University Library (2000*). Library services for persons with disabilities policy statement*. Retrieved May 5, 2004, from *http://www.library.yale.edu/Administration/SQIC/spd1.html*

Yu, H. (2002). Web accessibility and the law: Recommendations for implementation. *Library Hi Tech*, *20*(4), 406-419.

KEY TERMS

Access Board Standards: Technical and functional performance criteria developed by the Architectural and Transformation Barriers Compliance Board (the “Access Board”), a U.S. government agency, under Section 508. Only electronic and information technology conforming to these standards is considered accessible.

Accessibility: As defined within Section 508, accessibility is achieved when individuals with disabilities can access and use information technology in ways comparable to those available to people without disabilities. A narrower, operational definition conceptualizes accessibility in terms of conformance to certain accessibility criteria, such as the *Web Content Accessibility Guidelines* or the *Access Board Standards*.

Accessible Web Design: Also sometimes referred to as “barrier-free Web design”. Web design that strives to accommodate the needs of people with disabilities, including those using assistive technology, to access the Web environment.

Americans with Disabilities Act: U.S. civil rights legislation passed in 1990 that prohibits discrimination against people with disabilities in the areas of employment, transportation, telecommunications, and public accommodation.

Assistive Technology: Specialized software or hardware, such as screen readers, magnification software and a modified keyboard, used by some people with disabilities to interact with the computer.

Audio Browser: Also referred to as “talking browser”. Software that interprets the html code of Web pages and provides speech output for text-based components, along with information provided by the html mark-up tags. Typically, it also enables users to navigate the Web page through alternative keystrokes.

Print Disabilities: Comprises all those disabilities that make it difficult, or impossible, to read printed text. The term includes visual impairment and blindness; cognitive disabilities, such as dyslexia; and certain motor-control impairments.

Screen Reader: Software that interprets the signals sent to the computer screen and reads aloud the displayed text with the help of a speech synthesizer.

Section 508: A provision within the Rehabilitation Act of 1973, as amended by Congress in 1998, that mandates that the electronic and information technology developed, maintained, procured or used by the U.S.

government must be accessible to people with disabilities.

Universal Design: A concept similar to accessible design. Its meaning is broader in that it refers to design that strives to create products that are usable by all people, regardless of age, gender, (dis)ability, handedness, and so forth. Its meaning is narrower in that it seeks one solution to accommodate the needs of all people.

Web Content Accessibility Guidelines (WCAG): Guidelines for accessible Web design developed by the World Wide Web Consortium's Web Accessibility Initiative. WCAG 1.0 were passed in 1999. A working draft of a revised set of guidelines, WCAG 2.0, is currently under review.

Dimensions of Database Quality

John A. Hoxmeier
Colorado State University, USA

INTRODUCTION

The ultimate objective of database analysis, design, and implementation is to establish an electronic repository that faithfully represents the conceptual and logical model of the manageable aspects of a user's information domain. Enterprise and Web-enabled databases must satisfy a wide set of demands and constituents. Software engineering in general and database development in particular can be a complex, complicated process. There is probably no other product development process that faces the same amount of uncertainty, which may account for the high failure rate of software projects. This chapter expands on the growing body of literature in the area of data quality by proposing additions to a hierarchy of database quality dimensions that include model and behavioral factors in addition to the process and data factors.

BACKGROUND

While data quality has been the focus of a substantial amount of research, a standard definition does not exist in the literature (Wang & Madnick, 2000). The International Organization for Standardization (ISO) supplies an acceptable definition of data quality using accepted terminology from the quality field. These standards are documented agreements containing technical specifications or other precise criteria to be used consistently as rules, guidelines, or definitions of characteristics to ensure that materials, products, processes, and services are fit for their purpose. Applying the term *database quality* in this context would build on the ISO definition of quality, that is, "conformance to requirements" and "fitness for use." ISO 8402 as a quality management and quality assurance metric provides a formal definition of quality: the characteristics of an entity that represent its ability to satisfy stated and implied needs. This definition is consistent with the notion of customer satisfaction prevalent in the quality literature (Crosby, 1995; Juran, 1989). Thus, a database can be defined to be of the required quality if it satisfies the requirements stated in a specification, and the specification reflects the implied needs of the user. Therefore, an acceptable level of quality has been achieved if the database conforms to a defined specification, and the specification correctly reflects the intended use. Unfortunately, neither of these definitions is adequate for the purposes of assessing database quality. A database must also be judged by how closely it represents the world of the data consumer (the model), its ability to respond to both routine and unanticipated requests within the domain it is expected to manage (the behavior), and maintain this representation over time. The framework presented herein expands on work previously proposed (Hoxmeier, 1997) and incorporates data quality dimensions put forth by several prominent data quality researchers (Ballou & Pazar, 1995; Krogstie, Lindland, & Sindre, 1995; Lindland, Sindre, & Solvberg, 1994; Orr, 1998; Storey & Wang, 1994; Strong, Lee, & Wang, 1997; Wand & Wang, 1996; Wang, 1998; Wang, Kon, & Madnick, 1993; Wang et al., 1994; Wang, Storey, & Firth, 1995; Wang, Strong, & Guarascio, 1996; Wang, Strong, Kahn, & Lee, 1999). The framework is important because it expands the definition of strict data quality to that of a broader context of database quality and incorporates the importance of process management.

THE CHALLENGE

Many database applications are ultimately unsuitable to the consumer. The process must incorporate three conceptually distinguishable domains: the modeling, the performance, and the enactment domains. Designers attempt to conceptualize the problem domain into a suitable physical model. The proposed physical model is subject to many performance constraints including the physical representation, the network topology, system configuration, and system administration. Finally, and what may be the most difficult to administer, the information is presented to the consumer for interpretation and enactment. The representation of the database after each of these domain layers all contribute to the quality of the solution by the information consumer. The critical elements below are the bases for the discussion on database quality dimensions.

- The cycle process must be managed toward a successful outcome.
- The model itself must represent a usually diverse and fuzzy problem domain.

- The quality of the data in the database must be of sufficient grade.
- The application must behave or have the ability to behave in a way the consumer understands.

To ensure a quality database application, should the emphasis during model development be on the application of quality-assurance metrics (designing it right)? It is hard to argue against this point, but there are a significant number of studies and anecdotal evidence that suggests that a large number of database applications fail, are unusable, or contribute to negative organizational consequences (Abate, Diegert, & Allen, 1998; Redman, 1998; Stackpole, 2001; Standish Group, 1997; Wand & Wang, 1996). The Data Warehousing Institute estimates that businesses lose billions each year attributable to bad data (Eckerson, 2002; Trembley, 2002). A quality process does not necessarily lead to a usable database product (Arthur, 1997; Hoxmeier, 1995; Redman, 1995). There are also many examples of database applications that are in most ways well formed with high data quality but lack semantic or cognitive fidelity (the right design; Motro & Rakov, 1999). Additionally, determining and implementing the proper set of database behaviors can be an elusive task.

While researchers have developed a fairly consistent view of data quality, there is little available in the literature on the evaluation of overall database quality including other considerations such as semantic fidelity (model), behavioral, and value factors.

A Database Quality Framework

It is proposed that through the hierarchical framework presented in Figure 1, one can consider overall database quality by assessing four primary dimensions: process, data, model, and behavior. Portions of the hierarchy draw heavily from previous studies on data and information quality, and documented process quality standards (Arthur, 1997; Department of Commerce, 2004; Wang, 1998). A dimension is a set of database quality attributes or components that most data consumers react to in a fairly consistent way (Wang et al., 1996). Wang et al. define data quality dimension as a set of data quality attributes that represent a single data quality abstract or construct. The use of a set of dimensions to represent a quality typology is consistent with previous quality research (Dvir & Evans, 1996; Strong et al., 1997; Wang et al., 1996). The framework presents the four dimensions in a dimension-attribute-property hierarchy.

Process Quality

Much attention has been given over the years to process quality improvement. *ISO-9000-3, total quality management* (TQM), *quality function deployment* (QFD), and the *capability maturity model* (CMM) are approaches that are concerned primarily with the incorporation of quality management within the process of systems development (Dvir & Evans, 1996; Herbsleb, 1997; Hill, 2003; Schmauch, 1994). *Quality control* is a process of ensuring that the database conforms to predefined standards and guidelines using statistical quality measures. *Quality assurance* attempts to maintain the quality standards in a proactive way. In addition to using quality control measures, quality assurance goals go further by surveying the customers to determine their level of satisfaction with the product. Conceivably, potential problems can be detected early in the process.

Database Data Quality

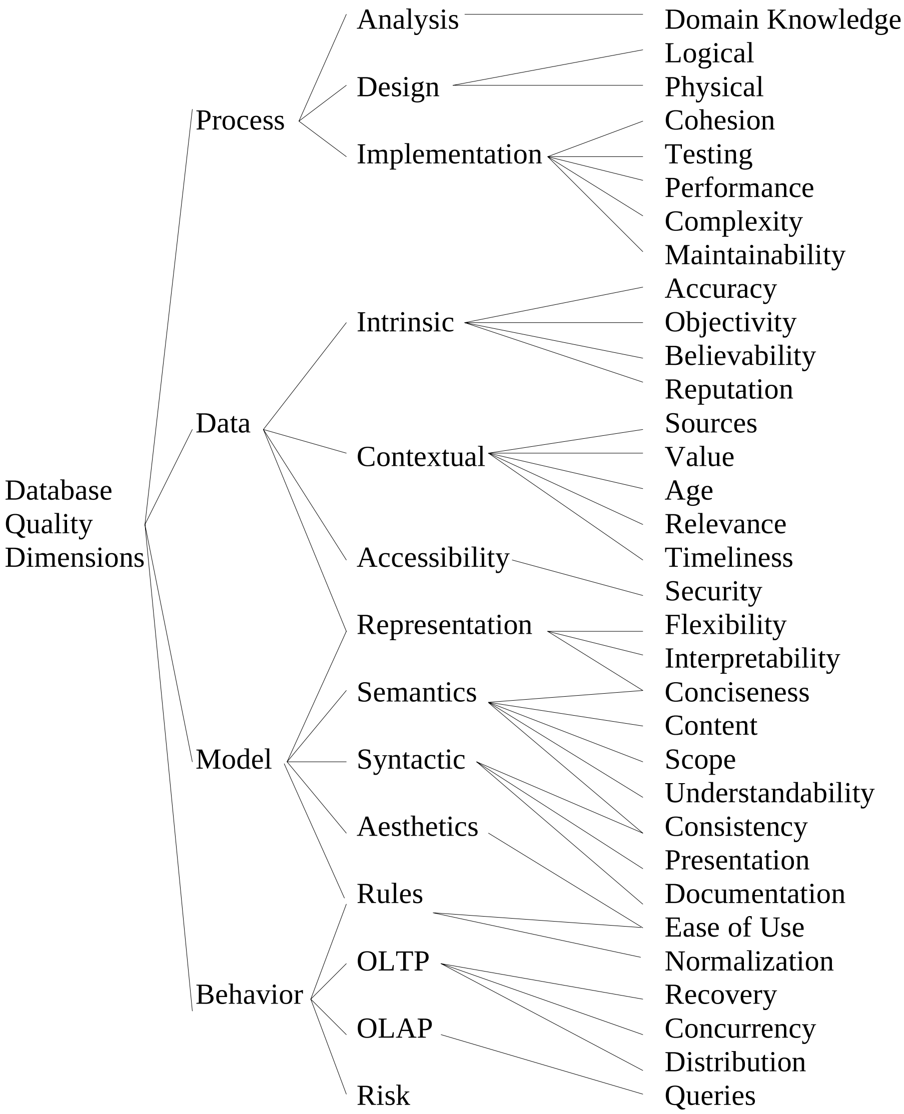
Data integrity is one of the keys to developing a quality database. Without accurate data, users will lose confidence in the database or make uninformed decisions. While data integrity can become a problem over time, there are relatively straightforward ways to enforce constraints and domains and to ascertain when problems exist (Moriarty, 1996). The identification, interpretation, and application of business rules, however, present a more difficult challenge for the developer. Rules and policies must be communicated and translated and much of the meaning and intent can be lost in this process.

Data Model Quality

As has been presented, data quality is usually associated with the quality of the data values. However, even data that meet all other quality criteria is of little use if they are based on a deficient data model (Levitin & Redman, 1995). Data model quality is the third of the four high-level dimensions presented above. Information and an application that represent a high proportionate match between the problem and solution domains should be the goal of a database with high semantic quality. Representation, semantics, syntax, and aesthetics are all attributes of model quality (Levitin & Redman; Lindland et al., 1994).

The database design process is largely driven by the requirements and needs of the data consumer, who establishes the boundaries and properties of the problem domain and the requirements of the task. The first step in the process, information discovery, is one of the most difficult, important, and labor-intensive stages of database development (Sankar & Marshall, 1993). It is in this stage where the semantic requirements are identified, prioritized, and visualized. Requirements can rarely be defined in a serial fashion. Generally, there is significant uncertainty over what these requirements are, and they only

Figure 1. Database quality dimensions



become clearer after considerable analysis, discussions with users, and experimentation with prototypes.

Qualitative and quantitative techniques can be used to assist the developer to extract a strong semantic model. However, it is difficult to design a database with high semantic value without significant domain knowledge and experience (Moody & Shanks, 1998; Navathe, 1997). These may be the two most important considerations in databases of high semantic quality. In addition, conceptual database design remains more of an art than a science. It takes a high amount of experience, creativity, and vision to design a solution that is robust, usable, and that can stand the test of time.

Database Behavior Quality

Many databases are perceived to be of low quality simply because they are difficult to use. Developers tend to focus on aspects of data quality at the expense of behavioral quality. What constitutes a database of high behavioral quality? Are the criteria different than those used for software applications in general? Clearly the behaviors for a database that is used to support transaction processing (OLTP) are different than those of a database used to support analytical processing (OLAP). Software development, in general, is very procedure or function driven. The objective is to build a system that works (and does it quickly). Database development, on the other hand, should be more focused on the content, context, behavior, semantics, and persistence of the data. The process of behavior implementation consists of the design and construction of a solution following the identification of the problem domain and the data model.

FUTURE TRENDS

The framework presented above offers a typology for assessing the various dimensions of database quality. The purpose of this paper was to expand on the existing research on data and process quality in an attempt to provide a more comprehensive view of database quality. The area is of great concern as information is viewed as a critical organizational asset, and knowledge management and the preservation of organizational memory has become a high priority. It has been estimated by the Data Warehousing Institute, the Gartner Group, Tom Redman, and others that organizations are losing billions of dollars due to poor data quality, and the problem is exacerbated by integrated systems. Most organizations realize that their databases may contain inaccurate data, but they underestimate the business risk of the result: poor information quality (Loshin, 2000). Yet most organizations are unwilling to spend the time and resources necessary to improve the situation. Tom Redman (2004) describes such a scenario in his fictitious case study, *Confronting Data Demons*. The case study describes a manufacturing firm's struggle to understand and correct data quality problems. Poor data quality is just one aspect of the case. The problem is magnified by the nature of the integrated database (ERP, CRM) and by poorly structured processes.

Part of the difficulty has been in the justification of such a project. Like many productivity improvement initiatives, a methodology based on ROI is required. Perhaps techniques such as the Data Quality Scorecard hold promise (Campbell & Wilhoit, 2003). Further research is required to continue to understand the risks, quantify the costs, improve the model, validate the frameworks, and identify additional data and database quality dimensions.

CONCLUSION

How does one ensure a final database product that is of high quality? Database quality must be measured in terms of a combination of dimensions including process and behavior quality, data quality, and model fidelity. By organizing attributes into database quality dimensions, many difficulties encountered when dealing with singular attributes can be effectively addressed. So, not only are dimensions more comprehensive, but organizing attributes into dimensions both organizes and minimizes the material that must be comprehended. Moreover, by analyzing dimensions, a data quality researcher may discover systemic root causes of data errors.

REFERENCES

Abate, M., Diegert, K., & Allen, H. (1998, September). A hierarchical approach to improving data quality. *Data Quality, 4*(1). Retrieved October 2000 from *http://www.dataquality.com/998abate.htm*

Arthur, L. (1997). Quantum improvements in software system quality. *Communications of the ACM, 40*(6), 47-52.

Ballou, D., & Pazer, H. (1995). Designing information systems to optimize the accuracy timeliness tradeoff. *Information Systems Research, 6*(1), 51-72.

Campbell, T., & Wilhoit, Z. (2003). *How's your data quality? A case study in corporate data quality strategy*. Proceedings of the Eighth International Conference on Information Quality (ICIQ-03). Retrieved June 2004 from *http://www.iqconference.org/Documents/IQ%20Conference%202003/Papers/HowsYourDataQualityACaseStudy.pdf*

Crosby, P. (1995). *Quality is still free*. New York: McGraw-Hill.

Department of Commerce, Bureau of Information Security. (2004). *Information quality guidelines*. Retrieved June 2004 from *http://www.bxa.doc.gov/pdf/qualityguidelines.pdf*

Dvir, R., & Evans, S. (1996). *A TQM approach to the improvement of information quality*. Retrieved October 2000 from *http://web.mit.edu/tdqm/www/wpaper.htm*

Eckerson, W. (2002). *Data quality and the bottom line* [The Data Warehouse Institute report series white paper]. Retrieved October 2002 from *http://dw-institute.com*

Herbsleb, J., Zubrow, D., Goldenson, D., Hayes, W., & Paulk, M. (1997). Software quality and capability maturity model. *Communications of the ACM, 40*(6), 30-40.

Hill, A. (2003, July). Quality in the boardroom. *TickIT International, 3Q*, 3-4.

Hoxmeier, J. (1995). Managing the legacy systems reengineering process: Lessons learned and prescriptive advice. *Proceedings of the Seventh Annual Software Technology Conference*, Ogden ALC/TISE, Salt Lake City, UT.

Hoxmeier, J. (1997). A framework for assessing database quality. *Proceedings of the Workshop on Behavioral Models and Design Transformations: Issues and Opportunities in Conceptual Modeling, ACM Sixteenth International Conference on Conceptual Modeling*, Los Angeles, CA.

Juran, J. M. (1989). *Juran on leadership for quality: An executive handbook.* New York: Free Press.

Krogstie, J., Lindland, O., & Sindre, G. (1995). Towards a deeper understanding of quality in requirements engineering. *Proceedings of the Seventh CaiSE*, 82-95.

Levitin, A., & Redman, T. (1995). Quality dimensions of a conceptual view. *Information Processing and Management, 31*(1).

Lindland, O., Sindre, G., & Solvberg, A. (1994, March). Understanding quality in conceptual modeling. *IEEE Software, 11*(2), 42-49.

Loshin, D. (2000). *Enterprise knowledge management: The data quality Approach.* Morgan Kaufmann.

Moody, D., & Shanks, G. (1998). What makes a good data model? A framework for evaluating and improving the quality of entity relationship models. *The Australian Computer Journal, 30*(4), 97-110.

Moriarty, T. (1996, May). Barriers to data quality. *Database Programming and Design, 61*.

Motro, A., & Rakov, I. (1999). Estimating the quality of databases. In *Lecture Notes in Computer Science*. Heidelberg, Germany: Springer Verlag.

Navathe, S. (1997). Conceptual modeling in biomedical science. *Proceedings of the ACM Entity Relationship 97 Modeling Preconference Symposium*, Los Angeles, CA.

Orr, K. (1998). Data quality and systems theory. *Communications of the ACM, 41*(2), 66-71.

Redman, T. C. (1995). Improve data quality for competitive advantage. *Sloan Management Review, 36*(2), 99-107.

Redman, T. C. (1998). The impact of poor data quality on the typical enterprise. *Communications of the ACM, 41*(2), 79-82.

Redman, T. C. (2004). *Confronting data demons.* Retrieved June 2004 from *http://www.dataqualitysolutions.com/Confront%20Your%20Data%20Demons.pdf*

Sankar, C., & Marshall, T. (1993). Database design support: An empirical investigation of perceptions and performance. *Journal of Database Management, 4*(3), 4-14.

Saviano, J. (1997, June 1). Are we there yet? *CIO*, 87-96.

Schmauch, C. (1994). *ISO-9000 for software developers*. ASQC Quality Press.

Standish Group. (1997). *The chaos report.* Retrieved October 2000 from *http://standishgroup.com/visitor/chaos.htm*

Storey, V., & Wang, R. (1994). Modeling quality requirements in conceptual database design. *Total data quality management* [Working paper series: TDQM-02-94]. Retrieved October 2000 from *http://web.mit.edu/tdqm/www/wp94.html*

Strong, D., Lee, Y., & Wang, R. (1997). Data quality in context. *Communications of the ACM, 40*(5), 103-110.

Trembley, A. C. (2002). Poor data quality: A $600 billion issue? *National Underwriter, 106*(1), 21.

Wand, Y., & Wang, R. (1996, November). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM.*

Wang, R. (1998). A product perspective on total data quality management. *Communications of the ACM, 41*(2), 58-65.

Wang, R., Kon, H., & Madnick, S. (1993). Data quality requirements analysis and modeling. *Ninth International Conference on Data Engineering*, 670-677.

Wang, R., & Madnick, S. (2000). *Introduction to the TDQM program: MIT data quality program.* Retrieved October 2000 from *http://web.mit.edu/tdqm/www/intro.html*

Wang, R., Storey, V., & Firth, C. (1995). A framework for analysis of data quality research. *IEEE Transactions on Knowledge and Data Engineering, 7*(4), 349-372.

Wang, R., Strong, D., & Guarascio, L. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems, 12*(4), 5-34.

Wang, R., Strong, D., Kahn, B., & Lee, Y. (1999). An information quality assessment methodology: Extended abstract. *Proceedings of the 1999 Conference on Information Quality*, 258-263.

KEY TERMS

Data: The raw material that feeds the process of information generation

Database: A self-describing collection of data that represents a model of an information domain

Database Quality: Includes dimensions of data, process, model, information, and behavioral characteristics

Domain Knowledge: Expertise in a given application area

Information: Data with context and utility

Quality: An encompassing term comprising utility, objectivity, integrity, and value

Semantic Model: The conceptual representation of the problem domain

Discovery of Classification Rules from Databases

Graeme Richards
University of East Anglia, UK

Beatriz de la Iglesia
University of East Anglia, UK

INTRODUCTION

In descriptive data mining, the objective is to build an *understandable* model that provides insight into the behaviour or characteristics of some data. The data comprise a set of records, each of which assigns values to a set of features or attributes. One of these features is designated the target feature; the value of this feature is described as the class of the record.

For example, given a set of motor insurance records including features describing car and driver details and claims history, we may wish to build a model that will classify drivers as high or low risk with respect to their claims history. This model could then be used when assessing premiums for new customers. We may also wish to understand what characterises low risk drivers so that new marketing campaigns can aim to attract them.

BACKGROUND

Complete classification, such as that usually produced by decision trees, assigns a class to each record in the data. This is often unsuitable for the descriptive data mining task as the models built are often very large and difficult to understand. Also, overall classification accuracy, often used as the guiding criterion to construct the classifier, does not guarantee accurate classification of minority classes (i.e., classes with few representative records, for example high insurance risk).

Partial classification (also known as *nugget discovery*) seeks to find simple and understandable patterns that represent "strong" descriptions of a particular class. It is often convenient to use *rules* to express such patterns (Ali et al., 1999). Rules are of the general form

antecedent* $\Rightarrow$ *consequent

where the *antecedent* and *consequent* are predicates that are used to define subsets of records from the database and the rule underlines an association between these subsets. In partial classification, the consequent is fixed to be a particular named class. The *strength* of the rule may be expressed by various measures, as described in the following sections.

We are concerned here with the task of partial classification, specifically with the problem of rule discovery. Firstly, we describe the structure of the classification rules used and how rules may be evaluated. We then go on to describe the various techniques that have been developed for the discovery of classification rules. These are:

- Modern Heuristic Methods - The use of optimisation algorithms.
- Multi-Objective Methods – The use of multi-objective evolutionary algorithms.
- All Rules Search – The use of constrained search algorithms.

RULE STRUCTURE

The number of rules that may be constructed is usually very large and often infinite, but imposing constraints on the structure of rules might reduce this. Highly flexible formats allow a rich expression of patterns, which may encapsulate stronger descriptions of a class, but the size of the search space may be very large. Conversely, if the format is too restrictive it will not be possible to express patterns of sufficient interest.

Many rule discovery techniques are restricted to a rule format where the antecedent comprises a conjunction of attribute tests, ATs, and the consequent comprises a single AT representing the class description.

Even with this restriction on rule format, the size of the search space is usually immense for any real-world problem. It is not normally possible to find all rules. Consequently it is necessary to use rule finding techniques that can search effectively within the search space, as described earlier.

D

EVALUATION OF CLASSIFICATION RULES

Two common measures of rule strength are *confidence* and *coverage,* which are described next.

Given a record, t, $antecedent(t)$ is true if t satisfies the predicate, *antecedent*. Similarly, $consequent(t)$ is true if t satisfies the predicate, *consequent*. Then the subsets defined by the *antecedent* or *consequent* are the sets of records for which the relevant predicate is true.

We define three sets of records:

- $A = \{t \in D \mid antecedent(t)\}$, (i.e., the set of records defined by the *antecedent*),
- $B = \{t \in D \mid consequent(t)\}$, (i.e., the set of records defined by the *consequent*),
- $C = \{t \in D \mid antecedent(t) \wedge consequent(t)\}$.

The support for any conjunction of ATs, M, $sup(M)$ is the number of records which satisfy M.

Given a rule, r, we designate the antecedent of the rule r^a and the consequent r^c.

Then, the support for the antecedent, $sup(r^a) = |A| = a$

and the support for the consequent, $sup(r^c) = |B| = b$, (i.e., the cardinality of the target class).

The *support* for r, $sup(r)$, is defined as $sup(r^a \wedge r^c) = |C| = c$

The *confidence* of r, $conf(r)$, is defined as

$$conf(r) = \frac{sup(r)}{sup(r^a)} = \frac{c}{a}$$

The *coverage* of r, $cov(r)$, is defined as

$$cov(r) = \frac{sup(r)}{sup(r^c)} = \frac{c}{b}$$

A strong rule may be defined as one that meets certain confidence and coverage thresholds. Those thresholds are normally set by the user and are based on domain or expert knowledge about the data. Strong rules may be considered interesting if they are found to be novel and useful. That type of criteria, which may be defined subjectively, can only normally be assessed by interpretation of the rule against the domain knowledge, and against the expectations and needs of the data owner, and so forth. In nugget discovery we are therefore interested in presenting a set of strong rules (possibly interesting rules) to the user for further subjective evaluation.

TECHNIQUES FOR THE DISCOVERY OF CLASSIFICATION RULES

Modern Heuristics

Modern heuristic optimisation techniques, namely simulated annealing, genetic algorithms and tabu search, may be used to extract the best classification rules according to a specified measure of interest (de la Iglesia et al., 1996, 2000). In this approach to nugget discovery the problem of finding strong class descriptions becomes an optimisation problem. We represent a conjunctive classification rule as a solution to this problem, and all the classification rules available given a particular rule format constitute the search space. We then evaluate classification rules using some measure of interest so that the search can be guided towards the most interesting rules according to that measure. One such measure is the fitness measure,

$$f(r) = \lambda c - a \text{ where } \lambda \in \Re$$

In this equation a and c are interpreted as described previously. This measure is capable of partially ordering rules according to confidence and coverage under certain constraints. Under the defined partial ordering, if two rules have the same confidence the rule of higher coverage is preferred, and if two rules have the same coverage the rule of higher confidence is preferred. It follows that if a rule has both higher coverage and confidence than another, then the first rule is preferred. The partial ordering defines a high confidence/coverage boundary from which the heuristic techniques would search for solutions. Variations in the λ parameter allow the algorithms to explore different areas of the upper confidence/coverage boundary, by encouraging the search for rules of high confidence or high coverage.

In the implementation given in de la Iglesia et al. (1996, 2000), a solution or rule is represented as a bit string. Each attribute is assigned a number of bits, with numerical attributes defined by a lower and upper limit and categorical attributes defined by a number of labels. The class label does not need to be represented, as it is fixed. Evaluation is conducted by examining the database to count the support for the antecedent and consequent of the rule, and then by calculation of the evaluation function used. Modern heuristic optimisation algorithms are then used to drive the search towards promising areas of the search space.

Multi-Objective

The problem solved by the modern heuristics can be redefined as a multi-objective optimisation problem. In this instance, instead of using an individual evaluation function representing a combination of the individual characteristics of the rule such as the fitness measure, we will treat each measure of interest or strength of the rule as a separate objective to be optimised. We can use the same rule encoding as was used for the modern heuristic approach. However, we will use a multi-objective evolutionary algorithm (MOEA) as the search engine.

A number of Pareto-based MOEAs have been reported in recent years, for example the Niched Pareto Genetic Algorithm (Horn et al., 1994), the PAES algorithm proposed by Knowles and Corne (1999) and, more recently, the Fast Elitist Non-Dominated Sorting Genetic Algorithm (NSGA II) proposed by Deb et al. (2000).

The multi-objective approach has various advantages with respect to the previous approach. First, instead of finding an individual rule at each run of the algorithm, we can find a set of solutions that represent an approximation to the upper confidence/coverage boundary. Second, the utilisation of measures of strength or interest in a multi-objective approach allows for a more flexible search. For example, if simplicity of the pattern obtained is important, having a third optimisation objective based on the number of ATs in the rule could be incorporated.

An implementation of this approach using NSGA-II showed that the approach is efficient on large databases and finds strong nuggets spread over the confidence/coverage boundary (de la Iglesia et al., 2003).

All Rules

All rules search methods attempt to find *all* classification rules that satisfy certain constraints. The problem is an extension of the *association rule* problem, which was introduced by Agrawal et al. (1993). Methods of finding association rules were given by Agrawal and Srikant (1994) and, more recently, by Han and Pei (2000a, 2000b). One of the first effective classification rule search algorithms was the Dense Miner algorithm (Bayardo & Agrawal, 1999, 2000) in which three constraints were used: the confidence constraint, the coverage constraint and the improvement constraint. The values of three parameters, *minConf*, *minCov* and *minImp* are specified by the user. A rule, r, satisfies the confidence constraint if $conf(r) \geq minConf$; it satisfies the coverage constraint if $cov(r) \geq minCov$; and satisfies the improvement constraint if $imp(r) \geq minImp$. The improvement of a rule, r, $imp(r)$ is defined as $conf(r) - conf(r')$ where r' is the subrule of r with the greatest confidence. A subrule of r, r^*, is any rule that has the same consequent as r and where the antecedent of r^* is a subset of the antecedent of r. Thus, the improvement constraint ensures that all ATs in a rule antecedent contribute significantly to its strength.

Dense miner is suitable for finding rules in classification data comprising only categorical data (numeric data must be discretized so that is they are, effectively, categorical). It explores the search space of rules by expanding a set enumeration tree (SE-Tree). Each node in the SE-tree is represented by a structure called a group. A group, g, has a head, $H(g)$, and a tail, $T(g)$, and $g = H(g)|T(g)$. The head is a conjunction of ATs that represents the antecedent of a rule. The tail is a vector of ATs that can be added to the head to form a new antecedent.

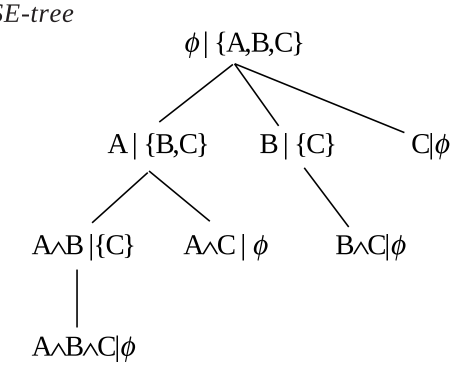
The head and tail of the group at the root of the SE-tree are $H(root)$, which is empty, and $T(root)$, which comprises all ATs in U except those based on the same attribute as the consequent AT, c. A group, g, is expanded to create a set of groups in the next level of the SE-tree by producing one child for each of the ATs in $T(g)$.

Figure 1 shows an SE-tree for the set of ATs {A, B, C}. It can be seen that the antecedent of every possible rule that can be constructed from the ATs is enumerated in the heads of the groups.

During the search, rules that potentially satisfy the constraints are added to a set of rules. A post- processing stage fully enforces the improvement constraint on this set.

To expand the entire SE-tree is usually infeasible. Therefore, at each node, pruning functions are applied. If it can be determined that none of the rules derivable from

Figure 1. SE-tree



a node can satisfy *all* of the constraints then the node can be pruned from the tree without compromising the completeness of the search. The use of these pruning functions is critical to the successful operation of the algorithm. Good pruning functions can ensure that large areas of the SE-tree do not need to be explored.

Algorithm ARA was developed by Richards and Rayward-Smith (2001); this extended the pruning functions used in dense miner (Bayardo & Agrawal, 2000), and a further constraint, the antecedent constraint, was introduced. This constraint restricts the number of ATs that may participate in the antecedent and is incorporated into the pruning functions to allow more efficient search. Subsequently ARA was modified to search effectively for rules that include numerical features, without any preprocessing.

Other rule finding algorithms are described by Megiddo and Srikant et al. (1998) and Brin et al. (1999).

FUTURE TRENDS

Algorithms for finding all rules are being made more efficient, allowing the analysis of larger and more complex databases. In addition, the multi-objective methods are being improved in order to handle ever more complex databases. An emerging technique is the use of these two methods in combination in a complimentary fashion. The multi-objective method may be used to find an approximation to the upper confidence/coverage boundary and then the all-rules method used for a detailed analysis of specific areas of interest.

One of the problems of finding rules is that very many, very similar rules may be found. Therefore, research is being undertaken to develop methods of clustering such rules and representing clusters with single descriptive rules.

CONCLUSION

Nugget discovery is an important descriptive data mining task. Nuggets are simple, understandable, strong, and potentially interesting patterns for a predefined class. A number of algorithms have been devised to find nuggets from databases. All rule algorithms can find all rules that underlie a database subject to certain constraints. The number of rules found might be high, and there may be limitations in their use for large databases. Heuristic methods can sample certain areas of the search space and may be adequate for large databases or for initial searches where an exhaustive search is considered unnecessary.

REFERENCES

Agrawal R., Imielinski T., & Swami A. (1993). Mining association rules between sets of items in large databases. *Proceedings of the ACM SIGMOD International Conference on Management of Data* (pp. 207-216). ACM Press.

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. *Proceedings of the 20th International Conference on VLDB* (pp. 487-499). Morgan Kaufmann.

Ali, S., Manganaris K., & Srikant R. (1997). Partial classification using association rules. *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining* (pp. 115-118). AAAI Press.

Bayardo, R., & Agrawal, R. (1999). Mining the most interesting rules. *Proceedings of the 5th International Conference on Knowledge Discovery and Data Mining* (pp. 145-153). AAAI Press.

Bayardo, R., & Agrawal, R. (2000). Constraint based rule mining in large, dense databases. *Data Mining and Knowledge Discovery Journal*, 217-240. Kluwer.

Brin, S., Rastogi, R., & Shim, K. (1999). Mining optimized gain rules for numeric attributes. *Proceedings of the 5th International Conference on Knowledge Discovery and Data Mining* (pp. 135-144). AAAI Press.

Deb, K., Agrawal, S., Pratap, A., & Meyarivan, T. (2000). A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. *Proceedings of the Parallel Problem Solving from Nature VI Conference, Lecture Notes in Computer Science, 1917,* 849-858. Springer.

de la Iglesia, B., Debuse, J.C.W., & Rayward-Smith, V.J. (1996). Discovering knowledge in commercial databases using modern heuristic techniques. *Proceedings of the Second International Conf. on Knowledge Discovery and Data Mining* (pp. 44-49). AAAI Press.

de la Iglesia, B., Philpott, M.S., Bagnall, T.J., & Rayward-Smith, V.J. (2003). Data mining rules using multi-objective evolutionary algorithms. *Proceedings of 2003 IEEE Congress on Evolutionary Computation*, Camberra, Australia.

de la Iglesia, B., & Rayward-Smith, V.J. (2000). The discovery of interesting nuggets using heuristic techniques. *Data mining: A heuristic approach* (pp. 72-96). Hershey, PA: Idea Group Publishing.

Han, J., & Pei, J. (2000b). Mining frequent patterns by pattern-growth. Methodology and implications. *SIGKDD Explorations, 2*(2), 14-20.

Han, J., Pei, J., & Yin, Y. (2000a). Mining frequent patterns without candidate generation. *Proc. 2000 ACM SIGMOD Int. Conf. on Management of Data* (pp. 1-12).

Horn, J., Nafpliotis, N., & Goldberg, D.E. (1994). A niched pareto genetic algorithm for multiobjective optimization. *Proceedings of the First IEEE Conference on Evolutionary Computation. IEEE World Congress on Computational Intelligence, 1,* 82-87.

Knowles, J., & Corne, D. (1999). The pareto archived evolution strategy: A new baseline algorithm for pareto multi-objective optimisation. *Proceedings of the Congress on Evolutionary Computation, 1,* 98-105.

Megiddo, N., & Srikant, R. (1998). Discovering predictive association rules. *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining* (pp. 274 – 278). AAAI Press.

Richards, G., & Rayward-Smith, V.J. (2001). Discovery of association rules in tabular data. *IEEE International Conference on Data Mining* (pp. 465-472). IEEE.

KEY TERMS

All Rules Search: An algorithm that finds all rules of a defined format that satisfy a defined set of constraints.

Classification Rules: Associations that occur in classification data between an antecedent containing a set of predicates expressing the values of particular attributes or features and a consequent expressing a class label. Classification rules represent concise descriptions of the target class.

Confidence: The proportion of records that belong to the target class from those that satisfy the antecedent of the rule. This is also often called the *accuracy* of the rule. An alternative definition is the proportion of records for which the rule makes the correct prediction of the target class.

Metaheuristic Techniques: Also known as modern heuristics, this refers to methods for the solution of complex optimisation problems that seek good (i.e., near optimal) solutions at a reasonable computational cost. Generally this type of method is based on the imitation of some natural process in the context of optimisation. The best-known modern heuristics are genetic algorithms, simulated annealing and tabu search.

MOEA: Multi-objective evolutionary algorithm. An evolutionary algorithm in which there are multiple objectives to be optimised simultaneously.

Multi-Objective Optimisation: The problem of finding a vector of decision variables that satisfies constraints and optimises a vector function whose elements represent the objective functions. There is often some conflict between the objectives to be optimised; hence the solution to such a problem tends to be a compromise solution that satisfies to some extent the objectives.

Nugget Discovery: The search for strong partial classification rules that are simple, understandable, novel and ultimately useful.

Support: The count of records that satisfy a particular predicate in classification rules.

Distance Education Success Factors

Cathy Cavanaugh
University of North Florida, USA

INTRODUCTION

While effective distance education has been practiced and studied for centuries, it has been in just the last decade that networked digital technology has been employed. Technologies and teaching techniques continue to evolve, and the options continue to expand, emphasizing the need for information that will assist distance education planners and participants in making decisions that will result in optimal learning experiences.

BACKGROUND

The process of developing and implementing effective distance education happens in an iterative cycle. Broadly considered, the three stages in the cycle are (1) procurement and preparation of the resources necessary to meet the distance education goals; (2) delivery of instruction using the best practices from education, business, and research; and (3) analysis of the results of distance education to gauge achievement of the goals. Each stage of the Resources-Practices-Results (RPR) cycle continually revisits lessons learned in the other stages and builds upon the successes realized in the other stages. Each stage requires participation of all stakeholders, including students, instructors, support and design professionals, administrators, and the community. The success factors discussed in each stage are based on decades of research and experience with learners from professions, higher education, and K-12 education (Barker, 1999; Bruce et al., 2000; Cavanaugh, 2001; Educational Development Associates, 1998; Fredericksen et al., 2000; Institute for Higher Education Policy, 2000; Mantyla, 1999).

THE RESOURCES PHASE OF THE RPR CYCLE

The resources required to sustain a quality distance education program exist to support students, faculty, and the program or institution toward achieving the goal of effective and appropriate learning. Responsive and flexible human resources, knowledge, skills, policies, procedures, and technical infrastructure enable quality practices and contribute to quality results. Procurement, development, and adaptation of resources are ongoing processes.

Institutional and Program Resources

To provide a vigorous quality distance education program, an institution begins with a policy that values distance education as an endeavor that integrates seamlessly with the institution's mission and goals. In creating a strategic plan, distance education administrators and instructors engage in continuous dialogue with a broad range of stakeholders in specifying quality benchmarks (Vaughan, 2000). The strategic plan is a financial and philosophical commitment that gives direction to personnel who make specific decisions regarding program implementation. It is a commitment to team support for distance educators and students, technology led by the program's current and future goals, and the development of program standards. Course developers and instructors need target standards to guide course design and delivery. As a partner to the standards, program review procedures must be developed, implemented, and revised frequently to ensure that all components of the program meet standards, and to ensure that the standards contribute to program goals.

Administration of a quality distance education program depends on clear and accurate communication to students. Qualified instructors and support staff must be recruited; they must be provided with development opportunities related to instruction, content knowledge, and technical skill; and they must receive feedback on their teaching. Qualitative input about student performance, satisfaction, and success is at least as important as quantitative data such as enrollment, costs, utilization of technology, and hiring rates. The elements of the comprehensive program evaluation process should be communicated to all stakeholders in advance, and the results should be reported completely and efficiently.

Faculty and Course Support Resources

Qualified and experienced distance education instructors are likely to have the desired attitudes and understanding of the distance education teaching and learning process. For faculty members to succeed in distance education, they need to be supported with accurate and complete information and training in order to develop their skills and understanding. Successful distance educators understand the distance learning environment and the options that exist for instruction.

In support of the design and delivery of quality courses, institutions are responsible for providing training and resources for instructors. Instructors need continual access to the physical resources and human support that will enable development of high quality teaching materials. The best distance learning courses use complete and up-to-date materials to increase the information literacy of students, while allowing opportunities for creative expression and mastery of concepts.

Student Support Resources

The focus of distance education is the students, whose work is made better when they receive well-designed instruction in a well-planned program. For students to maximize the time and effort they spend on their learning, they must minimize the time and effort they spend on solving non-academic problems and on seeking answers. Some students need hands-on technical training using the tools employed in courses and using general learning tools such as libraries and information archives. As students begin the work of learning, they need continual access to instructors, libraries, and other student resources. Students must have adequate access to resources appropriate to support their learning. The institution must assess the student's ability to succeed in online learning (SREB 2000).

Technical Resources

Even given the best plan, program, instructors, materials, and students, distance learning does not occur without the technology for delivery. Technology selection decisions involve all stakeholders. A technology plan guides decision makers in considering student outcomes, program goals, and technical feasibility (Council of Regional Accrediting Commissions, 2000). Support extends to all users of the technology for all facets of the learning process. Users require assistance with hardware and software uses.

Success Factors for the Resources Phase

- Institutional policy that values distance education
- Strategic plan for delivering distance education to students
- Stakeholder analysis to determine needs of graduates
- Financial commitment that gives direction regarding program implementation
- Team support for distance educators and students
- Appropriate technology infrastructure
- Program standards to guide course design and delivery
- Program review to ensure that all components of the program meet standards and to ensure that the standards contribute to program goals
- Effective communication of policies and expectations to students
- Student services: information, advising, orientation, and security
- Information privacy
- Qualified, experienced staff and faculty
- Community involvement in the program's goals, policies, and outcomes
- Information provided to faculty about teaching in the distance learning environment
- Instructor release time for course development
- Instructor training in distance education pedagogy and technology
- Course design and delivery assistance
- Well-designed and appropriate learning materials
- Student orientation and training
- Student access to learning resources and instructors
- Technical support for instructors and students
- Technology plan to communicate goals to all users

THE PRACTICES PHASE OF THE RPR CYCLE

With the right resources in place, the stage is set for dramatic distance learning performance. At this point, the spotlight shifts from the institution to the instructor. Quality distance teaching begins with the careful design of courses, materials, and learning activities. Next, the instructional practices employed during instruction will aim at developing independent learners with the ability to transfer their learning to novel situations. Throughout the course, effective communication and community building are essential foundations for all events.

Course Design Practices

Course design is a series of decisions regarding objectives and the most effective methods to ensure that students accomplish the objectives. Distance course design requires the methods to be effective in a technology-mediated environment. The requirements of the curriculum and the needs of the students lead the technological decisions in a well-designed course. A balance of comfort, control, and challenge can be difficult for distant instructors to achieve, and depends on psychosocial rather than academic strategies. The focus is on the needs of learners and the learning process, rather than on content. The quality service approach emphasizes the course structure and interactions in order to supply flexible scaffolding to learners as needed (Vaughan, 2000).

Communication Practices

Because learning is an interactive activity and constructed socially, a key to success lies in communication between students and others. A quality benchmark is to involve students in communication during 50% of the time they spend on the course. Frequent and active communication with the instructor, fellow students, or experts in the subject is essential in making students feel that they are part of the community of learners. Interaction in a distance learning course is most effective when it occurs through a variety of media, when it occurs with a variety of sources, and when it is integrated into the overall course design. Interactions are most effective when experienced within the context of other course activities. Communication in a course has the greatest value for students when it authentically approaches the kinds of communication students will experience beyond the course.

Instructional Practices

Successful distance educators understand that motivation is among the most important factors in promoting student learning. At the outset, instructors must clearly state the benefits of learning the course content to the student. The course activities should foster both knowledge construction and content understanding through active learning.

Success Factors for the Practices Phase

- Focus on content and students
- Relevant and important skills and knowledge addressed in courses
- Structured information presented in motivating context
- Social strategies to promote student comfort, control, and challenge
- Fast feedback from instructors to students
- Consistent design throughout each course
- Highly interactive activities for student engagement
- Authentic communication among students, instructors, and experts
- Course activities designed to maximize student motivation
- Activities focused on high-level cognitive skills
- Development of information literacy
- Development of applied technical skills

THE RESULTS PHASE OF THE RPR CYCLE

The only way to know whether a distance education program has achieved quality is to compare the program results to established quality benchmarks. Measures of quality are tied to institutional goals and account for the specific context of the program. To maintain success, a distance education program evaluation must account for institutional and instructional factors, as well as student factors. Evaluation of course and program results is a continual process that involves all stakeholders and requires a wide range of tools.

Assessing Learning

When experiencing quality learning, students shift roles from audience to actors as they acquire skills and display their abilities. The display of student abilities is the most important result of distance education. In the course of developing their abilities, successful students manage their learning by engaging in frequent self-assessment (Palloff & Pratt, 1999). Using varied assessment methods is a key to student assessment that gives an accurate picture of student abilities.

Program Review

Evaluation of course effectiveness by students is most useful when it is an ongoing feature of the course. Participation of students, instructors, and the institution is needed to determine the quality of the distance education program. Students should have the opportunity to offer feedback regarding their access to learning activities, course delivery, and technical support (Palloff & Pratt, 1999). The intended program outcomes must undergo review at the institutional level to ensure their clarity and their appropriateness to students who move into work or higher learning roles. Learning outcomes for distance education programs should be clear to instructors and students. Achievement of outcomes in specific courses should be observable and measurable against a known scale or set of criteria.

Accreditation

Accreditation gives an institution a seal of quality because educational standards have been met. A student who expects a distance education course to transfer to another school must be sure that a regionally accredited institution offers the course. Institutions with distance education programs approach accreditation in several ways. The accreditation process varies according to whether a distance education program is offered by a high school, a traditional institute of higher education, or dedicated distance education provider. The public needs information to help them distinguish among the accrediting bodies and the quality control they provide (Council of Regional Accrediting Commissions, 2000; Distance Education and Training Council, 2000; Web Based Education Commission, 2000).

Success Factors for the Results Phase

- Student independence developed through opportunities for self-assessment
- Peer review of student work as a professional experience
- Creation of student portfolios to showcase accomplishments
- Varied assessments for an accurate view of student abilities
- Open-ended assignments to increase thinking skills and reduce cheating
- Secure online testing
- Ongoing course evaluation by students
- Evaluation of program by students and faculty
- Review of program outcomes and components by all stakeholders
- Program accreditation

FUTURE TRENDS

Distance education technology and practice will continue to evolve, necessitating ongoing research into effective practice. The lines between distance and traditional education will blur as the points along the continuum become more finely refined in response to education customized to specific needs and contexts. Developers and learners will need more information to help them make the best choices.

CONCLUSION

A strength of distance education is its potential to focus the learning process on the student. Courses and programs that emphasize their focus on the student's strengths and needs will succeed in attracting students. In order to build their reputations and keep students, courses and programs must reach quality goals. High-quality distance education achievement is the outcome of the dedication of all constituents in the continual quest for the best possible resources, practices, and results. With an increased need for new career skills and improvement in delivery technology, distance education students will demand evidence of quality and authenticity in distance courses. It is imperative that distance education providers implement and review quality benchmarks regarding RPR in response to the needs of students, employers, and the community. Educational institutions must take the lead in developing and maintaining standards, and they must clearly communicate those standards to the public. When students benefit from an education program that meets their needs, the community benefits, as well.

REFERENCES

Barker, K. (1999). *Quality guidelines for technology-assisted distance education*. Washington, D.C: U.S. Department of Education Office of Learning Technologies.

Bruce, B., Fallon, C., & Horton, W. (2000). *Getting started with online learning*. Macromedia, Inc. Retrieved from *http://www.macromedia.com/learning/online_learning_guide.pdf*

Cavanaugh, C. 2001. The effectiveness of interactive distance education technologies in K-12 learning: A meta-analysis. *International Journal of Educational Telecommunications, 7*(1), 73-88.

Council of Regional Accrediting Commissions (2000). Statement of the regional accrediting commissions on the evaluation of electronically offered degree and certificate programs and guidelines for the evaluation of electronically offered degree and certificate programs. Retrieved from *http://www.ncacihe.org/resources/draftdistanceguide/*

Distance Education and Training Council (2000). Accreditation standards. Retrieved from *http://www.detc.org/content/ accredStandards.html*

Educational Development Associates (1998). *What quality distance learning courses for an institution?* Las Cruces, MN.

Fredericksen, E., Peltz, W., & Swan, K. (2000). Student satisfaction and perceived learning with online courses: Principles and examples from the SUNY learning network. *Journal of Asynchronous Learning Networks, 4*(2).

Institute for Higher Education Policy (2000). *Quality on the line: Benchmarks for success in Internet-based distance education*. Washington, D.C.

Johnstone, S. (2001). Does accreditation really mean accredited? *Syllabus, 14*(6), 22.

Kearsley, G. (2000). *Online education*. Belmont, CA: Wadsworth/Thomson Learning.

Mantyla, K. (1999). *Interactive distance learning exercises that really work*. Alexandria, VA: American Society for Training and Development.

Moore, M. (1989). *Effects of distance learning: A summary of the literature*. Washington, D.C: Office of Technical Assessment.

Moore, M., & Thompson, M. (with Quigley, A., Clark, G., & Goff, G.) (1990). *The effects of distance learning: A summary of the literature. Research monograph no. 2*. University Park, PA: The Pennsylvania State University, American Center for the Study of Distance Education.

Palloff, R., & Pratt, K. (1999). Bu*ilding learning communities in cyberspace*. San Francisco: Jossey-Bass Publishers.

Southern Regional Electronic Campus (2000). Principles of good practice. Retrieved from *http://www.srec.sreb.org/student/ srecinfo/principles/principles.html*

U.S. Department of Commerce (2000). Falling through the net: Toward digital inclusion. Washington, D.C.

Vaughan, M. (2000). Summary of quality issues in distance education. Retrieved from *http://www.lucent.com/cedl/sumqual.html*

Web-Based Education Commission (2000). The power of the Internet for learning. Washington, D.C: U.S. Department of Education.

Wilkes, C., & Burnham, B. (1991). Adult learner motivations and electronics distance education. *American Journal of Distance Education, 5*(1), 43-50.

KEY TERMS

Accreditation: Endorsement of quality performance by an outside agency.

Course Design: Decisions regarding objectives and the most effective methods of ensuring that students accomplish the objectives.

Distance Education: A teaching and learning system in which learning occurs in a time and/or place distant from the instructor.

Scaffolding: Cognitive and instructional supports for learning built into course design.

Standards: Benchmarks for quality performance.

Strategic Plan: A process by which quality will be improved and maintained in meeting the goals of the organization.

Technology Plan: Goals and benchmarks for technology systems and support in an organization.

Distributed Construction through Participatory Design

Panayiotis Zaphiris
City University, London, UK

Andrew Laghos
City University, London, UK

Giorgos Zacharia
MIT, USA

INTRODUCTION

This article presents an empirical study of an online learning community that collaborates with the course design team under the Participatory Design methodology. The different phases of this methodology were implemented using a four-stage participatory design process (Zaphiris & Zacharia, 2001):

1) building bridges with the intended users,
2) mapping user needs and suggestions to the system,
3) developing a prototype, and
4) integrating feedback and continuing the iteration.

We took advantage of the online and distributed nature of the student community to asynchronously design, implement, and study the course. We carried out the participatory design methodology by following the Distributed Constructionism pedagogical theory. During the different phases of the design process, we measured the student participation and the changes in their behavior when new design elements were introduced. We conclude that the most important element of this course was our discussion board, which helped us to promote student collaboration and the identification of the key community users who can participate productively in Participation Design activities.

There are three main sections to this article. After defining the key terminology, our Participatory Design approach is presented and its linkage to the Distributed Constructionism pedagogical theory specified. The article ends with ideas for future research and a set of conclusions.

BACKGROUND

Participatory Design

Participatory design (PD) refers to a design approach that focuses on the intended user of the service or product, and advocates the active involvement of users throughout the design process. PD is often termed as the "Scandinavian Challenge" (Bjerknes, Ehn & Kyng, 1987), since it was researchers from Scandinavian countries who pioneered its use in information systems development (Blomberg & Henderson, 1990; Bodker, Gronbaek & Kyng, 1993; Ehn, 1988).

User involvement is seen as critical both because users are the experts in the work practices supported by these technologies and because users ultimately will be the ones creating new practices in response to new technologies (Ellis, Jankowski & Jasper, 1998).

Blomberg and Henderson (1990) characterize the PD approach as advocating three tenets:

- The goal is to improve the quality of life, rather than demonstrate the capability of technology.
- The orientation is collaborative and cooperative rather than patriarchal.
- The process is iterative since PD values interactive evaluation to gather and integrate feedback from intended users.

By involving the users in the design process, the designers also gain knowledge of the work context, so that the new technology explicitly incorporates the values, history, and context of the work system (Ehn, 1988). The users take part in the entire design, implementation, and decision-making processes. Their involvement ensures that their activities are taken into account. Also by participating in the design, the users have a sense of "ownership" (Brown & Duguid, 2000), and the final system will have an increased user acceptance.

Distributed Constructionism

Simply put, Constructionism can be thought of as "learning-by-making" (Papert, 1991). It is both a theory of learning and a strategy for education (Papert, 1993). It focuses on the construction of a system rather than the information that will be used. The theory views computer networks as a new medium for construction, not as an information distribution channel. By embedding construction activities within a community, new ways for students to learn arise (Papert, 1993). Based on Piaget's constructivist theories, people don't get ideas, they make them. Learning is an active process where people construct knowledge from their experiences (Resnick, 1996).

Distributed Constructionism (Resnick, 1996) extends the Constructionism theory (Papert, 1991, 1993) to knowledge-building communities, where the online learning community (instead of one student) collaboratively constructs knowledge artifacts (Resnick, 1996). Distributed Constructionism asserts that "a particularly effective way for knowledge-building communities to form and grow is through collaborative activities that involve not just the exchange of information but the design and construction of meaningful artifacts" (Resnick, 1996). The three major activities of DC, within the context of an online learning community, are (Resnick, 1996):

- *Discussing Constructions:* Students discuss their constructions during the design, implementation, evaluation, and reiteration phases.
- *Sharing Constructions:* Web-based systems allow students to share their constructions and make them part of the shared knowledge.
- *Collaborating on Constructions:* The community can use online communication to collaborate on the design and development of the knowledge artifacts.
- *Distributed Constructionism:* Was enhanced among the users of the system, due to the iterative structure of our Participatory Design approach. Both the learning experience of the users and the content and functionality of the course itself were enhanced by the knowledge artifacts that were contributed to the course.

DESIGN APPROACH AND COURSE EVOLUTION

In this section, a case study applying the theories presented in the previous section is described.

Our focus has been to design an online learning community around a Computer Aided Language Learning (CALL) course. We believe that online interaction and community would increase users' motivation, commitment, and satisfaction with the online course. The Participatory Design methodology blends nicely with our goal. In particular, involving users during system development is thought to lead to greater user commitment, acceptance, usage, and satisfaction with the system (Baroudi, Olson & Ives, 1986).

In the design phase of the online course, we implemented PD as a four-step process (Ellis et al., 1998).

1) Building Bridges with the Intended Users

This step opened lines of communication between intended users and the development team. Specifically, this step involved the initialization of a multidisciplinary development team, identifying key groups of end users, and creating new methods of communication with users.

The development team in this project came out of the Kypros-Net (2002) group. Through their involvement in Cyprus and Greece related projects, they had longstanding relations with the intended user community.

The intended users have been especially people of the Greek Diaspora, travelers to Cyprus and Greece, and other Greek-speaking areas, and people who are generally interested in the Greek culture and language or languages in general. In our case, bridges with the intended users were build through our years of work at providing information about Cyprus through the Web pages of Kypros-Net, which primarily attracts the same user population as our intended Greek language online course.

2) Mapping User Needs and Suggestions to the System

Our conceptual design model has been "to design an effective online Greek language course that can build and sustain an online learning community of students." Based on the questions and inquiries we received from our users, we tried to match their needs (they wanted an easy-to-follow, both elementary and advanced course that they could attend at their own pace) with our conceptual design model.

3) Developing a Prototype

The project consists of 105 audio files, which were originally recorded as Radio lessons in Modern Greek for English speakers in the 1960s. The lessons were retrieved from the archives of the Cyprus Broadcasting Corporation, digitized in Real Audio 5.0 format, and published online through the course. Although, an optional textbook accompanied the original radio lessons, the online lessons were designed as a complete standalone course. We used several tools to assist students with the lessons, including an online English-Greek-English dictionary, a Greek spell checker, and a Web-based discussion board. The discussion boards served as the foundation for creating a community of online students and enhanced the learning experience with Distributed Constructionism.

4) Integrating Feedback and Continuing the Cycle

Feedback from our users and suggestions are continuously incorporated into our design through a series of additions and corrections. For example, we were asked to add an online notes section and to encode some files again because they were corrupted.

An important element in the participatory design methodology is the direct involvement of the users in all stages of the design process. We kept the users involved by participating in the discussion boards, and sharing with them design and development plans for the course.

The students of the audio courses included people with no knowledge of Greek language, bilingual members of the Greek Diaspora, as well as high-school professors of non-Greek language. These students created an open online community whose collaboration has boosted the learning experience of the whole community. The Web-based discussion board has proven to be the most constructive tool for the students' learning experience and the main source of feedback for the maintainers of the project. The experiences shared on the discussion board included tricks and tips on how to record the audio files, installation of Greek fonts, learning methodologies, and questions about the Greek language itself that arise from the lessons. The experienced users (some of them were retired teachers of foreign languages) had taken a lead role in the vast majority of the threads on the discussion board, answering most of the questions and encouraging the beginners to study the lessons further. They have also become the communication interface between the maintainers of the project and the community's requests.

At some point, the users started exchanging, through e-mail, written notes taken by the experienced users. They also used the discussion board to announce the availability of their personal notes. This behavior suggests that we must provide (and we did) the users with the capability to post their notes on the project's site.

The students had initiated Distributed Constructionism themselves. The course designers only provided technical support to facilitate the students' construction activities.

Discussing the Constructions

The course designers offered to provide publishing access to the online course to whomever wanted to contribute their material. Five users asked to be given access. Consequently, the five users, along with the two course designers, constituted the Participatory Design team. The PD team solicited contributions from the user community. The users suggested that they should transcribe the audio lessons, and compile verb lists, vocabulary lists, and grammatical notes for each lesson.

Sharing the Constructions

All the user contributions were shared in the common area of the online course. The user members of the PD team regularly posted notices on the discussion board about new material for the course. Also, other less active users chose to offer contributions for the course, by posting on the discussion board, rather than contacting the PD team. In their study on student involvement in designing an online foreign language course, Zaphiris and Zacharia (2002) state that the discussion board proved to be the most constructive tool for the students' learning experience and the main source of feedback for the maintainers of the project.

Collaborating on the Constructions

The user members of the PD team did not include any native speakers of Greek. They were all learning the language through the online course, and at that stage, they primarily depended on the audio lessons. In order to ensure the quality of the new material before publishing them on the course Web site, the user members of the PD team implemented a peer review process. A group of seven users, which included the five central user PD team members, reviewed and corrected all the material before posting them on the Web site. Each of the seven users offered to transcribe a number of the 105 audio lessons, and two of them also offered to provide verb and vocabulary lists. However, all materials were posted in a private area first, reviewed by the seven user members of the PD team, and posted on the Web site, when the five PD users were satisfied with the quality. Then the two PD course designers, who were both native Greek speakers, would go over the already published material and make sure it was correct. Most of the mistakes we had to correct were spelling mistakes, and we rarely had to correct grammatical mistakes.

Two months after the Distributed Constructionism effort started, students of the audio lessons managed to transcribe 81 out of the 105 lessons, correct them through the peer review process among themselves, and post them on the project's Web site. Six months later, the students had transcribed and peer reviewed all 105 lessons.

The knowledge constructed attracted significant user attention. The access to the audio lessons, the language tools, and the total access of the message board and the notes pages all kept increasing exponentially (Zaphiris & Zacharia, 2001). However, once we allowed our users to publish their own notes, there was a dramatic shift of traffic from the message board to the notes pages. In our view this is due to the fact that the users did not need to visit the discussion board any more to find out where other users had posted their notes. All the content was already aggregated and organized in a central location.

The course's popularity is apparent from the fact that the course currently has more than 25,000 registered students who actively participate in an online community that evolved around the course.

FUTURE TRENDS

Future work on this specific project will focus on a non-virtual, face-to-face participatory design team. Like the previous PD team, key stakeholders (teachers of Greek in the Diaspora, students, administrators, and designers) will work together, participating and interacting throughout the whole iterative design process. They will once again collaborate on the content and functionality development, peer review, and publish content contributions.

We believe that by encouraging the active involvement of the users, the product developed will be more enjoyable, more usable, and most importantly, more catered to their specific needs and requirements.

Also we anticipate that the expected benefit of this face-to-face PD team versus the virtual PD team will be that everyone involved will feel more like a team and have a stronger relationship with each other. Since the PD team will be a face-to-face one, communication will be better and there will be fewer misunderstandings or misinterpretations, and finally the collaboration results should be more immediate, and the final product more usable and acceptable by all the stakeholders.

From our analysis of existing literature, we observed that there is a need for additional research in areas like ethnography in participatory design and the application of our proposed methodology to new domains. As new delivery e-learning technologies are constantly emerging, research into Distributed Constructionism with the latest technology also remains important. Finally, evaluations of case studies of PD and DC will be very useful to case-specific applications of the theories.

CONCLUSION

By facilitating Distributed Constructionism in the iteration phase of a Participatory Design methodology, we enhanced the learning experience in our Web-based training. A questionnaire evaluation (Zaphiris & Zacharia, 2001) shows that the end system received high usability ratings from the users. Therefore, Distributed Constructionism enhanced the learning experience of both the PD team and the more passive users.

The students who participated actively in the design of the course also played a central role in the discussion board, answering other students' language questions, helping students to overcome technical problems, and helping them to find other resources to enhance their learning of the Greek language. These observations are with agreement the underlying goals of Participatory Design, which was an integral part of the development of this specific course.

Furthermore, the results of the analysis of the user questionnaire and the server logs shows that the final product (the course) meets—to a very large extent—the expectations and needs of the whole user population of this specific course. We believe that the direct involvement of the users in the development of the course helped in designing a more usable course that enhanced the learning of our users, and provided them with an enjoyable and rewarding experience.

REFERENCES

Baroudi, Olson & Ives (1986). An empirical study of the impact of user involvement on system usage and information satisfaction. *CACM, 29*(3), 232-238.

Bjerknes, G., Ehn, P. & Kyng, M. (Eds.). (1987). *Computers and democracy — A Scandinavian challenge*. Aldershot: Gower.

Blomberg, J.L. & Henderson, A. (1990). Reflections on participatory design: Lessons from the Trillium experience. *Proceedings of CHI'90* (pp. 353-359). Seattle, WA: ACM Press.

Bodker, S., Gronbaek, K. & Kyng, M. (1993). Cooperative design: Techniques and experience from the Scandinavian scene. In D. Schuler & A. Namioka (Eds.), *Participatory design: Principles and practices* (pp. 157-175). Hillsdale, NJ: Lawrence Erlbaum.

Brown, J.S. & Duguid, P. (2000). *The social life of information*. Boston: Harvard Business School Press.

Ehn, P. (1988). *Work-oriented design of computer artifacts*. Hillsdale, NJ: Lawrence Erlbaum.

Ellis, R.D., Jankowski, T.B. & Jasper, J.E. (1998). Participatory design of an Internet-based information system for aging services professionals. *The Gerontologist, 38*(6), 743-748.

Kypros-Net Inc. (2002). *The world of Cyprus*. Retrieved December 4, 2002, from *www.kypros.org*

Nielsen, J. (1993). *Usability engineering*. Chestnut Hill, MA: AP Professional.

Papert, S. (1991). Situating construction. In I. Harel & S. Papert (Eds.), *Constructionism* (pp. 1-12). Norwood, NJ: Ablex Publishing.

Papert, S. (1993). *The children's machine: Rethinking school in the age of the computer*. New York: Basic Books.

Perlman, G. (1999). *Web-based user interface evaluation with questionnaires*. Retrieved December 4, 2002, from *www.acm.org/~perlman/question.html*

Resnick, M. (1996). *Distributed Constructionism*. Retrieved December 4, 2002, from *Web.media.mit.edu/~mres/papers/Distrib-Construc/Distrib-Construc.html*.

Zaphiris, P. & Zacharia, G. (2002, September). Student involvement in designing an online foreign language course. *Proceedings of the British HCI Conference* (Volume 2, pp. 170-173), London.

Zaphiris, P. & Zacharia, G. (2001, October 23-27). User-centered evaluation of an online modern Greek language course. *Proceedings of the WebNet 2001 Conference*, Orlando, FL.

KEY TERMS

Computer Aided Language Learning (CALL): The use of computers in learning a language.

Distributed Constructionism (DC): An extension of the Constructionism theory to knowledge-building communities, where the online learning community (instead of one student) collaboratively constructs knowledge artifacts.

Ethnography: The branch of anthropology that provides scientific description of individual human societies.

Human-Computer Interaction: The study, planning, and design of what happens when humans and computers work together.

Participatory Design (PD): A design approach that focuses on the intended user of a service or product, and advocates the active involvement of users throughout the design process.

Pedagogy: The activities of education or instructing or teaching.

User-Centered Design: Puts the user into the center of the software design process.

Web-Based Training (WBT): Anywhere, anytime instruction delivered over the Internet, or a corporate intranet to learners.

D

Distributed Recommender Systems for Internet Commerce

Badrul M. Sarwar
University of Minnesota, USA

Joseph A. Konstan
University of Minnesota, USA

John T. Riedl
University of Minnesota, USA

INTRODUCTION

Recommender systems (RSs) present an alternative information-evaluation approach based on the judgements of human beings (Resnick & Varian, 1997). It attempts to automate the word-of-mouth recommendations that we regularly receive from family, friends, and colleagues. In essence, it allows everyone to serve as a critic. This inclusiveness circumvents the scalability problems of individual critics—with millions of readers it becomes possible to review millions of books. At the same time it raises the question of how to reconcile the many and varied opinions of a large community of ordinary people. Recommender systems address this question through the use of different algorithms: *nearest-neighbor algorithms* (Resnick, Iacovou, Suchak, Bergstrom, & Riedl, 1994; Shardanand et al., 1994), *item-based algorithms* (Sarwar, Karypis, Konstan, & Riedl, 2001), *clustering algorithms* (Ungar & Foster, 1998), and *probabilistic* and *rule-based learning algorithms* (Breese, Heckerman, & Kadie, 1998), to name but a few. The nearest-neighbor-algorithm-based recommender systems, which are often referred to as *collaborative filtering* (CF) *systems* in research literature (Maltz & Ehrlich, 1995), are the most widely used recommender systems in practice. A typical CF-based recommender system maintains a database containing the *ratings* that each customer has given to each product that customer has evaluated. For each customer in the system, the recommendation engine computes a *neighborhood* of other customers with similar opinions. To evaluate other products for this customer, the system forms a normalized and weighted average of the opinions of the customer's neighbors.

The emergence of the Internet and its far-reaching deployment is changing the way commerce is done. Economists and commerce experts are now suggesting companies to shift from the old world of mass production characterized by "...standardized products, homogenous markets, and long product life and development cycles..." to the new world where "...variety and customization supplant standardized products" (Pine, 1993). In his famous book *Mass Customization*, Joe Pine also suggests that building one product is simply not adequate anymore. Companies need to be able to develop multiple products that meet the multiple needs of multiple customers. The movement toward Internet commerce has allowed companies to provide customers with more options. However, in expanding to this new level of customization, businesses increase the amount of information that customers must process before they are able to select which items meet their needs. Traditional data-analysis techniques are often not sufficient to process this huge amount of data in real time as needed by the Internet sites. Recommender systems, by providing a "personalized" interface to each customer, can potentially automate personalization on the Internet. Personalization to this extent is one way to realize Pine's ideas of the "new world order" of Internet commerce.

As discussed by Schafer, Konstan, and Riedl (1999), recommender systems can help Internet commerce sites boost their business in several ways. First, by providing personalized recommendations on various products, they help convert browsers into buyers. Visitors to a Web site often look over the site without ever purchasing anything. Recommender systems can help customers find products they wish to purchase and can potentially increase sales. Second, recommender systems improve cross-sell by suggesting additional products for the customer to purchase. If the recommendations are good, the average order size should increase. For instance, a site might recommend additional products in the checkout process based on those products already in the shopping cart. Third, recommender systems help capture customer loyalty. Internet commerce is getting competitive day by day. Freed from large capital investment and recurring costs for physical storefronts, an unprecedented number of businesses are using the Internet to market and sell goods, potentially creating a vicious price war. As a consequence, gaining customer loyalty becomes an essential strategy for businesses to survive on the Internet nowadays (Reichheld, 1993; Reichheld & Sesser, 1990). Recommender systems improve loyalty by creating a value-added relationship between the site and the customer. Sites invest in learning about their users, use recommender systems to operationalize that learning, and present custom interfaces that match customer needs. Customers repay these sites by returning to the ones that best match their needs.

RECOMMENDATION INTERFACES

There is more than one way to display recommendations to a customer. The method selected may well depend on how the Internet commerce site wants the customer to use the recommendation. In the following we will examine several recommendation interfaces and how each assists the site in making money. While some of these methods have their roots in traditional commerce, each of them draws upon the strengths of ubiquitous Internet to provide more powerful recommendations. We present these interfaces in Table 1 (Sarwar, Konstan, & Riedl, 2001; Schafer et al., 1999).

DISTRIBUTED RECOMMENDER SYSTEMS

In the past, participating in commerce meant that the consumer had to travel to the location of the store from which he or she wanted to purchase a product. Today, participating in commerce may be as easy as moving a mouse and typing a few keystrokes. In the future, participating in commerce will become even easier. With the introduction of new wireless devices that enable commerce on the palmtop, consumers will be able to shop from wherever they happen to be. The result of technical improvements like wireless Web browsers is that consumers will come to expect the same shopping experience when travelling as they currently receive when directly connected to the Internet. For instance, a customer who uses a restaurant recommendation service in his or her hometown will expect to be able to use the same restaurant recommendation service while traveling. In fact, the recommendations will be even more valuable on the road since the customer will know less about what is available.

Creating good recommendations for a traveling consumer is challenging, though, especially for the leading recommendation technology, collaborative filtering. Many Internet commerce sites, including several of the largest ones, are now using CF recommender systems as part of their personalization effort. As the largest Internet commerce sites attempt to use these systems, however, they are discovering the drawbacks of today's centralized systems. While a centralized architecture may be useful for smaller applications, there are several key drawbacks to centralization as shown in Table 2. Together, these issues point toward distributed architectures for collaborative filtering (Sarwar, Konstan et al., 2001).

There are four fundamemtal components that are to be considered for designing a framework for distributed recommender system:

1) *Products*: Are the products of local or global interest?

Table 1. Recommendation interfaces for Internet commerce

Browsing: Recommended browsing helps the users narrow down their choices and makes them feel more confident about their decision buying decision by providing organized access to the recommendations.
Similar Item: Recommender systems display items based on the item in which a customer has shown interest. In so doing, sites increase customers' exposure to their product line and are ideally able to sell more items per order.
E-mail: Recommendations can also be delivered directly to customers through e-mail. This extension of the traditional direct-mail campaign is also expected to generate more sales.
Text Comments: Sites can provide customers with recommendations based directly on the text comments of other customers.
Average Rating: Recommender systems show the average ratings for particular items.
Top *N*: Once each site has learned details about a customer's likes and dislikes, each is able to provide the customer with a personalized list of the top-*N* unrated items for that customer. It helps sites convert browsers into buyers as well as helps customers in making a decision about a product that they originally held in doubt.

Table 2. Drawbacks of centralized recommender systems

Scalability: The largest Internet commerce sites have millions of customers and up to millions of products. Recommendation generation is becoming an increasingly computation-hungry process.

Fault Tolerance: This provides a central point of failure that is unacceptable.

Performance: Global businesses cannot deliver reliable, low-latency performance around the world from a single, potentially overloaded recomender system.

Security and Privacy: Consumers are becoming increasingly wary of businesses that attempt to "own" data about them.

2) *Ratings*: Are the ratings being used from local customers or all customers?
3) *Neighborhoods*: Are the neighborhoods formed by using local customers or all customers?
4) *Predictions*: Are the products for which predictions are formed local or are they all avialable products?

However, from an algorithmic perspective, the two most important of these components are the neighborhood-formation and prediction processes since these decide what data are used by the recommender system, where the data are stored, and how that data are used. Based on these two dimensions that can be either local or global, we find four potential models that present a taxonomy of distributed recommender-system applications. These are shown in Table 3 (Sarwar, Konstan, et al., 2001).

Using this taxonomy, we present three implementation frameworks for different types of distributed collaborative filtering. At the heart of all three of the frameworks are protocols for storage, exchange, and maintenance of customer preference profiles. These are listed below.

- **Local Profile Model:** The profile database is stored locally. A customer requests recommendations from recommender systems that retrieve the profile from the customer's client and computes prediction. The main advantage of this model is its simplicity. It also ensures security and privacy as the profile database is maintained solely at the customer site. It also provides the fastest recommendations compared to other models. However, the communication of profile data to a remote recommendation server is cumbersome.
- **Central Profile Model:** The profile data is stored in a central database. This solves the profile exchange problem, but the central profile storage can be a performance bottleneck and a security and privacy hazard. Also, the latency is higher due to increased communication with the profile storage.
- **Geographically-Distributed Profile Model:** The profile database is distributed on a number of interconnected but geographically distributed profile servers. Due to the fully distributed design, this model circumvents the bottleneck problem of the central profile model and provides a smoother way of sharing profile data. However, partitioning of the profile data according to the geographical distribution as well as consistency management can be a challenging task.

Table 3. Taxonomy of distributed recommender system

Global Neighborhood and Global Prediction: This is the most common application of RS on the Internet. All customers have opinions about all products and interest in all products. This model is likely to continue to be very successful in the Internet commerce domain.

Global Neighborhood and Local Prediction: This is most useful when the customer is traveling. Recommendations for local items will be produced based on a global opinion. Large cities with active Web populations are likely to have many entertainment venues that fit these criteria.

Local Neighborhood and Global Prediction: This is potentially useful for travelers. It automates the typical word-of-mouth recommendation. Since the neighborhoods are formed based on local ratings, the relationships between the neighbors are likely to be stronger, and hence these systems are expected to generate better recommendations

Local Neighborhood and Local Prediction: This is similar to the global-global case, but the items used for neighborhood formation and prediction generation are of local interest only.

CONCLUSION

Recent marketing trends suggest that personalized service is highly valued by the respective customers (Peppers & Rogers, 1997). One important component of personalized service is capturing and using customer preferences to treat each customer according to his or her taste. Recommender systems are capable of providing such value-added services for customers. These systems provide an Internet commerce site with most of the fundamental methods for achieving goals of mass customization listed in Joe Pine's book (Pine, 1993). Recommender systems achieve these goals by providing a customized service that enables Internet commerce sites to sell their largely commodity products more efficiently by creating "customizable products and services" and by directly "customizing the point of delivery" for the Internet commerce site. Recommender systems will become increasingly important in the future as modern businesses are increasingly focused on the long-term value of customers to the business (Peppers & Rogers). These systems are expected to be used in the future to "predict demand for products," enabling earlier communication to back the supply chain. Currently, most recommender systems are centralized and are run and controlled by a single Internet commerce merchant. Recommender systems will become more and more distributed over time, driven by user needs for less effort and better service. In this article, we have introduced the concepts of recommender systems and how they are being used to make money by capturing customer loyalty. We then presented a taxonomy of the application space for distributed recommender systems. We studied several frameworks for implementing distributed recommender systems and presented three different implementation models.

REFERENCES

Breese, J., Heckerman, D., & Kadie, C. (1998). Empirical analysis of predictive algorithms for collaborative filtering. *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, Madison, WI.

Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R., & Riedl, J. (1997). GroupLens: Applying collaborative filtering to Usenet news. *Communications of the ACM, 40*(3).

Maltz, D., & Ehrlich, K. (1995). Pointing the way: Active collaborative filtering. *Proceedings of ACM CHI '95*.

Peppers, D., & Rogers, M. (1997). *The one to one future: Building relationships one customer at a time.* Bantam Doubleday Dell Publishing.

Pine, J., II. (1993). *Mass customization*. Boston: Harvard Business School.

Pine, J., II, Peppers, D., & Rogers, M. (1995). Do you want to keep your customers forever? *Harvard Business School Review, 2*, 103-114.

Reichheld, F. (1993). Loyalty-based management. *Harvard Business School Review, 2*, 64-73.

Reichheld, F., & Sasser, W., Jr. (1990). Zero defections: Quality comes to services. *Harvard Business School Review, 5*, 105-111.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). GroupLens: An open architecture for collaborative filtering of Netnews. *Proceedings of CSCW '94*, Chapel Hill, NC.

Resnick, P., & Varian, H. R. (1997). Recommender systems. *CACM, 40*(3), 56-58.

Sarwar, B. M., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. *Proceedings of ACM WWW10 Conference*, Hong Kong, China.

Sarwar, B. M., Konstan, J., & Riedl, J. (2001). Distributed recommender systems: New opportunities in Internet commerce. In S. M. Rahman & R. J. Bignall (Eds.), *Internet commerce and software agents: Cases, technologies and opportunities.* Hershey, PA: Idea Group Publishing.

Schafer, J. B., Konstan, J., & Riedl, J. (1999). Recommender systems in e-commerce. *Proceedings of ACM Conference on Electronic Commerce (EC-99)*, Denver, CO.

Shardanand, U., & Maes, P. (1995). Social information filtering: Algorithms for automating "word of mouth." *Proceedings of ACM CHI '95*, Denver, CO.

Ungar, L. H., & Foster, D. P. (1998). Clustering methods for collaborative filtering. *Proceedings of the AAAI Workshop on Recommendation Systems*, Madison, WI.

KEY TERMS

Collaborative Filtering: A recommendation technique that uses k-nearest neighbor learning algorithm. Variations include user-based and item-based algorithms

Neighborhood: A list of like-minded customers that are discovered by looking into the ratings database and by using a similarity measure such as correlation

Prediction: An output interface of a recommender system. Given a customer and a product, the system generates a score expressing the predicted liking in a scale the same as the ratings.

Profile Database: Collection of customer preference data, possibly along with the customer's demographic information

Rating: An opinion of a customer about a product used by recommendation algorithms. This is usually expressed in a Likert scale (e.g., from 1 to 5).

Recommendation: An output interface of a recommender system. Given a customer, the system recommends a list of products the customer will be interested in (e.g., top-10 products).

Recommender System: An information-filtering technique based on machine learning and/or statistical methods. Widely used in Internet commerce nowadays

Dot-Coming SMEs in Singapore for the New Economy

Leo Tan Wee Hin
Singapore National Academy of Science, Singapore and Nanyang Technological University, Singapore

Ramesh Subramaniam
Singapore National Academy of Science, Singapore and Nanyang Technological University, Singapore

INTRODUCTION

Small- and medium-sized enterprises (SMEs) play a very significant role in the economies of all countries. They outnumber big corporations, and are conspicuous by their ubiquity in all sectors of society. The contribution of SMEs to a country's Gross Domestic Product is high by virtue of the large number of people they employ and the diversity of services that they offer (Unctad Secretariat, 2002). SMEs also offer a starting platform for entrepreneurs to realize their aspirations.

Globalization and the Internet have started to transform the global economic landscape. Capital mobility and the connectivity provided by the Internet for business processes are reengineering the contours of international trade significantly and giving rise to an international common market. Traditional structures of businesses are starting to atrophy to varying extents and new business models are emerging. These are affecting the competitiveness of nations. Since these affect the employment of workers (Tan and Subramaniam, 2001), they therefore impact on all business sectors. Those business outfits not able to compete in the emerging economic order face the possibility of being sidelined by the forces of globalization and Internet-based business practices. In particular, SMEs are vulnerable as they generally operate in a local environment. To compete in the new economy, these businesses need to reformat their work processes as well as adopt e-commerce.

BACKGROUND

There is no unambiguous definition in the literature of what e-commerce means, as the concept is still evolving and not many countries have embraced it in totality. Many countries are in different stages of economic development, and those that have warmed towards e-commerce have adopted it to varying extents. Broadly speaking, e-commerce refers to the use of the Internet to do business with suppliers, clients, and others in the business value chain (Zwass, 2003).

The definition of SMEs also varies among countries. In Singapore, which is the focus of this article, the following criteria is used to define an SME (http://www.singapore-sme.com):

- there must be local equity stake of at least 30 % in the company
- the net book value of the company's fixed assets, that is, factory building, machinery, and equipment, must not exceed S$15 million
- the workforce must not exceed 200 for non-manufacturing enterprises

The Internet offers a leveling field for SMEs to compete globally and overcome geographical barriers and other limitations. Whilst SMEs in the USA and Western Europe have adopted e-commerce significantly (Euro News, 2002; Fife and Pereira, 2001; Quayle, 2001), the same cannot be said of other countries. Among countries in Asia, only Singapore, Taiwan, Hong Kong, Australia, Japan and South Korea have good e-commerce infrastructures (Wong, 2001; OECD, 2000). In South-East Asia, Singapore is the only country with an advanced e-commerce infrastructure (Tan and Subramaniam, 2004a; 2004b).

This article examines the use of e-commerce as a business enabler for SMEs in Singapore. The rationale for selecting Singapore as a case study is as follows:

(a) Singapore was among the earliest adopters of e-commerce in the world, and thus its experiences can offer useful lessons for other countries.
(b) The use of state intervention strategies in getting SMEs to adopt e-commerce is a key feature of Singapore's approach as compared to other countries, where it is generally private sector-driven.
(c) Recognizing the differing levels of economic development in countries in the South-east Asian region, of which Singapore is a part, and the fact that they are on different learning curves and achievement bands as regards e-business initiatives, exploring the state of e-commerce in SMEs in a particular country is more meaningful and revealing.

HOW SMES ARE ENABLED IN SINGAPORE BY E-COMMERCE

Over 100,000 businesses in Singapore are SMEs – this represents 90% of all business establishments. They employ over 50 % of the workforce and generate about 30 % of the total value-added to the economy.

A good measure of the willingness of SMEs to try out e-commerce initiatives can be gauged from the extent to which the necessary infrastructure and frameworks for these are in place. The infrastructure relates to the presence of an advanced telecommunications network. A good infrastructure creates economic value for SMEs to ride on, as well as decreases their operating costs. The frameworks refer to the set of policies, e-payment protocols, and other enabling services that help companies use e-commerce in their businesses. These are elaborated upon in this section for Singapore.

Digital Telecommunications Network

An advanced telecommunications network leveraging three broadband platforms—Asymmetric Digital Subscriber Line, Hybrid Fibre Coaxial Cable Modem, and Asynchronous Transfer Mode—has been in operation since 2000. The technical aspects of these services have been addressed in detail by Tan and Subramaniam (2000, 2001). The maturation of the broadband market has led to Singapore being among the top 10 broadband economies in the world (Asia Pacific Telecommunication Indicators, 2002).

The cost of Internet access has been brought down drastically over the years through generous government subsidies for bandwidth. This has allowed telcos to pass on the savings to consumers and businesses, and has been an important factor in Singapore: having an Internet penetration rate of over 90 % (Tan and Subramaniam, 2003).

That Singapore is effectively networked for the new economy is shown by the fact that it has been ranked third, ahead of Finland and the USA, for networked readiness in the new economy (Dutta and Jain, 2003).

Frameworks for Electronic Commerce

Singapore was among the early adopters of e-commerce in the world. A comprehensive range of initiatives and frameworks (http://www.ec.gov.sg) has been put in place to allow businesses to capitalize on the potential that e-commerce presents. Fine-tuning of the frameworks on the basis of accumulated experiences in the local market and cognizance of best practices in overseas countries have been done on an ongoing basis. Some of the more important initiatives that have been put in place are:

(a) Alternative Payment Modes

For e-commerce to be entrenched in the business landscape, multiple payment systems of a secure nature are indispensable. Four payment channels are in common use in Singapore:

- A Secure Electronic Transaction (SET) system has been in place since 1997 to service credit card transactions over the Internet, this being a world-first
- NETSCash, the digital equivalent of the cash card, is available to service low-value purchases (S$0.01—S$500.00)
- Debiting via Internet banking has been available since 1997
- Electronic inter-bank payments via NETS Financial Data Interchange

(b) Use of digital certificates

Security of financial transactions on the Internet is important if businesses are to be encouraged to go online. In this context, two government-linked Certification Authorities (CA) have been established, and they have been empowered to issue digital certificates to businesses that require cross-checking of user credentials as well as assurance of high network security. On the certificates are digitally embedded details of the CA, the expiry date, encoded information about the user's public key, and details of the account.

(c) Application Service Providers

Application Service Providers (ASP) in the e-commerce industry provide services such as human resource management, sales and marketing, customer service, and even collaborative working with businesses. Their emergence have allowed SMEs to concentrate on their domain expertise whilst ancillary functions are outsourced to these ASPs.

(d) Business-to-Business Security Service Providers

For e-businesses that are hesitant to set up their own e-commerce infrastructure because of cost constraints and uncertainty of the business model, B2B service providers have emerged in the market. They provide a secure e-commerce network between companies, thus saving them considerable manpower and infrastructure costs as well as recurrent expenses.

(e) Legal and Policy Frameworks

In Singapore, there is recognition that for the support of businesses in the digital economy, the government must take the lead in laying down the administrative frameworks relating to regulatory and legal issues. These frameworks must not only be transparent and pro-business, they must also be seen to not stymie the growth of business enterprises or hamper innovation.

Currently, the following frameworks are in place – Intellectual Property Rights Act, Electronic Transactions Act, use of e-applications and e-licenses for businesses, commerce code for e-commerce transactions, Public Key Infrastructure, Computer Misuse Act, Electronic Evidence Act, and guidelines on tax issues in relation to e-commerce.

(f) E-Government

An e-government is in place to service the needs of the public and businesses. As of December 2001, all government services that need to be accessed by the public and businesses have been placed online. Some of these include: registering a business, tendering for government contracts, and e-filing of income taxes. Through these initiatives, the government aims to provide the public and businesses more convenience and prompt services 24/7. A key feature of the e-government is the Government-to-Business (G2B) portal, which streamlines various processes for businesses. (Table 1).

Of interest to note is the government's TradeNet initiative, which is an Internet-based system for processing 99% of all trade declarations made by businesses since 2002. This paperless processing system provides savings of about S$2.8 billion annually. Fisher (2000) has noted that enhancing business productivity through similar practices can boost GDPs of many Asian countries by 5-12% in the long run.

DISCUSSION

With an advanced infrastructure providing effective network connectivity and affordable access, and the frameworks promoting trust and user confidence, novel ways of doing business are emerging. It is not surprising that the take-up rate of e-commerce among SMEs has been very significant (Tan and Subramaniam, 2004b). For the period from January 2000 to March 2001, the growth rate was about 1800%, constituting nearly 20% of the total number of SMEs (Table 2) (Kwong & Lim, 2001).

Significant emphasis is also being placed on state intervention strategies to get more SMEs to embrace e-commerce. A key national agency, Spring Singapore, has been entrusted with this important responsibility. Some of the initiatives that it has introduced are addressed now.

Recognizing the high cost of setting up the necessary e-commerce infrastructure for a company (Koh, 2000), generic jump-start packages that enhance the online capabilities of businesses have been introduced. In contrast to customized solutions, this strategy has been found to be effective in getting more SMEs to embrace e-commerce since the package is less expensive and offers a quicker approach to joining the mainstream of e-commerce (Kwong and Lim, 2001). Further support for the efficacy of such jump start packages for implementing e-business solutions has been reiterated in another independent survey (Kwong and Lim, 2001) that further revealed the following findings:

- 21% of SMEs have e-commerce capabilities.
- About 50% of SMEs with e-commerce capabilities have performed some online transactions.

Table 1. Comparison of some services on civil service and e-government platforms

Item	**Civil Service Procedure**	**e-Government Procedure**
Submitting business plans	Entails separate submission of documents to 12 departments.	One-stop submission at G2B portal. Savings of S$450.
Incorporating new company	S$1,200 to S$35,000, depending on size of company. Approval time of 2 days.	Online incorporation at a cost of S$300. Approval time of 2 hours.
Opening new entertainment outlet	2 months for approval.	2 weeks for approval.

Source: http://www.egov.gov.sg

Table 2. E-commerce adoption rate among SMEs in Singapore

Year	Number of SMEs
Jan 2000	800
Mar 2000	4,000
June 2000	10,000
Sep 2000	16,000
Dec 2000	17,500
Mar 2001	19,000

Source: Interpolated from data presented in Kwong and Lim (2001)

- About 60% of SMEs believe that the Internet is an important tool for business.

Another important initiative is the establishment of industry-specific portals on the Internet. These portals are customized for the special needs of a particular industry. Besides catalyzing the evolution of online trading communities among the SMEs, they also facilitate a myriad of transactions. Examples of these portals include

http://www.Asiafurniturehub.com and http://www.kafe123.com. The portal operators also offer online links to overseas trading partners, thus opening up more opportunities for SMEs.

The government recognizes that a targeted and focused approach on SMEs would encourage more of them to embrace e-commerce. As it is, the sectoral mix of SMEs in Singapore indicates that many are in the brick-and-mortar business and only a small percentage are technology-based or are big enough to embark on e-commerce initiatives. Moreover, with Singapore's city/state structure, high population density, and presence of retail outlets and convenience stores within short traveling distances, the penchant for online purchases by consumers, and thus this aspect of promoting e-commerce, is not pronounced.

In the context of the foregoing, a national plan to boost the capabilities of SMEs in the new economy was instituted in 2000. Called SME 21, the ten-year strategic plan aims to raise the status of SMEs in a focused and selective way so that more of them can ride on the advanced telecommunications infrastructure and other frameworks set in place (http://www.singapore-sme.com). More specifically, the following initiatives have been put in place:

(i) Develop productive sectors in the SMEs

This plan aims to pivot SMEs up the value chain in the local economy through a process of restructuring, revitalizing, and upgrading. The aim is to double the productivity of the retail sector from S$28,000 per worker to S$56,000 per worker.

(ii) Groom innovative high growth SMEs

This plan aims to triple the number of local SMEs with annual sales turnover of S$10 million and above from 2,000 to 6,000 companies. That is, it aims to ensure that 1 in 15 companies will go on to become large corporations.

(iii) Institute a knowledge-based and pro-enterprise environment

This plan aims to create conditions where appropriate mindsets for business are inculcated, technopreneurship and innovation are encouraged, and obstacles to growth removed. In this way, it is expected to quadruple the numbers of SMEs transacting online from 8,000 to 32,000 companies. That is, one third of local SMEs will be plugged into the global economy.

To accelerate the realization of the above targets, schemes to provide grants and subsidies to SMEs whose corporate focuses are aligned with national objectives have been put in place.

Recognizing that the logistics and transportation sectors contribute significantly (8%) to the GDP, a three-year plan costing S$12 million was rolled out by the government in 2000 to enable SMEs in these industries to use e-commerce.

Promising SMEs are also being incentivised through the provision of grants and other assistance packages to morph from brick-and-mortar enterprises to click-and-mortar enterprises. As of 2001, a total of 165 SMEs have reinvented themselves in this way (Kwong and Lim, 2001).

To increase e-commerce practices among SMEs, focus has also been on large enterprises outside the SME fold. As they adopt e-commerce, SMEs, by virtue of their linkages with these enterprises, will be forced to adopt these practices in order to service them. For example, B2B turnover in Singapore has shown rapid growth (Tables 3 and 4), indicating that many big businesses have entered the e-commerce ecosystem (Ho, 2002). This is having implications for SMEs to embrace new economy paradigms.

FUTURE TRENDS

E-commerce is poised to be a key enabler for many SMEs to be competitive in the emerging new economy. The increase in penetration rates of the Internet, as well as the extent to which business practices are being reengineered to suit the contours of the competitive global economic landscape, means that those companies not in the economic value chain face the prospect of being side-lined if they are not responsive to the new ways of doing busi-

Table 3. Business-to-Business sales value in Singapore

Year	Value (S$ million)
1998	5,671
1999	40,425
2000	92,701
2002	109,460

Source: Ho, 2002

ness. For those which are in the brick-and-mortar business as well as those which do not have the necessary sales turnover, it would be a challenge to dot-com them. Further research is needed to evolve cost-effective e-business models for some of these SMEs to be brought into the e-commerce ecosystem.

CONCLUSION

The telecommunications infrastructure as well as the regulatory and policy frameworks for the support of e-commerce in Singapore are of a high standard. Whilst there is still quite some way to go for SMEs to embrace e-commerce, the Singapore experience shows that state intermediation strategies, especially with regards to business-friendly legislation, generous tax incentives, and technology grants are effective. More importantly, a targeted and focused approach in identifying promising SMEs to adopt e-commerce is more effective than targeting all, since many are brick-and-mortar companies whose lean operations do not justify adoption of e-commerce. In many cases, market developments will also generate competitive pressures for many SMEs to adopt e-commerce practices.

REFERENCES

Asia Pacific Telecommunications Indicators (2002). Retrieved April 10, 2003 from the World Wide Web at: *http://www.itu.int*

Dutta, S., & Jain, A. (2002). The networked readiness of nations. In S. Dutta, B. Lonvin, & P. Fiona (Eds.), *Global Information Technology Report 2002-2003* (pp. 2-25). London: Oxford University Press.

Fife, E., & Pereira, F. (2002). Small-and medium-size enterprises and the e-economy: Challenges and prospects, *The E-business*, 1-7.

Fischer, A. (2000, September 22). Race is on for supremacy. *Financial Times, London*, 27.

Ho, S. (2002). PKI - Towards building a secure e-business infrastructure. *Asia PKI Forum*, Beijing, China, July 4, 2002.

Koh, J. (2000, July 17). E-biz: Yet to touch small and medium size enterprises, *Business Times*,14.

Kwong, G.S., and Lim, W. (2001, June) Accelerating e-commerce adoption by SMEs, *Productivity Digest*, 2-8.

OECD. (2000). Enhancing SME competitiveness, OECD, Paris, 2001 (p. 158).

Quayle, M. (2001). E-commerce: The challenges for UK small and medium enterprises. *The 10th International Annual IPSERA Conference.*

Tan, W.H.L., & Subramaniam, R. (2000). Wiring up the island state. *Science*, 288, 621-623.

Tan, W.H.L., & Subramaniam, R. (2001). ADSL, HFC and ATM technologies for a nationwide broadband network. In N. Barr (Ed.), *Global communications* (pp. 97-102). London: Hanson Cooke.

Tan, W.H.L., & Subramaniam, R. (2004a). Infrastructure and policy frameworks for the support of intelligent enterprises: the Singapore experience. In J. Gupta & S. Sharma (Eds.), *Intelligent enterprises of the 21st century* (pp.176-187). Hershey, PA: Idea Group Publishing.

Tan, W.H.L., & Subramaniam, R. (2004b). E-commerce as a business enabler for small and medium size enterprises: issues and perspectives from Singapore. In B. Corbitt & N. Al-Qirim (Eds.), *eBusiness, egovernment and small and medum sized enterprises: Opportunities and challenges* (pp 65-90). Hershey, PA: Idea Group Publishing.

Table 4 Cross border electronic commerce

Year	Domestic (S$ million)	Export (S$ million)
1998	1,758	3,913
1999	25,468	14,958
2000	52,840	39,862
2001	51,440	58,014

Source: Ho, 2002

UNCTAD Secretariat. (2002). E-commerce and Development Report 2002. United Nations Conference on Trade and Development.

Wong, P.K. (2001). Globalization and E commerce: Growth and impacts in Singapore. Report for Centre for Research on Information Technology and Organizations, University of California. Retrieved April 11, 2003 from the World Wide Web at: *http://www.crito.uci.edu/git/publications/pdf/singaporeGEC.pdf*

Zwass, V. (2003). Electronic commerce and organizational innovation: Aspects and Opportunities. *International Journal of Electronic Commerce*, 7(3), 1-7.

KEY TERMS

Asymmetric Digital Subscriber Line: A technology which enables transmission of information at high rates over an ordinary phone line by capitalizing on the unused bandwidth outside the voice portion of the line.

Asynchronous Transfer Mode: A high-speed, low delay, multiplexing and switching technology that allows voice, image, data, and video to be transmitted simultaneously rather than through traffic-specific networks.

Bandwidth: Transmission capacity (in kilobytes per second) of a telecommunications line.

Broadband Network: A telecommunications network that can transmit information well above the normal rate (56K) on ordinary phone lines.

B2B: Business-to-Business

E-Commerce: The use of the Internet to conduct business and financial transactions.

E-Government: Public administration infrastructure of a government on the Internet; a range of services needed by citizens and businesses can be accessed here.

Hybrid Fiber Coaxial Cable Modem: A technology which enables transmission of information at high rates over the cable television network's infrastructure comprising optical fibers and coaxial cables.

Small- and Medium-Sized Enterprises: Companies which employ a limited number of people and which have a certain turnover, with the numbers varying among countries.

Supply Chain: The network of producers, wholesalers, retailers, distributors, transport operators, and storage hubs that are involved in the production and delivery of a product to the consumer.

DRM Technology for Mobile Multimedia

Sai Ho Kwok
California State University, Long Beach, USA

INTRODUCTION

Mobile multimedia has been promoted as a promising service and application in mobile e-commerce (m-commerce) by many mobile operators and mobile service providers, when high-speed mobile networks are expected to take off in the near future. However, at present, mobile multimedia is still in its infancy, accessed by relatively low-end mobile devices with limited bandwidth and resources. A typical example is Orange in Hong Kong which launched a low-grade multimedia service in 2000 to test the market with current mobile technologies. Due to the physical constraints of a 2.5G mobile network, audio broadcast is the best service that the network can offer up to date. However, in the near future, when advanced mobile networks and technologies become available, higher demands will be placed on the quality of mobile multimedia services. Such services support both audio and video data, for example, video conferencing, music video, video-on-demand and so on. Rights management deserves more serious concern because intellectual property of distributed multimedia content is as valuable as a company's physical assets (Doherty, 2002). This will become even more important when mobile multimedia services become marketable and an essential part of the business. The purpose of a digital rights management (DRM) system is to allow owners of digital assets (movies, songs) to distribute their products/services/contents electronically in a controlled way (Peinado, 2002). DRM technology makes various online payment schemes possible, such as pay-per-view, pay-per-download, pay-per-game and so on. Hence, mobile service providers are able to control end users' use of, and accessibility to, their products, and stand to gain huge profits from this capability with the DRM technology (Foroughi, Albin, & Gillard, 2002). A successful DRM system should address both business and technical issues (Grab, 2002), but this chapter only addresses and presents issues in the technical side due to the nature of this book. We present some critical issues of mobile DRM for mobile multimedia. A proposal of mobile DRM framework is presented to meet the urgent DRM needs with the existing 2.5G mobile technology. This chapter is concluded by presenting future directions of mobile DRM for mobile multimedia.

BACKGROUND

Internet Commerce

In the Internet domain, Vidius Incorporated estimates 450,000 to 580,000 downloads of unprotected full-length films are transferred over the Internet daily (Grab, 2002). Protection of distributed multimedia has been a growing concern to creators, distributors, copyright owners, publishers, and governments. DRM is considered to be one of the desirable solutions to this problem, and it can protect distributed media contents delivered over the Internet.

Several international standard organizations have been developing DRM solutions for various distributed multimedia, for example, digital music and video. The Secure Digital Music Initiative (SDMI) (SDMI, 2003), backed by the Recording Industry Association of America (RIAA) and 200 music and technology companies (as of October 2003), has been proposed to provide a secure environment for music distribution over the Internet. Another standard being developed by the Moving Picture Experts Group (MPEG) is known as MPEG-21 (Bormans & Hill, 2002) dedicated to distributing digital multimedia content. MPEG-21 defines an interoperable framework for Intellectual Property Management and Protection (IPMP). The IPMP can be interoperable with other MPEG standards, for example, MPEG-4. Therefore, the property protection will be also applicable to most of the MPEG video standards in the future. In addition, there are commercial DRM systems especially for the wired Internet business. They include Windows Media Rights Manager by Microsoft, and MetaTrust by InterTrust Technologies (InterTrust, 2000).

The above DRM standards and systems can be classified into two groups, namely, cryptographic-based and watermark-based DRM solutions (Kwok, 2003). Cryptographic systems permit only valid key-holders to access the encrypted data after receiving it from the authenticated senders. However, once such data is decrypted, it is impossible to track its reproduction or retransmission. Therefore, cryptography only provides protection during data transmission. Digital watermarking technology seems to complement the cryptographic process and to protect

Table 1. Summary of features of DRM solutions

- Media right protection and management
- Secure delivery and distribution of digital contents
- Processing authorization, data authentication and verification for content service
- Data security, integrity check, access control, and management for distributed systems and peer-to-peer (P2P) networks
- Multimedia watermarking for copyright protection, media authentication and integrity checking, finger-printing, and data annotation

Table 2. Problems and issues for mobile DRM

Mobile DRM standard: There is not yet a winner of mobile DRM standard. Open Mobile Alliance (OMA) DRM standard is one of the outstanding mobile DRM standards for mobile phones (Poropudas, 2003). However, other DRM standards, such as Windows Media DRM for Pocket PC (Microsoft, 2003), are highly competitive.

Trustful DRM protocol: Since DRM involves many parties, for example, technology service providers, mobile operators, service providers, creators, distributors and so forth, trust may not exist in all of these parties, for example, in a second-hand market. Hence, a trustful DRM protocol that can deal with DRM but without assuming mutual trust between involved parties is needed. A similar protocol was proposed by Cheung and Curreem (2002).

Robust and secure watermarking: A secure and robust watermarking algorithm is required to protect the distributed multimedia content. Such watermarking algorithm should resist attacks of any kinds. However, it cannot guarantee that a watermarking algorithm can resist all upcoming attacks (Tsang & Au, 2001).

Payment scheme: When trust does not exist, for example, in the second-hand market, a reliable payment scheme becomes an important issue.

Rights expression language: This is a need for a cross-platform rights expression language for all involved parties to specify and utilize their rights.

copyright ownership (Kwok, 2002). Digital watermarks can be visible but they are preferably invisible identification codes that are permanently embedded in the data and present within the data after any decryption process (Doherty, 2002).

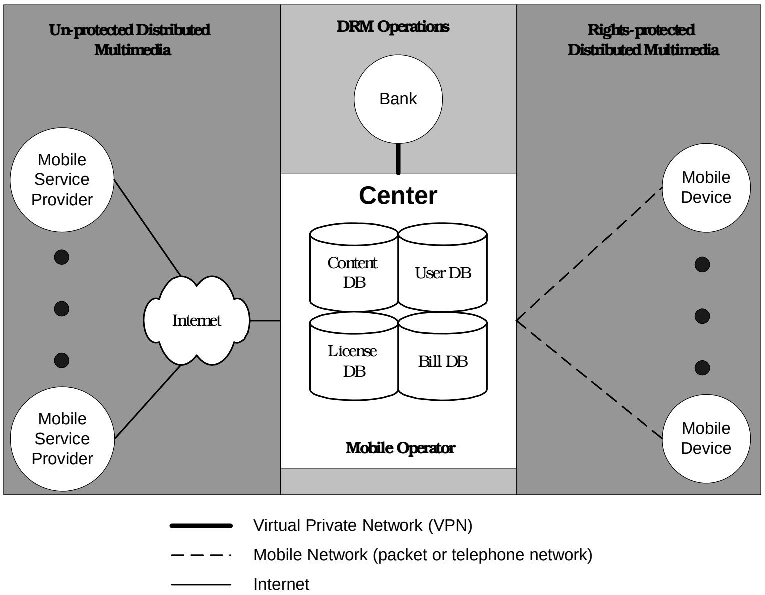
In order to manage digital rights effectively and efficiently, many commercial DRM solutions employ license management models (Kwok & Lui, 2002). A license management model consists of a digital license that keeps access and control rights. Corresponding rights enforcement DRM applications determine usage rights based on these digital licenses.

Mobile Commerce

The current 2.5G mobile technologies for mobile multimedia service are fundamentally different from those used for Internet commerce service, and they impose many limitations and constrains upon the sophistication of mobile multimedia service. This explains why existing DRM systems for Internet commerce cannot be applicable to DRM over the mobile environment in a straightforward way. Some of the most important technical and physical obstacles are summarized as follows:

1. *License management:* A mobile device usually has limited resources of both memory and processing power to handle and process license documents and rights-protected contents.
2. *Limited storage and processing power:* Due to the limited resources of the mobile device, it is not possible to download rights-protected contents to the mobile device and play it there.

Figure 1. DRM solution for mobile multimedia



3. *Rights insertion:* A sophisticated consumer's ID cannot be kept on the consumer's device due to the storage limitation, and it must be provided by another party or uploaded for rights insertion.
4. *Rights enforcement:* An active rights enforcement cannot take place at the mobile device because the device is not capable of intensive computation.
5. *Payment:* Mobile devices cannot support the elaborate computations required for the encryption and de-encryption process of electronic payment, and mobile networks may not be adequately secure to prevent the exposure of personal and credit card information.

CRITICAL ISSUES OF DRM FOR MOBILE MULTIMEDIA

For mobile multimedia, DRM involves specifying and associating rights with the distributed multimedia contents, placing controls on the content to enforce rights, enabling access checks, and tracking permissions usage and payment. For a general mobile service, the required capabilities include:

1. rights specification and rights label management;
2. rights authorization;
3. content protection, rights enforcement, and trusted rendering;
4. rights tracking; and
5. a security and commerce infrastructure.

Business transactions, such as payment, ordering, customer enquiry, and so forth, may occur between the concerned parties during content packaging, distribution, and usage. Managing rights in all these transactions is necessary. To support DRM operations in mobile multimedia, a DRM system needs to perform rights insertion and rights enforcement operations. In addition, a license management mechanism is also needed in managing license documents. A DRM solution for mobile multimedia should possess features stated in Table 1.

There are still many un-resolved technical problems and issues to be addressed before a successful DRM system for mobile multimedia can emerge. Some existing problems and issues are listed in Table 2.

A MOBILE DRM FRAMEWORK FOR MOBILE MULTIMEDIA

This chapter proposes a mobile DRM framework for mobile multimedia, derived from Kwok (2002). The framework is a closed system that hosts all DRM operations within its center and provides a platform for mobile service providers to conduct business with their customers. Apart from rights insertion and enforcement operations, the center can handle transactions with a highly secure payment scheme. It is assumed that the central party is mutually trusted by both the businesses and consumers. A trustful center can be a mobile operator. The distributed multimedia contents between the center and the mobile users are rights-protected with a digital license and watermarking, while the multimedia contents transferring to the center do not require rights protection.

Figure 1 presents the mobile DRM framework for multimedia distribution in a mobile environment. The center of the framework is a mobile operator that manages information to and from mobile devices, mobile service providers and other concerned parties. The principal components include: (1) a mobile network infrastructure; (2) a DRM system; (3) a payment system; and (4) databases. There are three types of parties involved in this framework: the mobile service providers (both official and unofficial sites), the bank, and the mobile users. The communication channels between different parties and the mobile operator are different from and independent of each other depending on the required security level. For example, a virtual private network (VPN) is used between the bank and the mobile operator, since highly confidential information is transferred through this channel, while the mobile operator relies on the packet network for multimedia content distribution, and the mobile service providers transfer multimedia contents to the mobile operator using the ordinary Internet.

The distinct features of the proposed framework include the following.

1. *DRM operation:* All DRM operations are performed by the center. It shifts all the processing and storage requirements to the center and relieves the burdens of the mobile devices and service providers.
2. *Independence of mobile devices:* The center can tailor the format of the distributed media for a specific mobile device. Besides, streaming technology is used in order to overcome the problems of processing power and storage requirement in the 2.5G mobile devices.
3. *Independence of mobile technology:* The framework can be applicable to 2.5G, 2.75G, 3G, 4G, and even higher because it does not depend on any specific mobile standard.
4. *Standardized rights expression language:* Since all DRM operations are managed by a single party—the mobile operator—the rights expression language can be standardized.
5. *Sharing and trading:* It facilitates media sharing and trading between users. Detail may be referred to Kwok (2002).
6. *Payment:* Transactions and payments are handled centrally through a secured channel.
7. *Ease of use and user satisfaction:* All DRM operations are completely transparent to mobile users and the mobile service providers.

FUTURE TRENDS

Mobile DRM for mobile multimedia is still at its infancy. The direction of mobile DRM is driven by the following factors.

1. *DRM standard:* One key player in mobile DRM standard, OMA mobile DRM has been proven successful in applying to music distribution. The standard is currently supported by some major labels, including Warner Music and BMG. However, the spectrum of mobile multimedia covers more than digital music, but also includes visual-audio data, such as movie, video conferencing, and so forth. It is still uncertain whether the market will accept OMA mobile DRM as the common standard for mobile multimedia.
2. *Mobile network:* Mobile multimedia demands a highly capable mobile network to support its services. An independent and constantly high transmission rate mobile network is the basic requirement for satisfactory mobile multimedia services. Unfortunately, the current mobile network, 2.5G or 2.75G, cannot provide a stable multimedia transmission. This problem will be overcome when 3G or 4G is launched.
3. *Mobile device:* The capabilities of mobile devices will be a major factor affecting the quality of mobile service. Pocket PCs usually perform better than smart phones when viewing mobile movie as their viewing screen and processing power are higher.

CONCLUSION

This chapter presents a mobile DRM framework for mobile multimedia. It is a practical and useful DRM framework when common mobile DRM standard and high bandwidth mobile channel are not available. This temporary but timely DRM solution could meet the urgent needs of DRM in mobile services. The primary objective of the framework is to impose DRM on mobile multimedia services without affecting the service providers and users. This is rather different from the emerging mobile DRM standards that require mobile users and service providers to adopt and apply their DRM technologies to mobile devices and distributed multimedia contents. However, privacy is the major problem of the proposed framework because the mobile operator possesses all of our transactions records. To respond to this problem, a possible solution may be encryption (Torrubia, Mora, & Marti, 2001) and an adapting system (Kenny & Korba, 2002).

ACKNOWLEDGEMENTS

The work described in this article was partially supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. HKUST6256/03E).

REFERENCES

Bormans, J., & Hill, K. (2002). MPEG-21 overview v.5. Retrieved October 19, 2003, from *http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm*

Cheung, S.C., & Curreem, H. (2002). Rights protection for digital contents redistribution over the Internet. Paper presented at the *26th Annual International Computer Software and Applications.*

Doherty, S. (2002). Managing your digital rights. *Network Computing, 13*(19), 65-68.

Foroughi, A., Albin, M., & Gillard, S. (2002). Digital rights management: A delicate balance between protection and accessibility. *Journal of Information Science, 28*(5), 389-395.

Grab, E. (2002). Applying DRM techniques to video on the Internet: Characterizing problems and solutions. *SMPTE Journal-Society of Motion Picture & Television Engineers, 111*(3), 154-158.

InterTrust. (2000). InterTrust, The MetaTrust utility, announces OpenRights Initiative. Mountain View, CA: InterTrust Press.

Kenny, S., & Korba, L. (2002). Applying digital rights management systems to privacy rights management. *Computers & Security, 21*(7), 648-664.

Kwok, S.H. (2002). Chapter 5: Digital rights management for mobile multimedia. In E.P. Lim, Z. Shen, & K. Siau (Eds.), *Mobile commerce: Current states and future trends* (pp. 97-111). Hershey, PA: Idea Group Publishing.

Kwok, S.H. (2003). Digital watermarking for digital rights management. In L. Jain, H.C. Huang, & J.S. Pan (Eds.), *Intelligent watermarking techniques*. Hauppauge, NY: Nova Science Publishers.

Kwok, S.H., Cheung, S.C., Wong, K.C., Tsang, K.F., Lui, S.M., & Tam, K.Y. (2003). Integration of digital rights management into Internet open trading protocol (IOTP). *Decision Support Systems (DSS), 34*(4), 413-425.

Kwok, S.H., & Lui, S.M. (2002). A license management model for peer-to-peer music sharing. *International Journal of Information Technology and Decision Making (IJITDM), 1*(3), 541-558.

Kwok, S.H., Lui, S.M., Cheung, S.C., & Tam, K.Y. (2003). Digital rights management with Web services. *Electronic Markets, 13*(2), 133-140.

Microsoft. (2003). Windows Media DRM. Retrieved October 27, 2003, from *http://www.microsoft.com/windows/windowsmedia/drm.aspx*

Paskin, N. (2003). On making and identifying a "copy". *D-Lib Magazine, 9*.

Peinado, M. (2002). Digital rights management in a multimedia environment. *SMPTE Journal-Society of Motion Picture & Television Engineers, 111*(3), 159-163.

Poropudas, T. (2003). OMA digital rights arrive. Retrieved October 26, 2003, from *http://www.mobile.seitti.com/print.php?story_id=3136*

SDMI. (2003). Secure Digital Music Initiative. Retrieved October 20, 2003, from *www.sdmi.org*

Torrubia, A., Mora, F.J., & Marti, L. (2001). Cryptography regulations for e-commerce and digital rights management. *Computers & Security, 20*(8), 724-738.

Trowbridge, C. (2003, 1995 [October 13]). Image protection for archives, special collection libraries and museums in the WWW environment. Retrieved April 15, 2003, from *http://sunsite.berkeley.edu/Imaging/Databases/Fall95papers/trowbridge.html*

Tsang, K.F., & Au, O.C. (2001). A review on attacks, problems and weaknesses of digital watermarking and the pixel reallocation attack. *Spie - the International Society for Optical Engineering, 4314*, 385-393.

KEY TERMS

Digital License: A digital license can be a separate file or message embedded in a media file. The license document states all of the terms and conditions concerning the use of the licensed media file. These terms and conditions can be static or dynamic depending on the payment scheme (Kwok, 2002).

Digital Rights Management (DRM): A set of technologies for content owners to protect their copyrights and stay in closer contact with their customers. In most instances, DRM is a system that encrypts digital media content and limits access to only those users who have acquired a proper license to play the content. That is, DRM is a technology that enables the secure distribution, promotion, and sale of digital media content on the Internet.

Identifiers and Metadata: Identifiers (unique labels for entities) and metadata (structured relationships between identified entities) are prerequisites for DRM. The essence of DRM is the control (licensing, etc.) of copies of entities; the identifiers and metadata are then essential to the management of this process, and to distinguishing and expressing relationships such as replicas and derivations (Paskin, 2003).

License Management: A mechanism to execute the terms and conditions stated in the license. This requires coordination among the media player, the media file, and other supporting modules; for example, the payment module. From the technical perspective, license management refers to issuing, hosting, and verifying the license (Kwok, 2002).

Rights Enforcement (or Verification): There are two types of rights enforcement: namely active enforcement and passive enforcement. The active enforcement takes place within the media player as a built-in function. The passive enforcement is an off-line ownership verification operation to check for the hidden owner identities (Kwok, 2002; Kwok, Cheung, Wong, Tsang, Lui & Tam, 2003; Kwok, Lui, Cheung, & Tam, 2003).

Rights Insertion: An operation to embed the identities of the concerned parties and assign business rules and conditions to the distributed multimedia content (Kwok, 2002; Kwok, Cheung, Wong, Tsang, Lui, & Tam, 2003; Kwok, Lui, Cheung, & Tam, 2003).

Watermarking: A technique for media authentication and forgery prevention. It is also viewed as an enabling technology to protect media from reuse without adequate credit or in an unauthorized way (Trowbridge, 2003). A watermarked media, M', can be mathematically represented as $M'=M + W$ where M is the original media content and W is the embedded watermark. It is common that the extracted watermark, W', could be different from the original watermark W because of the intentional or unintentional attacks or post processing. To detect the watermark, a watermark detector is used to evaluate the similarity between W and W'.

Dynamic Multidimensional Data Cubes for Interactive Analysis of Massive Datasets[1]

Mirek Riedewald
Cornell University, USA

Divyakant Agrawal
University of California, Santa Barbara, USA

Amr El Abbadi
University of California, Santa Barbara, USA

INTRODUCTION: DATA WAREHOUSING, OLAP, AND DATA CUBES

Rapidly improving computing and networking technology enables enterprises to collect data from virtually all its business units. The main challenge today is to extract useful information from an overwhelmingly large amount of raw data. To support complex analysis queries, data warehouses were introduced. They manage data, which is extracted from the different operational databases and from external data sources, and they are optimized for fast query processing. For modern data warehouses, it is common to manage Terabytes of data. According to a recent survey by the Winter Corporation (2003), for instance, the decision support database of SBC reached a size of almost 25 Terabytes, up from 10.5 Terabytes in 2001 (Winter Corporation, 2001).

Human analysts cannot "digest" such large amounts of information at a detailed level. Instead they rely on the system to provide a summarized and task-specific view of selected data. Consequently, *efficient summarization and aggregation of large amounts of data* play a crucial role in the analysis process. The goal of Online Analytical Processing (OLAP) is to support this style of analysis at interactive response times for massive data collections. OLAP applications often maintain aggregate information in data cubes. Pre-computed aggregate values that speed up query execution for group-bys, cross-tabs, and subtotals can be easily included in the model (e.g., CUBE operator, which is discussed later).

WHY MAKE DATA CUBES DYNAMIC

The primary goal of data warehouses is to support data analysis. Update costs were often not considered to be important, and hence systems are typically oriented towards batch updates, which are applied to the warehouse during times of low system load. The corresponding data cubes would be batch-updated as well, or even be recomputed from scratch. There are several arguments why incrementally maintainable *dynamic* data cubes are preferable.

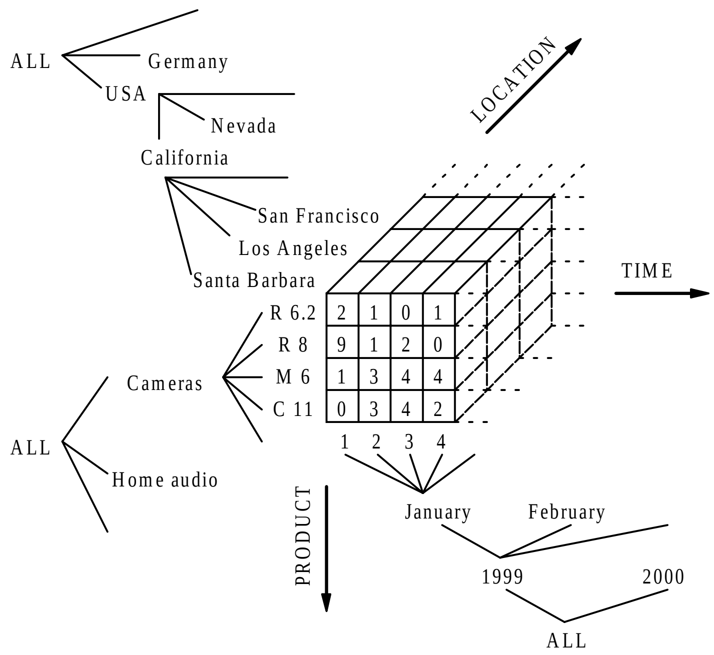
First, for very large data collections, re-computing a data cube from scratch is very costly, even if updates are batched. Second, some data structures that are heavily optimized towards query processing (e.g., the PS technique as discussed below) support only very limited update rates even if the updates are applied in batches. Third, OLAP is inherently an *interactive* information exploration process that includes what-if scenarios. What-if analysis requires real-time and hypothetical data to be integrated with historical data for the purpose of instantaneous analysis and subsequent action.

This article surveys dynamic solutions that offer different tradeoffs between query, update, and storage costs, allowing an administrator to find the matching approach for each application. It will be shown that efficient queries and updates can both be supported for multidimensional data cubes.

BACKGROUND

The terms "cube" and "data cube" are not used consistently in the data warehousing research literature. In this article they refer to a conceptual model similar to a multidimensional array. A simple example is a sales data cube SalesCube with dimensions product, time, and location, and the measure attribute amount (value of sales transaction). Figure 1 shows a possible instance with dimension hierarchies indicated (ALL represents the whole domain of an attribute). Note that we do not advocate any

Figure 1. Cube example



particular cube implementation (Roussopoulos, Kotidis & Roussopoulos, 1997; Sarawagi & Stonebraker, 1994) or system architecture like MOLAP or ROLAP.

Dominant aggregate query types on data cubes are roll-up/drill-down and slice-and-dice (Chaudhuri & Dayal, 1997). These queries are typically *range queries* or sequences of related range queries like hierarchical range queries (Koudas, Muthukrishnan & Srivastava, 2000). Hence we are mainly concerned with support for *aggregate range queries*. For the cube in Figure 1, an example is: "Compute the total sales of cameras in California for the first half of January 1999." The corresponding SQL query is

```
SELECT SUM(amount)
FROM SalesCube
WHERE time>='1-Jan-1999' AND time<='15-Jan-1999' AND
location.state='California' AND
product.category='Cameras'.
```

Note that location is a *categorical dimension* for which arbitrary range selections like "San Francisco < town < San Diego" are not meaningful. On the other hand it is very common to aggregate according to the dimension hierarchy. In the example, California corresponds to a range of selected cities (by making sure the dimension values are ordered according to the hierarchy). The same holds for other categorical dimension attributes like product.

For the following discussion we assume that the data cube has N cells in each dimension, that is, it has a total of N^d cells. Note that all techniques also work for cubes with different dimension sizes, but since our goal is to present the main tradeoffs of the techniques, we try to avoid notational clutter as much as possible. For the same reason we also omit ceiling and floor brackets from the formulas.

DATA CUBES FOR INVERTIBLE AGGREGATE OPERATORS

For invertible aggregate operators like SUM, COUNT, and AVG (when expressing it with SUM and COUNT), elegant techniques have been developed that do not require additional storage compared to materializing the original (dense) MOLAP cube. In the following we will survey techniques for the operator SUM. Other invertible operators are handled in a similar way.

Let A denote the original data cube. The cost of an operation is measured in terms of accessed cube cells, for example, updating a single cell in A results in a cost of 1. Answering an aggregate range query using A can be very costly. Even for a small cube with only three dimensions which contains data about 1,000 products, 100 store locations, and six years (about 2,000 days), a query in the worst case has to access and aggregate the values of 200 million cells. Hence using the original data cube *A* for OLAP represents a perfectly dynamic solution, but achieves only unsatisfactory query efficiency.

Trading off Query and Update Performance

Ho et al. (1997) addressed the problem of high query costs with the Prefix Sum (PS) technique. The main idea is to answer queries using a prefix sum cube *PS* whose cells *c* store the pre-computed sum of all cells in *A* which have smaller or equal dimension coordinates than *c*. Based on the principle of inclusion-exclusion, each range sum query on *A* can be computed from up to 2^d appropriately selected prefix sums in *PS*. Note that this cost is independent of the size of the data cube and the size of the selected query range. Unfortunately, the query efficiency comes at a price: in the worst case a single update could require the re-computation of the entire *PS* cube (at a cost of N^d).

To avoid PS's expensive cascading updates, Geffner and his colleagues (1999a, 1999b) introduced the Relative Prefix Sum (RPS). A later version of RPS removed the initial storage overhead (Riedewald, Agrawal, El Abbadi & Pajarola, 2000b). The main idea is to partition *A* into smaller hyper-rectangular chunks, called *boxes*. The *inner* cells of a box store prefix sums local to the box, while the *border* cells in selected surfaces also aggregate outside values. When the side-length of a box is set to $\sqrt{N}$ in each dimension, then the worst-case costs are 4^d for queries and $(2\sqrt{N})^d$ for updates.

The Dynamic Data Cube (DDC) proposed by Geffner, Agrawal, and El Abbadi (2000) balances query and update costs such that they are both polylogarithmic in the data size. Riedewald et al. (2000b) removed the considerable storage overhead of the original technique. Like RPS, DDC conceptually partitions the original cube into boxes. However, in contrast to RPS the partitioning is applied recursively to both inner and border cells. If the recursive box partitioning always divides boxes in half, the technique guarantees worst-case costs of $(2\log_2 N)^d$ and $(\log^2 N)^d$ for queries and updates, respectively.

Schmidt and Shahabi (2002) proposed ProPolyne, a wavelet-based pre-aggregation technique, which also guarantees polylogarithmic query and update costs. Depending on the selected wavelet encoding, all aggregates that are polynomials up to a fixed maximum degree over the measure values can be supported by a single data cube. The authors also show that high-quality approximate results can be obtained at a fraction of the cost of the exact result. Interestingly ProPolyne and DDC not only have the same asymptotic query and update cost; they also have the same storage requirements. Choosing wavelets that support higher-degree polynomials for ProPolyne has the same effect on storage as storing additional measure values of higher degrees in a DDC cube.

Chan and Ioannidis (1999) are the first to explicitly explore the possibility of letting the user choose a query-update tradeoff by parameterizing the construction of the pre-aggregated cube. They apply ideas from bitmap index design and define a class of so-called Hierarchical Cubes (HCs). The data cube is hierarchically partitioned into smaller boxes of equal size. According to the partitioning, the cells are assigned to classes, which determine the aggregate stored in a cell. Explicit, but complex, cost formulas enable an analyst to choose the best parameter setting for an application by experimental evaluation.

Combining Efficiency, Flexibility, and Simplicity

RPS, DDC, and HC are sophisticated multidimensional pre-aggregation techniques. Hence the corresponding algorithms are typically complex (Chan & Ioannidis, 1999; Geffner et al., 2000), and it is difficult to prove their correctness and to analyze their performance. The different dimensions of a data cube are treated uniformly in the sense that even though box sizes might vary, the same general scheme is used for each dimension. However, in practice, attributes have very diverse characteristics, for example, gender and income in a census data set. Also, for some attributes a priori information about the query workload could make one aggregation technique preferable over another.

This is the motivation behind the Iterative Data Cubes (IDC) approach proposed by Riedewald, Agrawal, and El Abbadi (2001). The approach provides a modular framework for combining *one-dimensional* pre-aggregation techniques to create space-optimal multidimensional pre-aggregated data cubes. An analyst can select the most appropriate technique for each dimension attribute independently, then the corresponding *multidimensional* IDC cube is automatically generated. Query and update algorithms for a *d*-dimensional IDC cube are a straightforward combination of the one-dimensional algorithms. Similarly, query and update *costs* for a *d*-dimensional IDC cube can be computed as the product of the corresponding costs for each single dimension. Table 1 shows selected techniques with their respective worst-case costs for vectors with *N* elements. The possibility of combining arbitrary one-dimensional pre-aggregation techniques allows generating a great variety of IDC instances.

Table 1. Query-update tradeoff for selected one-dimensional techniques

	Query cost (worst case)	**Update cost (worst case)**
Original array (A)	N	1
Prefix Sum (PS)	2	N
Relative Prefix Sum (RPS)	4	$2\sqrt{N}$
Dynamic Data Cube (DDC)	$2\log_2 N$	$\log_2 N$

DATA CUBES FOR OTHER AGGREGATE OPERATORS

Popular aggregate operators like MIN and MAX cannot be handled by the techniques discussed in the previous section due to the lack of an inverse operation. This section summarizes approaches that work for these two operators, and other commutative semigroup operators as well.

In Ho et al. (1997), a technique for the computation of range max queries is proposed. The data cube is recursively partitioned into smaller chunks. For each chunk, the position of its maximum value is stored. The corresponding tree is traversed top-down, using pruning techniques to reduce the number of accesses. Lee, Ling, and Li (2000) use a similar data structure, but a different pruning strategy. They obtain constant average query cost and logarithmic update cost in the size of the data cube (for fixed dimensionality). Poon (2003) develops an optimal range max data cube for the static case (no updates) that guarantees constant query cost and a storage cost that is linear in the size of the original cube.

Poon (2001) also proposes a general approach for commutative semigroups. Similar to Iterative Data Cubes, it combines one-dimensional schemes. Query, update, and storage costs for range max queries are $O((4L)^d)$, $O((12L^2 N^{\frac{1}{L}}\gamma(\mathrm{N}))^d)$, and $O((6N\gamma(N))^d)$, respectively. Here $L\in\{1,...,\log N\}$ denotes a user-controlled parameter and $\gamma(N)$ a slow-growing function. Applied to invertible operators, the query, update, and storage costs are $O((2L)^d)$, $O((2LN^{\frac{1}{L}})^d)$, and $O((2N)^d)$, respectively. Note that the greater generality of the technique comes at a cost: techniques like IDC, which are optimized to take advantage of the inverse operator, achieve better cost tradeoffs for invertible operators. For instance, for $L = 1$, an IDC that uses PS in each dimension achieves the same query, but lower update and storage costs (please see Table 1). Similarly IDC using RPS (DDC) in each dimension offers a better solution than the above technique for L=2 (L=log$N$).

THE DATA CUBE OPERATOR CUBE

CUBE, as proposed by Gray et al. (1997), generalizes SQL's GROUP BY operator. It computes the results of grouping a data set by all subsets of the specified d dimensions. Hence 2^d related groups, called *cuboids*, are generated. The overall result of CUBE will be referred to as *Cube*. In relational database terminology the cuboids are *views* generated by the corresponding GROUP BY queries.

Cube's main purpose is to support aggregate queries. A query that only selects a few attributes can be efficiently answered by processing the appropriate compact cuboid. Specialized index structures for Cube (Johnson & Shasha, 1999; Roussopoulos et al., 1997; Sismanis, Deligiannakis, Roussopoulos & Kotidis, 2002) further reduce query time. A large body of research focused on efficient computation, maintenance, and space reduction for Cube (Beyer & Ramakrishnan, 1999; Gupta, Harinarayan, Rajaraman & Ullman, 1996, 1997; Shukla, Deshpande & Naughton, 2000; Sismanis et al., 2002; Wang, Lu, Feng & Yu, 2002; Baralis, Paraboschi & Teniente, 1997).

CONCLUSION AND FUTURE TRENDS

We have discussed several approaches for supporting dynamic data cubes, focusing on MOLAP techniques. The commonality between all approaches is that they try to find a certain balance between query, update, and storage cost. While earlier work mostly focused on query and storage aspects, large data sets with frequent updates created a need for more dynamic solutions.

In the future, support for *sparse* and *high-dimensional* data will become increasingly important for supporting analysis of complex business processes, which depend on many attributes. Existing approaches from the computational geometry (de Berg, van Kreveld, Overmars & Schwarzkopf, 2000; Chazelle, 1988; Willard & Lueker, 1985) and database communities (Riedewald, Agrawal & El Abbadi, 2000a) have infeasible storage overhead or do not provide non-trivial performance guarantees. Independent of the data being dense or sparse, the current trend of rapidly growing data collections creates a strong incentive for developing new powerful *dynamic* approaches for maintaining data cubes *incrementally*.

REFERENCES

Baralis, E., Paraboschi, S. & Teniente, E. (1997). Materialized view selection in a multidimensional database. *Proceedings of the International Conference on Very Large Databases* (VLDB) (pp. 156-165).

Beyer, K. & Ramakrishnan, R. (1999). Bottom-up computation of sparse and iceberg CUBEs. *Proceedings of the ACM SIGMOD International Conference on Management of Data* (pp. 359-370).

Chan, C.-Y. & Ioannidis, Y.E. (1999). Hierarchical cubes for range-sum queries. *Proceedings of the International Conference on Very Large Databases* (VLDB) (pp. 675-686). Extended version published as Technical Report, University of Wisconsin.

Chaudhuri, S. & Dayal, U. (1997). An overview of data warehousing and OLAP technology. *SIGMOD Record*, *26*(1), 65-74.

Chazelle, B. (1988). A functional approach to data structures and its use in multidimensional searching. *SIAM Journal on Computing*, *17*(3), 427-462.

de Berg, M., van Kreveld, M., Overmars, M. & Schwarzkopf, O. (2000). *Computational geometry* (2nd edition). Springer-Verlag.

Geffner, S., Agrawal, D., & El Abbadi, A. (2000). The dynamic data cube. In *Proceedings of the International Conference on Extending Database Technology* (EDBT) (pp. 237-253).

Geffner, S., Agrawal, D., El Abbadi, A. & Smith, T. (1999a). Relative prefix sums: An efficient approach for querying dynamic OLAP data cubes. *Proceedings of the International Conference on Data Engineering* (ICDE) (pp. 328-335).

Geffner, S., Riedewald, M., Agrawal, D. & El Abbadi, A. (1999b). Data cubes in dynamic environments. *Data Engineering Bulletin*, *22*(4), 31-40.

Gray, J., Chaudhuri, S., Bosworth, A., Layman, A., Reichart, D., Venkatrao, M., Pellow, F. & Pirahesh, H. (1997). Data cube: A relational aggregation operator generalizing group-by, cross-tab, and sub-totals. *Data Mining and Knowledge Discovery*, 29-53.

Gupta, H., Harinarayan, V., Rajaraman, A. & Ullman, J.D. (1997). Index selection for OLAP. *Proceedings of the International Conference on Data Engineering* (ICDE) (pp. 208-219).

Harinarayan, V., Rajaraman, A. & Ullman, J.D. (1996). Implementing data cubes efficiently. *Proceedings of the ACM SIGMOD International Conference on Management of Data* (pp. 205-216).

Ho, C., Agrawal, R., Megiddo, N. & Srikant, R. (1997). Range queries in OLAP data cubes. *Proceedings of the ACM SIGMOD International Conference on Management of Data* (pp. 73-88).

Johnson, T. & Shasha, D. (1999). Some approaches to index design for cube forests. *IEEE Data Engineering Bulletin*, *22*(4), 31-40.

Koudas, N., Muthukrishnan, S. & Srivastava, D. (2000). Optimal histograms for hierarchical range queries. *Proceedings of the Symposium on Principles of Database Systems* (PODS) (pp. 196-204).

Lee, S.Y., Ling, T.W. & Li, H.G. (2000). Hierarchical compact cube for range-max queries. *Proceedings of the International Conference on Very Large Databases* (VLDB) (pp. 232-241).

Pendse, N. & Creeth, R. (2003). The OLAP report. Retrieved from *www.olapreport.com/Analyses.htm.*

Poon, C.K. (2001). Orthogonal range queries in OLAP. *Proceedings of the International Conference on Database Theory* (ICDT) (pp. 361-374).

Poon, C.K. (2003). Optimal range max datacube for fixed dimensions. *Proceedings of the International Conference on Database Theory* (ICDT) (pp. 158-172).

Riedewald, M., Agrawal, D. & El Abbadi, A. (2000a). pCube: Update-efficient online aggregation with progressive feedback and error bounds. *Proceedings of the International Conference on Scientific and Statistical Database Management* (SSDBM) (pp. 95-108).

Riedewald, M., Agrawal, D. & El Abbadi, A. (2001). Flexible data cubes for online aggregation. *Proceedings of the*

International Conference on Database Theory (ICDT) (pp. 159-173).

Riedewald, M., Agrawal, D., El Abbadi, A. & Pajarola, R. (2000b). Space-efficient data cubes for dynamic environments. *Proceedings of the International Conference on Data Warehousing and Knowledge Discovery* (DaWaK) (pp. 24-33).

Roussopoulos, N., Kotidis, Y. & Roussopoulos, M. (1997). Cubetree: Organization of and bulk updates on the data cube. *Proceedings of the ACM SIGMOD International Conference on Management of Data* (pp. 89-99).

Sarawagi, S. & Stonebraker, M. (1994). Efficient organization of large multidimensional arrays. *Proceedings of the International Conference on Data Engineering* (ICDE) (pp. 328-336).

Schmidt, R.R. & Shahabi, C. (2002). ProPolyne: A fast wavelet-based algorithm for progressive evaluation of polynomial range-sum queries. *Proceedings of the International Conference on Extending Database Technology* (EDBT) (pp. 664-681).

Shukla, A., Deshpande, P. & Naughton, J.F. (2000). Materialized view selection for multi-cube data models. *Proceedings of the International Conference on Extending Database Technology* (EDBT) (pp. 269-284).

Sismanis, Y., Deligiannakis, A., Roussopoulos, N. & Kotidis, Y. (2002). Dwarf: Shrinking the petacube. *Proceedings of the ACM SIGMOD International Conference on Management of Data* (pp. 464-475).

Wang, W., Lu, H., Feng, J. & Yu, J.X. (2002). Condensed cube: An effective approach to reducing data cube size. *Proceedings of the International Conference on Data Engineering* (ICDE) (pp. 155-164).

Willard, D.E. & Lueker, G.S. (1985). Adding range restriction capability to dynamic data structures. *Journal of the ACM*, *32*(3), 597-617.

Winter Corporation. (2001). Database scalability program. Retrieved from *www.wintercorp.com*

Winter Corporation. (2003). Top ten program. Retrieved from *www.wintercorp.com*

KEY TERMS

Aggregate Range Query: Selects a range of values in each dimension (dimensions not specified in the query are by default completely selected) and computes an aggregate, e.g., SUM, MAX, or COUNT, over the measure values of all data cube cells whose coordinates fall into the selected range. A typical aggregate range query is: "Compute the total sales of cameras in California for the first half of January 1999."

Cell: A point in the multi-dimensional data space. For a *d*-dimensional data cube, it is defined by a *d*-tuple of dimension values. A cell contains the values of the measure attributes of the data item that falls into that cell. If there is no such data item, then the cell is empty.

Data Cube: A data set is conceptually modeled as being embedded in a multidimensional hyper-rectangle, or data cube for short. The data cube is defined by its dimensions and stores the values of measure attributes in its cells.

Dense Data Cube: A data cube is dense if a significant number of its cells (typically at least 1 to 10%) are not empty.

Dimension: A dimension attribute of an entity is a functional attribute that describes an aspect of the entity, e.g., location or product. Dimension attributes can be hierarchical, e.g., year-quarter-month-day for the time dimension.

Measure: A measure attribute of an entity is an attribute of interest whose value depends on the dimensions of the entity. Measure values are aggregated and analyzed in OLAP. Examples are revenue and cost.

MOLAP: OLAP architecture that is based on a multidimensional database server. It typically maintains data cubes in compressed array-like data structures.

OLAP(Online Analytical Processing): A category of applications and technologies for collecting, managing, processing, and presenting multidimensional data for analysis and management purposes (Pendse & Creeth, 2003).

ROLAP: OLAP architecture that is based on a relational database engine. Data cubes are maintained in relational tables.

Sparse Data Cube: The vast majority (e.g., 99%) of the data cube cells are empty.

ENDNOTE

[1] This work was partially supported by NSF award IIS-0223022

E-Business Systems Security for Intelligent Enterprise

Denis Trcek
University of Primorska, Koper, Slovenia

INTRODUCTION AND BACKGROUND

Security became a topic of research with the introduction of *networked information systems*, or networked IS, in the early eighties. In the mid-nineties the proliferation of the Internet in the business area exposed security as one of the key factors for successful online business, and the majority of effort to provide it was focused on technology. However, due to lessons learned during this period, the paradigms have since changed, with increasing emphasis on human factors. It is a fact that security of information systems is becoming part of the core processes in all e-business environments. While data is clearly one of the key assets and has to be protected accordingly, IS have to be highly integrated and open. Appropriate treatment of these issues is not a trivial task for managers of intelligent enterprises and requires new approaches, especially in light of new technologies.

Proper management of security in *e-business systems* requires a holistic methodology with a two-plane approach, technological and organizational. In every case IS security management starts with the identification of threats and threats analysis - a typical example is based on risk probability and damage estimates (Raepple, 2001). Following this, the approach differs according to the plane:

- The technological plane takes into account machine-related interactions. This plane is about deployment of appropriate *security services* that are based on *security mechanisms*. To become operational, key management issues (i.e., handling of cryptographic algorithms' keys) have to be resolved. Finally, human-to-machine interactions have to be addressed carefully.
- In parallel, it is necessary to properly address the organizational plane where human resources management plays a central role. This plane emphasizes the organizational issues and socio-technical nature of contemporary IS, where modern methodologies play a central role.

FUTURE TRENDS: TECHNOLOGICAL PLANE OF IS SECURITY MANAGEMENT

From the technological point of view, the prevention of threats is achieved by use of security mechanisms and security services (ISO, 1995). Mechanisms include symmetric and asymmetric cryptographic algorithms, for example, AES (Foti, 2001) and RSA (RSA Labs, 2002); one-way hash functions such as SHA-1 (Eastlake, 2001); and physical mechanisms. For devices with weak processing capabilities like smart-card, elliptic curve-based systems such as ECDSA (ANSI, 1998) can be used. Regarding physical security, using cryptographic algorithms one can only reduce the amount of data that has to be physically protected, but the physical protection cannot be avoided.

To ensure that a particular public key indeed belongs to the claimed person, a trusted third party called *certification authority*, or CA, has to be introduced. The CA issues *public key certificates* that are digitally signed electronic documents, which bind entities to the corresponding public keys (certificates can be verified by CA's public key). CA also maintains certificate revocation lists, or CRL that should be checked every time a certificate is processed in order to ensure that a private/public key is still valid. The de iure and de facto standard for certificate format is X.509 standard (ITU-T, 2000).

By use of security mechanisms, the following security services are implemented:

- Authentication: Ensures that the peer communicating entity is the one claimed.
- Confidentiality: Prevents unauthorized disclosure of data.
- Integrity: Ensures that any modification, insertion, or deletion of data is detected.
- Access Control: Enables authorized use of resources.
- Non-Repudiation: Provides proof of origin and proof of delivery, where false denying of the message content is prevented.

- Auditing: Enables detection of suspicious activities and analysis of successful breaches, and serves as evidence when resolving legal disputes.

To enable these services, a certain infrastructure has to be set up. It includes a Registration Authority (RA) that serves as an interface between a user and CA, identifies users, and submits certificate requests to CA. In addition, a synchronized time base system is needed for proper operation, along with a global directory for distribution of certificates and CRLs. All these elements, together with appropriate procedures, form a so-called *public key infrastructure* or PKI (Arsenault, 2002).

To provide security, mostly commercial off-the-shelf solutions are used. Such solutions typically include firewalls, which are specialized computer systems that operate on the border between the corporate network and the Internet, where all traffic must pass through these systems (Cheswick & Bellovin, 1994). Further, real-time intrusion detection systems (Kemmerer, 2002) are deployed for detecting acts that differ from normal, known patterns of operation, or for detecting wrong behavior. Further, IPSec (Thayer, 1998), which is a security enhancement for IP protocol, can be used to prevent masquerade, monitoring of a communication, modification of data, and session overtaking. IPSec is suitable for Virtual Private Networks or VPNs, where one can establish secure private networks using public networks such as the Internet. Further, Secure Sockets Layer protocol or SSL (Freier, 1996) provides a common security layer for Web and other applications, and is available by default in Web browsers. It provides authentication, confidentiality, and integrity with the possibility of negotiating crypto primitives and encryption keys. Further, Secure/Multipurpose Internet Mail Extensions standard or S/MIME (Ramsdell, 1999) is often deployed as security enhancement for ordinary e-mail, and provides authentication, confidentiality, integrity, and non-repudiation.

Finally, the security of new paradigms has to be covered. These paradigms include objects, components, mobile code (computing), and *intelligent agents*. Every code (and object) can be treated as an electronic document. The creator defines its initial data and behavior (methods) and, optionally, signs it. The signature on the code gives a user the possibility to be assured of proper functioning of this object, where the problem is analogous to that of ensuring authentication and integrity for ordinary electronic documents. When considering intelligent mobile agents that are objects that satisfy certain conditions (Griss, 2001), the security paradigm is reversed. Agents operate in unpredictable environments and have to be protected from malicious hosts. These important issues have yet to be resolved (FIPA, 2001).

FUTURE TRENDS: ORGANIZATIONAL PLANE OF IS SECURITY MANAGEMENT

Even superior technological solutions will be in vain, if the complementary organizational issues are not treated properly. Therefore the second plane must be concentrated on organizational issues through human resources management, which has to be properly embodied in *security policy*. The basic standard in this area is BS 7799 (BSI, 1999), which recently became an international standard (ISO, 2000). It presents the main methodology, which is followed by the growing number of organizations for establishing security policy. It consists of two parts. The first part describes a code of practice for information security management, while the second specifies information security management systems.

This standard plays a central role as far as security policy is concerned. However, to implement successfully a concrete security policy, it is essential to support managers of intelligent enterprises with appropriate techniques. The organizational plane is characterized by a complex interplay between human factor and technology. The two constituent parts are coupled in many ways, such as by interactions - a large number of these interactions form various feedback loops. There are also soft factors that have to be taken into account, for example, human perception of various phenomena like trust. Therefore, to support decision making properly with regard to security, one has to deal with physical and information flows. Additionally, decisions are often to be made in circumstances where there is not enough time or resources to test decisions in a real environment; often such checks are not possible at all. Therefore support from computer simulations is highly desirable.

The methodology that can be used to support the resolution of the above-mentioned problems is *business dynamics* (Sterman, 2000). It enables qualitative and quan-

Table 1. Summary of basic security-related elements—technological plane

- *Security Mechanisms:* Symmetric and asymmetric algorithms, one-way hash functions, physical mechanisms
- *Security Services:* Authentication, confidentiality, integrity, non-repudiation, access control, auditing
- *Security Infrastructure:* Public key infrastructures, commercial off-the-shelf solutions (firewalls, intrusion detection systems, IPSec, SSL, S/MIME)

titative modeling of contemporary IS. Using this methodology, one starts with the identification of variables that are relevant to system behavior and defines the boundary distinguishing between endogenous and exogenous variables. Variables are connected by causal links which have positive polarity if increase/decrease of input variable results in increase/decrease of output variable. In a case when increase/decrease of input variable results in decrease/increase of output variable, the polarity is negative. Some variables are of a stock nature (also called levels), and these introduce persistency (inertia) to the system. They decouple inflows and outflows, and thus present a kind of buffer or absorber. When building a model of a system by this approach, so-called causal loop diagrams are obtained which provide a very useful means for management of e-business systems security. Finally, causal loop diagrams are upgraded with appropriate equations and tuned using real data to enable quantitative analysis, that is, computer simulation.

Business dynamics can be further enhanced by linking it to *business intelligence* (Ortiz, 2002) in order to achieve an appropriate architecture for qualitative and quantitative management of security in intelligent enterprises. The basis is operational security-related data that comes from various sources: general host logs, router logs, application logs, phone and fax logs, physical security system logs, and so forth. These operational, legacy data have to be transformed in the next step, which means preparation of data for a security data warehouse. The exact data that have to be captured in a security warehouse are defined in line with the causal loop diagram model of an organization. This also implies that such a solution will have to be tailored to the needs of each particular organization.

Summing up, security policy is the main document about risk management in every organization, and its introduction and maintenance has to be adequately supported by use of modern business intelligence techniques. Using causal loop diagramming, decision makers can get a holistic perspective on their systems, but to verify and properly adjust these models, they have to link them to appropriate sources to obtain real-time data for simulations. The resulting architecture provides an important tool for information security management of intelligent enterprises with emphasis on the human factors that are critical in the context of security policy.

CONCLUSION

Even in the era of intelligent enterprises, security management cannot avoid technical foundations. It is a fact that classical, cryptography-based approaches form the core of all security management, so that each IS security has to start with addressing these issues. However, this basis serves as a starting point for further development of methodologies for risk management that are concentrated on human resources. Experience shows that human factors play an increasingly important role. Taking this into account and due to the emergence of business intelligence, it is possible to further support the management of IS security. Using business dynamics and business intelligence techniques, decision makers can obtain data in real time and simulate the effects of their decisions in advance. This way they are armed with additional tools for successful protection of their IS.

REFERENCES

ANSI. (1998). *The Elliptic Curve Digital Signature Algorithm (ECDSA). X9.62 Standard.* Washington, DC: ANSI.

Aresenault, A. et al. (2002). *Internet X.509 public key infrastructure roadmap. PKIX Draft Standard.* Reston, VA: IETF.

BSI. (1999). *Code of practice for information security management. British Standard 7799.* London: British Standards Institute.

Cheswick, W. & Bellovin, S. (1994). *Firewalls and Internet security*. Reading, MA: Addison-Wesley.

Eastlake, D. & Jones, P. (2001). *Secure Hash Algorithm–1. RFC 3174 Standard.* Reston, VA: IETF.

Foti, J. (Ed.). (2001). *Advanced encryption standard. FIPS 197 Standard.* Washington, DC: DoC.

Foundation for Intelligent Physical Agents. (2001). *FIPA security SIG request for information. F-OUT-00065 Deliverable.* Concord, MA: FIPA.

Freed, N. (1996). *Multipurpose Internet mail extensions. RFC 2045 Standard.* Reston, VA: IETF.

Table 2. A summary of advanced security-related elements—organizational plane

- *Security Policy:* Concentration on human resources
- *Business Dynamics:* Quantitative and qualitative modeling of organizations for security policy
- *Business Intelligence:* Enabling integral capturing of data with knowledge extraction to support business dynamics

Freier, A.O. et al. (1996). *Secure Sockets Layer Protocol* (version 3). Mountain View, CA: Netscape Corp. Retrieved from *wp.netscape.com/eng/ssl3/index.html*

Griss, L.M. (2001). Accelerating development with agent components. *IEEE Computer, 5*(34), 37-43.

ISO. (1995). *IT, open systems interconnection: Security frameworks in open systems. IS 10181/1 thru 7*. Geneva: ISO.

ISO. (2000). *Code of practice for information security management. ISO 17799 Standard.* Geneva: ISO.

ITU-T. (2000). *Public key and attribute certificate frameworks. X.509 Standard.* Geneva: ISO.

Kemmerer, R.A. & Vigna, G. (2002). Intrusion detection: A brief history and overview. *IEEE Computer, Security & Privacy, 35*(5), 27-30.

Miller, S.K. (2001). Facing the challenge of wireless security. *IEEE Computer, 34*(7), 16-18.

Ortiz, S. (2002). Is business intelligence a smart move? *IEEE Computer, 35*(7), 11-15.

Raepple, M. (2001). *Sicherheitskonzepte fuer das Internet.* Heidelberg: dpunkt-Verlag.

Ramsdell, B. (1999). *S/MIME message specification. Standard RFC 2633.* Reston, VA: IETF.

RSA Labs. (2002). *PKCS–RSA Cryptography Standard, v 2.1*. Bedford: RSA Security.

Sterman, J.D. (2000). *Business dynamics*. Boston: Irwin-McGraw-Hill.

Thayer, R. et al. (1998). *IP security document roadmap. RFC 2411*. Reston, VA: IETF.

KEY TERMS

Business Intelligence: Deployment of (usually artificial intelligence-based) techniques such as On-Line Analytical Processing and data mining to analyze information in the operational data sources.

Certification Authority: An authority trusted by one or more users to create and assign public key certificates.

E-Business System: An organized, structured whole that implements business activities that are based on electronic technologies, methodologies, and processes.

Networked Information Systems: An IS that is strongly integrated in a global network. From the technological point of view, the difference between the IS and the global network is blurred; however, it exists from the administrative point of view.

Public Key Certificate: The public key of a user, together with some other information, rendered unforgeable by encipherment with the private key of the certification authority that issued it.

Public Key Infrastructure: The infrastructure capable of supporting the management of public keys able to support authentication, encryption, integrity, or non-repudiation services.

Security: A set of processes aimed at minimizing the vulnerability of assets and resources.

Security Mechanism: A basis for a security service—using particular security mechanism (e.g., cryptographic algorithm), the security service is implemented.

Security Policy: These are documented procedures that focus on the organization's management of security; it is about information confidentiality, integrity, and availability of resources.

Security Service: A service provided by an entity to ensure adequate security of data or systems.

System Dynamics: An interdisciplinary method to enhance learning of complex systems that is based on nonlinear dynamics and feedback control.

E-Business Transaction in Web Integrated Network Environment

V.K. Murthy
University of New South Wales, Australia

E.V. Krishnamurthy
Australian National University, Australia

INTRODUCTION

The World Wide Web (or the "Web") is revolutionizing the concept of service -oriented computing. It permits the integration of the local area networks (LAN) and wide-area networks (WAN), thereby providing servers that are interconnected at a worldwide or an intergalactic level. Such an integration serves as a backbone for the Web-based business or e-business to access information, as well as perform e-business transactions across the globe with adequate security and reliability. In this article, we describe the issues involved in the design of an e-business transaction processing system and the solutions that have been proposed for these problems using the techniques of AI, conventional database transaction processing methodology and protocol engineering principles. These techniques will be useful for improved transaction throughput and scalability in e-commerce (Brancheau & Shi, 2001; Menasce & Almeida, 2000; Murthy, 2002).

BACKGROUND

Transaction management is a well-established research area with many successful results having been achieved so far. Transactions that have ACID (atomicity, consistency, isolation and durability) properties have traditionally been used to ensure consistent database management through atomicity (all or none) of actions as well as isolation of user actions (Elmagarmid, 1995; Krishnamurthy & Murthy, 1992). Recent advancements in transaction management have relaxed some of these traditional properties of transactions. The appropriate properties for transactions in a *Web-integrated network environment* (we use the acronym WINE, for brevity) have been studied recently by Chen and Dayal (2000), Ghezzi and Vigna (1997), and Murthy (2002). These papers also discus issues and problems that mobility brings into transaction management, and describe new methods using workflow and mobile agents and various types of new software tools currently available such as Java, Java database connectivity and CORBA (Common Object Request Broker Architecture-Dignum & Sierra, 2001). A distributed multi-database system with many autonomous and heterogeneous component databases will provide support for the management of global e-business (also known as "intergalactic") transactions and data resources in a WINE with mobile and stationary hosts.

The WINE consists of a collection of dissimilar (heterogeneous) computers (fixed host computers (FH) and mobile client computers (MC)) connected through the fixed-wired networks (such as the Ethernet), as well as wireless (mobile) networks connected together via a fixed high-speed network (Mbps to Gbps). The mobile clients (MC) are capable of connecting to the fixed network via a wireless link. Fixed host (FH) provides mobile application services and coordinates tasks to mobile hosts. MC supports query invoking and information filtering from FH to provide personal information service. Since the computers may have different computational powers and may use different representations for data we need to take care of not only the incompatibility among their representations, but also about their interoperability in using different pieces of software. Also, to achieve high performance and reliability (that provides maximal concurrency and recovery under failure), we need suitable computational models to help understand and analyze their behavior.

MAIN THRUST OF THE ARTICLE

The following subsections bring out the main thrust of the chapter, and deal with the following aspects: requirements for e-business transactions, operational models for e-business systems, intergalactic computing for e-business, new logical modes for e-business environment, communication bandwidth management, agent technology, and e-business protocols.

REQUIREMENTS FOR E-BUSINESS TRANSACTIONS

E-business transactions need to have the following properties: attribute-sensitivity, attribute tolerance, time-criticality, and time-tolerance and eventual consistency. Attribute and time tolerant, and time critical transactions arise in planning or "what-if" programming mode (subjunctive) where we execute hypothetical or pseudo-transactions to test the intention of actions for trial-error design. For example, such transactions arise in real-time transactions in e-shopping carts.

The e-shopping cart holds a record of the selection the buyer intends to buy. At any point the buyer can review the items, remove items, or change their quantity, type and brand. It is useful to have the shopping cart in place even if the buyer leaves the Web site to do something else and come back later. Such a persistent e-shopping cart is very useful for grocery shopping. This of course requires a deadline on the availability, as well as pricing, since some items can go up or down in pricing. Thus the buyer does not only control the items selected, but the shop automatically changes its current availability and pricing at the time when intention commit becomes an action commit. The time-tolerant and eventual consistency property is used in e-tailing. To handle time-criticality, for example, inconsistency on deadlines, we need to ensure that the fixed host can meet the required deadline by determining a priori whether an incoming transaction or a part of a transaction from a mobile host is schedulable within that deadline. A bounded amount of inconsistency may be introduced to finish a task within a deadline or accept inconsistency only when the transaction is about to miss the deadline. This can be specified by condition-event-action or rule based systems. Thus in e-business, the traditional transaction model needs to be replaced by a more realistic model called "workflow" between the customer and the trader (Geppert, 2000; Murthy, 1998, 2001).

OPERATIONAL MODELS FOR E-BUSINESS SYSTEMS

Many applications in e-business require several different services: file service, transaction processing, database access, and applications including graphics. Since no one system can be versatile enough to provide all these services, we need to bring in a client-server relationship among the consumer and the service provider so that the required server processes run in separate machines and the client requests any one of these or a combination of these services through a network of interconnected servers. Typically, these services are built using middleware packages that tie the client to the servers by coordinating client-server interactions. A beautiful property of the client-server relationship is that it is flexible and malleable.

In fact, the distributed applications can be split between client, as well as server by appropriately adding extra functions to either one of them, making them "fat clients" or "fat servers" or even peers so that they are equals. Hence the name peer-to peer or P2P computing is currently widely used.

As examples of fat client and fat server, we can recall the functions such as Code on Demand (COD) and Remote Evaluation (REV) used in earlier days. The COD represents a simple fat client; here, to obtain a single value from a table of data, the searching routine is transmitted from the client to the remote server. In REV we have the simplest "fat server"; here, to obtain a single value from a table of data, the searching routine is implemented in the server and this routine is activated by remote invocation or by a trigger. In fat clients the bulk of the applications run in clients, while in fat servers the bulk of the applications run in several servers. To make the task easier, the two-tier client server is replaced with three or more tiers, where the middle tier or tiers handle the application services.

For example, in the N-tier system (NTS), the middle tier is not implemented as a monolithic program, but as a collection of components that are used in a variety of client-initiated business transactions. Each such component realizes a small business function.

The three-tier and N-tier models provide for excellent security, better performance, ease of development, hardware flexibility, rich communication choices and heterogeneous functionality support (Orfali, Harkey & Edwards, 1996, 1999).

INTERGALACTIC COMPUTING FOR E-BUSINESS

To develop a suitable inter-galactic computing model for e-business the client-server needs to provide a very high bandwidth for WINE so that it provides an information highway for an electronic bazaar. This requires handling architectural heterogeneity between communicating machines. Also appropriate choice of programming paradigms and related software tools are to be made available to the mobile host so that clients can install special purpose interfaces with appropriate properties they require at the remote fixed-host.

Some key technologies necessary for these developments are:

a. ***New logical modes of computation:*** These modes support a variety of trial and error transactions, long duration diagnostic transactions, and speculative business-to-business transactions (B2B).

b. ***Bandwidth requirements:*** Very high bandwidth is needed to connect to the access point on the WAN, high bandwidth for connecting wireless link to the LAN or WAN.

c. ***Agent technology:*** An agent is a code-containing object along with the data and the execution context. It can travel through the network, collect information, gather statistics, and perform a variety of tasks across different hardware platforms.

NEW LOGICAL MODES FOR E-BUSINESS ENVIRONMENT

An e-business environment needs to be rich in its problem solving capability, since it needs to serve as a virtual logical tool for the user. Such a tool is meant for effective decision making using different logical modes to cooperatively solve a business problem. Most conventional programming uses two types of logical constructs: declarative (declaration or assertion oriented) and imperative (command oriented). In a e-business environment, in addition to these constructs we need two other logical modes - subjunctive (what if I do this? or speculative) and abductive (how did this happen? or diagnostic) programming features that add additional power to the user and provide for various forms of reasoning to aid planning, analyzing, acquiring and arguing.

The subjunctive logical mode arises in practical situations such as: planning, reservation, purchases and forecasting. Abduction is useful in e-business as a diagnostic tool, especially in a failure diagnostic situation. For the subjunctive and abductive modes, the conventional transactional model is unsuitable due to ACID properties and serializability requirements. Atomicity is very restrictive, since all or no operations should take place indivisibly as one entity. Also, consistency may not be achievable in subjunctive and abductive reasoning situations. In addition, we may have to relax the isolation property, since it prevents each transaction from looking into what other transactions are doing at intermediate stages; here we need data exchanges to achieve consistency. Durability and recovery are to be supported under failures and voluntary disconnection and reconnection of the mobile hand fixed hosts and have to take place through data exchange that is not atomic. Thus disconnection and reconnection are essential functions and protocols are to be designed using suitable logs and recovery mechanisms.

COMMUNICATION BANDWIDTH MANAGEMENT

Bandwidth determines the information capacity of a network per unit of time. Text requires the lowest bandwidth, while audio and video data and signals require significant increase in bandwidth. Specification of bandwidth to be used and compression techniques used are to be laid down so that the data and images that are transmitted for e-business should be of high quality for advertising and marketing so as to serve their purpose appropriately.

Wireless networks deliver lower bandwidth than a wired network. Hence software techniques based on compression should be used for effective communication. Also scheduling communication intelligently can save bandwidth. For use in e-market, such techniques should be extremely reliable.

AGENT TECHNOLOGY

An agent is a code-containing object that along with data and execution context can migrate autonomously and purposefully within a computer network. Thus an agent knows what to do with the information obtained from its environment. They behave like actors and have intentions and actions. In addition, agents are flexible, proactive and have multithreaded control. These properties enable the consumers to place agents to look after their interests and business organizations to deploy agents to sell their products. Agent technologies are also useful for cross-platform operation and in efficient management of communication and the consequent reduction in bandwidth requirements, security enforcement, error-correction, transaction and notification services, and numerous other services (Nagi, 2001; Woolridge, 2002).

E-BUSINESS PROTOCOLS

We mentioned that the traditional transactional model is not well suited in WINE, since data consistency, deadlines and scheduling are to be taken care of. In such systems it is preferable to have a negotiation protocol between the client and host to carry out tasks reliably. This can go through two phases, called intention phase and action phase. These phases enable us to carry out subjunctive or “what-if” programming in which we execute hypothetical or pseudo-transactions to test the intention of each party and then perform actions. The subjunctive programming paradigm is widely applicable in mobile computing, as it allows the mobile client and fixed host to interact and negotiate and carry out transactions, workflows, as well as connection - disconnection protocols and propose-revise protocol for the abductive mode of tele-reasoning.

FUTURE TRENDS

New Software Tools

There are several available languages (Knabe, 1996; Yourdon, 1996) for programming the different operational models. Java seems to be ideally suited to our requirements. Java is a complete programming language, which offers all the basic mechanisms for communication and synchronization between processes and guarantees portability of the code across multiple architectures and operating systems in a high performance secure way. The distributed networking nature of Java permits the development of all the six models of distributed systems. Also the Java Database Connectivity Application Program interface (JDBC API) provides for simultaneous connection to several databases, transaction management, simple queries, and manipulation of precompiled statements.

Agent-based technology will provide new approaches to e-business transactions. Mobile agents in Java can also be very useful in mobile computing. Java is very suitable for CORBA (Common Object Request Broker Architecture) application development since Java and CORBA are similar in architecture; this similarity in architecture results in the near seamlessness between Java and CORBA (Lucena et al., 2004; Orfali, Harkey & Edwards, 1996; Rosenfeld & Morville, 1998; Siegel, 1996). Java provides to CORBA the capability to develop a client-side application once and run it on many platforms. Also, CORBA provides the benefit of cross-language interoperability to Java. CORBA provides an object-oriented abstraction that permits Java-based systems to communicate with applications written in almost any language.

CONCLUSION

We described the key issues involved in the design of online e-business transaction processing systems in a Web integrated network environment (WINE) and the solutions available for these problems using the techniques of AI, conventional database transaction processing methodology and protocol engineering. Agent technology will play a key role in e-business transaction management.

REFERENCES

Brancheau, J., & Shi, N. (2001). *Essential technologies for e-commerce*. Singapore: Prentice Hall.

Chen, Q., & Dayal, U. (2000). Multi agent cooperative transactions for e-commerce. *Lecture Notes in Computer Science, 1901,* 311-322. New York: Springer Verlag.

Dignum, F., & Sierra, C. (Eds.). (2001). Agent mediated e-commerce. *Lecture Notes in Artificial Intelligence, 1991, 2003*. New York: Springer Verlag.

Elmagarmid, A.K. (1995). *Database transaction models*. San Mateo, CA: Morgan Kaufmann.

Geppert, A. (2000). Modeling electronic workflow markets. *Lecture Notes in Computer Science, 1921,* 52-63. New York: Springer Verlag.

Ghezzi, C., & Vigna, G. (1997). Mobile code paradigms and technologies, A case study. *Lecture Notes in Computer Science, 1219,* 39-49. New York: Springer Verlag.

Knabe, F. (1996). An overview of mobile agent programming. *Lecture Notes in Computer Science, 1192,* 100-115. New York: Springer Verlag.

Krishnamurthy, E.V., & Murthy, V.K. (1992). *Transaction processing systems*. New York: Prentice Hall.

Lucena, C. et al. (2004). Software engineering for multi-agent systems II. *Lecture Notes in Computer Science, 2940*. New York: Springer Verlag.

Menasce, D.A., & Almeida, V.A.F. (2000). *Scaling for e-business*. New York: Prentice Hall.

Murthy, V.K. (1998). Transactional workflow paradigm for mobile computing. *Proc. 13th ACM Conference on Applied Computing* (pp. 424-432). New York: ACM Press.

Murthy, V.K. (2001). Seamless mobile transaction processing: Models, protocols and software tools. *Proc. Eight IEEE Intl. Conf. on Parallel and Distributed Systems* (pp. 147-154). CA: IEEE Computer Society Press.

Murthy, V.K. (2002). E-business transaction management in Web-integrated network environment. In *Architectural issues of Web-enabled electronic business* (pp. 198-213). Hershey, PA: Idea Group Publishing.

Nagi, K. (2001). Transactional agents. *Lecture notes in Computer Science, 2249*. New York: Springer Verlag.

Orfali, R., Harkey, D., & Edwards, J. (1996). *The essential distributed objects*. New York: John Wiley.

Orfali, R., Harkey, D., & Edwards, J. (1999). *Client/server survival guide*. New York: John Wiley.

Rosenfeld, L., & Morville, P. (1998). *Information architecture for the World Wide Web*. Sebastopol, CA: O'Reilly.

Siegel, J. (1996). *CORBA*. New York: John Wiley.

Woolridge, M. (2002). *An introduction to multi-agent systems*. New York: John Wiley.

Yourdon, E. (1996). Java, the Web and software development. *IEEE Computer, 29*, 25-39.

KEY TERMS

ACID Properties: Conventional online transaction processing (OLTP) requires the following properties called "ACID properties":

- **Atomicity (A):** All changes are totally done (committed) or totally undone (rolled back).
- **Consistency (C):** The effect of a transaction preserves the invariant properties of the system.
- **Isolation (I):** Intermediate results are not visible to other transactions. Transactions have the effect of executing serially, although they act concurrently.
- **Durability (D):** The effects of a transaction are persistent; changes are not lost except under catastrophic failure.

Agent: A system that is capable of perceiving events in its environment, or representing information about the current state of affairs and of acting in its environment guided by perceptions and stored information (current definition by AOIS, agent oriented information system community).

CORBA - Common Object-Request Broker Architecture: This is an object oriented architecture that provides a standard mechanism for defining the interfaces between components, as well as tools to facilitate the implementation of interfaces using the developer's choice of languages, thus providing for language and platform independence.

Programming Modes

- **Abductive Mode:** This mode permits programs to explore "how did this happen?" or provides for diagnostic programming features.
- **Declarative Mode:** Declaration oriented languages-such as Prolog, using declarative statements such as: x is greater than y.
- **Imperative Mode:** Command oriented languages-such as Fortran, Cobol, using commands, DO, GO TO, and so forth.
- **Subjunctive Mode:** This mode permits programs to do speculation, namely, "what if I do this?" mode.

Protocols: A set of rules that dictate the behavior of objects for communication and interaction.

Transactions

- **Attribute-Sensitive:** The transactions are to be committed or aborted locally or globally depending upon their exact attribute values.
- **Attribute Tolerant:** The transactions can be permitted to be locally relaxed in terms of certain constraints on attributes, but globally consistent eventually.
- **Time-Tolerant and Eventually Consistent:** Some transactions can wait until reconnection takes place, and are not time- critical, in the sense they will not create global inconsistency, but are only necessary to provide an eventual consistency with respect to the user and the relevant database.

Workflow: A collection of tasks organized to accomplish some business activity between the customer and the trader supported by suitable protocols.

E

E-Collaboration Support Systems: Issues to be Addressed

Nikos Karacapilidis
University of Patras, Greece

INTRODUCTION

Removal of communication impediments and provision for techniques that systematically direct the pattern, timing, and content of cooperative processes are two key prerequisites in the contemporary organization. Their establishment has been proven to facilitate the solution of ill-structured problems by a set of individuals working together as a team, through the interactive sharing of information between them. E-collaboration involves a variety of both communication and cooperation issues, in that it leverages the connective powers of a computer network to coordinate the efforts of a group of people. By using e-collaborative capabilities in an organization, people can operate as a single business entity, thus making joint decisions of added value.

Issues to be addressed in the establishment of an e-collaboration environment should have a strong organizational focus. These include work structuring in order to improve coordination, use of communication technology to make collaboration more efficient and effective, enforcing of rules and procedures for achieving consistency, and automating data processing in data intensive situations. One should further consider the conceptual, methodological, and application-oriented aspects of the problem. Conceptual focus is associated with the consideration of the nature of individual and organizational processes, methodological focus with the integration of existing computer-based tools, techniques and systems into the human decision making context, and application-oriented focus with the consideration of the real organizational needs by extending decision support to business teams (Angehrn & Jelassi, 1994).

BACKGROUND

The environment in which a collaborative process takes place sets different communication requirements. Issues to be taken into account in the design and implementation of an e-collaboration system include the following:

- The *spatial distance* between team members. This refers to whether full face-to-face communication among team members is possible. Depending on the group size and the proximity of members during a decision-making procedure, the following settings have been identified (DeSanctis & Gallupe, 1987): (i) the *decision room*, where an electronic version of a traditional meeting situation is established; (ii) the *legislative session*; (iii) the *local area decision network*, where group participants can communicate with each other and with a central processor through a local-area network; and (iv) the *computer-mediated conference*, where communication is provided between two or more remote groups by linking decision rooms together through audio and video facilities.
- The *temporal distance* among the activities performed by the individual group members. This refers to whether collaboration is taking place through meetings at a particular time, such as in conventional meeting or teleconferencing environments, or whether participants submit their input at different points in time, based on electronic mail, bulletin boards, newsgroups, and computerized conferencing concepts.
- The *type of participants' goals* distinguishes between an environment in which a group wants to solve its common problem cooperatively, and another in which bargaining takes place. Issues arising in the first case concern knowledge sharing, preference aggregation, and negotiation support. Depending on the degree of cooperativeness among the decision makers, three modes of reaching a decision have been reported (Jelassi & Foroughi, 1989): (i) the *pooled mode*, where there is so much cooperation that the individuals act almost as a single decision maker; (ii) the *cooperative mode*, where decision makers may have difficulties in understanding and accepting each other's positions; and (iii) the *noncooperative mode*, where a series of negotiations must integrate the diverse, often conflicting and incompatible individual problem representations into a common solution.
- The type of *control* over the collaborative process. There may be cases where the participants follow a democratic process in order to reach a solution, and

cases where the system is supported by a human group leader or mediator. In the former, communication and coordination are achieved by the users or directly by the system. The latter can be further distinguished in those where the human mediator cannot impose decisions on the participants, and those where there is compulsory arbitration from a group leader. Referring to a Group Decision Support System, for instance, three levels of control have been identified, namely, democratic participative decision making, semi-hierarchical decision making, and third-party arbitration (Jelassi & Foroughi, 1989).

- *Separating people from the problem*. The system designer has to evaluate the individual and group characteristics of the participants, as well as their motivations, disagreements, and conflicts, in order to reduce (if not avoid) the negative impact that misunderstandings, emotions, and bad communication may have. Different approaches to conflict resolution include (i) *contending or positional bargaining*, where a party is trying to convince the opponent(s) to accept its favorite position; (ii) *accommodating*, involving a party's effort to help another party meet its objectives; (iii) *compromising*, meaning a splitting of the differences between interested parties, that is satisfying but not optimizing; (iv) *collaborating*, involving parties working together to optimize their joint outcome, such as in group problem solving settings; and (v) *avoiding* the negotiation process for various reasons such as fear of conflict, not worth bargaining issues, or intention of negotiations' postponement (Lewicki & Litterer, 1985).
- The *type of communication* between the participants. Collaborative environments can be based either on *point-to-point communications* or on *broadcasting* of messages.

Furthermore, approaches for the development of a framework for e-collaboration have to address both behavioral and technical aspects (Zigurs et al., 1988). Behavioral issues concern the diffusion of responsibility, pressures toward group consensus, and problems of coordination. A framework that integrates behavioral and technical perspectives may reduce the negative impact and enhance the positive effects of the former ones. Issues involved in the design of such a framework are (i) support (or not) of anonymity, depending on the type of discussion; (ii) enforcement of participants' self-awareness; (iii) display of group inputs at any stage of the discussion; (iv) structure of the decision process (the actions the participants should follow may improve the efficiency of the system in terms of accuracy and response time); (v) ability to support communication, information sharing, and democratic control (provision of communication and information sharing helps participants to create a shared workspace on which the discussion will be based.

COMPUTER-SUPPORTED COOPERATIVE WORK

Computer-supported cooperative work (CSCW) has been defined as computer-assisted coordinated activity, such as communication and problem solving, carried out by a group of collaborating individuals (Greenberg, 1991). The multi-user software supporting CSCW is known as *groupware* (Ellis et al., 1991). Sometimes this term is broadened to incorporate the styles and practices that are essential for any collaborative activity to succeed, whether or not it is supported by a computer. CSCW may also be viewed as the emerging scientific discipline that guides the thoughtful and appropriate design and development of groupware (Greenberg, 1991). Key issues of CSCW are group awareness, multi-user interfaces, concurrency control, communication and coordination within the group, shared information space, and the support of a heterogeneous open environment that integrates existing single-user applications.

The most successful CSCW technology to date is undoubtedly electronic mail. Other well-developed technologies so far comprise computer conferencing, teleconferencing or desktop videoconferencing (the act of conferencing at a distance with the aid of audio and video links), group authoring (enabling cooperative writing with additions, revisions, comments, and annotations), and group decision support systems (problem solving is directed at the organization of the issues involved). The last category comprises mediating systems that support discussion, argumentation, negotiation, and decision-making in groups.

As illustrated in Table 1, most taxonomies of CSCW technologies distinguish them in terms of their abilities to bridge time and space (the table is a more elaborate version of the one appearing in Baecker (1993). As cited in Baecker (1993), groupware technologies of the future need to span all quadrants of this table. This is usually described as anytime/anyplace groupware. During the last few years, CSCW has been strongly supported and explored by both industry and academic research. Everybody speaks for the shifting role of computers in that they do not merely handle information processing issues, but they appear as tools for managing commitments and their fulfillment and as tools for producing and "listening to" the assertions and assessments that structure the organization (Winograd, 1992). Computers can make explicit the structure of human interaction in an organization,

Table 1. A taxonomy of CSCW technologies

	Synchronous Communications	**Asynchronous Communications**
One group site	Electronic Meeting Facilitation Decision Rooms	Media spaces Desktop Conferencing
Multiple individual or group sites	Teleconferencing Desktop Videoconferencing Broadcast Seminars	Electronic-Mail Voicemail Collaborative Writing Workflow Management Group Decision Support Cooperative Hypertext

providing new operational means for generating and monitoring workflows, being a more effective observer in what is going on, and determining what is needed when, for whom, and for what is to be done.

A principal aim for the designer of an e-collaboration framework is to apply state-of-the-art telematics and groupware technology to provide advanced support for the users over wide area networks—in particular, the Internet. Generally speaking, CSCW tools can harness the complexity of the social and knowledge processes involved, thus providing benefits in terms of speed and accuracy, and facilitating the development of business policies. Such tools can be used to support the group reasoning processes (i.e., to facilitate the evaluation of proposed solutions and their support; to structure the decision-making process through the implementation of specific methodologies; and to help group members to reach a shared understanding of the issue by supporting knowledge elicitation, knowledge sharing, and knowledge construction). Moreover, by exploiting intranet or Internet technologies, they can connect participants with similar interests, encouraging dialogue and stimulating the exchange of knowledge.

A plethora of systems that support capturing of decision rationale and argumentation for different types of user groups and application areas has been already developed. For instance, QuestMap, which is based on gIBIS hypertext groupware tool (Conklin & Begeman, 1987), can capture the key issues and ideas during meetings and attempt to create a shared understanding by placing all messages, documents, and reference material for a project on a "whiteboard." Euclid (Smolensky et al., 1987) is another system in this category, which provides a graphical representation language for generic argumentation. On the other hand, Janus (Fischer et al., 1989) is based on acts of critiquing existing knowledge in order to foster the understanding of design knowledge. Sepia (Streitz et al., 1989) is a knowledge-based authoring and idea-processing tool for creating and revising hyperdocuments that view authoring as a design process. QOC (Questions, Options and Criteria), based on a representation model of the rationale of reasoning in a decision-making process, allows users to represent and integrate rationale of varying degrees of stability at different stages in a design process (Shum et al., 1993). Finally, Sibyl (Lee, 1990) is a system that provides services for the management of dependency, uncertainty, viewpoints, and precedents. Generally speaking, the above systems meet the collaboration requirements concerning the type of control, conflict resolution, and behavioral issues, as discussed in the previous section, by providing a cognitive argumentation environment that stimulates reflection and discussion among participants. However, issues related to temporal and spatial distances are not fully addressed. These systems do not exploit any network infrastructure; thus, users can work in an asynchronous way only through a human mediator who receives their contributions and appropriately deploys them to the system (similar criticism holds for the display of each collaboration instance to all parties involved). Most important, this category of systems does not integrate any reasoning mechanisms to (semi)automate the underlying decision making and negotiation processes.

Increasing interest also has been developed in implementing Web-based conferencing systems, such as AltaVista Forum Center, Open Meeting and NetForum. Such systems exploit the platform-independent communication framework of the Web, as well as its associated facilities for data representation, transmission, and access. They usually provide means for discussion structuring and user administration tools, while the more sophisticated ones allow for sharing of documents, online calendars, embedded e-mail, and chat tools. Discussion is structured via a variety of links, such as simple responses or different comment types (e.g., qualify, agree, example in Open Meeting) to a previous message. This category of systems fully meets the requirements that are related to the spatial and temporal distances between members of a team. However, the above systems merely provide threaded discussion forums where messages are linked passively; this usually leads to an unsorted collection of vaguely associated comments. As pointed out by the developers of Open Meeting, there is a lack of consensus-seeking abilities and decision-making methods (Hurwitz & Mallery, 1995). Moreover, as in the previous category of systems, issues related to the appropriate storage of knowledge in order to be exploited in future collaboration settings are not addressed.

FUTURE TRENDS

We argue that services to be provided in a contemporary e-collaboration framework can be classified in three levels (see Table 2):

- The *information services* should deal with the interoperability of proprietary systems, providing efficient and cost-effective access to multimedia data in heterogeneous, distributed databases over wide-area networks. In particular, services should be included for finding relevant data and converting proprietary data to standard formats for data interchange. Additionally, these services should include ways of controlling remote servers from within compound documents and general-purpose electronic mail, conferencing systems, and hypermedia systems, such as the World Wide Web. Another major issue here concerns the provision for customized solutions that adapt to a team member's profile according to his or her preferences, abilities, experience, and collaboration mode, as well as aspects related to technical specifications of his or her platform, software available, and network connection. In order to be effective, such solutions have to remove barriers imposed by noninteroperable collaboration tools, inadequate infrastructure, undefined data sharing policies and standards, and differing priorities for presentation formats. What is often required is the generation of customized content through approaches such as document transformation, dynamic documents generation, adaptive hypermedia, and provision for personalized collaboration tools based on adaptive learning techniques that track a team member's activity and interactions with the system, analyze the feedback, and accordingly identify his or her needs or interests.
- The *documentation services* should provide a "shared workspace" for storing and retrieving the documents and messages of the participants, using appropriate document formats such as XML. As argued in Prahalad and Hamel (1990), an organization's only advantage in today's business environment is its ability to leverage and utilize its knowledge. While a firm comprises individuals and a set of objectified resources, its most strategically important feature is its body of collective knowledge (Spender, 1996). Such knowledge resides in an evolving set of assets, including the employees, structure, culture, and processes of the organization. Of these, employee knowledge (particularly tacit knowledge) is identified as the dominant one, which is decisive at all mental levels and has to be fully exploited (Nonaka, 1994). Such an exploitation refers to the transformation of tacit knowledge to codified information, which is considered a core process for economic activity and development. Security and privacy issues should be also addressed here. Moreover, controlled experimentation by simulation may augment the quality of a collaborative process by providing insight into the dynamic interactions and feedback loops formed by the problem elements (Sterman, 2000). A simulation

Table 2. E-collaboration services

Category of Services	**Purpose**
Information Services	Information search and retrieval, interoperability, adaptability
Documentation Services	Information transformation, knowledge management, meta-data, ontologies, experimentation, security, and privacy
Mediation Services	Conducting of debates, argumentation, negotiations, handling of conflicts, decision making

model can map organizational knowledge onto appropriate graphs, quantifying the problem under consideration and thus providing a clearer understanding of which alternative solution seems to be more prominent at the moment. Moreover, it can provide the means for an individual to conceptually define his or her position and perform experiments before asserting it to the "shared workspace." Taking into account the current state of the overall process, individuals may thoroughly contemplate on their next move to assure that it will have the best impact on the ongoing discussion. Finally, databases containing project documents may also become part of the collective memory of a community, facilitating the design and reuse of plans.

- The *mediation services* should regulate the group's activities and facilitate the underlying decision-making processes. Commercial workflow systems can be used to support well-defined, formal administrative procedures within organizations. Decisions should be considered as pieces of descriptive or procedural knowledge referring to an action commitment. In such a way, the decision-making process is able to produce new knowledge, such as evidence justifying or challenging an alternative or practices to be followed or avoided after the evaluation of a decision, thus providing a refined understanding of the problem. On the other hand, in a decision-making context, the knowledge base of facts and routines alters, since it has to reflect the ever-changing external environment and internal structures of the organization (Bhatt & Zaveri, 2002). Knowledge management activities such as knowledge elicitation, representation, and distribution (discussed in the previous category) influence the creation of the decision models to be adopted, thus enhancing the decision-making process (Bolloju et al., 2002).

The mediation services of the system are based on the specification of the underlying logic, the argumentation structure, and actions (i.e., duties and rights) of the team members. More specifically, mediation services should consist of the following four levels:

- the *Logic Level*, where the notions of consequence and contradiction are defined. This level formally specifies the notions of theory that will be used and provides the appropriate inference relations.
- the *Argumentation Framework Level*, where the concepts of positions, supporting arguments, counterarguments, and issues, as well as linguistic constructs for arguing about priority relationships among competing arguments are defined. The argumentation concepts at this level result in a kind of nonmonotonic formalism, founded on argumentation principles.
- the *Speech Act Level*, where the space of possible kinds of actions a participant may perform during a discussion is defined. Participants may alter the structure of the Argumentation Framework at the second level by, for example, adding and deleting claims or arguments.
- the *Protocol Level*, where norms and rules about duties and rights of the participants to perform actions defined at the previous level are specified. Such norms or protocols provide a means for structuring in advance demands for possible communication actions and should promote fairness, rationality and efficiency by taking into account the roles of participants, the types of their goals, and the type of control needed.

CONCLUSION

We have summarized a series of communication and cooperation issues to be considered in the development of systems supporting e-collaboration in the contemporary organization. Services to be provided by such systems have been classified in three levels: information, documentation, and mediation services. We argue that much more research and applied work need to be carried out on issues concerning the synergy of knowledge management and decision-making, while this should be further enhanced by providing advanced argumentation and experimentation features. Much attention should be also paid to adaptability issues by thoroughly taking into account an individual's profile during a collaborative process. Finally, we argue that more tools based on the concept of intelligent agents (Wooldridge, 2002) should be developed. Exploiting the basic characteristics of intelligent agents, such tools may perceive conditions holding in a dynamic e-collaboration environment, act with respect to these conditions, and reason to draw inferences and solve problems, thus facilitating the tasks of the individuals involved.

REFERENCES

Angehrn, A., & Jelassi, T. (1994). DSS research and practice in perspective. *Decision Support Systems, 12*, 267-275.

Baecker, R.M. (1993). *Readings in groupware and computer-supported cooperative work*. Amsterdam, the Netherlands: Morgan Kaufmann Publishers.

Bhatt, G., & Zaveri, J. (2002). The enabling role of decision support systems in organizational learning. *Decision Support Systems, 32*(3), 297-309.

Bolloju, N., Khalifa, M., & Turban, E. (2002). Integrating knowledge management into enterprise environments for the next generation decision support. *Decision Support Systems, 33*, 163-176.

Conklin, E.J., & Begeman, M.L. (1987). gIBIS: A hypertext tool for team design deliberation. *Proceedings of the Hypertext'89 Conference*, New York.

DeSanctis, G., & Gallupe, R.B. (1987). A foundation for the study of group decision support systems. *Management Science, 33*(5), 589-609.

Ellis, C.A., Gibbs, S.J., & Rein, G.L. (1991). Groupware: Some issues and experiences. *Communications of the ACM, 34*(1), 39-58.

Fischer, G., McCall, R., & Morch, A. (1989). JANUS: Integrating hypertext with a knowledge-based design environment. *Proceedings of the Hypertext'89 Conference*. New York.

Greenberg, S. (1991). *Computer-supported cooperative work and groupware*. London, UK: Academic Press.

Hurwitz, R., & Mallery, J.C. (1995). The open meeting: A Web-based system for conferencing and collaboration. *Proceedings of the 4th International World Wide Web Conference*, Boston, MA. Retrieved April 21, 2004, from *http://www.ai.mit.edu/projects/iiip/doc/open-meeting/paper.html*

Jelassi, M.T., & Foroughi, A. (1989). Negotiation support systems: An overview of design issues and existing software. *Decision Support Systems, 5*, 167-181.

Lee, J. (1990). SIBYL: A tool for managing group decision rationale. *Proceedings of the CSCW'90 Conference*, New York.

Lewicki, R.J., & Litterer, J.A. (1985). *Negotiations*. Homewood, IL: Richard D. Irwin Inc.

Nonaka, I. (1994). A dynamic theory of organizational knowledge vreation. *Organization Science, 5*(1), 14-37.

Prahalad, C.K., & Hamel, G. (1990). The core competence of the corporation. *Harvard Business Review, 68*(3), 79-91.

Shum, S., MacLean, A., Forder, J., & Hammond, N. (1993). Summarising the evolution of design concepts within a design rationale framework. *Adjunct Proceedings of InterCHI'93: ACM/IFIP Conference on Human Factors in Computing Systems*, Amsterdam.

Smolensky, P., Fox, B., King, R., & Lewis, C. (1987). Computer-aided reasoned discourse, or how to argue with a computer. In R. Guindon (Ed.), *Cognitive science and its applications for human-computer interaction* (pp. 109-162). Hillsdale, NJ: Erlbaum.

Spender, J. (1996). Organizational knowledge, learning and memory: Three concepts in search of a theory. *Journal of Organizational Change Management, 9*(1), 63-78.

Sterman, J.D. (2000). *Business dynamics: Systems thinking and modeling for a complex world*. New York: McGraw Hill.

Streitz, N., Hannemann, J. & Thüring, M. (1989). From ideas and arguments to hyper-documents: Travelling through activity spaces. *Proceedings of the Hypertext'89 Conference. ACM Press*, New York.

Winograd, T. (1992). Groupware and the emergence of business technology. In D.D. Coleman (Ed.), *Groupware 92* (pp. 69-72). Amsterdam, the Netherlands: Morgan Kaufmann Publishers.

Wooldridge, M. (2002). *Introduction to multi-agent systems*. New York: John Wiley & Sons.

Zigurs, I., Poole, M.S., & DeSanctis, G.L. (1988, December). A study of influence in computer-mediated group decision making. *MIS Quarterly*, 625-644.

KEY TERMS

Computer-Supported Cooperative Work (CSCW): A computer-assisted coordinated activity, such as communication and problem solving, carried out by a group of collaborating individuals. Key issues of CSCW are group awareness, multi-user interfaces, concurrency control, communication and coordination within the group, shared information space, and the support of a heterogeneous open environment which integrates existing single-user applications.

E-Collaboration: The process in which a set of individuals communicate through an intranet or Internet to coordinate their efforts towards the solution of a problem.

Group Decision Support System: An interactive, computer-based system that aids a set of decision-makers working together as a group in solving ill-structured problems. It enables decision-makers to analyze problem situations and perform group decision-making tasks.

Groupware: The multi-user software supporting CSCW. Sometimes this term is broadened to incorporate the styles and practices that are essential for any collaborative activity to succeed, whether or not it is supported by computer.

Intelligent Agent: A software entity that performs a set of operations on behalf of a user or another program. Such entities are embedded in computer-based information systems to make them smarter. This is usually achieved with the employment of artificial intelligence techniques.

Knowledge Management: The active management of the expertise in an organization involving collection, categorization, and dissemination of knowledge; the activity of representing and processing knowledge.

Ecological Models and Information Systems Curriculum

Arthur Tatnall
Victoria University, Australia

Bill Davey
RMIT University, Australia

INTRODUCTION

To those of us involved in research and teaching in information systems (IS) it is clear that curriculum innovation and change is complex, and anything but straightforward. The amount of control that individual IS academics have over the curriculum varies between universities. In some cases there is complete control over curriculum content, whereas in others just control over delivery with content determined externally. This chapter concentrates on the former situation but still has some relevance to the latter. All curriculum innovation is complex (Fullan, 1993) due to the involvement of a large number of human actors, but in information systems curriculum change this is particularly so due to the need to consider the part played by such non-human actors (Latour, 1996) as the technology itself.

We will argue that if you want to understand *how* IS curriculum is built, you need to use models and metaphors that relate to how people interact with each other, with the environment, and with non-human artefacts. One such approach is provided by the ecological metaphor described in this article in which we argue that systems of education may be seen as ecosystems containing interacting individuals and groups. The interactions between these will sometimes involve co-operation and sometimes competition, and may be interpreted in terms of these forces along with mechanisms for minimising energy expenditure. In this article we will examine the application of this metaphor to curriculum change in information systems.

BACKGROUND

Models of Curriculum Development

Nordvall (1982), building on the work of Havelock (1969, 1971), identifies several models for curriculum change that he suggests all have relevance, in the higher education context, at the subject, course, and institutional levels. These are:

- research, development and dissemination models;
- problem solving models;
- social interaction models;
- political and conflict models; and
- diffusion, linkage or adaptive development models (Tatnall, 2000).

Models of change based upon a process of research, development and dissemination (RDD) are probably the most common way of attempting an explanation of the process of curriculum development (Nordvall, 1982). In models like this, relying on logical and rational decisions, curriculum change depends on the use of convincing arguments based on programs of research. A rational and orderly transition is then posited from research to development to dissemination to adoption (Kaplan, 1991). These could then be considered as "manufacturing models", as they follow a fairly logical and straightforward mechanical approach with one thing leading directly to another and do not allow for or consider other influences such as those due to human interactions. If we were to accept a manufacturing model like this then we might expect some curriculum outcomes to be apparent across the world:

- As research would have shown that several specific programming languages were much more widely used and better to teach than others, all courses requiring programming would use just these few languages, and there would be no arguments regarding the best language to teach.
- As research would show the advantages of object-oriented methodologies all computing courses would teach only these and ignore other approaches.
- The content of courses around the world would be designed to achieve similar goals and outcomes, and contain similar content.
- Research would show the ideal method of teaching computing concepts and issues and classroom delivery of content would be moving towards this researched ideal. Everyone would then use these ideal delivery methods.

Figure 1. Research, development, diffusion and dissemination models

Research → Development → Production → Dissemination → Adoption

It is easy to illustrate that these predications are not borne out, in fact, as programs of study show wide variance within any given country and around the world. Many different programming languages and development methodologies are used, and a wide variety of techniques are adopted for classroom delivery. Some innovations seem to be accepted worldwide, but many are accepted only locally. Here, We will provide an alternative model that we believe better explains how IS curriculum is actually developed.

Metaphors and Models

Before proceeding however, we need to caution the reader on the limitations of models and metaphors. The dictionary describes a metaphor as a term "applied to something to which it is not literally applicable, in order to suggest a resemblance" (Macquarie Library, 1981, p. 1096). Metaphor are useful, not in giving a literal interpretation, but in providing viewpoints that allow us to relate to certain aspects of complex systems.

We contend that most curriculum models and metaphors are too simplistic to allow a useful view of a curriculum development as a complex system involving human and non-human interactions. In this regard, the ecological model offers two main advantages:

- A way of allowing for the inclusion of complexity.
- A new language and set of analytical and descriptive tools from the ecological sciences.

AN ECOLOGICAL MODEL OF CURRICULUM CHANGE

In ecology organisms are seen to operate within a competitive environment that ensures that only the most efficient of them will survive. In order to survive, they behave in ways that optimise the balance between their energy expenditure and the satisfaction they obtain from this effort. These two key principles underlie the discipline of ecology, which is concerned with the relationship of one organism to another and to their common physical environment (Case, 2000; Townsend, Harper & Begon, 2000). Habitat, ecological niches, and the exploitation of resources in predator-prey interactions, competition, and multi-species communities (Case, 2000) are all important considerations in ecology.

We have argued (Tatnall & Davey, 2002, 2003) that these ideas correspond to the process of curriculum development in that an educational system may be seen as an ecosystem, and that the interactions within this can then be analysed in terms of ecological concepts such as competition, co-operative behaviour and niche-development. Curriculum change can be interpreted in terms of mechanisms for minimising energy expenditure and decisions that individuals make about whether to co-operate or to compete.

In information systems curriculum development we should thus look at all the factors, both human and artefact, to see which could be expected to compete, and which to co-operate to become part of the surviving outcome. A non-human stakeholder such as a development tool or methodology must co-operate with the environment, compete successfully, or die out. This may mean a new curriculum element becomes incompatible with an old element and so replaces it. Alternatively it may mean that two new design tools can be used together, or that a particular curriculum element is compatible, or perhaps incompatible, with the desires and interests of a particular faculty member.

Ecological metaphors have been used in areas other than biology and IS curriculum change. An ecological framework has been used quite successfully in other areas including mathematics curriculum (Truran, 1997) and a study of the effects of violence on children (Mohr & Tulman, 2000). Ecology as a framework tells us to expect progress of a task through co-operative or competitive behaviours of the animate and inanimate factors in the environment. A factor that cannot compete or co-operate is inevitably discarded.

Ecosystems and Complexity

An ecosystem contains a high degree of complexity due to the large number of creatures and species living in it, and to the variety of interactions possible between each of these. The "ecosystem" represented by the curriculum in a university information systems department contains (at least) the following "species": lecturers, researchers, students, professional bodies, university administrators and representatives of the computer industry. The "environment" also contains many inanimate objects relevant to the formation of the curriculum, including: computers, programming languages, textbooks, lecture rooms, analysis and design methodologies, networks, laboratories, programming manuals, and so on.

Curriculum development can be seen as attempting to introduce change within an ecosystem. The problem, of course, is the large number of interested parties that must be contended with before change can be implemented. Curriculum development is more complex than resolving the conflicting needs of students, employers, academics and the academy. There is ongoing conflict between many things such as educational philosophies, pedagogical preferences, perceived resource constraints and personal issues. To investigate the interrelationships between these entities we will look now at competition, co-operation, niche formation and energy expenditure.

Competition

Competition in nature can occur both within and between species. In many species the males compete with each other for mates, while different species of fish compete for the best feeding areas. In IS curriculum we see many examples of competition, some of which are useful in determining the "fittest" topics and techniques best suited for survival (Darwin, 1958) in the curriculum, while others involve time-wasting clashes of personality between academics.

One example of competition seen in recent years in many IS departments is in programming between .Net and Java. The advocates of Java will contend that its use in producing Web-based applications and its non-proprietary nature mean that it is the best language to teach. .Net advocates, on the other hand, argue that while this may be so, .Net is easier to use and being backed by Microsoft has a considerable advantage in its use by industry. This, they contend, makes is the best vehicle to introduce students to programming. The result of this competition is, most likely, that one language will survive in the curriculum and the others die out. Similar examples can often be seen in competition between different methodologies and between software products. Most university courses now make use of Microsoft Office rather than Lotus, Word Perfect and the like, as Microsoft has clearly won the competition and become dominant in this area.

Co-operation

There are many examples of unexpected co-operation between organisms in nature: the oxpecker bird that lives with a rhinoceros, sharks and suckerfish, barnacles that attach themselves to whales, and dogs and cats living in close proximity with people. It is also possible to think of an organism living in co-operation with its environment, something the native peoples of many countries speak about.

In an educational program such as an IS degree some courses rely on earlier courses: that is, they have prerequisites. This can be seen as a form of co-operation in which each course benefits from the existence of the other. Another similar example is in software and programming languages where, for instance, the use of VB in a computer laboratory requires the presence (and co-operation) of Microsoft Windows. Likewise, subject material that relies on the use of a specific textbook could also be seen as an example of co-operation.

Ecological Niches

An ecological niche is a place where a particular species that is well suited to this environment is able to thrive, where other species may not. A curriculum example of this is in the teaching of the PICK operating system by a university in Australia. Some years ago PICK was a serious challenger to UNIX for the "universal operating system" in business, but PICK has now decreased in importance. Despite the fact that no other university in the region now teaches it, and its place being challenged by more recent operating systems, PICK has remained in the curriculum of this university. It has remained largely because an academic involved in its teaching was able to argue convincingly (Tatnall, 2000) that learning PICK allowed students to take up jobs in the small number of prominent local industries using this system: in other words, that it filled an important ecological niche.

Energy Expenditure

It is easy to find examples of minimisation of energy expenditure in curriculum development in the use of curriculum templates and the copying of curriculum from other institutions. Perhaps the greatest reduction in energy expenditure can be gained by using, without change, a model curriculum or the curriculum from another university. A related example is seen in choosing curriculum elements so that they fit in with existing university resources.

FUTURE TRENDS

Application of the Ecological Model

Please refer to the Application of the Ecological Model illustrated in Table 1.

CONCLUSION

Researchers investigating curriculum development, or any other field, must use language in framing their re-

Table 1. Application of the ecological model

Step	Ecological model	Example
1.	Examine the environment in which the IT curriculum change occurs.	This might be a university Department of information systems or computer science, a departmental sub-unit, or perhaps an entire university Faculty.
2.	Look for all relevant entities that might constitute this ecosystem.	Academics, students, university administrators, course advisory committees, local industry representatives, computer networks, programming languages, development methodologies, text books, courses of study, university handbooks and so on.
3.	Look at all interactions between entities and classify these as: cooperative, competitive or niche forming.	Some interactions may not easily fit a single category.
4.	Look for examples of potential cooperation and for cooperative entities.	One academic course can be seen to cooperate with another by acting as a pre-requisite. Visual Basic requires the cooperation of Microsoft Windows in laboratory computers in order to operate.
5.	Look for examples of potential competition and for competitive entities.	The Java and .Net programming systems can be seen to be in competition with each other. Another example is OO development methodologies and conventional structured methodologies.
6.	Look for potential niche applications	In one particular university the teaching of the Pick operating system and Pick basic constitutes a niche application as it prepared this institution's students for work in specific local companies.
7.	Look at the level of energy expenditure (both in keeping the current curriculum in place and in introducing change).	How much energy does a reactive Head of Department have to expend to prevent change? Alternatively, how much energy does an enthusiastic faculty member have to expend to bring about drastic change? (There are, of course, many other possibilities regarding energy expenditure between these extremes.)

search questions. The language used often reflects a general viewpoint of the field being investigated and will always embody some metaphor for the principle components of the field. The metaphor is not useful in *proving* relationships but can be used to convey meaning once relationships are discovered, and an appropriate metaphor can lead the researcher towards or away from useful possible conclusions. Many of the metaphors for curriculum development are simple ones from areas such as the manufacturing-type research, development and dissemination models described earlier. Any investigation of development processes in rapidly changing areas such as information systems shows that a common factor is complexity. This leads the search for a suitable metaphor to those disciplines that have accommodated complexity. One such area is ecology and we have shown how ecological principles appear to provide good descriptions of common curriculum development activities. The ease with which the metaphor can be used to describe actions within IS curriculum development shows that it can be useful as a set of language elements that might lead the researcher to framing useful questions that do not trivialise the complexity of the field.

IS curriculum development involves a complex process of negotiation between actors, and one that cannot be simply explained by reference to a set process of referring new ideas to a series of university committees. The choices of individual academics, or groups of academics, to adopt or ignore a new concept or technology, and to compete or co-operate, must also be considered. This inevitably involves a negotiation process between many different actors. We have argued that this negotiation process can be analysed in terms of ecological behaviour, and have utilised an ecological metaphor to assist in visualising the curriculum development process.

REFERENCES

Case, T. J. (2000). *An illustrated guide to theoretical ecology*. New York: Oxford University Press.

Darwin, C. (1958). *The origin of species*. New York: The New American Library.

Deakin University. (1985). *Curriculum design and innovation.* Geelong: Deakin University Press.

Fullan, M. (1993). *Change forces: Probing the depths of educational reform.* London: The Falmer Press.

Havelock, R. (1969). *The process and strategy of beneficial change: An analysis and critique of four perspectives.* Ann Arbour, Centre for Research on the Utilization of Scientific Knowledge, University of Michigan.

Havelock, R. (1971). *Planning for innovation through dissemination and utilization of knowledge.* Ann Arbor, Centre for Research on Utilization of Scientific Knowledge, Institute for Social Research, University of Michigan.

Kaplan, B. (1985). Barriers to medical computing: History, diagnosis, and therapy for the medical computing "lag". *The Ninth Annual Symposium on Computer Applications in Medical Care*, Silver Springs, MD, IEEE Computer Society.

Kaplan, B. (1991). Models of change and information systems research. In H.-E. Nissen, H.K. Klein & R. Hirschheim (Eds.), *Information systems research: Contemporary approaches and emergent traditions* (pp. 593-611).Amsterdam: Elsevier Science Publishers.

Latour, B. (1996). *Aramis or the love of technology.* Cambridge, MA: Harvard University Press.

Macquarie Library. (1981). *The Macquarie dictionary.* Sydney: Macquarie Library.

Mohr, W. K., & Tulman, L.J. (2000). Children exposed to violence: Measurement considerations within an ecological framework. *Advances in Nursing Science*, *23*(1), 59-67.

Nordvall, R.C. (1982). *The process of change in higher education institutions.* Washington DC, American Association for Higher Education.

Tatnall, A. (2000). *Innovation and change in the information systems curriculum of an Australian university: A socio-technical perspective.* PhD thesis. Education. Rockhampton, Central Queensland University.

Tatnall, A., & Davey, B. (2002). Information systems curriculum development as an ecological process. E. Cohen (Ed.), *IT education: Challenges for the 21st century* (pp. 206-221). Hershey, PA: Idea Group Publishing.

Tatnall, A., & Davey, B. (2003). ICT and training: A proposal for an ecological model of innovation. *Educational Technology & Society*, *6*(1), 14-17.

Townsend, C.R., Harper, J.L., & Begon, M. (2000). *Essentials of ecology.* MA: Blackwell Science.

Truran, J.M. (1997). Reinterpreting Australian mathematics curriculum development using a broad-spectrum ecological model. *Old Boundaries and New Frontiers in Histories of Education: Australian and New Zealand History of Education Society Conference*, Newcastle, Australia, The University of Newcastle.

Wetherbe, J. (1988). *Systems analysis and design: Traditional, structured and advanced concepts and techniques.* St. Paul, MN: West Publishing.

KEY TERMS

Competition: When two individuals or species are in competition with each other, they are each striving for the same thing. In biological systems is typically food, space or some other physical need, but in the model described in this article can be any matter relating to IS curriculum. When the thing they are striving for is not in adequate supply for both of them, the result is that both are hampered, or adversely affected, in some manner.

Co-operation: Occurs when one species works with another in order to achieve an outcome beneficial to one of both. Proto-cooperation is the situation in which both benefit by the co-operation, but can survive without it. Mutualism occurs when each benefits and cannot otherwise survive. Commensalism occurs when two species habitually live together, one species being benefited by this arrangement and the other unharmed by it.

Ecological Metaphor: A way of describing a complex situation, such as IS curriculum development, by providing a way of allowing for the inclusion of complexity, and a language and set of analytical and descriptive tools from the ecological sciences.

Ecological Niche: A place where a particular species that is well suited to this environment is able to thrive, where other species may not.

Ecosystem: In the context of this article, the ecosystem represented by the curriculum in a university information systems department contains (at least) the following "species": lecturers, researchers, students, professional bodies, university administrators and representatives of the computer industry.

Metaphor: A term applied to something to which it is not literally applicable, in order to suggest a resemblance.

Minimisation of Energy Expenditure: A principle of ecology in which a species uses the least possible amount of energy to achieve its purpose.

E–Commerce Curriculum

Linda V. Knight
DePaul University, USA

Susy S. Chan
DePaul University, USA

INTRODUCTION

This article begins by tracing the rapid development of e-commerce curricula in response to demand from employers and students, and the subsequent impact of the dot-com implosion on e-commerce degree programs. The main portion of the article then identifies the major approaches currently taken by universities with respect to e-commerce curricula and explores the three critical e-commerce curriculum issues facing universities. These issues concern (1) whether e-commerce is indeed a separate and distinct discipline, (2) appropriate e-commerce curriculum content, and (3) strategies that can facilitate implementation of an e-commerce program. In the next section of the article, five trends are identified that are critical to the immediate future of e-commerce curricula. Finally, conclusions are drawn concerning the long-term prospects for e-commerce degree programs.

BACKGROUND

In 1995, Vanderbilt University was the first in the world to establish a curriculum area with an emphasis in digital commerce (Vanderbilt University, 2003). By 1997, such schools as the University of Texas at Austin, Duke, Harvard, MIT, Stanford, the University of Michigan, and the University of Rochester were also active in the field (Lewis, 1997). By 1998, George Washington University was offering an electronic commerce class where students from a variety of disciplines bought and sold goods and services over the Internet while they designed and developed the supporting marketing and record-keeping infrastructure (Dhamija, Heller, & Hoffman, 1999). This type of integration of business principles with technology is a hallmark of many e-commerce curricula. In May 1999, Carnegie Mellon University began what is widely credited with being the first e-commerce master's degree in the United States, and over the next few years, e-commerce programs burst onto the academic scene. This rush was uncharacteristically rapid and inconsistent with the historical development of many schools' curricula. Joseph Alutto, business dean at Ohio State, indicated that "It took 20 years to get business schools to focus on international business, and some schools still aren't dealing with entrepreneurship….Yet, within a two- or three-year period, the rate of acceptance and integration of e-commerce into curriculum is much greater than anything we have seen before" (Fitzpatrick, 2000a). Several schools that were early adopters of e-commerce curricula, like Bentley College (Fedorowicz & Gogan, 2001) and DePaul University (Knight & Chan, 2002b), noted the importance of rapid curriculum development practices in responding to the swift emergence of the field.

Universities began offering e-commerce courses because of "the demands of technologically savvy students—and the businesses who want to hire them," according to Andrew B. Whinston of the University of Texas at Austin (Lewis, 1997). A secondary incentive for the development of e-commerce programs was the millions of dollars in funding offered by major corporations like IBM, Ford Motor Co., General Electric, Microsoft Corp, and iXL Enterprises (Dobbs, 1999). The appropriateness of simply responding to business and student demands has been questioned. Bailey and Dangerfield (2000) have maintained that business school curricula should not be driven by immediate customer demands, but rather should take a longer-range view, anticipating customers' expressed and latent needs. Bailey and Dangerfield attribute this approach to Slater and Narver's recommendation that organizations in general should be market-oriented, but not customer-led (Slater & Narver, 1998). Others, including Lightfoot (1999), have maintained that curricula should be driven primarily by educators with long-range perspectives, rather than by shorter-term demands of businesses or students.

With the widespread dot-com bankruptcies and drying up of venture capital funding that began in 2000, many universities began reexamining their e-commerce offerings. Georgia State University terminated its innovative global e-commerce master's program and incorporated e-business courses into other MBA programs (Georgia State University, 2003). DePaul University introduced more advanced technical courses on e-business enterprise architecture design (DePaul University, 2003). Bentley College added more user-centered Web development methods and mobile commerce courses (Bentley College, 2003). As institutions kept pace with industry, curricula moved from an emphasis on entrepreneurial, dot-com implementations to the use of Internet technology in traditional organizations.

ALTERNATIVE CURRICULUM APPROACHES NOW IN USE

The broad range of e-commerce programs has been categorized into a manageable number of models by profiling the programs along three dimensions—career target (a generalist or specialist focus), curricular thrust (a business or technology emphasis), and integration (the extent to which business and technology topics are integrated at the course level) (Chan, 2001). Examples of this classification scheme are depicted in Table 1.

CRITICAL ISSUES OF E-COMMERCE CURRICULUM

Is E-Commerce a Separate Academic Discipline?

As early as 1997, an issue arose that still plagues academics, "Is e-commerce actually a new discipline with a new set of rules and a unique knowledge base?" (Lewis, 1997). Many have argued that e-commerce is a distinct academic field. Donna L. Hoffman of Vanderbilt noted, "The Internet is radically different from traditional markets, and we're discovering that you need radically different approaches for teaching business." (Mangan, 1999). On the other hand, James Ho of the University of Illinois at Chicago stated, "Business is business. Everyone is talking about e-this and e-that, but we didn't say 't-business' when the telephone came along or 'f-business' when the fax came along." (Fitzpatrick, 2000b). The view that

Table 1. Sample classification of e-commerce curricula (Derived from Chan, 2001)

Institution	Career Target	Curricular Thrust	Integration
Bentley College (2003) MBA concentration in e-business	Specialists in accounting, IS, marketing, or finance	Business emphasis	Separate business and technology courses
Carnegie Mellon MS in e-commerce	Both generalists and specialists. Managers, planners, analysts, programmers	Equal emphasis upon business and technology	Separate business and technology courses. Integration of both in the practicum course
City University of Hong Kong (2004) MS in e-commerce	Both e-commerce managers and developers	Equal emphasis upon business and technology	Separate business and technology courses
DePaul University (2003) MS in e-commerce technology	Specialists in e-commerce development, project managers, and consultants	Somewhat greater emphasis upon technology	All e-commerce courses integrate business and technology
University of Westminster (UK) (2004) MS in e-commerce	Specialists in e-commerce development and technology managers	Stronger emphasis on technology	Primarily technology courses
Victoria University (2004) (Australia) Various bachelor of business degrees in e-commerce	E-commerce application developers in various industries	Technical skills, coupled with domain knowledge in a variety of fields	Separate business and technology courses

eventually e-commerce would be assimilated throughout traditional curricula has become increasingly popular as e-commerce curricula have matured. However, integration of new material within an already over-burdened curriculum is difficult. For many, the issue is what to remove in order to bring in more e-commerce concepts. Further, there is at least some evidence that traditional business faculty may not be effectively integrating e-commerce principles into traditional classes. A survey by Morrison and Oladunjoye (2002) of middle schools, secondary schools, community colleges, and colleges indicated, "business educators are not infusing e-commerce topics sufficiently into existing curricula to prepare their students for roles in companies where e-commerce is an integral part of operations." The study concluded that educators at all four teaching levels "reported a similar lack of involvement in e-commerce-related activities such as having read an e-commerce book within the last 6 months or having taken an e-commerce seminar within the past year."

What is the Appropriate Content for an E-Commerce Curriculum?

There is no widely recognized model curriculum for e-commerce, and considerable variation exists across universities. King, Frank, and Platt (2001) studied 65 syllabi from 47 institutions, and noted that most programs were graduate level, and electronic payment, security, and business models were the only topics covered by most institutions. Etheridge (2001) studied 77 e-business programs at AACSB-affiliated institutions, and concluded that the most commonly offered courses were e-business marketing at the graduate level and an introductory survey course at the undergraduate level. Durlabhji and Fusilier (2002) examined 67 North American programs and found that most offered more non-technical than technical content. In a follow-up study a year later, they found that, while the U.S. was "the treasure trove of e-business education," there were multiple programs in Australia, the UK, and Asia (Fusilier & Durlabhji, 2003). Further, technical programs predominated outside North America and were growing faster than more business-oriented programs. The most prevalent strategy among North American master's degrees remained a concentration or track within an existing Master of Business Administration (MBA) program.

Mechitov, Moshkovich, and Olson (2002) examined 26 U.S. MBA programs with emphasis in e-commerce and 10 U.S. M.S. programs in electronic commerce, and compared these programs with 16 non-U.S. programs. They concluded that the non-U.S. programs were similar to United States M.S. programs, although they found that the American M.S. programs were more likely to include courses on high technology management and e-commerce law. White, Steinbach, and Knight (2003) examined only more technically oriented programs, and concluded that graduate level courses commonly include courses in networking, client side Web development, supply chain management, security, and marketing, along with a project course or practicum. At the undergraduate level, they found that a survey of e-commerce, networking, business math and statistics, Web application models, database, marketing, and a senior project course were the most common.

The Mechitov (Mechitov et al., 2002) study proposed an e-commerce program model that included an introductory course, followed by six technology topics (Web site design and development, Web programming, networking and telecommunications, database management, e-commerce security, and systems analysis and design) and six business topics (e-commerce management, hi-tech management, e-commerce strategy, e-commerce marketing, supply chain management, and e-commerce law). The model is capped off with an internship or project course.

Two sets of authors have proposed specific sets of learning objectives for an e-commerce program. The categories of learning objectives identified here built upon those proposed earlier by Knight and Chan (2002a, 2002b) and Brookshire, Williamson, and Wright (2002). At the fundamental level, students should be able to exhibit knowledge and competency in the following areas: practical applications and evolving e-commerce business models within the context of organizational strategies; user-centered Web site design principles, techniques, and engineering processes; hands-on use of e-commerce technologies and tools; rapid Web engineering processes to integrate business models, strategies, design methodologies, and technologies; and individual and team-based solution development; privacy and ethical issues that surround e-commerce. At the advanced level, students should be able to exhibit knowledge and competency in the following areas: e-business architecture design; e-business security; e-business solutions for common business applications, including customer relationship management, supply chain management, B2B exchanges, and portal strategies; design, implementation and evaluation of Web-based marketing campaigns; mobile commerce; Web data mining for business intelligence; Web services; and peer-to-peer technologies.

What Strategies Can Facilitate the Implementation of an E-Commerce Program?

The multi-disciplinary nature of e-commerce requires integration of business strategies and technology issues. This is possible at two levels (Chan, 2001). Integration at the curriculum level expects students to integrate knowledge gained from separate business and technical courses. Course-level integration achieves a higher degree of integrated learning, but is more difficult to accomplish because it requires greater faculty efforts in curriculum design and course development, and greater faculty cooperation across traditional boundaries. Each institution must evaluate whether it possesses sufficient resources to integrate business and technology at the course level. If not, practicums may be used to facilitate student integration of discrete learning. Typically, practicums are introduced at the end of a program. However, some researchers have noted that introducing practicums early in an e-commerce curriculum can facilitate students' ability to integrate knowledge from multiple courses throughout the rest of their programs (Chan & Wolfe, 2000).

As Rob (2003) noted, "Without a core faculty taking full responsibility for its upkeep, an e-commerce program cannot survive." Knight and Chan (2002a, 2002b) recommend leveraging limited existing faculty resources by facilitating cross-disciplinary collaboration between business and technical programs, and providing opportunities for faculty development and retooling. Further, they note that creating a culture that encourages sharing instructional materials can lessen the burden on faculty straining to keep up with rapid technology change. Finally, involving industry partners in guest presentations, teaching, curriculum development, sponsorship of student internships or in-class projects, and lecture series and roundtable discussions is essential for augmenting faculty resources and keeping the curriculum current.

FUTURE TRENDS

Five trends are critical to the future of e-commerce curricula.

(1) **E-business will continue to expand worldwide.** Organizations throughout the world will continue to adopt Internet-based solutions for lowering operating costs, increasing productivity, and strengthening relationship with customers and trading partners.

(2) **Internet-enabled B2B collaboration will dominate e-business development.** Business is moving toward an e-business Web model where members of a participating network of suppliers, distributors, service providers, and customers collaborate. Such Internet-enabled industry value chains redefine e-business as collaborative commerce. Incentives for participants include cost and time reduction, real-time communication, lead-time reduction, and improved collaborative planning and forecasting (Lee & Whang, 2001).

(3) **Architecture for inter-enterprise collaboration is becoming a critical success factor.** Companies are seeking flexible, agile, and economical approaches to collaboration with trading partners. Enterprise architecture must consider external linkages and inter- and intra-enterprise communication through existing systems as well as public and private B2B exchanges (Genovese, Bond, Zrimsek, & Frey, 2001). Web service technologies are emerging as the foundation for a new generation of B2B applications and as the architecture of choice for integrating enterprise applications.

(4) **Inter- and intra-enterprise process will be redesigned.** The emerging framework for collaborative commerce may lead to more loosely coupled processes for component-based architecture and information exchange. Internal business processes in turn will become public through value chain integration. New approaches for process design and application development may emerge in the near future.

(5) **Academic programs need to bridge the gap in the e-commerce IT workforce.** Offshore outsourcing has decreased demand for lower level e-commerce IT jobs in programming, Web development, and design (ITAA, 2003), and companies are seeking more advanced IT skills in architecture, systems integration, collaborative systems development, and complex business process redesign. Yet, few academic programs emphasize advanced technology and business issues. Continuous curricular innovation and faculty development is critical to the future viability of e-commerce education.

The five future trends described above will produce major challenges for universities offering e-commerce curricula. As the technological environment continues to morph, universities will be pressed to keep up with the latest technologies and tools. At the same time, the increasingly complex nature of Web-based business applications will mandate more extensive projects to accurately simulate the intricate nature of real world e-business.

CONCLUSION

From an historical perspective, organizations are just beginning to realize the potential of the Internet. E-commerce solutions are increasingly technically sophisticated and involve increasingly more sophisticated business processes and strategies. In such an environment, e-commerce curriculum must continue to develop rapidly as well.

E-commerce has brought new dimensions to both business and information technology education. Business programs reflect new methods and strategies for organizations interacting with their customers, suppliers, and competitors. In IT programs, beyond the obvious addition of Internet technologies like markup languages, there is a new level of interest in cryptography, human-computer interaction, and networking, along with recognition of the need to combine business and technical knowledge (Ge & Sun, 2000). The latter, the coupling of technology with business, has long been a goal of information systems programs. Its realization may well be the primary lasting legacy of e-commerce curricula.

REFERENCES

Bailey, J.J., & Dangerfield, B. (2000, January/February). Applying the distinction between market-oriented and customer-led strategic perspectives to business school strategy. *Journal of Education for Business, 75*(3), 183-187.

Bentley College (2003). MBA programs. Retrieved on October 14, 2004, from *http://www.bentley.edu/graduate/academics/mba_programs/eb.cfm*

Brookshire, R.G., Williamson, K.C., & Wright, N.C. (Fall 2002). An interdisciplinary undergraduate degree program in electrical commerce. *Information Technology, Learning, and Performance Journal, 20*(2), 25-30.

Chan, S. (2001). Challenges and opportunities in e-commerce education. *Proceedings of the Seventh Americas Conference in Information Systems (AMCIS)* (pp. 1-7).

Chan, S. & Wolfe, R.J. (2000). User-centered design and Web site engineering. *Journal of Informatics Education and Research, 2*(2), 77-87.

City University of Hong Kong (2004). MS in e-commerce. Retrieved on October 14, 2004 from *http://www.fb.cityu.edu.hk/is/Default.cfm?folder menu=PROGRAMMES&smenu=MSEC*

DePaul University (2003). MS in e-commerce technology. Retrieved on October 14, 2004 from *http://www.cti.depaul.edu/programs/2004/gradECT2004.asp*

Dhamija, R., Heller, R., & Hoffman, L.J. (September 1999). Teaching e-commerce to a multidisciplinary class. *Communications of the ACM, 42*(9), 50-55.

Dobbs, K. (1999, December). New rage on campus: E-commerce degrees. *Training, 36*(12), 62-64.

Durlabhji, S., & Fusilier, M.R. (2002, January/February). Ferment in business education: E-commerce master's programs. *Journal of Education for Business, 77*(3), 169-176.

Etheridge, H.L. (2001, July/August). E-business education at AACSB-affiliated business schools: A survey of programs and curricula. *Journal of Education for Business, 76*(6), 328-331.

Fedorowicz, J., & Gogan, J.L. (2001, July/August). Fast-cycle curriculum development strategies for e-business programs: The Bentley College experience. *Journal of Education for Business, 76*(6), 318-327.

Fitzpatrick, M. (2000a, May 29). Inventing the new e-business curriculum. *Chicago Tribune*. Chicagoland Final, Business Section, 1.

Fitzpatrick, M. (2000b, May 29). New kind of teamwork emerges: Schools, businesses pool their knowledge. *Chicago Tribune*. Chicagoland Final, Business Section, 5.

Fusilier, M., & Durlabhji, S. (2003, February). No downturn here: Tracking e-business programs in higher education. *Decision Sciences, 1*(1), 73-98.

Ge, Y., & Sun, J. (2000). E-commerce and computer science education. *Proceedings of the 31st SIGCSE Technical Symposium on Computer Science Education* (pp. 250-255).

Genovese, Y., Bond, B., Zrimsek, B., & Frey, N. (2001). *The transition to ERP II: Meeting the challenges*. Gartner Research, R-94-0162, September 27, 2001.

Georgia State University (2003). GeM Executive MBA. Retrieved on October 14, 2004 from *http://www.eci.gsu.edu/gem/index%20old.htm*

Information Technology Association of America (ITAA). (2003). ITAA 2003 workforce survey.

King, C.G., Frank, S.L., & Platt, R.G. (2001, July/August). E-commerce courses: Overview of nature and content. *Journal of Education for Business, 76*(6), 332-337.

Knight, L.V., & Chan, S.S. (2002a). E-commerce curriculum development and implementation. In M. Dadashzadeh, A. Saber, & S. Saber (Eds.), *Information technology education in the new millennium*, (pp. 159-169).

Knight, L.V., & Chan, S.S. (2002b). E-commerce curriculum strategies and implementation tactics. In E. Cohen (Ed.), *Challenges of information technology education in the 21st century*, (pp. 187-205).

Lee, H., & Whang, S. (2001). E-business and supply chain integration. Stanford Global Supply Chain Management Forum. Retrieved on October 14, 2004 from *http://www.stanford.edu/group/scforum/Welcome/EB_SCI.pdf*

Lewis P.H. (1997, November 3). Business schools are gearing up for a new specialty: Electronic commerce. *New York Times*. Late Edition East Coast, Section D, 5.

Lightfoot, J.M. (1999, September/October). Fads versus fundamentals: The dilemma for information systems curriculum design. *Journal of Education for Business, 75*(1), 43-50.

Mangan, K.S. (1999, April 30). Business students flock to courses on electronic commerce. *Chronicle of Higher Education*, *45*(34), A25-26.

Mechitov, A.I., Moshkovich, H., & Olson, D.L. (2002, Summer). The master's degrees in e-commerce: A survey study. *The Journal of Computer Information Systems, 42*(4), 29-34.

Morrison, J.L., & Oladunjoye, G.T. (2002, May/June). E-commerce infusion into business education–Encompassing the realities of an emerging business model. *Journal of Education for Business, 77*(5), 290-295.

Rob, M. (2003, March). The rise and fall of an e-commerce program. *Communications of the ACM, 46*(3), 25-26.

Slater, S.F., & Narver, J.C. (1998, October). Customer-led and market-oriented: Let's not confuse the two. *Strategic Management Journal, 19*(10), 1001-1006.

University of Westminster (2004). MS in E-Commerce. Retrieved on October 14, 2004 from *http://www.wmin.ac.uk/courses/mscec.htm*

Vanderbilt University. (2003). eLab curriculum. Retrieved on October 14, 2004 from *http://elab.vanderbilt.edu/curriculum/index.htm*

Victoria University. (2004). Bachelors of business degrees in e-commerce. Retrieved on October 14, 2004 from *http://www.business.vu.edu.au/Infosys_Content/Course.htm#h2*

White, J., Steinbach, T., & Knight, L. (2003). E-commerce curriculum: After the fall. *Proceedings of 2003 Information Resources Management Association (IRMA) International Conference*, (pp. 611-614).

KEY TERMS

E-Commerce Security Course: Technologies, architectures, and infrastructure for securing electronic transactions over nonproprietary networks. Implementation and maintenance of mechanisms that secure electronic documents with confidentiality, authentication, integrity, and non-repudiation. Public key certificate. Digital signature.

E-Marketplace Technology Course: System development for online trading applications supporting complex interactions among a variety of users. Theoretical models of online information exchanges supporting negotiations, including auctions, brokerages, and exchanges.

Internet Supply Chain Management Course: System architectures, technologies, and infrastructure requirements in the context of supply chain systems. Design, development and implementation of systems that facilitate collaboration with customers and suppliers. Development of messaging-based collaborative frameworks using Web services.

Intranet and Portal Course: Intranet development methodology, data warehousing, and online analytical processing (OLAP). Enterprise information portals. Transforming information into knowledge. Decision support. Customer applications. Content personalization.

Mobile Commerce Course: Bandwidth, platforms, form factors, mobile data services, and security and transaction models. Web synchronization. Server-side content management. Wireless Application Protocol (WAP). Wireless Markup Language (WML). Handheld Device Markup Language (HDML). Wireless user interface design. Wireless Web development tools.

User-Centered Website Engineering Course: Web site engineering lifecycle and user-centered design. Site goals, business models, value propositions, user analysis, information architecture, interface and navigation design, usability guidelines, database, testing, hosting strategies, usage metrics, and collaborative development.

Web Data Mining Course: Data collection, data extraction, and knowledge discovery for e-business intelligence. Web usage mining, Web content mining, and Web structure mining. Site management. Personalization. User profiles. Privacy issues. Collaborative and content-based filtering.

E-Commerce Taxation Issues

Mahesh S. Raisinghani
University of Dallas, USA

Dan S. Petty
North Texas Commission, USA

INTRODUCTION

This article is designed to give the reader a balanced perspective on some of the issues surrounding the current discussions related to state and local taxation of Internet access fees and sales transactions. It attempts to express the issues being discussed and presents several viewpoints. The proponents of Internet taxation are searching for technological and administrative system to meet their goal. After much deliberation, the Advisory Commission on Electronic Commerce released its final recommendations to Congress in April 2000. Major emphasis is being placed on simplification, neutrality, avoiding double taxation and accepting the existing tax rules with no new taxes.

The United States economy has benefited tremendously by e-commerce. This escalation has created numerous highly skilled jobs, providing the consumer with goods and services at competitive prices. The Internet Tax Fairness Coalition and many other groups feel that implementing taxes on the Internet transaction can have an adverse affect on the businesses. According to the Supreme Court of United States, a vendor has a sales tax obligation only when the buyer and seller are in the same state or has a physical presence (nexus) in the buyer's state. These coalitions feel that entry barriers for new and old companies, who have yet to exploit the e-commerce, will slow the growth in this sector. With over 30,000 taxing jurisdictions, tax collection and payment can be a complex process. Many street retailers collect at a single rate, and prepare and file a single tax return at one place. Taxation of online transactions would require the vendor to identify and send forms to all taxing jurisdictions. Under the present circumstances, the ever-changing maze of state and local tax policies makes application of a single Internet transaction tax policy virtually impossible.

The complicated, complex and ever changing maze of state and local tax policies and laws make application of a sensible, fair and easily understood Internet transaction tax policy virtually impossible under the present circumstances. James Plummer, a policy analyst at Consumer Alert wrote, "Nefarious new taxes and regulations will kill many new start-up e-businesses before they even start up; denying consumers their chance to find the specialized products and services for their needs" (Plummer, 2000).

The anti-tax community and coalitions have a strong adversary in the National Governor's Association. The State is worried that the brick and mortar stores are jeopardized by the popularization of Internet commerce, which is tax-free. The Governors suggest that government tax policy offers a competitive advantage to Internet stores. Major brick and mortar retailers such as Sears and Wal-Mart are concerned that if unresolved, this issue may gain much public resistance, thus making the taxing of e-commerce politically impossible.

BACKGROUND

The United States Congress enacted The Internet Tax Freedom Act in 1998, imposing a three-year moratorium on new Internet taxation. It also established the Advisory Commission on Electronic Commerce to address the issues related to Internet taxation (Advisory Commission on Electronic Commerce, 2000).

The Advisory Commission has representatives from state and local governments and e-commerce industry. It is to conduct a study of federal, state, local and international taxation and tariff treatment of transactions using the Internet and Internet access, and other comparable sales activities. The Commission's recommendations are to be submitted to Congress no later than April of 2000. Based on testimonies, the Commission is reviewing barriers imposed in foreign markets on U.S. property, goods or services engaged in Internet and its impact on U.S. consumers, and ways to simplify federal, state and local taxes imposed on telecommunications services.

The National Governor's Association's Perspective

Today, 46 states have a sales tax of some sort. All of the 46 states that have a sales tax also have what is called a complementary use tax. Consumers pay the sales tax when they buy goods and services in their own state and use tax when they buy from other states. This strategy avoids double taxation. When the consumer buys from an out of state merchant, such as mail order or the Internet, tax is collected and sent to the consumer's state only if the merchant has a nexus in the consumer's state. According to U.S. Supreme Court 1967, National Bellas Hess and 1992 Quill decisions, the merchant is not required to collect the use tax and remit it to the state of residence of the consumer. Consumers then are responsible for paying taxes on goods they purchase through mail-order catalogues and over the Internet. This subsidizes one category of businesses at the expense of their competitors.

The Governors have suggested a streamlined sales tax system for the 21st century. Some of the features of the governor's proposed streamlined system include:

- Maintain the current definitions of nexus and eliminate collection of state and government taxes
- Simplify the current system and without any federal government intervention
- Eliminate the cost of compliance, tax returns and payments and tax audits
- Eliminate tax-rate monitoring and implementation.
- Eliminate risks for sellers exercising reasonable care

The states would implement uniform laws, practices, technology applications, and collections systems to achieve the goals and results. These goals, when implemented, would achieve the first step of the streamlined system. The second step would be for all state and local governments to adopt the same classification systems, definitions and audits

The overall concept of the streamlined system is to reduce the costs and burden of sales tax compliance by shifting sales tax administration to a technology oriented business model operated by trusted third parties (TTPs).

THE E-COMMERCE COALITION PERSPECTIVE

The e-Commerce Coalition is a broadly based national coalition dedicated to providing sound policy information on electronic commerce taxation, and includes AOL, Bank One, Cisco Systems, and the like. Can e-commerce step up the process of making each state responsible for administration of its own tax system and simplification? Time is of the essence because of the speed at which this industry is growing and changing.

GOVERNOR JAMES GILMORE'S PERSPECTIVE

Governor James Gilmore of Virginia is Chairman of the Advisory Commission on Electronic Commerce. Governor Gilmore submitted a proposal to the Commission on November 8, 1999 entitled "No Internet Tax" (Gilmore, 1999).

The Governor believes that American public policy should embrace the Internet and the borderless economy it creates.

- Prohibit all sales and use taxes on business-to-consumer interactions and protect companies from unfair taxes imposed due to their virtual presence
- Amend the Tax Freedom Act to prohibit all taxes on Internet access
- Abolish the federal 3% excise tax on telephone service
- No international tariffs or taxes on e-commerce

E-COMMERCE TAXATION FROM AN INTERNATIONAL PERSPECTIVE

The problem with international taxation is essentially that it is likely not possible to govern well when there is no international government to create an appropriate incentive structure to induce and compel good behavior (Bird, 2003). Molina and Michilli (2003) discuss how e-commerce is emerging in European regions. For instance, the region of Veneto has implemented "Bollo Auto" — the payment of car taxes through a digital network making use of lottery terminals located in the popular tobacco shops (*tabacchinos*). Since July 1, under the aegis of "leveling the playing field," the European Union (EU) has been imposing a value-added tax (VAT) on digital goods — namely games, music, and software — downloaded from non-EU companies via the Internet by EU citizens (Pappas, 2003). McLure (2003) concludes that any failure to apply value-added tax (VAT) to electronic commerce crossing borders between EU member states and other countries should not affect the value added-tax liability of registered traders, even if the reverse charge rule (taxation in the hands of recipients) is not applied. The study notes that the sales of digital content to consumers and unregistered traders that constitutes a minuscule fraction of purchases by households and unregistered traders (given the extremely low level of small-business exemptions) is problematic.

Li (2003) provides a technical and policy analysis of the Canadian Goods and Services Tax (GST) in the context of e-commerce and suggests some options for reform. Even though the GST has had a bad reputation in Canada and its integrity is now threatened by growing online cross-border shopping, based on the revenue potential of the GST, a replacement is highly unlikely, and a cleaned-up or reformed GST is more practical. Thus the government should take advantage of the opportunity presented by e-commerce to reform the GST.

In an application of the substance over form doctrine to the international e-commerce taxation issue, Ngoy (2003) proposes an approach consisting of applying what is called here the permanent establishment (PE) function test to e-commerce infrastructures in order to see whether they qualify for being fiscally treated as PE, if they pass the concerned test. The study concludes that some of them substantially have the same function as the category of office PE, and they should be fiscally treated as this category of PE no matter the form they have.

SUMMARY OF OPTIONS FOR RESOLUTION OF INTERNET TAXATION ISSUES

The European Union has worked out a system where all 15 members impose hefty value-added taxes and all retailers must collect the tax for all sales within the union. United States negotiators are working with groups internationally to come to some understanding regarding these complex issues (Landers, 2000). Senator Ron Wyden and Representative Christopher Cox state that they would work to extend the e-tax moratorium for 5 more years and permanently bar all access taxes (Business Week, 2001).

On the other hand, the taxing entities feel that the online retailers should have to collect the same taxes as brick and mortar businesses. They have signed on to the Streamlined Sales Tax Project, whereby they have agreed to pattern their tax system from a model code (The Wall Street Journal, 2001). Four states, California, Massachusetts, Virginia and Colorado, who are leading technology states, put forth the argument that taxing Internet access and commerce would harm the growth sector of their economies.

Congress has taken the first steps to make sure states cannot impose sales taxes on Internet access, use or content and cannot impose multiple or discriminatory new taxes. Congress has also encouraged and supported the work of the Streamlined Sales Tax Project so states can continue to receive sales tax revenues from Internet transactions, just as they would if the transactions took place in a brick and mortar location. Retail businesses that enter into Internet commerce will be required to collect sales taxes on Internet sales in accordance with the rules that state develops. Sales taxes will remain solely under the jurisdiction of state and local governments for the foreseeable future (Goold, 2003).

The Streamlined Sales Tax Agreement (SSTA) sets uniform definitions and other standards that will make it easier for retailers to collect tax from out-of-state purchasers. The SSTA would make it easier for states to collect taxes on Internet purchases — an extra-territorial money grab strikingly similar to the EU's plan. Although 31 states have already approved the new SSTA, the agreement is voluntary since remote Internet and mail-order sellers are still not legally obligated to collect any tax from out-of-state consumers. However the new Simplified Sales and Use Tax Act of 2003 introduced in Congress by Representative Ernest Istook (R-OK) and a companion bill sponsored by Senator Mike Enzi (R-WY) would effectively make the SSTA mandatory (Pappas, 2003; Rankin, 2003).

Most of the 45 states in the United States that impose sales and use taxes consider the advent and expansion of e-commerce the greatest threat to their financial stability since catalog merchandising. The Constitution's Commerce Clause grants Congress the power to regulate commerce among the several States. In Quill Corp. v. North Dakota, the Supreme Court revisited its long-standing rule that a state could not establish nexus with a remote seller unless the remote seller was physically present in the state. The U.S. Supreme Court reaffirmed that nexus cannot be established in the absence of a remote seller's physical presence in a state (Trealease & Storum, 2003).

FUTURE TRENDS

After reviewing some of the information available, it appears that an interim solution might evolve from the final report and recommendations of the Advisory Commission on Electronic Commerce, including an extension of the moratorium on taxes on Internet access. The infant industry argument is that a tax hurts an emerging industry that still needs a small boost to continue expanding and developing new products and new technology. Critics of Internet taxation charge that it is also unconstitutional and unfair. However, "unfair" is also what land-based retailers might cry, as they claim that Internet retailers have an undue advantage. For both catalog and Internet sales, states want to be able to tax those purchasers who live in their state (Jossi, 2003). Although there are serious and complex issues, it appears that a resolution can be constructed that will be favorable to consumers and businesses alike, including technology being applied to the collection process, standardization of tax systems, and state and local governments being responsible to pay the costs of newly developed and technologically sophisticated collection systems. In February 2003, high-profile retailers such as –INCLUDEPICTURE "http://proquest.umi.com/images/common/circlei3.gif" * MERGEFORMATINET—Wal-Mart, –INCLUDEPICTURE "http://proquest.umi.com/images/common/circlei3.gif" * MERGEFORMATINET —Target, Toys R Us, Marshall

Field's, and –INCLUDEPICTURE "http://proquest.umi.com/images/common/circlei3.gif" * MERGEFORMATINET —Mervyn's began voluntarily collecting sales taxes from their online customers. However, substantial nexus is the keystone of a state's taxing jurisdiction over remote sellers. If substantial nexus does not exist between the state and a remote seller (whether an Internet seller or not), the state may not validly impose tax liability on the remote seller or require it to collect and remit such tax. Perhaps, just as the Internet added new confusion to the area of sales and use taxation, emerging technology-based alternatives may eventually help Internet sellers efficiently carry some of the burdens of multistate tax compliance (Trealease & Storum, 2003). A system currently being used in Europe contains most of the features that states would find necessary for the proper and efficient collection of their taxes. The European Union has worked out a system where all 15 members impose hefty value-added taxes and all retailers must collect the tax for all sales within the union. That is, an Internet purchase made in Germany by a customer in Portugal gets taxed at the VAT rate for Portugal; the German seller collects it. United States negotiators are working with groups internationally to come to some understanding regarding these complex issues (Landers, 2000). State and local governments have acknowledged that their system of sales and use taxes must change in a substantial manner if they are to remain viable in the 21st century.

CONCLUSION

Seven criteria have been laid out for use in designing an acceptable cyber tax system. The system should be equitable and simple, ensure user confidence, prevent tax evasion and economic distortion, maintain a fair balance among countries, and not introduce a new form of taxation (Lee & Hwangbo, 2000). Without the assurance of a uniform nationwide approach, even the most sophisticated technological solution will collapse. It is critical to resolve the e-commerce taxation issue by finding a feasible way to implement a multi-state system for collecting taxes from literally hundreds of tax jurisdictions across the country. The Streamlined Sales Tax Project has been launched by some 30 state governments "to develop a radically simplified sales and use tax system that eases the burden of state use and tax compliance for all types of retailers, particularly those operating on a multi-state basis" (Rankin, 2000). The outcome will have long-term consequences for U.S. retailing, and, according to some, for the American system of government itself.

REFERENCES

Advisory Commission on Electronic Commerce. (2000). Retrieved April 29, 2000, from *http://www.ecommercecommission.org/FAQs.htm*

Bird, RM. (2003). Taxation and e-commerce. *The Canadian Business Law Journal, 38*(3), 466.

BusinessWeek, 49 (2001, February 19). The other tax battleground of 2001: The Internet.

The e-Freedom Coalition. (2000). Retrieved April 29, 2000, from *http://www.policy.com/news/dbrief/dbriefarc453.asp*

Gilmore, J., III. (1999). No Internet tax proposal. Retrieved April 18, 2000, from *http://www.ecommercecommission.org/proposal*

Goold, L (2003). Point, click, tax? *Journal of Property Management, 68*(5), 20.

The Internet Tax Fairness Coalition. (2000). Retrieved April 29, 2000, from *http://www.nettax.fairness.org/facts*

Jossi, F. (2003). The taxing issue of e-commerce. *Fedgazette, 15*(6), 9.

Kyu Lee, J., & Hwangbo, Y. (2000, Winter). Cyber consumption taxes and electronic commerce collection systems: A canonical consumer-delivered sales tax. *International Journal of Electronic Commerce, 4*(2), 6-82.

LaGesse, D. (2000). Governor George W. Bush. *The Dallas Morning News*, 1D.

Landers, J. (2000). Internet tax issues. *The Dallas Morning News,* 1D.

Li, J. (2003). Consumption taxation of electronic commerce: Problems, policy implications and proposals for reform, 38(3), 425.

McLure, Jr., C.E. (2003). The value added tax on electronic commerce in the European Union. *International Tax and Public Finance, 10*(6), 753.

Molina, A., & Michilli, M. (2003). E-commerce innovation in the Veneto region: Sociotechnical alignment in the context of a public administration. *International Journal of Entrepreneurship and Innovation Management, 3*(4), 415.

National Governor's Association. (2000). Retrieved April 29, 2000, from *http://www.nga.org/internet/overview.asp*

National Governor's Association. (2000). Retrieved April 29, 2000, from *http://www.nga.org/internet/facts.asp*

National Governor's Association. (2000). Retrieved April 29, 2000 from *http://www.nga.org/internet/proposal.asp*

National Tax Association Communications and Electronic Commerce Tax Project Final Report, vi-vii. (1999, September 7).

Ngoy, J.M. (2003). Is international e-commerce an HIV tax issue? *International Journal of Services Technology and Management, 4*(1), 53.

Pappas, M. (2003). Europe's global tax. *Foreign Policy, 139*, 92.

Parrish, R.L. (1999). *Tandy/Radio Shack corporation comments to advisory commission on electronic ecommerce.* E-mail correspondence forwarded by the author.

Plummer, J. (2000). Consumer alert. Interview. Retrieved April 29, 2000, from *http://www.policy.com/news*

Rankin, K. (2000, August). Race against time: Seeking a net sales tax solution. *ECWorld*, 26-28.

Rankin K. (2003). Tax-free Internet sales may be ending soon. *DSN Retailing Today, 42*(22), 9.

Releaser, N.T., & Storum, L.A. (2003). The gathering storm: State sales and use taxation of electronic commerce. *Corporate Taxation, 30*(3), 9, 16.

The Wall Street Journal. (2001, March 7). States at odds over Web taxes, B3.

Wyld, D.C. (2003). Don't shoot the Internet. *Computerworld, 37*(47), 21.

KEY TERMS

Double Taxation: When the same taxable item is taxed more than once by either the same or by different government agencies, there is said to be double taxation. The juridical type of double taxation happens when comparable taxes are imposed by two or more taxing jurisdictions on the same taxpayer in respect of the same taxable income or capital.

E-Commerce: Conducting commercial transactions on the Internet, where goods, information or services are bought and then paid for.

Moratorium: Temporary suspension of payments due under a financial or tax agreement. For example, a governmental body may offer a tax moratorium as an incentive to entice a business to locate and/or start up operations in its jurisdiction.

Nexus: The general concept of some connection or link to the taxing jurisdiction. In the U.S., jurisdiction for levying taxes has a Constitutional basis.

Sales Tax: An excise tax imposed on the transfer of goods, typically at retail. States vary as to whether the tax is imposed on the seller of goods or on the buyer; however, sales taxes are almost universally collected from the purchaser at the time of sale.

Use Tax: A complementary or compensating tax imposed by all states that impose a sales tax. Use taxes are typically charged on the "storage, use, or consumption" of goods in the taxing state. Liability to remit use taxes usually falls on the buyer of taxable property or services. Since it is administratively difficult to compel individual self-assessment of use taxes, most of those taxes will go uncollected unless the states can compel sellers to collect them. Significantly, a state may impose use tax collection responsibilities on Internet sellers if they have nexus with the state. The use tax is intended to stem the erosion of the sales tax base when a state's residents purchase taxable goods or services from sellers located outside of the state.

E-Commerce Training for SMEs

Yanqing Duan
University of Luton, UK

INTRODUCTION

E-commerce offers companies tremendous opportunities to improve their business performance in new and innovative ways. However, its potential benefit would only be realised by capable managers who can deal with these emerging technologies and implement them wisely. A skills shortage has been categorised as one of the challenges facing global e-commerce by Bingi and Khamalah (2000). The demand for highly knowledgeable and skilled managers and workloads places enormous pressure upon companies to improve or update their current knowledge and skills. This is particularly important in small and medium enterprises (SMEs),[1] as compared with their larger counterparts, they are often described as "lacking the expertise needed to set up the technologies necessary, despite having a great deal to gain from doing so" (Anonymous, 1998, p. 52). Training is often seen as the most effective way to help SMEs to cope with the increasing demand on improving their skills, while not increasing staffing.

BACKGROUND

E-Commerce Challenges for SMEs

E-commerce and e-business affect the whole business process and cover a wide range of business activities. As a result, its success poses a significant challenge to SMEs. The issues related to the e-commerce challenges within these areas can be examined from different perspectives. From the strategic perspective, Laudon and Laudon (1999) summarised them as:

- The strategic business challenge
- The globalisation challenge
- The information architecture challenge
- The information systems investment challenge
- The responsibility and control challenge

From the operational perspective, Bingi and Khamalah (2000) categorised e-commerce challenges along four major dimensions:

1. Economic: cost justification, Internet access, telecom infrastructure, skill shortage.
2. Technical: security, reliability and protocols, bandwidth, integration.
3. Social: privacy/security, cultural diversity, trust, absence of "touch/feel".
4. Legal: intellectual property right, legal validity of transactions, taxation issues, policing/regulation.

Despite the widespread use of numerous electronic tools, SMEs are described as the slowest sector to embrace e-commerce (Poon & Swatman, 1999). For many small businesses, e-commerce seems like a confusing nightmare. They are not able to react to the rapid changes brought about by this emerging technology, but on the other hand they are scared to be left behind and therefore eager to embrace the technology (Hobson, 2000). The fear of being left behind was also recognised by focus group discussions conducted by Mullins et al. (2000).

Cragg and King (1993) discover that the strongest inhibiting factors for small firms' implementation of information technologies are lack of IS knowledge, lack of managerial time, poor support and limited financial resources. Corbitt (2000) argues that if management is not made aware of new and enabling technology, then it is not surprising that they are unwilling to adopt e-commerce. Timmers (2000) also concludes that a lack of awareness and understanding is one of the reasons for hesitation among many companies about committing any major effort to e-commerce. Lack of skills is perceived to be the most significant barrier to uptake of information and communication technologies (ICTs) (Auger & Gallaugher, 1997; Duan & Kinman, 2000).

Skills Shortage and Importance of Training

Research by Corbitt (2000) found that there was an obvious concern that electronic commerce was not fully understood and the "I do not know" response was the prevailing view on e-commerce understanding in companies studied. Considering the barriers for e-commerce adoption in SMEs, a number of studies (e.g., Drew, 2003; Jones et al., 2003; Lawson et al., 2003; Matlay & Addis, 2003; Quayle, 2003; Ramsey et al., 2003) have been conducted to address this issue and identified that lack of skills and knowledge is one of the barriers for the uptake and use of Internet and e-commerce. There is sufficient evidence from literature on skills and knowledge deficiency in SMEs, which is and will continue to be a significant impediment to the uptake of new technology, and will increasingly disadvantage the competitiveness of SMEs. Thus, it is essential to address the issues of improving skills by effective training and support.

"Skills shortage and training: a forgotten dimension in new technology" is a concern raised by Foley and Watts (1994) nearly a decade ago, but the relationship between skills shortage and training in ICTs deserves even more attention presently. Skills and training issues were often forgotten or misjudged during the new technology appraisal process (Foley & Watts, 1994). Poon and Swatman (1998) found out that research on the topic of the Internet and small business points to the importance of training and demonstrated benefits. Education and training can bridge the gap between development and successful implementation of new technology (Singh, 2000). The rate of change in ICTs means that the training of IT staff is a continual challenge (Dench, 1998). It is particularly essential to keep managers educated on what is going on in order to make informed decisions in today's competitive environment. Although training has been highly regarded as an effective tool for addressing skills shortage, small businesses are particularly reluctant to train (Elbadri, 2001; Johnson & Loader, 2003).

A critical task that all SME managers are facing is how to respond to e-commerce challenges. The initial challenge is to address the current lack of appropriate skills. These can be broadly defined in two areas: technology understanding and ability to facilitate successful technology implementation through appropriate strategic thinking and business planning. Therefore, there is a need for better education and support for SME managers to ensure successful adoption and running of their e-business activities. Although lack of proper training and support is not a specific problem associated with e-commerce adoption in SMEs, it is becoming more prominent for e-commerce success, as managers not only need to become equipped with technical awareness and understanding, but more importantly, with its profound business implications. Evidence from the literature (Drew, 2003; Jones et al., 2003; Lawson et al., 2003; Matlay & Addis, 2003; Quayle, 2003) appears to suggest that there is a gulf between the level of skills and knowledge required for e-commerce success and the current level of skills that managers possess. To reduce the gap, effective training and education is paramount. However, to provide training in the most effective way and at most appropriate level, better understanding of what SME managers really need is deemed critical.

SMEs E-COMMERCE TRAINING AND SUPPORT NEEDS

To investigate the current training provisions and identify managers' perceived training needs for adoption and implementation of e-commerce, a study (Duan et al., 2002) was conducted with SMEs across five European counties, including Germany, Poland, Portugal, Slovak Republic and the United Kingdom (UK). In their studies, it shows that SME managers in general would like to receive training in e-commerce. The majority of SMEs indicate their training needs in e-commerce, Internet, EDI, and related business issues. In more detail, on average, 82% would like to have training on e-commerce/business, 83% on Internet, 77% on the strategic and managerial issues on the use of ICTs, 75% on general knowledge of ICTs and their use in SMEs, 72% on business Web page writing and 72% on EDI. As expected, e-commerce is ranked as a top priority for training, which may reflect the SME's strong willingness to embrace the technology. The least needed areas highlighted are teleworking (55%), video conferencing (60%), mobile communications (57%) and CD-ROM and electronic storage (58%). It seems that some SMEs show lower interest in applying these advanced technologies at the present time. SMEs are not yet convinced that the cost can be justified by the benefits (Spectrum, 1999).

As the implementation of e-commerce will have a profound impact on the overall business process in a company, business issues should always come before the technical issues. Most importantly, "a clear strategy for an e-commerce solution is the key to the door of success" (Cunningham, 1998), as e-commerce takes more than the technology and needs to be treated as a strategic business decision, not just a technology decision (Golderg & Sifonis, 1998; Turban et al., 2004). This issue is clearly reflected in the managers' needs for training and support by the survey conducted by Duan et al. (2002). On average the most demanded area for e-commerce training and support is "business strategies for successful e-commerce", followed by "managing e-commerce operations" and "security". Compared with traditional IT training, which is mainly concerned with technical aspects of information systems, training in e-commerce needs to address not only technical skills, but more importantly business issues related to performance analysis, strategy development and implementation. It is evident that managers are aware that the appropriate e-commerce strategy and management is vital for any business and should be considered as the most important area for training and support. These findings are also confirmed by focus groups conducted by Mullins et al. (2000) which show that strategic issues are very important for SMEs, but managers are not sure what strategy they should follow due to a lack of knowledge. It is not surprising to find that even if the right strategy has been implemented, managing e-commerce operations is also considered a vital area which needs further support. E-commerce is not a one-off event. Effective measures should be implemented to ensure its continuing effective operation.

In terms of training levels for those who have indicated the need for training, survey results by Duan et al. (2002) show that most SMEs would like to be trained at the beginner and intermediate level, although the levels vary according to the training areas and countries. The areas that require more advanced levels of training are e-mail and mobile communications.

Training can be delivered in different ways. How would the managers like to be trained? Some results from the survey (Duan et al., 2002) suggest that the most preferred way of receiving training is on site by external resources (48%), followed by off site training by training organisations (37%). Computer-based self-training delivered by CD-ROM comes as the most preferred self-training method (24%). This may reflect the research finding by Riemenschneider and Mykytyn (2000) which indicates that smaller firms could investigate newer forms of IT training, such as computer-based training or even an intelligent CBT system that contains expert-like capabilities to guide the trainee. It appears that Web-based self-training through the Internet is least popular, especially in Poland. Considering the hurdles, such as limited bandwidth and slow speed of Internet access, it is no surprise that managers would like to have off-line training with CD-ROM. However, it is anticipated that online training, particularly at the advanced level, will become more popular with improvements on Internet access speed. It is also possible that the lower enthusiasm for the Web-based training method is closely linked with the lower level of computer literacy among SME managers.

CONCLUSION

Rapid changes in information and communication technologies have brought enormous opportunities as well as challenges. To handle the challenges effectively and turn the opportunities into real benefits, managers in SMEs need to be equipped with new skills and better knowledge. Companies that are successful in traditional business transactions will be affected if they are left behind. SMEs need to address skills challenges faced by modern business managers, to understand the importance of training for e-commerce success, and more importantly, to identify training needs adequately. Though managers believe that effective training is very important or essential for the success of e-commerce, most SMEs only provide occasional training for their employees. Research suggests that the most needed training areas for e-commerce are e-commerce strategy development, managing e-commerce operations and Internet security. SMEs would like to be trained at beginner or intermediate levels. The most popular human training method is on-site training using external resources and the most preferred computer-based training method is self-training delivered by CD-ROM.

REFERENCES

Anonymous. (1998). E-commerce for SMEs. *Accountancy*, *122*(1259).

Barry, H., & Milner, B. (2002). SMEs and electronic commerce: A departure from the traditional prioritisation of training? *Journal of European Industrial Training, 26*(7), 316-326.

Bingi, P., & Khamalah, J. (2000). The challenges facing global e-commerce. *Information Systems Management*, *17*(4), 26-34.

Corbitt, B.J. (2000). Developing intraoganizational electronic commerce strategy: An ethnographic study. *Journal of Information Technology, 15,* 119-130.

Cragg, P.B., & King, M. (1993). Small-firm computing: Motivators and inhibitors. *MIS Quarterly, 17*(2), 47-59.

Dench, S. (1998). *Keeping it together: Skills for information technologists.* Report of The Institute for Employment Studies. Report 346. Published by The Institute for Employment Studies, University of Sussex, Brighton, UK.

Drew, S. (2003). Strategic use of e-commerce by SMEs in the East of England. *European Management Journal*, *21*(1), 79-88.

Duan, Y., & Kinman, R. (2000). Small manufacturing business: Meeting decision support needs. *Journal of Small Business and Enterprise Development*, *7*(3), 272-284.

Duan, Y., Mullins, R., & Hamblin, D. (2001). Training for e-commerce success in SMEs. In S.S. Burgess (Ed.), *Managing information technology in small businesses: Challenges and solutions* (pp. 334-348). Hershey, PA: Idea Group Publishing.

Elbadri, A.N.A. (2001). Training practice of Polish companies: An appraisal and agenda for improvement. *Journal of European Industrial Training, 24*(2-4), 69-79.

Foley, P., & Watts, D. (1994). Skills shortages and training: A forgotten dimension in new technology. *R & D Development, 24*(3), 279-290.

Goldberg, B., & Sifonis, J.G. (1998). Focusing your e-commerce vision. *Management Review, 87*(8), 48-51.

Hobson, S. (2000, August). Making an e-fit. *Conspectus*, 20-21.

Johnston, K., & Loader, K. (2003). Encouraging SME participation in training: Identifying practical approaches. *Journal of European Industrial Training, 27*(6), 273-280.

Jones, C., Hecker, R., & Holland, P. (2003). Small firm Internet adoption: Opportunities forgone, a journey not begun. *Journal of Small Business and Enterprise Development, 10*(3), 287-297.

Laudon, K.C., & Laudon, J.P. (1999). *Essentials of management information systems* (3rd ed.). NJ: Prentice-Hall Inc.

Lawson, R., Alcock, C., Cooper, J., & Burgess, L. (2003). Factors affecting adoption of electronic commerce technologies by SMEs: An Australian study. *Journal of Small Business and Enterprise Development, 10*(3), 265-276.

Matlay, H., & Addis, M. (2003). Adoption of ICTs and e-commerce in small business: An HEI-based consultancy perspective. *Journal of Small Business and Enterprise Development, 10*(3), 321-335.

Mullins, R., Duan, Y., & Hamblin, D. (2000, January 19-21). An analysis of factors governing the use of e-commerce in SME's. *Proceedings of the First World Congress on the Management of Electronic Commerce*, Hamilton, Canada.

Poon, S., & Swatman, P. (1998). A combined-method study of small business Internet commerce. *International Journal of Electronic Commerce, 2*(3), 31-46.

Poon, S., & Swatman, P. (1999). An exploratory study of small business Internet commerce issues. *Information & Management, 35*, 9-18.

Quayle, M. (2003). E-business in a turbulent world: Usage in European small and medium size enterprises. *International Journal of Electronic Business, 1*(1), 41-52.

Ramsey, E., Ibbotson, P, Bell, J., & Gary, B. (2003). E-opportunities of service sector SMEs: An Irish cross-border study. *Journal of Small Business and Enterprise Development, 10*(3), 250-264.

Riemenschneider, C.K., & Mykytyn, P.P., Jr. (2000). What small business executives have learned about managing information technology. *Information & Management, 37*, 257-269.

Timmers, P. (2000). *Electronic commerce: Strategies and models for business-to-business trading*. Chichester, England: John Willey & Sons Ltd.

Turban, E., Lee, J., King, D., & Chung, H.M. (2004). *Electronic commerce: A managerial perspective*. Upper Saddle River, NJ: Prentice Hall, Inc.

KEY TERMS

CD-ROM: A computer storage device offering a relatively high capacity. The acronym name CD-ROM stands for Compact Disc –Read Only Memory, denoting the fact that CD-ROM discs are read-only devices; data cannot be written to a CD-ROM by a conventional player.

Computer-Based Training (CBT): It can be defined as training and education delivered or enhanced using a computer-based system.

E-Commerce: Refers to the process of buying and selling goods and services electronically involving transactions using Internet, network, and other digital technology.

E-Learning: It can be broadly defined as the use of information and communication technology (ICT) for online delivery of educational courses as part of traditional on-campus learning, distance learning or corporate training. E-learning offers a new opportunity to widen student access to education and vocational training, enabling students to take control over their learning environment and providing enhanced student learning experience through greater interaction between teachers and students, and students with students.

Electronic Data Interchange (EDI): A standard used by businesses to transmit documents such as invoices and purchase orders to each other electronically. The parties who exchange EDI messages (which can be encrypted and decrypted) are refereed to as trading partners.

Small and Medium Sized Enterprises (SMEs): The definition of SMEs varies in different countries. It is normally defined as having between 10 and 249 employees in the UK and Europe.

ENDNOTE

[1] The definition of small and medium-sized enterprises (SMEs) adopted for the purpose of this research is that of the EU's recommendation (see Barry & Milner, 2002), where SMEs are defined as having between 10 and 249 employees.

eCRM in a Manufacturing Environment

Aberdeen Leila Borders
University of New Orleans, USA

Wesley J. Johnston
Georgia State University, USA

INTRODUCTION

This article examines the issue of electronic customer relationship management (eCRM) in a manufacturing context. eCRM has been described as the fusion of a process, a strategy, and technology to blend sales, marketing, and service information to identify, attract, and build partnerships with customers (Bettis-Outland & Johnston, 2003; Jaworski & Jocz, 2002). Although some customers still pay a premium for face-to-face or voice-to-voice interaction in today's high-tech world, through external (e.g., advertising) and internal (e.g., word-of-mouth) influence, the diffusion of the use of eCRM to build and sustain customer loyalty as a firm's strategy is on the rise.

BACKGROUND

Before the 1930s, the production era in which firms pushed to be the "provider" of products, whether customers needed, wanted, or could afford them, was prevalent. From the 1930s to the 1960s, the selling era dictated the commerce arena in which salespersons were encouraged to make sells, regardless of costs. The onset of the 1960s to the 1990s portrayed the infancy of the marketing era in which the marketing concept (or satisfaction of the customer) laid the historical foundation of eCRM. From the 1990s, the partnering era has predominated, and some functions previously performed by marketing have become absorbed into other functional areas, such as manufacturing. Today, value lies within customer relationships that are satisfying to both the customer and the company.

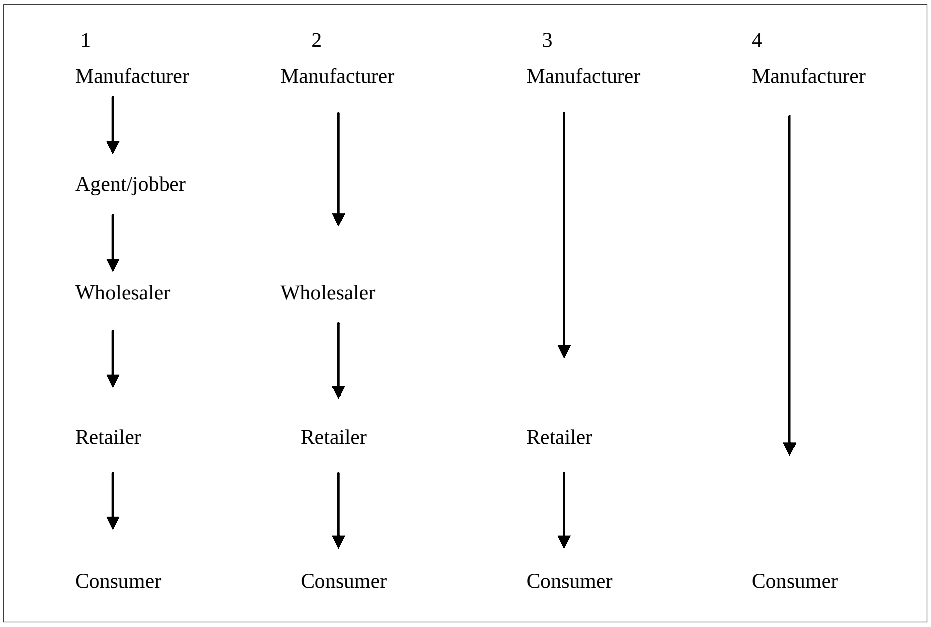
VALUE OF ONLINE SPACE

In reaching the maximization of value, businesses must define value, deliver it, and communicate it to customers. Value to the customer is his or her perception of the use of a product or service in relation to expectations. Drivers of value differ in physical (offline) and online places. Krishnamurthy (2003) viewed online operations as either profit centers (sources of income) or loss centers (offered as service to consumers). The 4 Ps (product, price, place, and promotion) primarily drive physical places. Online, the 6 Cs are the drivers—commerce, content, communication, connectivity, community, and computing (Krishnamurthy, 2003). *Commerce* describes the selling of products from the manufacturing, distribution, and retail firms to customers. Included in this category are the large businesses buying from other businesses in electronic marketplaces. *Content* is applicable to the news publishers (e.g., CNN, *New York Times*, etc.), e-books, or companies using the Internet to educate their customers (e.g., Procter and Gamble at Crest.com). *Communication* involves Web-based seminars, Internet company meetings, and e-mail-based customer service. *Connectivity* refers to the interconnections that employees and users have through the use of the Internet or other knowledge management tools. *Community* is portrayed through special user groups. *Computing* is manifested through tools, such as mapping software, tracking software, and other portfolio management tools that empower customers. Manufacturers build online trust and commitment and potentially increase their value to customers by designing interactive Web sites (Merrilees, 2002). Integrating online operations with physical operations and leveraging company assets provides synergy between physical and online stores that is key to effective eCRM.

DIGITALITY

Digitality refers to the proportion of a company's business that is online (Krishnamurthy, 2003). The digitality of a business lies between zero and one. A business that is completely online with no physical components has a digitality of one. An example is one in which all employees telecommute, digital products such as software are sold, and customers communicate directly with the company's Web site. Alternatively, businesses with no representation in the online space have a digitality of zero. Most manufacturing firms would have digitality close to zero, except those that have incorporated online activities, like Dell or Boeing.

Figure 1. Channel types



MANUFACTURING PROCESSES

Manufacturing processes are the most likely places for sources of innovation and are probably 10 years ahead of service or customer-facing processes (Dixon & Duffy, 1990). Although speed of production still reigns as important in manufacturing processes, the quality of the manufactured product, the flexibility to manufacture different types of products, reliable and predictable adherence to manufacturing timetables, and lowering of the cost and price of products must be matched against the marketing, engineering, and manufacturing capabilities for firms to become world-class competitors in the eCRM world (National Center for Manufacturing Services, 1990).

Typically, sales, manufacturing, and logistics are tightly woven. In coordinating manufacturing with sales, companies attempt to manufacture products and quantities to customer specifications and to minimize delays in delivery. This process has been described as the "lean production system" in automobile manufacturing (c.f. Davenport, 1993). In consumer foods, sales and manufacturing are driven to retail, wholesale, and distribution outlets by consumer demand. A common eCRM tool used is the salesperson's handheld computer that assists with the aggregation of store-level data, enabling linkages to materials and inventory systems, logistics, and sales departments.

Equipment maintenance is another key area in which knowledge and information must be shared in a manufacturing environment to avoid downtime or scheduling and resource requirements conflicts. Radical changes, even lofty customer-initiated improvements, have to be phased in incrementally due to interfaces with legacy systems and logistical concerns in manufacturing arenas, regardless of the company's eagerness to be customer responsive. Many companies innovating with eCRM to coordinate the procurement and delivery of goods, on the outbound logistics side, find it advantageous to use just-in-time delivery or electronic data interchange to shorten the order-to-delivery cycle (Borders, Johnston, & Rigdon, 2001). Finished goods customization, in some instances, is created only to fill customer orders and to ship goods to the customer, eliminating the need for warehousing. Customers with great bargaining power relative to their suppliers often initiate influence tactics that force suppliers to deliver on rapid, short terms (Borders, 2002).

DISINTERMEDIATION

In many cases, the Internet has become another sales channel with complementary features for bricks-and-mortar stores. The role of intermediaries is to facilitate buyer–seller transactions. Understanding disintermediation (the process of eliminating intermediates) requires familiarity with traditional business-to-business channels. Figure 1 illustrates business-to-business channels 1, 2, and 3.

In Figure 1, channel 1 is common in food markets, channel 2 in beer markets, and channel 3 in drugstore lines. Channel 4 is pervasive in the catalog line. Because the Internet makes it easy and relatively inexpensive to interact with customers directly, disintermediaries must prove their value. Channel 4 represents a business-to-consumer (B2C) disintermediated channel in which brokers or infomediaries prevail, as follows:

B2C intermediaries:
- Brokers (facilitate transactions, charge a fee on transactions)
- Infomediaries (serve fulfillment roles, act as filters between companies and consumers, sell market research reports and help advertisers target their ads, e.g., ETrade, virtual malls that help consumers buy from a variety of stores, etc.)

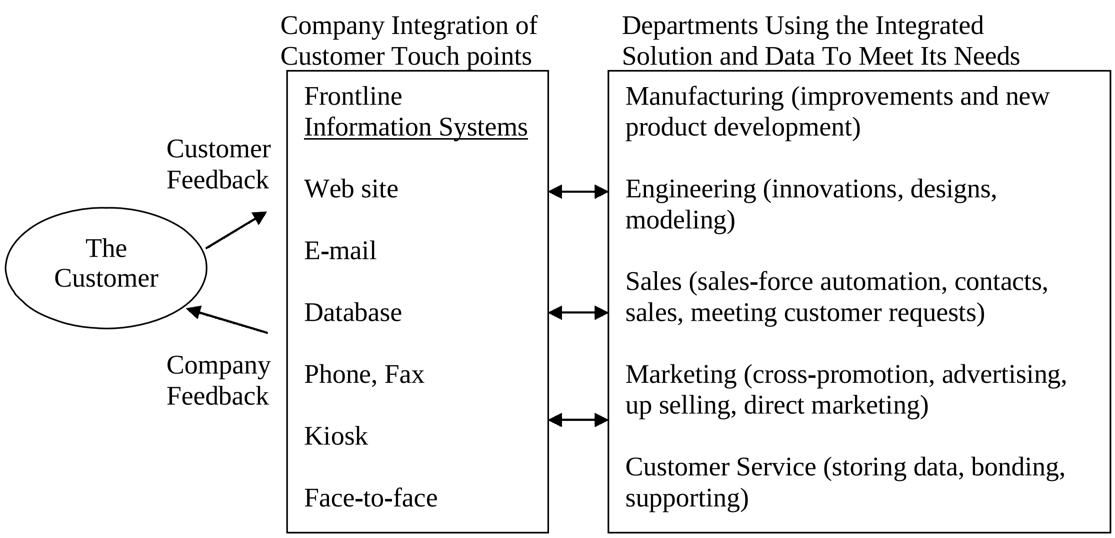
Other disintermediated channels are consumed by metamediary figures to encourage procurement activity. These channel members offer consumers access to a variety of stores or provide transaction services and include search agents (e.g., My Simon, that helps consumers compare different stores); bounty brokers that charge a fee to locate a person, place, or idea (e.g., Bounty Quest); or advertisement-based businesses (pop-ups, banners, other Internet linkages). In disintermediated channels, the major challenges that manufacturing firms face include building online traffic and sustaining customer loyalty and derived demand.

DERIVED DEMAND

Once a manufacturer builds customer traffic, switching costs for the customer increase. The provision of detailed personal and payment information, and "mental transaction costs" relating to trust and privacy by using too many vendors, can be costs that customers will choose to avoid. In employing eCRM, firms should use the appropriate mix of low-tech, middle-tech, and high-tech tools to acquire customers and satisfy them, and the appropriate metrics should be in place for low-profit or high-profit customers.

Krishnamurthy (2003) recommended that firms steer clear of too much eCRM customization and characterized online customers as simplifiers, surfers, bargainers, connectors, routiners, and sportsters. *Simplifiers* respond positively to easier and faster on- rather than off-line experiences. *Surfers* like to spend a lot of time online, and thus, companies must have a huge variety of products and constant updates. *Bargainers* are looking for the best price. *Connectors* like to relate to others. *Routiners* want content. *Sportsters* like sports and entertainment sites. Manufacturers target and build positive brand associations through opinion leaders that identify and attract customers most economically (Krishnamurthy, 2003). This subset of market segmentation is described as "customer profiling." Once these customers begin using the products, they serve as role models and influencers of others in their communities and subcultures in the persuasion of purchasing products.

Figure 2. How eCRM works



Source: Adapted from Timm and Jones (2005) and Sheth and Sisodia (2002).

HOW ECRM WORKS

Successful e-commerce operations involve three elements (Figure 2)—Internet technology, a business model, and marketing (Krishnamurthy, 2003). eCRM describes an Internet-enabled system that leverages the power of the Web to deliver the best possible customer experience. The cost per customer runs from a low of $5 for individual customers to a high of $6,244 for business customers (Chatham, 2001).

Effective eCRM solutions create synergy among sales, marketing, and customer-service activities. With multiple customer contact points, time frames, and systems, customer conversations can be disruptive and content can be lost, if they are not managed properly. eCRM consolidates customer information, such as personal data, preferences, inquiries, and order information and eliminates customers being given the run-around from department to department. eCRM is a supportive technology, whereby timely, targeted, personalized information and solutions are possible for customers.

ONLINE KNOWLEDGE

The quality of the online customer relationship is influenced by the *interactivity* between the customer and the firm, as mediated by the computer (Merrilees, 2002). Interactivity is described in terms of communication as follows:

1. The ability to address someone
2. The ability to gather information and responses from someone
3. The ability to re-address the individual with a unique response (Deighton, 1996)

It is prudent to use a database to target different consumers for different direct mail campaigns. However, junk mail (online or offline) should not be directed at customers merely because the firm has customer contact information. On the other hand, some customers such as “surfers” welcome any contact with the firm. What might be right for a financial institution may be entirely wrong for a small appliances manufacturer. eCRM allows for the creation of the type of relationship that best suits and satisfies the customer’s needs. When the information is “siloed” or stored in different places and is not easily accessible or compatible, there is a call for eCRM.

eCRM is embraced through automated quote generations, where customers can provide specifications for products, e.g., computers, sales referral tracking, travel expense reporting, account management, and sales activity. By studying Web site traffic, companies are in better shape for advertisement placement, budget preparation, and demographic campaign management.

ECRM AS A SOLUTION

“The true business of every company is to make and keep customers” (Drucker, 1954). The Web has become an outstanding channel for delivery of customer assistance and effective customer support activities. In an eCRM environment, the Web allows customers to troubleshoot their own problems with a computerized assistant guide. In a manufacturing environment, when moving customer support to the Web, a firm might start with a static Web page, similar to an electronic version of a marketing brochure, or an information base of common questions and answers. From the company’s perspective, these “low-assistance delayed communication” forms are the least expensive to implement. Self-serve personalized answers, such as tracking status, personalized pages, and real-time data are also eCRM features tailored to meet customers’ needs. Self-service can be provided through quick, simple, and accurate global 24/7 representations. Moving past the self-serve common-answer category is a self-learning knowledge base. Manufacturers that encourage customers to use online knowledge to answer their own questions share in the wealth of information gained from service resolution, abandonment rates, average site connect time, and they are able to better serve their customers by improving content site and contact information (Timm & Jones, 2005).

When customers cannot find the answers they need from self-serve sites, they can then turn to e-mail, in which customer service representatives respond. In comparing transaction costs per channel, Timm and Jones (2005) contended that firms have come to realize that the average transaction cost for Web self-service and interactive voice response (IVR) self-service ($0.24 and $0.45, respectively) compares more favorably than e-mail and phone customer care ($5.00 and $5.50, respectively).

THE BRICKS-AND-CLICKS BUSINESS MODEL IN THE ECRM MANUFACTURING ENVIRONMENT

In several manufacturing areas, it makes sense to combine online capabilities (clicks) with the advantages of traditional stores (bricks). The Internet allows an additional channel for customers to reach businesses and, in turn, through which businesses can reach their customers. For several product categories, such as furniture and apparel, individuals like to touch, feel, order, or try on the product before buying. Manufacturers set up "virtual dressing rooms" that provide customers the opportunity to try on a dress or shirt or other product before ordering it. Therefore, the bricks-and-click models can assimilate the experience of the product. The Internet could be used to locate the store or the variety of merchandise and to keep track of the status of the order.

Partnerships with brick-and-mortar stores can help alleviate delivery problems by providing another outlet for a customer to pick up a delivery that may have been placed online. Likewise, brick-and-mortar outlets that assort several items for shipping, and reduce the cost of shipment per item can aid in returning items to manufacturers. Functionality that makes it convenient for the customer is the motivation behind several eCRM efforts. Deshpandé and Farley (2002) agreed that the appropriate level of customer attention [market orientation] should be what the customer thinks it should be.

WHEN A MANUFACTURER MUST CHOOSE

When manufacturers diversify to create new revenue streams (by creating online tangential sales), they expose themselves to new levels of competition and vulnerability. Firms enhance the shopping experience by remembering relevant information about customers that can establish switching costs. They then use the positive brand-associated name to introduce other products through e-mail alerts, and they provide recommendations.

When implementing eCRM, economics of scale can be garnered when spreading costs across greater categories and leveraging the same brand name and customer base. In the electronics business, manufacturers have stringent requirements for retailers on how they will display and sell their products. Thus, after becoming an authorized dealer, one can get lower prices, money for cooperative advertising, and the right to sell warranties. Improved cross-marketing, copromotion programs, and customer acquisition are provided when manufacturers such as metal companies, thermoplastics companies, and automobile and airplane manufacturers participate in *buy-centric markets*, whereby large, influential buyers find a place where small and fragmented sellers can sell their goods.

EFFECTIVENESS OF ECRM

The hallmark of eCRM is perceived personalization. It promotes cross selling of additional products, and up selling of more profitable products. Determining the needs and wants of customers and satisfying these needs better than the competition while keeping and strengthening the customer relationship bond is the business of eCRM (Kalwani & Narayandas, 1995). eCRM allows firms to solve the problem (competence), apologize for inconvenience (care), and offer a peace token such as waiving fees (comfort).

FUTURE TRENDS

An interesting aspect of some manufacturers' eCRM strategy is to implement programs similar to Amazon's associates or affiliates program. The thrust behind these programs is to have small sites generate traffic by having content with a link to Amazon or other manufacturers. The originating site receives a commission for referred purchases and a smaller commission for other purchases made by the customer. Amazon's brand enhances the smaller site's presence. Partnering with other businesses that sell products the manufacturer does not is a profitable revenue stream (Timm & Jones, 2005).

Manufacturing operations and other backroom processes are now absorbing previous frontline customer-facing functions. Large numbers of food products previously sold in bulk to retailers are now being shipped and sold by manufacturers in smaller packages or multiple size packages, as customers demand (Sheth & Sisodia, 2002). Although many agree that eCRM is a business strategy to select and manage the most valuable customer relationships, business processes are aligning with customer strategies to build customer loyalty and increase profits over time (Rigby, Reichheld, & Schefter, 2002).

CONCLUSION

Three areas of importance that cannot be neglected in eCRM manufacturing processes are the leveraging of database technology, the value of frontline information systems, and the importance of having the right employees (Sheth & Sisodia, 2002). Leveraging database technology leads to better targeting of and maintenance of customers. The frontline information systems (FISs) approach allows customers and frontline employees to be at the cutting edge of information technology that supports relationships and directly impacts customer satisfaction. Having employees that are responsive, courteous, professional, and competent improves a firm's productivity, profitability, and employee and customer satisfaction. Considering the operating costs of serving customers over time and the costs of customer turnover, manufacturers may find the value from customer loyalty, behavior, and satisfaction far exceeds the costs of implementing and maintaining appropriate eCRM systems (Jones & Sasser, 1995; Kundisch, Wolfersberger, & Kloepfer, 2001).

REFERENCES

Bettis-Outland, H., & Johnston, W. J. (2003). Electronic customer relationship management (eCRM) in a business-to-business marketing setting. In T. Reponen (Ed.), *Information technology enabled global customer service*. Hershey, PA: Idea Group Publishing.

Borders, A. L. (2002). *The impact of customer-initiated influence tactics*. Unpublished doctoral dissertation, Georgia State University, Atlanta.

Borders, A., Johnston, W., & Rigdon, E. (2001). Beyond the dyad: Electronic commerce and network perspectives. *Industrial Marketing Management*, *30*(2), 199-206.

Chatham, B. (2001). CRM: At what cost? *The Forrester Report* [online]. Available from *http://www.forrester.com/ER/Research Report/0,1338,11224,FF.html*

Davenport, T. H. (1993). *Process innovation: Reengineering work through information technology*. Boston, MA: Harvard Business School Press.

Deighton, J. (1996). The future of interactive marketing. *Harvard Business Review, 74*, 150–151.

Deshpandé, R., & Farley, J. (2002). Looking at your world through your customer's eyes: Cross-national differences in buyer–seller alliances. *Journal of Relationship Marketing, 1*(3/4), 3-22.

Dixon, J. R., & Duffey, M. R. (1990). The neglect of engineering design. *California Management Review 32*(2,Winter), 9-23.

Drucker, P. (1954). *The practice of management*. New York: Harper and Row.

Jaworski, B., & Jocz, K. (2002, September/October). Rediscovering the customer. *Marketing Management, 11*(5), 22-27.

Jones, T. O., & Sasser, W. E., Jr. (1995). Why satisfied customers defect. *Harvard Business Review,* (November–December), 88-99.

Kalwani, M., & Narayandas, N. (1995). Long-term manufacturer–supplier relationships: Do they pay off for supplier firms? *Journal of Marketing, 59*(January), 1-16.

Krishnamurthy, S. (2003*). E-commerce management: Text and cases*. Cincinnati, OH: Thomson Southwestern.

Kundisch, D., Wolfersberger, & Kloepfer, E. (2001). Enabling customer relationship management: Multi-channel content model and management for financial eServices. *Journal of Marketing Management, 3*(11), 91-104.

Merrilees, B. (2002). Interactivity design as the key to managing customer relations in e-commerce. *Journal of Relationship Marketing, 1*(3/4), 111-125.

National Center for Manufacturing Services. (1990). *Competing in world-class manufacturing: America's twenty-first century challenge*. Homewood, IL: Richard D. Irwin.

Rigby, D., Reichheld, F., & Schefter, P. (2002). Avoid the four perils of CRM. *Harvard Business Review, 80*(2), 101-108.

Sheth, J. N., & Sisodia, R. S. (2002). Marketing productivity issues and analysis. *Journal of Business Research, 55*, 349-362.

Timm, P. R., & Jones, C. G. (2005). *Technology and customer service: Profitable relationship building*. Upper Saddle River, NJ: Pearson-Prentice Hall.

KEY TERMS

Commerce: The selling of products from the manufacturing, distribution, and retail firms to customers.

Connectivity: The interconnections that employees and users have through the use of the Internet or other knowledge management tools.

Customer Profiling: Selecting customers you want to find, going after them, and keeping them.

Digitality: The proportion of a company's business that is online.

Disintermediation: The process of eliminating intermediaries in the channels of distribution.

eCRM: The fusion of a process, a strategy, and technology to blend sales, marketing, and service information to identify, attract, and build partnerships with customers.

Interaction Management: An entire system that monitors customer communications at every possible contact point, regardless of source, i.e., Web, telephone, fax, e-mail, kiosks, or in person.

Knowledge Base: An online repository of information that represents the collective wisdom regarding a product or service.

Educating the Business Information Technologist

John Mendonca
Purdue University, USA

INTRODUCTION

The ways in which computers have been applied to business have evolved over the past 40 years. In the commonly accepted three-era model of applied computing, the first era focused on automating processes, and the second on providing management information. The current era embraces information technology (IT) as a strategic resource and critical asset of the organization (Jessup & Velacich, 1999; Ward & Griffiths, 1996; and others). IT is a transformation enabler or driver that allows organizations to meet both their short-term and long-term objectives. Under this latter model, organizations expect IT to create new ways to compete, new products, new processes, new marketing channels, and even new organizational forms that promote "better-faster-cheaper." Strategic IT thus provides an anchor for competitive advantage—enabling the efficiencies, innovation, market expansion, speed, and alliances that differentiate one company from another.

The pressures of IT globalization have also posited new challenges for the education and employment of traditional non-management IT workers. The growth of offshore outsourcing and insourcing (foreign workers employed overseas by American companies) seems inevitable. Perhaps upwards of 100,000 new jobs that serve American markets have been created in India in the past 4 years at a time when employment in some sectors of the American software industry has seen notable decline (Howell, 2004). The Gartner, Inc., consulting group has projected that, while 5% or fewer of IT jobs have been globally sourced as of the first quarter of 2004, by 2010 nearly 25% of those jobs will be so situated (Gaudin, 2004). While the extent to which global sourcing will be, or should be, embraced is somewhat debatable, IT managers will certainly consider, and at least partially adapt to, this cost-reducing opportunity.

The strategic use of IT, combined with the movement for global outsourcing, forces non-management IT workers to be more knowledgeable about how IT contributes to the organization and to have higher-level IT skills. The framework suggested in this article will help to advance the skill level of such workers by a building a strategic perspective for IT work.

BACKGROUND

Courses in management of information systems and technology, particularly leadership courses at the graduate level, have adapted to the paradigm of IT as strategic transformation agent. On the one hand, these courses, designed for managers, regularly include the use of information as a strategic resource and are amply supported by texts and other academic and professional literature. On the other hand, general undergraduate education in IT—that which produces the basic business information technology worker—continues to focus on producing programmers, systems developers, and systems analysts with first and second-era approaches to their work.

As both the expectations for IT increase and offshore sourcing becomes a reality, information technology education in the U.S. will need to adapt by moving beyond developing basic skills such as programming to developing higher-level skills, such as those required for communication, integration, project evaluation, and value assessment. In addition, the demands on IT to deliver better-faster-cheaper and the need for domestic IT workers to function as leaders and facilitators will require that *all* information technology workers adopt an entrepreneurial perspective consistent with third-era expectations. Organizations need more than traditionally educated programmers and analysts. They need business information technologists who understand such concepts as the organizational impact of IT, enterprise information architecture, enterprise application integration, rapid application development, and process re-engineering, and have good knowledge of analytical and value assessment tools. A basic characteristic of business information technologists should be that they are business information strategists who have, and apply, a strategic perspective toward their work.

The context for discussing and pursuing strategic IT has been developed primarily within the conceptual framework of strategic information systems (SIS). Basically, an SIS is one that supports an organization's competitive strategy—that is, its strategy for gaining advantage over its competitors (Ward, 1996). This concept, therefore, is primarily externally focused. The extension of that concept internally, where strategy is implemented through systems design, redesign, and implementation, is a necessary requirement for fulfillment of better-faster-cheaper in today's IT environment.

Set within the framework of IT as a strategic resource, these realities dominate the contemporary environment, and are serious challenges to the pace and manner in which all IT workers work:

- The rapid pace at which new technologies (software and hardware) are introduced
- The demand for expedited development and implementation of new systems and technologies, leading to new processes and techniques
- Emphasis on value returned on investment
- Telecommunications integrated into, and inseparable from, the computing environment
- Process re-engineering
- Need for integration of seemingly incompatible diverse systems and technologies

The significant driving force behind these realities is the strategic use of IT, and the result is a heightened focus on value contributed by IT. In response, an effective information technologist adopts a strategic perspective that manifests itself in these work attributes:

- An appreciation of IT within the context of business value
- A view of information as a critical resource to be managed and developed as an asset
- A continuing search for opportunities to exploit information technology for competitive advantage
- Uncovering opportunities for process redesign
- Concern for aligning IT with organizational goals
- A continuing re-evaluation of work assignments for added-value
- Skill in adapting quickly to appropriate new technologies
- An object/modular orientation for technical flexibility and speed in deployment

A FRAMEWORK FOR IT EDUCATION

Traditional tools for IT valuation, such as return on investment (ROI) and net present value (NPV), continue to be primarily the domain of financial analysts. IT managers have supplemented these tools with others that add qualitative assessment in addition to financial ones, for example Kaplan and Norton's (1996) balanced scorecard approach, which has gained wide attention. It attempts to apply metrics to an organization's value contribution components, such as customer service (Berkman, 2002).

The tools noted previously are management tools and many require complex analysis. The challenge of developing a strategic perspective for the non-manager business information technologist is to find simple but effective frameworks that can be used to develop the necessary work attributes within the context of SIS. The key question is, therefore: How can organizations get their non-management information technologists to think and act strategically?

The phrase "better-faster-cheaper" used earlier is a good characterization and summation of the concept "strategic". Because of its brevity and clarity, it is easily remembered and readily used internally, at the application implementation level. Information technologists with a strategic perspective can not only use better-faster-cheaper as a standard for examining the value added by information technology activities, but also to uncover IT strategic opportunities. This paradigm suggests important questions technologists should ask about their work assignments. These questions include: How does this IT activity create better products or processes for the customer? How does this IT activity support faster processes? How does this IT activity increase efficiency?

A framework for developing a strategic perspective should include these three elements in its foundation: 1) understanding of the business value of IT; 2) knowledge of key processes within the organization; and 3) understanding the precepts of business process reengineering (BPR) and its application.

The base for the framework, the first element, begins with understanding the potential strategic value of IT. IT is arguably the most disruptive force in organizations within many decades (Mendonca, 2003). IT can be either a driver, such as when it is *the* critical component of an organization's business function (Amazon.com and Ebay are examples); or it can be an enabler, in the case where it serves a significant support function (Federal Express Corp. is an example). In either case it can act as a change agent by being an enabler/driver of success or a significant inhibitor. In the extreme case, "IT can have a transformational effect on a business; IT can change a business in the areas of process, service, management, and even environment" (Luftman, 2004, p. 14).

In their classic work, which focused on the external (environmental) value of IT, Porter and Millar (1985) argued that IT could be used as an enabler of competitive advantage through the creation of new products and processes and the transformation of market forces. Charles Wiseman's strategic thrusts model (Wiseman, 1988) brings an internal perspective to the examination of the business value of IT. It recognizes that individuals and groups within an organization often have opportunities to contribute significant (competitive) value. Business value can be gained through process or product changes, for example in efforts that reduce costs significantly or improve product quality to the extent that it changes the behavior of customers or suppliers and has an impact on the competitive environment. Wiseman's "thrusts" include the following: differentiation, cost, innovation, growth, and alliance. Business information technologists can use both perspectives to assess the value of their own work, to improve that value, and, especially, to look for other opportunities for value enhancement. The significance of this element in the framework is that it answers the question: *What* is business value?

The second element of the framework is knowledge of key processes within the organization. Michael Porter's value chain model (1980) provides a good basis for identifying the fundamental interdependent activities that add value to a product or service as the organization brings the product or service to the customer. IT can profoundly impact the effectiveness and efficiency of these value-adding processes by altering the activities themselves or the relationships between activities. The model, presented in the context of a manufacturing organization, identifies five major functions: inbound logistics; operations; outbound logistics; marketing and sales; and customer service. While some adjustment might be made for other types of organizations, the basic concept holds. The significance of this element in the framework is that it answers the question: *Where* can value be added to the organization through enabling IT?

The third element of the framework is understanding the precepts of business process reengineering (BPR) and how to use it. BPR, developed and popularized by Hammer and Champy (1993), was a shining star of the business environment in the early '90s, but had faded somewhat by the end of that decade, probably because of its over-hyped expectations and poor application. Davenport, Prusak and Wilson (2003) see its resurgence, particularly for the redesign of business-to-business processes necessary for e-commerce. And Garland (2001) makes a case for the survival of the basic concepts as they apply to IT, especially as a critical element in development of enterprise-wide strategic systems such as enterprise resource planning, supply chain management and customer relationship management systems. In this IT education context, BPR is a good mechanism for "triggering" process improvement thinking because it provides guidelines for action. The basic objectives of BPR include:

- Elimination of non-value added processes
- Minimization of value-added processes (reduce the number of activities)
- Simplification of processes, forms, communications
- Integration of jobs, customers, suppliers
- Automation of difficult or hazardous processes

The significance of this element in the framework is that it answers the question: *How* can value be achieved?

The elements of the framework interact to achieve strategic value by considering what value is, where in the organization it can/should be applied effectively, and how it can be achieved. For example, strategic IT may be engaged through significant cost reduction (what) applied to operations (where) through the elimination of a non-value-added process. Another example of strategic IT engagement might be innovation (what) applied to customer service (where) that increases integration with customers (how). The three elements of the framework can be integrated into a structure for activating IT value thinking (Mendonca, 2003). The framework supports IT workers not only in thinking about their current work from a better-faster-cheaper perspective, but also in uncovering opportunities for value contribution.

FUTURE TRENDS

Change in formal IT education, like change in other areas of academia, will probably be slow. In the early 2000s, the impact of the weak U.S. economy, combined with offshore sourcing trends, enforced a perception and predictions that future IT jobs will not be as plentiful as once anticipated. Whether or not extensive global outsourcing will be a significant factor in IT employment is less important, ultimately, than the dominance of IT as possibly the best means by which organizations can achieve "better-faster-cheaper" to survive and thrive in a competitive and fast-changing environment. If educational institutions do not formally adopt new ways to develop entrepreneurial skills and strategic attitudes among their graduates, surely corporations will expand their training in this area to support non-managerial IT staff in their incessant drive for increased value through IT.

CONCLUSION

Because it is a strategic resource that is complex, costly and permeates all functions of the modern organization, information technology is closely scrutinized for its contributed value in relation to resources expended. Consequently, developing a strategic perspective among IT workers is a crucial enabling factor for successfully exploiting IT's transformation capabilities. Information technology within a fast-paced, fast-changing competitive environment places the burden of strategic thinking, and of achieving better-faster-cheaper, on all IT workers, not just managers and leaders.

REFERENCES

Berkman, E. (2002, May 15). How to use the balanced scorecard. *CIO Magazine.*

Davenport, T., Prusak, L., & Wilson, H. (2003, June 23). Reengineering revisited. *Computerworld, 37,* 48-50.

Garland, R. (2001). CPR on BPR: Long live business process reengineering. Retrieved April 15, 2004, from *http://www.technologyevaluation.com*

Gaudin, S. (2004). Gartner: One quarter of U.S. IT jobs outsourced by 2010. Retrieved April 15, 2004, from *http://itmanagement.earthweb.com/career/article.php/3331751*

Hammer, M., & Champy, J. (1993). *Reengineering the corporation; a manifesto for business revolution.* New York: HarperCollins Publishers.

Howell, D. (2004, March 25). Think tank: Offshoring hits high-end workers. *Investor's Business Daily,* A4.

Jessup, M., & Valacich, J. (1999). *Information systems foundations.* Indianapolis: Que Education and Training.

Kaplan, R.S., & Norton, D.P. (1996). *The balanced scorecard.* Boston: Harvard Business School Press.

Luftman, J.N. (2004). *Managing the information technology resource.* New Jersey: Pearson Education, Inc.

Mendonca, J. (2003). A model and sample case for teaching the business value of information technology. *Journal of Information Technology Education, 2,* 61-72.

Mendonca, J. (2003). Organizational impact. In H. Bidgoli (Ed.), *The Internet encyclopedia* (vol. 2, pp. 832-840). John Wiley & Sons.

Porter, M.E. (1980). *Competitive strategy.* New York: Free Press.

Porter, M., & Millar, V. (1985). How information gives you competitive advantage. *Harvard Business Review, 63*(4), 149-160.

Ward, J., & Griffiths, P. (1996). *Strategic planning for information systems.* London: John Wiley & Sons.

Wiseman, C. (1988). *Strategic information systems.* New York: Richard Irwin.

KEY TERMS

Balanced Scorecard: A valuation methodology, here applied to IT, that assigns metrics to non-financial value contributions for the organization, including customer service, innovation, and others.

Better-Faster-Cheaper: A shorthand label for processes that contribute value through better quality, less costly products and services, at a faster production pace.

Business Process Reengineering (BPR): The analysis and redesign of processes within and between organizations. Usually differentiated from process improvement, which is less transformational.

Global Outsourcing/Insourcing: In this context, the sourcing of IT jobs to foreign countries and corporations (outsourcing) or employment of foreign nationals in foreign countries by U.S. corporations (insourcing).

Strategic Information Systems: As opposed to operational systems, those information systems that support and promote the long-term goals of an organization, generally in terms of its competitive advantage in the business environment.

Strategic Information Technology: Computer-based technologies that specifically support and promote the long-term objectives of an organization, both financial and otherwise.

Value Chain: The fundamental organizational processes that add value to the product or service that is produced by a company.

Effective Learning Through Optimum Distance Among Team Members

Bishwajit Choudhary
Information Resources Management Association, USA

INTRODUCTION

For several years, researchers have argued that too much closeness or distance among the team members inhibits intellectual debate and lowers the quality of decision-making. In fact it is often said that if two people always agree, then one is useless and if they always disagree, then both are useless. While too much "closeness" leads to copycat attitude, too much "distance" among the team members results in incompatibility. Creating teams in which the members experience "optimum distance" is not easy.

In this backdrop, we have identified certain gaps in the contemporary organizational learning theories and developed conceptual constructs and conditions that are likely to cause optimum distance in teams.

BACKGROUND

Organizational learning (OL) gained currency when interpreting market information ahead of competitors was seen as a source of competitive advantage (DeGeus, 1988). Organizations increasingly realize the need to maintain a right degree of balance between exploiting the existing and exploring new knowledge base (Cox, 1993; Jackson et al., 1995). Concepts such as double loop learning (Argyris, 1977) and generative learning (Senge, 1990) have underlined the need for innovation and creativity in learning processes.

Research in organizational networks has primarily focused on knowledge creation at organizational levels (Nonaka et al., 1994). Almost all the analyses of networks have focused on inter-organizational groupings (Van De Ven & Walker, 1984). Andersen et al. (1994) define a business network as a set of two or more inter-connected business relationships and claim that the parties in networks have traditionally been shown to come from the same industry.

MAIN THRUST OF THE ARTICLE

In spite of pioneering attempts to conceptualize OL, lately, the researchers have expressed concerns. Ritcher (1998) remarks that the current literature does not adequately explore the dynamics of learning process. Nonaka et al. (1995) claim that "There is very little research on how knowledge is actually created *and hence there is a need to understand the dynamics of knowledge creation*" (italics added).

Alter and Hage (1993) have argued that new theories should be developed to encompass knowledge creation as a result of inter-firm collaboration. Macdonald (1995) claims that the current theories have neglected external-to-firm factors. The aim of OL should be to enhance innovation and not learning merely for the sake of it (Nonaka et al., 1994). D'Aveni (1995) argues that businesses need breakthrough innovations through industry-oriented learning processes and adequately respond to the dynamic external environment.

We now summarize the critical overview of the OL literature presented previously:

- Absence of external-to-firm factors in OL processes.
- Unclear conceptualization of optimum distance in teams.

WHAT IS OPTIMUM DISTANCE?

We delve deeper into OL processes by understanding the factors that constitute perceived distance among the team members by defining the relevant concepts.

Member Distance (MD)

Inkpen (1988) argues that in inter-organizational teams, distrust among members from the participating firms (who perceive each other as competitors) inhibits learning. We believe that this distrust among the team members is the result of the so-called "member distance". Member distance (or MD) reflects overall differences among the members due to objective factors (e.g., members' experience and education) and subjective factors (e.g., members' behavior, values and personality).

Extending Inkpen's (1988) classification of inter-organizational teams, we propose three team compositions comprising managers from:

- Different departments within the same firm (cross-functional teams).
- Same industry-sector but different firms (forums comprising partners).
- Different industries, but similar department (e.g., coordination forums for inter-sector policies or standards body, etc.).

Knowledge Distance (KD)

Managers in different industries need to know some basic industry-specific issues. For example, in the banking sector, managers need the knowledge of payment systems, customer support, and so forth, while in the telecommunication sector, managers need the knowledge of communication networks, mobile devices and so on. Knowledge distance (KD) conceptualizes industry-specific knowledge differences among managers from different sectors.

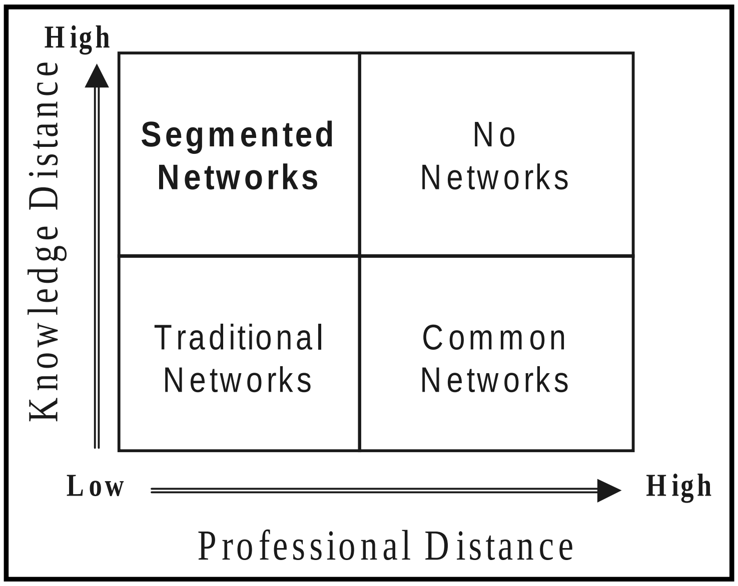
Professional Distance (PD)

Prolonged working and dedicated experiences within a specific department can influence managers' behavior at the workplace. Zuboff (1988) cites several examples (showing the impact of automation on employees' behavior). We refer to job-specific behavioral differences among managers as professional distance. Stated formally, professional distance (PD) comprises intuitive and often subjective personality differences among managers from different departments.

We now summarize some important observations on KD and PD:

- KD captures dissimilarities among managers due to external-to-firm and knowledge specific factors. PD conceptualizes department-specific, behavioral differences among managers.
- KD represents the member-distance at a macro (inter-sector or firm) level. PD represents subjective and more complex personality-based differences at a micro (or department) level.
- Since PD depends on the department dynamics (which impact managers' behavior at the workplace), PD will be low between managers of similar departments even if they come from different industries. KD in a team will be low only when the managers come from firms within the same industry.
- The unit of analyses of learning processes in an individual (manager).

Figure 1. Balancing member distance for effective learning in various teams



Since the large firms usually have a number of specialized departments, we may conceptualize such departments as micro-level "professional personality domains" and firms as "macro-level knowledge domains". KD and PD can then be used to conceptualize different team compositions as shown in Figure 1.

It is evident that the traditional networks occur when both KD and PD are low. These are the internal-department teams and found in all organizations. We refer to them as the "traditional networks".

When integrating knowledge from different departments is needed (in projects, for example), cross-functional teams are often created. In such teams, members have different behavioral approaches to problem solving (hence, high PD). However, their behavioral approaches are often complementary as well given their respective dedicated experience in different functional areas (as technology, business, finance and so on). However, as the managers in cross-functional teams come from the same firm, they share similar broad-based knowledge on industry-level issues (hence, low KD). In addition to low KD, the managers also share common objectives as they work for the same firm. Since such teams are very common, we refer to them as the "common networks".

When KD and PD are both high, no teams can (and should) be formed. Such teams would be difficult to manage and lack common interest areas. We do not discuss this scenario further.

SEGMENTED NETWORK

Varying knowledge-based competencies and similar professional personalities mark the conditions in this team. The presence of high KD creates knowledge-based differences, while low PD implies that members essentially share similar behavior at workplace. This so-called optimum balance between members' distance and closeness is achieved through knowledge-based differences but behavior-based similarities. High KD (through task-oriented conflicts) will lead to higher quality decisions as well (Schweiger et al., 1986) as well.

Segmented network (SN) is hence a team with conditions for optimum distance. It can typically be a forum to create best practices and inter-sector policies (e.g., standards, benchmarks, quality requirements). Use of macro (firm and industry) level and micro (department and individual) level factors facilitate our conceptual understanding of the relation between manageability of knowledge flow and team compositions. Such cross-sector forums are emerging fast (e.g., standards bodies, benchmarking forums, professional services (as consulting, law) firms). Although the participants in SNs come from different sectors, they usually serve the same (or same type of) user base, putting customers at the center-stage.

FUTURE RESEARCH

We now summarize the limitations of this work and identify areas of future research.

Measuring KD and (especially) PD may not be easy, but must be done one way or another. Conceptualizing learning processes with the assumptions of discrete activity is not realistic in a network sector where the value-creating activities are highly information-intensive and overlapping. We also ignored the impact of hierarchy and power on team dynamics for the sake of simplicity. Finally, the challenge in managing segmented networks would be high-level political issues and in turn, that would test the leadership skills of the team's coordinator.

CONCLUSION

By using the conceptual constructs as KD and PD we tried to plug the gaps in the contemporary organizational learning research. As the impact of external forces on firms' performance becomes increasingly pronounced, organizations will need learning processes that are inter-sector and not merely internal to the firm. Segmented networks fulfill this very important premise. Past research also supports our view that high degree of task-oriented conflicts (or high KD, as is the case in segmented networks) increases the quality of decision (Schweiger et al., 1986, 1989). Teams with a high degree of task-oriented conflicts are easier to manage than and may indeed be preferred to teams with a high degree of personality-oriented conflicts. Manageability of teams is a key issue.

Note that the common networks (or the cross-functional teams) also carry the important element of balance within them and their usefulness is well known. SNs offer a similar possibility at an inter-sector level, thereby complementing the internal-to-firm focus of the common networks. SNs are novel and manageable. Inter-sectors forums are growing especially fast in the network industries and so is the need for OL theories to conceptualize and understand SNs. In the network industries (such as banking, telecom sectors), interoperability of systems and procedures is a key success factor.

REFERENCES

Alter, C., & Hage, J. (1993). *Organizations working together*. Newbury Park, CA: Sage Publications Inc.

Anderson, J.C., Håkansson, H., & Johanson, J (1994, October). Dyadic business relationships within a business network context. *Journal of Marketing*, *58,* 1-15.

Argyris, C. (1977, September/October). Double loop learning in organizations. *Harvard Business Review*, *55,* 115 – 125.

Cox, T.H. (1993). *Cultural diversity in organizations: Theory, research and practice.* San Francisco: Barrett-Koehler.

D' Avneni, R.A. (1995). Coping with hypercompetition: Utilizing the new 7S's framework. *Academy of Management Executive*, *9* (3), 45-60.

Dodgson, M. (1993, March-April). Organizational learning. *Harvard Business Review*, *75,* 375-394.

Inkpen, A. (1998). Learning, knowledge acquisition, and strategic alliances. *European Management Journal*, *16*(2), 223-229.

Jackson, S, May, K.E., & Whitney, K. (1995). Understanding the dynamics of diversity in decision-making teams. In R.A Guzzo & E. Salas (Eds.), *Team effectiveness and decision-making in organizations* (pp. 204-261). San Francisco: Jossey-Bass.

Macdonald, S. (1995, September/October). Learning to change: An information perspective on learning in the organization. *Organization Science*, *6*(5), 557-568.

Nonaka, I., Byosiere, P., Borucki, C.C., & Konno, N. (1994). Organizational knowledge creation theory: A first comprehensive test. *International Business Review*, *3*(4), 337-351.

Nonaka, I., & Takeuchi, H. (1995). *The knowledge creating company*. New York: Oxford University Press.

Richter, I. (1998). Individual and organizational learning at the executive level. *Management Learning*, *29*(3), 299-316.

Schweiger, D.M, Sandberg, W.R.,& Ragan, J.W. (1986). Group approaches for improving strategic decision making: A comparative analysis of dialectical inquiry, devil's advocacy, and consensus. *Academy of management Journal*, *29*(1), 51-71.

Schweiger, D.M., Sandberg, W.R., & Rechner, P (1989). Experimental effects of dialectical inquiry, devil's advocacy and consensus approaches to strategic decision-making. *Academy of Management Journal*, *32,* 745-772.

Senge, P.M. (1990). *The fifth discipline: The art and practice of the learning organizations.* New York: Doubleday.

Van de Ven, A.H., & Walker, G. (1984). The dynamics of inter-organizational coordination. *Administrative Science Quarterly*, *29*(4), 598-621.

Zuboff, S. (1988). *In the age of smart machine-The future of work and power.* Basic Books.

KEY TERMS

Knowledge Distance (KD): Conceptualizes industry-specific knowledge differences among managers from different sectors.

Member Distance (MD): Reflects overall differences in approach among the members due to objective factors (e.g., members' experience etc.) and subjective factors (e.g., members' values etc.).

Optimum Distance (OD): Appropriate degree of closeness (or distance) among the team members achieved through the combination of KD and PD, leading to the so-called creative tension.

Professional Distance (PD): Comprises behavioral differences among managers from different departments.

Segmented Network (SN): Inter-sector teams with high KD and low PD, leading to conditions for the so-called optimum distance.

Efficient Multirate Filtering

Ljiljana D. Milić
University of Belgrade and Mihilo Pupin Institute, Serbia and Montenegro

INTRODUCTION

Efficient multirate filters have been developed during the past three decades for implementation of digital filters with stringent spectral constraints (Crochiere & Rabiner, 1981, 1983; Ansari & Liu, 1993; Bellanger, 1984, 1989; DeFata, Lucas & Hodgkiss, 1988; Vaidyanathan, 1990, 1993; Fliege, 1994; Zelniker & Taylor, 1994; Proakis & Manolakis, 1996; Mitra, 2001; Milić & Lutovac, 2002; Hentschel, 2002).

A multirate filter can be defined as a digital filter in which the input data rate is changed in one or more intermediate points. With the efficient multirate approach, computations are evaluated at the lowest possible sampling rate, thus improving the computational efficiency, increasing the computation speed, and lowering the power consumption. Multirate filters are of essential importance for communications, digital audio, and multimedia.

MULTIRATE FILTERING TECHNIQUES

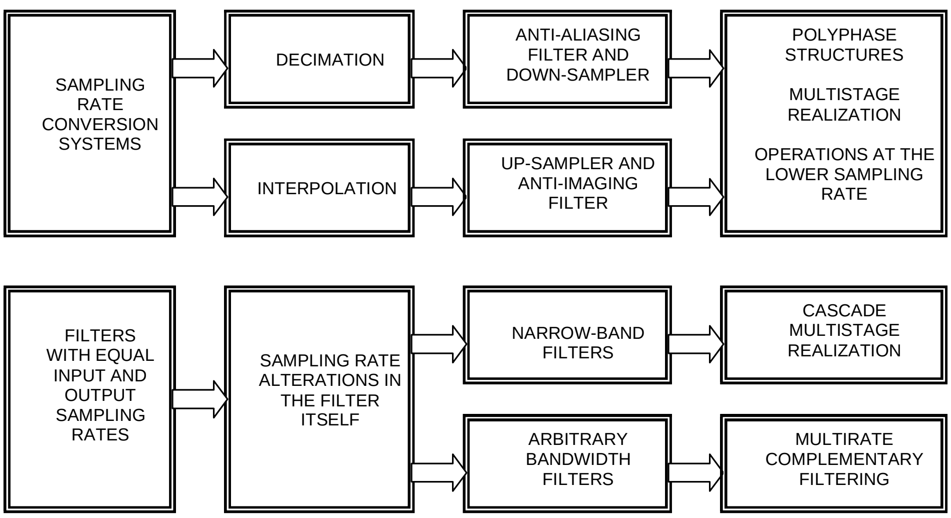
Multirate filtering is one of the best approaches for solving complex filtering problems when a single filter operating at a fixed sampling rate is of a very high order. With a multirate filter, the number of arithmetic operations per second is considerably reduced. The multirate technique is used in filters for sampling rate conversion where the input and output rates are different, and also in constructing filters with equal input and output rates. For multirate filters, FIR (finite impulse response) or IIR (infinite impulse response) transfer functions can be used. An FIR filter easily achieves a strictly linear phase response, but requires a larger number of operations per output sample when compared with an equal magnitude response IIR filter. Multirate techniques significantly improve the efficiency of FIR filters that makes them very desirable in practice.

Figure 1 depicts an overview of different multirate filtering techniques.

Polyphase Implementation

Polyphase realization is used to provide an efficient implementation of multirate filters. A polyphase structure is obtained when an Nth order filter transfer function is decomposed into M polyphase components, $M<N$. For FIR filters, polyphase decomposition is obtained simply by inspection of the transfer function (Crochiere & Rabiner, 1983; Vaidyanathan, 1993; Fliege, 1994; Proakis & Manolakis, 1996; Mitra, 2001). For multirate IIR filters, several approaches to polyphase decomposition have been developed (Bellanger, Bonnerot & Coudreuse, 1976;

Figure 1. An overview of multirate filtering techniques



Crochiere & Rabiner, 1983; Drews & Gaszi, 1986; Renfors & Saramäki, 1987; Russel, 2000; Krukowski & Kale, 2003).

Filters for Sampling Rate Conversion

Filters are used in decimation to suppress aliasing and in interpolation to remove imaging. The performance of the system for sampling rate conversion is mainly determined by filter characteristics. Since an ideal frequency response cannot be achieved, the choice of an appropriate specification is the first step in filter design.

Reducing the sampling rate by a factor of M is achieved by omitting every M-1 sample, or equivalently keeping every Mth sample. This operation is called down-sampling. In order to avoid aliasing, a low-pass anti-aliasing filter before down-sampling is needed. Therefore, a decimator is a cascade of an anti-aliasing filter and a down-sampler. To increase the sampling rate (interpolation by factor L), L-1 zeros are inserted between every two samples (up-sampling). An interpolation filter has to be used to prevent imaging in the frequency band above the low-pass cutoff frequency. An interpolator is a cascade of an up-sampler and an anti-imaging filter.

The efficiency of FIR filters for sampling rate conversion is significantly improved using the polyphase realization. Filtering is embedded in the decimation/interpolation process, and a polyphase structure is used to simultaneously achieve the interpolation/decimation by a given factor, but running at a low data rate.

Due to the polyphase multirate implementation, the number of arithmetic operations in linear phase FIR filters is decreased by a factor M (or L). An effective method, which leads to high efficiency for a high-order FIR filter, is proposed in Muramatsu and Kiya (1997). Efficient decimation and interpolation for the factor M=2 (L=2) is achieved with FIR half-band filters, since the number of constants is a half of the filter length.

Polyphase IIR filters require lower computation rates among the known decimators and interpolators (Renfors & Saramäki, 1987). If a strictly linear phase characteristic is not requested, an IIR filter is an adequate choice. Moreover, an IIR transfer function can be designed to approximate a linear phase in the pass-band (Jaworski & Saramäki, 1994; Lawson, 1994; Surma-Aho & Saramäki, 1999). An IIR decimator or interpolator is particularly useful in applications that cannot tolerate a considerably large delay of an adequate FIR decimator or interpolator. For a restricted class of filter specifications, an attractive solution based on all-pass sub-filters can be used, leading to very efficient implementation (Renfors & Saramäki, 1987; Krukowski & Kale, 2003). The most attractive solution is an IIR half-band filter implemented with two all-pass sub-filters (Renfors & Saramäki, 1987; Johansson & Wanhammar, 1999; Milić & Lutovac, 2002; Krukowski & Kale, 2003). For a rational conversion factor L/M, a very efficient decomposition of IIR filter is proposed in Russel (2000).

Filters with Equal Input and Output Rates

Digital filters with sharp transition bands are difficult, sometimes impossible, to be implemented using conventional structures. A serious problem with a sharp FIR filter is its complexity. The FIR filter length is inversely proportional to transition—width and complexity becomes prohibitively high for sharp filters (Crochiere & Rabiner, 1983; Vaidyanathan, 1993; Saramaki, 1993; Fliege, 1994; Proakis & Manolakis, 1996; Mitra, 2001). In a very long FIR filter, the finite word-length effects produce a significant derogation of the filtering characteristics in fixed-point implementation (Mitra, 2001). IIR filters with sharp transition bands suffer from extremely high sensitivities of transfer function poles that make them inconvenient for fixed-point implementation (Lutovac, Tošić & Evans, 2000). In many practical cases, the multirate approach is the only solution that could be applied for the implementation of a sharp FIR or IIR filter. Thus to design a multirate narrowband low-pass FIR or IIR filter, a classic time-invariant filter is replaced with three stages consisting of: (1) a low-pass anti-aliasing filter and down-sampler, (2) a low-pass kernel filter, and (3) an up-sampler and low-pass anti-imaging filter (Crochiere & Rabiner, 1983; Fliege, 1994; Mitra, 2001; Milić & Lutovac, 2002). The total number of coefficients in a multirate solution is considerably lower than the number of coefficients of a single-rate time invariant filter.

Multistage Filtering

For decimation and interpolation filters, and for multirate narrowband filters, additional efficiency may be achieved by cascading several stages, each of them consisting of a sub-filter and down-sampler for decimation and an up-sampler and sub-filter for interpolation (Fliege, 1994; Mitra, 2001; Milić & Lutovac, 2002). Design constraints for sub-filters are relaxed if compared to an overall filter. Hence, by using the multistage approach, the total number of coefficients is significantly reduced when compared with the single stage-design. The effects of finite word-length in sub-filters are low in comparison with the single-stage overall filter. When a decimation/interpolation factor is expressible as a power-of-two, the application of half-band filters improves the efficiency of the system.

Multirate Complementary Filters

This method can be used in designing filters with any pass-band bandwidth. The multirate techniques are included to reduce the computational complexity. Using the complementary property, the multirate, narrow pass-band filter designs can be used to develop high-pass and low-pass filters with wide pass-bands (Ramstad & Saramäki, 1990; Fliege, 1994; Mitra, 2001). When the output of a low-pass multirate filter is subtracted from the delayed replica of the input signal, the result is a wideband high-pass filter. The delay has to be selected to exactly equal the group delay of the multirate filter. For a low-pass wideband filter, the multirate narrowband high-pass filter has to be used.

Efficient FIR filters with an arbitrary bandwidth can be designed using multirate and complementary filtering (Ramstad & Saramäki, 1990; Fliege, 1994; Johansson & Wanhammar, 2002). The overall design is evaluated by cascading complementary multirate filtering two-ports composed of two series branches and one parallel branch. The cascade is terminated with a simple kernel filter. One series branch of the cascade is a decimator (filter and down-sampler), while the other is an interpolator (up-sampler and filter). The parallel branch is a delay. The most efficient solution is obtained when half-band filters are used in the cascade.

Recently, the complementary filtering approach is extended to IIR filters (Johansson, 2003). The overall filter makes use of an IIR filter as a kernel filter, the periodic all-pass filters for constructing complementary pair, and linear phase FIR filters for the sampling rate alterations.

Half-Band Filters

Half-band filters are basic building blocks in multirate systems. A half-band filter divides the basis band of a discrete-time system in two equal bands with symmetry properties. The FIR filters are most often used as half-band filters. For a linear phase FIR half-band filter, half of the constants are zero valued when the filter order is an even number (Saramäki, 1993; Mitra, 2001). A half-band IIR filter can have fewer multipliers than the FIR filter for the same sharp cutoff specification. An IIR elliptic half-band filter, when implemented as a parallel connection of two all-pass branches, is an efficient solution (Milić & Lutovac, 2002). The main disadvantage of elliptic IIR filters is their very nonlinear phase response. To overcome the phase distortion, one can use optimization to design an IIR filter with an approximately linear phase response (Surma-Aho & Saramäki, 1999), or one can apply the double filtering with the block processing technique for real-time processing (Powel & Chau, 1991; Lutovac et al., 2000).

For the appropriate usage of digital filter design software in half-band filter design, it is necessary to calculate the exact relations between the filter design parameters in advance (Milić & Lutovac, 2002). The accurate FIR half-band filter design methods can be found in Vaidyanathan and Nguen (1987), Saramäki (1993), and Wilsson (1999). For the IIR half-band filter design, see Schüssler (1998) and Milić and Lutovac (2002, 2003).

Multiplierless Solutions

The efficiency of multirate filters is significantly improved by simplifying arithmetic operations. This is achieved by replacing a multiplier with a small number of shifters-and-adders. Generally, implementing multiplierless design techniques in sub-filters, at the cost of a slight derogation of filtering performances, increases the efficiency of the overall multirate filter. For instance, one can use the optimization technique (Kaakinen & Saramäki, 1999), the multiplier block approach (Dempster & Macleod, 1995), or design based on EMQF (Elliptic Minimal Q-Factors) transfer functions (Milić & Lutovac, 1999, 2003; Lutovac & Milić , 2000). A well-known solution for large conversion factors in decimation is a cascaded integrated comb (CIC) filter, which performs multiplierless filtering (Hogenauer, 1981; Mitra, 2000; Hentschel, 2002).

CONCLUSION

The multirate filtering techniques are widely used in sampling rate conversion systems, and for constructing filters with equal input and output sampling rates. Various multirate design techniques provide that the overall filtering characteristic is shared between several simplified sub-filters that operate at the lowest possible sampling rates. Design constraints for sub-filters are relaxed if compared to a single rate overall filter. Hence, by using the multistage approach, the total number of coefficients is significantly reduced. As a consequence of the reduced design constraints, the effects of quantization (finite word-length effects) in sub-filters are decreased. Multirate filters provide a practical solution for digital filters with narrow spectral constraints that are very difficult to solve otherwise.

REFERENCES

Ansari, R. & Liu, B. (1993). Multirate signal processing. In S.K. Mitra & J.F. Kaiser (Eds.), *Handbook for digital* *signal processing* (pp. 981-1084). Hershey, PA: Idea Group Publishing.

Bellanger, M. (1984, 1989). *Digital processing of signals: Theory and practice.* New York: John Wiley & Sons.

Bellanger, M.G., Bonnerot, G. & Coudreuse, M. (1976). Digital filtering by polyphase network: Application to sample-rate alteration and filter banks. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 24*(April), 109-114.

Crochiere, R.E. & Rabiner, L.R. (1981). Interpolation and decimation of digital signals – A tutorial review. *Proceedings of the IEEE, 78*(March), 56-93.

Crochiere, R.E. & Rabiner, L.R. (1983). *Multirate digital signal processing.* Englewood Cliffs, NJ: Prentice-Hall.

DeFata, D.J., Lucas, J.G. & Hodgkiss, W.S. (1988). *Digital signal processing: A system design approach*. New York: John Wiley & Sons.

Dempster, A.G. & Macleod, M.D. (1995). General algorithms for reduced-adder integer multiplier design. *Electronic Letters, 31*(October), 1800-1802.

Drews, W. & Gaszi, L. (1986). A new design method for polyphase filters using all-pass sections. *IEEE Transactions on Circuits and Systems, 33*(March), 346-348.

Fliege, N.J. (1994). *Multirate digital signal processing.* New York: John Wiley & Sons.

Hentchel, T. (2002). *Sample rate conversion in software configurable radius.* Morwood, MA: Artech House.

Hogenauer, E.B. (1981). An economical class of digital filters for decimation and interpolation. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 29,* 155-162.

Jaworski, M. & Saramäki, T. (1994, May). Linear phase IIR filters composed of two parallel all-pass sections. *Proceedings of the IEEE International Symposium on Circuits and Systems* (Volume 2, pp. 537-540), London.

Johansson, H. (2003). Multirate IIR filter structures for arbitrary bandwidth. *IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications, 50*(December), 1515-1529.

Johansson, H. & Wanhammar, L. (1999). High-speed recursive filter structures composed of identical all-pass subfilters for interpolation, decimation, and QMF banks with perfect magnitude reconstruction. *IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing, 46*(January), 16-28.

Johansson, H. & Wanhammar, L. (2002). Design and implementation of multirate digital filters. In G. Jovanović - Doleček (Ed.), *Multirate systems*: *Design & applications* (pp. 257-292). Hershey, PA: Idea Group Publishing.

Kaakinen, J. & Saramäki, T. (1999). Design of very low-sensitivity and low-noise recursive filters using a cascade of low-order lattice wave digital filters. *IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing, 46*(July), 906-914.

Krukowski, A. & Kale, I. (2003). *DSP system design: Complexity reduced IIR filter implementation for practical applications.* Boston: Kluwer Academic.

Lawson, S.S. (1994). Direct approach to design of PCAS filters with combined gain and phase specification. *IEEE Proceedings of Vision, Image and Signal Processing, 141*(June), 161-167.

Lutovac, M.D. & Milić , L.D. (2000). Approximate linear phase multiplierless IIR half-band filter. *IEEE Signal Processing Letters, 7* (March), 52-53.

Lutovac, M.D., Tošić , D.V. & Evans, B.L. (2000). *Filter design for signal processing using MATLAB and Mathematica.* Upper Saddle River, NJ: Prentice-Hall

Milić , L.D. & Lutovac, M.D. (1999). Design of multiplierless elliptic IIR filters with a small quantization error. *IEEE Transactions on Signal Processing, 47*(February), 469-479.

Milić , L.D. & Lutovac, M.D. (2002). Efficient multirate filtering. In G. Jovanović -Doleček (Ed.), *Multirate systems: Design & applications* (pp. 105-142). Hershey, PA: Idea Group Publishing.

Milić , L.D. & Lutovac, M.D. (2003). Efficient algorithm for the design of high-speed elliptic IIR filters. *AEÚ International Journal of Electronics and Communications, 57,* 255-262.

Mitra, S.K. (2001). Digital signal processing: A computer-based approach. New York: McGraw-Hill.

Muramatsu, S. & Kiya, H. (1997). Extended overlap – Add and save methods for multirate signal processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 45,* 2376-2380.

Powel, S. & Chau, M. (1991). A technique for realizing linear phase IIR filters. *IEEE Transactions on Signal Processing, 39*(November), 2425-2435.

Proakis, J.G. & Manolakis, D.G. (1996). *Digital signal processing: Principles, algorithms, and applications.* London: Prentice-Hall.

Ramstad, T.A. & Saramäki, T. (1990, May). Multistage, multirate FIR filter structures for narrow transition-band filters. *Proceedings of the 1990 IEEE International Symposium on Circuits and Systems* (pp. 2017-2021), New Orleans, Louisiana, USA.

Renfors, M. & Saramäki, T. (1987). Recursive Nth-band digital filters – Part I: Design and properties. *IEEE Transactions on Circuits and Systems, 34*(January), 24-39.

Russel, A.I. (2000). Efficient rational sampling rate alteration using IIR filters. *IEEE Signal Processing Letters, 7*(January), 6-7.

Saramäki, T. (1993). Finite impulse response filter design. In S.K. Mitra & J.F. Kaiser (Eds.), *Handbook for digital signal processing* (Chapter 4, pp. 155-277). New York: John Wiley & Sons.

Surma-Aho, K. & Saramäki, T. (1999). A systematic technique for designing approximately linear phase recursive digital filters. *IEEE Transactions on Circuits and Systems-II, 46*(July), 956-963.

Vaidyanathan, P.P. (1990). Multirate digital filters, filter banks, polyphase networks, and applications: A tutorial. *Proceedings of the IEEE, 78*(January), 56-93.

Vaidyanathan, P.P. (1993). *Multirate systems and filter banks*. Englewood Cliffs, NJ: Prentice-Hall.

Vaidyanathan, P.P. & Nguen, T.O. (1987). A trick for the design of FIR half-band filters. *IEEE Transactions on Circuits and Systems, 34*(March), 297-300.

Wilsson, Jr., A.N. & Orchard, H.J. (1999). A design method for half-band FIR filters. *IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications, 45*(January), 95-101.

Zelniker, G. & Taylor, F.T. (1994). *Advanced digital signal processing: Theory and application.* New York: Marcel Dekker.

KEY TERMS

Decimation: Decreasing the sampling rate. Decimation process consists of filtering and down-sampling.

Down-Sampling: Discarding every M-1 samples (retaining every Mth sample).

FIR Filter: A finite impulse response digital filter.

Half-Band Filters: A low-pass or high-pass filter that divides the basis band in two equal bands, and satisfies prescribed symmetry conditions.

IIR Filter: An infinite impulse response digital filter.

Interpolation: Increasing the sampling rate. Interpolation consists of up-sampling and filtering.

Multirate Filter: A digital filter, which changes the input data rate in one or more intermediate points in the filter itself.

Multistage Filtering: Cascade of filters and decimators (interpolators and filters).

Polyphase Decomposition: Decomposition of a transfer function in M (L) polyphase components that provide sequential processing of the input signal at the lower sampling rate.

Up-Sampling: Inserting L-1 zeros between every two samples.

E-Government Interoperability

Dirk Werth
German Research Center for Artificial Intelligence, Germany

INTRODUCTION

The rise of the Internet has structurally changed not only the business area, but also governments and administrative authorities. The usage of information and communication technologies (ICT) influenced the organizational behavior and the daily work of public administrations. In parallel, a new management paradigm has grown in governments and administrations: The New Public Management (NPM) aims to a new orientation on the impact of public activities and on the benefit of public services for its customers, namely citizens and businesses (Barzelay, 2001). It puts the administration and its activities in the triangular relationship between politics, administration and citizens (Osborne & Gaebler, 1992). Within this "ecosphere", decentralized steering models (Reichard, 2002) as well as market mechanisms are introduced and emphasized (Pollitt & Bouckaert, 2000).

Both factors, ICT and NPM, have resulted in the electronic (E-) Government. E-Government is the support of public service processes through ICT. It affects all areas of governmental work and acting. For this article, we focus on the production and distribution of public services (Reinermann, 2002). However, even in this limited area, the implementations vary from case to case. In order to rate the effects and benefits of E-Government, that strongly depend on the interaction abilities already usable, a measurement is needed. Currently, the realization degree indicator is commonly used for this purpose. Table 1 summarizes the commonly used categories of realization steps for E-Government:

BACKGROUND

Integration requires a seamless and customer-oriented integration of public services across organizational borders and through different application systems (Wimmer & Traunmueller, 2002). This results in the need for public administrations to open themselves towards other authorities as well as to their customers, namely citizens and businesses and to actively collaborate with them. The area E-Government Interoperability covers strategies, organizational concepts and information technology to link administrative business processes and to interconnect application systems.

Conceptual and technical E-Government Interoperability infrastructures are considered as prerequisite for the creation of advanced public services and the distributed processing of multi-agency business processes (Werth & Zangl, 2004). Conceptual solutions mainly cover Enterprise Architectures and reference models for specific use in public authorities. In this context, Enterprise Architectures are representing the public authority in terms of organisation and operations (e.g. processes, behaviour, activities, information, decision and object flows, resources and organisation units, system infrastructure and architectures), in order to reach some finalities. Finalities here indicate the creation of explicit facts and knowledge that add value to the enterprise or can be shared by business applications and users for the sake of improving the performance of the organization (Interop, 2003). Especially, operations require the management of the public services as the external view (Glassee, Van Engers, & Jacobs, 2003) and of the business processes as the internal implementation (Seel, Guengoez & Thomas, 2004). Furthermore, reference process models represent the unifying abstraction of a variety of process models in different authorities. They can be used in multiple administrative scenarios to support the realization of services described by these reference models (Martin, Seel, Kaffai & Thomas, 2004).

However, these concepts also demand on capable and effective interoperational infrastructures for public administrations to operate (Fernandez, 2002). Furthermore, they foster "the transition from the current paradigm of highly fragmented, isolated applications and islands of functionality to a situation promoting consolidation to an integrated, collaborative and secure architecture" (Werth, 2003). Most approaches specify components for different tasks within the infrastructures, usually "definition/configuration", "discovery", "connection", "processing" and "monitoring/administration". First approaches base on workflow management architectures (Reinermann, 1997). The resulting technical infrastructures cover most or all of those components and organize them either in a centralistic or distributed way. In centralistic architectures, there is only one instance of a component that controls the activities. Distributed environments miss this single control instance in favour of self-organizing and –controlling techniques. Both approaches may be adequate for integration scenarios depending on the structure of the orga-

Table 1. Stage-wise realization model of e-government

I. *Information (enhanced)* – This first stage is the easiest to implement. It comprises electronic availability and provision of thematically structured information. This represents an unidirectional communication relationship between public administrations and their customers. A website is a typical example for this stage.
II. *Communication (interactive)* – extends the information stage by a feedback channel. Hence a bidirectional communication is established. Common technologies are e-mail, chats and forums.
III. *Transaction (transactional)* – The transaction stage describes the online availability of public services, i.e. it becomes possible for a citizen to trigger online a legally binding public service. It can also include electronic payment and all relevant phases of a transaction.
IV. *Integration (networked)* – This highest stage of realization, only rarely used, describes the integration of the customer himself into public administration processes as well as the collaboration of public services. The customer can influence the execution of public services without having the knowledge about the way of working and processing of the respective services. At this stage, full electronic support is required, covering all public services and the according processes.
The attributes in brackets state the presence measurement expressions of the United Nations (cf. United Nations, 2003).

nizational system to be integrated. In centralistic environments, (e.g. hierarchical ordered authorities) centralistic technical architectures are mainly suitable, whereas in decentralized organizational networks (e.g. in federative states), distributed architectures will fit the best.

A suitable approach to decompose E-Government Interoperability is given by three layers, as shown in Table 2.

APPROACHES AND TECHNOLOGIES

Currently, two major research streams address this topic from different directions and with diverging intentions:

The *Front-end interoperability* approach is user-centric and tries to realize an easy and direct access to exposed public services. It focuses on a consistent presentation and thematic structuring, based on the interests of the citizen. Mainly the attempt is to group those public services that are assumed relevant for a specific situation of a citizen's life (schooling, full age, building a house, retirement, etc.) or a business event (founding, location, economic simulation, etc) (Reinermann, 2002). This concept is called "live event oriented" for citizens resp. and "business episode oriented" for companies (Vintar & Lebel, 2002). To realize this concept, Internet-based portal technologies access services located at different authorities, without regarding institutional boundaries (Klischewski, 2001). Such integrated portals result in one-stop-shop E-Government solutions. They enable the triggering of single services or groups of them via one interaction step using the normal web browser (Tambouris, 2001).

The scope of the *Back-end interoperability* approach is the cross-organizational business process. The intension is to link business processes and their executing application systems. Therefore, it corresponds to the area of Enterprise Application Integration (Linthicum, 1999). On the backend side of service processing, adequate public services have to be identified, discovered and accessed. Implementing infrastructures mostly rely on middleware (Pasic, Diez & Espinosa, 2002) or agent tecnnologies (Carvalho, Moreira & Sa-Soares, 2003). Middleware is mostly used for data replication, transformation and transportation. Thus, single messages representing data items are passing through the applications following a specific rule set. This technology realizes simple data integration. Using agent technologies, the integration can be augmented to the functional level. Here, autonomous software components (agents) try to form a predefined behavior by calling functions (methods) of other agents. Emerging developments on Service-Oriented Architectures (SOA) and Web Services foster a new dimension of back-end interoperability, due to open standards and interfaces. By exposing encapsulated functionalities by self-describing Web Services, a public authority is enabled to use business functionalities that are processed by other authorities. Hence, it becomes

Table 2. Layer decomposition of e-government interoperability

- *Business Interoperability* – covers the exchangeable specification of public services, business processes and organizational structures, their transformation into compatible ones and the mechanisms to interconnect them.
- *System Interoperability* – means the enabling of communications and data exchange of application systems in order to realize a common, distributed application behavior.
- *Semantic Interoperability* – the need for this layer came from the fact, that "discrete resources use different terms to describe similar concepts or use identical terms to mean different things" (Miller, 2000). This layer includes data interpretation as well as knowledge representation and exploitation in order to create common semantic meanings.

possible to compose different services available within the network in order to realize its own business process. This approach is called business process orchestration (Gortmaker & Janssen, 2004).

FUTURE TRENDS

The domain of E-Government Interoperability is currently emerging. It is expected that several national solutions will arise that propose solutions to this problem domain. Actually, several national-driven E-Govern-

Table 3. International e-government interoperability initiatives

- *Federal Enterprise Architecture Framework (FEAF, USA)* – aims to create an enterprise architecture description, specifying the overall organization and relationships of public administrations. It comprises a business architecture (process and organization), a data architecture (data types), a application architecture (applications and capabilities) and a technology architecture (hard- and software, physical factors). (CIO Council, 1999)
- *E-Government Interoperability Framework (e-GIF, UK)* – is a collection of policies and standards for governmental IT systems. It was created and is updates by the British Cabinet Office. The framework is decomposed in two parts: Part One containing high level policy statements as well as management, implementation and compliance requirements and Part Two including technical policies and specifications in the five major categories "Interconnection", "Data integration", "Content management metadata", "Access" and "Business areas". (Office of the e-Envoy, 2003a & 2003b)
- *European Interoperability Framework (EIF, EU)* – defines "a set of recommendations and guidelines for Government services so that public administrations, enterprises and citizens can interact across borders, in a pan-European context. The target audience of the EIF is the managers of eGovernment projects in Member States Administrations and EU bodies. Member States Administrations should use the guidance provided by the EIF to supplement their national eGovernment Interoperability Frameworks with a pan-European dimension and thus enable pan-European interoperability." (IDA, 2004, p.3)

ment initiatives exist worldwide, addressing the field of interoperability. The most advanced are being developed in the United States, the United Kingdom and the European Union. Table 3 provides an overview on these activities:

CONCLUSION

E-Government Interoperability will be the enabler for the future development of E-Government. However, it should be considered not as the aim but as the instrument. The improvement of public services towards integrated and collaborative provision and processing techniques requires sustainable interoperability infrastructures. They comprehend Enterprise Architectures, addressing networked governmental structures on a conceptual level as well as application integration systems, realizing Internet-connected, decentralized, service processing grids. The key success factor for interoperability is the holistic development of business-, system- and semantic-layers and consequently, the creation of solutions integrating the three layers. Intensive further research on this topic may – in some years – discover the extension of the E-Government stage model, the fifth stage.

REFERENCES

Barzelay, M. (2001). *The new public management - Improving research and policy dialogue*. Berkeley: UC Press.

Carvalho, J., Moreira, H. & Sa-Soares, D. (2003). An architecture for European public administration systems interoperability. *Proceedings of the 3rd European Conference on e-Government*, Reading, MCIL, 65-68.

CIO Council (1999). *Federal enterprise architecture framework*, Washington: CIA Council.

Fernandez, A. (2002). Towards interoperability amongst European public administrations. In K. Lenk & R. Traunmueller (Eds.), *Electronic government* (pp. 105-110). Heidelberg: Springer.

Glassee, E., Van Engers, T.M. & Jacobs, A. (2003). POWER: An integrated method for legislation and regulations from their design to their use in e-government services and law enforcement. In M. F. Moens (Ed.), *Digitale wetgeving, digital legislation* (pp. 175-204). Brugge: Die Keure.

Gortmaker J. & Janssen, M. (2004). Business process orchestration in e-government: A gap analysis. In M. Khosrow-Pour (Ed.), *Innovations through information technology* (pp. 1243-1245). Hershey, PA: Idea Group Publishing.

IDA. (2004). European interoperability framework for Pan-European e-government services (version 4.1) [Working document]. Brussels: European Communities.

Interop Consortium. (2003). *Interoperability research for networked enterprises applications and software*. Brussels: Interop NoE.

Klischewski R. (2001). Infrastructure for an e-government process portal. In D. Remenyi & F. Bannister (Eds.), *European Conference on e-Government* (pp. 233-245). Reading: MCIL.

Linthicum, D.S. (1999). *Enterprise application integration*. Boston, MA: Addison-Wesley.

Martin, G., Seel, C., Kaffai, B. & Thomas, O. (2004). Administrational excellence – A transaction-based reference architecture for e-government. In Macik, K. (Ed.), *Proceedings of the 6th International Conference on "The Modern Information Technology in the Innovation Processes of the Industrial Enterprises (MITIP) 2004"*. Prague.

Miller, P. (2000). Interoperability-What is it and why should I want it? *Ariadne* (24).

Office of the e-Envoy. (2003a). E-government interoperability framework – Part one: Framework (Version 5.0). London: Cabinet Office.

Office of the e-Envoy (2003b). E-government interoperability framework – Part two: Technical policies and specifications (Version 5.1). London: Cabinet Office.

Osborne, D. & Gaebler, T. (1992). *Reinventing government: How the entrepreneurial spirit is transforming the public sector*. Reading: Addison Wesley.

Pasic, A., Diez, S., Espinosa J.A. (2002). IMPULSE: Interworkflow model for e-Government. In K. Lenk & R. Traunmueller (Eds.), *Electronic government* (pps. 472-479). Heidelberg: Springer.

Pollitt, C. & Bouckaert, G. (2000). *Public management reform: A comparative analysis*. Oxford: Oxford Press.

Reichard, C. (2002). The "new steering model" and the citizen. *German Journal of Urban Studies, 41*(2).

Reinermann, H. (1997). Administrative informatics and administrative reform. In M. Davies (Ed.), *New state, new millennium, new public management: Coping with the problem of transition* (pp. 137-146). Liverpool: Liverpool University.

Reinermann, H. (2002). The electronic city hall. *German Journal of Urban Studies*, *41*(2).

Seel, C., Guengoez, O. & Thomas, O. (2004). Towards an integrated e-government. In M. Khosrow-Pour (Ed.), *Innovations Through Information Technology* (pp. 943-946). Hershey, PA: Idea Group Inc.

Tambouris, E. (2001). An integrated platform for realising online one-stop government: The eGOV project. *Proceedings of the DEXA International Workshop "On the Way to Electronic Government"*. Los Alamitos, CA: IEEE Computer Society Press.

United Nations. (2003). *UN global e-government survey 2003*. New York: United Nations.

Vintar, M. & Leben, A. (2002). The concepts of an active life-event public portal. In K. Lenk & R. Traunmueller (Eds.), *Electronic government* (pp. 383-390). Heidelberg: Springer.

Werth, D. (2003). Future Research Directions. In L. Guijarro (Ed.), *E-government interoperability in the 2005-2010 horizon* (p. 17). Brussels: European Commission.

Werth, D. & Zangl, F. (2004). Enabling cross-border public services by Pan-European e-government interoperability. In Martin Bichler et al. (Eds.), *Coordination and agent technology in value networks* (pp. 109-123). Berlin: GITO.

Wimmer, M. & Traunmueller, R. (2002). Integration - The next challenge in e-government. In B. H. Far et al. (Eds.), *EurAsia-ICT 2002-Advances in Information and Communication Technology* (pp. 213-218). Vienna: Austrian Computer Society.

KEY TERMS:

Back-End Interoperability: Business-process-centric integration approach that interconnects different application systems in order to enable the execution of cross-organizational business processes.

E-Government Integration Stage: Also partially referred as Participation stage; means the formation of networked governments and the integration of public services both internally (i.e. electronically interconnected, multi-tier-transaction-enabled business processes) and towards the customers, citizens and businesses (i.e. offering of unified user access interfaces directly linked to integrated services).

E-Government Interoperability: The ability of a public service to collaborate (i.e. to work together) with other public services without special effort on the part of the processing unit. This comprehends the collaboration of the service-executing business processes as well as the interaction of the process supporting application systems.

Front-End Interoperability: User-centric integration approach, that presents consolidated information retrieved from different sources and that enables the access to multiple application systems in order to perform a set of coherent transactions at once.

Interoperability Enterprise Architecture: A set of consistent methods to specify the interaction behavior of a network of public organizations. It includes organizational structures, public services, business processes as well as interaction interactions, protocols and patterns.

Life-Event / Business-Episode Concept: The approach to classify and cluster public services by repeating, specific situations that often occur during the life of a citizen resp. the different periods of an enterprise. Typical examples are marriage for a citizen and VAT declaration for a company.

One-Stop-Shop E-Government Solutions: Are Internet portals of public authorities, offering to citizens and enterprises the life-event- resp. business-episode-structured access to a group of coherent public services via a single user interface. All services within this group can be triggered by one interaction ("one click").

E-Government, E-Democracy and the Politicians

Birgit Jaeger
Roskilde University, Denmark

INTRODUCTION

The development of Electronic or Digital Government (E-Government) has varied throughout the world. Although we give it the same name, we know from different studies that, for example, the concept of information society can be interpreted in different ways in different cultural settings (Sancho, 2002; Williams & Slack, 1999). This article provides a general outline of the development of E-Government in the West and is primarily based on European and Scandinavian experiences.

It is only possible to give an introduction to E-Government if we can define what we are talking about. E-Government is still a rather new concept, thus its definition is not yet completely accepted, but for the time being, most people agree that E-Government includes the following features:

- E-Government is based on Information and Communication Technologies (ICTs)
- E-Government is taking place in public administration
- E-Government concerns electronic ways to perform all kinds of internal administrative tasks
- E-Government also concerns the communication between the public administration and the citizens and other actors in the surrounding society (Jæger, 2003)

BACKGROUND

Based on the first part of this definition the history of E-Government starts in the beginning of the 1960's when the magnetic tape replaced the punched card. During the 1960's and 1970's big central databases were built and were run on big mainframe computers. The databases mostly contained administrative data from fields where the law and regulations were clear and there was a large amount of data to process. In this period, large registers concerning, for example, public institutions were formed; and software systems for the government of the economy including salaries, taxes and pensions were developed. These activities were often run centrally and the results were delivered to the relevant authority.

When we turn to the second characteristic, we have to include the development of the public administration as well. During the 1980's and 1990's most Western countries have experienced a profound modernization of their public administration. At first, this modernisation was marked by reforms that have since been collectively labelled New Public Management. According to Rhodes (1997), New Public Management can be said to involve two different types of initiatives, the first of which relates to the management itself. These initiatives include a focus on management by objectives, clear standards, and an evaluation of the quality of service, while at the same time granting greater attentiveness towards the users of the public service in question. The other type of initiative deals with the introduction of economic incentive structures. This involves the dissection of the public administration in demarcated services, contracting out some services, and other services that are sought and arranged in competitive-like situations by establishing quasi-markets in which the consumers of the services are provided with an opportunity to choose between different services.

These alterations have had a more or less unintended consequence – the emergence of new policy networks around the provision of public services (Rhodes, 1997; Stoker, 1998; Heffen, Kickert, & Thomassen, 2000). These policy networks draw new agents into the management of the tasks in question, including agents from the business community as well as from civil society. Now we see private companies carrying out publicly-commissioned services. We also see civic groups in the local community, NGO's, sports clubs or interest organisations take over more social and caring-type tasks, which were earlier defined as public. (Again, we have to be aware of different traditions in different countries, but especially in the Scandinavian countries many of these tasks have been defined as public, whereas in other Western countries the family and local community have played a much bigger part in taking care of these things.) These agents are now engaged in relations with the public administration in collective, binding policy networks. This general development is often described in terms of the transformation of public sector regulation from government to governance.

These reforms differ from country to country (Rhodes, 1999), but the general picture is that these reforms have had great impacts on the way the public administrative tasks are preformed. Here the development of E-Government plays a significant role. Many of these reforms would have been very difficult to realize without ICTs. An example of this is the decentralization of administrative tasks from town halls to public institutions in Denmark. This reform was based on the use of PCs and the development of an internal electronic network between the town hall and all the public institutions. Today we describe it as the start of the development of the intranet in the authority in question (Jæger, 2003).

This development has continued and today we have a wide range of different software systems for all kinds of administrative tasks. These include electronic archives, systems for handling electronic documents, systems for consideration of different cases and so forth. Garson (2000) provides an overview of this field as well as a review of the literature.

Rather early on, it became clear that the development of E-Government was not just the design of an information system and its implementation in an organization. Thus, over the years, a lot of effort has been put into developing methods for the process of design and implementation (Bødker, Kensing & Simonsen, 2000). Based on different analyses of failures, it was acknowledged that it is very important to draw on the experience of the potential users in the design process. Otherwise, it is easy to produce systems that do not fit their needs. The experience also showed that the implementation of the system is very important if the organization is going to harvest the benefits of the system. A lot of parameters have to be taken into consideration in this process. The staff has to be informed and drawn into the process. It also is likely that some training is needed. In addition, the way to organize the consideration of cases has to be carefully examined, and it is perhaps necessary to draw in other competences than those that already exist in the organization. The inclusion of all these factors is important if one wishes to ensure that the design and implementation of a new information system is to be a success for the public administration.

During the last couple of decades, public authorities on different levels have developed their own information systems for performing their tasks. This has lead to a situation where many public agencies are unable to communicate electronically because they use different technological standards. Thus, there is a need today for developing common standards for electronic communication between public agencies at different levels. In recent years, this has become a barrier to the development of E-Government and therefore a large amount of resources is now spent on solving this problem.

E-DEMOCRACY

Nevertheless, public authorities do not only communicate internally or with other public agencies, they also communicate with citizens, private companies and other users of public services, which is the last of the features of E-Government listed above. With the introduction of the World Wide Web in the 1990’s, the public authorities were given a tool for this external communication. During the late 1990’s, and since, most public agencies have developed their own website where they place a lot of information, and electronic forms citizens have to fill out to apply for a public service and so forth.

Also in this area of the E-Government, we find different kinds of development. In a study of the development of Digital Cities in Europe (Bastelaer, Henin, & Lobet-Maris, 2000), (Williams, Stewart, & Slack, Forthcoming) it became clear that in some cities (e.g. Copenhagen (Jæger, 2002)) the website was developed as a part of the E-Government and interpreted as a tool for communication between the public authorities and the citizens, while in other cities (e.g. Amsterdam (Van Lieshout, 1999)) the website was developed as a tool for communication between citizens and did not involve the public authorities very much.

In terms of definitions, this is the most debated aspect of E-Government. Some people interpret a public website only as being a tool for administrative tasks for use between the public agency and the citizens, while others see the website as a place for debate and a tool for democracy as well. The first group defines the objectives of E-Government as a way to rationalize public administration and increase its efficiency, thus democratic debate should, in their understanding, not be a part of a public authority’s website but should be developed as something else – as E-Democracy. The last group defines the objectives of E-Government as a tool for all the tasks a public authority has and, consequently, also a tool for the democratic process. In this understanding, the democratic use of ICTs should be developed side by side with the administrative use.

Whether E-Democracy is developed as an integrated part of E-Government or as a special service, it is the least developed use of ICTs. To further this development, it is necessary to define what kind of democracy the technology should support. Without going into a theoretical discussion on the concept of democracy, it is possible to state that we have at least two different kinds of democracy: representative democracy and participatory democracy. Representative democracy is what the Western world mostly defines as democracy, and functions through a parliament where all the citizens in a country have the possibility to elect some people to represent them. Participatory democracy is defined as the wide range of activities in which citizens participate in the political process. It is important to remember that these two kinds of democracy are not mutually exclusive. In most Western countries, the two kinds of democracy exist side by side.

Technology will be developed differently in accordance with the type of democracy it should support (Hoff, Horrocks, & Tops, 2000). If E-Democracy is going to be a tool for the representative democracy, it can be used for electronic voting or online referendums (Hauge & Loader, 1999). Or it can be used to support the political parties in their dialog with the voters (Löfgren, 2001; Nixon & Johansson, 1999). If E-Democracy is going to be a tool for participatory democracy, it should be used to support political debate, for example, in a local community or in other political processes. Then it is a question of making websites containing information about the political process and chat rooms or newsgroups where people can discuss different political subjects and where politicians have an opportunity to argue for their opinions.

In recent years, there have been a lot of small-scale experiments where ICTs have been tried out as a tool for E-democracy. Some of these have been promising (especially the experiments with electronic voting), but many have more or less failed (especially the experiments with political debates on the Internet). Even if some of these experiments have been promising, we still lack a convincing example of large-scale use of ICTs in E-democracy.

THE POLITICIANS AND E-GOVERNMENT

The development of E-Government has a great impact on the people who are affiliated to the public administration. During a recent Danish study of the role of politicians in the development of E-Government (Jæger, 2003, in press), it became clear that they play a very limited part in this development. It is mostly the civil servants who decide how to apply ICTs in E-Government. Hesitancy among politicians towards becoming involved in the development of E-Government and determining the means by which ICT is utilised in administration, is largely due to their role conception, for example, whether or not they perceive involvement in such concrete, technical tasks, as being a part of their role.

The role of the politician builds on a sharp distinction between politics and administration. The politician is to represent the people and put forward grand political visions for societal development. Conversely, the administrator's role is to serve the politicians by carrying out their visions and generally administrating society on the basis of the politically defined framework. The means by which administrators implement and administer political decisions is regarded as a technical matter. The administrators themselves can therefore choose the methods and instruments they consider to be most suitable and efficient in the given situation. According to this perception of the division of roles, it is clearly up to the administrators to determine how E-Government should be shaped.

This distribution of roles between politicians and administrators has long historical roots and has been working well for generations. But due to the shift in the public administration from government to governance, the politicians have to change their role if they want to maintain their positions as governors of the public sector. Sørensen (2002; see also Kooiman, 2000) describes the new requirements of politicians under governance regulation as different means of exercising meta-governance. The requirements of politicians exercising meta-governance concern the creation of frames to allow other actors to participate in the process. The role of the politicians as meta-governors is then to specify the competence that various actors have to perform, and to make decisions. In order to get this range of independent actors to work in the same direction, the politicians are also responsible for standing forth to offer political leadership capable of creating meaning via a common understanding of the general goals.

The development of E-Government can serve as a possibility for the politicians to fulfil the requirements for exercising meta-governance. The manner in which public authorities are presently co-ordinating their utilisation of ICTs will determine the electronic infrastructure these authorities must use internally and with external actors for many years to come. Accordingly, E-Government can be configured in a way that supports the politicians as meta-governors. The use of ICTs can make it easier for the politicians to communicate with the external actors, and by doing so make the framework for co-operation visible, just as they can use the technology for debates that are important in their efforts to create common understanding and meaning. But this development will not happen by itself. The politicians must perceive E-Government as a dimension of the institutional framework for network regulation, just as they must conceive of technology as part of their work to create meaning and identity, as well as in their work with the construction and support of various policy networks. At the same time, the involvement of politicians in the E-Government design process could probably lead to a further development of E-Democracy where the democratic use of ICTs would receive greater emphasis.

FUTURE TRENDS

The potential of E-Government is not yet fully discovered. In the coming years, we will see new ways of utilising ICTs, which will enhance the performance of E-Government. The field of communication between different public agencies especially can prove to increase the effectiveness of E-Government if the developers succeed in agreeing to common standards for communication and the exchange of information.

Furthermore, the field of a safe and accountable identification of the user has great potential for increased effectiveness. With the successful development of one it will be possible to design a long row of forms, which require a safe identification of the user, and as a result shift a lot of work from the civil servants in the public administration to the individual user of public services. At the same time, an accountable identification of the user will make it possible for the public authority to give the citizens access to the information about them and to follow the consideration of their own cases. By doing this, it will be possible to have a much more open administration than the one we know today.

CONCLUSION

E-Government is here to stay. Even though we still lack the results from a general study, which proves that E-Government is a more efficient way to perform public administration and establishes how much money E-Government has saved, some case studies have shown that a public administration using ICTs at least in these cases can perform more tasks, with the same amount of staff, serving a bigger population than the public administration without using ICTs (Jæger, 2003, pp. 107-110).

The further success of the development of E-Government is of course also dependent on the attitude of the citizens. If they do not have access to the technology, do not know how to utilize a computer, do not accept public services in an electronic way, or if they do not trust the electronic services but are afraid of misuse, then it will not be possible to realize the potentials in E-Government. In this way, it becomes a task for the public authorities to prevent a digital divide among the citizens.

It is still rather unclear what will happen to the development of E-Democracy. As it looks today, it is unlikely that E-Democracy will be developed as an integrated part of E-Government. On the other hand, the many small experiments with different forms for E-Democracy point to an independent development where the democratic potentials of ICTs are tested and new applications are developed. This situation is even more likely to take place if the politicians revise their role and start to interpret the development of E-Government as a part of their role.

REFERENCES

Bastelaer, B. V., Henin, L., & Lobet-Maris, C. (2000). *Villes virtuelle: Entre communauté et cité. Anslyse de cas*. Paris: L'Harmattan.

Bødker, K., Kensing, F., & Simonsen, J. (2000). *Professionel IT-forundersøgelse - grundlaget for bæredygtige IT-anvendelser*. Copenhagen: Samfundslitteratur.

Garson, D. G. (Ed.). (2000). *Handbook of public information systems*. New York: Marcel Dekker, Inc.

Hauge, N. B., & Loader, B. D. (Eds.). (1999). *Digital democracy: Discourse and decision making in the information age*. London: Routledge.

Heffen, O. V., Kickert, W. J. M., & Thomassen, J. J. A. (2000). *Governance in modern society: Effects, change and formation of government institutions*. Dordrecht: Kluwer Academic Publishers.

Hoff, J., Horrocks, I., & Tops, P. (Eds.). (2000). *Democratic governance and new technology*. London: Routledge.

Jæger, B. (2002). Innovations in public administration: Between political reforms and user needs. In J. Sundbo & L. Fuglsang (Eds.), *Innovation as strategic reflexivity* (pp. 233-254). London: Routledge.

Jæger, B. (2003). *Kommuner på nettet. Roller i den digitale forvaltning*. Copenhagen: Jurist- og Økonomforbundets Forlag.

Jæger, B. (in press). Digital visions - The role of politicians in transition. In V. Homburg (Ed.), *The information ecology of e-government*. Rotterdam: IOS Press.

Kooiman, J. (2000). Societal governance: Levels, models and orders of social-political interaction. In P. J. (Ed.), *Dating governance:Authority, steering and democracy*. (pp. 138-166). Oxford: Oxford University Press.

Löfgren, K. (2001). *Political parties and democracy in the information age. The cases of Denmark and Sweden*. Unpublished doctoral dissertation, Copenhagen University, Copenhagen.

Nixon, P., & Johansson, H. (1999). Transparency through technology: The internet and political parties. In B. N. Hauge & B. D. Loader (Eds.), *Digital democracy. Discourse and decision making in the information age* (pp. 135-153). London: Routledge.

Rhodes, R. A. W. (1997). *Understanding governance - Policy networks, governance, reflexivity and accountability*. Philidelphia: Open University Press.

Rhodes, R. A. W. (1999). *Understanding governance: Comparing public sector reform in Britain and Denmark* (Working paper 17/1999). Copenhagen.

Sancho, D. (2002). European national platforms for the development of the information society. In J. Jordana (Ed.), *Governing telecommunications and the new information society in Europe* (pp. 202-227). Cheltenham: Edward Elgar.

Stoker, G. (1998). Governance as theory: Five propositions. *International Social Science Journal, 155*, 17-28.

Sørensen, E. (2002). *Politikerne og netværksdemokratiet. Fra suveræn politiker til meta-guvernør*. Copenhagen: Jurist- og Økonomforbundets Forlag.

Van Lieshout, M. (2001). Configuring the digital city of Amsterdam: Social learning in experimentation. *New Media & Society, Under publication, (3)2*, 131-156.

Williams, R., & Slack, R. (Eds.). (1999). *Europe appropriates multimedia. A study of the national uptake of multimedia in eight European countries and Japan* (Vol. no. 42). Trondheim: Senter for teknologi og samfunn.

Williams, R., Stewart, J., & Slack, R. (in press). *Experimenting with information and communication technologies: Social learning in technological innovation*: Edward Elgar.

KEY TERMS

Accountable Identification: A way to identify a person in an electronic interaction and to give legal status to electronic documents. Different technologies have been tried out like e.g. chip cards and digital signatures.

Digital City: Usually a website, which is centered on a city, where public authorities, business and citizens can communicate and exchange information.

E-Democracy: Depending on what type of democracy it should support, ICT can be used for electronic voting, online referendums, or to support the political parties in their dialog with the voters. It can also be used to support political debate in a local community or in other political processes.

E-Government: Based on ICT, taking place in public administration, concerns electronic ways to perform administrative tasks, and the communication between the public administration and the citizens.

Meta-Governance: A new way to govern independent policy networks. The politicians have to create the frames that make it possible for other actors to participate in the policy process. They have to specify the competence of the various actors and to create meaning via a common understanding of the goals for the performance of the network.

New Public Management: Includes initiatives which relate to management of the public administration like e.g. management by objectives, clear standards, and evaluation of the quality of service. It also includes initiatives that deal with the introduction of economic incentive structures like e.g. outsourcing of public tasks and establishing of quasi-markets for public services.

Policy Network: Are centered on the provision of public services and include, beside the public administration, agents from the business community as well as from civil society, like e.g. NGO's, sports clubs or interest organisations. These agents are engaged in interdependent relations with the public administration.

E

Eight Key Elements of Successful Self-Funding E-Learning Programs

Yair Levy
Nova Southeastern University, USA

Michelle M. Ramim
Nova Southeastern University, USA

INTRODUCTION

The Greek philosopher Aristotle indicated that learning is the outcome of both teaching and practice. Clearly, learning is not confined exclusively to classroom lectures. In the past several decades, educators explored the possibilities of providing learning experiences to remote students. With improvement in technology and the growing popularity of Internet usage, e-learning caught the attention of both corporations and educational institutions. However, traditional learning methodology began transforming when elite universities embraced the Internet as a vehicle for their degree programs (Forelle, 2003). Progress in e-learning has increased its popularity in the past decade (Levy & Murphy, 2002). Consequently, it is carving a new brand of universities causing traditional schools to rethink their business model. Furthermore, some elite schools have developed specialized online degree and certificate programs. In doing so, these schools strive to compete on this new learning medium and create a new source of revenue, especially due to the declining enrollment and lower government funding resulting from the events on September 11, 2001 (Roueche, Roueche, & Johnson, 2002, p. 10). This paper provides definitions of the eight key elements any institutions should have in order to successfully implement self-funding e-learning systems.

BACKGROUND

In the past few decades, universities and colleges have been faced with a growing demand to graduate qualified students. At the same time, however, universities and colleges were faced with increased demand by local communities and governments to provide more scholarships and financial aid for local students, in spite of the reduction in financial support allocated to academic institutions (Cusick, 2003). As a result, higher education administrators have been seeking to increase their overall revenues from corporate sponsors and investors by crafting specialized degree and certificate programs. Not surprisingly, universities and colleges have been relying on international students to compensate by admitting a large number of full-fee paying foreign students (Surek, 2000). Since this has become such an important revenue stream, many business schools have gone beyond designing attractive specialized programs for international students and have even collaborated with international universities around the world to create joint programs.

In the post-September 11, 2001 era, because some of the hijackers came to the U.S. on student visas, new tougher Immigration and Naturalization Service (INS) regulations were installed to control and evaluate the issuing of student visas. These new regulations are dramatically affecting the number of international students seeking U.S. education. That impact is a result of a decrease in overall student visas issued by the INS and the sluggish process of new student visa seekers. At the same time, current international students who were already in the U.S. were forced to return to their home countries and reapply for student visas under the new regulations, and then wait months for permission to come back. As a result, online learning programs have become an alternative solution for international students seeking U.S. academic degrees.

In the past few years, Information and Communication Technologies (ICT), such as online learning, has grabbed the attention of many higher education administrators. In the late 1980s, Canadian schools invested enormous amounts of time and resources to develop learning programs for a distance delivery. U.S. schools quickly followed with some top business schools like Duke and Michigan implementing online learning programs in the 1990s. As the use of the Internet increased during the second half of the 1990s, many other U.S. universities, headed by their business and engineering schools, implemented online learning programs where almost all included one version or another of MBA programs (Davids-Landau, 2000; Forelle, 2003).

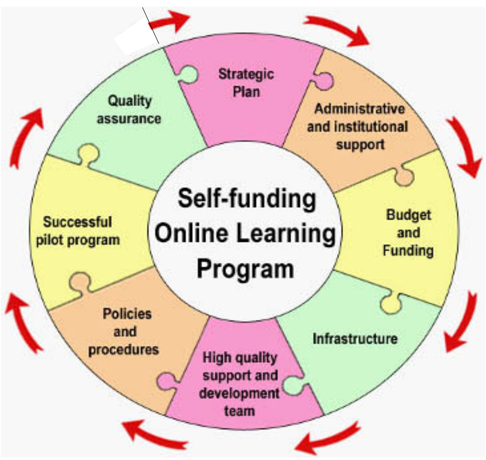
Today, more than ever, higher education administrators are very much interested, in online learning programs, as they face declining student enrollments, an aging student population, and a reduced level of federal, state, and local funding. This has resulted in a growing number of institutions that are looking for new innovative ways, mainly through the use of ICT, to attract both U.S. and international students in remote or distance locations. However, literature suggests that faculty members without prior online teaching experience have very little understanding of the skills needed to make their online teaching effective (Conrad, 2004). Moreover, literature suggests that tremendous efforts are needed also from institutions in order to prepare their faculties to effectively teach online (Chacon, 2001). Thus, the next section of this paper concentrates on the key elements needed in order to deploy a successful online learning program.

THE KEY ELEMENTS

Overview

Implementation of e-learning systems can be challenging, as only a limited number are successful in sustaining programs to the point of self-funding. In order to achieve a successful self-funding e-learning program, investment of both time and capital is essential in the period prior to implementation. Nevertheless, successful implementation of an e-learning system is a success ticket for a self-funding program. This paper will present the eight key elements of successful implementation (See Figure 1). Additionally, it will provide a roadmap that can help institutions to develop self-funding e-learning programs, along with the justifications for the importance of each key piece that comprises the overall success of the project. In the following sections, definitions of the eight key elements of successful self-funding e-learning programs are provided.

Figure 1. Eight key elements for self-funding e-learning program.



Strategic Plan

A vital step in the project is to devote adequate time for proper planning. A good strategic plan should include an analysis of all key elements presented in this paper, along with the development of a detailed blueprint—or strategic plan—of the implementation process. Such a plan should also include foreseeable problems and some suggested solutions or avenues of findings solutions to such challenges. A viable plan should be based on a gradual development process, rather than implementing a full-fledged program without proper adjustment by faculty, students, and administrators. This suggested methodology will allow institutions to progress with their projects one step at a time, while building and improving based on feedback and constructive comments from users, faculty, and administrators. At the same time, a detailed strategic plan would provide a solid plan for scalability and ability to reach a self-funding stage.

Administrative and Institutional Support

Administrative and institutional support is another key piece of successful implementation in the pursuit of a self-funding e-learning program. It is important to emphasize to the administrators the potential benefits associated with such programs and, at the same time, candidly present the challenges foreseen in such projects. The lack of knowledge related to the benefits and limitations of such technologies may cause some misconceptions and reservations about such projects resulting in roadblocks for the successful implementation.

Numerous institutions that seek to provide e-learning programs invest tremendous efforts in the development of e-learning courses, faculty training, and equipment, but lack the overall united institutional support for e-learning students. Institutional support for e-learning students goes beyond the access to the e-learning system and interactions with the professor. Students enjoy all of the benefits as if they are on-campus, but in a format that is available via the Internet or the web (Levy & Ramim, 2003). Such a support activities and benefits should include online access to: registration, financial aid, library, bookstore, advisors, student organizations, and virtual communities. In most institutions, these support functions are already available and just need integration or a single point of link to provide students' with a centralized point of entry. Such a centralized point of entry, or portal, can be created via the e-learning platform where links to the institutional support functions are provided along with links to the online courses. Furthermore, a centralized point of entry would enable a seamless integration of online and on-campus services.

Budget and Funding

Funding and budgetary issues are major key challenges for many e-learning programs. On one hand, the development of a professional e-learning program is dependent on adequate funding for leading-edge hardware and software, highly trained staff, and incidental expenses. On the other hand, many e-learning programs initially encounter either minimal funding or even insufficient funding. Such a state can trigger a "domino" effect, where the limited or decreased funding reduces the quality of support, development, and other needed resources (Alavi. Yoo, & Vogel, 1997). Consequently, funding sources must be incorporated as a key component of the implementation plan to avoid such derailing. One solution for the funding source includes an application for an institution, state, or federal grant that provides the seed funding for the project.

E-learning courses require a tremendous amount of development, student and faculty support, infrastructure, and other resources, which in most cases is not provided by the university or academic institution. Therefore, an additional fee is needed to cover such operations. This fee for e-learning courses should be collected and utilized for the maintenance of the program. For private universities, such additional fees can be added onto the general term fees in the same fashion that parking, health, or lab fees are added onto students' tuition each term. For state or public universities, the common and recommended way of collecting such funds is by adding an added-value charge for all e-learning courses as an "over tuition fee" similar to a computer lab fee. *Over tuition fee* is commonly defined as an assessed fee charged for certain classes in addition to the current state fees in order to cover the special expenses that such a course or program accrued. In some universities, both private and public, a small portion of such fee must be shared with their overhead charges, while the rest of the funding can be channeled into the program.

Infrastructure

Software

When evaluating in-house development of a customized platform, the school must be aware of both the costs and benefits associated with such avenues. If special features or customized features are needed, a customized solution of some form is probably necessary, although not always. Schools should determine to what degree their desired features are unique or non-standardized and how much effort will be needed to develop such features. If the unique features desired by the school are not addressed by any available commercial packages, a thorough investigation is needed. One aspect of the investigation should look into the costs associated with such development, including: time to develop, salaries for developers, upgrades and scalability, proprietary versus open code, training material, etc. The other aspect of the investigation should look at the benefits that will result from the development of customized platforms in comparison to commercially available solutions. If the desired features are standardized, the school would likely be better off with a commercially available platform than going through a system development life cycle.

Higher educational institutions are in the business of providing education and not in the business of software development (Levy, 2001). Therefore, a wiser way to pursue such implementation is to evaluate current commercially available packages and determine which platform best addresses the needs of the school, provides better support, allows easy upgrades, integrates best with current platforms on-campus, and is scalable. Choosing a good and credible solution can be a very critical decision. Taking the time to effectively examine and evaluate commercially available platforms may make the difference between a successful program and a costly failure.

Hardware

Evaluation of existing hardware or a proposal for needed hardware must be included early in the strategic planning stages of the project. Outsourcing and in-house are two main avenues to consider for the hardware infrastructure solution. In the event that competent employees with experience in hardware infrastructure are not available or difficult to obtain, outsourcing may be the appropriate solution. It is recommended to investigate the cost associated with large-scale operations at the initial stage. Some hosting services will provide a reduced rate for initial implementation, but dramatically increase rates once the program is growing and expanding, both in the magnitude of the hardware and bandwidth needed.

In the event that competent employees can be hired or utilized in-house, the avenue of in-house existing infrastructure and in house development should be pursued (Levy, 2001). Looking at in-house implementation of such a solution requires attention to the subcomponents needed to link the hardware piece. Such subcomponents include: networking and telecommunications, server architecture, along with backup devices and backup policies and procedures. The increased use of network traffic is a commonly overlooked component by many institutions and may cause a bottleneck effect, thereby reducing the confidence of faculty and students. Network testing and feasibility analyses prior to implementation are recommended to adequately support such a project in-house. The emerging use of minicomputers (or small to medium servers) for e-commence solutions has accelerated the development of server scalability and modular component-based methodology to hardware support. This allows institutions to start with a small installation and scale up by integrating more servers to the cluster as needed.

High Quality Support and Development Team

Institutions should not spend time and money on training professors to develop their own courses, as we don't train vehicle drivers to assemble cars. Rather, they should train professors to effectively use the tools and to deliver their course content via the Internet in the same way we train vehicle drivers to drive and obey road signs (Levy, 2001). It is best to assemble a professional support and development team to provide student support and to coordinate online course development with professors.

Support and development team can be created using in-house expertise. Often, retaining IT and instructional media developers in educational institutions is a challenge. However, institutions can provide some incentives, such as partial- or full-tuition reimbursement or other benefits in order to retain competent employees. In some cases, using an inside group can provide great first-mover advantages, as these employees know the environment and the settings of the institution, with a limited learning curve for these issues as compared with an outside group. Using an outside consulting company or an Internet course development company can be quite expensive and may require considerable coordination efforts. Furthermore, most publishers today provide canned content (electronic package or "e-pack" in short) that is based on the textbooks they published, thereby extensively reducing the development time of the course content.

Policies and Procedures

The first step to establish a high quality program should be the establishment of guidelines for official policies and procedures similar to the ones used for on-campus courses. Teaching requirements should be defined in terms of academic and professional experience essential for course instruction. Both professors and students should maintain a minimum technical proficiency to prevent the technology from becoming a hurdle in the education process (Levy, 2001). Some institutions have established a mentoring program where new faculty members are mentored by experienced online faculty members during their first online term. Such mentoring has proven vital in the event that other initial training or a web-assisted period is not feasible.

Policies and procedure manuals usually include documentation on issues such as:

- Teaching Requirements, including academic experience, professional experience, technical proficiency, and other competencies.
- New Faculty Hire Mentorship, including contacts, FAQs, technology setup, best practices.
- New Course Development, including learning objectives, online assessments types, student motivation techniques, information sources, syllabus development, and textbook selection.
- Facilitating an Online Course, including course orientations, instructor accessibility, grading, motivational techniques, continuous improvement, and evaluations.
- Student Issues, including e-registration, e-advising, e-bookstore, e-library, grievance, and online student code of conduct

SUCCESSFUL PILOT PROGRAM

The temptation to launch a full-fledged e-learning program quickly is tremendous. Nevertheless, a proper gradual process is recommended. It is recommended to start with web-assisted or web-enhanced capabilities for current on-campus courses, rather than with full-fledged online courses (Hiltz, Coppola, Rotter & Turoff, 2000). This will allow the faculty to learn the system and get comfortable with the technologies without the pressure of swiftly adjusting to a new medium of teaching. It is also recommended to start with faculty who are recognized "first adopters" of technology, as such programs rely heavily on technology and provide close monitoring and support to all students participating in online courses.

Incorporating feedback and making adjustments to policies and business processes is a valuable outcome of the pilot program. As part of the quality improvement process, feedback and adjustments for learning objectives and teaching methodologies (what works and what doesn't) is needed. Doing so right after the pilot program not only helps the institution in developing an effective and successful program with potential for increased funds, but is also useful for the accreditation purposes.

Quality Assurance

Pedagogy

A high quality course includes clearly defined course objectives and assessment tools such as the ones used in a traditional on-campus course. However, creation of an incremental assessment is recommended as a means of encouraging students' learning pace, constant interaction with the professor, and discouragement of cheating. Useful assessment measures include weekly quizzes, and multiple small and large assignments during the course of the semester. Quality assurance improves over time as the e-learning course infrastructure expands and best practices emerge. For increased effectiveness, continuous improvement and evaluations must be part of the program culture. As the program progresses, a yearly faculty retreat can facilitate discussion within the institution on best practices and improving the quality of teaching provided by the whole program.

Technology

Technology should not become an obstacle for delivering the educational experience. By the same token that schools are required to maintain their physical classrooms and to make students comfortable when learning on-campus, technology should be transparent and free of errors or down-time periods. The institution should spend time to evaluate current commercially available e-learning systems to find one that best matches their expertise and budget.

Once an e-learning system is selected, the institution should create a standardized interface thereby defining the page appearance, the institution's logo, the professor's image and biography, and common links to the institution's key sites, such as library, registration, and technical support sites. Color palettes and banners should also be uniform, as should the selection of communication tools available to the learners. This will assure consistency and reliability of the learning medium as users move from one course to another, providing a high quality experience from a technology prospective, while simplifying the job of the programming team. This is similar to the "brick and mortar" classroom features that are uniform in nature.

FUTURE TRENDS

The future trends in this area should include the exploration of successful e-learning programs and validation of this model through various case studies beyond the one documented by Levy and Ramim (2003). Moreover, as additional traditional higher educational institutions are implementing e-learning programs, it will be fascinating to explore the difficulties and challenges they face as a result of the lack of one or more of the eight key elements proposed in this article. Additional research should also focus on augmenting the model proposed here with newer formats of e-learning, such as the ones that include a mixture of synchronous and asynchronous learning tools for the same e-learning courses.

CONCLUSION

In this paper, the eight key elements for successful implementation of self-funding e-learning programs were defined and articulated. The tremendous growth and potential of e-learning in the business world and corporate training centers in the past few years is making its way to the academic world. As more and more individuals are stressed for time and the demand to expand their education and knowledge increases, greater pressure is put on administrators of higher education institutions and universities to come up with new, innovative ways and new technologies to satisfy those needs. Consequently, higher educational institutions have to spend more time investigating and planning the appropriate methods to implement their self-funding e-learning programs in the era of decreased funding.

Successful implementation of self-funding e-learning programs requires careful research, strategic planning, and forecasting. This paper reviewed and defined the eight key elements needed to achieve a self-funding level of e-learning programs. The key elements include: strategic planning, administrative and institutional support, budget and funding, infrastructure, high quality support and development team, policies and procedures, successful pilot program, and quality assurance.

REFERENCES

Alavi, M., Yoo, Y., & Vogel, D. (1997). Using information technology to add value to management education. *Academy of Management Journal*, *40*(6), 1310-1333.

Chacon, F. (2001). Transforming classroom professors into virtual class mentors. *Proceedings of the General Conference of International Associations of Universities*, Skagen, Denmark. Retrieved June 9, 2004, from: *http://www.unesco.org/iau/tfit_denmark-chacon.html*

Conrad, D. (2004). University instructors' reflections on their first online teaching experiences. *Journal of Asynchronous Learning Networks*, *8*(2). Retrieved June 9, 2004, from: *http://www.aln.org/publications/jaln/v8n2/v8n2_conrad.asp*

Cusick, R. W. (2003). College planning and alternative funding sources. *Journal of Financial Service Professionals*, *57*(1), 5-9.

Davids-Landau, M. (2000). Corporate universities crack open their doors. *The Journal of Business Strategy, 21*(3), pp. 18-23.

Forelle, C. (2003). Elite colleges finally embrace e-learning. *Wall Street Journal*, January 15, B.1.

Hiltz, S. R., Coppola, N., Rotter, N., & Turoff, M. (2000). Measuring the importance of collaborative learning for the effectiveness of ALN: A multi-measure, multi-method approach. *Journal of Asynchronous Learning Networks, 4*(2). Retrieved May 12, 2004, from: *http://www.aln.org/publications/jaln/v4n2/v4n2_hiltz.asp*

Levy, Y. (2001). E-Learning: An overview of next-generation Internet-based distance learning systems. *Proceedings of the WebNet 2001*, Orlando, Florida, October 23-27, 780-786.

Levy, Y. & Murphy, K. (2002). Toward a value framework for e-learning systems. *Proceedings of the Thirty-Fifth Hawaii International Conference on System Sciences (HICSS-35)*, Big Island, Hawaii. January 7-10, pp. 1-9. Retrieved May 12, 2004, from: *http://csdl.computer.org/comp/proceedings/hicss/2002/1435/01/14350005.pdf*

Levy, Y. & Ramim, M. (2003). Building successful self-funding online learning program: Think big…start small…build fast… *Proceedings of the Business Information Systems Conference (BIS2003)*, Colorado Springs, Colorado, June 4-6, 2003, 273-285.

Roueche, J. E., Roueche, S. D., & Johnson, R. A. (2002). At our best: Facing the challenge. *Community College Journal, 72*(5), 10-14.

Surek, B. (2000). Funding problems of technical education in developing countries. *Proceedings of the Technological Education and National Development (TEND) Conference* (2nd, April 8-10, 2000, Abu Dhabi, United Arab Emirates), ERIC Document No. ED447290.

Swan, K. (2003). Learning effectiveness: what the research tells us. In J. Bourne & J. C. Moore (Eds) *Elements of Quality Online Education, Practice and Direction*. Needham, MA: Sloan Center for Online Education, 13-45.

KEY TERMS

Administrative and Institutional Support: All the benefits students enjoy when they are on-campus, but in a format that is available via the Internet or the Web, beyond the access to e-learning courses and interactions with the professor. These include e-registration, e-financial aid, e-library, e-bookstore, e-advisors, e-student organizations, and virtual communities.

E-Learning Program: The entire organizational, technological, and administrative structure that enables student learning via the Internet.

E-Learning Program Strategic Plan: The blueprint of the e-learning program implementation process that also includes foreseeable problems and solutions to such challenges.

E-Learning Support and Development Team: The team includes: program director, a program coordinator or assistant, instructional designer, system administrator(s), multiple developers/programmers, multiple support staff (for faculty and students), graphics and video production artist, and marketing coordinators.

E-Learning System: The technological and management system that facilitates and enables student learning via the Internet.

Over Tuition Fee: An assessed fee charged for certain classes in addition to the current state fees to cover the special expenses such course or program accrued.

Policies and Procedures of E-Learning Program: A set of guidelines and requirements for faculty to be able to teach online and for students to learn online. For faculty these may include: teaching requirements, new faculty hire mentorship, new course development, facilitating an online course, and student issues; for students: a minimum technical proficiency, new students' orientation, grievance process, and an online student code of conduct.

Quality Assurance of E-Learning Program: The inspection and maintenance of a high quality program that is based on two major components: 1) the pedagogy aspect and the effectiveness of the learning experience; and 2) the technology aspect and the quality of the online course.

E-Learning Environment

Mizue Kayama
Senshu University, Japan

Toshio Okamoto
University of Electro-Communications, Japan

OVERVIEW OF E-LEARNING

Nowadays, the concept/system of *e-learning* (or *eLearning*) is widespread with the advent and prevalence of the Internet. Via the Internet, people can communicate with each other at anytime and from anywhere. People can also share, rebuild, stock, and reuse various kinds of information. Here, it is clear that e-learning gets citizenship in the educational society instead of CAI (computer-assisted instruction) and CMI (computer-managed instruction). As a response to society's advance, it is necessary to construct a new learning ecology, such as a learning organization or a learning community. To date, the need for an understanding of e-learning issues has not been met by a coherent set of principles for examining past work and plotting fruitful directions. Obviously, it would be difficult to document the many seeds sown now.

The e-learning environment is cataloged as follows (Okamoto, 2000):

- Individual learning environment with learning materials
- Group learning/collaborative learning environment with some shared tools/applications
- Classroom learning (lecturing)

This learning ecology has the mixed mode of either synchronous or asynchronous by using any teaching/learning contents and audio/visual devices, such as videoconference and other communications tools.

e-Learning is a learning/education/training style that uses information technologies. In the past, this type of learning/education/training was called various names, like "distance learning," "distance education," "cyber learning," "virtual learning," "Web-based training (WBT)," "Web-based learning (WBL)," "online learning," and so on. Nowadays, e-learning is innovated by using the latest information technologies, including WWW technology for e-learning course delivery, movie/speech compression technology for e-learning content production, and the learning technology standards (SC36, 2004), like LOM (learning object metadata), SCORM (sharable content object reference model), and collaborative technology, for keeping the interoperability of e-learning systems/contents/courses.

The main advantages of e-learning are, as is well known, from "to any place, at any time" attributes. Often, the free education aspect also appears, although much of the educational software offered today is not free, and many educational institutions offer e-learning programs at a price. Plain, text-based course materials are not enough anymore. The recent increases in bandwidth made more avenues of expression possible, images on the Internet are commonplace, soundtracks and videos are used with growing frequency, and other (multi- and mixed) media types evolved (animation, simulation, collaboration, etc.).

Before now, based on learner modeling, adaptation of teaching strategies and intelligent user adaptation in intelligent tutoring system (ITS) were developed. More recently, the field of adaptive hypermedia (De Bra et al., 1999) emerged, at the crossroads of hypertext/hypermedia and user modeling. Adaptive presentation of the educational material can mean one or more of the following: providing prerequisite, additional, or comparative explanations; conditionally including fragments and stretchtext; providing explanation variants; reordering information; etc. Adaptive navigation support can mean one or more of the following: direct guidance, sorting of links, links annotation (Brusilovsky, 1999), link hiding, link disabling, link removal, and map adaptation. Another main advantage of the Internet is that it favors collaborative work, which, in turn, favors learning (Dillenboug, 1999).

Moreover, we regard e-learning as meaningful self-development of an environment for lifelong learning. The recent technological changes are influencing our society, and people are asked to acquire new knowledge all the time. The opportunities to take education with high quality have to be provided for all sorts of people who have different backgrounds, different abilities and knowledge, and various needs. E-Learning is one answer to the rigidity of the present Web-based courses and courseware.

THE DESIGN OF AN E-LEARNING ENVIRONMENT

When we think modality of computerization on education, it is generally categorized as follows:

1. Self-study entity through electronic information media-based materials and courseware
2. Learning entity, with electronic information media (e.g., computer) as learning/problem-solving/representing/knowledge-transmitting tools
3. Learning entity about information and communication technology, social problems, etc.
4. Computerizing entity of education

The relationships among those entities should be compensated for mutually, and an e-learning cycle can be developed. The idea here is in line with building the environment for "anybody" to learn something from "anywhere" and at "anytime" in the e-society. There are two purposes for this expansion: on one hand to enlarge the study opportunity, and on the other, to develop people's new competencies.

When we build an e-learning environment, at least three issues should be considered (Okamoto, 2000). The first is the pedagogical goal representing ability and knowledge as learning objectives. The second is the subject contents. The third is the learning forms, defined by seven learning environments:

1. Distance individual learning environment for mastery learning. This environment provides courseware for knowledge and skills acquisition, i.e., the typical e-learning course, such as WBT/VOD (video on demand) systems (Hui, 2000).
2. Distance individual learning environment for discovery learning using various search engines (VOD search and navigation mechanism).
3. Distance individual learning environment for problem-solving learning using simulations, such as ILE (interactive learning environment), and so on.
4. Videoconference system in the classroom environment for discussion, instructional presentation, questions and answers sessions, and telecommunications (Chen, 2001; Nieminen, 2001).
5. Collaborative learning environment for a small group or pairs using videoconferencing, some kind of communication tool, or various applications accompanied by a screen-shared viewer and learning log tracking mechanism.
6. Collaborative simulation learning environment for different learners performing different functions in a teamwork learning pattern, and as such, forming a special skill in the learner's own domain, e.g., a collaborative activity within the jet plane's cockpit.
7. Linkage and coordination among different organizations and areas, e.g., access the online school library, online museum, and so on.

In the establishment of e-learning environments, the most important idea is to start by defining the instructional goal and then classifying learning contents that are best equipped to build the learning environment. Moreover, the research on the method is required in order to build the asynchronous collaborative learning contents (Dryden, 2001). Further research directions should be placed on the study of the learning environment, with emphasis on the virtues of individualized learning and collaborative learning. In this case, transmission of real images and voice data is required. The fundamental environment components for e-learning systems include the whole information system related to e-learning environments. It consists of several management functions, such as curriculum and learning-materials management, learners' profile and log-data management, LMS (learning management system) and LCMS (learning contents management system). In order to construct those educational management systems, we need, technologically, several data/file-processing modules, such as a distributed file system, synchronous data communications, and so on. If any applications and tools related to e-learning can be plugged in the core framework, we would build an integrated e-learning environment where learners can share and operate these software/data in real time. In addition, the total management system of e-learning is required in order to execute a real educational project or practice, which means research project management, learning schedule management, courseware development, and so on.

COLLABORATIVE LEARNING UNDER E-LEARNING ENVIRONMENT

Collaborative learning is a participants' initiative learning form that has been stressed with the paradigm shift from the teaching side to the learning side in the current learning ecology. The objectives of collaborative learning are the effective and efficient group activity and the collaborative mutual interdependence relations within the group. In collaborative learning, each learner is submitted a subtask, and he or she is expected to accomplish it. As the result, the group goal and each learner's learning goal would be achieved.

Distributed collaborative learning is a type of collaborative learning that can take place in the Internet environment (e.g., e-learning environment) with multiple learners geographically far from one another (O'Malley, 1994). Geographically, a distanced situation can mean remote or far physically, but this also covers cases where direct interaction and dialogue are not possible among participants due to other reasons. The urgent research topic is how to support distributed collaborative learning, including how to support the collaboration among plural learners according to a teacher's educational objectives. This type of learning is called CSCL (computer-supported collaborative learning). CSCL focuses not on the efficiency of group work, but on a deep and comprehensive understanding with self-reflecting and self-monitoring (Dillenboug, 1999). In general, the CSCL management software provides two types of activity space: a private working space and a collaborative working space, where the learners can exchange information in a synchronous or an asynchronous manner. Many researches are discussing these two types of activity space: the information exchange types that exist, and those that are necessary (Synnes, 1999).

The resources required in a collaborative learning (CL) environment are taken up as follows (Okamoto, 2001):

- Technologically mediated dialogue channel
- Shared workplace for a group
- Personal workplace
- Learning materials and learning tools
- Analyzing tools of data/information
- Repository/memory for data/information revealed in CL
- Reference channel for the collaborative repository
- Modeling tools for monitoring the process of CL

In order to support collaborative learning in an e-learning environment, many platforms are proposed. For instance, Timbuku (Netopia), NetMeeting (Microsoft), Media-Fusion (Apple Computer), Habanero (University of Illinois/NCSA), MatchMaker (University of Duisburg), SimPLE (University of Maryland), CSILE/Knowledge Forum (University of Toronto/Learning in Motion) and REX (University of Electro-Communications). We can also mention the performance support system (PSS). In such a system, learning activity is directly connected to problem solutions in the real world. The PSS in a virtual environment is used in many fields, such as flight (flight simulator), shipping (navigation simulator), and fire (fire simulator).

FUTURE WORK

The traditional classroom teaching method is limited in time, and it is difficult to supervise all students. In order to realize individual learning, educators must prepare and provide a tailored curriculum teaching method for each student's needs. However, it is difficult for a teacher to manage such a teaching activity when considering students' individual differences. On the other hand, the e-learning paradigm provides a solution to these problems. Adaptive/personalized/individualized learning environments seem to be more useful than the traditional classroom teaching environment. In addition, we can enhance the quality of the course contents by effectively using learning media.

However, when we build a desirable learning environment to promote learning effectiveness, we have to tackle the following questions:

- **Should it be an e-learning system?** Who are the target learners, and how far are they from the site? Are these features appropriate for the e-learning paradigm? Can we expect to include the number of learners from a time and location point of view? We should also consider the needed type of e-learning system and the technical tasks of the e-learning system based on those aspects mentioned above.
- **Should it be an intelligent, adaptive system?** Are the students very different? Do their backgrounds vary? Do the students belong to different age groups? Are they full-time and part-time students? Are there company workers among them? Does the subject of the course include features that might be relevant to different people? Do all students have to study the whole material, or should there be alternatives according to their needs? How granular is the course presented? In this case, we have to consider both the predicted student attention span and the smallest educational unit containing useful information. These kinds of questions, again, do not only point to whether an adaptive system is needed or not, but they also show what kind of adaptation is necessary.
- **Are various media needed?** Does text suffice for the presentation? Can we expect a better student reaction from different media? May the media motivate the learners? May the media make progress in their understanding of the contents? What kind of media is appropriate? What part of the e-learning system (contents presentation module, testing module, education tactics planning module, and so on) will be best suited for the media?

CONCLUSION

The retrospective of education in the 20th century is portrayed by classroom teaching and learning, group learning, and individual learning in traditional teaching and learning forms and e-learning forms. Education is changing as technology advances. Internet technologies are highly influential in developing and presenting teaching and learning activities. The Internet is used as a place to look for information, communicate, and realize self-development. Moreover, the innovation of technology will prospectively increase the speed and capacity of processing and the availability of access.

Learning needs are dynamic and vary in the computerization age. As a response to this society advance, e-learning is here to stay, as it involves the methodology of instruction, cooperative and collaborative learning, self-study, and all kinds of media, and furthermore, it engages various evaluation ideas. Nevertheless, it is important for people to form new competencies and abilities and to create and obtain sound attitudes toward the new era. In keeping with these purposes, we need to design the learning environment to foster deep and wide competency. Moreover, teachers need to put their expertise to this learning environment through their educational practices. The method of group formation and organization, the preparation of the learning contents and media, and teamwork among teachers are the most eloquent and meaningful matters.

Finally, we would like to portray some essential matters in order to achieve real success in the e-learning world:

1. Quality of contents based on well-examined curriculum (or competency library) reflecting the specified learning goals
2. Quality of services from educational and technological perspectives
3. Involvement and commitment of human mentors
4. Copying technology changes
5. Seamless information flow and easy access
6. Knowledge repository for reusing resources
7. Integration of authoring and delivering
8. Sharing functions such as “plug-ins” for applications and tools

These essential matters set the stage for the 21st century to host collaborative development of infrastructure for large-scale use by the profession, and conduct essential research into the advancement of education.

REFERENCES

Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia. *User Modelling and User-Adaptated Interaction*, *6*, 87–129.

Chen, N., & Shin, Y. (2001). Stream-based lecturing system and its instructional design. In Okamoto T., Hartley, R., Kinshuk & Klus, J. (Eds.) *Proceedings of International Conference of Advanced Learning Technologies*, IEEE Computer Society, CA, USA, (pp. 94–95).

De Bra, P., Brusilovsky, P., & Houben, G. J. (1999). Adaptive hypermedia: From systems to framework. *ACM Computing Surveys*, *31*(4), 12.

Dillenboug, P. (1999). *Collaborative learning, cognitive and computational approaches*. Advances in learning and instruction series. Oxford; Elmsford, NY: Pergamon Press.

Dilley, J., & Arlitt, M. (1999). Improving proxy cache performance: Analysis of three replacement policies. *IEEE Internet Computing*, *3*(6), 44–55.

Dryden, G., & Vos, J. (2001). *The learning revolution*. Stafford, UK: Network Educational Press Ltd.

Hui, S. (2000). Video-On-Demand in education. Retrieved from http://www.cityu.edu.hk/~ccncom/ net14/vod2.htm

ISO-IEC JTC1 SC36. (2004). SC36 HomePage. Retrived from http://jtc1sc36.org/

Nieminen, P. (2001). Videolecturing for international students. In Ruokamo, H., Nykänen, O., Pohjolainen, S., Hietala, P. (Eds.), *Proceedings of International PEG Conference*, Tampere University of Technology, (pp. 162–168).

Okamoto, T. (2000). A distance ecological model to support self/collaborative-learning via Internet. In S.S. Young, J. Greer, H. Maurer, Y. S. Chee (Eds.) *Proceedings of the International Conference of Computer on Education 2000*, AACE-APC/National Tsing Hua University, Taiwan, (pp. 795–799).

Okamoto, T., Kayama, M., & Cristea, A. (2001). Considerations for building a common platform of collaborative learning environment. In Lee, C. H., Lajoie, S., Mizoguchi, R., Yoo, Y. D., and du Boulay, B. (Eds.) *Proceedings of the International Conference of Computer on Education 2001*, AACE-APC/Incheon National University of Education Publications, Seoul, Korea, (pp. 800–807).

O’Malley, C. (Ed.). (1994). *Computer supported collaborative learning*. NATO ASI series, Vol. F-128. Berlin: Springer-Verlag.

Synnes, K., Parnes, P., Widen, J., & Schefstroem, D. (1999). Student 2000: Net-based learning for the next millennium. In P. De Bra and J. Leggett (Eds.)*Proceedings of the World Conference on the WWW and Internet 1999*, AACE, VA, USA, (pp. 1031–1036).

KEY TERMS

Collaborative Learning: This is a form of learning that involves collaborative learning processes. It is designed for coaches, helpers and faculty, and groups of learners to fulfill the learning objectives of groups and of each learner through sharing resources and interacting.

E-Learning: A learning/education/training form/shape using information technologies. E-learning can provide learning/education/training services for anyone, from/to anywhere, and anytime. By network technology and learning technology, the following matters work at the e-learning environment: blending of learning and teaching methods (virtual classroom, simulation, collaboration, community, and classroom); supporting learning and teaching services (from assessment through testing and certification); online administration (handling user and course registration and monitoring learner progress); and so on.

Learning Contents Management System (LCMS): This is a system that is used to create, store, assemble, and deliver personalized content in the form of learning objects in an e-learning environment. LCMS includes and combines the functions of a content repository/database or CMS (content management system) and a LMS. Although the LMS has functions to manage the learning courses, the LCMS has functions to manage the content and permit locating stored content, authoring new content, attaching metadata to content, and managing versions of content.

Learning Management System (LMS): An e-learning infrastructure with real-time databases that deal with user (learner, coach, faculty, and so on) information, including the user's learning competencies, learning objects for each type of learning style and form, and learning activity and performance log tracking. An extended LMS may also support authoring, performance assessment, classroom management, competency management, knowledge management, certification and compliance training, personalization, mentoring and coaching, and communication.

Learning Object: This is any entity (digital or nondigital) that may be used for learning/education/training. A learning object is usually the smallest unit of instruction managed by a LMS. However, a learning object may grow increasingly complex, have any internal structure, and may get more size or granularity. In order to reuse, learning objects are described by metadata (LOM : learning object metadata).

Learning Technology Standard: This is a *de jure* standard for learning/education/training with information technology. This type of standard includes a formal accredited normative specification or set of guidelines, typically including conformance criteria. A standard is created by a formal standards development organization (SDO), like the European Committee on Standardization (CEN), the International Organization for Standardization (ISO), and the IEEE Learning Technology Standards Committee (IEEE LTSC).

Electronic Commerce Policies for Dutch SMEs

Maria-Eugenia Iacob
Telematica Instituut,[1] The Netherlands

Piet Boekhoudt
Telematica Instituut,[1] The Netherlands

Freek Ebeling
Telematica Instituut,[1] The Netherlands

INTRODUCTION

Dutch small and medium enterprises (SMEs) are one of The Netherlands' most important sources of wealth creation and employment. No less than 99% of private enterprises consist of medium and small-scale businesses. They provide employment for 2.3 million people (60% of the Dutch labor force) and account for 52% of the national income generated in the private sector (www.mkb.nl/mkbnederland/english.shtml).

Today, the unused potential for SMEs, from adopting electronic commerce, is huge, although this varies by industry, size, and segment of the diverse SMEs. Therefore, enabling the SMEs to engage in the digital economy is one of the key priorities of the Dutch government. This line of thinking is promoted through the "Nederland gaat digitaal" ("The Netherlands go digital") program (www.nederlandgaatdigitaal.nl).

In this article we present an overview of the approach taken by the Dutch government regarding the development of the information society and the diffusion of electronic commerce in small and medium organizations. Our analysis also includes an assessment of the current situation of SMEs with respect to the adoption of electronic commerce and their position with respect to the governmental policies. Furthermore, we refer to policy implementation instruments ("iMPact" and "ASPect" projects) in the area of electronic markets and application service provision.

BACKGROUND: THE PLACE OF SME IN THE ICT POLICY IN THE NETHERLANDS

Policies and Governmental Initiatives

"Nederland gaat digitaal" ("The Netherlands go digital") is the central slogan of a broad campaign the Dutch government started in 2000 with the intention to modernize and change many aspects of Dutch society. The idea of this campaign was to proactively stimulate the use of information and communication technology (ICT) in all the layers of the society: by citizens, government, education, healthcare, and companies. Five ministries are coordinating and financing programs that fall within their areas of responsibility:

- The Ministry of Interior Affairs is taking care of the diffusion of ICT in the relationship between the citizen and the state (e.g., city halls are opening digital offices).
- The Ministry of Justice is coordinating the development of a legal framework that regulates the use of Internet for communication and business.
- The development of a communication infrastructure is under the supervision of the Ministry of Transport, Public Works, and Water Management.
- The Ministry of Education, Culture, and Sciences is coordinating the diffusion of ICT in the educational system.
- The Ministry of Economic Affairs has developed, together with a number of organizations—ECP.NL–Electronic Commerce Platform Netherlands, Syntens,[2] MKB[3] Nederland,[4] and industrial branch organizations—its own SME campaign "Nederland gaat digitaal" (for 2000-2001), aiming to enhance the involvement of SMEs in e-commerce. Furthermore, specific programs were defined to support the industrial branch organizations to develop strategies with respect to ICT, such that these, in their turn, can inform and advise their members accordingly.

During 2001, the Ministry of Economic Affairs collected information from Syntens, MKB Nederland, and other sources showing that there exists a great interest from SMEs for a continuation of the program (see the NIPO Survey, Table 1, Ministerie van Economische Zaken, 2001; NIPO, 2001). The result was that 71% of the SMEs

Table 1. NIPO evaluations (see Snelders & Eecen, 2003)

Category	**4-th quarter 1999 (before the start of "Nederland gaat digitaal")**	**1-st quarter 2001 (half way evaluation)**	**2003**
Percentage of SMEs with internet connection	55%	68%	83%
Percentage of SMEs that do electronic business	18%	36%	66%

were aware of the program and of its goals. Consequently, a second phase of "Nederland gaat digitaal" was initiated.

The new program (2002-2005, 22,7 millions EUR), "The SMEs in the Digital Delta," is targeting companies with less than 100 employees with concrete goals: in the end of the program, two-thirds of the SMEs should be present on the Internet with at least a "shop window" Web site, and two-thirds of the SMEs should do business transactions via the Internet. The plan identifies four obstacles that are in the way of these targets: SMEs do not know what business opportunities the Internet can offer, they don't have in-house *ICT expertise* and knowledge, they don't *trust* the Internet (legally and technologically), and often they are not financially able to support innovation. The approach taken by this new program is dealing only with the first two of the aforementioned problems and in short is the following:

1. To increase the *awareness of SMEs related to ICT opportunities* and to provide them with tailored knowledge, an intensification of the former MKB campaign "Nederland gaat digitaal" is planned. Namely, supplementary founding is directed towards workshops, consultancy, information materials, training and support of branch organizations and of their members, and development of example pilot projects for a number of sectors, also via two projects, iMPact and ASPect, in the areas of e-markets and ASP.
2. With respect to the *enhancement of trust*, the following activities were planned:

   • The government must play an active role in the development of the legal framework for e-commerce. Although, the Dutch legislation can support electronic transactions within acceptable limits (see the white paper on "Legislation for the Electronic Highway," Lower House of Dutch Parliament, 1998), efforts are being made for implementing the EU directives. Several important Internet and e-commerce regulations are already operational, and a number of bills are currently discussed in the Parliament: Distance Sales Act (see Staatsblad, 2000b), Personal Data Protection Act (see Staatsblad, 2000a), Act Regarding the Legal Protection of Databanks, (see Staatsblad, 1999), Electronic Signatures Act (see Staatsblad, 2003), Electronic Commerce Bill (see Lower House of Dutch Parliament, 2001b), Bill regarding Electronic Money (see Lower House of Dutch Parliament, 2001a).

   • Apart from legislation, an important role is reserved for self-regulation. Basically, self-regulation assumes that companies conducting business on the Internet adhere to and act according to a code of conduct. The task of defining and disseminating such a Code of Conduct was carried out in The Netherlands by a non-profit organization, The Electronic Commerce Platform Netherlands (ECP.NL). We will mention here only two of the products that relate ECP.NL to the needs of SMEs: the "*Model Code of Conduct for Electronic Business," version 4.0* (see ECP.NL, 2001; this Dutch initiative was acknowledged as a model by a number of international organizations such as OECD, UN, and the European Commission); and "*Nederland gaat digitaal, netjes volgens het boekje*" ("Netherlands go digital, by the book") (see ECP.NL, Syntens & EZ, 2002, and www.nederlandgaatdigitaal.nl/regels/). The latter explicitly addresses the SMEs. It offers practical guidance related to the e-commerce acts and regulations, by means of questions, answers, and practical examples.

   • The reliability of the communication infrastructure and the security of information exchange over the Internet are targeted in two other national programs (NACOTEL and the campaign "Internet and Security").

The European E-Commerce Policy Landscape

Most of the European governments have recognized the fact that e-business has drawn a clear separation line

Table 2. The use of the Internet in SME–2002 (see CBS, 2002, 2003; Snelders & Eecen, 2003)

	2002 (%)		2003 (%)	
Category	**5-20 empl.**	**20-100 empl.**	**5-20 empl.**	**20-100 empl.**
With internet connection	78	88	85	97
With a Website	48	61	51	85
Electronic banking	66	78	76	83
Ordering	34	44	50	71
Payments	16	19	3	5
Offering product information	47	59	96	
Receiving orders online	33	37	48	
Receiving payments online	10	10	10	
Delivering digital products	9	10	16	
Customer support	14	21	24	

between big companies and SMEs. The costs of introducing and maintaining ICT in small companies is significant when compared with their turnover, and therefore they cannot afford experiments and expensive mistakes. Therefore the *rationale* behind all the e-business policy programs was to help SMEs not to make such mistakes and eventually to support them to make the right step towards e-business. Governments (e.g., in the UK, The Netherlands, Greece) have argued the need for such actions based on substantial surveys and large consultations with SMEs, SME (branch) associations, business and IT experts, and academia. Most of these programs have explicitly set as a primary *target* increasing the e-business awareness. Specialized "awareness programs" have been started by several other member states (UK/Scotland—*First Steps Workshop Series*, Austria—*Ecaustria (Let's E-Biz)*, Sweden—*SVEA*, Germany—Lower Saxony (the *B-On-Line Project)*, Spain—*Catalunya on the Net*. In terms of ambitions, each national framework program strives to secure a leading position for that nation in the digital economy.

Significant differences between these programs appear in the selected *implementation* mechanisms, which take various forms: broad consultation forums, consulting, general training, Web-based e-learning tools and toolboxes, tailored e-business strategies, or direct financial support for IT introduction.

MAIN THRUST OF THE ARTICLE: E-COMMERCE AND SMES – THE STATE OF AFFAIRS

In what follows we will present several statistics and figures that will bring some light over the current situation regarding the presence of e-commerce in SMEs.

The introduction of electronic commerce can theoretically lead to three types of improvements in an organization: efficiency (doing things better), effectiveness (doing things differently), and innovation (doing other things).

Table 3. E-commerce innovations (Snelders & Eecen, 2003)

	2002 (%)		2003 (%)	
Category	**5-20 empl.**	**20-100 empl.**	**5-20 empl.**	**20-100 empl.**
With internet connection	78	88	85	97
With a Website	48	61	51	85
Electronic banking	66	78	76	83
Ordering	34	44	50	71
Payments	16	19	3	5
Offering product information	47	59	96	
Receiving orders online	33	37	48	
Receiving payments online	10	10	10	
Delivering digital products	9	10	16	
Customer support	14	21	24	

- *Efficiency*: In 2003, 75% of the SMEs stated that by using electronic commerce, they have more efficient business processes (in terms of time and lower costs), 58% that they have less administration work, and 12% that for them doing business electronically leads to an increase of the turnover (see Snelder & Eecen, 2003).
- *Effectiveness*: An enhancement of effectiveness almost always assumes a partial or total redesign of the business processes or even an entire new business model. From this point of view, there still is little progress for SMEs (see Table 2): only 34% of them have changed their internal business processes. However, the number of companies that perform some form of electronic transactions (even occasionally) rose in 2003 from 59% to 66% (see Snelders & Eecen, 2003). Most of this growth is due to businesses with less than 20 employees.
- *Innovation:* The most significant novelties brought by e-commerce are those of providing a new transaction channel and accessing new market segments. As a consequence, many companies have reconsidered their strategies with respect to marketing and procurement/sales, and even have developed new products/services to strengthen their position in the market. Still, things are not moving very fast in the SME community. According to MKB Nederland, FHI, and Nederland-ICT (2002), until now most SMEs see e-commerce (in particular having a Web site or an online storefront) only as a cheaper alternative to advertising.

Table 3 gives more detailed information regarding the penetration of the different types of e-commerce innovations among SMEs.

FUTURE TRENDS

Obstacles

Surveys completed during the period 1999-2003 (see NIPO, 2001; MKB Nederland et al., 2002; Snelders & Eecen, 2003) have revealed that SMEs are *still* confronted with several "e-commerce-related" problems. Again, the most important one is the *insufficient expertise and lack of accurate information*: 37% of SMEs still believe that for them this is the biggest obstacle in the way of e-commerce (in 2003 this percentage decreased to 11%; see Snelders & Eecen, 2003). This is perceived as high risk, since SMEs are often not able to express clearly what they need from their ICT suppliers. Moreover, the opinion of SMEs is that the ICT market is not transparent because of a multitude of reasons:

- They cannot have any means to measure the quality of ICT products on the market, and therefore they cannot establish a relation between the price and quality or quantify them.
- They do not understand the ICT-jargon.
- There are too many different ICT products/services.
- They don't know how to compare two IT products and to see how they differ.
- Things are changing too fast and too often in ICT.

The majority (60%) of SMEs think that their ICT suppliers should more seriously consider products tailored to specific branches. With respect to their need for objective information, SMEs have expressed various requests: 20% need more information on ASPs, 22% on e-business, 25% on e-logistics, 34% on administration, and 41% on CRM. However, most need more information on prices and costs (54%) and functionality of ICT products/services (66%).

A second obstacle perceived by 6% of the businesses (Snelders & Eecen, 2003) is the *maintenance of information systems*: half of the SMEs administrate their ICT-infrastructure themselves; a quarter of them use the services of a person, which is either an acquaintance or a hardware supplier; and the rest occasionally hire a system administrator. Although SMEs have declared that in general they are satisfied with government programs like "SME in the Digital Delta," they feel that more attention should also be paid to problems related to internal administration of information systems. SMEs often do not have them completely under control, and if however they decide to transact online, they take great risks.

Finally, SMEs consider that the program of *the government does not stimulate enough the effective implementation* of e-commerce. "Nederland gaat digitaal" succeeded to raise the SMEs interest in electronic commerce (67% of businesses are aware of the program, and 63% acknowledge Syntens as the preferred source of information regarding Internet and e-business; Snelders & Eecen, 2003), but an investment in e-commerce implementations is not part of it.

Policy Implementers: iMPact and ASPect

Providing the Dutch SMEs with tailored knowledge is the most direct way to help them make informed decisions concerning the suitability of conducting business online. For this purpose, the program "SME in the Digital Delta" has (among others) selected two themes:

- The *electronic marketplace model*, as a virtual intermediary between potential buyers and sellers. The government considers that electronic marketplaces can be very attractive for many SMEs, not only because it facilitates potentially more commercial transactions, but also because it makes them more efficient, allowing in many ways significant decreases of costs.
- The *application service providing model*, as a relatively new solution for outsourcing applications and related services and making these applications remotely accessible via the Internet on a rental basis. The ASP model offers a solution for SMEs willing to outsource certain ICT solutions. This will allow them to focus on their core activities, instead of spending resources on the operational administration of ICT applications and the specific requirements (such as expertise, professionals, financial investments, etc.) for it. Besides, such a model can fill the need for otherwise very expensive applications (such as ERP solutions), while paying, in principle, only for the effective usage of them. Finally, SMEs no longer have to worry about acquisition of newer software versions and can easily experiment with new functional features of such updates. Since SMEs have limited abilities to cope with ICT, and considering the aforementioned advantages, the government considers ASP to be an interesting solution.

The two research projects (iMPact and ASPect) developed around these themes are to be carried out over two years (2002-2004) by Telematica Instituut and its partners: Syntens and the branch organizations. Their roles are explained below.

- Telematica Instituut covers the practical research work for the dissemination and introduction of e-marketplace and application service, providing knowledge in the SME sector.
- Syntens is acting as an intermediary between Telematica Instituut and the SMEs. All iMPact and ASPect deliverables are transferred into Syntens hands, which in turn will disseminate them among SMEs via its national SME consultancy network.
- Branch organizations represent the interests of their SME members. On one hand, together with Syntens, they help SMEs to receive all necessary advice regarding e-marketplaces and ASP. On the other hand, they are providing the project research teams with branch-specific information. That will allow them to define a branch strategy for electronic commerce by building pilots and examples tailored to the needs of each branch.

Supporters of these projects are ECP.NL and MKB Nederland.

The work in iMPact and ASPect is carried out along three directions:

1. *Knowledge Management:* This direction is devoted to the development of business, organizational, legal, and technological knowledge.
2. *Education:* This direction will develop training materials and courses for Syntens advisors and staff members of the branch organizations.
3. *Exploration and Analysis of Branches:* Within this direction several industrial branches are surveyed. The aim is to define adequate adoption models for e-marketplaces and ASP. The project teams are attempting to involve primarily large branches in these activities.

CONCLUSION

Without any doubt the policies of the Dutch government are leading to important changes in way the branch organizations, and subsequently the small and medium enterprises, are looking at electronic commerce. However, for SMEs there still is a way to go and several milestones to leave behind. Some of them, such as expertise and skills, can be improved trough projects like iMPact and ASPect. For others, like standardization and security technologies, the ICT forums and industry are striving to find better solutions. Financial obstacles, however, must be taken over by SMEs themselves.

REFERENCES

CBS. (2002). *De digitale economie 2002*. Centraal Bureau voor de Statistiek, Voorburg/Heerlen, 2002, ISSN 1568-2773, ISBN 9035730259.

ECP.NL. (2001, October). *Model code of conduct for electronic business, version 4.0,* ISBN 90-76957 04 5, ECP.NL, Leidschendam.

ECP.NL, Syntens & EZ. (2002, September). *Nederland gaat digitaal, netjes volgens het boekje, spelregels voor elektronisch zakendoen*. T. Wagemans & P. Koudstaal (Eds.). ISBN: 90-76957-07-X (in Dutch).

Lower House of Dutch Parliament. (1998). *Wetgeving voor de Elektronische Snelweg*. Parliamentary Documents 1997-1998, 25880, No. 1-2 (in Dutch).

Lower House of Dutch Parliament. (2001a). *Elektronisch geld*. Parliamentary Documents 2001-2002, 28189, No. 1-2, 3, A (in Dutch).

Lower House of Dutch Parliament. (2001b). *Elektronisch handle*. Parliamentary Documents 2001-2002, 28197, No. 1-2, 3, A, B (in Dutch).

Ministerie van Economische Zaken. (2001, July). *Het MKB in de digitale delta, plan van Aanpak van de toepassing van Internet en electronic commerce door het MKB.* Gravenhage (in Dutch). Retrieved December 1, 2003 from *www.ez.nl/upload/docs/Kamerbrieven/PDF-Documenten/01033996-bijlage.pdf*

MKB Nederland, FHI & Nederland-ICT. (2002). *ICT voor productiviteitssprong in MKB.* Delft, September 9 (in Dutch).

NIPO. (2001). *Nederland gaat digitaal, 1-metin*g. Survey completed for the Ministry of Economic Affairs in the first quarter of 2001. NIPO Business Monitor (in Dutch).

Snelders, H.A.J. & Eecen, A.M.D. (2003). *Rapport Elektronisch Zakendoen 3-meting, Nipo het marktonderzoekinstituut.* Amsterdam (in Dutch).

Staatsblad. (1999). Act regarding the Legal Protection of Databanks. *Bulletin of Acts, Orders and Decrees 1999, 303* (in Dutch).

Staatsblad. (2000a, July). Personal Data Protection Act. *Bulletin of Acts, Orders and Decrees 2000, 302* (in Dutch).

Staatsblad. (2000b, December). Distance Sales Act. *Bulletin of Acts, Orders and Decrees 2000, 617* (in Dutch).

Staatsblad. (2003). Electronic Signatures Act. *Bulletin of Acts, Orders and Decrees 2003, 199* (in Dutch).

KEY TERMS

Application Service Provision (ASP): The provisioning to individuals and companies of software applications and ICT-related services via the Internet, or other data networks, that are to be paid on a rental/usage base.

Electronic Business (E-Business): The conduct of business through the Internet. It is a more generic term than e-commerce because it refers to not only buying and selling, but also servicing customers and collaborating with business partners with the scope of enhancing the efficiency and effectiveness of business processes. An example of activity falling in the area of e-business is the *front office* (client-oriented business processes such as sales, catalogues, CRM, etc.) *back office* (internal business processes such as administration and logistics) *integration*.

Electronic Commerce (E-Commerce): The sale of products and services over the Internet.

Electronic Markets (E-Markets): An online environment where buyers and sellers use the Internet as a communication platform in order to exchange information, goods, and services, independent from time and space.

Online Storefront: A Web site, available on a public network such as the Internet, which offers goods and services for sale. An online storefront is the virtual electronic version of a physical store/place of business that a customer would visit to purchase goods and services.

Small and Medium Enterprises (SMEs): A heterogeneous group, including a wide variety of enterprises, having very different levels of sophistication (in terms of products and services offered, and skills of the workers), and operating in various markets and institutional environments. The quantitative definition of SMEs varies by country and is usually based on the number of employees and the value of assets or turnover. An example of such a definition is given in the "Commission Recommendation of 3 April 1996 (96/280/EC) concerning the definition of small and medium-sized enterprises."

Self-Regulation: "Self-regulation means that businesses involved in electronic business voluntarily undertake to comply with certain rules of conduct when dealing electronically with others. Self-regulation can take different forms, e.g., adopting a code of conduct or participating in a national or international trustmark scheme" (Recommendation 32: E-Commerce Self-Regulatory Instruments; retrieved December 1, 2003, from *www.unece.org/cefact/rec/rec32/rec32_2001_%20final.pdf*).

ENDNOTES

[1] Telematica Instituut is a non-profit research institute managed and funded by top companies and the government. It is a partnership between the business community, research centers, and government, to translate fundamental knowledge into market-oriented research, in the field of telematics for the public and private sectors.

[2] Syntens is a national innovation network for SMEs founded by the Dutch Ministry of Economic Affairs. It delivers expertise, consultancy, and guidance in the area of technological innovation.

[3] MKB is the Dutch abbreviation for SME.

[4] The Royal Association MKB-Nederland is one of the largest employer organizations in The Netherlands. MKB-Nederland promotes the interests of small and medium Dutch enterprises as an influential negotiating partner for local, national, and European authorities.

Electronic Government Strategies and Research in the U.S.

Charles C. Hinnant
The University of Georgia, USA

Steve Sawyer
The Pennsylvania State University, USA

INTRODUCTION

Since the mid-1990s, adoption of wide-area computer networks, such as the Internet and the World Wide Web (WWW), by the public, educational institutions and private sector organizations has helped spur an interest in using these new Information and Communication Technologies (ICT) as a means to increase the efficiency and effectiveness of organizational processes. Private sector firms have focused on using Internet-based technologies, especially the browser-based technologies of the WWW, as a means to conduct business transactions. The use of such electronic transmission technologies in carrying out business activities has generally been dubbed electronic commerce, or e-commerce (Schneider, 2003). Attempts to reinvent public organizations in the United States during the 1990s were heavily grounded in the belief that the adoption of new forms of ICT will streamline both service generation and delivery (Osborne & Gaebler, 1993; Gore, 1993). Some government actors and observers, such as the National Science Foundation, have more recently referred to the overall use of ICT to carry out the activities of government institutions as digital government. The term *digital government* has in many respects grown to refer to the development, adoption or use of ICT as a key component of a public organization's internal information and control systems, as well as any use of ICT to facilitate interaction with external stakeholders. Some scholars have attempted to examine how governments have used ICT systems, such as the Internet and WWW, as a means to facilitate interactions with citizens and other stakeholders in an attempt to foster democratic processes via electronic media. These activities have been called electronic democracy, or *e-democracy*. This broad concept is then usually subdivided into two subsets of activities, electronic politics and electronic government. Electronic politics, or *e-politics*, centers on activities that facilitate civic awareness of political processes, as well as the ability of citizens to participate in those processes. Electronic government, or *e-government*, includes the use of ICT by government agencies to provide programmatic information and services to citizens and other stakeholders (Watson & Mundy, 2001).

CURRENT E-GOVERNMENT STRATEGIES AND APPLICATIONS

While the use of ICT in public organizations is far from a new phenomenon, using them to communicate directly with, and provide services directly to, government stakeholders is a relatively new occurrence. Since the Internet and the hypertext-based WWW provide a ready-made communication system that can be accessed and navigated with the use of graphical user interfaces (GUIs), such as the now ubiquitous browser, a great deal of attention has been focused on how such ICT can be employed strategically to alter traditional linkages with various government stakeholders, as well improve intra-organizational activities. In regards to external linkages, these strategies can be organized into one of many activities focused on altering government interaction with citizens, business, or other government entities.

Government-to-Citizen (G-to-C) Activities

Communicating with the citizens served by public organizations, may include developing an organizational website with information about government programs, as well providing additional methods of communicating with public employees. In addition, it is now quite common for citizens and other stakeholders to pay taxes, pay license fees, or conduct other transactions with public organizations via ICT. Such activities streamline traditional transactions between government and stakeholders, therefore, providing an opportunity for increased levels of government accountability, as well intra-organizational efficiency. More advanced G-to-C activities include providing another means for citizens to become involved in various political processes. For example, several federal regulatory agencies, such as the Environmental Protection Agency (EPA) and Federal Communications Commission (FCC), are developing new methods of allowing citizens to participate in regulatory activities, such as rulemaking, through the use of Internet-based systems.

Government-to-Business (G-to-B) Activities

While business firms also benefit from the ready accessibility of basic programmatic information, the potential to conduct electronic transactions via the Internet is perhaps of greatest interest to businesses that sell goods and services to public organizations. It is important to realize that various government purchases account for a sizeable percentage of all goods and services sold each year. Federal executive agencies alone purchase over $200 billion worth of goods and services each year. Agencies such as the General Services Administration (GSA), with its Federal Procurement Data Center, are making use of new forms of ICT in order to process and track governmental purchases. Moon (2002a) indicates that many state governments are also attempting to use ICT to improve procurement systems. For example, many states now post solicitations for bids online, make use of electronic ordering, are automating procurement systems, and are adopting purchasing cards. State governments, however, have been slower to adopt the use of digital signatures, Internet-based bidding, or reverse auctions.

Government-to-Government (G-to-G) Activities

The ability of new ICT to enhance communication and information sharing among public organizations with similar goals is often seen by public agencies as a major reason for adopting such technologies. This is especially true in certain functional areas, such as law enforcement and national security. For example, many law enforcement organizations maintain crime databases that are not readily shared outside of their jurisdiction. Some state-level initiatives, such as the Justice Net project in Pennsylvania, have attempted to overcome the social and technical barriers of sharing databases maintained by local, state, and federal government agencies. At the federal government level, the perceived need for increased coordination in response to national security concerns was a primary reason for the reorganization of several established agencies under the institutional umbrella of the Department of Homeland Security (Yen, 2004). As such institutional reorganizations take place, increased ICT compatability is sought as a means of enhancing social coordination both within and between government organizations.

Development of E-Government Applications

The traditional approach to developing and deploying ICT to support and enable government activities was to devise purpose-built specialty systems (Danziger, Dutton, Kling, & Kraemer, 1982). In recent years, governments have increasingly employed Commercial Off-The-Shelf (COTS) products as well. This latter trend in systems development has been spurred on as digital technologies have become more powerful, incorporated Internet capabilities, and become more reliable. This dual approach of ICT development – purpose-built systems to support the specialized functions of government (such as voting and public safety) and attention to COTS (to support human resources, payroll, and other standard operational processes) – is representative of current attempts to development e-government applications.

The use of COTS in e-government has often focused on the emphasis on implementing Enterprise Resource Packages (ERP) to support large scale transaction processing and operational activities. Such efforts with respect to ERP are emblematic of the move towards developing and adhering to concepts of enterprise-wide architecture and systems standards within large public organizations (NASCIO, 2002). The adoption of such technologies in an attempt to enhance government operations has also focused attention towards efforts to integrate commercial ICT into an organization's social structure. Purpose-built systems are often developed for a myriad of advanced Internet-based applications since their goals are usually unique to governmental functions. For instance, the delivery of complex government information to external stakeholders has been a focus of specialized application systems such as those used to allow for enhanced visualization of large federal data sets (MacEachren, Hardisty, Dai, & Pickle, in press). Likewise, increased interest in online voting has initiated development of ICT-based voting systems that are intended to support e-democracy and bring about greater civil engagement (Chadwick, 2004). Whether designed to enhance G-to-C, G-to-B, or G-to-G activities, increased interest in using ICT to support governmental activities made the development of e-government applications a vibrant and innovative sector of the broader IT industry.

Researchers believe that the adoption of e-government strategies and applications will evolve over time, noting that certain types of activities, such as posting basic program information, are relatively simple practices to adopt; while other activities, such as the processing of financial transactions (taxes, license payments, etc), require more technical capacity and sophisticated knowledge of information management. West (2004) highlights four stages of e-government evolution that correspond to real transformation in regards to how public organizations operate. These stages including the *billboard* stage in which public organizations focus primarily on providing programmatic information to stakeholders, the *partial-service delivery* stage in which public organizations offer some online services, the *portal* stage in which information and services are integrated, and *interactive democracy* which would offer online methods of public outreach, such as two-way communication. Other researchers, such as Moon (2002b), conceptualize e-government evolution into five stages. These include *information dissemination, two-way communication, service transactions, vertical and horizontal integration, and political participation*. Regardless of the conceptual framework, the core belief is that public organizations will adopt more advanced e-government strategies and applications over time as they are perceived to be beneficial and as core resources allow.

CURRENT STATUS OF E-GOVERNMENT RESEARCH

The growing accessibility of the Internet and WWW to citizens, stakeholder groups, and a variety of public and private sector organizations brought about a growing belief that such technology could be used to transform traditional bureaucratic arrangements. For example, some researchers have examined how website characteristics, such as levels of interactivity and organizational transparency, might serve as indications of how open or transparent a public organization is to various stakeholders. By examining the websites of national-level agencies, such research has served to document the rapid adoption of WWW-based technology by national governments and has sought to examine how the use of ICT may lead to greater levels of accountability in regards to governmental activities (La Porte, Demchak, & Friis, 2001). Other researchers have adopted a dual approach to examining the extent to which national governments have adopted e-government. For instance, a research study conducted by the consulting firm Accenture examined the level of e-government services offered by 22 countries and also conducted citizen surveys in 12 countries to assess how citizens viewed e-government. The results indicate that Canada, Singapore, and United States lead in the quality and availability of e-government services (Accenture, 2004).

Other scholars have attempted to explore how ICT may serve to transform broader institutional relationships. For example, some scholars have examined how administrative reform movements, such as the reinventing government movement that was initiated during the Clinton Administration, are based at least partially on the presumption that ICT can be used to alter bureaucratic arrangements and, therefore, increase the efficiency and effectiveness with which public services are produced (Gore, 1993). In her examination of federal agencies during the 1990s, Fountain (2001) suggests that WWW-based technologies, such as agency websites, internal ICT networks, and cross-agency systems, can at least potentially bring about increased levels of efficiency and effectiveness. By examining public organizations as they undergo broader administrative reforms, this research highlights the underlying assumption that successful institutional reforms are dependent on harnessing ICT.

While some researchers examine national governments, other researchers have focused primarily on studying the e-government activities of state or local governments. West (2004) has examined the extent to which state governments have adopted new forms of ICT and the extent to which such technologies have influenced the delivery of public services, democratic responsiveness, and public attitudes. While acknowledging that most state governments have adopted the early stages of e-government, he believes that many of the purported benefits of e-government have yet to be realized. For example, most state government now have program information available via the WWW and many states allow for online transactions, and yet most state governments still have not been able to adopt more advanced e-government strategies which would allow for heightened levels of political participation.

Examining the adoption and use of e-government activities by local governments leads to similar findings. Moon (2002b) employed survey data collected by the International City Management Association (ICMA) in order to assess the level of adoption of many e-government practices. His evaluation of over 1400 survey responses also indicates that many local governments have adopted at least rudimentary attempts to post information and provide basic online services. However, few local governments have adopted more advanced forms of ICT that foster high levels of interactivity, communication and actual political participation. This study indicates that the amount of resources available to public organizations may be a primary factor in the adoption of more advanced technologies. Other studies of local governments have also indicated a reluctance to adopt more advanced online technologies. Hinnant and O'Looney's (2003) investigation of one online innovation, the personalization of online services, also hints that the adoption of more advanced technologies may be slowed by resource limitations, such as a lack of technical expertise and budgetary considerations. Furthermore, many advanced e-government technologies may also pose public organizations with new social challenges. For example, some ICT practices, such as personalization, require tracking the online activities and interests of individual citizens. Managing and using behavioral information may pose privacy issues that some public organizations have heretofore not been forced to address.

FUTURE CHALLENGES AND TRENDS

Overall, research indicates that the adoption of e-government strategies is still in its infancy. While some innovations, such as information dissemination and even service provision through integrated online WWW portals, has been widely accepted by federal, state, and local governments, more needs to be achieved in order to realize the purported benefits of e-government (Gant, Gant, & Johnson, 2002). Government faces many challenges in regards to realizing the potential of e-government activities. One challenge that is often discussed is one of ICT access. While many citizens have now adopted ICT such as the Internet and WWW, some demographics groups may still be without sufficient access to the very communication systems that e-government often relies upon (Thomas & Streib, 2003). Some research also indicates that even when citizens have sufficient access, some demographic groups may not possess the basic technical skills required to truly make use of e-government systems (Mossberger, Tolbert, & Stansbury, 2003). Private sector firms may not view such issues as overly important as long as their message and online commerce sites are available to their target market. However, government must necessarily be concerned with issues of access and use since it, as an institution, is accountable to the greater public. The access issue may eventually resolve itself as ICT is even more widely adopted within society but it does currently pose an important challenge to public organizations.

Other challenges to the adoption and use of ICT include resource limitations. For example, many governments face severe resource constraints that serve as obstacles to employing ICT. Since governments often face fluctuating budgets and sometimes inconsistent leadership, it is often difficult to support technical and administrative reforms that take years of nurturing and support in order to bring about lasting change. Governments should seek to develop funding, as well as professional resources, that will serve to maintain and advance e-government initiatives well past the initial adoption of such innovations. Similarly, many governments are being forced to deal with technical and administrative issues that arise from the employment of ICT itself. For example, employing ICT to deliver services to the public requires more attention to issues such as information assurance, ICT systems security and overall organizational accountability. If government wishes stakeholders, such as citizens and private sector firms, to be willing participants in e-government activities, it must provide assurances that such information-intensive activities are well planned out and relatively secure from potential abuse. In essence, governments must seek to promote trust on the part of its stakeholders (Welch & Hinnant, 2003).

CONCLUSION

The adoption of new forms of ICT leads to potential benefits for government organizations. As noted, e-government activities are potentially important technical, as well as administrative, reforms. They may eventually bring about greater levels of efficiency and effectiveness in regards to government operations and service production. Furthermore, they may even bring about heightened levels of government accountability and greater levels of participation in political processes. However, it is important to realize that while the Internet and WWW may serve as paths to better government, observers should not ignore prior research regarding the adoption of ICT by complex organizations. While ICT may be an important driver of change, its impact within the greater social structure of organizations or institutions is sometimes difficult to predict. For instance, previous research examining the adoption of new ICT within public organizations indicates that ICT reinforce preexisting social arrangements or shift power towards higher levels of management (Kraemer & King, 1986). Moreover, adoption of new ICT may shift power towards those in the organization who manage the technology itself (Bugler & Bretschneider, 1993; Kraemer & King, 1986; Kraemer, King, Dunkle, & Lane, 1989). Such findings should serve as caution to those who believe that the adoption of ICT as part of broader e-government initiatives yield only expected results. As government further adopts and even relies upon e-government strategies as a core part of its operations, it must develop institutional mechanisms that better anticipate and manage unexpected results.

REFERENCES

Accenture (2004). eGovernment leadership: High performance, maximum value. Retrieved June 20, 2004, from *http://www.accenture.com/xdoc/en/industries/government/gove_egov_value.pdf*

Bugler, D. & Bretschneider, S. (1993). Technology push or program pull: Interest in new information technologies within public organizations. In B. Bozeman (Ed.), *Public management: The state of the art* (pp. 275-294). San Francisco: Jossey-Bass.

Chadwick, A. (2003). Bringing e-democracy back in. *Social Science Computer Review*, *21*(4), 443-455.

Danziger, J., Dutton, W., Kling, R., & Kraemer, K. (1982). *Computers and Politics: High Technology in American Local Governments.* New York: Columbia University Press.

Fountain, J. (2001). *Building the virtual state: Information technology and institutional change*. Washington DC: Brookings Institution Press.

Gant, D.B, Gant, J.P. & Johnson, C. (2002). *State web portals: Delivering and financing e-service*. Arlington, VA: The PricewaterhouseCoopers Endowment for the Business of Government.

Gore, A. (1993). *From red tape to results: creating a government that works better and costs less: The report of the National Performance Review*. New York: Times Books.

Hinnant, C.C. & O'Looney, J. (2003). Examining pre-adoption interest in online innovations: An exploratory study of e-service personalization in the public sector. *IEEE Transactions on Engineering Management*, *50*(4), 436-447.

Kraemer, K.L. & King, J.L. (1986). Computing in public organizations. *Public Administration Review*, *46*, 488-496.

Kraemer, K.L., King, J.L., Dunkle, D. & Lane, J.P. (1989). *Managing information systems: Change and control in organizational computing*. San Francisco: Jossey-Bass.

La Porte, T.M., Demchak, C.C. & Friis, C. (2001). Webbing governance: Global trends across national-level public agencies. *Communications of the ACM*, *44*(1), 63-67.

MacEachren, A.M., Hardisty, F., Dai, X. & Pickle, L. (in press). Supporting visual analysis of federal geospatial statistics. *Communications of the ACM.*

Moon, M.J. (2002a). *State government e-procurement in the information age: Issues, practices, and trends*. Arlington, VA: The PricewaterhouseCoopers Endowment for The Business of Government.

Moon, M.J. (2002b). The evolution of e-government among municipalities: Rhetoric or reality. *Public Administration Review*, *62*(4), 424-433.

Mossberger, K., Tolbert, C.J. & Stansbury, M. (2003). *Virtual inequality: Beyond the digital divide*. Washington, D.C.: Georgetown University Press.

National Association of State Chief Information Officers. (2002). Enterprise architecture development program: Enterprise architecture development tool-kit (Vol. 2.0). Retrieved June 20, 2004, from *https://www.nascio.org/hotIssues/EA/AEADTool-Kitv2.pdf*

Osborne, D. & Gaebler, T.A. (1992). *Reinventing government: How the entrepreneurial spirit is transforming the public sector*. New York: Penguin Books.

Schneider, G.P. (2003). *Electronic commerce*. Boston: Thomson Publishing.

Thomas, J.C., & Streib, G. (2003). The new face of government: Citizen-initiated contacts in the era of e-government, *Journal of Public Administration Research and Theory*, *13*(1), 83-101.

Watson, R.T. & Mundy, B. (2001). A strategic perspective of electronic democracy. *Communications of the ACM*, *44*(1), 27-31.

Welch, E.W. & Hinnant, C.C. (2003). Internet use, transparency, and interactivity effects on trust in government. *Proceedings of the 36th Hawaii Conference on System Sciences (HICSS'03)*, IEEE Computer Society. Retrieved from *http://csdl.computer.org/comp/proceedings/hicss/2003/1874/05/187450144a.pdf*

West, D. M. (2003). E-government and the transformation of service delivery and citizen attitudes. *Public Administration Review*, *64*(1), 15-28.

Yen, J. (2004). Emerging technologies for homeland security. *Communications of the ACM*, *47*(3), 33-35.

KEY TERMS

Digital Government: The development, adoption, or use of ICT by government organizations and actors.

Electronic Commerce (E-Commerce): Conducting business and communication transactions over electronic computer networks. This may include the selling of goods and services as well as the transfer of funds from customers. This may also include inter-firm (business-to-business) transactions as well as intra-firm computer-based activities.

Electronic Democracy (E-Democracy): The use of ICT such as the Internet or WWW to foster democratic processes such as citizen participation.

Electronic Government (E-Government): The use of ICT such as the Internet and WWW by public organizations such as government agencies to provide information and public services to citizens and other government stakeholders.

Electronic Politics (E-Politics): The use of ICT such as the Internet and WWW by political actors to inform and facilitate public participation in the political process.

Electronic Procurement (E-Procurement): The use of ICT such as the Internet and WWW by government agencies to facilitate the purchasing of goods and services.

Internet: A large system of interconnected computer networks composed of backbone networks, mid-level networks, and local networks. This includes networks owned and managed by public, private, and nonprofit sector organizations.

World Wide Web (WWW or Web): A hypertext-based client-server system that is one component of the larger Internet. The hypertext-based format of the WWW allows users to navigate through the system by using graphical user interface software know as browsers.

Electronic/Digital Government Innovation, and Publishing Trends with IT

Yukiko Inoue
University of Guam, Guam

Suzanne T. Bell
University of Guam, Guam

INTRODUCTION

With scientific achievement and technological advancement, "American society is poised for a radical shift in the manner in which individuals learn and work, conduct their business, and are entertained and informed, arising from the innovation of new information technologies" (Schorr & Stolfo, 1998, p. 15). One of the significant innovations in information technology in this digital age has been the creation and ongoing development of the Internet: The Internet increases communication flexibility while reducing cost by permitting the exchange of large amounts of data instantaneously regardless of geographic distance (McNeal, Tolbert, Mossberger, & Dotterweich, 2003). Internet technology has changed rules about how information is managed, collected, and disseminated in commercial, government, and private domains. The Federal Government is increasing the reliability and efficiency of information services and resources provided to the public since the development of Internet technology as follows (Aldrich, Bertot, & McClure, 2000):

From the initial steps to establish the Internet in the late 1960s (originally ARPANET) to the establishment of the National Information Infrastructure (NII) and National Performance Review (NPR) initiatives in 1993, the federal government has adapted progressive computer and telecommunication technologies both operationally and in policy to harness computing power to improve government performance and enhance citizen access to government and other information services and resources. (p. 349)

BACKGROUND

For over one hundred years, the U.S. government agencies have collected and provided data and information (such as statutes and regulations, court decisions, votes by Congress, and the records of hearings) for citizens. For making government more convenient and accessible, during the 1980s, federal information dissemination began a transition to significant use of electronic formats (Notess, 1998). "Information technology, already an essential part of government operations, will continue to be vitally important to administration, decision making, and direct service delivery" (Dawes, Bloniarz, Kelly, & Fletcher, 1999, p. 6). As Coglianese (2004) confirms, information technology holds the potential for improving the process by which government makes regulatory decisions effecting vital aspects of society and the economy.

Digital government, however, is not just about using Web-based technology to deliver services, or converting paper-based processes into electronic ones. Digital government is a fundamental shift in government culture, allowing those interested in public policy and government to respond much more quickly ("What is digital," Department of Information Service [DIS], 2003). The DIS defines digital government in five ways: 1) replacing old bureaucratic service lines with accessible information and service, available 24-hours a day, directly from the desktop, using powerful new technologies such as digital signatures and electronic forms; 2) offering a "one-stop-shop" to many government services through the state's Internet Portal Access Washington; 3) making the process of accessing government services immediate, simple, seamless and intuitive; 4) reducing paperwork and its costs within government in order to move funds into direct delivery of services; and 5) improving service delivery to all segments of the U.S. population by reducing traditional counter lines with those accessing government services through the Internet. Specifically, digital government transforms government activities in two ways: 1) by improving service delivery and associated costs; and 2) by enhancing communication between citizens and government (McNeal, et al., 2003).

DIGITAL GOVERNMENT INNOVATIONS AND CRITICAL ISSUES

Digital Government Initiatives

The government is a dynamic mixture of structures and functions; the government's following six digital initiatives are complex efforts intended to use new technologies to support a transformation in the effectiveness and efficiency of the government (Pardo, 2000):

1. **Citizen access to government:** includes establishing mechanisms that deliver information based on the customer's perspective rather than a functional perspective.
2. **Facilitating compliance with rule**s: provides electronic access to services that facilitate compliance with a set of rules or regulations (e.g., driver's license renewal, hunting and fishing licenses, and business permits).
3. **Citizen access to personal benefits:** electronic access to citizens' personal benefits, for example, online applications for public assistance and workers' compensation.
4. **Procurement including bidding, purchasing, and payment:** procurement applications allow government agencies to reap the benefits being realized in the private sector through electronic commerce applications.
5. **Government-to-government information and service integration:** integrating service delivery programs across government agencies and between levels of government require electronic information sharing and integration.
6. **Citizen participation:** online democracy includes access to elected officials, discussion forums, "town meeting," voter registration, and ultimately online voting (and this particular one must face issues of the digital divide and security). (pp. 3-4)

Three Aspects of Digital Government

Digital government enables constituents to access information and services from home, which reduces traffic flow, and improves the environment. There are at least three distinct aspects of digital government: 1) information dissemination, 2) interactive service delivery, and 3) online monetary transactions. Atkinson and Ulevich (2000) suggest what government policy makers should be doing in each of these areas (see Table 1). The authors further point out: 1) most applications still focus on information dissemination from the government to user but several agencies are beginning to combine resources to better carry out tasks, record transactions, and benefit the consumer; 2) many agencies allow individuals to obtain

Table 1. A summary of what the government should be doing with information dissemination

• Developing enterprise-wide information architecture. • Implement a standardized information tagging system. • Create an entryway/portal to government services. • Expand the amount of information accessible on searchable databases. • Use "information on request" to provide people with government information. • Develop "expert systems" to access information. • Make the Web the first place to put information, not the last. • Measure customer satisfaction. (pp. 17-19)
Interactive service delivery
• Expand and standardize the number of applications for online forms. • Whenever possible, use Web-based technology. • Online forms should use shared information about the submitter. • Integrate form (putting form online is one thing, streamlining and consolidating information is another). • Focus on intergovernmental solutions. (pp. 20-21)
Online monetary transactions
• Use EFT in all monetary transactions. • Implement the use of electronic checks. • Develop government-wide electronic procurement systems. • Make it easier for citizens and business to directly file their taxes online. • Attach digital signature functions. (pp. 22-23)

forms online, print them, and then mail the paper copies, where the information is then either entered by hand or scanned into a computer, while a few agencies are beginning to allow citizens and businesses to file forms online; and 3) monetary functions increasingly rely on electronic funds transactions (EFT), and virtually all of these activities could be done electronically, replacing paper.

Critical Issues of Internet Adoption

Through the Web revolution, more effective, convenient, and flexible digital government services are happening, yet *security* and *privacy* are the two critical issues in digital government applications. Building secure digital government systems requires a careful balancing between providing convenient access and appropriately monitoring permission, and technology-based solutions are still at their infancy and the existing alternatives consist essentially of enforcing privacy by law or self-regulation of operational practices ("Security and privacy," CISC, 2004).

Concerning barriers to Internet use (e.g., cost, systems incompatibility, legal issues such as intellectual property and security, human resistance, and privacy), Lan and Falcone (1997) identify four factors in terms of positive and negative impacts of policy decisions on Internet adoption: 1) the technical characteristics of information technologies; 2) user preferences for information resources; 3) the institutional arrangements that define, govern, and constrain provider/user relationships; and 4) psychological factors affecting information processing. Other legal issues include: electronic signatures and records must still meet formal legal requirements for paper documents; and agencies must ensure that the electronic signature or records and all relevant information will be available as needed (Champion, 2003).

Web Site-Specific Policy

Halchin (2002) recommends developing a government-wide disclosure policy that applies specifically to agency Web sites, chiefly because such guidelines could help agency officials determine what types of information are potentially compromising, and could include procedures for reviewing and reposting documents and information that were removed or altered. Based on the fact that currently 54 federal agencies have Chief Information Officers but the Federal Government as a whole does not, it is extremely important to establish the position of a Chief Information Officer (CIO) for the Federal Government so that a federal CIO would report directly to the President and direct the process of developing a concerted digital government conversion plan (Atkinson & Ulevich, 2000).

Online electronic government access contains a tremendous amount of information on a plethora of subjects. The material is available virtually in one place, a personal computer, yet U.S. federal agencies have altered their Web sites particularly since the events of terrorist attacks on American soil, September 11, 2001(Halchin, 2002); for example:

The Department of Energy (DOE) removed from its Web site detailed maps and descriptions of all ten nuclear facilities with weapons-grade plutonium and highly enriched uranium; The National Aeronautics and Space Administration's (NASA) John Glenn Research Center has limited public access to many of its Web sites; The Nuclear Regulatory Commission (NRC) removed its Web site, then reposted selected content; and The Agency for Toxic Substances and Disease Registry (ATSDR) removed a report on chemical site security from its Web site. (pp. 246-247)

As Internet information is selectively provided to citizens, the Internet's open nature remains an ideal arena for dissemination of misinformation. The issue is also how to deal with false information on the Web and how to decide whether a source is reliable. Ultimately each individual is responsible for his or her use of technology and for decisions taken based on information gathered from the Web (Thompson, 2004).

FUTURE TRENDS

The Government Printing Office

The U.S. Government Printing Office (GPO) keeps America informed. The GPO began operations accordance with Congressional Joint Resolution 25 of June 23, 1860 ("About the U.S.," GPO, 2003) with the mission to inform the Nation by producing, procuring and distributing publications from Congress, executive departments and establishments of the federal government. The GPO is part of the legislative branch of the federal government and operates under the authority of the public printing and documents chapters of Title 44 of the U.S. Code.

The Federal Depository Library Program

In order to organize, manage, and disseminate the federal government's information, the Depository Library Act of 1962 expanded the Federal Depository Library Program (FDLP), which is administered by GPO. For more than 100 years, FDLP has been an important channel through which the U.S. government has made information available to the public (Lopresti & Gorin, 2002). Withstanding this historic goal of information dissemination, a compelling letter dated August 25, 2000, was sent to all 1330 of the Nation's depository libraries from the Director of the FDLP and then Superintendent of Documents, Francis J. Buckley. ("SuDocs Letter," GPO, 2000). This document announced a most profound change in how the American public would access government information in the future via electronic format. Buckley's announcement stated that as a result of the proposed Congressional appropriation for the FDLP, the shift to a primarily electronic program would accelerate. Consequently, both diminished funding for federal publishing and developing technologies greatly accelerated the migration and rapid growth of online electronic content delivery ("Library programs," GPO, 2000). The Superintendent of Document's announcement rippled through the FDLP community, gathering reactions such as "the government's target of making all its services available electronically by 2005 is a milestone in the application of electronic communications" (Ryan, 2002, p. 48). As new materials are produced in electronic formats with no printed equivalents, apprehension regarding weaknesses in the current state of online government information has become evident; some researchers, libraries, and governments are concerned about the accessibility, authentication, cataloguing, and long-term preservation of electronic records (Carvin, 2004).

While printing will not disappear, the University of Arizona, based on their pilot project with GPO, has decided to become the first all-electronic selective depository library (James, 2003). James continues that it should be determined what services GPO can offer exclusively to depository libraries that will attract and keep libraries in the program when citizens can obtain *everything* free on the Internet. The trend toward government information online definitely continues to be a major one. However, the widespread utilization of electronic information dissemination, fewer staff and financial resources, and other barriers all present new and difficult challenges for the regional libraries (Barkley, 2002). To access the electronic information, the librarians must remain current with computer and Internet skills, in addition to knowledge of the legal system and the legislative process (Yang, 2002).

CONCLUSION

Predictions for the future include that citizens can log onto an Internet site, easily find the government services they are looking for and proceed to conduct an online transaction (Atkinson & Ulevich, 2000). In addition to Web portals to provide information and services, electronic/ digital government innovations offer management and policy-making tools. Government has been at the forefront of information technology research and, as Coglianese (2004) states, "efforts to help improve the ability of agencies to manage the rulemaking process as well as increase public participation in process deserve careful consideration" (p.85). Digital government and regulatory policy making capabilities are only beginning to emerge but in the future it can be expected to fall within the main four areas (information technology, agency management of rule making, public involvement in the rulemaking process, and regulatory compliance) and these areas of the research will be enhanced through coordinated research efforts that involve perspectives from both the social and information science (Coglianese, 2004).

Digital government will cut the cost and improve the quality of government, but real progress in a timely manner depends upon digital government rapidly becoming a priority of Congress and the Administration, both in terms of funding and leadership (Atkinson & Ulevich, 2000). The WWW technologies (specifically, intranets using Web browsers such as Netscape, or Microsoft Internet Explorer) are changing rapidly, creating enormous opportunities for central and local governments; however, there are diverse problems such as new methodologies of authentication, record keeping, security, and self-service access (Safai-Amini, 2000). Finally, in today's electronic information age, "libraries must act as expert service providers rather than warehousing physical collections" (James, 2003, p. 19).

REFERENCES

About the U.S. government printing office. (2003). Washington, DC: The U.S. Government Printing Office. Retrieved [2004] from *http://www.gpoaccess.gov/about/index.html*

Aldrich, D., Bertot, J. C., & McClure, C. R. (2002). E-Government: initiatives, developments, and issues. *Government Information Quarterly, 19*(4), 349-355.

Atkinson, R. D., & Ulevich, J. (2000). *Digital government: The next step to reengineering the federal government (progressive policy institute)*. Washington, DC: Progressive Foundation.

Barkley, D. (2002). Collection development in a regional depository library. *A Quarterly Journal of Government Information, 30*(2), 12-13.

Carvin, P. (Ed.). (2004). *The United States government Internet manual*. Lanham, MD: Bernan Press.

Champion, M. T. (October 2003). Electronic signatures: What rights and duties do North Carolina agencies possess under the current statutory scheme. *Digital govern-*

ment innovation bulletin, The University of North Carolina at Chapel Hill.

Coglianese, C. (2004). Information technology and regulatory policy. *Social Science Computer Review, 22*(1), 85-91.

Dawes, S. S., Bloniarz, P. A., Kelly, K. L., & Fletcher, P. D. (1999). *Some assembly required: Building a digital government for the 21st century*. Center for Technology in Government, University at Albany, SUNY.

Halchin, L. E. (2002). Electronic government in the age of terrorism. *Government Information Quarterly, 19*(3), 243-254.

James, B. R. (2003). New Directions for the FDLP. *A Quarterly Journal of Government Information, 31*(3/4), 17-20.

Lan, Z., & Falcone, S. (1997). Factors influencing Internet use—a policy model for electronic government information provision. *Journal of Government Information, 24*(4), 251-257.

Library programs service FY2000 annual report (2000). Washington, DC: U.S. Government Printing Office. Retrieved [2004] from *http://www.access.gpo.gov/su_docs/fdlp/pubs/annrprt/001psar.html*

Lopresti, R., & Gorin, M. (2002). The availability of US government depository publications on the World Wide Web. *Journal of Government Information, 29*(1), 17-29.

McNeal, R.S., Tolbert, C. J., Mossberger, K., & Dotterweich, L. J. (2003). *Innovating in digital government in the American States, 84*(1), 52-70.

Notess, G. R. (1998). *Government information on the Internet*. Lanham, MD: Bernan Press.

Pardo, T. (2000). Realizing the promise of digital government: It's more than building a web site. Retrieved [2004] from *http://www.netcaucus.org/books/ego2001/pdf/realizin.pdf*

Ryan, P. (2002). 'e' is for education. *Update, 1*(3), 48-49

Safai-Amini, M. (2000). Information technologies: Challenges and opportunities for local governments. *Journal of Government Information, 27*(4), 471-479.

Schorr, H., & Stolfo, S. J. (1998). A digital government for the 21st century. *Communications of the ACM, 41*(11), 15-19.

Security and privacy in digital government application. (2002). Harrisonburg, VA: Commonwealth Information Security Center (CISC), James Madison University. Retrieved [2004] from *http://www.cisc.jmu.edu/research/bouguettayal.html*

SuDocs letter to director: Changes in FDLP. (2000, September). *Administrative Notes, 21*(13). Washington, DC: The Government Printing Office.

Thompson, P. Cognitive hacking and digital government: Digital identity. Retrieved [2004] from *http://ists.dartmouth.edu/IRIA/project/semantic/coghacker.pdf*

What is digital government? (2003). Department of Information Service (DIS). Retrieved [2004] from *http://www.wa.gov/dis/role/whatisdiggov.htm*

Yang, Z. Y. (2002). An assessment of education and training needs for government documents librarians in the United States. *Journal of Government Information, 28*(4), 425-439.

KEY TERMS

ARPANET: Advanced research projects agency network that the Internet has roots in, developed by the Department of Defense.

Electronic Signature: The Electronic Signature Act that President Clinton signed into law on June 30, 2001 facilitates electronic contractual arrangements involving cryptography principles, resulting in a digital approval process, allowing for a handwritten signature or digital certification.

Federal Depository Library Program (FDLP): Federal program of disseminating government program information products in all media to participating libraries.

Internet: World's largest network, a worldwide collection of networks that link together millions of businesses, governments, educational institutions, and individuals using modems, telephone lines, and other communications devices and media. Also called the Net.

Intranet: A network based Internet belonging to an organization or groups where information is accessed with authorization.

National Information Infrastructure (NII): A futuristic network of high-speed data communications links that eventually will connect virtually every facet of our society.

National Performance Review (NPR): A management reform initiatives established by the national administration to identify ways to make the government work better and cost less.

Public Printer: The head of the Government Printing Office (GPO).

Web Portal (Portal): Provides Internet access to a broad array of Web pages and linked services.

World Wide Web (WWW): Worldwide collection of electronic documents on the Internet that have built-in hyperlinks to other related documents. Also called the Web.

E-Mail and Communication

Dianne Willis
Leeds Metropolitan University, UK

INTRODUCTION

Electronic mail is the most frequently used application of the Internet. IDC research in 2002 suggested that the daily output of e-mails will reach 35 billion by 2005.

Sands (2003) stated than in the past 5 years, e-mail has matured until almost everyone in the developed world is expected to be able to be contacted via e-mail. He also stated that it is now as taken for granted as the telephone within a few short years of its introduction. E-mail has become a vital aspect of the communication process.

E-mail is being increasingly adopted as a major communication tool in U.K. higher education establishments (colleges of higher and further education and universities).

As the use of e-mail grows, the effect on communication patterns needs to be established. This article looks at communication theories, identifying when e-mail is an appropriate medium, and current patterns of e-mail usage within a higher education institution in the United Kingdom (the author's own).

COMMUNICATION AND E-MAIL

Communication takes place on an ongoing basis between individuals and groups using a variety of media. However, not all communication is clear to all parties—misunderstandings arise. To prevent these problems, people develop communication strategies. Te'eni (2001) suggested that communication complexity is a major aspect of strategy selection, because it reflects difficulties in communication, while the selected strategy is the means by which complexity is reduced. Te'eni introduced the terms "cognitive," "dynamic," and "affective" as being the components of communication complexity. Cognitive complexity is defined as a function of the intensity of information exchanged and the multiplicity of views held. Dynamic complexity refers to how far the communication process depends on time constraints and unclear processes that may increase the likelihood of misunderstanding. Affective complexity refers to how sensitive communication is to attitudes or changes in disposition toward the communication partner or subject matter.

He suggested a model depicting the relationship between communication goals and strategies (Figure 1).

Te'eni (2001) also suggested that certain medium attributes are more effective for certain strategies and produce a model of the relationship (see Figure 2). Te'eni takes on board the variety of communication technologies available today, including e-mail, which leads to a greater choice in terms of communication medium.

E-mail, like a letter, has low channel capacity, but like face-to-face communication, it enables instant response,

Figure 1. Adapted from: The complexity of communication goals affects strategy selection, Te'eni (2001), p. 269.

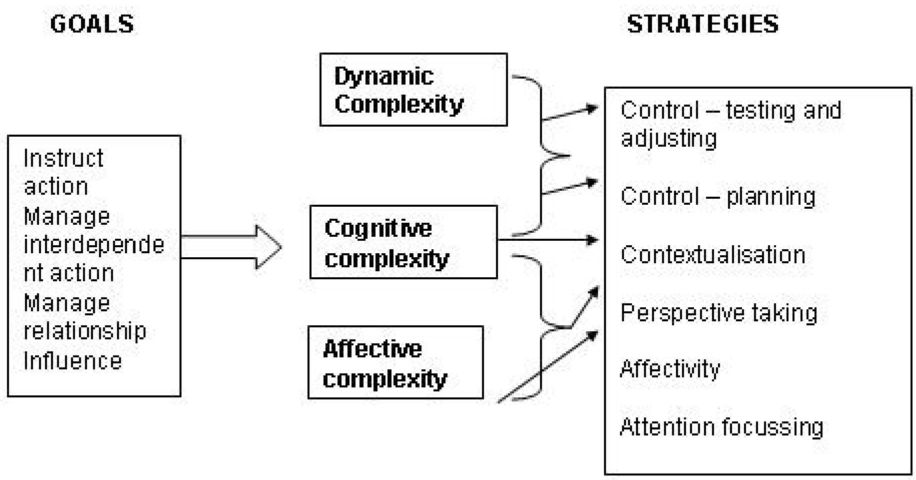
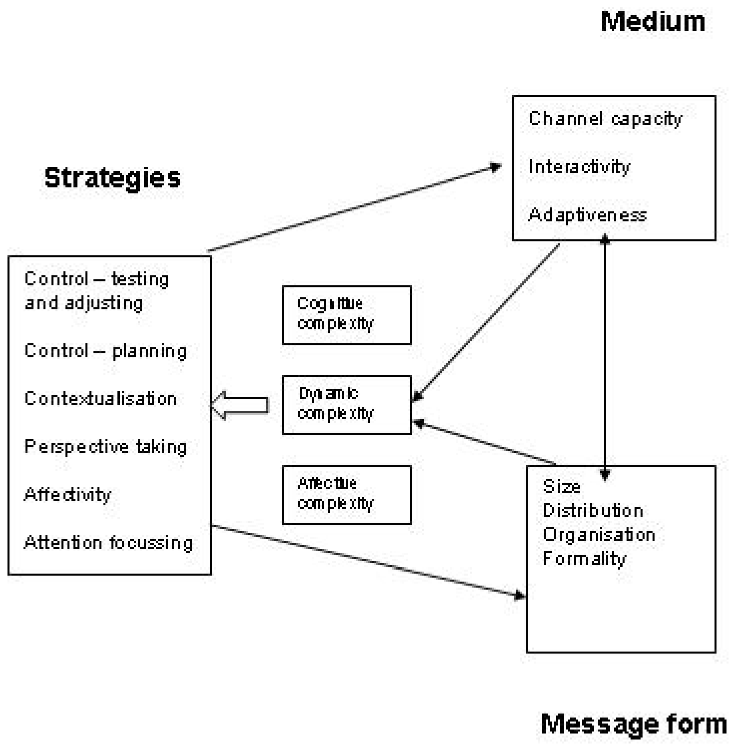


Figure 2. Adapted from: Strategies affect and are affected by medium and message, Te'eni (2001), p. 271.



though without the nonverbal feedback used for control through testing and adjusting. Markus (1994) and Romm and Pliskin (1998) suggested that e-mail is uncontrolled and disturbs relationships, even when it does not reduce the level of social interaction. Sproull and Kiesler (1992) posited that e-mail increases the ability to distribute messages en masse, and at the same time, reduces the costs involved, making it a very attractive proposition for organisations, even if the desired richness of the communication is not met.

There is also a general belief that synchronous media are richer than asynchronous media, because the lack of delay can assist in the interpretation of the message. E-mail is ranked by some writers as somewhere between the telephone and nonelectronic written communications (Trevino et al., 1987). This would lead to the selection of other media in preference to e-mail. Consideration of the modern working environment would suggest that this is not the case, and some discussion of possible underlying reasons would be useful.

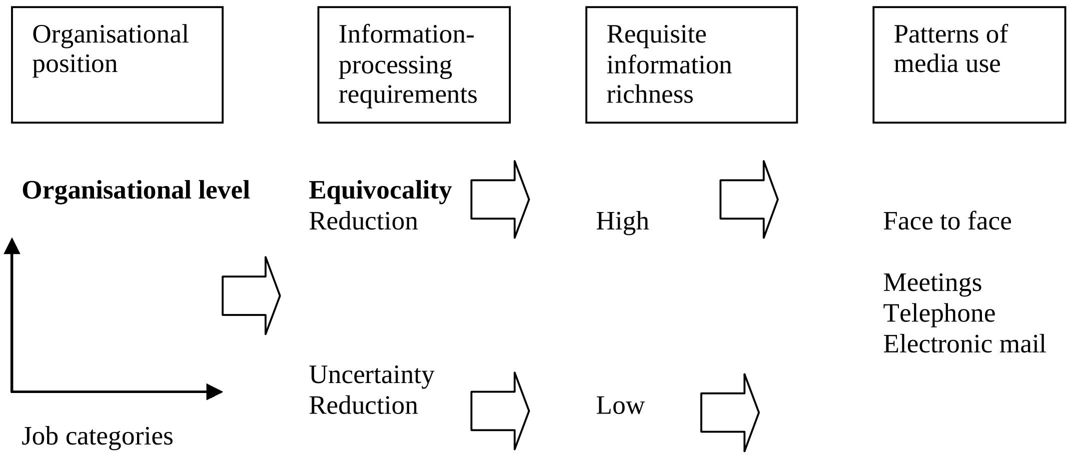
E-mail does have some specialist characteristics that might lead to it being chosen where theory would dictate the choice of another medium. Multiple addressability, the capacity to keep a permanent record, and the ability to field and search for information contained within e-mails adds much to the concept of richness. Markus (1994) and Panteli (2002) both suggested that this may be sufficient to make it a richer medium than the telephone. A communicator will generally have a choice in how to communicate. What may be more useful to study is the complexity of the communication to explain the choices of strategies, messages, and media. Communication complexity arises from trying to ensure the message is received when conditions are not ideal, i.e., when there is some form of interference in transmission of the message. Wood (1986) identified the intensity of communication as an important issue in the communication process, whereas Te'eni (2001) concentrated on interdependency.

Rice and Shook (1990) produced a model of the communication process, illustrated in Figure 3, that identifies appropriate media for communication in terms of media richness:

E-mail ranks toward the low end of the richness scale, supposedly limiting its communication possibilities.

E-mail can be viewed primarily as a sociotechnical system. A working definition of a sociotechnical system states that changes in one part of the system, be it technical or social, will affect the other parts and, thus, the system as a whole. A sociotechnical system is concerned with the interactions between the psychological and social factors and the needs and demands of the human part of the organization (its structural and technological requirements) (Mullins, 2002). E-mail satisfies the criteria for a sociotechnical system in that it consists of technology and the software needed to run the system as well as operate within the social norms adopted by the

Figure 3. Adapted from: Visual model of theoretical argument motivating the meta-analysis and individual level analyses (Rice and Shook, 1990, p. 196).



organisation choosing to use e-mail as a communication tool. These social norms will help to determine the way in which e-mail is used and the extent to which it will become a dominant communication form. The relative importance of each part of the system, the social and the technical, will vary between different institutions. When any new technology is introduced, it is important that the user population have a positive attitude toward the new technology. For it to be successful, any new technology needs to be understood and valued by the users.

Increasingly, e-mail has become a common mode of communication for many people, though exceptions must be made for those who do not have easy and regular access to the technology required. Anderson et al. (2000) considered in-depth universal access to e-mail and the personal and social benefits this can bestow. They argued that e-mail is a mechanism that provides global communication possibilities. However, the lack of access by certain sections of a society leads to the formation of an information elite, often referred to as the "information rich." Those without access are known as the "information poor." The extent to which e-mail can impact global communications must be considered in the light of equal access for all. Anderson et al. (2000) posited that there are no fundamental technical barriers to providing universal access to e-mail, so the responsibility for pushing forward equal access lies with society rather than with the technology. Cost implications are a major consideration here. Equal access to e-mail and, hence, to the communication space, would require a program of investment in access points in a variety of places, including schools, libraries, and even supermarkets. The maintenance of such access points would also need to be carefully considered. While these issues are under consideration in the United Kingdom, from the point of view of access to government offices, little is being considered from the viewpoint of universal access.

From the preceding discussion, it can be argued that e-mail is manifestly a less "rich" medium than is face-to-face communication, and careful consideration must be given to whether there are conditions in which it is totally inappropriate as a medium. Consideration must be given to what would happen to the social side of the system if e-mail is used inappropriately. Would this change patterns of usage or bring about new norms of social behaviour? There have been several examples recently cited in the UK press, where staff have been dismissed for inappropriate use of e-mail. Recent UK legislation, The Regulation of Investigatory Powers Act 2000, gives employers the power to intercept employee e-mail. This legislation is commonly known as the "Snooping Bill" and is hotly debated in terms of the underlying concepts of privacy and freedom of speech.

E-mail lends itself to informal communication, ensuring interaction and conveying important information quickly. Remembering that the main advantage of e-mail is speed, messages should be kept short and to the point to facilitate a quick response. Sherwood (2000) believed that e-mail is a fundamentally different form of communication than paper-based communication, with the speed of the process making it lean toward a more conversational tone. As recipients of e-mail can quickly respond, ask questions, or seek clarification, there is a tendency for e-mails to be not so rigorously or painstakingly constructed as letters or memos. E-mail also does not convey emotions nearly as well as face-to-face or telephone conversations, as it lacks the cues available in these forms of interaction. E-mail can be seen merely as a facilitating tool for the communication process.

Kraut and Attewell (1997) found that the use of e-mail correlated with the use of other media in that people classified as "heavy communicators" are likely to use all forms of media more often. One of the advantages of e-mail is in terms of its relatively low cost when compared with alternative communication forms, so it is often regarded as a cost-saving method. It may be the case that rather than substituting for other communication forms, e-mail may actually stimulate new relationships within the social system that will eventually lead to an increase in the total amount of communication taking place, which may then not constitute a cost savings. People generally have a preference in communication method. From an organisational standpoint, it is helpful when these preferences are for the most cost effective form of communication.

Problems with e-mail identified by Woolston and Lipschutz (1998) include the following:

- Nonverbal cues are limited, leading to possible misinterpretation of the feelings associated with e-mail
- Discussion of confidential information is not always secure
- E-mails are context deficient, there is no simple way of determining that the sender of an e-mail is who they claim to be
- Overload

There is increasing concern about overload in the new information world, given the numbers of e-mail messages in circulation on a daily basis. It is possible that the ease of use of the technology and the ready availability of multiple mailing lists adds to this feeling of overload.

One of the major advantages claimed for e-mail is that it provides a record of the communication, which was not always available in face-to-face or telephone conversations. This record seems to be of increasing importance in modern academic life. One respondent in an earlier study (Coakes & Willis, 2000) cited this record as a useful defence against possible litigation.

THE INSTITUTIONAL EXPERIENCE

The author has been studying the effect of e-mail in U.K. higher education institutions over the past 6 years. The present study of an individual department in the author's institution was prompted by comments such as "the institution is totally e-mail driven." This was backed up by staff saying that if their e-mail was down, they were unable to find out what was happening. There was also a concern that e-mail was used increasingly in instances where face-to-face communication was required.

Large quantities of information are passed around the institution using e-mail with little to no checking as to whether or not the message has been read. There is an implicit assumption that if something has been sent round in an e-mail, it has been communicated adequately. This work challenges that assumption and looks at the reality as seen from a staff viewpoint in a single department.

Short and long individual interviews were conducted over a 2 year period with 48 staff (Willis, 2003) to ascertain the following:

1. Whether they felt that e-mail was being used appropriately, i.e., if the right medium was being chosen for the message that needed to be communicated
2. Whether staff primarily instigated or responded to e-mail communication
3. How comfortable staff were with that mode of communication

Early results indicate that e-mail is often used when face-to-face communication is needed for a variety of reasons, including time and the need for a record of the communication. All 48 staff interviewed agreed that e-mail was the dominant form of communication within the department, though it was not the preferred method. In response to the question on preferred method of communication, 80% chose face-to-face, mainly because of the richness of the medium. Interestingly, although it was not the preferred method, 75% said they tended to use e-mail, as it was much faster, saving time, and was more efficient than trying to arrange a meeting.

Unsurprisingly, given the culture in the institution, e-mail overload is becoming an issue, as some staff are receiving in excess of 200 e-mails a day. In terms of this volume of messages, it must be questioned whether e-mail is actually a suitable medium to get any message through, despite the advantages claimed for e-mail.

A second issue of concern was that of the time taken to answer the large volume of e-mails. Some staff interviewed reported that they were much more desk bound since the advent of e-mail and actually had to spend so much time answering messages that they were unable to walk along the corridor and seek out colleagues for discussion.

A final issue identified was that of the increasing volume of SPAM. One interviewee receives an average of 75 SPAM e-mails per day. This has led in the past to the accidental erasure of legitimate messages by the setting up filters that do not work as effectively as was hoped.

Research by the law firm Masons (www.out-law.com) showed that in May 2003, over half of all e-mails sent were

Table 1. Research comments

"I regard e-mail as part of my job—I must answer it on a daily basis."	"E-mail enables me to work more efficiently."
"E-mail ties me to the office, and I am unable to find time for face-to-face communication."	"I get much more work done as I can deal with my e-mail as and when I like."
"E-mail is a way of passing on large amounts of work without consideration for the individual."	"I love e-mail, but I'm very nosy—I have to look to see what I've been sent even when I should be working on something else."
"SPAM is becoming a real problem, I get 75–100 messages a day."	"I can contact colleagues abroad much more easily for collaboration on projects."

classified as SPAM, and this figure is expected to continue rising. The reality of this is that it will become increasingly difficult to sort legitimate messages (HAM) from the rest. Obviously, this will have an effect on the time problem identified earlier.

Table 1 illustrates some comments taken from the research.

FUTURE TRENDS

It is likely that in the near future e-mail will take on a greater role in terms of a mechanism for student support. This would obviously increase the load on individual lecturers who are already reaching the limit of their ability to deal with ever-increasing volumes of e-mail. While e-mail can provide a useful method of communication between students and staff, the overload issue and the efficiency of the method of communication are called into question.

The use of a Virtual Learning Environment (VLE), WebCT, is being explored as an alternative to e-mail. Messages can be posted to whole groups and discussions can be held with the group of students attached to the module to address problems that previously would have been sorted on an individual basis.

Bulletin boards are another mechanism being examined to potentially reduce e-mail overload. Individuals will access information on an interest basis, which will prevent the common problems of sending information to the whole institution when it is only relevant to some members.

In terms of e-mail generally, it appears that the increase in volume is set to continue, for the near future at least, and it is to be hoped that this will not mean that the medium will become virtually unusable, though this issue is raised frequently in the U.K. press.

CONCLUSION

The theory states that e-mail is a relatively lean medium, though it has better interactive opportunities than letters and enables asynchronous working. The ability to communicate without resorting to "telephone tag" is very helpful to people who are busy. In the author's work environment, the inconvenience of e-mail tends to be overlooked, as the culture of the organisation drives its use.

Given that e-mail is a major communication method within the organisation, care must be taken that the volume of e-mail received by each member of staff does not reach the point where it becomes impossible for them to deal with it.

Guidelines for the effective use of e-mail include the following:

- Take care to avoid overload.
- Speed is the biggest advantage for e-mail; make sure the culture is such that this is adhered to in communication.
- Look for ways of using e-mail to facilitate student support.
- Use Web sites or the intranet to provide large quantities of information.
- Target the message carefully to increase efficiency.
- Think before replying.
- Use high priority sparingly.
- Delete unwanted material to facilitate effective use and save server space.

To sum up, it can be seen that e-mail has many of the advantages claimed by the theory. However, care must be taken to ensure that it does not become the sole communication medium. Appropriateness must be the watchword, or the quantity of material communicated will increase but the quality will not.

REFERENCES

Anderson, R. H., & Bikson, T. K., Law, S.A. & Mitchell, B.M. (Eds.), *Universal access to email, Feasibility and social implications*. Santa Monica, CA: RAND.

Coakes, E., & Willis, D. (2000). Computer mediated communication in universities and further education establishments—A comparison of use and utility. IRMA 2000, May 2000, Anchorage, Alaska.

Kraut, R. E., & Attewell, P. (1997). Media use in a global corporation: Electronic mail and organisational knowledge. In S. Keisler (Ed.), *Culture of the Internet* (pp. 321-341). Mahweh, NJ: Lawrence Erlbaum Associates.

Markus, L. (1994). Electronic mail as the medium of managerial choice. *Organisation Science, 5*(4), 321-340.

Mullins, L. J. (2002). *Management and organisational behaviour* (6th ed.). London: Pitman.

Panteli, N. (2002). Richness, power cues and email text. *Information and Management, 40*(2), 75-86.

Rice, R. E., & Shook, D. E. (1990). Relationships of job categories and organisational levels to use of communication channels, including email. *Journal of Management Studies, 27*(2), 195-229.

Romm, C., & Pliskin, N. (1998). Email as a coalition building information technology. *ACM Transactions on Information Systems, 16*(1), 82-100.

Sands, M. (2003). Integrating the Web and email into a push–pull strategy. *Qualitative Market Research: An International Journal, 6*(1), 27-37.

Schwartz, D. (1999). When email meets organisational memories: Addressing threats to communication in a learning organization. *International Journal of Human-Computer Studies, 51*, 599-614.

Sherwood, K. D. (2000). A beginners guide to effective email. Retrieved September 28, 2000, from *www.webfoot.com/advicce/email.top.html*

Sproull, L., & Kiesler, S. (1986). Reducing social context cues: Electronic mail in organisational communication. *Management Science, 32*(11), 125-130.

Te'eni, D. (2001). A cognitive-affective model of organisational communication for designing IT. *MIS Quarterly Review, 25*(2), 251-312.

Trevino, L. K., Lengel, R. H., & Daft, R. L. (1987). Media symbolism, media richness and choice in organisations. *Communication Research, 14*(5), 553-574.

Willis, D. (2003). Impact of email on working practices in an UK Higher Education environment: A case study. Unpublished research paper (work in progress).

Wood, R. E. (1986). Task complexity. *Organisational Behaviour and Human Decision Processes, 37*(1), 60-82.

Woolston, D., & Lipschutz, W. (1998). Using email effectively in academic advising. Pre-conference workshop at the *22nd Annual National Conference, National Academic Advising Association*, October.

www.out-law.com. IT and legal advice and support from international law firm. Masons. Accessed November 11, 2003.

KEY TERMS

Affective Complexity: How much communication is sensitive to attitudes or changes in disposition toward the communication partner or subject matter.

Channel Capacity: The potential to transmit a high variety of cues and languages.

Cognitive Complexity: A function of the intensity of information exchanged and the multiplicity of views held.

Communication Strategies: The means by which communication goals can be fulfilled.

Computer-Mediated Communication: The dialogue that unfolds between users of electronic media.

Dynamic Complexity: How much the communication process depends on time constraints, unclear or deficient feedback, and changes during the process.

E-Mail: Messages sent and received electronically via telecommunication links.

HAM: Legitimate e-mail communications.

Sociotechnical System: The relationship between technical and social variables in a system. Changes in one part, technical or social, will affect the other parts and, thus, the whole system.

SPAM: Unsolicited marketing messages sent by e-mail or posted on newsgroups.

E-Mail as a Strategic Tool in Organizations

Denise Johnson McManus
Wake Forest University, USA

Houston Hume Carr
Auburn University, USA

INTRODUCTION

During the past decade, strategic information technology (IT) applications, such as electronic mail (e-mail), have been used to gain competitive advantage over rivals or prevent rivals from gaining a competitive advantage. The use of e-mail has been considered a component of overall business strategy, and its application depends more on understanding unique business opportunities than on competitive benefits from technological features that are easily duplicated. Since the first message was sent in 1971 by Ray Tomlinson, e-mail has emerged as an important technological requirement in business. E-mail is as common as the U.S. postal mail in the twentieth century. Businesses, schools, and universities are insisting that employees and students have an e-mail account to check updates or download information concerning business tasks or class activities. Given the widespread acceptance and use of this technology, the strategic value of e-mail within an organization should be apparent.

Companies have discovered that they must aggressively pursue new information technologies to exploit their own core competencies. Contemporary organizations have found ways to not only build more functionality onto their core mail systems, but also combine their core capabilities with strategic and creative ways of using e-mail. Clearly the definition and parameters around e-mail are changing, and the use of electronic communication is revolutionizing how firms do business.

BACKGROUND

E-mail, as a specific information technology, not only has proven to be one of the primary technologies by which an organization can begin to achieve radical and beneficial organizational change, but also has become an integral part of the corporate culture in many organizations (Carroll, 1993). This technology, which is considered the electronic transmission of messages, documents, data, and images, has enabled people to send electronic correspondence to others with e-mail addresses anywhere in the world, and is fast, inexpensive, and easy to use, making it a convenient and flexible way to improve corporate communications. As a competitive information technology, e-mail offers many advantages to the dynamic environment of today's businesses (Reark, 1998).

Literature suggested that e-mail may be used to improve relationships with customers, to link organizations with their strategic business partners, and to enhance communication between managers responsible for business redesign (Kirschner, 1995). This technology may be used to facilitate better coordination of a firm's communication between departments and divisions—*intraorganizational* uses of e-mail. Although internal communication was considered to be the core use of e-mail, firms are increasingly using it for outside communications—*interorganizational* uses of e-mail which include improving relationships with customers. Lastly, researchers have suggested that the primary benefit of e-mail have been its efficiency (Palme, 1995). For example, e-mail decreased the use of paper, eliminated time delays, and allowed the sender to transmit messages to a list of specific individuals; we refer to these as the organizational benefits of e-mail.

In 2000, more than seven trillion e-mail messages traveled the wires in the U.S. alone, up from a mere four trillion in 1999, and thus supporting the argument that e-mail correspondence is the fastest growing communication medium in the world. The most recent findings report that the average businessperson sends and receives approximately 90 e-mail messages daily (Casperson, 2002). As international competitiveness concerns have forced businesses to right-size, e-mail has surfaced as one of the facilitators that permit savings in middle-management expenditures. The increasing ubiquity of e-mail lessens telephone tag and makes direct communication more efficient, even when there are no other sophisticated systems supports (King & Teo, 1994).

Furthermore, the utilization of e-mail to communicate with customers and suppliers indicates the need for interorganizational linkages. Customers prefer to communicate with key people in a firm through the use of e-mail, because it produces a record of their communication as opposed to multiple phone calls. This perceived personalized service allows direct access to a person who could answer their questions and solve their problems. Providing this dependable link to the customer gave the company an edge over its biggest competitors.

STRATEGIC USES OF E-MAIL

Technology reaches its greatest potential when it supports the strategic operations of the organization. It is not necessary for a technology to be new for it to be used strategically. An established technology can provide a competitive advantage when it is viewed from a strategic perspective. For example, the often touted success of American Hospital Supply (AHD) was the result of using old technology in a new way. AHD simply placed tried-and-true terminals on the desks of purchasing agents, supplanting catalogs. Thus, even a simple, well-established IT has been used as a competitive tool when corporate management develops a strategic viewpoint about its computing and telecommunications resources.

E-mail was seen as a facilitator of business change as early as 1995 (Garai & Pravda, 1995) and has become as indispensable a business tool as the telephone, the postal service, or the fax machine (Harper, 2002). The increasing use of e-mail systems resulted from the desire of users to share information. For these and other reasons, e-mail has become a vital technology in today's organizations, thus improving a firm's competitive position by increasing productivity and enhancing communication inside (intraorganizational) and outside (interorganizational) the organization (McManus, Carr, Sankar & Ford, 2002).

E-mail is an example of an information technology that has potential as a strategic tool because of its capacity to increase efficiency, decrease costs, and improve productivity. Intraorganizational uses of e-mail include what has been termed productivity use and functional coupling use. Productivity use represents the use of e-mail to increase the speed, efficiency, and effectiveness of communication (e.g., scheduling meetings) (McManus et al., 2002). The utilization of e-mail for these tasks improved the productivity of individuals, giving them the ability to concentrate on more critical tasks.

As literature has indicated, many businesses have restructured themselves into high-performance teams to accomplish strategic and tactical goals. Such restructuring places a premium on reliable and timely communication between team members and functional departments. This coordination of efforts between functional areas avoids duplication and supports the timely completion of critical tasks. E-mail has been shown to support functional coupling to accomplish tasks or to achieve communications that typically involve multiple departments or divisions (e.g., polling opinions on a topic). Therefore, e-mail facilitated functional coupling by providing better coordination of a firm's internal processes (McManus et al., 2002).

Furthermore, it is not uncommon for organizations to read employees' e-mail that is believed to be inappropriate for the business environment. These findings have been supported in other studies, which refer to such applications as "inappropriate use" of e-mail technology (Rice, 1994). This stands to reason that if employees are utilizing time for social activities, the company is not achieving its business goals. Previous literature indicated that many office workers are tempted to do a little business of their own on company time, with e-mail and fast Web connections at their fingertips (Whitford, 1998). The misuse of e-mail for non-business activities has the potential to negatively impact the competitive performance of the company. Therefore, as e-mail is viewed as a strategic information technology resource, consideration of the misuse of the resource is important to avoid negating benefits.

E-mail supports the drive for competitive advantage by reducing cost and increasing productivity, providing linkages with suppliers and customers, and changing the way goods and services are provided and supported. It is these competitive uses of e-mail that make it a strategic resource. If e-mail were removed from many modern organizations, they would likely fail to remain competitive. "Building an e-mail database should be central and basic to business because this means of communications is both extremely effective and very economical" (American Gas, 2002, p.10). These strategic uses of e-mail ultimately provide a method of gaining a competitive advantage over rivals and should be a component of overall corporate strategy. Therefore, to consider e-mail other than a strategic resource places it back into a utilitarian category.

FUTURE TRENDS

During the height of the technology boom of the mid-1990s, information technology (IT) investments were in excess of 50% of capital budget expenditures in U.S. organizations (Rockart, Earl & Ross, 1996). Managers and researchers agree that IT must be appropriately utilized within these organizations in order to achieve increased worker productivity, better decision making, or other expected benefits. Thus, researchers continue to develop new theories in an effort to inform IS professionals who design and manage information technology, such as e-mail, which support managerial communication.

CONCLUSION

The strategic use of information technology capabilities within the broader corporate framework may well be the key to the future success of many businesses. The notion of strategic integration and communication of information systems with the production, marketing, financial, and other functional systems in the organization should influence technology selection and the strategic direction in some firms. Today's companies have discovered that they must aggressively pursue new information technologies, such as e-mail, to exploit their own core competencies. Successful organizations have found ways to combine their core capabilities with the available information technologies to remain competitive.

Although computer and telecommunications equipment have been used extensively in modern firms, these technologies are not new. Their widespread adoption and continued application in business have established computers as crucial vehicles for organizational effectiveness. The impact of IT has been pervasive, reaching into the decision-making processes in both small and large organizations. E-mail, as an IT capability, has been used to change a product or service, differentiate a product or service, open new markets electronically, and create barriers to entry by competitors (Stephens & Loughman, 1994).

Cooperative systems for exchanging information electronically across organizational boundaries allowed organizations to coordinate and share information when pursuing common objectives. It is vital that management realize the power of IT to build a sustainable competitive advantage, by leveraging unique corporate abilities with IT to achieve long-term performance gains (Kettinger, Grover, Guha & Segars, 1994). For example, properly implementing internal e-mail systems in multi-office organizations saves time and money over conventional courier services and fax (Vice, 1994). Therefore, organizations must realize that electronic communication is a necessity in the current dynamic global marketplace, and the effective use of any information technology can create a competitive advantage.

REFERENCES

American Gas. (2002). The power of e-mail. *84*(6), 10-11.

Casperson, D. (2002). E-mail etiquette: How to make sure your message gets across. *The American Salesman, 47*(7), 10-13.

Carroll, J. (1993). Electronic mail, workflow and you. *CMA Magazine, 67*(5), 9-11.

Garai, G. & Pravda, S. (1995). Pros and cons of using e-mail. *Electronic Business Buyer*, *21*(1), 19.

Harper, D. (2002). The rise of e-mail marketing. *Industrial Distribution, 91*(7), 56.

Kettinger, W.J., Grover, V., Guha, S. & Segars, A.H. (1994). Strategic information systems revisited. *MIS Quarterly, 18*(1), 31-58.

King, W.R. & Teo, T.S. (1994). Facilitators and inhibitors for the strategic use of information technology. *Information and Management*, *27*(2), 71-87.

Kirschner, E.M. (1995). Chemical companies discover a weapon for globalization and reengineering. *Chemical and Engineering News*, *73*(13), 62-71.

McManus, D.J., Carr, H.H., Sankar, C.S. & Ford, F.N. (2002) Intraorganizational versus interorganizational uses and benefits of electronic mail. *Information Resources Management Journal*, *15*(3), 1-13

Reark, R. (1989). Electronic mail speeds business communications. *Small Business Reports*, *14*, 73-75.

Rice, R.E. (1994). Relating electronic mail use and network structure to R&D work networks and performance. Journal of Management Information Systems, *11*(1), 9-29.

Rockart, J., Earl, J. & Ross, J. (1996). Eight imperatives for the new IT organizations. *Sloan Management Review, 38*, 43-55.

Stephens, C. & Loughman, T. (1994). The CIO's chief concern. *Communication. Information and Management*, *27*(2), 129-137.

Whitford, D. (1998). Confessions of an online moonlighter. *Fortune*, *137*(8), 443-444.

Vice, P. (1994). Electronic connection. *Canadian Insurance*, *99*(10), 12-13.

KEY TERMS

Competitive Advantage: The ability to gain a disproportionately larger share of a market because of cost leadership or product or service differentiation.

Electronic Communication: From a human perspective, it is the transfer of information—textual, graphic, oral, or visual—from one point/person to another via electric, electromagnetic, or photonic means. Machine-to-machine communications deals in bits and bytes only.

Electronic Mail: Allows users to communicate electronically with other users as if two typewriters were connected by a channel. E-mail adds a new dimension to the office environment, replacing paper copies and reducing time of transmittal. E-mail is the transmission of messages, textual or graphic, over various communications networks.

Information Technology: Computer systems and applications that include the organization's hardware, software, networking, and telecommunications.

Interorganizational Communications: Communications that include entities that are legally external from the organization where legal recourse may occur due to the content of the message.

Intraorganizational Communications: Communications internal to the legal boundaries of an organization, though it may cross departmental or divisional boundaries. There is no legal recourse for the content of a message.

Strategic: A tactical plan of action that is essential to the strategy of an organization, such as the use of communication to gain a competitive advantage.

Uncertainty: Not known or established; questionable. Uncertainty is generally caused by a lack of information.

Email Usage in South Pacific Distance Education

Jonathan Frank
Suffolk University, USA

Janet Toland
Victoria University of Wellington, New Zealand

Karen D. Schenk
K.D. Schenk and Associates Consulting, USA

INTRODUCTION

The impact of cultural diversity on group interactions through technology is an active research area. Current research has found that a student's culture appears to influence online interactions with teachers and other students (Freedman & Liu, 1996). Students from Asian and Western cultures have different Web-based learning styles (Liang & McQueen, 1999), and Scandinavian students demonstrate a more restrained online presence compared to their more expressive American counterparts (Bannon, 1995). Differences were also found across cultures in online compared to face-to-face discussions (Warschauer, 1996). Student engagement, discourse, and interaction are valued highly in "western" universities. With growing internationalization of western campuses, increasing use of educational technology both on and off campus, and rising distance learning enrollments, intercultural frictions are bound to increase.

This research adds to the body of knowledge by evaluating e-mail effectiveness as a communication medium in facilitating meaningful class participation in two distance education institutions. This work investigates online student interaction in the South Pacific region. The scattered geography of the South Pacific has produced immense variations in culture among a relatively low population base. For example, in the tiny island group of Vanuatu with a population of less than 200,000, over 110 different languages are spoken (Myers & Tan, 2002). This makes the South Pacific an ideal laboratory in which to explore the impact of cultural differences on online learning, as the region contains a broad representation of the cultural characteristics found throughout the world.

Subjects were drawn from business information systems and computer information technology classes at the University of the South Pacific and Central Queensland University. Three research questions were addressed:

- Does cultural background affect the extent to which distance education students use e-mail to communicate with educators and other students for academic and social reasons?
- Does cultural background affect the academic content of e-mail messages from distance education students?
- Does cultural background influence distance education students' preference to ask questions or provide answers using e-mail instead of face-to-face communication?

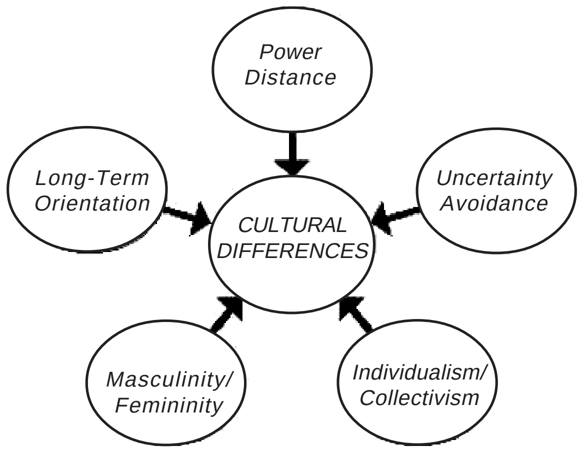
BACKGROUND

There have been a number of papers that have examined the impact of cultural diversity and group interaction in computer-mediated communication environments (Jarvenpaa & Leidner, 1998). Hofstede's (1991) well-known model categorizes different cultures according to five pairs of dimensions (Figure 1)

Although not exhaustive, Hofstede's model has been widely used, and it provides a useful starting point for exploring the influence of cultural backgrounds (Holden, 2002; Myers & Tan, 2002). For this research the focus was on the dimensions of individualism versus collectivism, and high power distance versus low power distance. These two dimensions were considered to have the most impact on learning style, the individualism/collectivism dimension will affect the way students interact with their peers, and the power distance dimension will influence the way they interact with their professor.

Hofstede's work indicated that there was a strong correlation between a country's national wealth and the degree of individualism in its culture. Richer countries tend to have an individualistic style, whereas poorer countries are more collectivist. As a poorer country becomes wealthier, it tends to move towards an individualistic pattern. Additionally, people from a rural background tend to be more collectivist than those from an urban background. Countries which fall into the low power distance, individualist category are Australia, New

Figure 1. Hofstede's model of cultural differences



Zealand, North America, England, and Scandinavia; countries in Asia, India, and South America would be considered high power distance and collectivist (Hofstede, 1991).

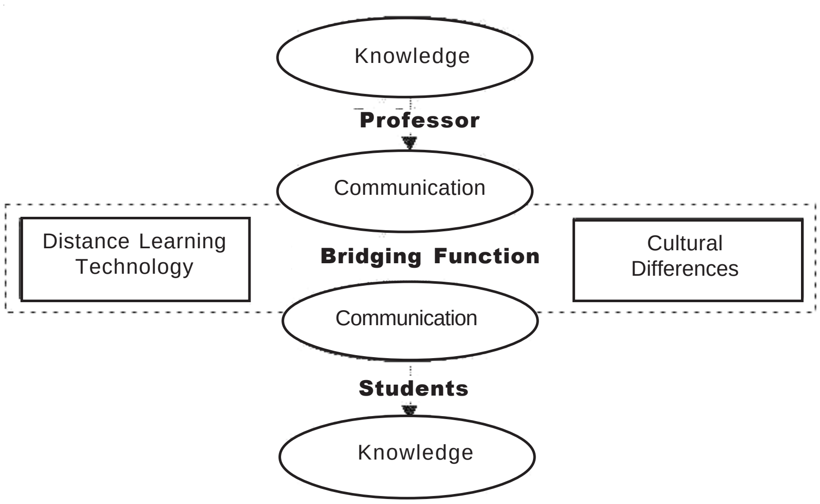
If a country is collectivist, it is also likely to exhibit characteristics of a high power distance country, where the views of senior people tend not to be questioned. Pacific Island people are in the high power distance category with their system of chiefs and their tradition of not questioning the chief's decision. South Pacific society is also collectivist with the custom of "Kere Kere" or not being able to refuse a favor that is asked of you by a member of your own in-group.

There have been a number of recent publications reviewing aspects of the development of IT in the South Pacific (Davis, McMaster & Nowak, 2002; Olutimayin, 2002; Purcell & Toland, 2004), however no research has yet been published that maps Hofstede's model on the many South Pacific cultures. Lynch, Szorengi, and Lodhia (2002) have explored Hofstede's framework with respect to Fiji, hypothesizing where the indigenous Fijian population and the Indo Fijian population would fit into the framework, however they are still in the process of collecting empirical evidence to validate their theories. This research forms a useful starting point to locate South Pacific cultures on the dimensions of individualism, collectivism, and power distance.

Most research on the effect of cultural differences has focused on traditional face-to-face teaching, rather than distance education. The literature has often cited difficulties in motivating students from collectivist (as opposed to individualistic) cultures to "speak up" in a face-to-face learning situation. Students from a collectivist culture prefer to listen, reflect, and discuss material with their peers, before preparing a written response. In common with many other collectivist cultures, it would be considered undesirable for students to speak up in class, as communication is mostly teacher centered. In Fiji, lecturers have widely commented on the "quietness" of their students (Handel, 1998). Additionally, in some pacific cultural norms, student silence is seen as a sign of respect for teachers (Matthewson et al., 1998).

More research is needed to understand fully the cultural contexts in which distance education programs are situated and how distance students process materials, especially in a second language (Guy, 1991). One study

Figure 2. Container model of knowledge transfer



concluded that students from collectivist cultures would be less receptive to distance education than students from individualistic cultures. Cultural differences also affect the way students interact with different types of messages (Anakwe, 1999). It has been shown that students from individualistic cultures are more willing to respond to ambiguous messages, which may result in a different approach to e-mail (Gudykunst, 1997).

THE SOUTH PACIFIC STUDY

The two distance education institutions studied have different approaches to education. In the University of the South Pacific (USP), knowledge transfer tends to occur in a one-directional mode from professor to student. This familiar model is often called the container or transfer model of knowledge transfer, or *migratory* knowledge (Badaracco, 1991). USP's distance education model reflects the traditional correspondence course model, in that learning materials are packaged for students, and little interaction is anticipated between student and teacher. Figure 2 proposes an adaptation of the container model, incorporating a bridging function (Jin et al., 1998) as well as components to reflect the effect of distance learning technology and cultural factors on knowledge transfer.

In contrast, Central Queensland University (CQU) distance education pedagogy is extraordinarily dependent on e-mail interaction between students and professors. Fifty percent of students' grades are based on group exercises. Groups consist of 5-10 students from 12 countries. Students are assigned to groups by the course coordinator to maximize in-group diversity. Students are required to post within group and between group evaluations to a threaded discussion forum each week. Students are encouraged to learn from each other, as well as their professor (Romm, 2001; Jones, 1999).

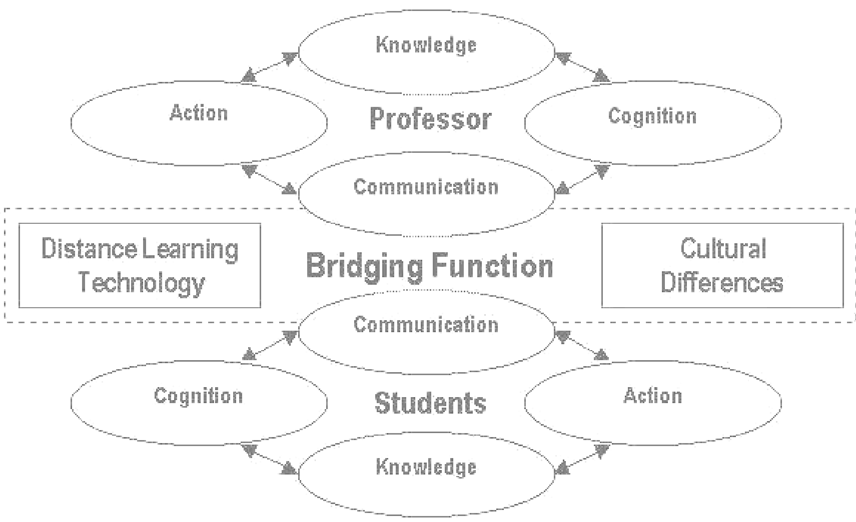
The *social construction* model of learning and knowledge transfer (Figure 3) fits well with the CQU approach to learning. This model represents knowledge as one part of a process. It considers knowledge, cognition, action, and communication as inseparable.

"The term enactment captures this interrelationship among the different aspects of knowing, acting, communicating, and perceiving. Knowledge takes on meaning as the entity interacts with its environment through communicating with other entities, acting (and thereby changing the environment) and interpreting cues arising from these interactions." Weick (1979)

FINDINGS

To assess the impact of these two approaches to distance education on students from different cultures, two studies were conducted in parallel. The University of South Pacific study examined how distance education students from different cultural backgrounds used e-mail to communicate with other students and teachers. Four hundred USP students located at different regional centers were surveyed about their e-mail usage. In the Central Queensland University study, postings to threaded dis-

Figure 3. Social construction model of knowledge transfer



cussion groups by 867 distance education students based in Fiji and Australia were analyzed. For a detailed discussion of research methodology and findings, see Frank, Toland, and Schenk (2004).

Study findings suggested differences in usage of e-mail by students from individualistic and collectivist cultures both in terms of the quantity of messages sent, and in the nature of those messages. Australian students (individualist—low power distance culture) send significantly more posts to their discussion board than do Fijian students (collectivist—high power distance culture). Australian students appear more ready to respond to questions than Fijian students do. Fijian students volunteer fewer answers to the threaded discussion group. One explanation might be associated with fear of "losing face" among their peer group. Another possibility might be Fijian students view participation on the list as not directly affecting their grade, and therefore see no reason to volunteer answers.

An analysis of Fijian messages confirmed that a large percentage of messages were social in nature; students seem to use the lists more for forming groups than Australian students. Further investigation of questions asked by Fijian students indicated a need for reduction in ambiguity about assignment specifications.

CONCLUSION

The South Pacific region, isolated, vast, and culturally diverse, was selected as an appropriate research environment in which to study the effect of cultural differences and educational technology on distance education. The research context was two competing distance education institutions in Fiji: the University of the South Pacific and Central Queensland University. E-mail was used for teaching and learning in different ways at these institutions. At Central Queensland University, interactive e-mail was incorporated into distance education pedagogy across all courses, whereas at the University of the South Pacific e-mail was used more informally.

Three research questions were addressed: Does cultural background affect the extent to which distance education students use e-mail to communicate with educators and other students for academic and social reasons? Does cultural background affect the academic content of e-mail messages? Does cultural background influence students' preference to ask questions or provide answers using e-mail?

Two studies were conducted in parallel. Subjects were drawn from business information systems and computer information technology taught by distance educators at the University of the South Pacific and Central Queensland University. Four hundred University of the South Pacific students located at different regional centers were surveyed about their e-mail usage. In the Central Queensland University study, postings to course discussion lists by 867 students based in Fiji and Australia were analyzed. The results of these studies suggest that there are significant differences in the use of e-mail by distance education students from different cultural backgrounds.

High power-distance/collectivist students are more likely to use e-mail to interact socially with their peers than they are to use it for contacting their professors. These distance education students tend to ask more questions than low power-distance/individualist students, and their questions are more likely to focus on group formation or reduction of assignment ambiguity. They also tend to volunteer fewer answers than low power-distance/individualistic students. Anxiety over "loss of face" may contribute to collectivist students' reluctance to volunteer answers.

FUTURE TRENDS

Notwithstanding the geographical limitations of these studies, the findings may have wider relevance as distance education classes worldwide become more culturally diverse. Future researchers might wish to pursue a more qualitative study in different geographic settings, interviewing students and academics to gain a more in-depth understanding of the reasons for some of the observed behaviors. Additional research in this area is important as we expand teaching across cultural boundaries through the use of distance education.

REFERENCES

Anakwe, U. et al. (1999). Distance learning and cultural diversity: Potential users' perspective. *International Journal of Organizational Analysis, 7*(3), 224-244.

Badaracco, J.L. Jr. (1991). *The knowledge link.* Boston: Harvard Business School Press.

Bannon, L.J. (1995). Issues in computer-supported collaborative learning. In C. O'Malley (Ed.), *Computer-supported collaborative learning* (pp. 267-281). Berlin: Springer-Verlag.

Davis, C.H., McMaster, J. & Nowak, J. (2002). IT-enabled services as development drivers in low-income countries: The case of Fiji. *The Electronic Journal on Information Systems in Developing Countries, 9*(4), 1-18.

Frank, J., Toland, J. & Schenk, K. (2004). The effect of culture on email use: Implications for distance learning In C. Howard, K. Schenk & R. Discenza (Eds.), *Distance learning & university effectiveness: Changing educational paradigms for online learning* (pp. 213-233). Information Science Publishing

Freedman, K. & Liu, M. (1996). The importance of computer experience, learning processes, and communication patterns in multicultural networking. *Educational Technology Research and Development, 44*(1), 43-59.

Gudykunst, W.B. (1997). Cultural variability in communication. *Communication Research, 24*(4), 327-348.

Guy, R.K. (1991). Distance education and the developing world: Colonisation, collaboration and control. In T.D. Evans & B. King (Eds.), *Beyond the text: Contemporary writing on distance education* (pp. 152-175). Geelong, Australia: Deakin University Press.

Handel, J. (1998). *Hints for teachers.* Centre for the Enhancement of Teaching and Learning, University of the South Pacific.

Holden, N.J. (2002). *Cross-cultural management: A knowledge management perspective.* London: Prentice-Hall.

Hofstede, G. (1994). *Cultures and organizations*. Harper Collins.

Jarvenpaa, S.L. & Leidner, D. (1998). Communication and trust in global virtual teams. *Journal of Computer Mediated Communication, 3*(4).

Jin, Z. et al. (1998). Bridging US-China cross-cultural differences using Internet and groupware technologies. *Proceedings of the 7th International Association for Management of Technology Annual Conference*. Retrieved from *www.cim-oem.com/bridge_8c18c.html*

Jones, D. (1999). Solving some problems with university education: Part II. *Proceedings of AusWeb99, the Fifth Australian World Wide Web Conference*. Retrieved from *ausWeb.scu.edu.au/aw99*

Liang, A. & McQueen, R.J. (1999). Computer-assisted adult interactive learning in a multi-cultural environment. *Adult Learning, 11*(1), 26-29.

Lockwood, F. et al. (2000, August). *Review of distance and flexible learning at the University of the South Pacific*. Report submitted to the Vice Chancellor and senior staff at the University of the South Pacific.

Lynch, T., Szorengi, N. & Lodhia, S. (2002). Adoption of information technologies in Fiji: Issues in the study of cultural influences on information technology adoption. *Proceedings of the Conference on Information Technology in Regional Areas,* Rockhampton, Australia.

Matthewson, C. et al. (1998). Designing the Rebbelib: Staff development in a Pacific multicultural environment. In C. Latchem & F. Lockwood (Eds.), *Staff development in open and flexible learning.* London: Routledge.

Myers, M. & Tan, F. (2002). Beyond models of national culture in information systems research. *Journal of Global Information Management, 10*(1), 24-32.

Olutimayin, J. (2002). Adopting modern information technology in the South Pacific: A process of development, preservation, or underdevelopment of the culture. *Electronic Journal of Information Systems in Developing Countries*, *9*(3), 1-12.

Purcell, F. & Toland, J. (2004). Electronic commerce for the South Pacific: A review of e-readiness. *Electronic Commerce Research Journal, 4*(3).

Romm, C. et al. (2001). Searching for a "killer application" for on-line teaching—or are we? *Proceedings of AMCIS2001,* Boston.

Warschauer, M. (1996). Comparing face-to-face and electronic discussion in the second language classroom. *CALICO Journal, 13*(2), 7-26.

Weick, K.E. (1979). *The social psychology of organizing* (2nd ed.). New York: McGraw-Hill.

KEY TERMS

Container Model: Knowledge flows directly from the teacher to the learner, independently of the learner's environment.

Cultural Diversity: The way that people have different values and attitudes depending on where they were born and the society that brought them up.

Enactment: Knowledge only takes on meaning as it interacts with the learner's environment.

Individualism/Collectivism: An individualist society is one where each person is expected to be self-sufficient and look after themselves and their immediate family. A collectivist society is one where every person is a member of a group, and loyalty is to that group. Typically, in an individualist society, any money earned will be kept by the individual that earns it, whereas in a collectivist society, earnings will be shared among the group

Long-Term Orientation: This refers to a society's attitude towards time, do they tend to plan for a long- or a short-term time horizon.

Masculinity/Femininity: Societies that are masculine would favor values such as assertiveness, competitiveness, and toughness. Societies with a more feminine focus would be more nurturing, cooperative, and concerned with the quality of life.

Migratory Knowledge: Knowledge that can travel directly from teacher to learner without changing in form or substance.

Online Group Interaction: The use of computer-mediated communication such as e-mail, chat, or a threaded discussion by a group to communicate for the purposes of carrying out a task.

Power Distance: The way authority figures are perceived: in countries with a high power distance, a leader is an authoritarian figure or a benevolent dictator and their authority will not be questioned; in a low power distance country, individuals are prepared to argue with leaders, who must be able to justify their decisions.

Social Construction: In this model of learning, knowledge, cognition, action, and communication are seen as inseparable.

Uncertainty Avoidance: This is how a society reacts to change or something that is unknown; a society with a high uncertainty avoidance will resist anything that is different.

Empirical Study of E-Commerce Adoption in SMEs in Thailand

Chalermsak Lertwongsatien
Ministry of Finance, Thailand

Nitaya Wongpinunwatana
Thammasat University, Thailand

INTRODUCTION

It was predicted that in 2004, e-commerce would generate worldwide revenue as high as $6.9 trillion, and the number of Internet users would grow as high as 765 million users in 2005 (CommerceNet, 2003). Though about 40% of online spending originates in the United States, this proportion is predicted to fall to about 38% by 2006, due to increased online spending of residents in Asia and Western Europe (Virgoroso, 2002). In addition, it was expected that online buying in Asia will grow about 89% in 2002 (Virgoroso, 2002), and the Asia Pacific will be the second most profitable, with a value of $1.6 trillion (CommerceNet, 2003).

As one of the fast growing countries in Asia, Thailand has initiated and implemented a series of national plans and activities to promote diffusion of e-commerce in both public and private sectors. Despite optimism towards e-commerce, evidence from a recent survey by the EIU (Economic Intelligence Unit) suggested that Thailand's adoption rate is still slow. According to the 2002 EIU e-readiness survey, Thailand was ranked 46 from the 60 main economies of the world (EIU, 2002). It is therefore important to understand factors affecting a firm's decision on e-commerce adoption. A deeper and clearer understanding on such issues would help design appropriate e-commerce models to target consumers in these economies and also to formulate appropriate policies to accelerate e-commerce diffusion in other countries.

This article examines the factors influencing e-commerce adoption decisions in small and medium enterprises (SMEs) in Thailand. We classify firms into three main groups—*adopters*, *prospectors*, and *laggards*—based on their extent of e-commerce innovativeness (Roger, 1995), defined as the extent to which an organization is relatively earlier to adopt e-commerce than others.

To begin, adopters are firms that have already implemented and used e-commerce in their business activities. These firms are creative and innovative in applying leading-edge technologies, such as e-commerce, compared with their competitors. Second, prospectors are firms that have not yet implemented e-commerce, but they have a specific plan in the near future to adopt and implement e-commerce. Prospectors tend to avoid the immediate application of leading-edge technologies; however, after a certain period of time, they readily adopt innovations that have been proven effective. Finally, laggards are firms that have not implemented e-commerce and have no plan or intention to adopt e-commerce in the near future. Laggards typically are slow in adopting new innovation; however, they may decide to adopt the technologies when forced by business competition.

Most of the IT innovation studies have been conducted in the US. However, Asian firms are different from U.S. firms in many respects, such as geographic, political, and cultural aspects. The research findings from this study can help in determining whether the organizational innovation theory can be generalized across other settings, particularly in Asian settings.

FACTORS INFLUECING E-COMMERCE ADOPTION

Organizational innovation was adopted as a theoretical foundation for this study. Organizational innovation can be defined as the development and implementation of ideas, systems, products, or technologies that are new to the organization adopting it (Rogers, 1995). Innovations are means of changing an organization, either as a response to changes in the external environment or as a preemptive action to influence the environment. The adoption of innovation is a process that includes the generation, development, and implementation of new ideas or behaviors (Rogers, 1995). Innovations can be categorized by a broad range of types, including new products or services, new process technologies, new organizational structures or administrative systems, or new plans or programming pertaining to organizational members (Poutsma, Van Uxem & Walravens, 1987). Adoption of e-commerce, hence, can be regarded as one form of innovation adoption.

The innovation literature has identified various groups of variables that are possible determinants of organizational adoption of an innovation (Fichman & Kemerer, 1997; Kimberly & Evanisko, 1981; Tornatzky & Fleischer, 1990; Chau & Jim, 2002). Based on a synthesis of the organizational innovation literature, Kwon and Zmud (1987) identified five sets of factors that may influence IT innovation. These sets include user characteristics, task characteristics, innovation characteristics, organizational characteristics, and environmental characteristics. Kimberly and Evannisko (1981) proposed three clusters of predictors for innovation adoption: characteristics of organization, characteristics of leader, and characteristics of environment.

Recently a number of IT innovation studies (e.g., Boynton, Zmud & Jacobs, 1994; Tornatzky & Fleischer, 1990) have adopted an emerging theory from the strategic management literature—absorptive capacity (Cohen & Levinthal, 1990)—to explain a firm's abilities in adopting and assimilating an innovation. Boynton et al. (1994), for example, argued that a firm's ability to effectively use IT is influenced by the development of IT-related knowledge and processes that bind them together the firm's IT managers and business managers. They pointed to the organizational climate as the key factor influencing the ability of firms to absorb new knowledge and technology. Fichman and Kemerer (1997) found that organizations are more likely to initiate and sustain the assimilation of software process innovations when they have a more extensive existing knowledge in areas related to the focal innovation. Table 1 summarizes factors potentially influencing a firm's decision in adopting information technology.

Table 1. Factors influencing information technology adoption

Organization Factors
• Organization Size • Organization Readiness • Employee IT Knowledge
Leadership Factors
• Top Management's IT Knowledge • Top Management's Innovativeness • Top Management Support
Technology Factors
• Relative Advantage • Compatibility • Complexity
External Factors
• Environment Uncertainty • Information Intensity

Drawing from organizational innovation literature, we identify and test six research variables, representing three major groups, *organizational factors*, *technology factors*, and *external factors*, potentially influencing a firm's decision in e-commerce adoption. First, organizational factors have been the most widely used and tested as the key determinants of innovation (Grover & Goslar, 1993; Thong, 1999). In this study, we focus on three sets of variables: structural variable (size), process variable (top management support for e-commerce), and IT context variables (existence of IT department). Size is one of the most widely investigated variables for innovation adoption. Large firms are more likely to adopt innovation since they are capable of absorbing the risk associated with innovation, and have sufficient resources and infrastructure to facilitate the implementation of innovation (Fichman & Kemerer, 1997). Process factors have also frequently been adopted as a key determinant of IT-related innovation adoption, especially the roles of top management. The IT innovation literature generally reported a positive effect of senior management support on IT-related innovation (Orlikowski, 1993; Rai & Patnayakuni, 1996). The common rationales provided include influencing the allocation of slack resources, and generating enthusiasm and commitment toward changes among organizational members. Finally, evidence from the innovation literature recently suggests that the role of a firm's ability to absorb new knowledge related to innovation can play an important role in innovation adoption (Cohen & Levinthal, 1990). SMEs that are familiar with IT skills and knowledge might find it easier to acquire additional knowledge necessary for adopting e-commerce. Hence, it is conceivable that the existence of IT departments in SMEs could promote e-commerce adoption.

The second group of variables is technology factors. Specific factors related to innovation characteristics are frequently used as a key determinant of innovation adoption intention. This study investigates the effects of two innovation characteristics: perceived compatibility and perceived benefits. Different organizations may face different innovation opportunities. Whether these opportunities can be exploited depends on the degree of match between the innovation's characteristics and the infrastructure currently available in the organization (Rogers, 1995). In addition, not all innovations are relevant to an organization. The degree of relevance depends on the potential benefits organizations received.

The third group of variables is an external factor. Past studies have stressed the importance of environments. When organizations face a complex and rapidly changing environment, innovation is both necessary and justified (Pfeffer & Leblebici, 1977). Environmental factors, especially market factors (i.e., competitiveness), cannot be controlled by organizations; rather, they affect the way

Table 2. Examined variables

Factors	Theoretical Representation	Variables
Organizational Factors	Organizational Structure	Size
	Organizational Process	Top Management Support for E-Commerce
	IT Context	Existence of IT Department
Technology Factors	Technological Context	Perceived Benefits
		Perceived Compatibility
Environmental Factors	Organizational Environment	Competitiveness

Table 3. Pair-wise analysis of the mean difference among the three organization types

Variables	Mean Difference Between		
	Adopters & Laggards	Adopters & Prospectors	Prospectors & Laggards
Size	Significant*	Significant*	Not Significant
Top Management Support for E-Commerce	Significant***	Not Significant	Not Significant
Existence of IT Department	Significant***	Significant***	Not Significant
Perceived Benefits	Significant**	Not Significant	Significant*
Perceived Compatibility	Significant***	Not Significant	Significant***
Industry Competitiveness	Significant**	Not Significant	Not Significant

firms conduct their businesses. Thus, it is conceivable that environmental factors create a need for firms to adopt IT-related innovation such as e-commerce. This study examines the effect of competitiveness on e-commerce adoption. Table 2 summarizes the research variables used in this study.

DATA COLLECTION

Survey was the primary research methodology of this study. The target firms for this study are SMEs in Thailand. Based on the definition accepted by the Ministry of Industrial in Thailand, SMEs are businesses that have overall asset values less than or equal to 200 million baht for manufacturing and service firms, 100 million baht for wholesalers, and 60 million baht for retailers. While not officially being defined, the size of SMEs, reflected by the number of employees, has generally been accepted to be fewer than 200 (Sevilla & Soonthornthada, 2000). We adopt this number as the maximum size of SMEs.

Data was collected through a national survey in several major provinces in Thailand. Respondents were those who influenced or were part of a decision-making process of e-commerce adoption. In total, 1,200 packages of questionnaire were distributed, and 452 questionnaires were returned. Sixty-six questionnaires were unusable. The total response rate of this study is 32.16%.

Respondent firms were subsequently classified based on their e-commerce adoption action. Firms were classified as *adopters* if they had already adopted e-commerce; as *prospectors* if they had not adopted e-commerce, but had a specific plan to adopt e-commerce within one year; and as *laggards* if they neither had a specific plan nor intention to adopt e-commerce. From the total 386 responses, 107 firms (27.7%) were classified as adopters, 52 firms (13.47%) as prospectors, and 227 firms (58.8%) as laggards.

FINDINGS

A post-hoc multiple comparison (Scheffee's) was employed to perform a pair-wise comparison of the mean difference on the key factors among the three organization types (see Table 3). The results suggest that adopters and laggards were significantly different in all factors. However, the results of the mean difference between adopters and prospectors and between prospectors and laggards are mixed. In particular, adopters and prospectors are significantly different in size and existence of IT department. Prospectors and laggards are significantly different in perceived benefits and perceived compatibility.

FUTURE RESEARCH OPPORTUNITIES

For future research, while this study incorporates a number of key variables identified from the literature, future studies may expand the research model by incorporating various variables to cover more comprehensive aspects of the phenomenon, such as variables reflecting key aspects of e-commerce, variables reflecting SME context, and variables representing a gap between Asian and Western cultures.

In addition, this study examines only one aspect of e-commerce adoption, the extent to which an organization is relatively earlier to adopt e-commerce than others. Future study may attempt to test other aspects of e-commerce adoption. The diffusion of innovation literature can be used as a foundation to develop dependent variables. Fichman (2001), for example, identified a number of measures of organizational innovation such as earliness of adoption, infusion, and assimilation. While these variables are widely used specifically in the context of information technology, they can be applied in the context of e-commerce. For instance, future study may develop a research model to test the infusion of e-commerce in SMEs (i.e., the extent to which e-commerce is used in a complete and sophisticated way). By adopting a more sophisticated measure, future study might apply a more sophisticated technique in testing the relationships in the research model, such as linear regression and structural equation modeling techniques (i.e., LISREL, PLS).

CONCLUSION

It is unarguable that e-commerce is becoming one of the key technologies driving businesses in the current dynamic environment. A study of e-commerce would expand our understanding on the rational underlying the thinking logic of firms in adopting the technology.

Results from our statistical analysis reveal an inside look at key factors that influence adoption decision of SMEs in Thailand. Overall, the results strongly support that organizations differ in key variables identified from the innovation literature, reinforcing that key variables identified from the organization innovation theory are applicable in the context of Asian settings and in the context of e-commerce innovation.

A number of conclusions can be drawn based on these results. First, the amount of strategic emphasis firms give to IT (i.e., IT-related knowledge, resources, and capabilities) has an impact on e-commerce adoption intention. Firms that strongly support the use of information technology, by formally establishing an IT department, are more likely to adopt e-commerce earlier than firms with less IT support. Second, we found that prospectors significantly differ from laggards in technology factors—perceived benefits and perceived compatibility—whereas there is no difference between prospectors and adopters in the technology factors. These results imply that technology factors do have an influence on attitudes toward e-commerce (i.e., changing from unfavorable—laggards—to favorable tendency—prospectors), but have no influence on a relative earliness of adoption stage (i.e., no change from prospectors to adopters). Third, e-commerce adopters are more likely to operate in a more competitive environment, compared with the other two organization types. This also implies that SMEs operating in a competitive environment are constantly scanning and implementing new technologies such as e-commerce.

For practitioners, this study highlights the importance of IT skills and knowledge in influencing e-commerce adoption. This study shows that knowledge about IT plays a major role in influencing a firm's adoption decision. SMEs that strongly support IT deployment (i.e., high level of IT investment, existence of IT department) are more likely to adopt e-commerce earlier than those with less IT support. Since e-commerce is a key technology driving businesses in a current competitive environment, the earlier the firms adopt e-commerce, the faster the firms can extend their reach to customers and secure their share in the market by using a market access capability of e-commerce.

REFERENCES

Boynton, A.C., Zmud, R.W. & Jacobs, G.C. (1994). The influence of IT management practice on IT use in large organizations. *MIS Quarterly*, (September), 299-318.

Chau, P.Y.K. & Jim, C.C.F. (2002). Adoption of electronic data interchange in small- and medium-sized enterprises. *Journal of Global Information Management*, *10*(4), 61-85.

Cohen, W. & Levinthal, D. (1990) Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 128-152.

CommerceNet. (2003). *Industry statistics—world wide statistics.* Retrieved March 2003 from *www.commerce.net/research/stats/indust.html*

EIU (Economic Intelligent Unit). (2002). *The Economist Intelligence Unit e-readiness rankings.* Retrieved July 2002 from *www.eiu.com*

Fichman, R. & Kemerer, C.F. (1997). The assimilation of software process innovations: An organizational learning perspective. *Management Science*, *43*(10), 1345-1363.

Grover, V. & Goslar, M. (1993). The initiation, adoption, and implementation of telecommunications technologies in U.S. organizations. *Journal of Management Information Systems*, *10*(1), 141-163.

Iacovou, C., Benbasat, I. & Dexter, A. (1995). Electronic data interchange and small organizations: Adoption and impact technology. *MIS Quarterly*, (December), 465-485.

Kimberly, J.R. & Evanisko, M.J. (1981). Organizational innovation: The influence of individual, organizational, and contextual factors on hospital adoption of technological and administrative innovations. *Academy of Management Journal*, *24*(4), 689-713.

Kwon, T.H. & Zmud, R.W. (1987). Unifying the fragmented models of information systems implementation. In R.J. Boland Jr. & R.A. Hirschheim (Eds.), *Critical issues in information systems research* (pp. 227-251). New York: John Wiley & Sons.

NECTEC. (2002). *National Electronics and Computer Technology Center*. Retrieved from *www.nectec.or.th/home*

Orlikowski, W.J. (1993). CASE tools as organizational change: Investigating incremental and radical changes in systems development. *MIS Quarterly*, *17*(3), 309-340.

Pfeffer, J. & Leblebici, H. (1977). Information technology and organizational structure. *Pacific Sociological Review*, *20*(2), 241-261.

Poutsma, E.F., Van Uxem, F.W. & Walravens, A.H.C.M. (1987). *Process innovation and automation in small and medium sized business*. Delft, The Netherlands: Delft University.

Rai, A. & Patnayakuni, R. (1996). A structural model for CASE adoption behavior. *Journal of Management Information Systems*, *13*(2), 205-234.

Rogers, E.M. (1995). *Diffusion of innovation* (4th ed.). New York: The Free Press.

Sevilla, R.C. & Soonthornthada. (2000). *SME policy in Thailand: Vision and challenges*.: Institute for Population and Social Research, Mahidol University, Thailand.

Thong, J. (1999). An integrated model of information systems in small businesses. *Journal of Management Information Systems*, *15*(4), 187-214.

Tornatzky, L.G. & Fleischer, M. (1990). *The process of technological innovation*. Lexington.

Virgoroso, M. (2002). The world map of e-commerce. *E-Commerce Times*, (April). Retrieved from *www.ecommercetimes.com*

KEY TERMS

Absorptive Capacity: The ability to value, assimilate, and apply new knowledge.

Compatibility: The degree to which an innovation is perceived as consistent with the existing values, past experiences, and needs of potential innovation adopters.

Diffusion of Innovation: The process by which an innovation is communicated through certain channels over time among the members of a social system.

Electronic Commerce (E-Commerce): The sharing of business information, maintaining business relationships, and conducting business transactions by means of telecommunication network.

Electronic Data Interchange (EDI): The movement of specially formatted standard business documents, such as orders, bills, and confirmations sent between business partners.

Electronic Readiness (E-Readiness): The aptitude of an economy to use Internet-based computers and information technologies to migrate traditional businesses into the new economy.

Innovativeness: The degree to which an individual or other unit of adoption is relatively earlier in adopting new ideas than others.

Perceived Benefits: The extent of management recognition of the relative advantage that innovation can provide to firms.

End–User Computing Success Measurement

E

Conrad Shayo
California State Univeristy, San Bernardino, USA

Ruth A. Guthrie
California State Polytechnic University, Pomona, USA

INTRODUCTION

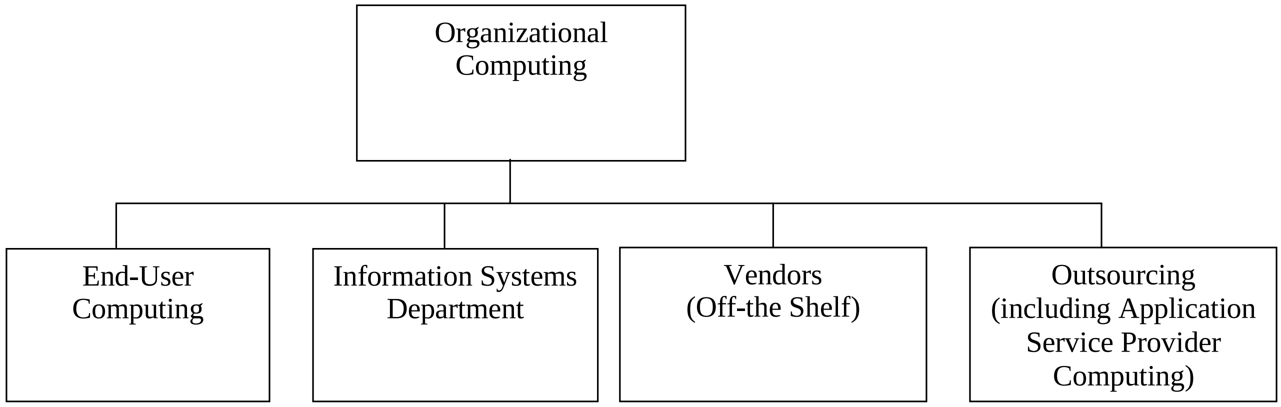
End-user computing (EUC) is the optional development of computer applications and models by personnel (individuals or groups) outside the MIS department. The emergence of EUC in the 80s and early 90s can be traced to the proliferation of computers, increased organizational computing needs, more sophisticated user application development tools coupled with higher computer and information literacy among staff and professional workers. Prior to the arrival of personal computers and graphical user interfaces, end users relied on data processing (now information technology) personnel to assist in meeting their information needs (Inman, 1986). Programming a mainframe was beyond the skills of most workers. Problems identified during this era of computing include:

- Failure to meet end-user needs.
- Cost of developing end-user applications was high.
- Large backlog of end-user applications made the development very slow.

As users required more information for decision making and highly user-friendly applications became available, end users began developing customized solutions to the needs that the data processing departments could not (Ahituv, Neumann & Riley, 1994). Today, EUC has become commonplace. Small, customized applications with spreadsheets and databases are commonplace in end-user departments. At present, EUC is just one contributor to overall organizational computing. As shown in Figure 1, the other sources of computing include applications developed by the information systems department (ISD), applications developed by vendors (off-the-shelf), and outsourcing—including application service providers (ASPs). EUC success is therefore just one contribution to overall organizational computing success. For our purposes, we define EUC success as the degree to which the organizational EUC strategy contributes to individual, group, and organizational computing success in an environment that includes applications developed by the information system department (ISD), application service providers, outsourcing parties, and of-the-shelf vendors. This means EUC complements the other components of organizational computing.

The type of applications developed by end users include transaction processing systems, manufacturing systems, expert systems, executive information systems, decision support systems, and online application processing systems (McLean, Kappelman & Thompson,

Figure 1. Components of organizational computing success



1993). There are problems, however: although end-user developed applications are low risk, localized and quickly meet user needs, unlike applications developed by the ISD or vendors, they lack integration, standardization, documentation and quality control. They also lack security, data consistency, and may result in duplication of effort. Table 1 compares the characteristics one would find in the different types of organizational computing.

In this article, we review the major research studies on EUC success measurement focusing on what has been accomplished and what remains to be done. We conclude that the measurement of EUC success seems to be an intractable problem. For example, we identified among others that there is shortage of longitudinal EUC measurement studies. There is lack of studies that have controlled for task, technology and work context. Also, there is lack of research about the relationship between EUC and other forms of organizational computing.

BACKGROUND

Measuring the Elements of EUC Success

Figure 1 indicates that EUC success should be measured as an embedded unit of organizational computing success. One problem is that the specific objectives of EUC are often invisible to the end user and to the company. The extent to which the objectives are attained is also unknown because end users often develop applications without knowledge of how their actions impact the other embedded units of organizational computing. End-user developed applications are rarely tracked by organizations. At the same time, it is not difficult to find organizations where an end- user developed application (e.g., DSS) is critical to daily operations. Furthermore, end users may be unwilling to allow objective measurement of the

Table 1. Characteristics of types of organizational computing

	End-User Computing	Information Systems Department	Vendors (Off-the-Shelf)	Outsourcing
Cost	Low/None	Moderate	Moderate	High
Schedule	Immediate	Slow	Fast	Slow
Size	Small	Moderate	Large	Moderate/Large
Control	Low	High	High	High
Risk	Low	Moderate	Low	High
Influence	Local	Local to Organizational Levels	Local to Organizational Levels	Local to Organizational Levels
Fit to User Task	High	Moderate	Low	Moderate
Support (help desk, documentation, training)	Low/None	Moderate	Moderate	Moderate
Integration	Low	Moderate	Low	High
Security	Low	High	High	High
Data Quality and Integrity	Questionable, Sometimes higher quality than corporate systems	High	Moderate	High
Duplication of Effort	High	Low/None	Low/None	Low/None
Acceptance	High	Low	Moderate	Moderate

efficiency or effectiveness of their applications, especially from an outsider, for fear of job loss. Although benign measures such as end-user satisfaction are less threatening and easier to obtain, this is problematic because end users are asked to place a value on something about which they are far from objective.

According to Gerrity and Rockart (1986), EUC success will occur if there is:

- An increase of effectiveness in the individual using the developed application.
- A move to formalize the informal system that was developed so that it can be used by other end users.
- An increase in learning on the part of the users in their ability to accomplish work.
- An increase in competitive advantage through support of new products, markets, and opportunities.
- An improvement in organizational effectiveness as users access data to improve decisions they make.

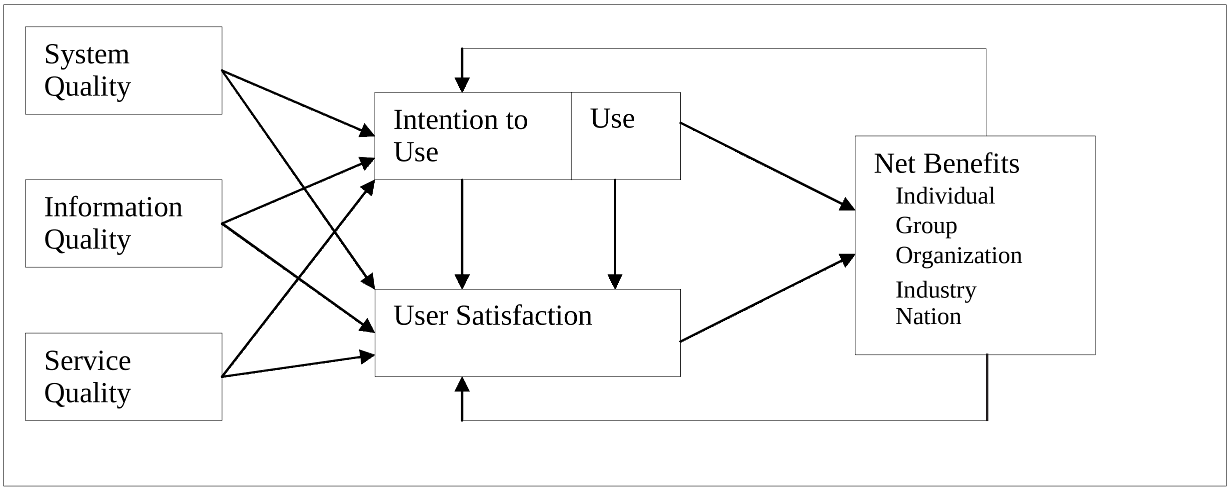
Scott Morton (1990) added a sixth item to this list: EUC success is observed if an overall increase in national wealth due to increase in knowledge of workers and information handlers exists.

The investigation of the payoff from IT investment in general has been relentless (Brynjolfsson, 1993; Seddon, Graeser & Willcocks, 2002). Since EUC became an accepted part of organizational computing, researchers have investigated whether there are significant returns from investment in EUC (Amoroso & Cheney, 1992; Blili, Raymond & Rivard, 1998; Brynjolfsson, 1993; Doll & Xia, 1997; Fabey, Land & Targett, 1999; Smithson & Hirschheim, 1998; Torkzadeh & Doll, 1991). Has the investment of time and money in EUC been worthwhile? Has the investment increased the effectiveness, efficiency and competitiveness of the organization? In reviewing the literature, we found out that most articles that did measure EUC success focused on user satisfaction, system use and end-user productivity. The measurements were mainly based on subjective self-reporting. The dependent and independent variables varied depending on the objectives of the researchers (Brancheau & Brown, 1993; DeLone & McLean, 2003; Seddon et al., 1999). Table 2 presents the independent variables and Table 3 the dependent variables that have been used to measure EUC success. The variables can be captured by the DeLone and McLean's Updated Model for measuring IS success (DeLone & McLean, 2003).

Information Systems Success and EUC Success

Several articles discuss components of IS success though its economic and objective measurement is often elusive (DeLone & McLean, 1992, 2003; Seddon, 1997; Seddon et al., 1999). However, consensus on specific measures of IS success seem to center on intention to use the system, actual system use, user satisfaction, and the net payoffs or benefits that accrue to the various stakeholders. As shown in Figure 2, success measures will depend on the unit of analysis that stands to benefit from the use of the target information system (DeLone & McLean, 2003; Seddon et al., 1999). The literature also indicates that higher levels of system use (intended or actual) and user satisfaction will occur when there is high system quality, high information quality, and high service quality. Higher system use and user satisfaction will lead to higher net benefits, which in turn will encourage more usage and higher user satisfaction.

Figure 2. DeLone and McLean's 2003 model of IS success



DeLone and McLean (2003) conclude that the six categories of success clearly indicate that IS success is a complex, multidimensional and interdependent construct and that it (IS success) should be measured as such. As shown in Figure 2, although "Net Benefits" is seen as the ultimate measure of IS success it also influences system usage and user satisfaction. DeLone and McLean recommend that more field research is needed to investigate and incorporate the "Net Benefit" measures. They also recommend that user satisfaction should always be used when "IS Use" is mandatory. However, the researcher's selection of what IS dimensions to focus on will depend on the objectives and context of the study.

Table 2. Measures used as the independent variable in EUC success measurement

Management Support for Planning and Control ✳ Top Management Understanding of IT ✳ Development of Appropriate Strategies/Policy ✳ Top Management Integration of the organizational EUC strategic plan with the IS master plan ✳ Provide a budget for educational and training programs in-house or at remote location using company personnel (software, OS, communication training, general and functional computer literacy) ✳ Encourage experimentation with various emerging technologies and applications ✳ Encourage use of IT to support a wider variety of tasks ✳ Reward efforts of using IT to meet set goals at sectional, department, divisional and corporate levels ✳ Develop core internal experts who will train and mentor others (local resident experts) ✳ Provide software library services ✳ Access to an Information Center (IC), Help Desk or Hotline **Information Center Support** ✳ Guidance in selecting hardware and software ✳ Hardware setup/configuration ✳ Software installation/backup/recovery ✳ Access to corporate data ✳ Applications maintenance/troubleshooting **Organizational Structure** ✳ Centralized ✳ Decentralized ✳ Distributed **User Characteristics** ✳ Years of Education ✳ Cognitive Style ✳ Computer Skills (command level, programming) ✳ Self-efficacy expectation ✳ Demographics (Gender, Age) ✳ Inputs to EUC: programming time, flow diagram, 4GL, pseudocode, and so forth ✳ Personality ✳ Computer Attitudes ✳ Computer Anxiety ✳ Math Anxiety ✳ Experience ✳ Skill Variety ✳ Autonomy ✳ End-user computing sophistication/Competence: ability, usage intensity, application, customization	**System Characteristics** ✳ High/Low end ✳ Quality: Security, Functionality, Ease of Use, Reliability, Availability, Portability, Importance, Integration, Adaptability, Usability, Response Time, Documentation ✳ Type of Application ✳ Value and Usefulness of system terminology **Information Characteristics** ✳ Quality of output: content, structure, correctness, accuracy, format, ease of use, completeness, timeliness, security, relevance ✳ Number of defects/function point ✳ Quality attribute models ✳ Value and usefulness of screen displays/user interface **User/System/Task Interaction** ✳ End-user participation in analysis and design ✳ End-user involvement in analysis and design: perceived risk, degree of pleasure, status value ✳ End-user usage of the system (frequency, time, number of accesses/transactions, usage pattern, dependency) ✳ Time on project in task categories ✳ LOC/hour ✳ Function points/hour ✳ Outcome expectancy ✳ Perceived usefulness ✳ Intention to use ✳ Perceived fun ✳ Satisfaction **External Support** ✳ Good relationship with hardware/software vendors/consultants (positive feelings, realistic expectations) ✳ Technical support (Service Level Agreements, Resource Level Agreements) ✳ Training by external consultants, friends, vendors, or educational institutions **Task Characteristics** ✳ Task identify ✳ Task significance ✳ Task uncertainty (complexity, volatility) **Service Characteristics** ✳ Quality: Reliability, Responsiveness, Assurance, Empathy **Other characteristics** ✳ EUC stage of growth ✳ Social pressure

Table 3. Measures used as the dependent variable in EUC success measurement

✳ End-User Satisfaction	✳ Quality of Work
✳ End-User Productivity	✳ Decision Making Performance
✳ End-User Computer Skill	✳ Job Performance
✳ End-User Ability/Competence	✳ Work/Job Effectiveness
✳ End-User Success	✳ IS Acceptance
✳ Motivation/Intention for System Use	✳ EUC Management Effectiveness
✳ Work Effectiveness	✳ System Effectiveness
✳ System Usage	✳ Competitive Advantage
✳ Customer Satisfaction	✳ Strategic Alignment
✳ Task Innovation	✳ Net Return on Investment

FUTURE TRENDS

The Future of EUC Measurement

Measuring EUC success seems to be an intractable problem. Although research studies have contributed to our understanding of the EUC success measures, there is still lack of consistency in the specification of the measures, research design, and technology used. Problems associated with EUC success measurement are:

Control and Clarity—There is a need to control for task, technology and context in studies that measure EUC success. Considerations for task variety and complexity were rarely found in the literature. Given the wide variety of skill, position, and types of computing work, it is imperative that we consider controlling for task complexity. Similarly, the research studies have examined and reported on a wide range of technologies. The problem of control is exacerbated by rapid changes in technology that make it difficult to repeat similar tests and measures over time. Context also needs to be clearly defined so that it is understood whether use of the information system is mandatory or voluntary or whether a competing information system exists.

Creation of Meta and Longitudinal Data—There is shortage of longitudinal EUC measurement studies. Most studies identified in the literature are cross-sectional. Longitudinal studies will enable us to learn how end users perform as they develop higher technology skills and how technology changes impact performance of complex tasks and the other components of organizational computing over time.

Unit of Measurement—Most studies tend to focus on the individual end user as the unit of analysis. However, if we are concerned with the degree to which EUC contributes to overall organizational computing success, we need to apply group, departmental and ultimately organizational measures of EUC. For instance, if satisfaction is a measure of EUC success, that success should be measured at many other levels. A CEO or departmental manager may have a very different perspective on an IS success than a specific end user has.

Lack of Objective Measures of End-User Performance in a field setting—End-user computing activities are rarely visible to the rest of the organization. Such activities are therefore difficult to observe, document and measure unobtrusively (Shore, 1998).

Need for Confirmatory Studies—More studies are needed to test and challenge the new DeLone and Mclean model (2003) using the context variables suggested by Seddon et al. (1999) and Brancheau and Brown (1993), and work system framework variables suggested by Alter (2002).

Lack of Conceptual Definitions—The operational definition of the measuring variables must correspond with the conceptual definition. Researchers should develop and use existing measures that are well established and validated in IS and other disciplines. For example, are we really measuring end-user satisfaction? If so, how is this different from satisfaction with artifacts researched in other fields such as marketing or engineering?

Lack of Interrelationship Studies—There is a lack of research about the relationship between EUC and other forms of organizational computing. EUC development, implementation and maintenance can cause an organization to formalize informal systems, hoping to gain productivity or control where corporate systems have failed. The existence of EUC developed systems may be an important step towards successfully developing more formalized systems. EUC developed systems may play an educational role in moving user groups towards understanding of requirements and acceptance of more formal systems. The growth of EUC communities of practice may accelerate this trend.

Lack of Global View—Most EUC research has been conducted in North America, Europe, and Australia. There is belief that the results obtained from these settings are generalizable to the rest of the world. This belief may be ill advised given differences in culture, socio-work roles, level of IT sophistication and access to technology. Models of EUC need to be checked for external validity across cultures. This is even more important in the future as companies employ a global workforce.

CONCLUSION

When future research uses existing well-defined EUC concepts and measures, it will be possible to conduct meta-analyses and improve our overall understanding of EUC success. There is also a need to pay attention to the contextual factors of end-user computing. Too many studies are focused at the individual level, ignoring departmental, work-group, organizational and even national effects. A broader view of the implications of EUC in organizational computing success could give insight into its contribution to national productivity, quality and competition.

REFERENCES

Ahituv, N., Neumann, S., & Riley, H.N. (1994). *Principles of information systems management* (4th ed.). IA: Wm. C. Brown Communications, Inc.

Alter, S. (2002). The work system method for understanding information systems and information systems research. *Communications of AIS, 9*, 90-104.

Amoroso, D.L., & Cheney, P.H. (1992). Quality end user developed applications: Some essential ingredients. *DATABASE, 23*(1), 1-11.

Blili, S., Raymond, L., & Rivard, S. (1998). Impact of task uncertainty, end user involvement, and competence on the success of end user computing. *Information and Management, 33*, 137-153.

Brancheau, J.C., & Brown, C.V. (1993). The management of end user computing: Status and directions. *ACM Computing Surveys, 25*(4), 437-482.

Brynjolfsson, E. (1993). The productivity paradox of information technology. *Communications of the ACM, 36*(12), 67-77.

DeLone, W.H., & McLean, E.R. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research, 3*(1), 60-95.

DeLone, W.H., & McLean, E.R. (2003). DeLone and McLean model of information systems success: A ten year update. *Journal of Management Information Systems, 19*(4), 9-30.

Doll, W.J., & Xia, W. (1997). A confirmatory factor analysis of the end user computing satisfaction instrument: A replication. *Journal of End User Computing, 9*(2), 24-31.

Fabey, B., Land, F., & Targett, D. (1999). Moving IS evaluation forward: Learning themes and research issues. *Journal of Strategic Information Systems, 8*(2), 189-207.

Gerrity, T.P., & Rockart, J.F. (1986). End user computing: Are you a leader or laggard? *Sloan Management Review, 27*(4), 25-34.

Inman, W.H. (1986). *Management of end user computing in information organizations.* Homewood, IL: Dow Jones-Irwin.

McLean, E.R., Kappelman, L.A., & Thompson, J.P. (1993). Converging end user and corporate computing. *Communications of the ACM, 36*(12), 79-92.

Scott Morton, M.S. (1991). Introduction. In M.S. Scott Morton (Ed.), *The corporation of the 1990s: Information technology and organizational transformation.* New York: Oxford University Press.

Seddon, P.B. (1997). A re-specification and extension of the DeLone and McLean model of IS success. *Information Systems Research, 8*(3), 240-253.

Seddon, P.B., Graeser, V., & Willcocks, L.P. (2002). Measuring organizational IS effectiveness: An overview and update of senior management perspectives. *The DATABASE for Advances in Information Systems, 33*(2), 11-28.

Seddon, P.B., Staples, S., Patnayakuni, R., & Bowtell, M. (1999). Dimensions of information systems success. *Communications of AIS, 2*(20).

Shore, B. (1998). Managing end user challenges. *Information Systems Management, 15*(1), 79-83.

Smithson, S., & Hirschheim, R.A. (1998). Analyzing information system evaluation: Another look at an old problem. *European Journal of Information Systems, 7*(3), 158-174.

Torkzadeh, G., & Doll, W.J. (1991). Test-retest reliability of the end-user computing satisfaction instrument. *Decision Sciences, 22,* 26-37.

KEY TERMS

End-User Computing (EUC): The optional development of computer applications and models by personnel (individuals or groups) outside the MIS department.

EUC Net Benefits: A measure that captures the balance of positive and negative impacts that result from EUC activities in an organization.

EUC Satisfaction: An affective measure of an end user's opinion about the net benefits derived from EUC activities in an organization.

EUC Success: The degree to which the organizational EUC strategy contributes to individual, group, and organizational computing success in an environment that includes applications developed by the information system department (ISD), application service providers, outsourcing parties, and off-the-shelf vendors.

Information Quality: The success of the information being meaningful and useful to the receiver**.**

Intention to Use: An attitude measure of an end user's goal to use a specific information system.

Service Quality: A measure of end-user opinions on the overall support provided to them by top management, ISD, and others.

System Quality: The reliability and efficiency of an information system.

System Usage: The extent of actual use of a specific information system—(mostly voluntarily).

Engineering Emotionally Intelligent Agents

Penny Baillie-de Byl
University of Southern Queensland, Australia

Mark Toleman
University of Southern Queensland, Australia

INTRODUCTION: EVOLUTION OF EMOTIONAL AGENTS

Traditionally, philosophers have defined emotions to be interruptions to otherwise logical states of being (Smith & Kirby, 2000). The recent resurgence of research in the emotional realm in both psychology and cognitive science agrees with the view developed by the late Charles Darwin who, in the late 1800s, conceived that emotions play an important part in our cognition and serve to provide us with the mechanisms for adaptive behaviour in a dynamically complex world (Smith & Ellsworth, 1985).

One relevant line of research is the realm of affective computing. This relatively new domain examines the effect that emotions have on human intelligence and endeavors to use this to further enhance the field of artificial intelligence (AI). How the concept of emotions might heighten the intelligent functioning of artificial beings is still unclear, but through the variety of research programs that currently exist, areas that might benefit are being identified. In this domain, much work is being done to develop artificial intelligences capable of identifying, processing and synthesizing emotions. Picard (1997) suggests that emotions are an integral part and a natural progression for AI. She further states that: "...the inability of today's computers to recognize, express, and have emotions, severely limits their ability to act intelligently and interact naturally with us".

Emotional decision making provides a good solution for computer systems that face the problem of enumerating and evaluating multitudinous choices within an acceptable time frame. One application of AI that is benefiting by integrating emotional decision making mechanisms is that of intelligent agents. The word *agent* is used within the AI domain to refer to a number of different applications. The most popular use of the term pertains to an autonomous artificial being that has the ability to interact intelligently within a temporally dynamic environment.

BACKGROUND

Much of the current research for achieving these types of results with artificial agents is based on appraisal theories. Models such as the Affective Reasoner (Elliot, 1992, 1998), the Oz project (Bates, Loyall & Reilly, 1992), PETEII (El-Nasr, 1998), Silas (Blumberg, Todd & Maes, 1996), Yuppy (Velasquez, 1999), extensions of the INES intelligent tutoring system (Hospers, Kroezen, op den Akker, Nijholt, & Keylen, 2003; Heylen, Nijholt, op den Akker, & Vissers, 2003) and systems based on PMFserv (Silverman, 2001; Silverman, Cornwell, & O'Brien, 2003) are just a few examples of how appraisals (cognitive and non-cognitive) are used to generate emotional states and behaviors within an artificial intelligence.

Cognitive appraisal theory and the OCC Model developed by Ortony, Clore and Collins (1988), although not originally intended for synthetic generation of emotions, is a basis for several of these projects. Table 1 displays the values used in the OCC model and subsequent AI applications to appraise an event and determine an appropriate emotion. The table is a subset of the OCC appraisals, displaying eight of the possible 22 emotional states mentioned in the theory.

Heylen et al. (2003) report a proposed extension to INES to incorporate social and emotional intelligence skills of human tutors using emotional axes defined in Kort, Reily & Picard (2001). Implementation is still at the conceptual stage. Baillie (2002) developed an emotionally motivated artificial intelligence (EMAI) architecture that has the capacity for decision making influenced by simulated human emotional intelligence. The architecture consisted of several major processing and knowledge representation areas working together in a complex network of information gathering, manipulation and update. A focal point of the architecture was an "affective space" which acted as an emotional filter influencing an EMAI agent's perception of its beliefs about the environment and, as a consequence, how it behaved. EMAI agents have been

Table 1. A subset of the OCC cognitive structure of emotion

Consequential Disposition	Focus Object (self or other)	Consequences for Focus Object	Emotion
pleased	other	desirable	Happy For
pleased	other	undesirable	Pity
displeased	other	desirable	Resentment
displeased	other	undesirable	Gloating
pleased	self	relevant	Hope
pleased	self	irrelevant	*Joy*
displeased	self	relevant	Fear
displeased	self	irrelevant	Distress

implemented and evaluated for reasonableness in their simulation of emotions. Silverman, Cornwell & O'Brien (2003) report on progress towards an open agent architecture (PMFserv) to allow exploration of a range of human attributes, including emotions. The architecture has many similar structural components providing similar functionality to the EMAI architecture, but its application domains and integration efforts hinder experimentation and simulated scenarios have not been evaluated.

AN EMOTIONALLY INTELLIGENT AGENT ARCHITECTURE

The EMAI architecture (shown in Figure 1) is one example of recent emotional agent architectures. It is a complex set of mechanisms that process emotional concepts for their use in affective decision-making and reasoning. There are two types of emotion mechanisms integrated in the EMAI architecture. The first mechanism emulates fast primary emotions (Koestler, 1967) otherwise known as motivational drives. These drives can be classified according to their source, pleasure rating and strength. In an EMAI agent, these drives are used to initiate behaviour in the agent. They can include concepts such as hunger, fatigue or arousal. The strength of the drives is temporally dynamic and at particular threshold levels the agent will set goals, that when successfully achieved, will pacify the drives. For example, over time the strength of the hunger drive will increase. At a certain point, the agent will become so hungry that it will set appropriate goals to ensure it obtains food. On the consuming of food, the strength of the agent's hunger drive will decrease.

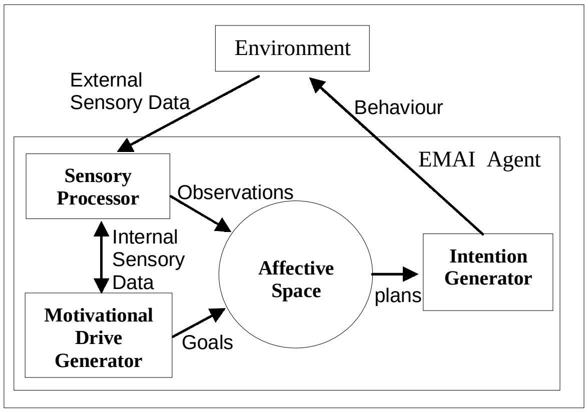
The agent's goals are generated by a motivational drive generator which consists of drive mechanisms and a set of internal state registers representing the primary emotions. Each register is represented by a single gauge and stores the value for a particular drive, for example hunger. The number of internal state registers implemented depends on the application for which the EMAI agent is being used.

The second type of emotion implemented in the EMAI architecture is secondary emotion. This category of emotion refers to the resultant mental (and in turn physical) states generated by attempts to satisfy the goals. These emotions include feelings such as happiness, anger, sorrow, guilt and boredom. Secondary emotions are represented in the EMAI architecture as values in an affective space.

The affective space is a six-dimensional space defined by six appraisal dimensions. The affective space, based on the psychological model of Smith and Ellsworth (1985), defines 15 emotions (happiness, sadness, anger, boredom, challenge, hope, fear, interest, contempt, disgust, frustration, surprise, pride, shame and guilt) with respect to the six dimensions of pleasantness, responsibility, effort, certainty, attention and control.

In addition to representing the agent's emotional state, the agent uses the affective space to associate emotions with all stimuli both internal to the agent (as internal sensory data) and within its environment (as external sensory data). The stimuli are perceived by the agent as part of an event. An event is a behavioral episode executed by the agent. Stimuli can be any tangible element in the agent's environment including the actions being performed, smells, objects, other agents, the time of day or even the weather.

Figure 1. Summary illustration of the EMAI architecture



The sensory processor of the agent is where high level observation takes place. This information is filtered through the affective space before it is used by the agent to generate outward behaviour (determined by the intention generator). It is at this point that the information has been perceived. Therefore, all information perceived by the agent is influenced by the agent's emotional state.

Before a stimulus or event can be perceived by the agent, the agent must calculate an associated emotion for each. The emotion associated with a stimulus is determined by examining each of the appraisal dimensions with respect to the agent's last encounter with the stimulus.

FUTURE TRENDS & KEY ISSUES

Synthesizing and modeling emotions is a complex task. Most affective computing research concentrates on specific, often single-dimensional aspects of generating emotions rather than adopting a holistic view or providing integrated solutions necessary to engineer emotional agents. Various key issues, problems and limitations remain for developers attempting to engineer emotional agents.

- **Emotion Range:** Although humans have a plethora of complex emotional states, research has found that viewers can best recognize six universal emotions, namely, sadness, anger, joy, fear, disgust and surprise (Ekman, Friesen & Ellsworth, 1972). Many current models simply represent these discrete emotions using levels that rise and fall as the agent's state changes. This is a simplistic view and fails to allow complex or mixed emotional states. Current models also lack the ability to maintain a temporally dynamic emotional experience with an environmental element be it an object or another agent. Humans can also express more ambiguous emotions such as pride combined with physical states such as tiredness; however, these are often interpreted in different ways by different people. Agent architectures must allow more flexible representations of emotions if modeling is to reflect reality.
- **Knowledge Representation:** An agent's knowledge representation should be deeply coupled with the knowledge-base rules. Emotions are often processed by a mechanism other than the inference engine, however it may be pertinent to include emotions in the rule-base and adjust inference engine processes to naturally handle emotion with the rules. This way emotion is a fully integrated part of the agent and can be better incorporated with the agent's knowledge and behavior allowing agents to revise their beliefs based on experience gained. Production rules usually associated with agent development (see, for example, Heylen et al., 2003) are not flexible enough for this level of knowledge representation. Contemporary affective agent architectures provide no means of attributing partial blame or credit for resulting emotions to individual elements that were part of an event. However, conceptual graphs have been shown by Baillie (2002) to provide the appropriate interrelated knowledge of goals and known concepts allowing agents to extrapolate to new situations that are related through its ontology.
- **Emotional State Combining and Decay:** The natural decrease in emotion strength or emotional decay in models is usually dealt with by using constant decay rate functions that act over time and require considerable programming and coordination effort. Baillie (2002) used the idea of a mood, represented by a point in an affective space, and modeled changing moods as movement from point to point in that space.
- **Feedback and Adaptation:** For humans, the experiencing of emotional states provides feedback that influences future behavior. Such a mechanism is not provided in many current agent architectures to the extent that it would allow an agent to adapt to its environment on an emotional basis. Extension of the relationships inherent in the Theory of Reasoned Action (Fishbein & Ajzen, 1975), viz., beliefs, assessments and intentions, or consideration of similar theoretical depictions of emotion synthesis is vital if agents are to show these traits.
- **Emotion Emergence:** Models that simulate or evaluate the emergence of emotion in an artificial agent are rare. Contemporary agent models do not apply emotional bias in decision making. While they do provide the means to generate agents with personalities and preprogrammed behaviors associated with differing dispositions, they do not allow the temporal changes in decision making considering emotions.
- **Common Standards:** The natural sciences domain is divided in its definitions and causes of emotion. This could also be a potential problem in the youthful area of affective computing where a plethora of models and theories have arisen. An emotional language standard is needed to gather efforts in the domain and continue the forward thrust in research.

Research on these issues and limitations and consideration of them by agent developers will influence the quality of engineered emotional agents and contribute significantly to the domains of affective computing and artificial intelligence. This will lead to improved affective agent architectures that integrate new approaches to emotion synthesis and affective decision making.

CONCLUSION

What the future holds for the field of affective computing is unclear. Although emotions have been thought to impede rational thinking they are now acknowledged as a vital part of rational intelligent behavior and necessary not only for survival, but also intellectual success. As affective computing is very much in its infancy, researchers need to continue to examine and assess the elementary concepts of emotion generation. No single theory stands out from the rest as being the ideal. The complexities of human emotions may be too extreme to include them holistically within an artificial intelligence at this time. Only those segments of emotional behaviour that are advantageous to the goals of an artificial being should be considered. Further categorization of emotional behaviours is necessary to identify domains where particular aspects of emotions are of advantage.

REFERENCES

Baillie, P. (2002). *The synthesis of emotions in artificial intelligences*. Unpublished Ph.D. Dissertation, University of Southern Queensland, Australia.

Bates, J., Loyall, A.B., & Reilly, W.S. (1992). An architecture for action, emotion, and social behaviour. *Proceedings of Artificial Social Systems: Fourth European Workshop on Modeling Autonomous Agents in a Multi-Agent World*, Pittsburg, (pp.55-68). Berlin: Springer-Verlag.

Blumberg, B.M., Todd, P.M., & Maes, P. (1996). No bad dogs: Ethological lessons for learning in Hamsterdam. *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behaviour*, Cape Cod, (pp.295-304). Cambridge, MA: MIT Press/Bradford Books.

Ekman P., Friesen W.V., & Ellsworth P. (1972). *Emotion in the human face*. New York: Pergamon Press.

Elliot, C. (1992). *The affective reasoner: A process model of emotions in a multi-agent system*. Unpublished PhD Dissertation, Northwestern University.

Elliot, C. (1998). Hunting for the holy grail with emotionally intelligent virtual actors. *SIGART Bulletin, 9*(1), 20-28.

El-Nasr, M.S. (1998). *Modeling emotion dynamics in intelligent agents*. Unpublished M.Sc. Dissertation, American University in Cairo.

Fishbein, M. & Ajzen, I. (1975). *Belief, attitude, intention and behaviour*. London: Addison-Wesley.

Heylen, D., Nijholt, A., op den Akker, R., & Vissers, M. (2003). Socially intelligent tutor agents. In R. Aylett, D. Ballin & T. Rist (Eds.), *Proceedings of Intelligent Virtual Agents (IVA 2003), Lecture Notes in Artificial Intelligence 2792*, (pp.341-347). Berlin: Springer-Verlag.

Hospers, M., Kroezen, E., op den Akker, R., Nijholt, A., & Keylen, D. (2003). Developing a generic agent-based intelligent tutoring system. In V. Devedzic, J.M Spector, D.G. Sampson, & Kinshuk (Eds.), *Proceedings of 3rd IEEE International Conference on Advanced Learning Technologies (ICALT 2003)*, (pp. 443). Los Alamitos: IEEE Computer Society Press.

Koestler, A. (1967). *The ghost in the machine*. London: Penguin Books.

Kort, B., Reily, R., & Picard, R. (2001). External representation of learning process and domain knowledge: Affective state as a determinate of its structure and function. In *Proceedings of AI-ED 2001 Workshop*, San Antonio, Texas.

Ortony, A., Clore, G.L., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge: Cambridge University Press.

Picard, R. (1997). *Affective computing*. London: The MIT Press.

Silverman, B.G. (2001). More realistic human behavior models for agents in virtual worlds: Emotion, stress and value ontologies, Technical Report, University of Pennsylvania.

Silverman, B.G., Cornwell, J.B,. & O'Brien, K. (2003). Progress to date on the human performance moderator function server (PMFserv) for rapidly generating reusable agents, forces and crowds, Technical Report, University of Pennsylvania.

Smith, C.A., & Ellsworth, P.C. (1985). Attitudes and social cognition. *Journal of Personality and Social Psychology*, *48*(4), 813-838.

Smith, C.A., & Kirby, L.D. (Eds.) (2000). *Consequences require antecedents: Towards a process model of emotion elicitation. Feeling and thinking: The role of affect in social cognition*. London: Cambridge University Press.

Velasquez, J.D. (1999). From affect programs to higher cognitive emotions: An emotion-based control approach.

Proceedings of Workshop on Emotion-Based Agent Architectures, Seattle, USA, (pp.10-15).

KEY TERMS

Affective Computing: A domain of computer science research combining fields such as Artificial Intelligence and Human Computer Interaction in the endeavor to develop computers that can recognize, express and have emotions.

Cognitive Appraisal Theory: The categorization of emotions through the evaluation of stimuli.

Emotion: A subjective response, accompanied by a physiological and behavioral change in an individual.

Emotion Blending: The determination of a final emotion being the result of several emotional states being experienced simultaneously an individual.

Emotion Synthesis: The process of representing emotions in computing devices using cognitive appraisal or other emotion classification theories.

Emotional State Decay: The rate at which the valence of an emotion decreases with time.

Valence: The strength of an experienced emotion.

Enhanced Knowledge Warehouse

Krzysztof Węcel
The Poznan University of Economics, Poland

Witold Abramowicz
The Poznan University of Economics, Poland

Pawel Jan Kalczynski
University of Toledo, USA

INTRODUCTION

Enhanced knowledge warehouse (eKW) is an extension of the enhanced data warehouse (eDW) system (Abramowicz, 2002). eKW is a Web services-based system that allows the automatic filtering of information from the Web to the data warehouse and automatic retrieval through the data warehouse. Web services technology extends eKW beyond the organization. It makes the system open and allows utilization of external software components, thus enabling the creation of distributed applications.

The enhanced knowledge warehouse is an add-in to the existing data warehouse solutions that offers a possibility of extending legacy data-based information resources with processed information acquired from Web sources of business information.

eKW is characterized by transparent filtering and transparent retrieval. The system automatically acquires interesting documents from Web and internal sources and enables business users, who view data warehouse reports, to access these documents without the necessity of formulating any document retrieval queries. Instead, the system takes advantages of meta-information stored in business metadata and data warehouse reports, and builds and runs appropriate queries itself. The system also has an event-alerting capability.

eKW, by design, is a part of the semantic Web. The main role of eKW in the semantic Web is mediation between the world (external information) and internal information systems utilized within the organization. One of the most popular ways of conforming to semantics is to use ontologies, which provide a means for defining the concepts of the exchanged data. eKW is a kind of data mediator that employs ontologies as a conceptualization layer (W'cel, 2003).

BACKGROUND

Document management systems may be divided into two major categories: information retrieval (IR) and information filtering (IF) systems.

The classical models of information retrieval systems were defined by Salton (1983). As distinct from structured data management, information retrieval is imprecise and incomplete. This is due to the inaccurate representations of document contents and user information needs. The distinctive feature of information retrieval systems is a relatively constant document collection. The collection stores documents and their representations (indices). When a user submits a query to the information retrieval system, the query is compared against all indices in the collection. Documents whose indices match the query are returned as the resulting subset of the collection. Most search engines on the Web and digital libraries available on the market can serve as examples of information retrieval systems (Baeza-Yates, 1999).

The main objective of information filtering systems is to block unwanted information from the incoming stream of documents. As distinct from IR systems, filters do not have any fixed collection of documents. Instead, they hold a collection of standing queries (profiles). The profiles represent long-term information interests of their owners. When a new document arrives at the filter its representation is created and compared against all profiles in the collection. Some libraries maintain systems that inform users about new volumes included in the library. Such systems are usually based on the server-side filtering architectures and their profiles usually consist of semi-structured elements (e.g., author, subject, title). This idea is referred to as selective dissemination of information (SDI).

JUSTIFICATION

Possessing and maintaining large information resources does not itself provide any guarantees that users will manage to find the piece of information (document) they need. First of all, users must be aware that the particular piece of information already exists in their resources. Yet, even if they are aware, searching may be very labor consuming. Such a situation is highly undesirable for organizations, as they do not exploit their potentials to increase the effects. That is why knowledge management is nowadays considered the capability of re-use of information and is becoming an increasingly vital issue.

Usually, users are not eager to learn how to operate in several different information systems. Integration of information resources enables the exploitation of the capabilities of many systems through a single user interface (UI). Data warehouse users should be capable of finding interesting documents through a single UI. Not only should the users be aware that the desired information does exist, but also relevant documents should be disseminated to them mechanically. Such a mechanical distribution of the relevant information gives more time for other tasks.

Business users have certain problems with finding information sources, building correct retrieval queries and formulating proper filtering profiles. Thus, introducing the system capable of relieving the users from the necessity of formulating queries and solving the most commonly reported problems with accessing information on the Web would increase productivity of those seeking relevant external information.

The constantly growing number of content providers and the exploding volume of business information are not accompanied by the corresponding growth of capabilities of exploitation of the resources by the contemporary organizations. Therefore, automatic acquisition, organization and presentation of information became not only possible but also essential for the performance of today's businesses.

THE IDEA OF eKW

The term *data warehouse* was defined by Bill Inmon in 1992 as *a collection of integrated, subject-oriented databases to support the DSS function, where each unit of data is relevant to some moment in time*. Since then a rapid growth of this relatively new idea has been observed.

In the eKW model, Web documents are assumed to be business news published in English by major content providers on the Web (e.g., Reuters, Cnn.com), because this type of external information is the one in favor of business people (Abramowicz, 2000; Webber, 1999). The solution proposed is based on the existing standards in the area of data warehousing, document management, communication among software agents, internetworking and storage.

The basic idea of data warehousing is to create a data model common for the whole organization (metamodel). The model provides a framework for uploading legacy data, stored in heterogeneous databases across the organization, to the warehouse. Before data are uploaded they need transformation that includes filtering, consistency check, field mapping and aggregation.

Data warehouses proved to be useful for the purpose of legacy data analysis and they are commonly implemented by organizations. They provided novel data processing techniques like data mining, basket analysis, dimensional data analysis, drill-down, drill-through or slice-and-dice. These techniques are based on the characteristic features of data and information stored in the data warehouse: non-volatility, relevance to some moment in time (timeliness), correctness, consistency, completeness, business subject orientation, and business descriptions in metadata (Adelman, 1997; Kimball, 1996).

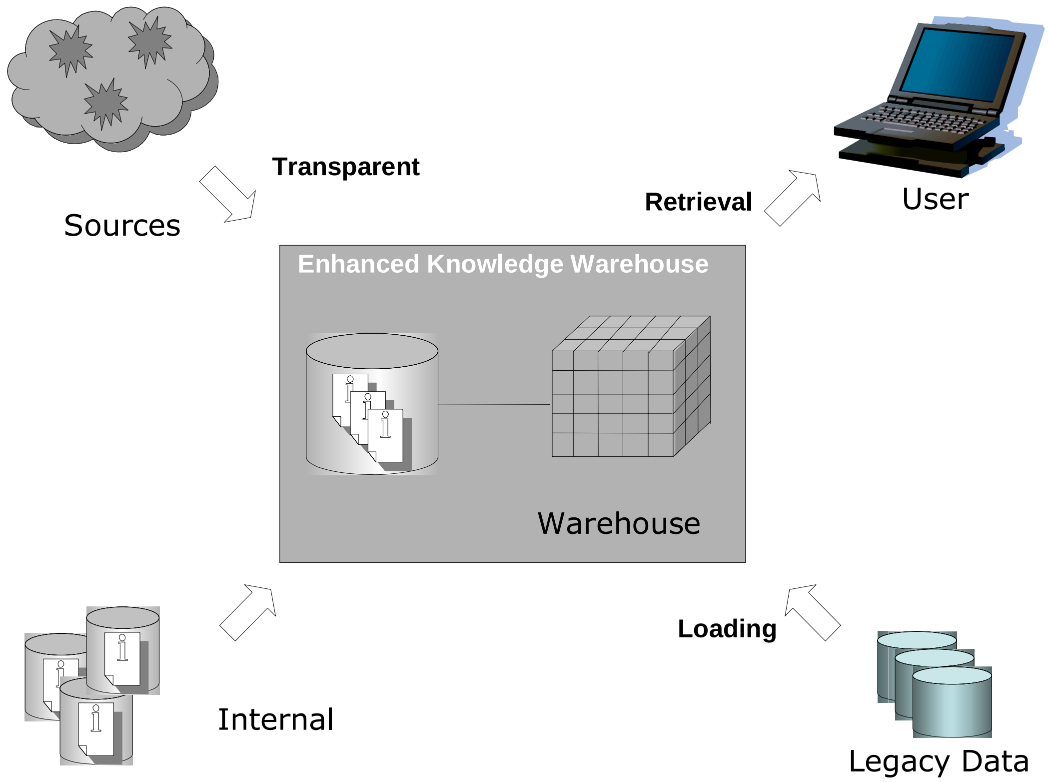
In terms of the data warehouse, metadata1 (warehouse metadata) are data about data. Metadata consist of facts and events concerning database objects. Metadata are usually stored in a database referred to as the metadata repository or the metadata collection. Nowadays, metadata are considered to be an integral part of the data warehouse (Gleason, 1997; Inmon, 1999).

Due to its destination and the sophisticated methods of data processing, the enterprise data warehouse is often claimed to be the corporate knowledge repository. However, we argue that information derived only from structured information (data) produced by the organization is just a fraction of corporate knowledge that may be stored in the repository.

As distinct from database systems, document management systems deal with unstructured and semi-structured information. Because machines are not capable of understanding text, they must rely on mathematical representations of the document content and user information needs.

The most distinctive feature of eKW is that information acquisition, organization and dissemination are performed without any additional actions taken by data warehouse users. In particular, no formulation of keyword-based queries is necessary to access external business news relevant to user information needs. Instead, the queries are formulated mechanically based on business metadata and information stored in the enterprise

Figure 1. Overview of the enhanced knowledge warehouse



data warehouse. The system mechanically collects external documents that are potentially interesting to users and stores them in the local collection. Business users who access corporate data are mechanically supplied with documents that match both the context of the data accessed and user interests at the same time. In this way users have easy access to relevant information resources acquired from outside the organization. These resources are being continuously extended and integrated with the data warehouse. In this way enhanced knowledge warehouse provides a framework for more effective (faster and based on more relevant information) decision making in the organization.

Based on the previous experience with augmenting documents and data, we decided to create a model that would relieve business users from the necessity of connecting to diverse and volatile sources on the Web, formulating queries, and browsing through the results. In order to do this, our system must be capable of mechanical information acquisition from Web sources, adding appropriate descriptions (indices) and precise information retrieval. The first two features imply a continuous series of actions, while the latter requires a trigger that makes the retrieval part of the system run.

The four arrows in Figure 1 represent main processes in the system. The process of loading legacy data to the data warehouse is a typical warehouse loading process defined by the organization. The process of loading internal documents is based on the principles of loading a digital library. In other words, there are two processes that distinguish the enhanced data warehouse model: transparent filtering (continuous) and transparent retrieval (spontaneous). The prefix "transparent" indicates that both processes are see-through for business users.

The enhanced knowledge warehouse is composed of the following Web services that can act independently:

a) Library Service, an implementation of data warehouse library
b) Profiling Service, responsible for data warehouse profiling
c) Filtering Service, responsible for retrieving business information from the Web utilizing filtering machine
d) Indexing Service, responsible for organizing the documents in DWL (Indexing parser)
e) Reporting Service, the front office of the system, responsible for delivering the results to the users.

FUTURE TRENDS

Although both document management and data warehousing systems are designed to supply business users with relevant information that is necessary to solve business problems, they are usually implemented and used separately. Most often, there is no system to store, organize and disseminate information published on the Web. This forces users, who need a broader view of business, to switch between data warehouse applications and content providers on the Web. The latter are usually overloaded with information, require navigation and formulating queries, entering passwords and flood users with commercials. Researchers report that the two most wanted attributes of external information sources are speed and convenience (Webber, 1999), whereas most sources cause frustration and discourage business users from taking any advantages of information available on the Web.

Therefore, the main future trend will be the integration of documents and data. In the eKW model, the problem of consistency of warehouse data and Web documents is addressed.

CONCLUSIONS

Enhanced knowledge warehouse model has certain advantages for business users over non-integrated solutions: instant/extended search, cross-organizational knowledge exchange, personalized event-alerting, and continuous information acquisition.

Because the amount of information produced by content providers on the Web is growing, it is impossible for a human user to process it in the traditional way. Web searches cause frustration, rarely lead to success and waste precious time. eKW was designed to help business people extend their knowledge about the organization by supplying them automatically with documents acquired from Web sources, relevant to their current context and information needs. Transparent filtering function mechanically collects documents potentially interesting to the organization, describes them with commonly accepted terms, and organizes them according to the existing business metadata. Transparent retrieval function supplies data warehouse users with documents relevant to their contexts and their personal interests in business information. The contexts are mechanically extracted from the state of the currently browsed report and personal needs are based on implicit feedback.

The enhanced knowledge warehouse system supplies business users with information based on the legacy data extended by information acquired from the business environment. This gives additional possibilities of business data analysis and provides a framework for better understanding of business processes related to the organization.

REFERENCES

Abramowicz, W., Kalczynski, P., & Węcel, K. (2000). Information filters supplying data warehouses with benchmarking information. In W. Abramowicz & J. Zurada (Eds.), *Knowledge discovery for business information systems* (pp. 1-28). Kluwer Academic Publishers.

Abramowicz, W., Kalczynski, P.J., & Węcel, K (2001a). Profiling the data warehouse for information filtering. *Proceedings of IRMA 2001 International Conference,* Toronto (pp. 810-814).

Abramowicz, W., Kalczynski, P.J., & Węcel, K (2001b). Time consistency among structured and unstructured contents in the data warehouse. *Proceedings of IRMA 2001 International Conference,* Toronto (pp. 815-818).

Abramowicz, W., Kalczynski, P.J., & Węcel, K (2001c). Common semantic layer to support integration of the data Warehouse and the Web. *Human Computer Interaction – HCI 2001*, Sopot, Poland.

Abramowicz, W., Kalczynski, P., & Węcel, K. (2001d, November 27-29). Information ants to filter business information from Internet sources. *Proceedings of the CAINE-2001 Conference*, Las Vegas (pp. 134-137).

Abramowicz, W., Kalczynski, P., & Węcel, K. (2002). *Filtering the Web to feed data warehouses*. London: Springer-Verlag.

Adelman, S. (1997). Data quality. In J. Bischoff & T. Alexander (Eds.), *Data warehouse – practical advice from the experts* (pp. 122-134). Prentice Hall Inc.

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval.* NY: Addison-Wesley, ACM Press.

Gleason, D. (1997). Metadata. In J. Bischoff & T. Alexander (Eds.), *Data warehouse – practical advice from the experts* (pp. 135-150). Prentice Hall Inc.

Inmon, W. (1999). *Metadata in the data warehouse environment.* Bill Inmon's Library.

Kalczynski, P. (2002). *Software agents to filter business information from the Internet to the data warehouse.* PhD dissertation. The Poznan University of Economics, Poland.

Kalczynski, P., Abramowicz, W., Węcel, K., & Kaczmarek, T. (2003). Time indexer: A tool for extracting temporal references from business news. *2003 Information Resource Management Association International Conference* (pp. 832-835).

Kimball, R. (1996). *The data warehouse toolkit.* NY: John Willey & Sons, Inc.

Llido, D., Berlanga, R., & Aramburu, M. (2001). Extracting temporal references to assign document event-time periods. In H. Mayr et al. (Eds.), *DEXA 2001 Conference Proceedings* (pp. 62-71). Berlin: Springer-Verlag.

Salton, G., & McGill, M. (1983). *Introduction to modern information retrieval.* McGraw-Hill Book Company

Webber, S., Allcock, S., Plenty, A., & Yeates, R. (1999). Business information and the Internet: Use of the Internet as an information resource for small and medium-sized enterprises: Final report. *British Library Research and Innovation Report, 136.*

Węcel, K. (2002). *Profiling the data warehouse for business information filtering.* PhD dissertation. The Poznan University of Economics, Poland.

Węcel, K. (2003). Towards an ontological representation of knowledge on the Web. In W. Abramowicz (Ed.), *Knowledge-based information retrieval and filtering.* Kluwer Academic Publishers

E

KEY TERMS

Common Semantic Layer: The ontology for eKW is implemented with common semantic layer (CSL). CSL is a relatively stable set of concepts (key words and phrases) with hierarchical and non-hierarchical relations utilized by the organization to describe information. A typical CSL consists mostly of proper names (e.g., products, shares, vendors or competitors). It is a foundation for building structures that facilitate the access to information resources (Abramowicz, 2001c).

Consistency of Warehouse Data and Web Documents: In order to introduce consistency among documents and data, data warehouse library must enjoy data warehouse features. Hence, documents are never removed from DWL; all documents are properly described with metadata and linked to data warehouse objects. There are four levels of consistency between DWL and DW that enable the eDW system to build retrieval queries: subject consistency, temporal consistency, semantic consistency and personalization. These four levels together provide constraints for sub-setting DWL and producing relatively small ranked lists of relevant documents associated with data warehouse reports.

Context Query: Context query represents the short-term user information needs. It is created each time a business user launches a data warehouse report. The context query consists of three parts: subject constraints, time constraints and semantic constraints. Subject constraints define the warehouse objects that are part of the report. Time constraints are represented by the actual time range grasped by the most current state of the report. Semantic constraints are represented as a set of CSL-based weighted keywords. The context query is executed on the data warehouse library and personalized results are immediately returned to the user.

Data Warehouse Library: The data warehouse library (DWL) is a repository of documents acquired from Web sources or located within the organizational intranet.

Apart from the contents, the repository stores additional metadata about the documents, mainly temporal and semantic indices. Basically, DWL is an extensive digital library (DL) and enjoys all capabilities of a typical information retrieval system. However, there are three distinctive features that make DWL something more than a typical DL: temporal indexing, CSL-based indices and direct links to warehouse business metadata. All these features provide novel possibilities of sub-setting the collection of documents to be searched and establishing links among documents and data.

Data Warehouse Profiling: Data warehouse profiling is a way of discovering the information needs of the data warehouse users. It is the first stage of the transparent filtering. The profiling action is triggered by a change in business metadata. The process results in a set of warehouse profiles that represent long-term information needs of data warehouse objects logically represented in metadata. It is assumed that warehouse profiles are transitive; that is, they represent information needs of users who utilize the objects while looking for information (Abramowicz, 2001a; W'cel, 2002).

Filtering Machine: Filtering is performed autonomously by a set of specialized source agents dispatched by the agent-based filter. The filter has access to the current warehouse profiles. It stores the list of sources subscribed to by the organization with their properties (Abramowicz, 2001). We introduced a new automatic navigation technique: information ants. The ant-like agents carry the warehouse profiles. New documents are parsed, indexed and matched with all profiles respectively. If a document matches the profile, it is downloaded and re-indexed in order to store it in DWL.

Indexing Parser: Because of the heterogeneous nature of Web documents, we argue that indexing must also be performed at the stage of parsing Web documents. Hence, each source agent has a built-in indexing parser capable of building a document representation (index) based on the internal structure and the content of a given Web document. Typically the hierarchy of tags is taken into account while building the document index. Another important part of the process is taking into account source-specific lists of stop words, thesauri and ontologies. Each resulting index consists of a weighted set of terms that occurred in the analyzed document. This format is a standard vector space model format that is commonly utilized in document filtering and retrieval.

Temporal Indexing: Extracting temporal references from documents (Llido, 2001), also referred to as temporal indexing or time indexing, aims at describing the content of textual documents with a set of time ranges (Abramowicz, 2001a; Kalczynski, 2002). The time ranges may be weighted. Time index is based not only on firm dates extracted from the content, but also on weaker references, like "next Monday," "last month" or "today," and inexplicit references such as tense operators "will," "was" or "is". Weaker references are resolved according to the date the document was published (reference date). The final outcome of temporal indexing is the time index associated with the document. The index consists of a finite number of non-overlapping weighted time ranges. Detailed description of the time indexer applied in the eKW system is given in Kalczynski (2003).

Transparent Filtering: Transparent filtering is a continuous process, transparent for business users, in which documents are acquired from Web sources and stored in the data warehouse library with appropriate descriptions. The principle of filtering is that documents are analyzed only once; hence only new documents are processed by the system. Transparent filtering can be further divided into profiling and filtering. Those two sub-processes are mutually independent.

Transparent Retrieval: Transparent retrieval is a spontaneous process triggered by data warehouse users each time they launch a data warehouse report. In this process the data warehouse library is mechanically searched for documents that match the current context of a given data warehouse report. The context is extracted from the report metadata and data. The ranked list of matching documents is compared against the user profile and presented to the business user as an extension of structured information displayed on the report.

ENDNOTE

[1] Greek meta means going beyond.

Enhancing Competitiveness of B2B and B2C E-Commerce

Henry Aigbedo
Oakland Unversity, Michigan, USA

INTRODUCTION

A framework that is often used to describe the transformation of society relates to the percentage of people involved in or impacted by a given paradigm at some point in time. Even though some of these paradigms co-exist at a given point in time, one of them tends to dominate and exert considerable influence on the others. Several authors have described such evolution in different contexts. For example, Fitzsimmons and Fitzsimmons (2004) suggest that societies have undergone the following transformation, one leading to the next in succession: agrarian, manufacturing, service, and information. Since its introduction a little over a decade ago, the Internet, which exemplifies the leap into the information age, has had a profound impact on many areas of human endeavor, indeed, far more than had been imagined or anticipated. This impact on society is apparent through its widespread use in government, education, medicine, engineering, as well as in business. Considering the number of people involved and the dollar amount of sales transacted over the Internet, it seems reasonable to infer that it has had the greatest impact in business during the past millennium. The Internet continues to play a significant role in making enterprises to be more competitive.

The progression of the information age is underscored by projections about its growth and impact on the economies of many countries. For example, Forrester and Gartner research groups project unprecedented growth in Internet-related business transactions globally in the coming years. In particular, Gartner estimated that worldwide B2B e-commerce will surpass 7.3 trillion dollars by 2004 (Noyce, 2002). It is interesting to observe that the order of growth in electronic transactions is similar to the projected growth in computer processing speed. Further, technological advancements in the computer and communication industries have led to reduced prices of computers and Internet service, respectively. These developments are playing a significant role in facilitating the use of the Internet for commerce. For example, several polls indicate that more than half of Americans own personal computers. While not all people who own computers have Internet connectivity at home, many gain access through secondary means such as public libraries and work computers. University of California Los Angeles (UCLA) Internet-related studies indicate that over 70% of Americans are online (Lebo & Wolpert, 2004; UCLA, 2001). Also, Barnes (2003) reports that, as of fall 2002, estimates for people around the world who use the Internet were about 600 million, about 10 percent of the world population.

This article discusses how firms are using Business-to-Business (B2B) and Business-to-Consumer (B2C) e-commerce strategies to gain competitive advantage. Several related initiatives that firms can avail themselves of in order to be more effective will also be briefly examined. The article is organized as follows. First, we provide some background by describing some issues that have been identified in the literature concerning the subject. Next, we discuss the importance of B2B and B2C e-commerce to firms and their clientele. We then examine the future of B2B and B2C e-commerce initiatives as potent weapons and how they should be deployed to foster success. Finally, we summarize and conclude with our final thoughts on B2B and B2C e-commerce.

BACKGROUND

Several models of electronic commerce have been developed to represent or formalize the various practices of various entities (Aigbedo, 2004; Barua, Prabhudev, Winston, & Yin, 2001; Mahadevan, 2000). Also, Kaplan and Sawhney (2000) present a two-by-two matrix that describes activities among businesses along two dimensions: “What businesses buy” (operating inputs or manufacturing inputs) and “How businesses buy” (systematic sourcing or spot sourcing). Under this classification scheme, Internet exchanges such as e-STEEL fall within the manufacturing inputs/spot sourcing quadrant.

From a definitional perspective, electronic commerce includes Internet-based and non-Internet-based means of transmitting business information. However, we restrict our discussion to Internet-based e-commerce because of its ubiquity and potential for growth in the coming years. The principal users of Internet-based e-commerce can be classified as Consumers, Businesses, and Governments, with interactions occurring among pairs of these classes or between a class and its like. The two major areas that have experienced the most growth and that are projected to capture a significant proportion of Internet use for commerce are Business-to-Business (B2B) and Business-to-Consumer (B2C) e-commerce. Now let us see two examples of these transactions. When an Original Equipment Manufacturer places orders for parts from its suppliers through the supplier's Web site, the parties are engaged in B2B e-commerce. On the other hand, when a consumer purchases a book through an online bookstore such as Barnes and Nobles' Web site, the interaction between this customer and the company is B2C e-commerce.

Amazon.com was one of the firms that blazed the trail in e-commerce, in general, and B2C e-commerce, in particular, by selling books through its Web site. Since its inception, it has expanded its offerings to include a wide variety of items such as compact disks, electronic appliances, apparel, and accessories. Following Amazon's success, many "brick-and-mortar" firms have either fully or partially transferred their operations online. For example, Wal-Mart, the retail giant, now also offers for sale through its Web site many items that it sells through its conventional stores (Wal-Mart, 2004) Also, the Internet has greatly facilitated Dell Computer's direct business model, which entails sale of computers directly to consumers, as the company generates a significant proportion of its sales through its Web site. In 1999, Dell Computer earned over 1 billion dollars through its B2C e-commerce initiative (Chopra & Van Mieghem, 2000).

So far, B2C e-commerce appears to have had the greatest impact from the point of view of the number of people directly affected. However, as Kenjale and Phatak (2003), for example, observe, B2B e-commerce has a greater potential as far as number of transactions and dollar volumes are concerned. A case in point is DaimlerChrysler's 2001 purchase of highly engineered parts for two future car models through Covisint for more than 3 billion dollars under a B2B reverse auction framework (Konicki, 2001).

From the foregoing discussions, it would seem that it is only well-known, large corporations that are engaged in e-commerce. However, this is not the case, as many companies, including small ones, are transforming part or all of their operations to the virtual environment. Chen, Haney, Pandzik, Spigarelli, and Jesseman (2003) report on the initiatives taken by a small firm in implementing e-commerce. They also describe some critical success factors that need to be followed by such companies. These include gradual transformation into e-commerce (especially considering cost implications and unavailability of technical expertise in small firms), proper allocation of resources, outsourcing of critical startup services, modification of business practices, collection and analysis of customer-related data to help monitor buying patterns, thus facilitating promotions and better service offerings.

Just as understanding consumer behavior has been the focus of traditional marketing research for decades, it is extremely important to understand it in the context of e-commerce as well. By synthesizing several published work in this respect, Saeed, Hwang, and Yi (2003) proposed an integrative framework for understanding the basic factors that influence customer behavior in e-commerce environments. They identified four predictor variables: system quality (relates primarily to web design and technological features that facilitate the customer's shopping experience), information quality (relates primarily to content accuracy, relevance, and reliability), service quality (relates to the customer's perceptions of the experience), and vendor and channel characteristics (e.g., size and reputation).

B2B E-COMMERCE INITIATIVES

Companies are finding the Internet a very useful tool for facilitating transactions among themselves. It is gradually replacing traditional methods such as facsimile and Electronic Data Interchange (EDI). Although B2C applications spurred the use of the Internet for commerce, a far greater volume of transactions occurs among companies. There is much more to transactions among companies than simply buying and selling, which are central to B2C applications. Customer relationship management, exchange of product and production-related information, and collaboration in product design and development are some of the areas where B2B e-commerce is deployed. More and more companies now understand that they have reached their limits in terms of improvements in performance and profits that they can achieve by themselves. Therefore, in recent years we have seen increased emphasis on supply chain management, as companies seek to work more closely with their partners on the downstream side of the chain (towards the customer) and the upstream side of the chain (towards the parts suppliers). In the retail industry, for example, Wal-Mart's tremendous success in pursuing its low-cost strategy is due largely to its supply chain initiatives that provide for effective exchange of information among its distribution centers, the manufacturers, and its stores, in relaying real-time demand information. Dell Computer has also achieved similar success by building powerful links with its suppliers, thus creating what is generally referred to as the virtual supply chain, characterized by virtual integration between Dell and its partners. Based on the results of a study of ten companies in Australia, Power and Sohal (2002) discuss the firms' perspectives on the benefits of e-commerce to their operations. Tangible benefits included reduction of finished goods and work-in-process inventories, increased flexibility, and overall cost reductions. Some intangible benefits were also identified that included the ability to promote improvement in various operations with their business partners.

In recent years, there has been a move in certain industries to form Internet exchanges – virtual locations on the Internet where buyers and sellers meet to exchange information and carry out transactions. A classic example is Covisint, the automotive exchange formed a couple of years ago by the "Big Three" US automotive companies and Nissan, Renault, Commerce One, and Oracle. In addition to the fact that this arrangement provides visibility to both buyers and sellers, it has facilitated auctions and reverse auctions of products or capacity. As noted above, DaimlerChrysler spent more than 3 billion dollars to purchase parts for two future car models through Covisint. (Konicki, 2001). Industry experts are increasingly recognizing how important it is for exchanges to provide sustainable value propositions to their members—value that transcends only technological offerings. Indeed, exchanges that have thrived so far (e.g., Pantelos Group for the utility and energy industry in the United States and Worldwide Retail Exchange) have put mechanisms in place that facilitate this laudable objective (Ulfelder, 2004). Volkswagen is also following suit with the exchange it has established to streamline processes with its numerous suppliers (Hofmann, Sloan, & Malykhina, 2004).

As of now, there is no general agreement about how these exchanges will evolve over the next couple of years. For example, Wise and Morrison (2000) had argued that competitive bidding practice, which places emphasis only on price, does not account for many other important factors such as quality, delivery speed, and customization, in certain manufacturing environments. Furthermore, such arrangements often offer little benefits to the suppliers (Aigbedo & Tanniru, 2002; Wise & Morison, 2000). Talluri and Ragatz (2004) discuss various types of auction arrangements that can be used to facilitate B2B e-commerce, and they propose an Analytical Hierarchy Process (AHP) framework for practical application of the concepts when auction is based on multiple attributes such as price, quality, and flexibility. On the other hand, Wyld (2002) used recent data compiled by the Institute of Supply Management (ISM) to analyze the state of industry in the application of the Internet for procurement. He observes that there has been steady growth, especially in the use of the Internet to procure indirect goods and services. Other areas of use emerging from that study included identification of new suppliers and collaboration with suppliers.

Although not all companies use the Internet as the medium to coordinate with their partners, the enhanced information flow that the Internet provides has several benefits. This includes reduced cost of modification of product and pricing information, since a central system such as an exchange can make the same information available to all parties at the same time without the need for duplication. Furthermore, the hub-and-spoke connection that a central system provides reduces linkage costs, as only one link needs to be made each time a new member is included in the supply network. This will result in savings in setup and maintenance costs. Increased visibility that is enabled by the Internet is very crucial for environments where product mixes change very rapidly; examples include fashion goods, electronic equipment such as computers, and automobiles. The Internet helps substitute information for inventory, thus reducing operational costs. This visibility also reduces or eliminates the "bullwhip effect" (upstream amplification of demand requirements) in supply chains. The bullwhip phenomenon, which has begun to gain attention lately, results mainly from information asymmetry among companies in the successive stages of the supply chain.

B2C E-COMMERCE INITIATIVES

The airline industry is a major sector that has seen tremendous growth in B2C e-commerce. Most would-be passengers now purchase airline tickets online. Some airlines offer better prices or provide incentives to customers purchasing tickets online as compared to those purchasing through their sales representatives (Southwest Airlines, 2004). Recently, airlines such as Northwest Airlines have even gone as far as making it possible for customers to print their boarding passes from the airline's Web site, thus saving on check-in time at the airport. Other sectors that have experienced the benefits of B2C e-commerce include the hospitality industry (Carroll & Siguaw, 2003), the mail-delivery industry such as Federal Express, United Parcel Service, and the United

States Postal Service, the print media (Rose, 2003), and the retail industry (Karp, 2003; Merrick, 2003).

B2C e-commerce provides value propositions to customers as well as the companies engaged in it. Among other things, it offers customers the opportunity of comparing features and prices and purchasing items from the comfort of their own homes. In addition to vast savings that can be realized in operational costs, this provides a company with the opportunity to reach a very large customer base. Furthermore, it reduces errors that could arise from multiple handling of data and provides a convenient way of aggregating and analyzing customer-related information. One novel way companies started using e-commerce initiatives to improve their profits is working with online companies to aggregate service offerings in the form of packages that would be more attractive to customers. For example, Marriot Hotel chains recently entered into such an agreement with Expedia and Hotels.com (Binkley, 2004).

The above examples may seem to suggest that e-commerce initiatives are always successful. However, this is not the case as shown by the failure and subsequent bankruptcy of Webvan (an online grocer) in 2001. Among other things, Starr (2003) analyzed the operations of Webvan over a three-year period, 1998-2000, and sought to understand the reasons for Webvan's failure and whether or not it could have been averted. He observes that Webvan relied too much on its technological prowess without adequately assessing operational factors, including inventory-related issues for their short-lifespan products. This result corroborates Porter's (2001) assertion that it is extremely important for firms to perform adequate groundwork, as the use of the Internet does not obviate the need for sound business models.

THE FUTURE OF B2B AND B2C E-COMMERCE

Just as it is with the adoption of any new technology or method of operation, e-commerce has had its challenges, failures, and successes. Since 2000, we have seen the demise of many dot.com companies that proved to be unviable. Many others are learning from mistakes they made in the past, and others are waxing stronger by appropriately positioning themselves and adopting strategies that are consistent with the dynamic nature of the business environment.

Further, one important issue that needs to be borne in mind is that the Internet and the e-commerce technology that comes along with it are simply tools. Porter (2001) aptly describes this issue by observing that moving over to the Internet to transact business should not replace sound business principles. In other words, doing so would simply amount to transferring the problems of the system into a virtual environment. In a similar light, Kendall (2003) discusses the importance of planning in making e-commerce initiatives successful. This can be especially important when one understands that the Internet, by virtue of its ubiquity, essentially lowers the barrier to entry for firms in many industries. Thus, while it is true that the Internet potentially provides some operational advantages, the key for any firm is not just the use of the Internet, but how to use it in innovative ways, thus creating distinctive competency.

The above no doubt underscores the need for firms to spend a considerable amount of time in developing suitable business models that will yield significant value propositions and that are operationally effective and efficient before implementing them. Adequate market research to assess potential for viability, as well as performing appropriate simulation studies, where applicable, will certainly prove to be very invaluable in this respect.

One of the key issues that will increasingly become important in B2B e-commerce is effective coordination among various firms in the supply chain. Coordination here means information sharing and visibility relating to the spectrum of individual and inter-firm activities. While there have been advances in this area over the past couple of years, with firms such as Dell Computer leading the way with virtual integration, there is still a lot that can be done in many other industries such as the automotive industry. Information can be used to replace inventory thus enabling the firms in the supply chain to maintain the appropriate amounts of inventory at the appropriate locations. A major area of difficulty in some industries, especially the automotive industry, is that of trust among the partners. In some cases, shared information has been used in a detrimental way by partnering firms to leverage their competitive position or when asking for price concessions. Still along this line, the formation of consortium exchanges by firms in certain industries (such as Covisint in the automotive industry) is useful. However, there is need for them to be more proactive in articulating the value that all participants (especially low leverage participants such as suppliers) stand to gain. Otherwise, many supplier firms will be reluctant to participate, thus negatively impacting the benefit to the entire supply chain.

More than ever, quality management will play a very vital role in the success of B2C e-commerce in the coming years. For example, amidst the many firms providing similar products over the Internet and the numerous brick-and-mortar competitors, a company will need to pay particular attention to on-time delivery and e-fulfillment.

Although the customer is aware that there is need for trade-off between convenience and delivery time offered by an e-tailer (retailer who operates using the Web and Internet technology), a company needs to make the necessary efforts to ensure that the difference is not so much as to outweigh the advantages. Particular attention should be given to building partnerships with efficient logistic firms and designing effective inventory management systems. Depending on the circumstances, it may even be more cost-effective for a firm to partner with local stores as outlets, requiring the customer to "complete the last mile" of the transaction by going to pick up the items (Lee & Whang, 2001). This phenomenon is sometimes referred to as a "click-and-mortar" framework.

Another important issue is how an e-tailer presents information to would-be customers on its Web site. Just as proper layout of facilities is important in a brick-and-mortar setting, it is essential that would-be customers have little difficulty locating the items they want to purchase. Furthermore, procedures relating to purchase transactions (such as shipping costs and return policies) need to be clearly indicated and appropriately placed on the Web site, otherwise it could lead to customer dissatisfaction, which will have serious negative impact on the firm.

The process of tracking customer or prospective customer behavior in e-commerce transactions has continued to receive attention (Moe & Fader, 2001). Not all customers who visit a company's Web site or check on a product eventually make purchases. Effectively forecasting demand for a given product or related products has continued to be a daunting endeavor. Development of sophisticated data-mining techniques, coupled with their integration with forecasting techniques, should lead to better forecasting, thus enabling a firm to respond faster, without having to hold excessive amounts of items in inventory.

B2B and B2C companies need to be increasingly vigilant in the area of Internet security, as recent events has shown substantial losses incurred by many firms due to the activities of miscreants. Investments in appropriate technology, proper training of personnel, and constant monitoring of these systems will go a long way in preserving them so they can perform as desired. Heightened interest in the development and deployment of intrusion-detection devices for addressing these issues is an important step in this direction (Richmond, 2002).

CONCLUSION

The significant growth that has been experienced in B2B and B2C e-commerce in recent years suggests great potential for continued growth in the coming years. Many companies have now introduced, to varying degrees, the use of the Internet for commerce, and several studies have indicated that many more are following suit. However, it is important for managers to recognize that the Internet is primarily a means to facilitate business transactions, and by no means does it obviate the need to develop and operate models that are based on sound business principles. In the B2B e-commerce arena, special attention needs to be accorded to effective coordination across the supply chain, appropriately replacing inventory with information in ways that will lead to better value for the entire supply chain. Firms involved in B2C e-commerce should especially pay attention to quality of services to their customers, especially as it pertains to fulfillment of their needs. Furthermore, the issue of Internet security needs to be given high priority, as it appears to pose the most serious threat to success of e-commerce operations.

REFERENCES

Aigbedo, H. (2004). Managing operations in the e-commerce era: requirements and challenges. In J. Gupta and S. Sharma (Eds.), *Intelligent enterprises for the 21st century*. Hershey, PA: Idea Group Publishing.

Aigbedo, H. & Tanniru, M. (2002). Electronic markets in support of procurement processes along the automotive supply chain. Research Paper, Oakland University School of Business Administration, Rochester, MI.

Barnes, S. (2003). Location based services: The state of the art. *E-service Journal*, *3*(2), 59-70.

Barua, A., Prabhudev, K., Winston, A., & Yin, F. (2001). Driving e-business excellence. *Sloan Management Review*, *43*(1), 36-44.

Binkley, C. (2004). Marriott signs deal with web sites; Expedia, hotels.com gain access to more hotel rooms but chain gets to set prices. *Wall Street Journal* (Eastern edition), February 4, D3.

Carroll, B. & Siguaw, J. (2003). The evolution of electronic distribution: Effects on hotels and intermediaries. *Cornell Hotel and Restaurant Administration Quarterly*, *44*(4), 38-50.

Chen, L., Haney, S., Pandzik, A., Spigarelli, J., & Jesseman, C. (2003). Small business internet commerce: A case study. *Information Resources Management Journal*, *16*(3), 17-31.

Chopra, S. & Van Mieghem, J. (2000). Which e-business is right for your supply chain? *Supply Chain Management Review*, July/August, 32-41.

Fitzsimmons, J. & Fitzsimmons, M. (2004). *Service management: Operations, strategy and information technology* (4th ed.). New York: McGraw Hill.

Hofmann, M., Sloan, E., & Malykhina, E. (2004). VW revs its B2B engine. *Optimize,* March, 22-26.

Kaplan, S. & Sawhney, M. (2000). E-hubs: The new B2B marketplaces. *Harvard Business Review*, *78*(3), 97-103.

Karp, J. (2003). From bricks to clicks: A Brazilian discount retailer finds success where so many others have failed: E-commerce. *Wall Street Journal* (Eastern Edition), September 22, R7.

Kendall, K. (2003). Rescuing e-commerce or e-commerce to the rescue? *Information Resources Management Journal*, *16*(3), 1-3.

Kenjale, K. & Phatak, A. (2003). B2B exchanges: How to move forward from here, *World Trade*, *16*(6), June, 26-28.

Konicki, S. (2001). Covisint's big deal: Daimler spends $3B in four days. *Information Week*, *34*(828), 34-34.

Lebo, H. & Wolpert, S. (2004). First release of findings from the UCLA World Internet Project shows significant 'digital gender gap' in many countries, January 14, *http://newsroom.ucla.edu/page.asp?RelNum=4849&menu=fullsearchresults*

Lee, H. & Whang, S. (2001). Winning the last mile of e-commerce. *Sloan Management Review*, *42*(4), 54-62.

Mahadevan, B. (2000). Business models for internet-based e-commerce: An anatomy. *California Management Review*, *42*(4), 55-69.

Merrick, A. (2003). Sears to sell clothing on its Web site. *Wall Street Journal* (Eastern Edition), September 26, B2.

Moe, W. & Fader, P. (2001). Uncovering patterns in cybershopping. *California Management Review*, *43*(4), 106-117.

Noyce, D. (2002). eB2B: Analysis of business-to-business e-commerce and how research can adapt to meet future challenges. *International Journal of Market Research*, *44*(1), 71-95

Plotnikoff, D. (2001). Number of Americans online grows to 72.3 percent. *Arizona Daily Star*. Retrieved from: *http://www.azstarnet.com/public/startech/archive/120501/wire2.html*

Porter, M. (2001). Strategy and the Internet. *Harvard Business Review*, *79*(3), 63-78.

Power, D. & Sohal, A. (2002). Implementation and usage of electronic commerce in managing the supply chain: A comparative study of ten Australian companies. *Benchmarking: An International Journal*, *9*(2), 190-208.

Richmond, R. (2002). E-commerce: B2B—security—beating the hackers: Martin Roesch is a pioneer in creating software designed to thwart computer attacks; but can he make a business of it? *Wall Street Journal* (Eastern Edition), June 10, R12.

Rose, M. (2003) Wall Street Journal circulation rises 16% with web additions. *Wall Street Journal* (Eastern Edition), November 4, B2.

Saeed, K., Hwang, Y., & Yi, M. (2003). Toward an integrative framework for online consumer behavior research: A meta-analysis approach. *Journal of End User Computing*, *15*(4), 1-26.

Southwest Airlines (2004). Southwest to extend bonus credit to fliers booking via Web site. *Wall Street Journal* (Eastern Edition), Mar 17, 1.

Starr, M. (2003) Application of POM to e-business: B2C e-shopping. *International Journal of Operations and Production Management*, *23*(1), 105-124.

Talluri, S. & Ragatz, G. (2004). Multi-attribute reverse auctions in B2B exchanges: A framework for design and implementation. *The Journal of Supply Chain Management*, Winter, *40*(1), 52-60.

UCLA (2001). The UCLA Internet Report 2001: Surveying the digital future, Year 2, UCLA Center for Communication Policy (p.17). *http://ccp.ucla.edu/pdf/UCLA-Internet-Report-2001.pdf*

Ulfelder, S. (2004). B2B survivors. *ComputerWorld*, *38*(5), February 2, 27-28.

Wal-Mart (2004). Wal-Mart opens online music store: Undercutting competitors. *Wall Street Journal* (Eastern Edition), March 24, D2.

Wise, R. & Morrison, D. (2000). Beyond the exchange: The future of B2B. *Harvard Business Review*, *78*(6), 86-96.

Wyld, D. (2002) The electric company: How the supply chain is being reinvented through the rapid application of e-procurement processes in the business-to-business arena. *Management Research News*, *25*(2), 22-53.

KEY TERMS

B2B E-Commerce: Business-to-Business (B2B) e-commerce is transaction that occurs between and among firms that are related to the procurements of goods and/or services through electronic medium. The typical medium of transaction is the Internet and World Wide Web.

B2C E-Commerce: Business-to-Consumer (B2C) e-commerce is transaction that occurs between a firm and individual consumer or user of the firm's goods and/or services through electronic medium. The typical medium of transaction is the Internet and World Wide Web.

Brick-and-Mortar: Operations that take place in conventional off-line settings as opposed to those that take place online. For example, offering items for sale at a conventional store as opposed to selling them online.

Bullwhip Effect: Demand amplification from its source across the supply chain. This is largely caused by information asymmetry among the entities in the supply chain.

Click-and-Mortar: A firm that operates both online and off-line or a hybrid operation that requires a combination of online and off-line aspects to complete a transaction. For example, a company offers items for sale online and requires that the customers pick up the items from a nearby store.

Consortium Exchange: A group formed by companies in the same industry, bringing their supply chains together for the purpose of facilitating transactions among themselves over the Internet.

E-Fulfillment: A company meets the needs of its customers with whom it carries out some transactions through the Internet. It is most often associated with B2C e-commerce.

Internet Security: The phenomenon by which a company assures that its data and entire Internet infrastructure are protected from damage caused by miscreants through activities such as hacking.

Original Equipment Manufacturer (OEM): A company that manufactures the main equipment or product that is used by the final consumer. An OEM can be considered as the core or focal organization that bridges the supply side and the demand side of the full supply chain linking all the way from the raw materials to the final consumer.

Supply Chain: A network of companies interacting among themselves with the ultimate aim of providing needed goods and/or services to the final consumer. Three main items that flow among the entities are goods/services, information, and funds.

Virtual Supply Chain: A supply chain with little or no ownership bond, but so closely linked together and operations very well coordinated as though the various entities in the chain were all part of the same company.

ENDNOTE

[1] Revised Manuscript (April/June 2004).

Enhancing Workplaces with Constructive Online Recreation

Jo Ann Oravec
University of Wisconsin-Whitewater, USA

INTRODUCTION

Organizations have become more permeable— integrating more influences from the outside world— as participants engage in such online diversions as trading stocks, engaging in multiplayer games, or viewing images of their children in daycare. Ready availability of these activities has brought the potential for abuse but also new opportunities. Constructive uses of online recreation and play can enhance many workplaces (especially high-tech and information-saturated ones) and perhaps ultimately make them more productive. This article proposes that these complex issues be resolved through participatory approaches, involving workgroups in discussions as to what constitutes "constructive recreation" as well as in development and dissemination of effective and fair policies. This discourse can also ultimately increase levels of trust among team members and between employees and management.

BACKGROUND

Issues concerning the boundaries between work and play have provided continuing struggles for managers and employees. Workplaces have become more "porous" and permeable— integrating more influences from the outside world— as individuals engage in such online diversions as trading stocks, playing games, or viewing images of their children in daycare. Everyday workplace life is becoming more diverse and chaotic. Although many organizational roles today demand high levels of creativity and mental flexibility, they can also fail to provide the means through which individuals can gain fresh perspectives. In the "information age," playful, exploratory, and spontaneous interaction can also facilitate the exchange of ideas for tackling workplace problems. Managers who expect employees not to use the Internet for some amount of off-task activity severely misjudge the nature of workplace life— which is solidly infused in online interaction. Depriving employees of opportunities for Internet recreation in some cases excludes the possibility of nearly any form of diversion from assigned responsibilities.

Workplace use of the Internet for activities that are not directly authorized by management is often considered as the "theft" of human and computer resources, while construed as a just reward by employees (Lim, 2002). Even though many managers consider the personal use of the Internet as an ethical lapse (Greengard, 2000), the "moral high ground" concerning these issues is not entirely clear. Much of the rhetoric and advertising copy associated with workplace computing incorporates recreational imageries and motifs, which can send misleading signals to employees. A number of individuals have already had significant experience combining work with online recreation; convincing them that hard work cannot be combined with online play is thus a tough sell. Telecommuters returning to organizational settings are often not entrusted with the autonomy to engage in online breaks at appropriate times— latitude they take for granted when doing the same tasks in their home offices. Many young people became comfortable with computing through video games and online interpersonal interaction and took online breaks during their demanding college studies (Colkin & George, 2002). Individuals must find ways to cope psychologically with increased pressures on the job (Weil & Rosen, 1997) and management should explore creative but feasible ways to assist them in these efforts.

Wireless Internet applications add more complexities, further increasing the porousness of organizations and making employees' access to recreation less dependent on systems controlled by their managers. Daniels (2000) reports how wireless technologies (such as PDAs with Internet access) are used within meetings to amuse and distract participants, often resulting in productivity losses. Since wireless technologies are still in the early stages of adoption in many organizational contexts, placing severe restrictions on their use (and penalties for misuse) could be counter-productive. Personal computers became familiar workplace additions in the 1980s in part because of their use for gaming, an activity that encouraged employees of a variety of ages and backgrounds to explore the various dimensions of the devices and to become more comfortable with them.

If engaged in constructively, online recreation can aid in awakening creativity and increasing wellbeing, just as appropriate and timely face-to-face diversions have restored employees' energies over the past decades. However, some individuals may not be able to deal with online recreation constructively. They indeed will use it in ways that affect their organizations and themselves negatively, just as some individuals cannot perform adequately on the job for other reasons. Forms of "positive discipline" can be utilized if employees choose to exceed reasonable, agreed-upon limits; implementing such discipline "requires that the supervisor and employee work together to correct the problem behavior" (Guffey & Helms, 2001). Managers and employees should strive together to harness online recreation toward positive ends, rather than condemning or seeking to stifle it completely.

WHAT IS "CONSTRUCTIVE RECREATION"?

Online recreation has already served many supportive purposes in organizations; games can be used to help decrease computer anxiety as well as encourage experimentation and the early stages of learning (Kendall & Webster, 1997; Oravec, 1999; Webster & Martocchio, 1992). What would make online recreation optimally beneficial to individuals, project teams, and the organization as a whole? To start the discussion: recreation is "constructive" when it is in synch with pending work responsibilities, allowing individuals to use time not consumed by workplace demands in ways that equip them to face future tasks with greater energy and expanded perspectives. Constructive recreation is also in keeping with technological constraints, as exemplified by the organizations that allow online recreation but place limits during certain hours to avoid system overload (Verton, 2000). Policies established are developed in participatory ways, and are disseminated broadly. Constructing ways of assigning tasks and evaluating employees so that significant and meaningful measures of productivity are involved can lessen an emphasis on the "surface" behavior of employees. Other characteristics of constructive recreation initiatives include:

- ***fostering flexibility:*** A major impetus behind constructive recreation initiatives is facilitating the rapid adaptation of individuals to changing circumstances. Constructive recreation affords individuals the means to maintain their flexibility in workplace environments that place increasing demands on their capacities to withstand change.
- ***manifesting sensitivity to cultural concerns:*** Workplace recreation is also "constructive" to the extent in which it is responsive to the overall culture of the organization and sensitive to the needs and values of other organizational participants (including freedom from harassment). Requirements of project team members in terms of scheduling are especially critical to recognize since the synchronization and sustained involvement of everyone are required during critical periods.
- ***providing stimulation and refreshment:*** Along with its other aspects, recreation is constructive if it provides intellectual and psychological stimulation or support, the sustenance often needed to take on tough challenges. "Reclaimed moments" that individuals spend in such activity can allow them to reestablish senses of control in otherwise stressful and constraining contexts. Ability to access such recreation and thus momentarily escape can provide a safety valve for those who face unyielding situations or put in long work hours, thus putting the porousness of today's Internet-supported workplaces to good use.

FUTURE TRENDS

The value of recreation and play in adult realms is not well understood. Play has been given an assortment of definitions in the academic and research literatures (with examinations in the fields of social psychology, philosophy, and anthropology); it is often considered in both its adult and child modes as a "cognitive and symbolic act that is fundamental to the human representational process" (Myers, 1999). Across species as well as cultures, play has been shown to help individuals prepare for the unexpected by presenting varying streams of novel or challenging situations (Spinka, 2001). Play is generally considered as a support for children's intellectual and social development, but its role in adult lives is less clear. Research initiatives on what kinds of recreation and play are most efficacious in different workplace environments—as well as on individual and group "play styles"—could enlighten constructive recreation efforts (although they cannot be expected to provide definitive results).

Simulation is indeed an aspect of play that has some direct implications for employee readiness in the workplace, and it has received some research treatment (Myers, 1999). Michael Schrage's (1999) *Serious Play* examines how simulations expand the intellectual capacities of knowledge workers; forms of online play may equip individuals to utilize an organization's "serious" computer simulations more effectively, thus reinforcing skills applicable in many workplace contexts. Many powerful simulation games with societal or political themes are widely available to the public and have considerable audiences;

the Sims series and other popular single- and multiplayer games have been used to entertain and educate in a variety of contexts (Pillay, Brownlee & Wilss, 1999).

Constructive recreation initiatives will also be a part of many organizational efforts to build cohesion. Managers have often used organizationally sanctioned recreation as a perquisite, a bonus for acceptable conduct. It has served as an extension of the workplace, providing a form of "social capital" (part of the "glue" that holds the at-work community together). Through the past century, many organizations have sponsored picnics and celebrations with the strategy of increasing workplace cohesion (Putnam, 2000).

As employees (including many white collar as well as knowledge workers) telecommute or put in long and irregular hours, the adhesive that binds organizations has been increasingly conveyed through electronic channels. However, it is unclear what kinds of online activity can foster social capital (Uslaner, 2000). Just as human resource experts struggled early in the twentieth century to integrate face-to-face recreation into workplace contexts, organizations should attempt similar feats in online realms, thus making online recreation a shared and open resource rather than a secretive endeavor. Unlike many early human relations experiments, the recreational activities involved should be developed in a participatory (rather than patriarchal) fashion. Whether organization-approved fantasy football, discussion group and collaborative filtering forums, joke-of-the-day contests, or other recreations are ultimately successful will depend on how they fit into everyday working experiences.

CONCLUSION

Can we indeed construct a "level playing field"? As workplaces have evolved, so have the issues that have divided employers and managers. Conflict has ensued for decades on an assortment of matters relating to the quality of worklife, often leading to dysfunctional confrontations. Today, employees who guess wrong about online recreation standards— or choose to violate them— often pay large penalties, even being demoted or fired. Some managers have devised negative sanctions for these infringements far more severe than those applied to comparable face-to-face interaction. Office workers paging through paper catalogues in idle minutes rarely face the harsh penalties that those caught shopping online often encounter.

Hard-line positions against forms of online recreation may be required in some instances and directly related to important organizational goals. For instance, air traffic controllers should be expected to keep focused on landing real airplanes rather than escape into fantasy games during assigned hours. However, some hard-line restrictions can reflect fear or lack of understanding of online realms. Management may assume that online recreation will foster or encourage Internet addiction or related concerns. "Internet addiction" has become a widely identified syndrome, although its medical underpinnings are still in question (Beard, 2002; Oravec, 1996, 2000).

Ambiguities concerning online work and play in virtual realms are increasingly adding complexities to these issues (Broadfoot, 2001). It is often difficult to tell which Web sites are related to business needs and which are recreational; many have dual purposes, combining amusement with news and other serious pursuits. Slashdot.org has humorous material as well as valuable technical commentary, and abcnews.com has stories on upcoming movies as well as current economic results. Helpful intelligent agents (some with cartoon-like manifestations) can add levity to everyday tasks. Surfing the Internet for an answer to a question or fiddling with various programs can interfere with productive effort, as individuals dwell on technological nuances. Managers and employees need to deal not only with recreational concerns but also with broader issues of how to integrate computing into workplaces in ways that are engaging yet productive. However, online recreation should not be exploited as a means to keep individuals glued to workstations for indefinite periods in lieu of reasonable work schedules and functional work-life balances.

Solutions as to how to couple online work and play are emerging in organizations that are tailored to specific workplace contexts. Managers and employees are gaining important experience in resolving these issues as individuals perform activities away from direct supervision via mobile computing or virtual office configurations. Managers are learning how to perform their functions without direct employee surveillance. Employees are learning higher levels of self-discipline and the skills of balancing on-line work and play— just as they have learned to balance face-to-face schmoozing with task orientation in the physical world. Thus setting severe restrictions on online recreation can serve to slow down the process of understanding how to migrate the organization into virtual realms and establish trust. Responsibility and respect for others in these realms can be difficult to acquire, and many employees will indeed need direction.

Allowing for reasonable and humane amounts of online recreation can indeed have considerable advantages, both for the individuals involved and the organization as a whole. It can serve to open blocked creative channels and possibly relieve stress as well. Online recreation can also extend the limits of individuals' working days by providing extra dimensions to workplace activity.

Rather than going through the emotional labor of looking busy, employees can utilize spare moments on the job in recharging their mental batteries. Constructive use of recreation will require a number of changes, such as increases in managerial flexibility and employee empowerment (Boswell, Moynihan, Roehling & Cavanaugh, 2001; Kanter, 2002). Organizational participants must learn how to handle the distractions and opportunities of increasingly porous workplaces, with their many external influences. Education and training can be useful in these initiatives: novice employees can be aided to couple work and recreation in ways that increase overall effectiveness. Constructive recreation strategies can bring these complex matters into the open, rather than allow them to be objects of rumor and fear.

Forms of online diversion are already becoming integral elements of everyday workplace life, often serving to humanize and enhance organizations. Negotiation and discourse on constructive recreation issues can increase mutual trust and respect concerning online as well as face-to-face activity. With effort on everyone's part, the constructive use of online recreation can help the entire organization work harder and play harder.

REFERENCES

Beard, K. (2002). Internet addiction: Current status and implications for employees. *Journal of Employment Counseling, 39*(1), 2-12.

Boswell, W., Moynihan, L., Roehling, M., & Cavanaugh, M. (2001). Responsibilities in the 'new employment relationship': An empirical test of an assumed phenomenon. *Journal of Managerial Issues, 13*(3), 307-328.

Broadfoot, K. (2001). When the cat's away, do the mice play? Control/autonomy in the virtual workplace. *Management Communication Quarterly*, *15*(1), 110-115.

Colkin, E., & George, T. (2002, March 25). Teens skilled in technology will shape IT's future. *InformationWeek, 881,* 72-73.

Daniels, C. (2000, October 30). How to goof off at your next meeting. *Fortune, 142*(10), 289-290.

Greengard, S. (2000). The high cost of cyberslacking. *Workforce, 79*(12), 22-23.

Guffey, C., & Helms, M. (2001). Effective employee discipline: A case of the Internal Revenue Service. *Public Personnel Management, 30*(1), 111-128.

Kanter, R. (2002). Improvisational theater. *MIT Sloan Management Review, 43*(2), 76-82.

Kendall, J., & Webster, J. (1997). Computers and playfulness: Humorous, cognitive, and social playfulness in real and virtual workplaces— introduction to the special issue. *DATA BASE, 28*(2), 40-42.

Lim, V. (2002). The IT way of loafing on the job: Cyberloafing, neutralizing and organizational justice. *Journal of Organizational Behavior, 23*(5), 675-694.

Myers, G. (1999). Simulation, gaming, and the simulative. *Simulation & Gaming*, *30*(4), 482-490.

Oravec, J. (1996). *Virtual individuals, virtual groups: Human dimensions of groupware and computer networking.* New York: Cambridge University Press.

Oravec, J. (1999). Working hard and playing hard: Constructive uses of on-line recreation. *Journal of General Management, 24*(3), 77-89.

Oravec, J. (2000). Internet and computer technology hazards: Perspectives for family counselling. *British Journal of Guidance and Counselling, 28*(3), 309-324.

Pillay, H., Brownlee, J., & Wilss, L. (1999). Cognition and recreational computer games: Implications for educational technology. *Journal of Research on Computing in Education, 32*(1), 203-217.

Putnam, R. (2000). *Bowling alone: The collapse and revival of American community*. New York: Simon & Schuster.

Schrage, M. (1999). *Serious play*. Cambridge, MA: Harvard Business School.

Spinka, M. (2001). Mammalian play: Training for the unexpected. *Quarterly Review of Biology*, *76*(2), 141-169.

Uslaner, E. (2000). Social capital and the net. *Communications of the ACM, 43*(12), 60-64.

Verton, D. (2000, December 18). Employers OK with e-surfing. *Computerworld*, *34*(51), 1-2.

Webster, J., & Martocchio, J. (1992). Microcomputer playfulness: Development of a measure with workplace implications. *MIS Quarterly*, *16*(2), 201-226.

Weil, M., & Rosen, L. (1997). *TechnoStress: Coping with technology @ work @ home @ play*. New York: John Wiley & Sons.

KEY TERMS

Flexible Workplace: Organizational settings that can quickly take external and internal changes into account in their processes.

Internet Addiction: Use of the Internet and network resources that undermines the fulfillment of some of an individual's basic human needs.

Organizational Policies: Openly-stated, officially-sanctioned rules for organizational resource usage and other kinds of organization-related conduct.

Participatory Management: Management in which the input of employees as well as managers is thoughtfully taken into account in setting organizational policies and developing organizational structures.

Play: Activities in which individuals and groups engage that stimulate various aspects of personal and social functioning without necessarily being related to particular utilitarian outcomes.

Simulation Games: Games in which important aspects of a system are modeled so that game participants can engage in activities and deal with events that are comparable to those that system participants would encounter.

Social Capital: Social closeness, mutual knowledge, and cohesion that are a product of a wide assortment of different kinds of informal, volunteer, and partially-structured social interactions.

Enterprise Resource Planning and Integration

Karl Kurbel
European University Viadrina Frankfurt (Oder), Germany

INTRODUCTION

Enterprise resource planning (ERP) is a state-of-the-art approach to running organizations with the help of comprehensive information systems, providing support for key business processes and more general, for electronic business (e-business). ERP has evolved from earlier approaches, in particular, materials requirement planning (MRP) and manufacturing resource planning (called MRP II) in the 1980s.

The focus of MRP and MRP II was on manufacturing firms. The essential problem that MRP attacked was to determine suitable quantities of all parts and materials needed to produce a given master production schedule (also called a "production program"), plus the dates and times when those quantities had to be available. Application packages for MRP have been available from the 1960s on. In the beginning, they were mostly provided by hardware vendors like IBM, Honeywell Bull, Digital Equipment, Siemens, etc.

MRP was later expanded to *closed-loop MRP* to include capacity planning, shop floor control, and purchasing, because as Oliver Wight (1884) puts it: "Knowing what material was needed was fine, but if the capacity wasn't available, the proper material couldn't be produced" (p. 48).

The next step in the evolution was *MRP II (manufacturing resource planning)*. According to the father of MRP II, Oliver Wight, top management involvement in the planning is indispensable. Therefore, MRP II expands closed-loop MRP "to include the financial numbers that management needs to run the business and a simulation capability" (Wight, 1984, p. 54).

Enterprise resource planning (ERP) has its roots in the earlier MRP II concepts, but it extends those concepts substantially into two directions. ERP takes into account that other types of enterprises than those producing physical goods need comprehensive information system (IS) support as well, and even in the manufacturing industry, there are more areas than those directly related to the production of goods that are critical for the success of a business.

BACKGROUND OF ERP: THE NEED FOR INTEGRATION

The key issue of ERP is integration (Langenwalter, 1999). Whereas stand-alone solutions—sometimes sophisticated information systems for various areas of a business—have been available before, ERP takes an holistic approach. Instead of isolated views—on procurement, on manufacturing, on sales and distribution, on accounting, etc.—the focus is now on integrating those functional areas (Scheer & Habermann, 2000).

The need for integrated systems has been recognized by many, but Germany-based SAP AG was the first to put them into reality. SAP's early success as worldwide market leader comes largely from the fact that this company actually designed and implemented business-wide integrated information systems.

The lack of integration of information systems has created a variety of problems. The most serious ones among those problems are the following:

- Redundancy, i.e., the same information is stored and maintained several times.
- Inconsistency, i.e., information about the same entity stored in different places is not the same.
- Lack of integrity, i.e., databases where such information is stored are not correct.

Mistakes, wrong decisions, and additional work are some of the consequences resulting from these problems. Consider, for example, data about customers. Such data are often entered and maintained in a sales and distribution information system (customer orders), then again in the dispatching system (delivery orders), and perhaps once more in a financial accounting system (invoices). Not only is this redundant and mean additional work, but also the same attributes may even stand for different things. For example, an "address" field in the sales and distribution system may represent the address of the customer's procurement department, whereas "address" in the dispatching system is the place where the goods have to be delivered.

Integration of information systems can be considered from several perspectives: from the data, the functions, the operations, the processes, the methods, and the software perspectives. The most important aspects are data integration, operations integration, process integration, and software integration:

- Integration of *data* means that data models and databases are unified so that all departments of an enterprise use the same data entities, with the same values.
- Integration of *operations* requires connecting individual operations, or steps of a business process, with preceding or succeeding operations, respectively.
- Integration of *processes* means that interfaces between different business processes are explicitly considered (e.g., connections between order processing and flow of material control).
- Integration of *software* means that different programs, e.g., information systems for different business functions, can run together and use each other's data and operations.

Those aspects of integration have always been considered important requirements for effective business information processing, but how does one actually obtain enterprise-wide integrated information systems?

Because most organizations have been using information systems in various business areas for quite some time, one way is to integrate those stand-alone systems subsequently. This approach has been discussed and practiced under the concept of "software reengineering," often related to the term "legacy systems" for the information systems to be integrated (Miller, 1997; Seacord et al., 2003).

The other approach to obtain integrated information systems is obviously to start developing them from scratch. In such a situation, information structures can be modeled and designed on the drawing board in an enterprise-wide manner, at least in theory. Practical experiences have shown that developing comprehensive information systems for all areas of a business is a giant task. That is why such systems have rarely been developed as individual solutions. Not only is the investment needed very high, but also manpower and know-how to develop such systems are often beyond the means of a single company. Therefore, comprehensive integrated information systems have mostly been developed by dedicated software and consulting companies. In the 1970s and 1980s, those systems were named with rather general terms, like standard packages or integrated business information systems, until the terms "enterprise resource planning" and "ERP system" emerged in the 1990s. In fact, the term "enterprise resource planning" has been coined by the software industry and not by academia.

Today there is a common understanding of what the term stands for. The definition used in this chapter is as follows: An *enterprise resource planning system (ERP system)* is a comprehensive information system that collects, processes, and provides information about all parts of an enterprise, automating business processes and business rules within and across business functions partly or completely.

Alternatively, an ERP system may be defined as a set of integrated information systems rather than as one system. This depends on the perspective of the viewer. For the user, an ideal ERP system will behave like *one* enterprise-wide information system, with one database, and one common user interface. Nevertheless, such a system may be composed of many subsystems and many databases, as long as they are well integrated.

COMPONENTS OF AN ERP SYSTEM

Horizontal and Vertical Views of Enterprise Resource Planning

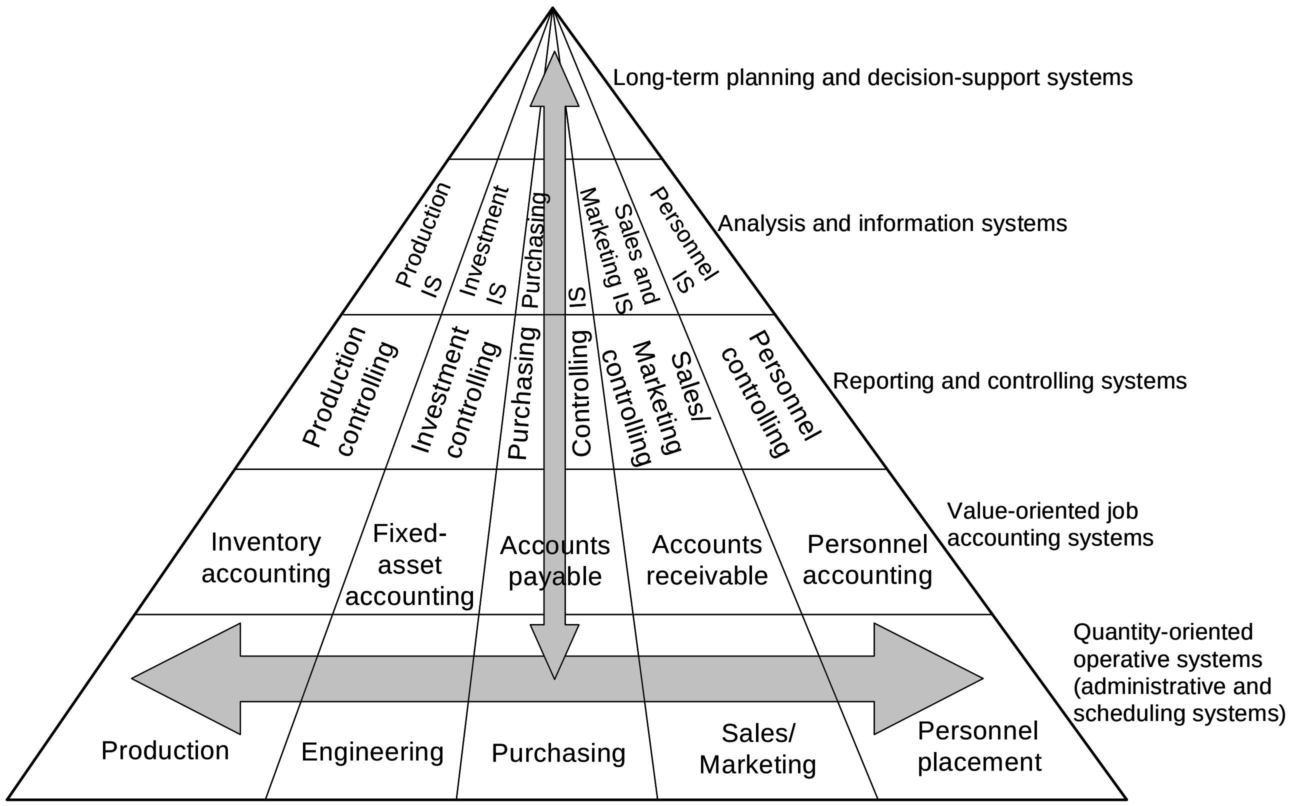
An ERP system integrates information, processes, functions, and people into one coherent system (Brady et al., 2001). Such a system supports all horizontal business functions and all vertical levels of a business (operational, tactical, and strategic). Figure 1 illustrates this view in a simplified information systems pyramid. Each component may be seen as a functional subsystem. In a horizontal perspective, systems are integrated along the value chain. The vertical direction asks for integration of operative systems with their corresponding value-oriented accounting systems; reporting and controlling systems; analysis and management information systems; and long-term planning and decision support systems (Scheer, 1994, p. 5).

A typical ERP system provides components like the ones shown in Figure 1, arranged and extended in one way or another. As an example of integrated information systems, the mySAP ERP system is described subsequently.

An Example: mySAP ERP

mySAP ERP is the market leader's current ERP system. This system encompasses all levels of the pyramid. It is structured into the following modules (short descriptions are taken from SAP, 2003, 2004):

Figure 1. Integrated business information systems (Source: Scheer, 1994, p. 5)



Analytics

- Strategic enterprise management: Supports the top level of the pyramid in Figure 1: integrated strategic planning, performance monitoring, business consolidation, and stakeholder communication; provides tools for planning and executing the strategies: balanced scorecard, value-based management, financial statement planning, risk management, investment planning, and more.
- Business analytics: Supports managers with methods and tools for financial and management reporting, financial planning, budgeting and forecasting, profitability management, product and service cost management, overhead cost management, working capital and cash-flow management, etc.; provides analytical functions for procurement, inventory and warehouse management, manufacturing, transport, sales, customer service, quality management, enterprise asset management, program and project management, and more. [In Figure 1, those functions, tools, and methods belong mostly to levels 2 (analysis and information systems) and 3 (reporting and controlling information systems).]

Figure 2. Application domains and modules of mySAP ERP (Source: SAP, 2004)

Analytics	Strategic Enterprise Management	Financial Analytics	Operations Analytics	Workforce Analytics	
Financials	Financial Accounting	Management Accounting	Financial Supply Chain Management	Corporate Governance	
Human Capital Management	Employee Life-Cycle Management	Employee Transaction Management	HCM Service Delivery	Workforce Deployment	
Operations: Value Generation	Procurement	Inventory & Warehouse Management	Manufacturing	Sales Order Management	Customer Service
Operations: Support	Life-Cycle Data Management	Program & Project Management	Quality Management	Enterprise Asset Management	
Corporate Services	Travel Management	Environment, Health & Safety	Incentive & Commission Management	Real Estate Management	

Financials

The Financials module supports several application areas that are located first on level 4 of the pyramid:

- Financial accounting: Processing of incoming and outgoing payments, cash flows; provides general ledger, accounts receivable, accounts payable, fixed assets accounting, inventory accounting, tax accounting, financial statements, and more; helps to monitor financial transactions; supports business analysis through combined planning, reporting, and analysis of competitive measures.
- Managerial accounting: Provides profit center accounting, cost center and internal order accounting, project accounting, product cost accounting, profitability accounting; supports investment management, revenue and cost planning, transfer pricing, etc.
- Financial supply chain management: Supports financial collaboration within the enterprise and its business networks; provides credit management, cash and liquidity management, treasury and risk management, and more.
- Manager self-service: Provides managers with access to all relevant business information as well as related services.

Human Resources

The Human Resources module is comprehensive, including functionalities of the operative, administrative level (level 5) and the value-oriented level (level 4).

- Employee life cycle management: Supports recruiting and talent management, performance management, compensation management for various modes

(e.g., performance- and competency-based pay)
- Employee transaction management: Provides the central repository for employee data; integrates the information with other SAP business applications, especially Financials and Operations; supports time and attendance processing (planning, managing, and evaluating the working times and activities of internal and external employees); handles working-time provisions determined by companies themselves, by standard agreements, or required by law; handles all payroll processes, supports current legal regulations and collective agreement specifications, and ensures compliance with regulatory changes
- Workforce deployment: Provides project resource planning, resource and program management (i.e., resource management, project portfolio management, project execution, and skills management), and specific solutions for retail personnel and call centers
- HCM (human capital management) self-services: Provides self-services for managers and employees

Operations

This module supports mainly the bottom level through quantity-oriented operative subsystems for daily operations, including support for planning and execution.

- Purchase order management: Provides conversion from demands to purchase orders, issuance, and confirmation of purchase orders; supports purchasing of materials and services (for example, subcontracting for components).
- Inventory and warehouse management: Comprises inbound processing (all the steps of an external procurement process that occur when the goods are received) and outbound processing (all steps to prepare and ship goods to their destination), warehousing and storing (warehouse-internal movements and storage of materials); manages physical inventory for the company's own stocks (periodic, continous, etc.).
- Manufacturing planning: Provides typical MRP (materials requirement planning) functionality: computing quantities and due dates for production orders and purchase requisitions through lead-time scheduling, depending on buffers, operation times, lot-sizing rules and so on.
- Manufacturing execution: Supports the process of capturing actual production information from the shop floor to support production control and costing processes; supports a variety of concepts: make-to-order, repetitive manufacturing, flow manufacturing, shop-floor manufacturing, lean manufacturing, process manufacturing, and batch manufacturing.
- Transportation: Provides transportation planning (routing, carrier selection, etc.) and execution (shipment orders), freight costing, and legal services.
- Sales order management: Supports quotation and order management (creating and processing orders, including pricing and scheduling orders for fulfillment) including inquiries and follow-up orders; provides mobile-sales, billing, and contract-management functionalities.
- Customer service: Helps to manage a customer's product configuration (installation and configuration management, including definition of the product hierarchy, management of serial numbers, measurements, document management, and engineering change management), service contracts, planned services, warranties, and so on.
- Life cycle data management: Provides document management, product structure management (including bills of materials), recipe management, integration of CAD (computer-aided design), PDM (product data management), and GIS (geographical information system) data, and more.
- Quality management: Supports quality engineering according to the ISO 9000 standard, quality assurance and control (quality inspections, statistical process control, traceability, etc.), quality improvement (audit management according to ISO 19011, problem/complaint management, corrective and preventive action, etc.).
- Project management: Provides functions for project planning (project structures, costs, budgets, workforce and resource planning, scheduling activities, etc.), project execution (monitoring project progress, progress analysis/earned value analysis, progress tracking, etc.).

Corporate Services

The Corporate Services module provides comprehensive support for resource-intensive corporate functions:

- Real estate management: Provides tools to support real-estate property acquisition and disposal, property portfolios, functions to help users lease and manage the real estate portfolio, etc.
- Incentive and commission management: Helps to design incentive compensation plans, calculate vari-

able compensation (e.g., direct sales commissions), carry out evaluations of performance and cost results, and more.
- Travel management: Provides functions for travel request and pre-trip approval, travel planning and online booking, travel expense management, services for mobile staff, and more.
- Environment, health, and safety: Supports a variety of functions for product safety, handling of hazardous substances, transportation of dangerous goods, waste management, etc.; allows companies to take preventive care of their employees' health; schedules medical examinations and testing for workers; manages emissions for air, water, and soil; monitors and controls plant emission sources.

This overview of components of the mySAP ERP system illustrates the wide range of functions that support ERP nowadays. The reader interested in details of ERP functionality is encouraged to study the products' Web sites. Descriptions of the above functions can be found on SAP's Web site (SAP, 2003, 2004). ERP systems by other vendors' ERP systems are outlined below.

THE MARKET FOR ERP

ERP systems have been around for about two decades now. All large and medium-size companies use such systems today, and more small companies are catching up. Well-known ERP systems in use include R/2, R/3, R/3 Enterprise, and mySAP ERP (all by SAP AG, Germany), Oracle's E-Business Suite (by Oracle, USA), iBaan Enterprise (formerly by Baan, Netherlands, now SSA, USA), Movex (by Intentia, Sweden), System21 (by GEAC, Canada), J.D. Edwards (USA), PeopleSoft (USA), Navision (by Microsoft, USA), Business Solutions (by Sage, UK), and infor:COM (by infor, Germany).

However, names and vendors are changing, as there is plenty of dynamic in the ERP market. Mergers and acquisitions and an ongoing market concentration can be observed year by year. Some of the above-mentioned systems will have been renamed because of those processes, or will have gone up in the new parent company's ERP system by the time this article is published.

Although there is still a fairly large number of ERP vendors, a handful of them dominate the ERP market worldwide. The global market leader by far is SAP AG from Walldorf (Germany). Followers are Oracle Corp., PeopleSoft (having completed the acquisition of J.D. Edwards), Sage, and Agilisys (haven taken over infor:COM).

FUTURE TRENDS: EXTENDING AND ENHANCING ERP

While ERP continues to be the core of any integrated business software, the focus has shifted toward advanced user support, like Business Intelligence (Biere, 2003) and Knowledge Management (Davenport et al., 1998; Earl, 2001), and interorganizational support of electronic business, in particular, supply chain management and customer relationship management. As those areas are closely related to ERP, all major vendors have extended their systems to support them as well.

Customer Relationship Management (CRM) is an approach to develop a coherent, integrated view of all relationships a firm maintains with its existing and potential customers (Laudon, 2004, p. 61). Nowadays, many channels are available for enterprises and customers to be in contact with each other: retail stores, telephone, e-mail, electronic shopping on the Web, mobile devices, etc. CRM systems try to consolidate customer information from all those channels and integrate the firm's diverse customer-related processes.

The major focus of ERP is to support the internal business processes of an organization. However, business activities do not end at the limits of one's own company. A natural extension of ERP is, therefore, s*upply chain management* (SCM; Ayers, 2001). SCM looks at the organization's business partners, in particular at the suppliers and their suppliers. In addition, many methodological and technical shortcomings of ERP have been removed or at least improved in SCM. Those improvements have been discussed in the literature under "Advanced Planning and Scheduling" (e.g., Meyr et al., 2002) and were implemented in SCM solutions by SCM vendors.

One example of such improvements is the use of optimization methods, like linear programming, mixed-integer programming, constraint propagation, and heuristics like genetic algorithms to solve production and distribution planning problems. Another example is pegging, i.e., creating, maintaining, and evaluating relationships between purchasing orders, production orders, transportation orders, and customer orders across entire supply networks worldwide.

CONCLUSION

ERP is a comprehensive approach to running organizations with the help of computer-based information systems. Such systems, so-called ERP systems, support all major areas of a business. ERP systems are integrated systems, with respect to information, processes, functions, and people. Usually, those systems are very large, developed by specialized software firms. The worldwide market leader is, by far, Germany-based SAP AG, with its current systems R/3 Enterprise and mySAP ERP. Except for small enterprises, almost all of today's companies use ERP systems for their ongoing business.

Extensions of ERP systems can be observed at the former limits of ERP: in CRM, the focus is on dedicated customer support; the focus of SCM is to help organizations plan worldwide supplier–buyer networks and to act successfully within such networks; business intelligence enables enterprises to behave in an "intelligent" way (e.g., addressing the most promising customers well-aimed); knowledge management helps to formalize and preserve important knowledge in an organization; and information system support for key areas, like the before-mentioned areas, is provided both by ERP vendors and by specialized software and consulting firms.

REFERENCES

Arnold, R. S. (1993). *Software reengineering*. Los Alamitos, CA: IEEE Computer Society Press.

Ayers, J. B. (2001). *Handbook of supply chain management*. Boca Raton, FL: St. Lucie Press.

Biere, M. (2003). *Business intelligence for the enterprise*. Upper Saddle River, NJ: Prentice Hall.

Brady, J., Monk, E. F., & Wagner, B. J. (2001). *Concepts in enterprise resource planning*. Boston, MA: Course Technology.

Davenport, T. H., Delong, D. W., & Beers, M. C. (1998). Successful knowledge management projects. *Sloan Management Review, 39*(2), 43-57.

Earl, M. (2001). Knowledge management strategies: Toward a taxonomy. *Journal of Management Information Systems*, *18*(1), 215–233.

Langenwalter, G. A. (1999). *Enterprise resources planning and beyond: Integrating your entire organization*. Boca Raton, FL: Saint Lucie Press.

Laudon, K. C., & Laudon, J. P. (2004). Management information systems. Managing the digital firm (8th ed.). Upper Saddle River, NJ: Prentice Hall.

Meyr, H., Wagner, M., & Rohde, J. (2002). Structure of advanced planning systems. In H. Stadtler & C. Kilger (Eds.), *Supply chain management and advanced planning* (2nd ed.) (pp. 99–104). New York: Springer.

Miller, H. (1997). *Reengineering Legacy Software Systems*. Woburn, MA: Digital Press.

SAP. (2003, June). mySAP ERP press fact sheet. Retrieved March 16, 2004, from http://www.sap.com/company/press/factsheets/solution/erp.asp

SAP. (2004). mySAP ERP business maps. Retrieved March 16, 2004, from http://www.sap.com/solutions/businessmaps/

Scheer, A. -W. (1994). *Business process engineering—Reference models for industrial companies* (2nd ed.). Berlin: Springer.

Scheer, A. -W., & Habermann, F. (2000). Making ERP a success. *Communications of the ACM, 43*(5), 57–61.

Seacord, R. C., Plakosh, D., & Lewis, G. A. (2003). *Modernizing Legacy Systems: Software technologies, engineering processes, and business practices*. Reading, MA: Addison-Wesley.

Wight, O. W. (1984). *Manufacturing resource planning: MRP II. Unlocking America's productivity potential* (revised ed.). New York: John Wiley & Sons.

KEY TERMS

Data Integration: Unifying data models and databases so that all departments of an enterprise use the same data entities, with the same values.

Enterprise Resource Planning (ERP): The current state-of-the-art approach to running organizations with the help of information systems that provide support for key business processes and, more general, for electronic business (e-business).

ERP System (Enterprise Resource Planning System): A comprehensive information system that collects, processes, and provides information about all parts of an enterprise, automating business processes and business rules within and across business functions, partly or completely.

Operations Integration: Creating logical connections of individual operations, or steps of a business process, with preceding or succeeding operations, respectively.

Process Integration: Defining and automating interfaces between different business processes explicitly.

SAP: Software company based in Walldorf (Germany); market leader in ERP software worldwide. SAP is an abbreviation of the company's German name "Systeme, Anwendungen, Produkte in der Datenverarbeitung" (systems, applications, products in data processing).

Software Integration: Connecting different programs so that they can run together and use each other's data and operations.

Enterprise Resource Planning and Systems Integration

Mark I. Hwang
Central Michigan University, USA

INTRODUCTION

Systems integration has been an important topic ever since businesses started using mainframes to run their back-office operations. These systems specialized in common tasks found in functional areas such as accounts receivable and accounts payable, inventory, purchasing, and ordering. However, getting information from these fragmented systems to get a whole picture view of the business was extremely difficult if not impossible. By integrating back-office operations into a single system, enterprise resource planning (ERP) was supposed to solve that problem. In reality, however, very few companies fully implement all the modules of an ERP package and most continue to rely on legacy systems and other specialized software for their processing needs Integration of ERP with various enterprise applications remains a challenge. Moreover, it is not uncommon to have ERP software from different vendors or multiple copies of ERP software from the same vendor running in the same company. Integrating multiple instances of ERP software will be the predominant ERP project in most large corporations in the next half decade. Finally, many companies pursue merger and acquisition as a major growth strategy. A critical task in merging two companies nowadays involves integrating their ERP systems. This paper addresses these integration issues involving ERP systems. Common tools for integration and success factors for integration projects are discussed.

NEED FOR SYSTEMS INTEGRATION

The trends in globalization, mergers and acquisitions, and the advent of e-commerce and e-business have all contributed to the intensification of the competitive landscape. Companies need to find better ways to interact with their customers and provide better services. More than ever before, customers require information on various relations they have with a business; information about their accounts, their balances, recent purchases, and bills (Slater, 2000a). Traditionally, all that information, even if available, was locked in disparate systems. Consequently, companies that succeeded in integrating their systems enjoyed tremendous competitive advantage and reaped huge rewards in sales and market shares by offering unprecedented customer values. A good example is Dell, who was able to integrate not only its internal but also suppliers' systems. As a result, it has cut down its inventory to four day's worth of supplies, compared with from 30 to 50 days of its competitors. Its integrated supply chain system is a major contributor to its number one position in the worldwide PC market (Hildebrand, 2003).

Systems integration has been an important topic ever since businesses started using mainframes to support their back-office operations (Kumar & Hillegersberg, 2000; Hildebrand, 2003). Companies developed computer systems to automate common tasks in functional areas such as accounts receivable, accounts payable, inventory, purchasing and ordering. However, for the most part, these departmental systems were un-integrated and, therefore, not capable of providing a whole picture view of the business. Integration traditionally was done in a piece meal fashion and required custom coding that was both difficult and expensive. Attempts to develop enterprise-wide integrated systems for the most part have failed (Kumar & Hillegersberg, 2000). Commercial enterprise resource planning (ERP) systems, first appeared in the 1980s, were considered a major solution to the integration problem. However, it was soon found out that ERP could create its own integration problems, as discussed next.

Integration through ERP

ERP represents a major commercial solution that enables companies to integrate business operations across functions. Expanding from their roots in manufacturing and operations, vendors such as SAP and Baan continue to add business processes to their ERP offerings in areas including order management, marketing, purchasing, warehouse management, human resources and finance. In the mid-1990s sales of ERP software got a major boost as corporations rushed to replace their homegrown systems with Y2K compliant ERP systems. The implementation of ERP, however, was enormously difficult and expensive. Due to the extreme complexity of the software and the major changes required in the associated business processes, many ERP projects were abandoned or had their scope dramatically reduced. As a result, a typical company only implements a very small portion of an ERP package. Many companies continue to rely on their legacy systems or special software to support their operations. Integration of ERP with various enterprise applications remains a challenge (*Themistocleous, Irani, & O'Keefe, 2001)*.

ERP Consolidation

Many companies have also adopted the best-of-breed approach to the implementation of ERP: Picking and choosing the best modules on the market to create their ERP system. For example, a company may use PeopleSoft for human recourses, SAP for finances and manufacturing, and JD Edward (now part of PeopleSoft) for purchasing. In fact, some have estimated that as many as 90 percent of companies have ERP software from different vendors (Worthen, 2003). A 2003 Hackett Group survey, for example, found that an average company had 2.7 copies of ERP systems (2003). Integrating various ERP systems with each other and with other enterprise applications makes systems integration an even more daunting job.

Even if there is only one ERP package from one vendor, integration can be complicated. In the rush to meet the project deadline, especially during the Y2K crisis, many companies made another mistake: Instead of having one instance or copy of, say SAP, serving the whole company, they installed multiple instances of the software for different business units, different geographical locations, etc. Due to a lack of time or will, instead of standardizing, companies allowed different units and locations to keep their idiosyncratic work processes, which required customization of the software in different ways in different units and locations. This has resulted in a proliferation of ERP systems that, even when purchased from the same vendor, are unable to talk to each other. One company reportedly has as many as 64 copies of SAP running in different business units (Sliwa, 2000)! Integration of these multiple instances of ERP is very expensive and will be the major systems implementation project for large corporations in the next decade (Berinato, 2003).

Mergers and Acquisitions

As more and more companies deploy ERP systems, their integration becomes a top priority in mergers and acquisitions (Stedman, 2000). Because these systems are very complex and difficult to implement, their integration with different cultures and management styles from two merging companies presents enormous hurdles (Radcliff & LaPlante, 1999; Stedman, 1999). In addition, since this kind of projects is so resource intensive, it may compromise the implementation of other IT initiatives. Exxon/Mobil's merger in 1998 was especially noteworthy as it resulted in the largest SAP systems integration project at a time when the two companies had to contend with the Y2K issue (King & Nash, 1998).

Due to customizations and release variance, merging ERP systems from even the same vendor can be extremely difficult if not impossible (Kubilus, 2003; Stedman, 1999). Consequently, some firms chose not to integrate their ERP systems initially (Caldwell, 1998; Sliwa, 2000). The decision on which ERP system to keep has also been made irrespective of technical or even financial considerations. When Standard Register acquired Uarco, the former had PeopleSoft's ERP whereas the latter had Baan's. After the merger the new company stayed with Standard Register's legacy system because it was Y2K compliant. After 2000, the company moved to Baan's because the combined company had more experience with it than with PeopleSoft (Caldwell, 1998). Another example is the merger of Dow Chemical with Union Carbide. Dow was, and still is, an SAP R/2 user whereas Union Carbide had implemented SAP R/3. Some industry observer believed that the new company would move to R/3 since it was the newer version (Collett, 1999). In the end, however, the merged company decided to standardize on R/2, which has been an integrated system that supports Dow Chemical's global business operations in 135 countries since 1998.

To summarize, ERP is a solution to the systems integration problem. However, the installation of an ERP package does not create an integrated enterprise (Slater, 2000a). In fact, as discussed above, ERP amplifies the need for systems integration. It is, therefore, imperative that ERP implementers understand systems integration issues and be involved in integration projects. The next two sections discuss tools and success factors for systems integration.

TOOLS OF THE TRADE

Like all information systems project, systems integration takes people, resources and proper project management, some of which will be discussed in the next section. For the software component, most companies rely on middleware to integrate various applications. Middleware may be defined as software products that connect applications to enable data sharing (Slater, 2000b). The advantages of middleware and different types of middleware are discussed next.

Advantages of Middleware

Technically, middleware offers three advantages (Slater, 2000b):

- *Simplicity*: In the past the dominant approach to integration was point-to-point integration, which involved custom coding two or more programs so that they could talk to each other. If one of these programs needed to talk to another program, another point-to-point integration was created. Instead of connecting every application with every other application, the middleware approach allows one connection only from each application—to the middleware. Systems integration becomes much cleaner and more robust. The savings in development and maintenance rapidly multiply when the number of applications in need of integration increases.
- *Persistence*: Some middleware can capture and then hold data until it is processed by the applications or databases involved. This allows asynchronous or batch integration, which is usually easier and cheaper to implement then real-time integration (Hildebrand, 2003).
- *Services*: Middleware can offer commonly needed services such as data checking, validation, and manipulation. Using middleware means companies don't have to duplicate those functions in every application that needs them. Again, the savings in development and maintenance quickly multiply as the number of applications involved increases. As the middleware market evolves, more services will become available to companies that use them. The most promising framework for providing services is web services, which will be discussed in the next section.

Types of Middleware

Middleware can be loosely classified into four groups (White, 2003). Table 1 lists the four groups and their respective integration foci.

- *Object- and Component-Oriented Middleware*: This type of products is based on object-oriented architectures such as Common Object Request Broker Architecture (CORBA), Object Request Brokers (ORBs) and platforms such as Java 2 Platform Enterprise Edition (J2EE) and .NET. Object-oriented architectures describe the interactions between applications and networks and are a key part in building services and interoperable applications (Slater, 2000b). Iona Technologyies' Orbix 2000 is a middleware product based on CORBA release 3.0.
- *Message-Oriented Middleware*: This type of products relays messages or data from one application to another by putting them in a queue waiting for delivery. This is an example of persistence mentioned above and it also allows the receiving application to retrieve the message at a later time, thus permitting asynchronous integration of data. An example is a customer who may have several accounts with a bank, which are tracked in different databases. When the customer updates his address in one account, the middleware can update his other accounts automatically (Slater, 2000b). IBM's MQSeries is a prime example of message-oriented middleware.
- *Web-based Middleware*: This type of products utilizes web technologies such as web browsers, web servers and Extensible Markup Language (XML) as the data exchange protocol. In the business-to-business (B2B) space, XML is expected to replace the current standard, electronic data interchange (EDI), as the data exchange method (White, 2003).

Currently, the most important class of web-based middleware is known as web services, technologies that allow easy integration of applications over the Internet or Internet protocol based networks. In addition to XML, three other standards are involved in web services (Kalin, 2002):

- *Simple Object Access Protocol or SOAP*: Describes how an application talks to a Web service and asks it to perform a task and return an answer,
- *Universal Description, Discovery and Integration or UDDI*: A virtual yellow pages for Web services that lets software discover what Web services are available and how to hook up to them, and

Table 1: Types of middleware

Middleware Grouping	**Integration Focus**
Object- and component-oriented middleware	Sharing (reuse) of business logic
Message-oriented middleware	Data integration
Web-based middleware	Data and user interface integration; newer web services also provide application integration
Application integration middleware	Application integration

- *Web Services Description Language or WSDL*: The little blurb associated with each entry in the UDDI that describes what kind of work a Web service can do.

Web services offer tremendous potential as an integration tool (Kalin, 2002). First, they run on the Internet or Internet protocol based networks, which are the preferred media for e-commerce and e-business. Second, they have the backings of major technology vendors including BEA systems, IBM, Hewlett-Packard, Microsoft, Sun, and Oracle. Some consider web services one of the two dominant ways to consolidate ERP systems; the other being one instance ERP, a monolithic system from one vendor, say SAP or Oracle, that runs all enterprise-wide systems including ERP, customer relations management (CRM), and supply chain management (SCM) (Worthen, 2003).

- *Application Integration Middleware*: This group includes older technologies such as screen scraping, database gateways, and transaction processing monitors. It also includes a new class of software products known as enterprise application integration (EAI), which incorporate technologies for communications/messaging, XML, data transformation, brokers, process models and development frameworks, and business rules. EAI products are comprehensive software suits that allow connection to an array of applications including ERP, CRM and to various databases. Using EAI, big corporations with a large number of applications to integrate can save as much as 80 percent over custom coding (Slater, 2002). EAI has also received the most attention in academic journals as an enterprise integration tool (e.g., Lee, Siau, & Hong, 2003; Themistocleous, Irani, & *O'Keefe*, 2001). It should be noted that the term "EAI" or simply application integration (AI) has been used broadly to refer to an approach to systems integration or narrowly to mean a type of software products. For example, Lee, Siau, and Hong (2003) defined EAI as "a business computing term for plans, methods, and tools aimed at modernizing, consolidating, and coordinating the overall computer functionality in an enterprise" (P. 57). Themistocleous, Irani, and *O'Keefe* (2001), on the other hand, reported various middleware products including EAI that have been used to integrate ERP systems with other applications.

SUCCESS FACTORS FOR SYSTEMS INTEGRATION

In addition to proper software, other factors affect the success of systems integration. Summarized below are guidelines that have appeared in practitioners' journals. Some success factors such as human capital are common to all IT implementation projects while others such as the will to share information and work processes are specific to integration efforts. Empirical validation of these and other factors in a more systematic way seems a promising research avenue.

1. Robust infrastructure: The volume of data to be stored, transported and processed will grow greatly so plan for storage, bandwidth and processing power accordingly (Slater, 2000a).
2. The will to share information and work processes: Integration means free flow of information across organizational boundaries. People must be willing to unlock their information vaults or all the investments in technology will not create an integrated enterprise (Slater, 2000a). Integration also brings new ways of doing things or more tasks for existing users. Consequently, users must be educated and trained properly for them to accept the new responsibilities (Hildebrand, 2003).
3. Clearly defined and communicated vision: Total integration is great but it's also very expensive. Different companies have different integration needs. For companies that are technology laggards, a better strategy may be to choose an ERP from a big enterprise software vendor such as SAP or PeopleSoft and wait for them to roll out add-on modules to integrate additional functions (Hildebrand, 2003). For technology leaders and large companies that have a vast number of existing applications, the payoffs from integration can be huge so EAI or web services make better sense, but the ROI must be communicated to and accepted by business executives. (Slater, 2000a)
4. Prioritization: Analyze each business area and process to determine which should be integrated first and the level of integration. Customer facing applications such as e-commerce and e-business are good candidates for initial integration. Also, some applications require real-time integration whereas others make do with batch or asynchronous integra-

tion, which is cheaper and easier to implement than the former (Hildebrand, 2003).

5. Human capital: Regardless of technology, it is people that make a system work. Retention of integration expertise in house is the key to integration success (Iansiti, 2003). An integration project can fail due to inexperienced project managers (Bass, 2003). Many companies have also created a program office to help manage systems integration projects (Slater, 2000a). A program office is a high-level group that brings visibility to systems integration and creates a forum for knowledge sharing and issues resolution across business areas.

CONCLUSION

Systems integration is an on-going issue in the use of information technology by businesses. Whereas in the past, integration may be viewed as part of the maintenance and upgrade of software and performed in a piece meal fashion, competitive pressures dictate that companies take a proactive view today and develop a comprehensive strategy for integrating all applications in a systematic way. The introduction of ERP is a first step towards integration but it also brings new challenges in creating a truly integrated enterprise. Companies that have the vision, strategy, and support mechanisms for achieving integration of all applications including ERP will be rewarded with sizable increases in sales, profits and market share; those that don't will fall into oblivion.

REFERENCES

Bass, A. (2003). Cigna's self-inflicted wounds. *CIO Magazine*, *16*(11), 1.

Berinato, S. (2003). ERP consolidation. *CIO Magazine*, *16*(7), 1.

Cadwell, B. (1998). Which information systems package to keep. *Computer Reseller News*, (805), 50.

Collett, S. (1999). Merger marries "old" ERP to new. *Computerworld*, *33*(32), 1, 95.

Hildebrand, C. (2002). Putting two and two together. *Darwin Magazine*, *2*(1), 30-35.

Iansiti, M. (2003). Integration the right way, the wrong way. *CIO Magazine*, *16*(15), 1.

Kalin, S. (2002). The essential guide to web services. *Darwin Magazine*, *2*(1), 38-43.

King, J., & Nash, K. (1998). Exxon/Mobil sets up mega SAP project. *Computerworld*, *32*(49), 1-2.

Kubilus, N. (2003). Mergers: IT disaster or IT opportunity? *Computerworld*, *37*(38), 46.

Kumar, K., & van Hillegersberg, J. (2000). ERP experiences and evolution. *Communications of the ACM*, *43*(4), 23-26.

Lee, J., Siau, K. & Hong, S. (2003). Enterprise integration with ERP and EAI. *Communications of the ACM*, *46*(2), 54-60.

Radcliff, D., & LaPlante, A. (1999). When merger mania hits ERP. *Computerworld*, *33*(48), 44-46.

Slater, D. (2000a). The integrated enterprise: The whole … is more than its parts. *CIO Magazine*, *13*(15), 116.

Slater, D. (2000b). The integrated enterprise: Middleware demystified. *CIO Magazine*, *13*(15), 126.

Slater, D. (2002). Costly, painful and worth It. *CIO Magazine*, *15*(7), 70.

Sliwa, C. (2000). Drug giants' merger to bring systems integration hurdles. *Computerworld*, *34*(1), 12.

Stedman, C. (1999). ERP mix a challenge in drug merger. *Computerworld*, *33*(45), 6.

Stedman, C. (2000) Market pressures will make IT a priority in drug merger. *Computerworld*, *34*(4), 2.

Themistocleous, M., Irani, Z., & O'Keefe, R. (2001). ERP and application integration. *Business Process Management Journal*, *7*(3), 195-204.

White, W. (2003). What is middleware and what are the business implications of its use? *Darwin Magazine*, *3*(10).

Worthen, B. (2003). Extreme ERP makeover. *CIO Magazine*, *17*(4), 1.

KEY TERMS

Enterprise Application Integration: Comprehensive middleware software suits that allow connection to an array of applications including Enterprise Resource Planning, Customer Relationship Management and to various databases.

Enterprise Resource Planning: Configurable enterprise software that integrates business processes across functions.

Middleware: Software that connects different applications to allow data sharing.

Simple Object Access Protocol or SOAP: Describes how an application talks to a Web service and asks it to perform a task and return an answer.

Systems Integration: The process of tying together two or more computer systems for sharing data and functionality.

Universal Description, Discovery and Integration or UDDI: A virtual yellow pages for Web services allowing software to discover what Web services are available and how to hook up to them.

Web Services: Technologies that allow easy integration of applications over the Internet or Internet protocol based networks.

Web Services Description Language or WSDL: The little blurb associated with each entry in the Universal Description, Discovery and Integration that describes what kind of work a Web service can do.

XML (Extensible Markup Language): The de facto standard for data exchange among different applications, especially over the Internet.

Enterprise Resource Planning for Intelligent Enterprises

Jose M. Framinan
University of Seville, Spain

INTRODUCTION

Enterprise resource planning systems can be defined as customizable, standard application software that includes integrated business solutions for the core processes and administrative functions (Chan & Rosemann, 2001). From an operative perspective, ERP systems provide a common technological platform, unique for the entire corporation, allowing the replacement of mainframes and legacy systems. This common platform serves to process automation as well as to simplify current process either by an explicit reengineering process or by the implicit adoption of the system 'best practices' (Markus & Tanis, 2000). Finally, the common centralized platform allows the access to data that previously were physically or logically dispersed. The automation of the processes and the access of data allows the reduction of the operating times (thus reducing operating costs) while the latter serves to better support business decisions (see, e.g., Umble, Haft & Umble, 2003, for a detailed review of ERP benefits).

A widespread critique to ERP systems is their high total cost of ownership (Al-Mashari, Al-Mudimigh & Zairi, 2003) and hidden costs in implementation (Kwon & Lee, 2001). Besides, ERP systems impose their own logic on an organization's strategy and culture (Davenport, 1998), so ERP adopters must adapt their business processes and organization to these models and rules. Consequently, organizations may face difficulties through this adaptation process, which is usually carried out without widespread employee involvement. This may cause sore employees, sterile results due to the lack of critical information usually provided by the employees, and also that the new system is delivered late, with reduced functionality, and/or with higher costs that expected (Kraemmeraard, Moeller & Boer, 2003). Additionally, some analysts have speculated that widespread adoption of the same ERP package in the same industry might lead to loss of competitive advantage due to the elimination of process innovation-based competitive advantage (Davenport, 1998). This has been observed, for instance, in the semiconductor manufacturers sector (Markus & Tanis, 2000).

The early stage of ERP was carried out through Materials Requirement Planning (MRP) systems (Umble et al., 2003). The next generation of these systems, MRP II (Manufacturing Resources Planning), crossed the boundaries of the production functionality and started supporting not only manufacturing, but also finance and marketing decisions (Ptak & Schragenheim, 2000). Current ERP systems appeared in the beginning of the '90s as evolved MRP II, incorporating aspects from CIM (Computer Integrated Manufacturing) as well as from EDP (Electronic Data Processing). Therefore, ERP systems become enterprise-wide, multi-level decision support systems. ERP systems continue evolving, incorporating Manufacturing Execution Systems (MES), Supply Chain Management (SCM), Product Data Management (PDM), or Geographic Information Systems (GIS), among others (Kwon & Lee, 2001).

BACKGROUND

Most enterprise resource planning systems share a number of common characteristics, both from a technological as well as a business perspective. These include:

- **Client/server, open systems architecture.** Most ERP packages adopt an open systems architecture that separates data (database server), application (ERP server), and presentation (user interface/ERP client) layers, guaranteeing cross-platform availability and systems integration. In order to interoperate with existing business applications or information systems, most ERP packages adhere to the majority of common standards for data exchange or distributing processing.
- **Enterprise-wide database.** One of the most distinguishable characteristics of ERP is the strong centralization of all relevant data for the company (Al-Mashar et al., 2003). When physical centralization is not possible, communication and/or replication protocols among the different databases should be implemented in order to ensure data consistency and accessibility throughout the entire enterprise.
- **Kernel architecture.** Some ERP systems support more than 1,000 different business functionalities (Bancroft, Seip & Sprengel, 1998), covering nearly

all relevant business aspects for most of the enterprises. As all these functionalities cannot be loaded in the ERP server at the same time, the majority of ERP systems employ a so-called 'kernel architecture'. In this architecture, most functionalities are stored in the ERP database, usually in the form of source code of a proprietary, fourth-generation, programming language. When certain functionality is required by an ERP client, the ERP server loads it from the database and compiles the corresponding code so the functionality is made available for the clients. Once it is not required, the functionality is removed from the ERP server. Note that this mechanism also allows for an easy enhancement/updating of existing functionalities, as well as for the construction of new ones.

- **Process-oriented, business reference model.** ERP is process-oriented software that has been developed, starting from an implicit or explicit business reference model in order to appropriately describe the relevant business functions covered by the ERP system. For most ERP vendors, this model is explicit and takes the form of the 'best practices' extracted form the ERP vendor experience (Markus & Tanis, 2000). This can be used to analyze and evaluate current business processes in the enterprise prior to the implementation of the ERP package, serving thus as benchmark processes for business process reengineering (BPR).
- **Adaptation to the enterprise.** In order to meet the specific requirements of different enterprises, ERP systems are highly configurable. This potential for customization is considered to be one of the main differences between ERP and other standard software packages (Kraemerand et al., 2003). The customization process may take several months, or even years, depending on the enterprise.
- **Modularity.** Although the term 'ERP system' is usually employed to design a system covering all corporate functions (Slater 1998), generally an ERP system is composed of a set of ERP modules. An ERP module is a group of function-oriented, tightly integrated functionalities, which in many cases can be separately purchased and installed. Typical ERP modules are the financial-accounting module, production-manufacturing module, sales-distribution module, or human resources module. This allows enterprises to purchase only these modules, strictly required, as well as offers the possibility of integrating them with existing information systems.

An intelligent enterprise is an organization that acts effectively in the present and is capable of dealing effectively with the challenges of the future (Wiig, 1999). Since most enterprises operate today in a complex and dynamic environment, characterized by increasing competition and continuous changes in products, technology and market forces, an intelligent enterprise should be proactive, adaptable, knowledgeable, and well resourced (Kadayam, 2002). In order to achieve this behavior, it is expected that all employees in the intelligent enterprises not only deliver the work products that are directly associated with their functions, but that they also innovate to improve customer relationships and enterprise capabilities, and to envision opportunities for new products and services (Wiig, 1999). Therefore, it is clear that an intelligent organization should have timely access to all critical information in order to gain insight into its performance and should be able to provide effective decision support systems. Hence, one of the requisites for the intelligent enterprise is the availability of all relevant data in the organization. Indeed, access to the right information is considered to be one of the key characteristics of intelligent enterprises (Smirnov, Pashkin, Chilov & Levashova, 2003).

FUTURE TRENDS

If we adopt the generic intelligent enterprise architecture by Delic and Dayal (2002), ERP addresses issues of supply chain efficiency and back-office optimization, and provides the basis for Enterprise Knowledge Management (EKM). At the same time, the evolution of enterprises to the form of intelligent organizations requires the cooperation of independent companies into a virtual multi-tier enterprise (Olin, Greis & Kasarda, 1999), the Internet providing the glue for their heterogeneous information systems (Delic & Dayal, 2002). In order to achieve this, one of the main trends followed by most ERP systems vendors is the introduction of the Internet (Chan & Rosemann, 2001; Kwon & Lee, 2001). The adoption of the Internet can be seen from two viewpoints—the user interface viewpoint, and the internal/external communication viewpoint.

With respect to the user interface, ERP systems are transaction oriented. However, the connectionless nature of the Internet protocols (i.e., the connection between the Web server and the browser is not maintained after the former has sent the requested data to the later) makes it not well suited for transactions. Therefore, it is intrinsically difficult to adapt the ERP internal structure to the Internet. As a consequence, most of the ERP vendor's effort is on creating reliable gateways between the ERP system and an Internet server.

Regarding the internal or external communication of the ERP system, the emphasis is on the adoption of Internet standards for data exchange. This is done with respect to both the exchange of data among the different ERP modules and to the exchange of data among the ERP system and external applications. Hopefully, this effort will result in the adoption of a common communication standard that will allow the integration of the information systems of customers and/or providers in a supply chain. Additionally, it will make feasible the so-called 'component ERP'—that is, the acquisition of the 'best-of-breed' modules from every ERP vendor (Fan, Stallaert & Whinston, 2000). Since communications among ERP modules has been driven by proprietary protocols, the ERP market has been forcing the enterprises to purchase all modules of the ERP system from the same vendor or face huge costs in developing interfaces for modules from different ERP systems. This may be greatly simplified by the adoption of a public, common, protocol standard such as those on the Internet. Even in the most likely case that interfaces between modules from different vendors are still required, the decrease in the cost of their development may render it affordable for the enterprises (Appleton, 1997; Kwon & Lee, 2001).

THE ERP MARKET: VENDORS AND MARKET TRENDS

About 40% of large companies worldwide and 60% of smaller companies are deploying ERP systems, and 70% of Fortune 1,000 companies have implemented core ERP applications (Bingi, Sharma & Godla, 1999; Yen, Chou & Chang, 2002). The world ERP market has been growing at a rate higher than 30%, and most forecasts predict keeping the figures. Actually, the ERP market is the largest segment of a company's application budget (34%), and it is expected to remain so through 2004 (Scott & Shepherd, 2002; Somers & Nelson, 2004). This growth has been boosted by a number of reasons, such as globalization, market maturity in developed countries, and advances in information and communication technologies, among others.

With respect to the market players, it is not easy to offer precise information about market shares, since the comparative analysis of the published results show a great dispersion depending on the sources. All these analyses show the enormous fragmentation of the ERP market, where more than 100 products are available (see APICS, 2000). Additionally, there is consensus that SAP AG is the world market leader with the product SAP R/3, although its market share oscillates between 16% and 32%, depending on the sources. Along with SAP AG, five other products keep a market share oscillating around 2%. These are PeopleSoft, Oracle, Computer Associated, JD Edwards, and Baan.

THE ERP IMPLEMENTATION PROJECT: RISKS AND KEY SUCCESS FACTORS

ERP acquisition and implementation constitutes a risky project that may result, in a high number of cases, in unsatisfactory, if not failed, system implementations. It has been reported that nearly three-fourths of ERP implementation projects are judged unsuccessful by the ERP implementing form (Griffith, Zammuto & Aiman-Smith, 1999). Cases of failures in well-known organizations such as Boeing (Stein, 1997) or Siemens (Seidel & Stedman, 1998) have been described. Furthermore, these reported instances involve the software of all primary ERP vendors (Motwani, Mirchandani, Madan & Gunasekran, 2002). More detailed reports (Booz & Allen's 1999 report on ERP, cited by Buckhout, Frey & Nemec, 1999) confirm that nearly 35% of ERP implementation projects are cancelled, while in 55% of the projects, the final budget is doubled from the original one, and so happens with respect to the estimated project due dates.

Key success factors stem from the consideration of the ERP implementation project as a strategic project, in terms of time, costs, and expected benefits. Therefore, as with any strategic project, a commitment of the enterprise with the project is required (Appleton, 1997). The influence of the different success factors is discussed, for example, in Clemons (1998), Brown & Vessey (1999), Holland & Light (1999), Sumner (1999), Markus & Tanis (2000), and Parr & Shanks (2000).

The causes of failures can be fitted in three major groups:

- **Inherent complexity of ERP implementation project.** Implementation projects are extremely complex projects affecting key functional areas of the enterprise, usually carried out by a team composed of personnel external to the enterprise, supported by some company staff. To adequately handle this complexity, companies must be willing to dedicate some of their best employees with the right personal, business, and/or technological competencies, and hire external consultants who combine these qualities (Kraemmeraard et al., 2003).
- **Implementation strategy.** In some companies, big-bang implementation strategies (also known as all-at-once) have been followed, seeking to implement

all required features and modules at once, thus reducing the overall implementation time and minimizing the transient period between the former and the new information systems (Gill, 1999). However, this approach delays the visibility of the results for a very long time, leading in some cases to distrust in the overall ERP implementation project (Motwani et al., 2002). Instead, in a phased implementation strategy, the project is divided into milestones, each one representing an ERP package module or set of related functionalities. Then, these modules or sets are implemented one by one, not starting the implementation of a module until the previous one has been satisfactorily implemented and tested.

- **Organizational/cultural clash.** Since the implementation of ERP systems requires disruptive organizational changes (Hammer & Stanton, 1999), it is not surprising that some 60% of the failed implementation experiences arise from cultural or organizational clashes (Wheatley, 2000). Reports point at poor training as the main cause, due to underestimations on the hours of training, or to a training more focused on the technical aspects of the new information system, rather than in clearly explaining the new business processes. The solution to the poor training is obviously a higher awareness of the importance of the ERP training in the success of its implementation (Kraemmeraard et al., 2003). Optimal figures for training costs are between 15-20%. Besides, training must be more focused on a system's business processes rather than on a system's screens, so the users can sufficiently understand the logic of the new system (Wheatley, 2000).

CONCLUSION

ERP systems have been considered one of the most noteworthy developments in information systems in the past decade. ERP systems are present in most large companies that operate in the new millennium. Their advantages in terms of access to information or the integration of business functions has been outlined. However, ERP implementation projects are not risk-free: rates of ERP implementation failures are rather high. Although the failure figures may be partly explained by the intrinsic complexity of the ERP implementation project, some others may be minimized by the consideration of the ERP implementation as a strategic decision in the enterprise, resulting in a principal, long-term project rather than a 'single' information system change.

ERP systems will play a central role in the intelligent enterprise of the future. ERP vendors are continuously adding new features and providing an easy integration with other information systems, as well as among modules from different vendors. Success in the latter issue is claimed to be crucial for maintaining the now outstanding ERP position in the new enterprise.

REFERENCES

Al-Mashari, M., Al-Mudimigh, A. & Zairi, M. (2003). Enterprise resource planning: A taxonomy of critical factors. *European Journal of Operational Research, 146*, 352-364.

APICS. (2000). APICS survey on ERP 2000. Retrieved from *www.apics.com*.

Appleton, E.L. (1997). How to survive ERP. *Datamation,43* (3), 50-53.

Bancroft, N., Seip, H. & Sprengel, A. (1998). *Implementing SAP R/3*. Greenwich, CT: Manning,

Bingi, P., Sharma, M. & Godla, J.K. (1999). Critical issues affecting and ERP implementation. *Information Systems Management*, *16*(3), 7-14.

Brown, C. & Vessey, I. (1999, December 12-15). ERP implementation approaches: Towards a contingency framework. *Proceedings of the International Conference on Information Systems*, Charlotte, NC.

Buckhout, S., Frey, E. & Nemec, J. (1999). Making ERP succeed: Turning fear into promises. *Strategy and Business, 15*, 60-72.

Clemons, C. (1998). Successful implementation of an enterprise system: A case study. *Proceedings of the Americas Conference on Information Systems* (AMCIS), Baltimore, MD.

Chan, R. & Rosemann, M. (2001). Managing knowledge in enterprise systems. *Proceedings of the 5th Pacific Asia Conference on Information Systems* (pp. 916-932), Seoul, Korea.

Davenport, T. (1998). Putting the enterprise into the enterprise system. *Harvard Business Review,* (July-August), 121-131.

Delic, K.A. & Dayal, U. (2002). The rise of the intelligent enterprise. *Ubiquity—ACM IT Magazine & Forum, 3*(45). Retrieved from *www.acm.org*.

Fan, M., Stallaert, J. & Whinston, A.B. (2000). The adoption and design methodologies of component-based enterprise systems. *European Journal of Information Systems, 9*, 25-35.

Gill, P.J. (1999). ERP: Keep it simple. *Information Week*, (August).

Griffith, T.L., Zammuto, R.F. & Aiman-Smith, L. (1999). Why new technologies fail. *Industrial Management, 41*, 29-34.

Hammer, M. & Stanton, S. (1999, November-December). How process enterprises really work. *Harvard Business Review*, 108-118.

Holland, C. & Light, B. (1999). A critical success factor model for enterprise resource planning implementation. *IEEE Software, 16*(3), 30-35.

Kadayam, S. (2002). The new business intelligence. *KMWorld*, (January), S6-S7.

Kraemmeraard, P., Moeller, C. & Boer, H. (2003). ERP implementation: An integrated process of radical change and continuous learning. *Production Planning & Control, 14*(4), 338-348.

Kwon, O.B. & Lee, J.J. (2001). A multi-agent system for efficient ERP maintenance. *Expert Systems with Applications, 21*, 191-202.

Markus, M. & Tanis, C. (2000). The enterprise systems experience: From adoption to success. In R.W. Zmud (Ed.), *Framing the domains of IT research: Glimpsing the future through the past.* Cincinnati, OH: Pinnaflex Educational Resources.

Motwani, J., Mirchandani, D., Madan, M. & Gunasekran, A. (2002). Successful implementation of ERP projects: Evidence from two case studies. *International Journal of Production Economics, 75*, 83-96.

Olin, J.G., Greis, N.P. & Kasarda, J.D. (1999). Knowledge management across multi-tier enterprises: The promises of intelligent software in the auto industry. *European Management Journal, 17*(4), 335-347.

Parr, A.N. & Shanks, G. (2000, January). A taxonomy of ERP implementation approaches. *Proceedings of the 33rd Hawaii International Conference on System Sciences*, Hawaii.

Ptak, C. & Schragenheim, E. (2000). *ERP: Tools, techniques and applications for integrating the supply chain.* Boca Raton, FL: St. Lucie Press.

Scott, F. & Shepherd, J. *The steady stream of ERP investments*. AMR Research. Retrieved from *http://www.amrresearch.com*

Seidel, B. & Stedman, C. (1998). Siemens cuts PeopleSoft loose for SAP. *Computerworld* (Online), (October 5).

Slater, D. (1998). The hidden costs of enterprise software. *CIO Magazine, 12*, 30-37.

Smirnov, A.V., Pashkin, M., Chilov, N. & Levashova, T. (2003). Agent-based support of mass customization for corporate knowledge management. *Engineering Applications of Artificial Intelligence, 16*, 349-364.

Somers, T.M. & Nelson, K.G. (2004). A taxonomy of players and activities across the ERP project cycle. *Information & Management, 41*, 257-278.

Sprott, D. (2000). Componentizing the enterprise application packages. *Communications of the ACM, 43*(4), 63-69.

Stein, T. (1997). Boeing to drop Baan's software. *Information Week*, (August 25).

Sumner, M. (1999). Critical success factors in enterprise wide information management systems. *Proceedings of the American Conference on Information Systems* (pp. 232-234), Milwaukee, WI.

Umble, E.J., Haft, R.R. & Umble, M.M. (2003). Enterprise resource planning: Implementation procedures and critical success factors. *European Journal of Operational Research, 146*, 241-257.

Wheatley, M. (2000). ERP training stinks. *CIO Magazine*, (June).

Wiig, K.M. (1999). *The intelligent enterprise and knowledge management.* Knowledge Research Institute Working Paper.

Yen, D.C., Chou, D.C. & Chang, J. (2002). A synergic analysis for Web-based enterprise resource planning systems. *Computer Standards and Interfaces, 24*(4), 337-346.

KEY TERMS

Best Practices: Process procedures of recognized excellence, usually obtained from companies' experience and or process optimization analysis.

Big-Bang ERP Implementation: ERP implementation strategy consisting of implementing all required modules and features at once.

Business Process Reengineering (BPR): Radically re-thinks key enterprise process in order to achieve substantial process improvement.

Client/Server Architecture: Computer network model separating computers providing services (servers) from computers using these services (clients).

Enterprise Knowledge Management (EKM): Aimed to inject knowledge into business processes and to enable reuse of human expertise through the creation of common data objects and definitions that can be used with equal ease and success by all employees in the enterprise.

Enterprise Resource Planning (ERP): Packaged software to support corporate functions such as finance, human resources, material management, or sales and distribution.

Intelligent Enterprise: Organization capable of acting effectively in the present and dealing effectively with the challenges of the future by being proactive, adaptable, knowledgeable, and well resourced.

Modularity: Most ERP packages decompose their functionality in modules grouping typical business functions such as finance, sales, and manufacturing, among others.

Process Orientation: Recognition of series of functions that carry out an overriding task by providing the customer with a meaningful result.

Enterprise Resource Planning Maintenance Concepts

Celeste See-pui Ng
Yuan Ze University, Taiwan R.O.C

INTRODUCTION

Enterprise resource planning (ERP) is packaged software, and the accompanying set of best practices designed to integrate and serve different departments' needs, and to automate a corporation's business functions and processes on a single management system, database, and interface. ERP allows business to operate in a real-time, more effective and responsive manner and can facilitate better identification, storage, analysis, reuse and distribution of a company's business intelligence (e.g., performance, productivity, customer satisfaction, availability, inventory management, communication). SAP, Oracle, and PeopleSoft are the top three ERP software vendors. ERP software installation and popularity have grown over the past decade, especially in the late 1990s (it was once thought to be the panacea for Y2K problem). The motivations for ERP use are internal business integration, best business practices, competitive advantage, and operational cost reduction. Following the market trend, a large number of ERP implementation projects were conducted, especially by large enterprises. However, multi-million dollar ERP implementation projects are not without problems. In some companies, implementation takes much longer and requires a larger budget than expected, and a large number of modifications and configurations to the software are necessary. Some businesses have to change their existing processes to adapt to the software, and thus have to deal with significant internal resistance and change management. A few companies have even been bankrupted by unsuccessful ERP implementation projects.

A typical packaged software lifecycle, from the client perspective, involves implementation and installation, maintenance, and upgrade. Traditional software maintenance has been acknowledged by many researchers as the longest and most costly phase in the software lifecycle. This is also the case for the ERP context. It has been observed from experience and from the literature that there are inconsistencies in and confusion surrounding the terms used to describe the same ERP maintenance activity and tailoring options. Determining the most accurate and appropriate terms, together with their meaning, is a worthwhile endeavor as it can facilitate better communication among researchers and practitioners, and enable research results to be transferred more effectively to the real world. This paper aims to emphasize better understanding of ERP maintenance activities, and attempts to clear confusions and inconsistencies in the use of terms commonly applied in the ERP maintenance context. The discussion in this text focuses on SAP's ERP software. A Yankee Group survey of 350 business decision makers found that SAP is the most recognized name, followed by Oracle and PeopleSoft (Westervelt, 2004). The survey also revealed that 50% of the respondents preferred a large vendor. This indicates that the present paper will have relevance for a wide audience. In addition, the majority of research done to date has been based on the SAP software. This allows comparisons to be made with existing studies.

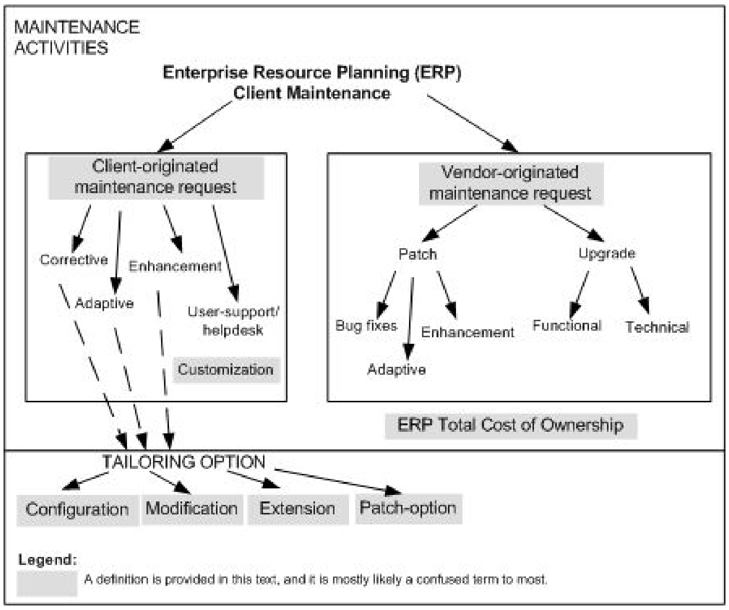
BACKGROUND

Overall, confusion about the terms used in the context of ERP maintenance, as well as inconsistencies in the terms used to refer to the same ERP maintenance activities and tailoring options, are observed in practice, in the trade press and in academic literature. For example, are patch maintenance and upgrade parts of the clients' ERP maintenance activities? Is ERP enhancement similar to ERP modification, and can the term "customization" be used interchangeably with "configuration"?

Client-Originated Maintenance Request Types

In an ERP environment, maintenance requests include those initiated both by the ERP client, and by the vendor (Ng, Gable, & Chan, 2002), see Figure 1. A survey conducted by Glass and Vessey (1999) on ERP maintenance has indicated that most maintenance requests introduced within an ERP-using organization are enhancement-driven. Enhancement is needed to change existing functionality in order to operate in a way that is desired and/or better than that offered by the original 'vanilla-flavored' software. Although in reality the clients' ERP maintenance also includes corrective, adaptive, and user-support (Ng et al., 2002), to the knowledge of the author

Figure 1. Terms defined in this text



very little has been researched and written about them. Previous research has been focused on the costly, lengthy, failure and critical success factors of ERP implementation projects. However, relative to maintenance that includes upgrading (even though some are supported by the vendor – at a cost), implementation issues are likely to be the tip of the iceberg of ERP software lifecycle costs and issues.

According to Ng et al., ERP maintenance is defined as:

post-implementation activities related to the packaged application software undertaken by the client-organization from the time the system goes live until it is retired from an organization's production system, to: keep the system running; adapt to a changed environment in order to operate well; provide help to the system users in using the system; realize benefits from the system; and

Table 1: SAP categorization of tailoring options

Tailoring option	Description*
Customization	— involves changes made within the standard vendor code by setting system parameters via SAP's configuration tables. In SAP R/3, the configurable elements are central functions, organizational elements, control elements, data validation, and system control (Bancroft, Seip, & Sprengel, 1998).
Enhancement using customer exits	— is for requirements that are not included in the standard SAP. Customer exits are incorporated in the standard as empty modification 'shells'. Customers fill the 'shell' with their own coding.
Enhancements using business add-ins	— allows ERP client organizations to add their custom code or the standard supplemental solutions provided by SAP, and also permits third-party software development to be incorporated in the standard code (Kessler, 1999).
Assisted modifications	— is done using the modification assistant to create customer-specific objects within the customer name range; e.g. screen layout, and function key assignment.
Modifications to the SAP standard	— involves modifying SAP standard objects.

**Source: Kessler, 1999; SAP, 1998*

Table 2: Researchers categorization of SAP software tailoring options

Tailoring option*	**Description***
Configuration	Setting of parameters (or tables) in order to choose between different executions of processes and functions.
Bolt-on	Implementation of a third-party package designed to work with the ERP system and to provide industry-specific functionality.
Screen masks	Creating new screen masks for input and output of data.
Extended reporting	Programming of extended data output and reporting options.
Workflow programming	Creating non-standard workflow.
User exits	Programming additional software codes in an open interface.
ERP programming	Programming additional applications without changing the original source code.
Interface development	Programming interfaces to legacy systems or third party products.
Package code modification	Changing the source-codes.

**Source: Brehm et al., 2001*

keep the system a supported-version and meet the vendor's requirements for standard code. These activities include: implementing internal change-requests; responding or handling user-support requests; upgrading to new versions/releases; and performing patches. (pg. 100)

Solutions: Tailoring Options

In order to implement an enhancement, corrective or adaptive maintenance request, knowledge of the tailoring options available and/or provided by the vendor is important. According to SAP (1998) and consultant Kessler (1999), there are five fundamental choices for tailoring to the SAP standard. The list and description of the SAP tailoring options are given in Table 1.

SAP tailoring categorization is technical-oriented and emphasizes to the performance of particular tailoring tasks, and what tool to use. A more detailed analysis of the tailoring option is given in the academic literature (see Brehm, Heinzl, & Markus, 2001). Brehm et al. (2001) describe the tailoring options from multi perspectives — how they are done, what is changed/modified, and the layers (i.e. interface, application and database) involved (see Table 2). This categorization is obtained from an analysis of the implementation literature and from the authors' interviews with MIS directors (from SAP client-organizations).

In another version of tailoring options given by Glass and Vessey (1999), the SAP term "customization" is used to refer to those configuration activities which involve changes made to ERP functionality via internal configuration switches. On the other hand, "extension" in (Glass & Vessey, 1999) is used to describe changes associated with user exits, bolt-ons and custom code add-ons.

Vendor-Originated Maintenance Request Types

The externally-originated maintenance activities comprise support packages or patches and upgrade versions, which are distributed by the vendor but implemented by the ERP client on its ERP system. A support package or patch contains corrections and further adjustments for an already installed version. ERP client organizations obtain maintenance support services through a maintenance contract with an ERP vendor, for example SAP. The annual maintenance fee is usually between 12 and 20% of the initial software license fee (Butler, 2001). In providing efficient maintenance support to its clients, SAP posts the patch support on the online system. This system offers various patch types dedicated to fixing program bugs in particular areas.

An ERP upgrade is part of ERP post-implementation activities. It involves replacing an installed ERP version with a newer version that is readily available (nearly always) from the same vendor in the market. In contrast to patch maintenance (which is meant for bug fix and, occasionally, minor enhancements), organizations typically upgrade to a new ERP version in order to realize the benefits of substantial new functionality (SAP, 2002; Stein, 1999), and new technologies or business opportunities (Callaway, 2000). There are basically two types of upgrade: functional and technical. Upon implementing a functional upgrade, the system users will experience new and/or additional functionality in the system. On the other hand, no added functionality is usually found from a technical upgrade. This type of upgrade is usually triggered by the need to migrate to a better software platform, to obtain better performance, and to maintain a viable migration path. A technical upgrade is a "like to

Table 3. ERP upgrade cost factors

Cost factor	Description*
Software	A new version that has more functionality, flexibility, and extensibility, will generally cost more to upgrade.
Hardware	Additional and more powerful hardware, or new hardware that is compatible with the new system (Jakovljevic, 2000).
User training	It is driven by changes in user interface and/or new functionality in a new system (Ohlson, 2000).
Consultant	Knowledge and expertise required in an ERP upgrade are not always readily available internally. Hence, most of the time, external consultants are required.
Upgrade implementation	Upgrade implementation involves the cost of data conversion, system analysis, system integration and testing, and post-implementation turmoil (Jakovljevic, 2001).

**Source: Ng, 2001*

like functionality replacement". On the other hand, a functional upgrade delivers major business improvements and enhancement benefits (Ng, 2001).

Unlike internally-originated maintenance, client organizations do not have to determine which tailoring option to adopt in performing a patch-maintenance or an upgrade. Specific program, tools, and methodology are provided by the vendor to assist clients in these activities.

In making the decision to upgrade, one has to consider several factors. These include the availability of the desired functionality (Collins, 1999; Ohlson, 2000), the right and sufficient personnel, the comfort and familiarity of its users with the system (400-Group, 1998), and the vendor's upgrade path availability and support window (Collins, 1999). Similar to the initial ERP implementation and acquisition cost, upgrade expenses are as shown in Table 3.

MAIN THRUST OF THE ARTICLE

This section will attempt to provide a clearer definition of which terms are appropriate to which situation. Figure 1 depicts the commonly encountered, and yet confusing, terms in the context of ERP client maintenance. A client-originated maintenance request originates within a client organization. It is meant to fix bugs, adapt internal and external changes to business processes, make enhancement to the installed system to meet new user requirements, and provide helpdesk/user-support. On the other hand, vendor-originated maintenance requests are maintenance needs that are created by a vendor in order to fix bugs in the packaged software, adapt to new legal requirements, impose the vendor's maintenance-support strategy and upgrade path availability, and provide additional enhancement to business process functionality and meet user requirements. In some cases, the terms "customization" and "enhancement" are used interchangeably, but, is this appropriate? "Customization" is more suitable as a generic term to refer to the need for changes, modifications or configurations to the standard ERP system. It should not be mistakenly associated with enhancement request or "configuration" tailoring options.

Also, as has been observed from the literature in the previous section, inconsistent use of terms to refer to the same tailoring option exists between the top ERP vendor and researchers, and among researchers. This is likely to

Table 4. Match between researchers and SAP tailoring typologies

Brehm et al. (2001) tailoring option	Corresponds to SAP's definition of
Configuration	Customization
User exits	Enhancement using customer exits
Bolt-on	Enhancement using business add-in
Screen masks, Extended reporting, Workflow programming, ERP programming, Interface development	Assisted modifications
Package code modification	Modifications to the SAP standard

Table 5. ERP tailoring options versus client-originated maintenance requests

		ERP client-originated maintenance request		
		Corrective	Adaptive	Enhancement
Tailoring option	Configuration	✓	✓	✓
	Modification	✓(for custom code only)	✓	✓
	Extension		✓	✓
	Patch-option	✓	✓	✓

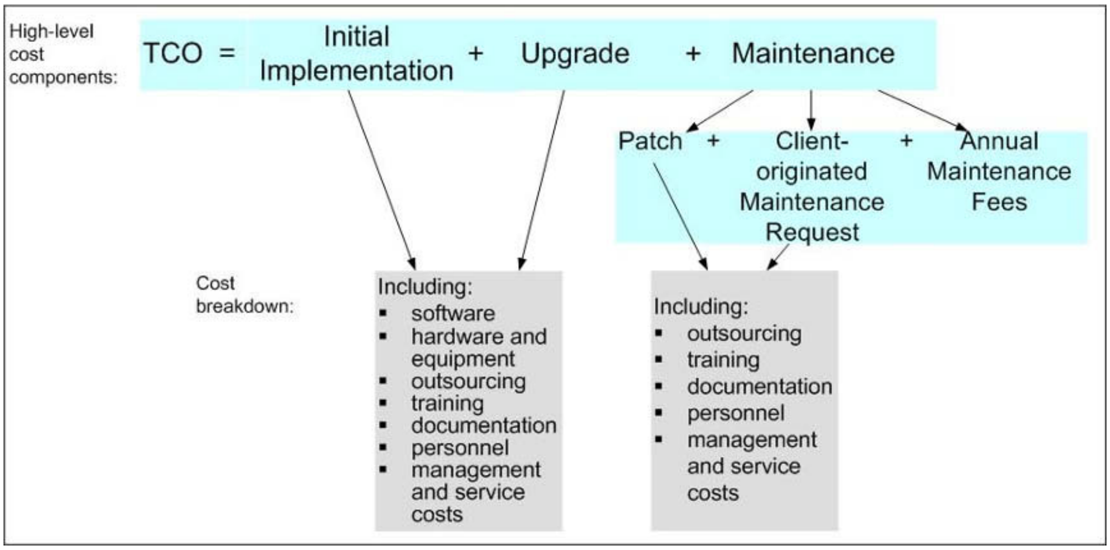
be due to a lack of a common understanding and a formal definition for ERP tailoring options among researchers. Moreover, in the author's opinion, this inconsistency exists because sometimes people associate a maintenance activity with a tailoring option and vice versa (e.g., "enhancement" with "modification", "customization" with "configuration").

By comparing the description of each tailoring option from the SAP (1998) and Brehm et al. (2001), overlaps are observed. The extent of correspondence in terms between Brehm et al. and SAP is illustrated in Table 4.

It is proposed that only the term "configuration" (as it is more intuitive than "customization") be used to refer to the tailoring option where changes are done, within the standard vendor code, by setting system parameters via vendor provided configuration tables. "Modification" is used to refer to the tailoring option where a vendor's standard objects and/or code are changed or modified. On the other hand, for tailoring options that involve adding custom or third-party codes to user-exits, add-ons, reports or user interfaces, without changing the vendor code, the proposed term is "extension". It is of note that the tailoring options reviewed here are non request-type specific or applicable only to enhancement requests. Besides this, an additional tailoring option – "patch-option" – is introduced. This is because occasionally general enhancement, adaptation, and correction requests can be satisfied by implementing the patches supplied by the vendor. Table 5 is a cross-tabulation between the four types of tailoring option and the three fundamental types of maintenance request. It shows the possible tailoring options that can be applied to the three maintenance request types.

In light of the typical packaged software lifecycle, the total cost of ownership (TCO) for ERP software comprises three main components: initial implementation, maintenance, and upgrade. This is illustrated in Figure 2.

Figure 2. ERP total cost of ownership



FUTURE TRENDS

Like other software systems, ERP is not a panacea for all. ERP is built to be generic, in order to meet a wide range of clients needs. Thus, while some clients try to change their business processes, some change the standard code in order to incorporate their unique business processes where they perceive critical for competitive advantages. Also, although ERP software provides a comprehensive set of business applications that can serve different departments under a single system, most companies still maintain their idiosyncratic legacy systems and some buy best-of-breed packaged software from different vendors. This is the result of their unwillingness to replace a system that is still working, and the packaged software does not meet their requirements or simply not good enough. Besides, such hesitancy could also be the result of an economic slowdown, wanting to avoid vendor lock-in, waiting for the web-application to mature, and so forth. However, multi-systems lead to integration problems. As a consequence, time and effort need to be invested in order to make disparate systems 'talk' to each other. Some perceive that the answer to this is through Web applications. But, as far as Web applications are concerned, there are a number of issues yet to consider and address, including cost, benefit, network bandwidth and infrastructure, threats and risks, security, and ethics. Moreover, in light of the growing customer base of the ERP system, it is valuable to investigate how ERP maintenance and upgrades could be better managed. How can the implementation of patches be better queued and implemented to achieve economies-of-scale? And what process model will best describe activities in ERP maintenance and upgrades?

CONCLUSION

ERP maintenance and upgrade issues become immediate important topics after client organizations have their ERP system in place – if not sooner. This paper suggests that inconsistency and confusion in some terms commonly used in the ERP maintenance environment can be minimized by avoiding the association between maintenance request types and tailoring options. The tailoring option called patch-option is introduced to better reflect the actual ERP maintenance environment. In light of the current ERP market situation, integration, Web-application, and maintenance and upgrade process model issues will remain and get more attention from researchers and practitioners in the future.

REFERENCES

400-Group. (1998). Boston beer co. minimizes coding changes, foregoes to keep SAP running smoothly. *I/S Analyzer, 37,* 13-15.

Bancroft, N. H., Seip, H., & Sprengel, A. (1998). *Implementing SAP R/3* (2nd ed.). Greenwich, CT: Manning Publications.

Brehm, L., Heinzl, A., & Markus, M. L. (2001). Tailoring ERP systems: A spectrum of choices and their implications. Paper presented at the *34th Hawaii International Conference on Systems Sciences*, Hawaii, USA.

Butler, J. (2001). Risk management skills needed in a packaged software environment. In J. M. Myerson (Ed.), *Enterprise Systems Integration* (2nd ed., pp. 439-448). Boca Raton, FL: Auerbach.

Callaway, E. (2000). *ERP - The next generation: ERP is web enabled for e-business* (1st ed.). Charleston: Computer Technology Research Corp.

Collins, K. (1999). Strategy and execution of ERP upgrades. *Government Finance Review*. *15*(4), 43-47.

Glass, R. L., & Vessey, I. (1999). Enterprise resource planning systems: Can they handle the enhancement changes most enterprises require? *The Software Practitioner, 9,* 1-12.

Jakovljevic, P. J. (2000). *Essential ERP - Its underpinning technology*. TechnologyEvaluation.Com. Retrieved February 14, 2001, from *http://www.technologyevaluation.com/*

Jakovljevic, P. J. (2001). *ERPbBeginner's guide in so many words*. TechnologyEvaluation.Com. Retrieved February 5, 2001, from *http://www.technologyevaluation.com/*

Kessler, K. (1999). Extending and modifying the SAP standard with business add-ins and the new modification assistant. *SAP Professional Journal*. *1*(1), 3-16.

Ng, C. S. P. (2001). A decision framework for enterprise resource planning maintenance and upgrade: A client perspective. *Journal of Software Maintenance and Evolution: Research and Practice*. *13*(6), 431-468.

Ng, C. S. P., Gable, G. G., & Chan, T. (2002). An ERP-client benefits-oriented maintenance taxonomy. *Journal of Systems and Software*. *64*, 87-109.

Ohlson, K. (2000). *Study: R/3 users face high costs for upgrades*. Computerworld. Retrieved January, 10, 2001, from *http://www.computerworld.com/*

SAP. (1998). *BC enhancements to the SAP standard*. SAP AG. Retrieved June 23, 2000, from *http://www.sap.com/*

SAP. (2002). *SAPcCustomers report significant return on investment from my SAP™CRM*. SAP AG. Retrieved June 28, 2002, from *http://www.sap.com/*

Stein, T. (1999). *ERP overhaul for lawson — Strategic ledger among new features*. InformationWeek. Retrieved November 16, 2001, from *http://www.informationweek.com/*

Westervelt, R. (2004). *Survey: ERP vendors need image boost*. SearchSAP.com. Retrieved March 12, 2004, from *http://searchsap.techtarget.com/*

KEY TERMS

Client-Originated ERP Maintenance Request: Originated from within a client organization. It is intended to fix bugs, adapt internal and external changes to business processes, make enhancements to the installed system to meet new user requirements, and provide helpdesk supports.

ERP Configuration: A tailoring option, which involves setting or configuring a generic/industry-specific ERP system using the switches/tables provided by the vendor in order to personalize the ERP system to support an organization's business practices, and requirements.

ERP Customization: Customization simply means that changes, modifications or adaptations are needed in order to meet some user-requirements. It can be carried out via configuration tables, adding extensions or making modifications to the standard code (but cannot be done by applying any patch provided by the vendor).

ERP Extension: A tailoring option, which involves adding custom or third-party codes to user-exits, add-ons, reports or user interfaces, without changing the vendor/standard code.

ERP Modification: A tailoring option, which results in changes being made to the existing ERP (standard) code and custom objects being created.

ERP Patch-Option: A tailoring option, where vendor's patch(es) are used to service a maintenance request.

ERP Total Cost of Ownership (TOC): ERP software lifecycle cost covering the initial implementation and installation, continuous maintenance, and upgrades to the system until the software is retired from the production system. It includes all the software, hardware and equipment, annual maintenance fees, outsourcing, training, documentation, personnel, management and service costs.

Tailoring Option: Describes how changes can be incorporated into the ERP system or how maintenance request can be serviced.

Vendor-Originated ERP Maintenance Request: Created by a vendor in order to fix bugs in the packaged software, adapt to new legal requirements, impose the vendor's maintenance-support strategy and upgrade path availability, and provide additional enhancement to business process functionality and meet user requirements.

ERP Adoption in Indian Organizations

Monideepa Tarafdar
University of Toledo, USA

INTRODUCTION

Enterprise resource planning (ERP) systems integrate various functions and processes in organizations. ERP software is developed in the form of different modules, each of which helps to perform distinct functions within the company. The modules interface with the same database and are integrated so that workflows can be designed across different modules. The software helps standardize business processes and ensures organization-wide availability of transaction data. ERP software evolved from earlier manufacturing resource planning (MRP) systems, which included inventory management, procurement and production planning functions. The implementation of ERP software started in the early 1990s and during the late 1990s, the growth rate of the ERP market was between 30 to 40%. As of 2001, 30,000 companies around the world had implemented ERP and the total value of the ERP market was at $25 billion. There is not much literature relating to ERP implementation and adoption in companies in Asia and other parts of the developing world. These organizations face issues that are significantly different from those faced by organizations in the developed world, because of differences in the sophistication of IT use, and in the cultural and social contexts. In this article, we describe some experiences that companies in India have gone through in implementing ERP systems. We present a framework for analyzing the critical factors and issues that influence the ERP adoption process, and highlight the areas of opportunity and risk. The framework is sufficiently general so as to be extended to other developing countries.

BACKGROUND

The implementation of ERP software is quite different from traditional software development. ERP software is a single program that is bought off the shelf and then configured to include the specific characteristics of processes of individual companies. Parts of the software have to be customized such that they can correctly represent the workflow and processes of the particular company. This is a complex process, requiring many activities to be carried out. Most studies have described the implementation process in terms of stages.

Ross and Vitale (2000) suggest that ERP implementation is done in five stages. The *design* phase deals with planning and standardization, where an organization chooses the specific package and decides on the extent of customization required. In the *implementation* stage, the software is implemented and goes live. In the *stabilization* phase, the firm adjusts to the new system and integrates it into its existing operations by identifying and smoothing out integration hiccups. The *continuous improvement* stage is marked by additions to the existing functionality of the ERP package through add-ons from other vendors. Organizations typically implement data warehousing, data mining, customer relationship management (CRM) and supply chain management (SCM) software to augment the transaction-based capabilities of the ERP software. In the *transformation* stage, the organization starts to see the benefits of the ERP system, in terms of more efficient processes and possibly better information exchange with partners, leading to flexibility and responsiveness.

Rajagopal (2002) proposes a six-stage model. In the *initiation* stage, organizations study possible business benefits, such as IT infrastructure integration and business process re-engineering, which might make ERP adoption necessary. The *adoption* stage consists of activities such as investment decisions and cost-benefit analysis related to choice of the package and the vendor. In the *adaptation* stage, the system is implemented and becomes available for use. In the *acceptance* stage, users become more comfortable with using the ERP system, their requirements are incorporated, and the overall benefits of the system become apparent. In the *routinization* stage, system integration is realized, users fully accept the system and its use becomes a routine activity. Finally, during the *infusion* stage, the organization looks to the next level of benefits that might be available. The organization moves beyond "just" using the system—it uses the available information to enhance the performance of different functions.

In the model proposed by Markus and Tannis (1999), the *chartering* stage comprises review and selection of the package and consultants, and clarifying the business related factors that make ERP a necessity. The *project* stage describes different aspects of the implementation process and consists of project management, software customization and process re-engineering. During the *shakeout*, managers familiarize themselves with the software. System bugs are reported and fixed and the operational effects on the business are felt. Finally in the *upward and onward* phase, strategic business benefits from ERP occur, additional technical skills are built and upgrades are planned for.

ERP implementation results in significant changes in the IS architecture, redesign of process, increased managerial competence and comfort with new technology and a greater role for IT in critical processes (Scott et al., 2000). A number of factors influence the ERP adoption process. First, strong leadership support is important for ensuring the availability of resources and manpower, and for signaling to the employees the importance of the software (Baskerville et al., 2000; Bingi et al., 1999; Markus, 1999; Parr et al., 1999; Sarkar et al., 2000). Second, open and honest communication about the ERP initiative results in greater understanding of organizational needs and hence quicker acceptance of the software (Davenport, 1998; Holland et al., 1999; Mendel, 1999). It also helps employees understand the rationale for ERP implementation and enables them to appreciate problems in existing systems. A third influencing factor is the skill and competencies of the implementation team. Implementation teams that are technically strong, empowered to make decisions, politically close to important people in the organization and have a good understanding of end-user needs, enable organizations to minimize implementation difficulties and time and cost overruns (Askenas et al., 2000; Baskerville et al., 2000; Ferranti, 1998).

THE ADOPTION OF ERP SOFTWARE IN INDIA: A PROCESS MODEL

Issues and Problems

The first ERP systems were adopted in India in the mid 1990s and the Indian ERP market grew at an average rate of 70% over the years 1995 to 20011. As of 2001, about 800 companies had implemented ERP software2. The study of ERP implementation experiences in Indian companies is an interesting case of new technology adoption because many Indian companies go from very rudimentary IT-based systems to sophisticated ERP systems in one quantum jump (Sharma, 2001). We studied the ERP adoption processes of 25 companies from 10 different industries in the manufacturing and service sectors, which had implemented ERP software, to understand the factors that characterize the ERP adoption process in Indian companies. This article presents the findings in the form of an overarching framework that describes different aspects of the ERP implementation process of companies in India.

A Framework for Analyzing the Adoption of ERP in Indian Organizations

The ERP implementation process in Indian companies can be modeled as a *phase-stage* process framework, as shown in Figure 1. Each cycle in the spiral represents a single phase of implementation, in which a given number of modules are implemented. The innermost cycle represents the first phase and subsequent phases are introduced as more modules are implemented. The number of cycles or phases that are required depends on the characteristics of the individual organization, like annual turnover, size, complexity of organizational processes and IT readiness. Each phase has three stages and the activities carried out within a stage are shown alongside the section representing the stage.

The *planning stage* consists of activities before the ERP software is implemented. These include drawing up a business case and providing justification for the implementation of ERP, such as improvement of critical operational parameters. They also involve making decisions about whether to re-engineer processes using the *"best practices"* feature of the software or to customize the software to existing needs and processes. Planning activities revolve around the functioning of the IS department. In many cases, the IS function is not considered a very important one, prior to the implementation of the ERP software. Since ERP systems have great reach and are technically very complex software, it is essential to change the status of the IS function. Related planning activities therefore include giving greater powers to existing IS functional heads or instituting new CIO positions, which are powerful and visible, and can take charge of the implementation process. IS professionals trained in both technology and business are recruited. On the technology front, anticipated changes involve migration of data from legacy systems into an integrated platform, integration of existing application programs into the ERP system, and connection of separate servers and terminals into a single networked configuration. There is also an accompanying need for training and technical skill acquisition for maintaining the new systems. External consultants are recruited in order to help with software configuration, customization and expertise transfer. Since the ERP project is usually much larger than any previous IS project in terms of the scale of resource commitment, range of operations affected, and the number of people involved, the risk of failure is perceived to be high. Therefore it is essential to plan ahead for extra resource allocation and increases in IT budget, in case of unforeseen problems during implementation.

The *implementation* stage deals with issues that are required to be addressed once the implementation pro-

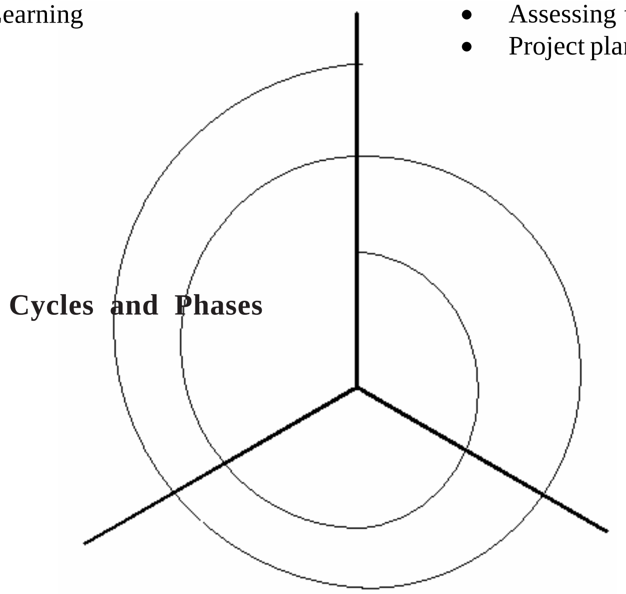
Figure 1. Phase-stage model for adoption of ERP in Indian organizations

Post-implementation Review Stage

Activities:
- Institutionalizing Process changes
- Formalizing Organizational benefits
- Assimilating Phase Learning

Planning Stage

Activities:
- Drawing up a business case
- Understanding business characteristics
- Assessing the IT readiness of the organization
- Project planning



Implementation Stage

Activities:
- Management of technical aspects
- Organizational change management

cess is underway. The ERP system is significantly different from existing systems, and requires advanced technical skills. Hence the implementation process is a joint exercise between external consultants and company employees. These consultants also train the IS professionals, who later become responsible for the maintenance of the system. Those IS professionals who are not directly involved in the ERP project or have not trained on the software are often cynical of the possible benefits of the projects and refuse to co-operate. This is attributed to a lack of motivation and a feeling of alienation amongst these employees. Hence there is a need for making sure that these employees support the project requirements and help in the overall effort.

ERP results in significant changes at various levels. Task definitions are changed and employees require new skills and understanding as different tasks are combined. Reorganization of tasks also results in changes in the reporting structure. These transformations are often difficult for employees to accept quickly because of the magnitude of change involved. We found that senior managers were apprehensive that if they were not able to communicate clearly the nature of, and rationale behind, the changes, software would be rejected. Hence the primary risk of failure of the ERP system stems from human change acceptance issues, and not technical problems.

The *post-implementation review* stage is concerned with analysis of benefits and changes accruing from the ERP system. It also includes an examination of key learnings and skills acquired from the implementation process, and their application to subsequent rounds of implementation of the software. In most of the organizations that we studied, there were significant changes in business processes and reduction in the number of overlapping activities, coupled with greater understanding of task requirements because of multi-tasking and the combination of different stages of tasks into single steps. There was an overall shift from being function driven to process driven. There were also enhancements in the technical capabilities of employees and IS professionals. Operational benefits included improved customer service, reduced costs, better inventory management, and the facilitation of real-time processing, flow and availability of financial and other information. All of this led to better decision-making and greater accountability. Overall, ERP was used to create a systematic structure for storing and retrieving information seamlessly across functions.

Differences between Developed and Developing Nations

In this section we describe differences in the ERP implementation process between companies in the developed and developing nations, as illustrated by the phase-stage model.

1. ***The Phase-Stage Framework:*** The most important aspect of the framework is the presence of phases. In each phase, a company selects specific modules, implements them and leverages the key learnings in subsequent phases of implementation. The process of phasing has allowed Indian companies to limit exposure to implementation risks. While the presence of stages has been described in existing studies from developed countries, the presence of phases has not been mentioned. This is probably because most companies in developed countries have had a relatively longer history of organization-wide IT usage and the level of general awareness and societal acceptance is higher. Hence they have taken a big-bang approach, implementing all major modules in one go, in order to reap the benefits of integration quickly. Indian companies, and possibly those in other developing nations, because of their low level of IT usage prior to ERP, have had to adopt a conservative approach towards implementation. This has allowed them to avoid committing vast resources that a full package implementation would have demanded, to infuse changes gradually and minimize the risks of failure. The phased implementation, in concept and in spirit, is similar to the spiral software development lifecycle model of Boehm (1986), which suggests a risk-driven approach to software development.

2. ***The Existence of Stages within each Phase:*** The second important aspect of the framework is the existence of different stages within a phase. The presence of stages is in broad theoretical agreement with the stage theories described in the earlier section. In addition, the phase-stage model also describes specific aspects of each stage that are typical for companies in India and possibly in other developing economies, and has described the implications of these aspects for managers.

3. ***Changes in the Information Systems Function:*** ERP implementation has resulted in significant changes in the skill sets and organizational importance of the IS department in many organizations in India. IS professionals who had earlier worked with very basic systems have acquired highly marketable implementation skills. This has resulted in a high attrition rate of IS professionals and a high level of turbulence in the IT humanpower pool. This aspect is not as pronounced in existing models of ERP adoption based on studies in developed societies.

FUTURE TRENDS

Future developments in ERP software need to be analyzed in the light of two contexts. The first relates to the enhancement of technical and functional capabilities of the ERP software. ERP packages now provide Web-enabled functionalities and e-business suites for use in B2B and B2C transactions. In this context, SAP has launched mySap.com, Baan has a product called iBaan, and Oracle has incorporated e-business functionalities into their new ERP software called Oracle 11i. Moreover, many ERP vendors have positioned themselves as one-stop vendors for different integration requirements, and are adding CRM and SCM functionalities into their software. This is in response to integration problems that organizations have had when they have tried to interface different best of breed solutions. Therefore, basic ERP packages are expected to serve as the back-end transaction-processing database, to which e-business modules incorporating functions relating to e-procurement, CRM and SCM would be added.

The second context that is expected to shape the future of the ERP industry, especially in developing nations, is the decreasing cost of ERP solutions. Between 1995 and 2000, most customers have been large companies who are market leaders with considerable slack resources. Most of the large companies have already implemented ERP. The future market is represented by companies in the small and medium enterprise (SME) segment. The phase-stage framework would be important for these companies because the financial outlay required for full-scale ERP implementation is significantly larger than the normal IT budget, for these companies. Hence a gradual and risk driven approach to ERP implementation is required. This framework would also be important for global companies with facilities located in developed and developing nations. Using this framework would help them find an implementation pace that suits both kinds of settings.

Future research efforts would therefore need to be directed at analyzing ERP adoption as part of an overall framework for adoption of e-business, along with the implementation of other packaged software such as CRM and SCM. The study of ERP adoption by SMEs and the accompanying specific problems and issues is another area of potential investigation.

CONCLUSION

Over the last few years, many organizations in India have benefited from improved processes and better information availability as a consequence of the implementation of ERP solutions. For many others, the adoption of ERP has resulted in a very painful transition and adaptation period, while the benefits have not been immediate or tangible. This model presents a general conceptual framework and serves as a useful starting point from where the ERP experience of Indian companies can be analyzed. It also presents some practical implications for managers for managing and controlling relevant aspects of different stages of the implementation process. The generality of the framework has been further enhanced because of the number of industries covered in the study. The framework can be used for organizations in other developing societies, as well as in SMEs.

REFERENCES

Askenas, L., & Westelius, A. (2000). Five roles of an information system: A social constructionist approach to analyzing the use of ERP systems. In Orlikowski et al. (Eds.), *Proceedings of the 21st International Conference on Information Systems.*

Baskerville, R., Pawlowski, S., & McLean, E. (2000). Enterprise resource planning and organizational knowledge: Patterns of convergence and divergence. In W. Orlikowski et al. (Eds.), *Proceedings of the 21st International Conference on Information Systems.*

Bingi, P., Sharma, M.K., & Godla, J.K. (1999). Critical issues affecting an ERP implementation. *Information Systems Management, 16*(3), 7-14.

Boehm, B. (1986, August). A spiral model of software development and enhancement. *ACM SIGSOFT Software Engineering Notes.*

Connor, S.J. (1999, September 15). The ERP dilemma. *Dataquest–India.*

Dasgupta, S. (2000, December 18.). The technology behind the Colgate smile. *Network Computing.*

Davenport, T.H. (1998, July/August). Putting the enterprise into the enterprise system. *Harvard Business Review.*

Ferranti, M. (1998). Debunking ERP misconceptions. *Info World, 20*(33).

Holland, C., & Light, B. (1999, May/June). A critical success factors model for ERP model for ERP implementation. *IEEE Software*, 30-36.

Markus, M.L., & Tanis, C. (1999). The enterprise systems experience—from adoption to success. In R.W. Zmud (Ed.), *Framing the domains of IT research: Glimpsing the future through the past.* Cincinnati: Pinnaflex Educational Resources.

Mendel, B. (1999). Overcoming ERP project hurdles. *InfoWorld, 21*(29).

Natarajan, G. (1998, November 11). Implementing ERP: A process-centric approach. *Dataquest–India.*

Parr, A.N., Shanks, G., & Darke, P. (1999). Identification of necessary factors for successful implementation of ERP systems. In O. Ngwenyama, L.D. Introna, M.D. Myers & J.I. DeGross (Eds.), *New information technologies in organizational processes: Field studies and theoretical reflections on the future of work* (pp. 99-119). Boston: Kluwer Academic Publishers.

Rajagopal, P. (2002). An innovation-diffusion view of the implementation of ERP systems and the development of a research model. *Information and Management, 40,* 87-114.

Ross, J.W., & Vitale, M.R. (2000). The ERP revolution: Surviving vs. thriving. *Information Systems Frontiers, 12*(2), 233-241.

Sadagopan, S. (1999, July). ERP honeymoon is over. *Computers Today*, 1-15.

Sandoe, K., Corbitt, G., & Boykin, R. (2001). *Enterprise integration.* New York: John Wiley and Sons.

Sarkar, S., & Lee, A.S. (2000). Using a case study to test the role of three key social enablers in ERP implementation. In W. Orlikowski et al. (Eds.), *Proceedings of the 21st International Conference on Information Systems.*

Scott, J.E., & Vessey, I. (2000). Implementing enterprise resource planning systems: The role of learning from failure. *Information Systems Frontiers, 2*(2), 213-232.

Sharma, M. (1999, March 21). Avon cycles: From Munims to ERP. *Dataquest–India.*

Tarafdar, M., & Roy, R. (2002). A framework for analyzing the adoption of enterprise resource planning systems in Indian organizations. In M. Khosrow-Pour (Ed.), *Issues and trends of information technology management in contemporary organizations, Proceedings of Information Resource Management Association (IRMA) Conference.*

Tarafdar, M., & Roy, R. (2003). Adoption of enterprise resource planning systems in Indian organizations: A process perspective. *Journal of Global Information Technology Management*, (4), 231-251.

Wagle, D. (1998). The case for ERP systems. *The McKinsey Quarterly, 2,* 130-138.

Willcocks, L., & Sykes, R. (2000). The role of the CIO and IT function in ERP. *Communications of the ACM, 43*(4).

KEY TERMS

Best Practices: Generic business processes that are programmed into ERP software. They are based on cumulative knowledge about widely accepted practices that have been seen to work effectively across organizations in generic industries.

Customer Relationship Management (CRM): An approach that manages in an integrated manner all business processes that directly involve customers, in an effort to build long-term and sustainable relationships with customers. A CRM system is a central repository of customer information that records information from all contact points with customers, and generates "customer profiles" available to everyone who wishes to "know the customer".

Developing Nations: Those nations, particularly in Asia and Latin America, which are not as industrially developed as countries in Europe and North America. They are characterized by underdeveloped and uneven infrastructure in areas of telecommunications, roads, transportation and electricity distribution. Other features of such societies include low-cost labor and a large portion of the population living in rural areas and employed in the agriculture sector.

Enterprise Resource Planning (ERP) Systems: Integrated systems that support a wide range of transaction processing functions common to most organizations. Applications are designed around modules, which are linked to a central database in such a manner that data entered once through one module are available to applications of all other modules.

Enterprise Software Configuration: The process of implementing customized information flows, business logic and database design in an ERP software. This requires activities such as designing forms, reports and screen layouts, modifying existing programs, writing code for add-on applications, designing and implementing databases, and converting data from legacy systems.

Process Re-engineering: An exercise by which organizational processes are radically re-designed and changed. Instead of automating existing processes, organizations are encouraged to question existing practices, eliminate unnecessary steps in the workflow, streamline activities and introduce a paradigm shift in the organization.

Small and Medium Enterprises: Companies that have an annual turnover of less than $6 million and less than 250 to 300 employees.

Supply Chain Management (SCM): The integration of all activities in the value chain that provide products, services and value to customers. These activities include purchasing, materials management, production planning and control, inventory control and distribution and delivery.

ENDNOTES

[1] According to figures published by the National Association for Software Companies, which is a body that analyzes and directs the growth and competence areas of the Indian software industry.

[2] The key ERP vendors in the Indian industry include SAP, Peoplesoft, Oracle and Baan.

Essentials of Functional and Object-Oriented Methodology

Peretz Shoval
Ben-Gurion University, Israel

Judith Kabeli
Ben-Gurion University, Israel

INTRODUCTION

Background on Traditional Approach to Information System Development

Many paradigms for system analysis and design have been proposed over the years. Early approaches have advocated the functional approach. Common methodologies that support this approach are SSA and SSD (DeMarco, 1978; Yourdon & Constantine, 1979). SSA is based on the use of data flow diagrams (DFDs), which define the functions of the system, the data stores within the system, the external entities, and the data flows among these components. Early SSA and similar methodologies emphasized the functional aspects of system analysis, neglecting somehow the structural aspects, namely the data model. This was remedied by enhancing those methodologies with a conceptual data model, usually the entity-relationship (ER) model (Chen, 1976), that is used to create a diagram of the data model, which is later mapped to a relational database schema.

SSD is based on the use of structure charts (SCs), which describe the division of the system to program modules as well as the hierarchy of the different modules and their interfaces. Certain techniques have been proposed to create SCs from DFDs (Yourdon & Constantine, 1979). The main difficulty of an approach where functional analysis is followed by structured design lies in the transition from DFDs to SCs. In spite of various guidelines and rules for conversion from one structure to the other, the problem has not been resolved by those methodologies (Coad & Yourdon, 1990).

Shoval (1988, 1991) developed the ADISSA methodology that solved this problem. It uses hierarchical DFDs during the analysis stage (similar to other functional analysis methodologies), but the design centers on *transactions* design. A transaction is a process that supports a user who performs a business function, and is triggered as a result of an event. Transactions will eventually become the application programs. Transactions are identified and derived from DFDs: A transaction consists of elementary functions (namely functions which are not decomposed into sub-functions) that are chained through data flows; and of data stores and external entities that are connected to those functions. A transaction includes at least one external entity, which serve as its trigger. The process logic of each transaction is defined by means of structured programming techniques, for example pseudo-code. Based on the hierarchical DFDs and the transactions, ADISSA provides structured techniques to design the user-system interface (a menus-tree), the inputs and outputs (forms and reports), the database schema, and detailed descriptions of the transactions, which will eventually become the application programs. The menus-tree is derived from the hierarchy of DFDs in a semi-algorithmic fashion, based on functions that are connected to user-entities. The design of the forms and reports is based on data flows from user-entities to elementary functions and from elementary functions to user-entities. The design of the relational database schema is based on the analysis of dependencies among the data elements within the data-stores. The data flows from elementary functions to data stores and from data stores to elementary functions serve as a basis for defining access steps, namely update and retrieval operations on the relations. Access steps are expressed as SQL statements that will be embedded in the program code of the respective transactions. The products of the design stages can be easily implemented using various programming environments.

Background on Object-Oriented Approach to Information System Development

The development of object-oriented (OO) programming languages gave rise to the OO approach and its penetration into system analysis and design. Many OO methodologies have been developed in the early 90s (e.g., Booch, 1991; Coad & Yourdon, 1990, 1991; Jacobson, 1992; Rumbaugh, Blaha, Premerlani, Eddy & Lorensen, 1991;

Shlaer & Mellor, 1992; Wirfs-Brock, 1990). In the OO approach the world is composed of objects with attributes (defining its state) and behavior (methods), which constitute the only way by which the data included in the object can be accessed. When using the OO approach, a model of the system is usually created in the form of a class diagram consisting of data classes with structural relationships between them (e.g., generalization-specialization), and each class having its attributes and methods.

While there are no doubts about the advantages of the OO approach in programming, as it supports information hiding (encapsulation), software reuse and maintenance, there are doubts with respect to the effectiveness of the approach for analyzing business-oriented information systems (as opposed to real-time systems). The early OO methodologies tended to neglect the functionality aspect of system analysis, and did not show clearly how to integrate the application functions (transactions) with the class diagram. Another difficulty with those methodologies was that they involved many types of non-standard diagrams and notations.

The multiplicity of diagram types in the OO approach has been a major motivation for developing the UML (Booch et al., 1999; Clee & Tepfenhart, 1997; Fowler, 1997; Larman, 1998; Maciaszek, 2001; UML-Rose, 1998). UML provides a standard ("unified") modeling language. It

Figure 1. The initial class diagram of Music Programs system

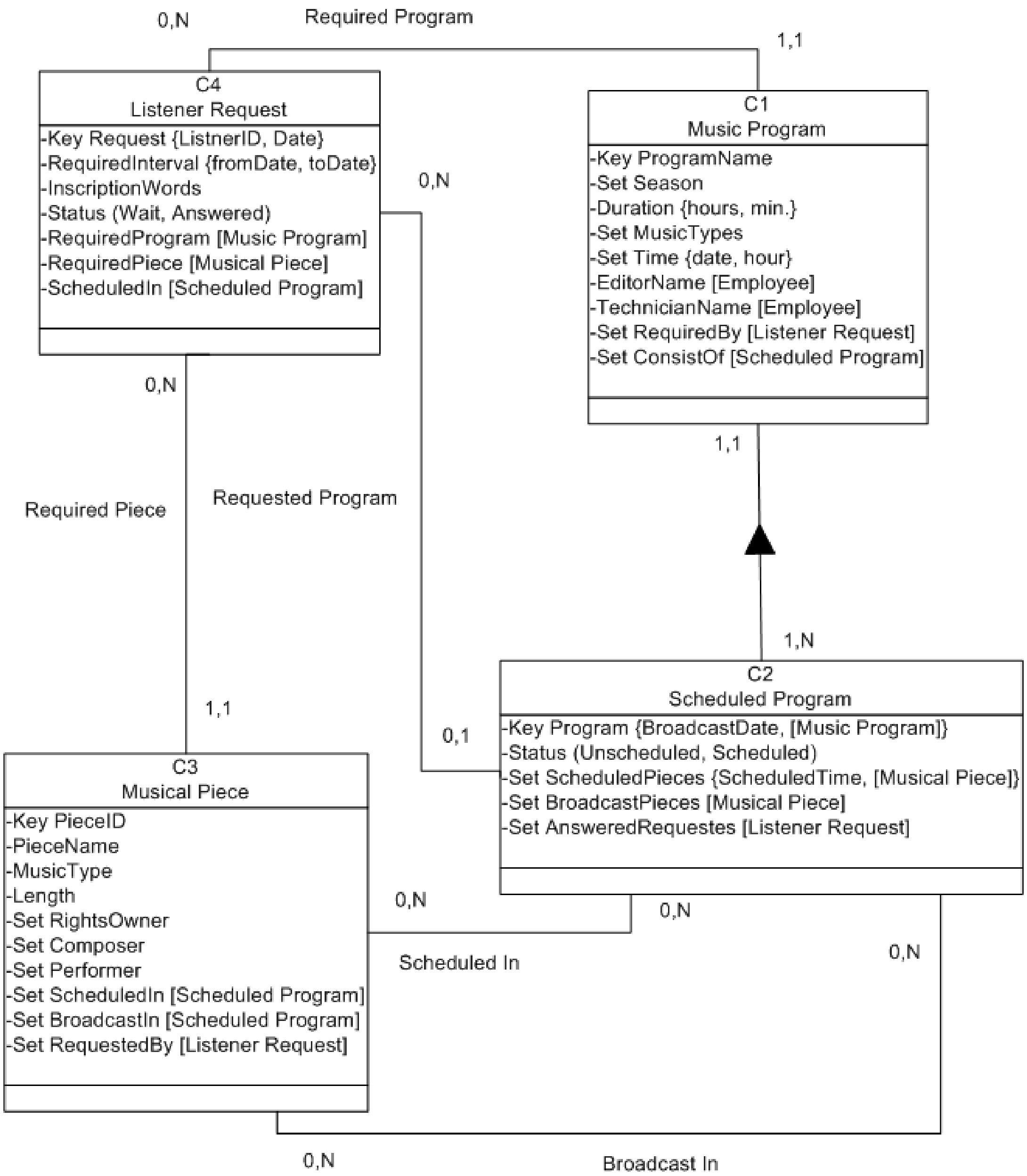
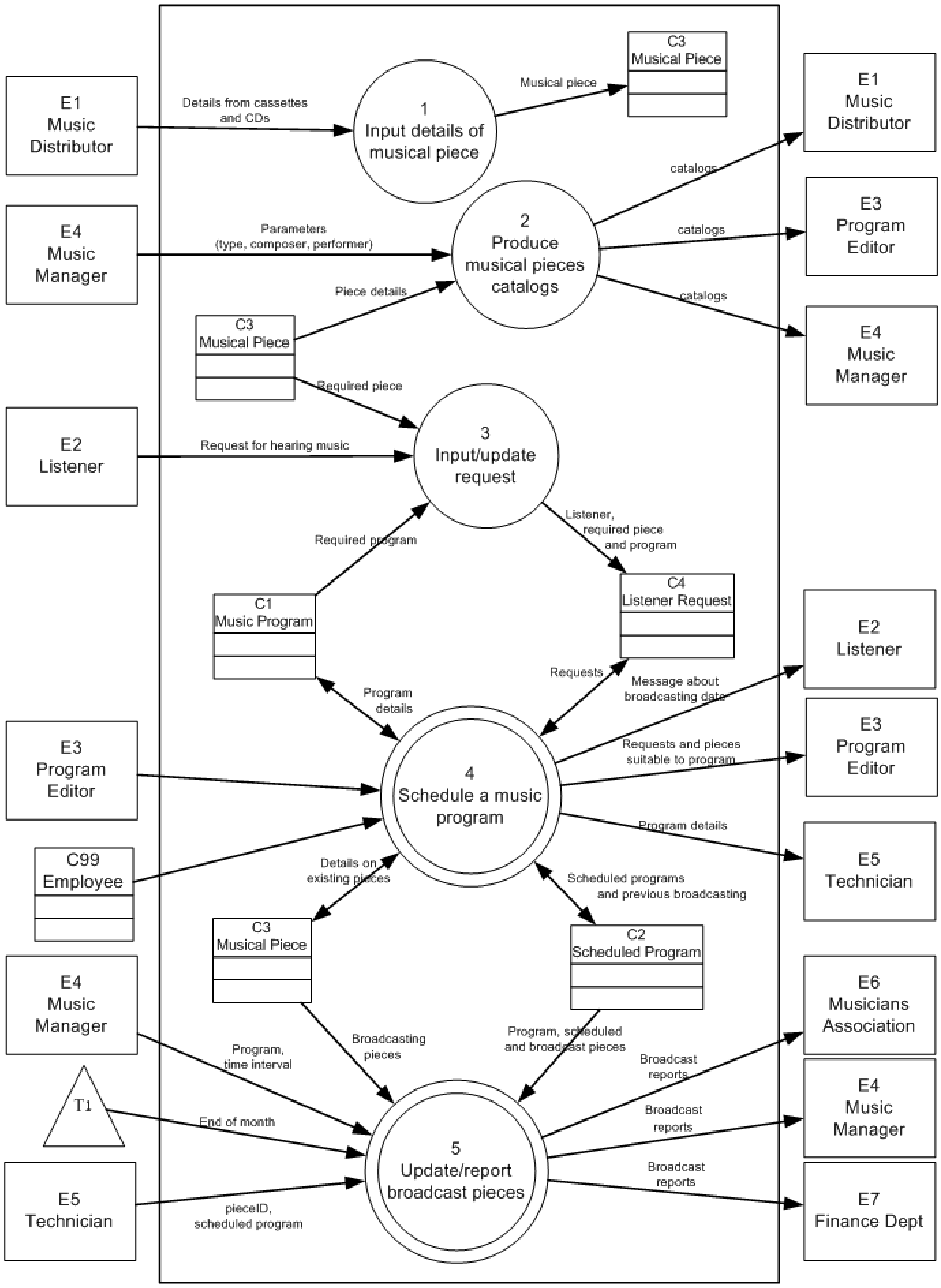


Figure 2. The top-level OO-DFD of Music Programs system



consists of several types of diagrams with well-defined semantics and syntax, which enable a system to be presented from different points of view. But UML does not offer a development methodology that guides the developer on *how* to do and which techniques to use. Moreover, some of the diagram types are redundant, and it is not always clear if and when they should be used in system development (e.g., sequence diagrams vs. collaboration diagrams).

Our approach is to integrate the two paradigms of system development. In our view, functions are as fundamental as objects, complementing each other, and a development methodology should combine the functional and the object-oriented approaches. FOOM is a system analysis and design methodology that presents this combination.

ESSENTIALS OF FOOM METHODOLOGY

FOOM is an integrated methodology for the development of business-oriented information systems, which combines the functional and the OO approaches. The essential steps of the methodology are as follows (for more details see Shoval & Kabeli, 2001):

The Analysis Phase

The analysis phase consists of two main activities: data analysis and functional analysis. The products of this stage are a data model, in the form of an initial class diagram, and a functional model, in the form of hierarchical OO-DFDs.

The initial class diagram consists of data (entity) classes, namely classes that are derived from the application/user requirements and contain "real world" data. Each class includes attributes of various types (e.g., atomic, multi-valued, sets, and reference attributes). Association types between classes include "regular" relationships, generalization-specialization (inheritance) links between super and subclasses, and aggregation-participation (part-of) links. Relationships are signified by links between respective classes, and by reference attributes to those classes. The initial class diagram does not include methods; these will be added at the design phase. An example of an initial class diagram is shown in Figure 1.

The OO-DFDs specify the functional requirements of the system. Each OO-DFD consists of general or elementary functions, external entities – mostly user-entities, but also time and real-time entities, data classes (instead of the traditional data stores), and data flows among them. An example of an OO-DFD is shown in Figure 2; it is the top-level OO-DFD. For each general function, signified by a double circle, there is a separate OO-DFD that details its sub-functions and related entities and classes (not shown).

Data analysis and functional analysis may be performed in any order. When starting with functional analysis, the analyst elicits the user requirements and based on that creates the OO-DFDs, which include (besides other components) data class rather than data stores. Then the analyst can create an initial class diagram – using the classes already appearing in the OO-DFDs. This means mainly defining proper class associations and attributes. Alternatively, the analyst may first create an initial class diagram (based on user requirements), and then create OO-DFDs, using the already defined classes. We investigated the pros and cons of the two alternative orders in an experimental setting, and found out that an analysis process that starts with data analysis provides better products and is preferred by analysts (Kabeli & Shoval, 2003). At any rate, the two product of the analysis stage are synchronized so that all classes appearing in the class diagram appear also in the OO-DFDs, and vise versa.

The Design Phase

Top-Level Design of Transactions

This stage is performed according to ADISSA methodology, where the application transactions are derived from DFDs. The products of this stage include transactions diagrams, as extracted from the OO-DFDs, and top-level descriptions of the transactions. In addition, a new class - Transactions – is added to the class diagram. This abstract class does not contain objects – it will contain transaction methods (which will be designed from the transactions' descriptions). Figure 3 shows an example of a "simple" transaction consisting of an elementary-function, a class and a user-entity; this transaction is derived from OO-DFD (Figure 2).

A top-level transaction description is provided in Structured-English (pseudo-code), and it refers to all components of the transaction: every data-flow from or to an external entity is translated to an "Input from..." or "Output to..." command; every data-flow from or to a class is translated to a "Read from..." or "Write to..." command; every data flow between two functions is translated to a "Move from... to..." command; and every function in the transaction is translated into an "Execute function.." command. The process logic of the transaction is expressed by standard process-logic patterns (e.g., if... then... else...; do-while...). The analyst and the user, who presents the requirements, determine the process logic of each transaction. This cannot be deducted automatically

Figure 3. A "simple" transaction diagram

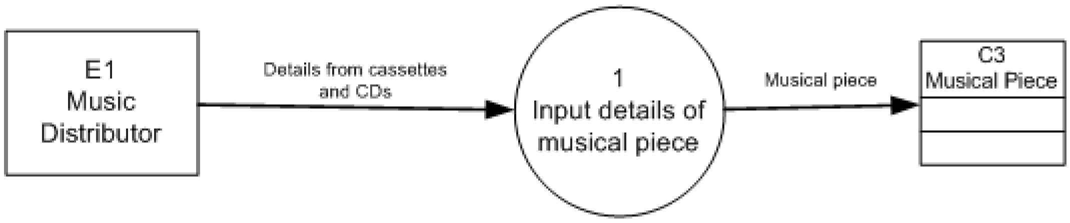


Figure 4. Top-level description of the transaction

Begin transaction_1
/ Input details of a Musical Piece */*
1. Input from E1: Music Distributor – details from cassettes and CDs.
2. Execute Function 1 – Input details of musical piece
3. Write to Class C3: Musical Piece – details of musical piece.

End transaction.

from the transaction diagram alone, because a given diagram can be interpreted in different ways, and it is up to the user to determine its proper semantics. The top-level description of this transaction is shown in Figure 4. This description will be used in further stages of design, namely input/output design and behavior design, to provide detailed descriptions of the application-specific class methods.

Design of the Interface – the Menus Class

As in ADISSA, the menus tree is derived in a semi algorithmic way from the hierarchy of OO-DFDs (Shoval, 1990). An example of a menus tree that is derived from our example is shown in Figure 5. The main menu contains five lines/items; three marked by "T" (for "trigger"), indicating lines that trigger transactions; and two marked by "S" (for "selection"), indicating lines that call other menus. (The numbers next to each line indicate the functions included in the transaction being triggered by the respective menu line). The menus tree is translated into a new class – Menus; the instances (objects) of this class are the individual menus. Note that at run time, a user who interacts with the menu of the system actually works with a certain menu object. He/she may select a menu line that will cause the presentation of another menu object, or invoke a transaction, which is a method of the Transactions class.

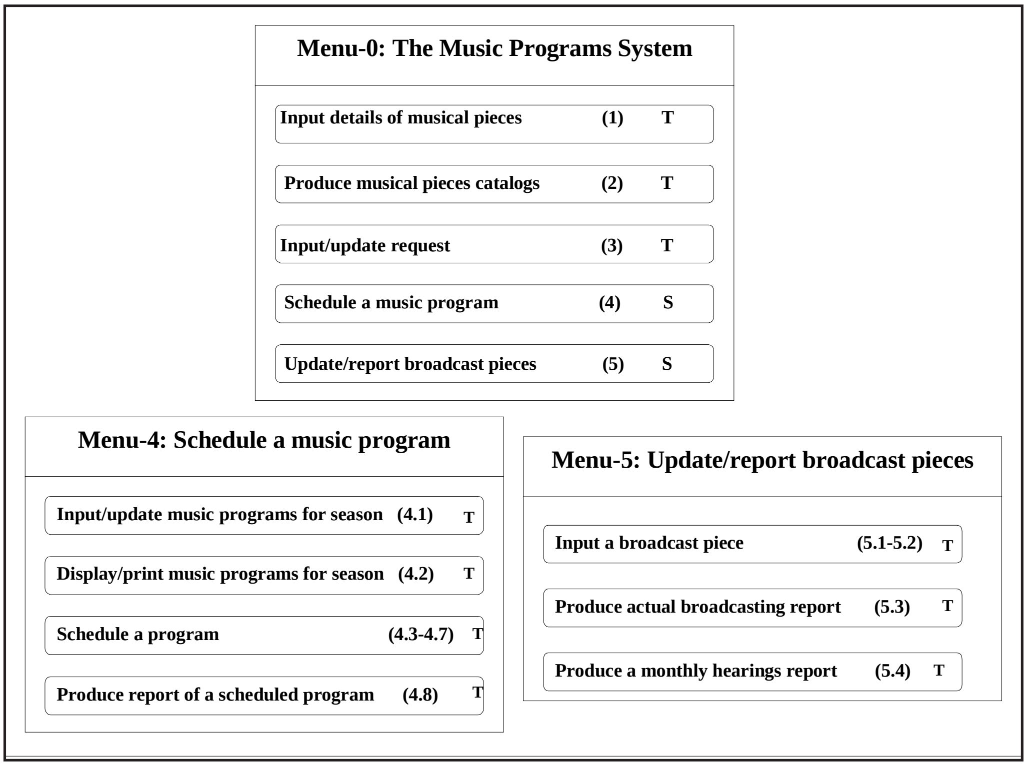
Design of the Inputs and Outputs – The Forms and Reports Classes

This stage is also performed according to ADISSA and is based on the input and output commands appearing in each of the transaction descriptions. For each "Input from..." command, an input screen/form will be designed, and for each "Output to..." command an output screen/report will be designed. Depending on the process logic of each transaction, some input or output screens may be combined. Eventually, two new classes are added to the class diagram: Forms - for the inputs; and Reports - for the outputs. Obviously, the instances (objects) of each of these classes are the input screens and output screens/reports, respectively.

Design of the System Behavior

In this stage, the top-level descriptions of the transactions are converted into detailed descriptions of application programs and class methods. The transition from a top-level description of a transaction to detailed descriptions of methods is done as follows: Every "Input from..." and "Output to..." command in the top-level description is translated to a message calling an appropriate method of the Forms or Reports class. Every "Read from..." or "Write to..." command is translated to a message calling a "basic" method of the appropriate class. Basic methods include Create, Read, Update and Delete (CRUD), and are assumed to exist in every class. Every "Execute-Function..." command can be translated to a basic method of certain classes, or to a "specific" method that will be attached to a proper class. A specific method may perform a procedure/function that is specific to the application/transaction, beyond any basic method (e.g., comparisons, computations). Each procedure that is defined as a basic or specific method of some class is removed from the transaction's description and replaced by a message to that class-method. A procedure/function that encom-

Figure 5. Interface design – the Menus tree



passes several classes must not be defined as a specific method of a certain class; rather it may remain part of the transaction. The remaining parts of the transaction's description are defined as "transaction method". This method is actually the "main" part of the transaction's program, and belongs to the Transactions class. Hence, when a user asks the system to perform some program/transaction, he/she fires the "main" method of the transaction (via proper menus selections); that method executes, and may call (send messages to) other methods (basic or specific) of respective classes – according to the process logic of the transaction.

Each transaction method or specific method can be described in two complementing forms: pseudo-code and message chart. Figure 6 shows the pseudo-code of the transaction. This description must not be too "formal"; it should include comments that will clarify the transaction to the programmer who will implement it in the proper programming language. A message chart (Figure 7) shows the classes, methods and messages included in a method, and the order/process logic of their execution. A message chart is actually a combination of a program flowchart and class diagram that includes only the classes involved in the method. Message charts supplement the pseudo-code; they must not be created for all transactions.

To summarize, the products of the design phase are: a) a complete class diagram, including Data, Menus, Forms, Reports and Transactions classes, each with various attribute types and method names (and parameters), and various associations among the classes; b) detailed menu objects of the Menus class; c) detailed form and report objects of the Forms and Reports classes; d) detailed descriptions of the transaction and specific methods, in pseudo-code and/or in message charts. At the implementation stage, the programmers will use these design products to create the software with any common OO programming language.

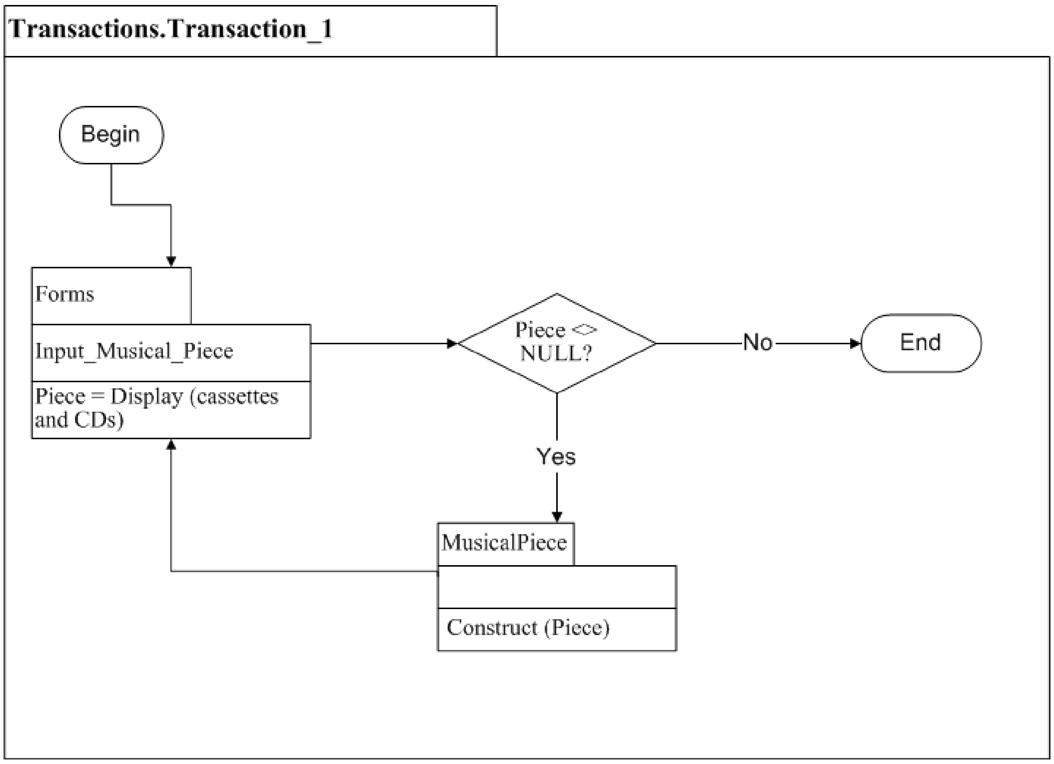
CONCLUSION

The advantages of the FOOM methodology are: (a) System analysis is performed in functional terms via OO-DFDs - a natural way for users to express their information needs, and in data terms via an initial class diagram. (b) System design follows the analysis and uses its products. The class diagram is augmented with a Menus class that is derived from the menus designed earlier from the OO-DFDs. Inputs and Outputs classes are also derived from the input forms and the outputs of the system (earlier

Figure 6. Pseudo-code of the transaction

```
Begin Transaction_1
   Repeat
      Piece=Forms.Input_Musical_Piece.Display (cassettes and CDs)
         /* Input_Musical_Piece is a form, which enable the user to fills in details of a musical piece. When
                                   the user completes the form it returns the filled piece object */
      If Piece <> NULL then        /* if the user selects to add a piece*/
               Musical_Piece.Construct(Piece)    /*Construct an object of Musical Piece*/
      End if
   Until Piece = NULL                    /* the user selects to stop adding pieces
End
```

Figure 7. Message chart of the transaction



products of the design stage). The transactions are defined as basic methods, specific methods and transaction methods. (c) The end products of the design phase can be easily implemented with any OO programming environment.

FOOM methodology is supported by a set of CASE tools. These include tools that enable creating and checking the correctness of class diagrams, hierarchical OO-DFDs and the menus of the application, and identifying transactions from DFDs. Other tools, which are under development, will enable designing the forms and reports, and creating pseudo-code and message charts of the methods – as based on the transactions.

REFERENCES

Booch, G. (1991). *Object-oriented design with applications*. Benjamin/Cummings.

Booch, G., Rumbaugh, J., & Jacobson, I. (1999). *The Unified Modeling Language user guide*. Addison-Wesley.

Chen, P. (1976). The entity-relationship model: Toward a unified view of data. *ACM Transactions on Database Systems, 1*(1), 9-36.

Clee, R., & Tepfenhart, W. (1997). *UML and C++ A practical guide to object-oriented development*. Prentice Hall.

Coad, P., & Yourdon, E. (1990). *Object-oriented analysis*. Englewood Cliffs, NJ: Prentice Hall.

Coad, P., & Yourdon, E. (1991). *Object-oriented design*. Englewood Cliffs, NJ: Prentice Hall.

DeMarco, T. (1978). *Structured analysis and system specification*. NY: Yourdon Press.

Fowler, M. (1997). *UML distilled*. Addison-Wesley.

Jacobson, I. (1992). *Object-oriented software engineering: A use case driven approach*. ACM Press.

Kabeli, J., & Shoval, P. (2003). Data modeling or functional analysis: What comes next? An experimental comparison using FOOM methodology. *Proc. of the 8th CAISE/IFIP8.1 Int'l Workshop on Evaluation of Modeling Methods in Systems Analysis and Design (EMMSAD'03)* (pp. 48-57).

Larman, C. (1998). *Applying UML and patterns- an introduction to object oriented analysis and design*.

Maciaszek, L.A. (2001). *Requirements analysis and system design – developing information systems with UML*. Addison-Wesley.

Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F., & Lorensen, W. (1991). *Object-oriented modeling and design*. Englewood Cliffs, NJ: Prentice Hall.

Shlaer, S., & Mellor, S. (1992). *Object life cycles: Modeling the world in states*. Englewood Cliffs, NJ: Yourdon Press.

Shoval, P. (1988). ADISSA: Architectural Design of Information Systems based on Structured Analysis. *Information System*, *13*(2), 193-210.

Shoval, P. (1990). Functional design of a menu-tree interface within structured system development. *Int'l Journal of Man-Machine Studies*, *33*, 537-556.

Shoval, P. (1991). An integrated methodology for functional analysis, process design and database design. *Information Systems*, *16*(1), 49-64.

Shoval, P., & Kabeli, J. (2001). FOOM: Functional- and object-oriented analysis & design of information systems - an integrated methodology. *Journal of Database Management*, *12*(1), 15-25.

UML Rose. (1998). *http://www.rational.com*

Wirfs-Brock, R., Wilkerson, B., & Wiener, L. (1990). *Designing object-oriented software*. Englewood Cliffs, NJ: Prentice Hall.

Yourdon, Y., & Constantine, L.L. (1979). *Structured design*. Englewood Cliffs, NJ: Prentice Hall.

KEY TERMS

ADISSA - Architectural Design of Information Systems based on Structured Analysis: An analysis and design methodology. In the analysis stage it utilizes hierarchical DFDs. In the design stage the DFDs are used to design the various components of the system. These include: a) top-level descriptions of the transactions, which eventually become detailed descriptions of the applications programs; b) the user interfaces (menus); c) the input and output screens and reports; d) the database schema in the form of normalized relations, and SQL commands for retrieving and updating the database.

DFD - Data Flow Diagram: A diagram used in functional analysis that specifies the functions of the system, the inputs/outputs from/to external (user) entities, and the data being retrieved from or updating data stores. There are well-defined rules for specifying correct DFDs as well as for creating hierarchies of interrelated DFDs.

ER - Entity-Relationship: A conceptual data model that defines the domain in terms of entities, attributes and relationships. ERD is an ER diagram, in which entities are represented as rectangles, attributes as ellipses and relationships between entities as diamonds.

OO-DFD – Object-Oriented DFD: A variant of DFDs introduced in FOOM methodology, which include object (data) classes rather than data stores.

SSA - Structured System Analysis: A traditional, functional-oriented methodology for analyzing information systems, which utilized data flow diagrams.

SSD - Structured System Design: A traditional methodology for designing information systems, which utilized structure charts.

SQL – Structured Query Language: A standard language for querying and updating a relational database.

UML – Unified Modeling Language: An "industrial standard" notation for object-oriented development. It consists of a several types of (mostly) diagrams that enable describing systems from different perspectives, including managerial, structural and behavioral (functional).

Ethical Implications of Investigating Internet Relationships

Monica T. Whitty
Queen's University Belfast, Northern Ireland

INTRODUCTION

The focus on Internet relationships has escalated in recent times, with researchers investigating such areas as the development of online relationships (e.g., McCown, Fischer, Page & Homant, 2001; Parks & Roberts, 1998; Whitty & Gavin, 2001), the formation of friends online (Parks & Floyd, 1996), representation (Bargh, McKenna & Fitzsimons, 2002), and misrepresentation (Whitty, 2002) of self online. Researchers have also attempted to identify those addicted to accessing online sexual material (Cooper, Putnam, Planchon & Boies, 1999). Moreover, others have been interested in Internet infidelity (Whitty, 2003a) and cybersex addiction (Griffiths, 2001; Young, Griffin-Shelley, Cooper, O'Mara & Buchanan, 2000). Notwithstanding this continued growth of research in this field, researchers have neglected to consider the more ethical ways in conducting this type of research.

While it is acknowledged here that some of the discussions in this article might be equally applied to the study of other Internet texts, such as religious or racial opinions, the focus in this article is on the concomitant ethical concerns of ongoing research into Internet relationships. Given that the development and maintenance of online relationships can be perceived as private and very personal (possibly more personal than other sensitive areas), there are potential ethical concerns that are unique to the study of such a topic area (Whitty, 2004). For a broader discussion of virtual research ethics, refer to Ess and Jones (2004).

REVIEW OF RESEARCH ON INTERNET RELATIONSHIPS AND SEXUALITY

Early research into this area has mostly focused on the similarities and differences between online and offline relationships. Researchers have been divided over the importance of available social cues in the creation and maintenance of online relationships. Some have argued that online relationships are shallow and impersonal (e.g., Slouka, 1995). In contrast, others contend that Internet relationships are just as emotionally fulfilling as face-to-face relationships, and that any lack of social cues can be overcome (Lea & Spears, 1995; Walther, 1996). In addition, researchers have purported that the ideals that are important in traditional relationships, such as trust, honesty, and commitment, are equally important online, but the cues that signify these ideals are different (Whitty & Gavin, 2001). Current research is also beginning to recognize that for some, online relating is just another form of communicating with friends and lovers, and that we need to move away from considering these forms of communication as totally separate and distinct entities (e.g., Vayreda, Galvez, Nunez & Callen, 2002). Moreover, McKenna, Green, and Gleason (2002) have found that when people convey their 'true' self online, they develop strong Internet relationships and bring these relationships into their 'real' lives.

Internet friendships developed in chat rooms, newsgroups, and MUDs or MOOs have been examined by a number of researchers. For example, Parks and Floyd (1996) used e-mail surveys to investigate how common personal relationships are in newsgroups. After finding that these relationships were regularly formed in newsgroups, Parks and Roberts (1998) examined relationships developed in MOOs. These researchers found that most (93.6%) of their participants had reported having formed some type of personal relationship online, the most common type being a close friendship.

Researchers have also been interested in how the playful arena of the Internet impacts on the types of relationships formed in these places (e.g., Whitty, 2003b; Whitty & Carr, 2003). Turkle's (1995) well-known research on her observations while interacting in MUDs found that the role-playing aspect of MUDs actually creates opportunities for individuals to reveal a deeper truth about themselves. Whitty and Gavin (2001) have also contended that although people do lie about themselves online, this paradoxically can open up a space for a deeper level of engagement with others.

Cybersex addiction and the available treatment for these cybersex addicts and their partners has been an area of research and concern for psychologists (e.g., Schneider, 2000; Young, Pistner, O'Mara & Buchanan, 1999). Research has also focused on what online acts might be considered as an act of infidelity. For example, Whitty (2003a) found that acts such as cybersex and hot-chatting were perceived as almost as threatening to the offline relationship as sexual intercourse. In addition to the above concerns, Cooper et al. (1999) identified three categories of individuals who access Internet erotic material, including recreational users, sexual compulsive users (these individuals are addicted to sex *per se*, and the Internet is but one mode where they can access sexual material), and at-risk users (these individuals would never have developed a sexual addiction if it were not for the Internet).

ETHICAL ISSUES PERTINENT TO THE STUDY OF INTERNET RELATIONSHIPS

Much of the research, to date, on Internet relationships and sexuality has been conducted online—either through interviews and surveys, or by carrying out analysis on text that is readily available online. There are many advantages to conducting research online, as well as collecting text or data available online for analysis in one's research (see Table 1).

In spite of the numerous advantages to conducting research online, investigators also need to be aware of the disadvantages (see Table 2).

What all studies that research Internet relationships have in common is that they are researching a sensitive topic, which requires individuals to reveal personal and often very private aspects of themselves and their lives. Given the sensitive nature of this topic area, it is crucial that researchers give some serious thought to whether they are truly conducting research in an ethical manner.

A common form of using the Internet for researching online relationships is to analyze the text produced by people online. The text can be produced in a number of different forums, including chat rooms, MUDs, newsgroups, and online dating sites. One way researchers collect data is by lurking in these different spaces in cyberspace. The development of online relationships (both friendships and romantic) and engaging in online sexual activities, such as cybersex, could easily be perceived by those engaging in such activities as a private discourse. Given the nature of these interactions, social researchers need to seriously consider if they have the right to lurk in online settings in order to learn more about these activities—despite the benefits of obtaining this knowledge.

There are fuzzy boundaries between what constitutes public and private spaces online, and researchers need to acknowledge that there are different places online. For example, a chat room might be deemed a more public space than e-mail. It is contended here that lurking in some spaces online might be ethically questionable. We must, as researchers, debate how intrusive a method lurking potentially is. As Ferri (1999, cited in Mann & Stewart, 2000) contends: "Who is the intended audience of an electronic communication—and does it include you as a researcher?" (p. 46).

Researchers also need to consider how the participant perceives the spaces online. As Ferri suggests, private interactions can and do indeed occur in public places. It has been theorized that the Internet can give an individual a sense of privacy and anonymity (e.g., Rice & Love, 1987). The social presence theory contends that "social presence" is the feeling that one has that other persons are involved in a communication exchange (Rice & Love, 1987). Since computer-mediated relating (CMR) involves less non-verbal cues (such as facial expression, posture, and dress) and auditory cues in comparison to face-to-face communication, it is said to be extremely low in social presence. Hence, while many others might occupy the space online, it is not necessarily perceived in that way. As researchers, we need to ask: Can researchers ethically take advantage of these people's false sense of privacy and security? Is it ethically justifiable to lurk in these sites and download material without the knowledge or consent of the individuals who inhabit these sites? This is especially relevant to questions of relationship development and sexuality, which are generally understood to be private matters. Therefore, good ethical practice needs to consider the psychology of cyberspace and the false sense of security the Internet affords.

Table 1. Practical benefits of conducting research online

-	Easy access to a population of individuals who form relationships online and who access sexual material
-	Internet provides researchers with a population that is sometimes difficult to research (e.g., people with disabilities, agoraphobia)
-	Contact people in locations that have closed or limited access (e.g., prisons, hospitals)
-	Requires relatively limited resources
-	Ease of implementation

Table 2. Disadvantages of conducting research online

- Security issues
- Possible duplication of participants completing surveys
- Difficult to ascertain how the topic area examined impacts on the participant
- Restricted to a certain sample

It is suggested here that researchers need to maintain personal integrity as well as be aware of how their online investigations can impact on the Internet relationships they study. For example, given researchers' knowledge of online relationships, interacting on online dating sites, chat rooms, and so forth could potentially alter the dynamics of these communities.

While it might be unclear as to how ethical it is for lurkers to collect data on the Internet, there is less doubt as to whether it is acceptable to deceive others online in order to conduct social research, especially with respect to online relationships and sexuality. Ethical guidelines generally state that deception is unethical because the participant is unable to give free and fully informed consent. For example, according to the Australian National Health and Medical Research Council (NHMRC), which set the ethical guidelines for Australian research:

"…as a general principle, deception of, concealment of the purposes of a study from, or covert observation of, identifiable participants are not considered ethical because they are contrary to the principle of respect for persons[,] in that free and fully informed consent cannot be given." (NHMRC, 1999)

Generally, ethical guidelines will point out that only under certain unusual circumstances is deception unavoidable when there is no alternative method to conduct one's research. However, in these circumstances individuals must be given the opportunity to withdraw data obtained from them during the research that they did not originally give consent to.

PRACTICAL CONSIDERATIONS

As with any other research conducted within the social sciences, some important ethical practices need to be adhered to when we conduct research on Internet relationships and sexuality (see Table 3).

Informed consent requires researchers to be up front from the beginning about the aims of their research and how they are going to be utilizing the data they collect. In offline research, individuals often sign a form to give their consent; however, this is not always achievable online. One way around this is to direct participants to a Web site that contains information about the project. This Web site could inform the participants about the purpose of the study, what the study entails, as well as contact details of the researcher and the university Human Ethics Committee.

In some cases, spaces on the Web are moderated. In these instances, it is probably also appropriate to contact the moderators of the site prior to contacting the participants. This is analogous to contacting an organization prior to targeting individuals within that organization. Researchers also need to be aware that some European countries require written consent. If written consent is required, then the participant could download a form and sign it offline, and return it by fax or postal mail (Mann & Stewart, 2000).

In research about relationships and sexuality, in particular, there is the risk that the interview or survey has created too much stress for the participant to continue. As with offline research, we need to consider up until what point a participant can withdraw consent. The end point of withdrawal of consent might be, for instance, after the submitting of the survey or at the conclusion of the interview, the interviewer might find confirmation that the participant is happy to allow the researcher to include the transcript in the study. Social scientists should also be aware that the lack of social cues available online makes it more difficult for them to ascertain if the participant is uncomfortable. Thus one should tread carefully and possibly make an effort to check at different points in the interview if the individual is still comfortable with proceeding.

There are other issues unique to Internet research with respect to withdrawal of consent. For example, the computer could crash midway through an interview or survey. Mechanisms need to be put into place to allow participants to rejoin the research if desired, and consent should not be assumed. In circumstances such as the computer or server crashing, we might need to have a system to enable debriefing, especially if the research is asking questions of a personal nature. Nosek, Banaji, and Greenwald (2002) suggest that debriefing can be made available by providing a contact e-mail address at the beginning of the study. They also suggest providing "a 'leave the study' button, made available on every study page, [which] would allow participants to leave the study early and still direct them to a debriefing page" (p. 163). In

Table 3. Ethical practices

- Informed consent
- Withdrawal of consent
- Confidentiality
- Psychological safeguards

addition, they suggest that participants be given a list of frequently asked questions (FAQs), since, they argue, there is less opportunity to ask the sorts of questions participants typically ask in face-to-face interviews.

There are various ways we might deal with the issue of confidentiality. As with offline research we could elect to use pseudonyms to represent our participants or even request preferred pseudonyms from them. However, a unique aspect of the Internet is that people typically inhabit the Web using a screen name, rather than a real name. Can we use a screen name given that these are not real names? While they may not be people's offline identities, individuals could still be identified by their screen names if we publish them—even if it is only recognition by other online inhabitants.

As mentioned earlier in this article, research into the areas of relationships and sexuality is likely to cause psychological distress for some. It is perhaps much more difficult to deal with psychological distress online and with individuals in other countries. Nevertheless, it is imperative that we ensure that the participant does have counseling available to them if the research has caused them distress—which sometimes might be delayed distress. This could mean that there are limits to the kinds of topics about which we interview participants online, or that we restrict our sample to a particular country or region where we know of psychological services that can be available to our participants if required.

Given that research into Internet relationships and sexuality is a relatively new area, future research might also focus on how to improve ethical practices. For instance, future studies might interview potential participants about how they would prefer social scientists to conduct research. Moreover, gaining a greater understanding of how individuals perceive private and public space could also influence how we conduct future studies in this topic area.

CONCLUSION

In concluding, while this article has provided examples of ways forward in our thinking about virtual ethics in respect to the study of online relationships, it is by no means prescriptive or exhaustive. Rather, it is suggested here that debate over such issues should be encouraged, and we should avoid setting standards for how we conduct our Internet research without also considering the ethical implications of our work. The way forward is to not restrict the debate amongst social scientists, but to also consult the individuals we would like to and are privileged to study.

REFERENCES

Bargh, J.A., McKenna, K.Y.A., & Fitzsimons, G.. (2002). Can you see the real me? Activation and expression of the "true self" on the Internet. *Journal of Social Issues, 58*(1), 33-48.

Cooper, A., Putnam, D.E., Planchon, L.A., & Boies, S.C. (1999). Online sexual compulsivity: Getting tangled in the Net. *Sexual Addiction & Compulsivity, 6*(2), 79-104.

Ess, C., & Jones, S. (2004). Ethical decision-making and Internet research: Recommendations from the AoIR Ethics Working Committee. In E. Buchanan (Ed.), *Readings in virtual research ethics: Issues and controversies* (pp. 27-44). Hershey, PA: Idea Group Publishing.

Griffiths, M. (2001). Sex on the Internet: Observations and implications for Internet sex addiction. *Journal of Sex Research, 38*(4), 333-342.

Lea, M., & Spears, R. (1995). Love at first byte? Building personal relationships over computer networks. In J.T. Wood & S.W. Duck (Eds.), *Understudied relationships: Off the beaten track* (pp. 197-233). Newbury Park, CA: Sage Publications.

Mann, C., & Stewart, F. (2000). *Internet communication and qualitative research: A handbook for researching online.* London: Sage Publications.

McCown, J.A., Fischer, D., Page, R., & Homant, M. (2001). Internet relationships: People who meet people. *CyberPsychology and Behavior, 4*(5), 593-596.

McKenna, K.Y.A., Green, A.S., & Gleason, M.E.J. (2002). Relationship formation on the Internet: What's the big attraction? *Journal of Social Issues, 58*(1), 9-31.

NHMRC. (1999). *National statement on ethical conduct in research involving humans.* Retrieved September 25, 2002, from *www.health.gov.au/nhmrc/publications/humans/part17.htm*

Nosek, B.A., Banaji, M.R., & Greenwald, A.G. (2002). E-research: Ethics, security, design, and control in psychological research on the Internet. *Journal of Social Issues, 58*(1), 161-176.

Parks, M.R., & Floyd, K. (1996). Making friends in cyberspace. *Journal of Communication, 46,* 80-97.

Parks, M.R., & Roberts, L.D. (1998). 'Making MOOsic': The development of personal relationships online and a comparison to their off-line counterparts. *Journal of Social and Personal Relationships, 15*(4), 517-537.

Rice, R.E., & Love, G. (1987). Electronic emotion: Socioemotional content in a computer mediated communication network. *Communication Research, 14,* 85-108.

Schneider, J.P. (2000). Effects of cybersex addiction on the family: Results of a survey. *Sexual Addiction & Compulsivity, 7,* 31-58.

Slouka, M. (1995). *War of the worlds: Cyberspace and the high-tech assault on reality.* New York: Basic Books.

Turkle, S. (1995). *Life on the screen: Identity in the age of the Internet.* London: Weidenfeld & Nicolson.

Vayreda, A., Galvez, A., Nunez, F., & Callen, B. (2002, October 13-16). Participating in an electronic forum: The difference gender makes. *Proceedings of Internet Research 3.0: Net/Work/Theory,* Maastricht, the Netherlands.

Walther, J.B. (1996). Computer-mediated communication: Impersonal, interpersonal and hyperpersonal interaction. *Communication Research, 23,* 3-43.

Whitty, M.T. (2002). Liar, liar! An examination of how open, supportive and honest people are in chat rooms. *Computers in Human Behavior, 18,* 343-352.

Whitty, M.T. (2003a). Pushing the wrong buttons: Men's and women's attitudes towards online and off-line infidelity. *CyberPsychology & Behavior, 6*(6), 569-579.

Whitty, M.T. (2003b). Cyber-flirting: Playing at love on the Internet. *Theory and Psychology, 13*(3), 339-357.

Whitty, M.T. (2004). Peering into online bedroom windows: Considering the ethical implications of investigating Internet relationships and sexuality. In E. Buchanan (Ed.), *Readings in virtual research ethics: Issues and controversies* (pp. 203-218). Hershey, PA: Idea Group Publishing.

Whitty, M.T., & Carr, A.N. (2003). Cyberspace as potential space: Considering the Web as a playground to cyberflirt. *Human Relations, 56*(7), 861-891.

Whitty, M. & Gavin, J. (2001). Age/sex/location: Uncovering the social cues in the development of online relationships. *CyberPsychology and Behavior, 4*(5), 623-630.

Young, K.S., Griffin-Shelley, E., Cooper, A., O'Mara, J., & Buchanan, J. (2000). Online infidelity: A new dimension in couple relationships with implications for evaluation and treatment. *Sexual Addiction & Compulsivity, 7,* 59-74.

Young, K.S., Pistner, M., O'Mara, J., & Buchanan, J. (1999). Cyber disorders: The mental health concern for the new millennium. *CyberPsychology & Behavior, 2*(5), 475-479.

KEY TERMS

Chat Room: A Web site, or part of a Web site, that allows individuals to communicate in real time.

Cybersex: Two or more individuals using the Internet as a medium to engage in discourses about sexual fantasies. The dialogue is typically accompanied by sexual self-stimulation.

Discussion Board/Group: A Web site, or part of a Web site, that allows individuals to post messages, but does not have the capacity for interactive messaging.

Hot-Chatting: Two ore more individuals engaging in discourses that move beyond light-hearted flirting.

Lurker: A participant in a chat room or a subscriber to a discussion group, listserv, or mailing list who passively observers. These individuals typically do not actively partake in the discussions that befall in these forums.

MUDs and MOOs: Multiple-user dungeons, or more commonly understood these days to mean multi-user dimension or domains. These were originally a space where interactive role-playing games could be played, very similar to Dungeons and Dragons.

Online Sexual Activity: Using the Internet for any sexual activity (e.g., recreation, entertainment, exploitation, education).

Screen Name: A screen name can be an individual's real name, a variation of an individuals' name, or a totally made-up pseudonym. Screen names are especially required on the Internet for applications such as instant messaging.

E

Ethics of New Technologies

Joe Gilbert
University of Nevada Las Vegas, USA

INTRODUCTION

Information processing has been done through telling stories, drawing on cave walls, writing on parchment, printing books, talking on telephones, sending messages via telegraphs, broadcasting on radio and television, processing data in computers, and now by instantaneous network dissemination. Since the mid-1990's, personal computers have been the instrument of choice for sending and receiving information, and for processing much of it. The technology is the latest in a long series, but social issues involved have not really changed. Issues of content (is it true? obscene?), ownership (whose picture/text/idea? whose parchment/telephone system/computer?), and impact (anti-government, anti-social, harmful to children) appear today just as they did hundreds or thousands of years ago.

BACKGROUND

New technologies enable people to do new things (send 20 copies of a memo at once) or to do old things in new ways, such as storing files (Freeman & Soete, 1997). Improvements in technology that are incremental do not usually introduce major social issues, but radical innovations frequently present new kinds of social opportunities and threats (Brown, 1997). Ethics is the branch of philosophy that studies interpersonal or social values and the rules of conduct that follow from them. Ethics deals with questions of how people should treat each other on a basic level (Berlin, 2000). It considers such issues as rights and duties and fairness or justice. Because ethics concerns itself with fundamental rules, its applications to specific new technologies might require both knowledge of the new technology and reasoning about its possible applications based on established principles of ethics (Burn & Loch, 2001; Halbert & Ingulli, 2002).

Philosophers have pondered and written about issues of ethics for thousands of years. Some of their writings on this subject continue to be read and debated generation after generation (LaFollette, 2000). Three basic approaches have been most common and most accepted in discussions of ethics.

- Utilitarianism maintains that the ethical act is the one that creates the greatest good for the greatest number of people.
- Rights and duties maintain that the ethical act is the one that acknowledges the rights of others and the duties which those rights impose on the actor.
- Fairness and justice hold that the ethical act is the one that treats similarly situated people in similar ways with regard to both process and outcome.

ETHICS AND TECHNOLOGY

John Stuart Mill and Jeremy Bentham are the two philosophers most closely associated with utilitarianism. This view of ethics puts a high value on results, and holds that we must consider whether and to what degree our actions will bring pain or pleasure not only to ourselves but to all others who will be impacted by what we do (Frey, 2000; Mill & Bentham, 1987). A utilitarian would argue that the harm done to many individuals and businesses by viruses and worms far outweighs any happiness brought to their authors, and thus creating and disseminating such code is unethical. Similarly, a utilitarian analysis of music file-sharing would consider whether widespread free file-sharing might result in composers and artists deciding that it is not in their financial interest to continue writing and performing music. If this result occurred, not only the composers and artists but also their listeners would end up suffering harm that might outweigh the good that they enjoy from free file-sharing. Finally, a utilitarian analysis would favor products and policies that increase the spread of computer literacy and availability, since the Internet can bring great good to its users and computer literacy and availability makes such use possible.

Many philosophers have written about rights and duties (Sumner, 1987). The basic idea of this approach is that individuals do have rights, and that these rights are, practically speaking, worthless unless someone or some group has a corresponding duty. Thus, if I have a right to privacy, you have a duty not to monitor my every move (Kelly & Rowland, 2000). There are four basic sources of rights, and we will consider each in turn.

Human rights are possessed by every human, simply by virtue of being human. Among these rights are the right to live (not to be randomly killed), to be told the truth, to own property, and to basic dignity (Ignatieff, 2001). Among these, the one that most often causes confusion is the right to be told the truth. Humans could not interact with each other in any meaningful way if lying and truth-telling were equally valid. Promises, contracts, and interpersonal relations all depend on the fact that the default setting for conversation is truth-telling. This does not mean that everyone always will or even should tell the whole truth all the time. It does, however, mean that we can and do start with an assumption of truth-telling (Bok, 1999). A right to property, whether physical or intellectual, means that others have a duty not to take or use my goods without compensating me.

Since property rights are human, they apply whether a given country's laws regarding such things as copyright and intellectual property are specific on a given issue or not. Music companies and movie studios, on behalf of individual artists, have a right to control and charge for distribution of their products. This right imposes a duty on individuals not to take such property without paying for it and recognizing the terms of distribution. Similarly, software companies have a right to charge for and control the distribution of their intellectual property. They paid programmers to develop a software product; others have a duty to respect the rights to this intellectual property.

Some rights are given to individuals by law. These citizen rights come by reason of membership in a community (nation, state, county, etc.). The right of citizens of the United States to free speech is not recognized by some other countries. Typically, dictatorships grant few citizen rights to those under their rule. These rights often coincide with human rights (right to live, to property, etc.) but frequently go beyond basic human rights. Copyright, as it exists in the United States, is not recognized equally in all countries. This is why it is important that the basic right to own property is a human right—it is valid whatever the laws of a particular jurisdiction.

A third source of rights is position. Policemen may apprehend and incarcerate suspected criminals. CEO's can speak for their companies on many issues. Purchasing agents can spend a company's money on goods or services within some limits. Managers can set rules for computer usage at work. People have these rights not just because they are human, or because they are particularly wise or knowledgeable, but because of the position they occupy. Since individuals have these rights, others have duties to respect the rights and follow their direction.

The fourth and final basic source of rights is by contract. Individuals or organizations can agree to contractual relations that create rights and impose duties that would not otherwise exist. If I agree to pay a certain amount of money each month in order to use an online service, I have a duty to pay and the service provider has a duty to make the service available to me under the terms of the contract.

The third basic approach to ethics is fairness and justice: it is ethical to be fair and unethical to be unfair. It is not fair that some individuals should purchase software and others obtain it free through sharing or piracy. It is fair for those who invest time, talent and money in producing software to be paid for the products resulting from their efforts and investments by all of those who use them, not just by some. Issues of fairness sometimes arise in the area of using computer technology for purposes of employee monitoring (Alder, 1998). In general terms, fairness involves treating similarly situated people in similar ways with regard to both process and outcome. However, justice is sometimes defined as equality, and at other times, as based on contribution, on needs and abilities, or on maximum freedom (Velasquez, 2002).

An issue that often arises in considering fairness and justice is the question of which individuals or groups are similarly situated. In the sense that all who access the Internet can view unrestricted sites, all who access the Internet are similarly situated. In the sense that some who access the Internet may choose to view pornography and others may choose not to (even inadvertently), we have at least two groups that are not similarly situated. Using this approach, one might argue against unrestricted availability of pornography on the Internet, but in favor of restricted access to Internet pornography. All who receive e-mail might be viewed as similarly situated. Spam reaching all e-mail accounts thus reaches similarly situated people. However, if most individuals who receive e-mail do not wish to receive spam, then this group (the unwilling) might be seen as not similarly situated with those who do wish to receive it. Such an argument could serve as the basis for something like an e-mail equivalent of the do-not-call list recently introduced for telemarketing.

The different approaches to ethics often produce the same result. If we consider the issue of hacking or gaining unauthorized access to another's system, utilitarianism concludes that more harm than good results from this activity. Those whose system is wrongfully accessed are faced with revising controls, checking to see what harm if any has been done, and correcting any problems caused by the hacking. Only the hacker gains. Those who have created or purchased the system have a right to limit access; the hacker has a duty to respect this right. It is not fair or just that some people go through the appropriate authorization to access or use a system while others hack into it. Thus from all three perspectives, hacking as defined can be judged unethical. If one does not accept the basic premises of the prevailing capitalist system, however, a defense of hacking can be devised (Halbert, 1994).

When the three approaches provide different results, rights and duties usually prevails as a way of determining whether an act is ethical, because rights are so basic. However, this is not always the case. In American copyright law, there is a "fair use" provision that allows an individual to make one copy for personal use of a copyrighted article without obtaining permission of the copyright holder. Whether this copy is made from a printed article on a photocopy machine or downloaded from a computer, the same basic principle applies (Halbert & Ingulli, 2003). A utilitarian analysis suggests that this provision allows the greatest good for the greatest number of people, because single-copy permissions for personal use would involve excessive transaction costs to both users and copyright owners. One could argue from a rights and duties perspective that the copyright owner has a right (and the user a corresponding duty) to payment for each and every copy made of the material, even for a single copy. Fairness and justice would suggest that the "fair use" exception is ethically acceptable as long as each owner and user play by the same rules.

FUTURE TRENDS

Each of the three views can be used to analyze issues regarding new information technology and its applications (Gilbert, 2001). Because information technology has developed so quickly in the last quarter century, and appears to be poised for continued rapid development, a good deal of ethical analysis is and will be needed of specific questions and issues. The basic approaches are clear from a long history of ethical theory. The specific issues are and will be fresh, but those concerned with using information technology ethically can find answers to their questions. The collecting, storing, transmitting and analysis of information will continue to be central to both commerce and society. Issues of ownership, access, privacy and social impact will continue to concern individuals and society.

It seems safe to assume that laws will evolve and court decisions will help to illuminate legal issues involved with information processing technology. Individual managers, technicians and citizens will ponder and debate the ethics of various uses of this technology. Basic philosophical approaches will remain the same; their application to individual situations will continue to require that thoughtful individuals work through the journey from abstract principles to particular applications and decisions.

CONCLUSION

Much of the discussion of right and wrong concerning information technology is based either on personal opinion or on legal interpretations of such topics as intellectual property rights and individual privacy. Philosophers have thought about, discussed, and written about right and wrong for thousands of years. While the history of philosophy has unfolded before the invention of electronic information technology, issues such as privacy, property ownership, truthfulness, and government intrusion on individual liberties have been the subjects of ethical inquiries for well over two thousand years. In attempting to cope with social issues raised by new technologies, the best thoughts of many generations of humans can be usefully brought to bear on current controversies. Doing this requires both knowledge of the discipline of ethics and at least some knowledge of new technologies and their social impacts. The application of such knowledge connects us with our ancestors who wrote on cave walls, told stories over campfires, printed books, and used the other means of processing information described at the beginning of this article.

REFERENCES

Alder, S. (1998). Ethical issues in electronic performance monitoring: A consideration of deontological and teleological perspectives. *Journal of Business Ethics*, *17*, 729-743.

Berlin, I. (2000). The pursuit of the ideal. In H. Hardy & R. Hausheer (Eds.), *The proper study of mankind: An anthology of essays* (pp. 1-16). New York: Farrar, Strauss & Giroux.

Bok, S. (1999). *Lying: Moral choice in public and private life* (Updated ed.). New York: Vintage Books.

Brown, J.S. (Ed.). (1997). *Seeing differently: Insights on innovation*. Boston: Harvard Business Review Books.

Burn, J., & Loch, K. (2001). The societal impact of the World Wide Web—Key challenges for the 21st century. In G. Dhillon (Ed.), *Social responsibility in the information age: Issues and controversies* (pp. 12-29). Hershey, PA: Idea Group Publishing.

Freeman, R., & Soete, L. (1997). *The economics of industrial innovation* (3rd ed.). Cambridge, MA: MIT Press.

Frey, R. (2000). Act-Utilitarianism. In H. LaFollete (Ed.), *The Blackwell guide to ethical theory* (pp. 165-182). Malden, MA: Blackwell Publishing.

Gilbert, J. (2001). New millenium; new technology; same old right and wrong. In G. Dhillon (Ed.), *Information security management: Global challenges in the new millennium*. Hershey, PA: Idea Group Publishing.

Halbert, T. (1994). Computer technology and legal discourse. *Murdoch University Electronic Journal of Law*, 2, May.

Halbert, T., & Ingulli, E. (2002). *Cyberethics*. Cincinnati, OH: West Legal Studies in Business.

Halbert, T., & Ingulli, E. (2003). *Law and ethics in the business environment* (4th ed.). Mason, OH: West Legal Studies in Business.

Ignatieff, M. (2001). *Human rights as politics and idolatry*. Princeton, NJ: Princeton University Press.

Kelly, E., & Rowland, H. (2000, May-June). Ethical and online privacy issues in electronic commerce. *Business Horizons*, 3-12.

LaFollette, H. (Ed.). (2000). *The Blackwell guide to ethical theory*. Malden, MA: Blackwell Publishers.

Mill, J.S., & Bentham, J.(1987). *Utilitarianism and other essays*. London: Penguin Books.

Sumner, L. (1987). *The moral foundation of rights*. Oxford: Clarendon Press.

Velasquez, M. (2002). *Business ethics: Concepts and cases* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

KEY TERMS

Citizen Rights: Those rights that an individual has by virtue of being a member of a government unit (country, state, province, etc.). They vary from government unit to government unit.

Contract Rights: Those rights that an individual has by reason of a valid contract that imposes duties on the other contracting party or parties. They are enforceable under legal systems, but are not the same as citizen rights.

Duties: The correlative of rights, since rights by their nature impose duties.

Ethics: The study of social or interpersonal values and the rules of conduct that follow from them.

Fairness and Justice: The philosophical view that the moral act is the one that treats similarly situated people in similar ways with regard to both process and outcome.

Human Rights: Those rights that all humans have simply by reason of being human, without regard to an individual government unit's laws.

Philosophy: The study of basic principles including what and how we know, rules for language and reasoning, and the basis for social interaction.

Position Rights: Those rights that an individual has by reason of the position that he or she occupies, such as police officer, chief financial officer, or parent.

Rights and Duties: The philosophical view that the moral act is the one that recognizes the rights of others and the duties that those rights impose on the actor.

Utilitarianism: The philosophical view that the moral act is the one that results in the greatest good or happiness for the greatest number of people.

Evaluating Computer–Supported Learning Initiatives

John B. Nash
Stanford University, USA

Christoph Richter
University of Hannover, Germany

Heidrun Allert
University of Hannover, Germany

INTRODUCTION

The call for the integration of program evaluation into the development of computer-supported learning environments is ever increasing. Pushed not only by demands from policy makers and grant givers for more accountability within lean times, this trend is due also to the fact that outcomes of computer-supported learning environment projects often fall short of the expectations held by the project teams. The discrepancy between the targets set by the project staff and the outcomes achieved suggests there is a need for formative evaluation approaches (versus summative approaches) that facilitate the elicitation of information that can be used to improve a program while it is in its development stage (c.p., Worthen, Sanders & Fitzpatrick, 1997). While the call for formative evaluation as an integral part of projects that aim to develop complex socio-technical systems is widely accepted, we note a lack of theoretical frameworks that reflect the particularities of these kind of systems and the ways they evolve (c.p., Keil-Slawik, 1999). This is of crucial importance, as formative evaluation will only be an accepted and effective part of a project if it provides information useful for the project staff. Below we outline the obstacles evaluation faces with regard to projects that design computer-supported learning environments, and discuss two promising approaches that can be used in complimentary fashion.

BACKGROUND

According to Worthen et al. (1997), evaluation is “the identification, clarification, and application of defensible criteria to determine an evaluation object’s value (worth or merit), quality, utility, effectiveness, or significance in relation to those criteria.” In this regard evaluation can serve different purposes. Patton (1997) distinguishes between judgment-, knowledge- and improvement-oriented evaluations. We focus on improvement-oriented evaluation approaches. We stress that evaluation can facilitate decision making and reveal information that can be used to improve not only the project itself, but also outcomes within the project’s target population. The conceptualization of evaluation as an improvement-oriented and formative activity reveals its proximity to design activities. In fact this kind of evaluative activity is an integral part of any design process, whether it is explicitly mentioned or not. Accordingly it is not the question if one should evaluate, but which evaluation methods generate the most useful information in order to improve the program. This question can only be answered by facing the characteristics and obstacles of designing computer-supported learning environments.

Keil-Slawik (1999) points out that one of the main challenges in evaluating computer-supported learning environments is that some goals and opportunities can spontaneously arise in the course of the development process and are thus not specified in advance. We believe that this is due to the fact that design, in this context, addresses ill-structured and situated problems. The design and implementation of computer-supported learning environments, which can be viewed as a response to a perceived problem, also generates new problems as it is designed. Furthermore every computer-supported learning experience takes place in a unique social context that contributes to the success of an intervention or prevents it. Therefore evaluation requires that designers pay attention to evolutionary and cyclic processes and situational factors. As Weiss notes, “Much evaluation is done by investigating outcomes without much attention to the paths by which they were produced” (1998, p. 55).

For developers designing projects at the intersection of information and communication technology (ICT) and the learning sciences, evaluation is difficult. Evaluation efforts are often subverted by a myriad of confounding variables, leading to a “garbage in, garbage out” effect;

the evaluation cannot be better than the parameters that were built in the project from the start (Nash, Plugge & Eurlings, 2001). Leaving key parameters of evaluative thinking out of computer-supported learning projects is exacerbated by the fact that many investigators lack the tools and expertise necessary to cope with the complexity they face in addressing the field of learning.

We strongly advocate leveraging the innate ability of members of the computer science and engineering communities to engage in "design thinking" and turn this ability into a set of practices that naturally becomes program evaluation, thereby making an assessment of the usefulness of ICT tools for learning a natural occurrence (and a manifest activity) in any computer-supported learning project.

Design-Oriented Evaluation for Computer-Supported Learning Environments

There are two approaches that inherently relate themselves to design as well as to evaluation. Therefore they are useful tools for designers of computer-supported learning initiatives. These two perspectives, discussed below, are scenario-based design and program theory evaluation. Both approaches assume that the ultimate goal of a project should be at the center of the design and evaluation discussion, ensuring a project is not about only developing a usable tool or system, but is about developing a useful tool or system that improves outcomes for the user. Beyond this common ground, these approaches are rather complementary to each other and it is reasonable to use them in conjunction with one another.

Scenario-Based Approaches

Scenario-based approaches are widely used in the fields of software engineering, requirements engineering, human computer interaction, and information systems (Rolland et al., 1996). Scenarios are a method to model the universe of discourse of an application, that is, the environment in which a system, technical or non-technical, will be deployed. A scenario is a concrete story about use of an innovative tool and/or social interactions (Carroll, 2000). Scenarios include protagonists with individual goals or objectives and reflect exemplary sequences of actions and events. They refer to observable behavior as well as mental processes, and also cover situational details assumed to affect the course of actions (Rosson & Carroll, 2002). Additionally it might explicitly refer to the underlying culture, norms, and values (see Bødker & Christiansen, 1997). That said, scenarios usually focus on specific situations, only enlighten some important aspects, and generally do not include every eventuality (e.g., Benner, Feather, Johnson & Zorman, 1993).

Beside their use in the design process, scenarios can also be used for purposes of formative evaluation. First of all, as a means of communication, they are a valuable resource for identifying underlying assumptions regarding the program under development. Stakeholder assumptions might include those related to instructional theories, the learner, the environmental context, and its impact on learning or technical requirements. Underlying assumptions such as these are typically hidden from view of others, but easily developed and strongly held within individuals developing computer-supported learning environments. Scenarios help to reveal the thinking of designers so that others can participate in the design process and questionable assumptions can come under scrutiny. The use of scenarios also allows identification of pros and cons of a certain decision within the design process. In this vein Carroll (2000) suggests employing "claim analysis." Claims are the positive or negative, desirable and undesirable consequences related to a certain characteristic of a scenario. Assuming that every feature of a proposed solution usually will entail both positive and negative effects helps to reflect on the current solution and might provoke alternative proposals. The analysis of claims is thereby not limited to an intuitive ad hoc evaluation, but also can bring forth an explicit hypothesis to be addressed in a subsequent survey.

Program Theory Evaluation

Program theory evaluation, also known as theory-based evaluation, assumes that underlying any initiative or project is an explicit or latent "theory" (or "theories") about how the initiative or project is meant to change outcomes. An evaluator should surface those theories and lay them out in as fine detail as possible, identifying all the assumptions and sub-assumptions built into the program (Weiss, 1995). This approach has been promoted as useful in evaluating computer-supported learning projects (Strömdahl & Langerth-Zetterman, 2000; Nash, Plugge & Eurlings, 2001) where investigators across disciplines find it appealing. For instance, for designers (in mechanical engineering or computer science), program theory evaluation reminds them of their own use of the "design rationale." And among economists, program theory evaluation reminds them of total quality management (TQM). In the program theory approach (Weiss, 1995, 1998; Chen, 1989; Chen & Rossi, 1987), one constructs a project's "theory of change" or "program logic" by asking the various stakeholders, "What is the project designed to accomplish, and how are its components intended to get it there?" The process helps the project stakeholders and the evaluation team to identify and come to consensus on the project's theory of change. By identifying and describing the activities, outcomes, and goals of the program, along with their interrelationships, the stakeholders are then in position to identify quantifiable measures to portray the veracity of the model.

Theory-based evaluation identifies and tests the relationships among a project's inputs or activities and its outcomes via intermediate outcomes. The key advantages to using theory-based evaluation are (Connel & Kubisch, 1995; Weiss, 1995):

- It asks project practitioners to make their assumptions explicit and to reach consensus with their colleagues about what they are trying to do and why.
- It articulates a theory of change at the outset and gains agreement on it, by all stakeholders reducing problems associated with causal attribution of impact.
- It concentrates evaluation attention and resources on key aspects of the project.
- It facilitates aggregation of evaluation results into a broader context based on theoretical program knowledge.
- The theory of change model identified will facilitate the research design, measurement, data collection, and analysis elements of the evaluation.

Both scenario-based design and program theory stress the importance of the social context while planning computer-supported environments. They also represent means to facilitate the communication among the stakeholders and urge the project team to reflect their underlying assumptions in order to discuss and test them. Furthermore, both approaches are particularly suitable for multidisciplinary project teams. Scenarios and program logic maps are not static artifacts; they are a starting point for discussion and have to be changed when necessary. With these similarities there are also differences in both approaches. The major difference between them is that program theory offers a goal-oriented way to structure a project, while scenario-based design proffers an explorative approach that opens the mind to the complexity of the problem, alternatives, and the diversity of theories that try to explain social and socio-technical process. That is, scenario-based design highlights the divergent aspects of project planning, and evaluation program theory stresses the convergent aspects. Program theory evaluation helps to integrate each scenario, decision, and predefinition into the whole process. Scenarios force users not just to use terms, but to give meaningful descriptions. They force users to state how they actually want to instantiate an abstract theory of learning and teaching. This helps to implement the project within real situations of use, which are complex and ill structured. Program theory helps to focus on core aspects of design and prevent getting 'lost in scenarios'. Scenarios and program theory evaluation can be used in an alternating way. Thereby it is possible to use both approaches and improve the overall development process.

The program theory of an initiative can be a starting point for writing scenarios. Especially the interrelations between the goals and interrelation between ultimate goal and inputs can be described with a scenario. The scenario can help to understand how this interrelation is meant to work and how it will look in a concrete situation. Scenarios on the other hand can be used to create program theory by pointing out main elements of the intended program. They can also be used to complete already existing program theory by presenting alternative situations of use. For developers of computer-supported learning environments, scenario-based design and program theory represent complementary approaches, which when used together or separately, can add strength to the implementation and success of such projects.

FUTURE TRENDS

It is clear that formative evaluation will become more important in the future, and it will be especially crucial to think about how to integrate evaluation into the design process. Essentially, designers will need to answer the question "Why does the program work?" and not just "Did it work?" It becomes obvious that the design of a computer-supported learning environment, like the development of any other complex socio-technical system, is a difficult process. In fact the necessity for changes in the original plan is practically preordained due to the ill-structured and situated nature of the domain. The mere act of engaging in a design process suggests that designers will engage in planned as well as evolutionary, unplanned activities. Therefore it is important that the project designers use methods that support divergent thinking and methods that support convergent processes. While scenario-based design and program-theory evaluation represent complementary views on the design and evaluation of computer-supported learning environments that can facilitate these processes, there is still room for improvement.

CONCLUSION

Formative evaluation is an important means to ensure the quality of an initiative's outcomes. Formative evaluation directed towards improvement of an initiative can be understood as a natural part of any design activity. While this is widely recognized, there is still a lack of program evaluation frameworks that reflect the uniqueness of the design process, the most crucial of which is the inherent ambiguity of design. In spite of great inspiration portrayed by project teams, usually manifested by visions of a certain and sure outcome, no project can be pre-planned completely, and midcourse corrections are a certainty. Scenario-based design and program theory evaluation provide a theoretical foothold for projects in need of collecting and analyzing data for program improvement and judging program success.

In sum, scenario-based design and program theory hold many similarities. The major difference between them is that program theory offers a goal-oriented way to structure a project, while scenario-based design provides an explorative approach that opens the mind to the complexity of the problem, alternatives, and the diversity of theories that try to explain social and socio-technical process.

Scenario-based design highlights the divergent aspects of project planning, and evaluation program theory stresses the convergent aspects. For developers of computer-supported learning experiences, scenario-based design and program theory represent complementary approaches, which when used together or separately can add strength to the implementation and success of ICT learning projects.

REFERENCES

Benner, K.M., Feather, M.S., Johnson, W.L. & Zorman, L.A. (1993). Utilizing scenarios in the software development process. In N. Prakash, C. Rolland & B. Pernici (Eds.), *Information system development process* (pp. 117-134). Elsevier Science Publishers.

Bødker, S. & Christiansen, E. (1997). Scenarios as springboards in design. In G. Bowker, L. Gaser, S.L. Star & W. Turner (Eds.), *Social science research, technical systems and cooperative work* (pp. 217-234). Lawrence Erlbaum.

Carroll, J.M. (2000). *Making use: Scenario-based design of human-computer interactions*. Cambridge: MIT Press.

Chen, H.T. (1989). Issues in the theory-driven perspective. *Evaluation and Program Planning, 12*, 299-306.

Chen, H.T. & Rossi, P. (1987). The theory-driven approach to validity. *Evaluation Review, 7*, 95-103.

Connell, J.P. & Kubisch, A. (1995). Applying a theory of change approach to the evaluation of comprehensive community initiatives: Progress, prospects, and problems. In K. Fulbright-Anderson et al. (Eds.), *New approaches to evaluating community initiatives. Volume 2: Theory, measurement, and analysis*. Washington, DC: Aspen Institute.

Keil-Slawik, R. (1999). Evaluation als evolutionäre systemgestaltung. aufbau und weiterentwicklung der paderborner DISCO (Digitale Infrastruktur für computerunterstütztes kooperatives Lernen). In M. Kindt (Ed.), *Projektevaluation in der lehre—multimedia an hochschulen zeigt profil(e)* (pp. 11-36). Münster, Germany: Waxmann.

Nash, J.B., Plugge, L. & Eurlings, A. (2001). Defining and evaluating CSCL evaluations. In A. Eurlings & P. Dillenbourg (Eds.), *Proceedings of the European Conference on Computer-Supported Collaborative Learning* (pp. 120-128). Maastricht, The Netherlands: Universiteit Maastricht.

Patton, M.Q. (1997). *Utilization-focused evaluation* (3rd Edition). Thousand Oaks, CA: Sage Publications.

Rolland, C., Achour, C.B., Cauvet, C., Ralyté, J., Sutcliffe, A., Maiden, N.A.M., Jarke, M., Haumer, P., Pohl, K., Dubois, E. & Heymans, P. (1996). *A proposal for a scenario classification framework.* CREWS Report 96-01.

Rosson, M.B. & Carroll, J.M. (2002). Usability engineering: Scenario-based development of human-computer interaction. San Francisco: Morgan Kaufmann.

Strömdahl, H. & Langerth-Zetterman, M. (2000). *On theory-anchored evaluation research of educational settings, especially those supported by information and communication technologies (ICTs).* Uppsala, Sweden: Swedish Learning Lab.

Weiss, C. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. In J. Connell et al. (Eds.), *New approaches to evaluating community initiatives: Concepts, methods, and contexts*. Washington, DC: Aspen Institute.

Weiss, C. (1998). *Evaluation research: Methods for studying programs and policies.* Englewood Cliffs, NJ: Prentice-Hall.

Worthen, B.R., Sanders, J.R. & Fitzpatrick, J.L. (1997). *Program evaluation—alternative approaches and practical guidelines* (2nd Edition). New York: Addison Wesley Longman.

KEY TERMS

Computer-Supported Learning: Learning processes that take place in an environment that includes computer-based tools and/or electronically stored resources. CSCL is one part of this type of learning.

Evaluation: The systematic determination of the merit or worth of an object.

Formative Evaluation: The elicitation of information that can be used to improve a program while it is in the development stage.

Program: A social endeavor to reach some predefined goals and objectives. A program draws on personal, social, and material resources to alter or preserve the context in which it takes place.

Program Theory: A set of assumptions underlying a program that explains why the planned activities should lead to the predefined goals and objectives. The program theory includes activities directly implemented by the program, as well as the activities that are generated as an response to the program by the context in which it takes place.

Scenarios: A narrative description of a sequence of (inter-)actions performed by one or more persons in a particular context. Scenarios include information about goals, plans, interpretations, values, and contextual conditions and events.

Summative Evaluation: The elicitation of information that can be used to determine if a program should be continued or terminated.

Evaluating IS Quality as a Measure of IS Effectiveness

Carla Wilkin
The University of Melbourne, Australia

INTRODUCTION

An enduring question in information systems research and practice concerns evaluation of the impact of information systems (IS). It endures, as to date there is no ready solution. Focusing on one aspect, measuring IS success or effectiveness, there are ranges of measures available. At one end of the scale we have perceptual measures like use and user satisfaction; somewhere along that scale we have the more objective measures like quality; whilst at the other end we have objective measures like increased market share, price recovery and increased product quality.

Measurement of IS success or effectiveness has been shaped by DeLone and McLean (1992), who proposed a taxonomy and an interactive model that conceptualized and operationalized IS success. However, this was based on theoretical and empirical work from the 1970s and 1980s, published in the period 1981-1988. Information systems, not being a static phenomenon, have progressed and changed. DeLone and McLean (2002, 2003) themselves acknowledged this in their recent revisitation, reexamination and reformulation of their IS success model. Their view correctly affirms that we cannot leave people outside this equation; meaning objective measures alone are not appropriate. Furthermore, the subjectivity of perceptual measures mean they are of questionable usefulness. Taking the middle ground, where quality is the measure, the question then becomes how best to measure quality of a delivered IS.

In an equation that seeks to define our understanding of the value of information technology (IT) to the business process, the system as a stand-alone object is worthless. The worth of the system lies in its role in the business process: and it is people who make it work in these processes. What is therefore required is a measure that takes account of human reactions to delivered systems. This can be evaluated by considering a variety of end-user stakeholder expectations and/or perceptions as measures of quality. In fact, much insight can be gained by measuring the disconfirmation of expectations of ideal service and perceptions of reality (Wilkin, 2001), particularly if this is assessed at various levels of seniority.

MEASURING QUALITY

Debate has surrounded measuring quality from a disconfirmation perspective (Carr, 2002; Peter, Churchill & Brown, 1993; Van Dyke, Prybutok & Kappelman, 1999). Justification for including expectations (Cronin & Taylor, 1992, 1994; Teas, 1993, 1994; Van Dyke, Kappelman & Prybutok, 1997) centred on the insight it provided about how users formulated perceptions or how significant such users saw each dimension or statement (Carman, 1990; Kettinger & Lee, 1997; Parasuraman, Zeithaml & Berry, 1986; Pitt, Watson & Kavan, 1995). Moreover, expectations are seen as essential to both understanding and achieving IS effectiveness, particularly given the different internal opinions held by different user stakeholders where a low or high perception rating could provide misleading information. A measure that includes expectations provides insight regarding changes in the system environment (Watson, Pitt & Kavan, 1998; Wilkin, 2001).

The perception's only measure, another approach to defining and evaluating quality, was proposed in a belief that a measurement of service quality derived by the difference score only captured factors that were related to service quality and did not measure customers' view of the concept itself (Cronin & Taylor, 1992). However, support can be found for the view that a single measure of performance provides little information about a user's thoughts in relation to product features, nor the process by which performance is converted into understanding by the consumer (Oliver, 1989; Spreng, MacKenzie & Olshavsky, 1996).

A definition of quality could have many contradictory functions: sometimes implicit/sometimes explicit; at times mechanistic/at times humanistic; and sometimes conceptually/sometimes operationally understood. In an IT context, there is not any single understanding of the term. Quality, being concerned with the totality of features, is best evaluated as a multi-dimensional construct using multiple statements to capture the quality of each dimension.

Applying a measure of quality to evaluate something as complex as a delivered IS requires consideration and understanding of the mechanisms that underpin an IS. The DeLone and McLean model conceptualized system quality (not system) and information quality (not information). Despite the complexity and technical nature of some IT products, in order to achieve success, we need to look beyond the process and delivery of the product, to the system as a whole, and ask whether benefits can be gained by focusing on customer views of the quality of the product, product delivery and associated concerns (Wilkin, 2001).

Quality has many elements. If we put this human evaluation of a delivered system into context, then it is not just measurement of the system itself (system quality), nor the information so generated (information quality) that is important, but a balanced evaluation that also takes account of service (service quality) and the role of an IS unit in contributing to the effectiveness of delivered IS, which is important (Wilkin, 2001). Support for the argument to include service quality in this evaluation can be found in the work of other researchers too (DeLone & McLean, 2002, 2003; Kettinger & Lee, 1994; Li, 1997; Pitt, Watson & Kavan, 1995; Wilkin & Hewett, 1999).

Assuming a multi-dimensional approach to evaluating quality of delivered IS encompassing the system, information and service aspects, the issue then is which dimensions are important for each aspect (component). Table 1 summarizes the important dimensions (Wilkin, 2001) in measuring each component (system quality, information quality and service quality). Following on, what are then required are indicators capable of measuring aspects of each component. These are many and vary from "*responds quickly to all commands*" (system quality), to "*quickly interpreted*" (information quality) and "*delivers support in a timely manner*" (service quality).

Under this multi-dimensional approach, ratings for the various aspects of quality, 1, 2 and so on, captured on a Likert scale of 1 to 7 (strongly agree to strongly disagree), highlight problematic areas, which when viewed in conjunction with organizational goals and objectives, can facilitate the establishment of priorities.

At a strategic level, the merits of this approach, where multiple dimensions and statements are used to evaluate the quality/effectiveness of an information system, relate to the ease and simplicity with which insight into the system in question is provided. Predecessors have captured quality or surrogates of quality in a single statement, thereby limiting insights provided to interested parties on the aspects of the business system/application stakeholders perceive as problematic. Thinking beyond the impact on the individual and organization, the value provided by such an approach is significant in light of the advancement of organizations to what Drucker (1988) forecast as the third period of change in organizational structure, namely to an information-based organization. Herein, "information is data endowed with relevance and purpose and knowledge, by definition, is specialized" (Drucker, 1988, p. 58). Thus, it is accordingly vital that the IS delivers information of the required quality.

In line with Drucker (1988), this multi-dimensional approach allows the evaluator to directly target and compile the views of a broad cross-section of stakeholders regarding the quality of the IS with respect to the performance of their duties.

At an operational level, the merits of the approach include:

- the flexibility to add and subtract dimensions for each component according to users requirements;
- the use of different dimensions to measure the different components of quality;
- the capability for benchmarking where expectations, measured at intermittent intervals, is balanced with more timely assessment and reassessments of perceptions;
- the opportunity, because of the use of dimensionality, to discover specific problematic areas, and then "drill down" into those areas; and
- improvement in the "usefulness" of the results through the addition of statements specific to the situation – something that is offset to a degree against the increase in length.

FUTURE TRENDS

Despite much work having been done on evaluation of the impact of IS, further investigation is warranted to balance subjective and objective measures of quality of these systems. The answers to this investigation will probably flow from the debate concerning the relative merits of

Table 1. Important dimensions in measuring system quality, information quality and service quality

System Quality	Information Quality	Service Quality
Functionality	Accuracy	Expertise
Integration	Availability	Credibility
Usability	Relevance	Availability
Reliability	Presentation	Responsiveness
Security	Promptness	Supportiveness

considering the desirability and relevance of monetary evaluations balanced with subjective judgments related to end-user stakeholder evaluations of IS performance and productivity.

CONCLUSION

Iacocca's (1998) words, quality "doesn't have a beginning or a middle. And it better not have an end" (p. 257), are as valid today as ever, since the realization of high quality/effectiveness is only achievable when it becomes an intrinsic part of business operations through every stakeholder's mindset.

The quality-based, multi-dimensional approach to evaluation of a delivered IS outlined here (comprising components, dimensions and indicators), enables problematic areas to be more accurately pinpointed. The magnitude of organizations' investment and commitment to IT, compounded by the increasingly complex and interwoven nature of IS, make evaluation of quality of a delivered IS a significant issue. In this regard, this article has discussed a number of critical issues, which offer to business and researchers alike implications and challenges. Hence, despite persistent difficulties in measuring the quality of these delivered systems (Davis, 1989), we should pursue work on balancing subjective and objective measures of quality in a timely manner.

REFERENCES

Brown, S.A. (1992). *Total quality service: How organizations use it to create a competitive advantage*. Scarborough, Ontario: Prentice Hall.

Carman, J.M. (1990). Consumer perceptions of service quality: An assessment of the SERVQUAL dimensions. *Journal of Retailing, 66*(1), 33-55.

Carr, C.L. (2002). A psychometric evaluation of the expectations, perceptions, and difference-scores generated by the IS-adapted SERVQUAL instrument. *Decision Sciences, 33*(2), 281-296.

Cronin, J.J., Jr., & Taylor, S.A. (1992, July). Measuring service quality: A reexamination and extension. *Journal of Marketing, 56,* 55-68.

Cronin, J.J., Jr., & Taylor, S.A. (1994, January). SERVPREF versus SERVQUAL: Reconciling performance-based and perceptions-minus-expectations measurement of service quality. *Journal of Marketing, 58,* 125-131.

Davis, F.D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 318-339.

DeLone, W.H., & McLean, E.R. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research, 3*(1), 60-95.

DeLone, W.H., & McLean, E.R. (2002). Information systems success revisited. In R.H. Sprague, Jr. (Ed.), *Proceedings of the 35th Annual Hawaii International Conference on System Sciences,* Big Island, Hawaii (p. 238).

DeLone, W.H., & McLean, E.R. (2003). The DeLone and McLean model of information systems success: A ten-year update. *Journal of Management Information Systems, 19*(4), 9-30.

Drucker, P.F. (1988, January-February). The coming of the new organization. *Harvard Business Review,* 45-53.

Iacocca, L. (1988). *Talking straight*. New York: Bantam Books.

Kettinger, W.J., & Lee, C.C. (1994). Perceived service quality and user satisfaction with the information services function. *Decision Sciences, 25*(5/6), 737-766.

Kettinger, W.J., & Lee, C.C. (1997). Pragmatic perspectives on the measurement of information systems service quality. *MIS Quarterly, 21*(2), 223-240.

Li, E.Y. (1997). Perceived importance of information systems success factors: A meta analysis of group differences. *Information and Management, 32*(1), 15-28.

Oliver, R.L. (1989). Processing of the satisfaction response in consumption: A suggested framework and research propositions. *Journal of Consumer Satisfaction Dissatisfaction and Complaining, 2,* 1-16.

Parasuraman, A, Zeithaml, V.A., & Berry, L.L. (1986). SERVQUAL: A multiple-item scale for measuring customer perceptions of service quality. *Marketing Science Institute,* 1-39, 86-108.

Peter, J.P., Churchill, G.A., Jr., & Brown, T.J. (1993). Caution on the use of difference scores in consumer research. *Journal of Consumer Research, 19*(4), 655-662.

Pitt, L.F., Watson, R.T., & Kavan, C.B. (1995). Service quality: A measure of information systems effectiveness. *MIS Quarterly, 19*(2), 173-187.

Spreng, R.A., MacKenzie, S.B., & Olshavsky, R.W. (1996). A reexamination of the determinants of consumer satisfaction. *Journal of Marketing, 60*(3), 15-32.

Teas, R.K. (1993, October). Expectations, performance evaluation, and consumers' perceptions of quality. *Journal of Marketing*, *57*, 18-34.

Teas, R.K. (1994, January). Expectations as a comparison standard in measuring service quality: An assessment of a reassessment. *Journal of Marketing*, *58*, 132-139.

Van Dyke, T.P., Kappelman, L.A., & Prybutok, V.R. (1997). Measuring information systems service quality: Concerns on the use of the SERVQUAL questionnaire. *MIS Quarterly*, *21*(2), 195-208.

Van Dyke, T.P., Prybutok, V.R., & Kappelman, L.A. (1999). Cautions on the use of the SERVQUAL measure to assess the quality of information systems services. *MIS Quarterly*, *21*(2), 195-208.

Watson, R.T., Pitt, L.F., & Kavan, C.B. (1998). Measuring information systems service quality: Lessons from two longitudinal case studies. *MIS Quarterly*, *22*(1), 61-79.

Wilkin, C. (2001). *Quality as the criterion for delivered IS effectiveness*. Unpublished doctoral thesis. Deakin University, Australia.

Wilkin, C., & Hewett, B. (1999). Quality in a respecification of DeLone and McLean's IS success model. In M. Khosrow-Pour (Ed.), *Managing information technology resources in organizations in the next millennium* (pp. 663-672). Hershey, PA: Idea Group Publishing.

KEY TERMS

Component: A term used to describe an information system and its composition for the purposes of this work. Specifically, the components in this work are: system quality, information quality and service quality.

Dimension: Refers to the determinants of quality of each of the three components, namely, system quality, information quality and service quality.

Expectations: These have a future time perspective and a degree of uncertainty. Expectations are a set of beliefs, held by targeted users of an information system, associated with certain attributes, beliefs or outcomes. They are associated with the eventual perception of a system and with the performance of the system.

Indicator: A term used to refer to something that would point to quality or a lack thereof.

Information Quality: A global judgment of the degree to which these stakeholders are provided with information of excellent quality with regard to their defined needs, excluding user manuals and help screens (features of system quality).

IS Success: A global judgment of the degree to which these stakeholders believe they are better off. The term is sometimes used interchangeably with IS effectiveness.

Perceptions: Contingent upon prior expectations, perceptions have been used by some as a reality check of expectations, where an assessment of quality is derived by the disconfirmation of the two. Moreover, they have also been proposed as a measure of adequacy (perceptions)/importance.

Quality: An elusive and indistinct construct defined in terms of customer perceptions and expectations. In arriving at a definition, one must take account of both audience and circumstance. There has been some attempt to define it as a global judgment about a product's (or service's) overall excellence. Quality can be measured on the basis of customer expectations and perceptions, along the lines of Brown's (1992, p. 255) definition: "[t]he totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs".

Service Quality: A global judgment or attitude relating to an assessment of the level of superiority or excellence of service provided by the IS department and support personnel.

System Quality: A global judgment of the degree to which the technical components of delivered IS provide the quality of information and service as required by stakeholders, including hardware, software, help screens and user manuals.

Evaluation of an Open Learning Environment[1]

Geraldine Clarebout
University of Leuven, Belgium

Jan Elen
University of Leuven, Belgium

Joost Lowyck
University of Leuven, Belgium

Jef Van den Ende
Institute of Tropical Medicine, Belgium

Erwin Van den Enden[2]
Institute of Tropical Medicine, Belgium

INTRODUCTION

Educational goals have generally shifted from knowing everything in a specific domain to knowing how to deal with complex problems. Reasoning and information processing skills have become more important than the sheer amount of information memorized. In medical education, the same evolution occurred. Diagnostic reasoning processes get more strongly emphasized. Whereas previously knowing all symptoms and diseases was stressed, reasoning skills have become educationally more important. They must enable professionals to distinguish between differential diagnoses and recognize patterns of illnesses (e.g., Myers & Dorsey, 1994).

BACKGROUND

Authentic or realistic tasks have been advocated to foster the acquisition of complex problem-solving processes (Jacobson & Spiro, 1995; Jonassen, 1997). In medical education, this has led to the use of expert systems in education. Such systems were initially developed to assist practitioners in their practice (NEOMYCIN, in Cromie, 1988; PATHMASTER in Frohlich, Miller, & Morrow, 1990; LIED in Console, Molino, Ripa di Meanan, & Torasso, 1992). These systems simulate a real situation and were expected to provoke or develop students' diagnostic reasoning processes. However, the implementation of such expert systems in regular educational settings has not been successful. Instead of developing reasoning processes, these systems assume them to be available. They focus on quickly getting to a solution rather than reflecting on possible alternatives. Consequently, it was concluded that students need more guidance in the development of diagnostic reasoning skills (Console et al., 1992, Cromie, 1988; Friedman, France, & Drossman, 1991); instructional support was lacking.

KABISA is one of the computer programs that, among other things, aims at helping students to develop their diagnostic reasoning skills (Van den Ende, Blot, Kestens, Van Gompel, & Van den Enden, 1997). It is a dedicated computer-based training program for acquiring diagnostic reasoning skills in tropical medicine.

DESCRIPTION OF THE PROGRAM

KABISA confronts the user with cases or "virtual patients". The virtual patient is initially presented by three "characteristics"[3], randomly selected by the computer. After the presentation of the patient (three characteristics), students can ask additional characteristics gathered through anamnesis, physical examination, laboratory and imaging.

If students click on a particular characteristic, such as a physical examination test, they receive feedback. Students are informed about the presence of a certain symptom, or whether a test is positive or negative. If students ask a "non-considered" characteristic, that is, a characteristic that is not relevant or useful in relation to the virtual patient, they are informed about this and asked whether they want to reveal the diagnosis they were thinking about. When they do so, students receive an overview of the characteristics that were explained by their selection and which ones are not. Additionally, they get the place of the selected diagnosis on a list that ranks diagnoses according to their probability given the characteristics at hand. If students do not want to show the diagnosis they were thinking about, they can just continue asking characteristics.

A session is ended with students giving a final diagnosis. KABISA informs them about the correctness. If it is correct, students are congratulated. If the diagnosis is not correct, students may be either informed that it is a very plausible diagnosis but that they do not have enough evidence, or they may get a ranking of their diagnosis and an overview of the disease characteristics that can and cannot be explained by their answer.

Additionally, different non-embedded support devices, that is, tools, are made available to support learners. These tools allow students to look for information about certain symptoms or diseases, to compare different diagnoses, or to see how much a certain characteristic contributes to the certainty for a specific diagnosis. Students decide themselves when and how they use these devices (for a more detailed description, see Clarebout, Elen, Lowyck, Van den Ende, & Van den Enden, 2004).

FUTURE TRENDS

In this section, some critical issues are put forward that raise discussion points for the future design and development of open learning environments.

A Learning Environment vs. a Performance Environment

KABISA is designed as an open learning environment, that is, students are confronted with a realistic and authentic problem; there is a large amount of learner control and tools are provided to learners to guide their learning (Hannafin, Land & Oliver, 1999). However, the performed evaluation study revealed some interesting issues. A first revelation was that students do not follow a criterion path when working on KABISA. Prior to the evaluation, two domain experts in collaboration with three instructional designers constructed a criterion path. This path represented the ideal paths students should go through to optimally benefit from KABISA (following the “normative approach” of Elstein & Rabinowitz, 1993), including when to use a specific tool. Only five out of 44 students followed this path.

A second issue relates to tool use. KABISA offers different tools to support students. These tools can help students in their problem-solving process. Results suggest that students consult some help functions more than others, but overall they do not consult them frequently and if they use them they do not use them adequately. Students also tend to not use the feedback that they can obtain when asking for a “non-considered” characteristic.

Although this environment can be described as an open learning environment, it seems that students do not perceive it as a learning environment, but rather as a performance environment. Thinking aloud protocols reveal that students think they are cheating or failing when consulting a tool. Giving the limited use of these tools, it becomes difficult to gain insight in the effect of tool use on the learning process.

However, in spite of the observation that in only a small number of consultations the criterion path was followed, students do find in 80% of the consultations the right diagnosis. It seems that by trial and error, by not following the criterion path, students can also obtain the right diagnosis.

The results of this evaluation suggest that students do not use KABISA to foster their diagnostic reasoning skills. Rather, KABISA enables them to train readily available skills.

The Use of Design Models for Designing Open Learning Environments

This evaluation shows the importance of an evaluation phase in the design and development of computer-based training programs. It reveals the valuable contribution of (linear) design models, such as the so-called ADDIE-model (Analyse-Design-Development-Implementation-Evaluation). Although it is argued that in open learning environments a linear design process cannot longer be applied, this evaluation shows that it still can contribute to the design. For instance, a more thorough analysis (first phase) of student characteristics could have provided a means to adapt the difficulty level to the level of the students or to identify what guidance students actually need. Apparently, the feedback given to students does not encourage them to adapt their problem-solving process. Being product- rather than process-oriented, feedback may not be adapted to students’ actual needs. Or, students’ instructional conceptions about computer-based learning environments or their perceptions about KABISA (game versus an educational application) may influence the use of the program. Students’ instructional conceptions should be taken into account through the design process of the program. One possible way to influence these conceptions might be the introduction of the program. In the introduction, the aims of the program, the different functionalities and the relationship with the different courses should be clearly defined (see Kennedy, Petrovi, & Keppell, 1998, for the importance of introductory lessons). This relates to the implementation phase.

Given the difficulty of anticipating potential problems and difficulties students might encounter in open learning environments, it might be considered to break the linearity of such design models and to introduce a formative evaluation after each phase. This would enable the redirection of the program while developing it, rather than after the implementation of the program. Rather than only evaluating a final product, the development process should be taken into consideration as well. Rapid prototyping for testing the program at different phases of the development might be indicated. This leads to a more spiral cycle rather than a linear design process.

Amount of Learner Control in Computer-Based Programs

In the design and development of KABISA, a lot of time and effort is spent in the development of tools, similar to other computer-based programs. However, results show that students do not (adequately) use these tools. Other authors have found similar results (see for instance, Crooks, Klein, Jones, & Dwyer, 1996; Land, 2000). This raises questions about the amount of learner control in open learning environments. Should the environment be made less open and provide embedded support devices instead of tools so that students cannot but use these devices? Or should students receive some additional advice towards the use of these tools? In the first case, support might not be adapted to the learners need. This might cause problems, given that either too much or too less support can both be detrimental (Clark, 1991). The second option leaves the environment open. But also here it can be questioned whether this advice should not also be adapted to the learners' needs. A possible solution with respect to this issue might come out of the animated pedagogical agent-research. These agents are animated figures that aim at helping learners in their learning process and adapt their support based on the paths learners followed (Moreno, 2004; Shaw, Johnson, & Ganeshan, 1999).

CONCLUSION

The evaluation of KABISA addressed some general issues important to consider in the design, development and implementation of open learning environments. Although these environments are advocated to foster the acquisition of complex problem-solving skills, there seems still to be gap between the intention of the designers and the use by the learners. This relates to the issue addressed by Winne and Marx (1982) about calibration. In order for an instructional intervention to be effective, calibration is needed between the conceptions of the different people involved. The introduction of a pedagogical agent might help to calibrate the conceptions of students to those of the designers. Moreover these agents might help in encouraging students to adequately use tools without reducing the openness of the learning environment.

REFERENCES

Clarebout, G., Elen, J., Lowyck, J., Van den Ende, J., & Van den Enden, E. (2004). KABISA: Evaluation of an open learning environment. In A. Armstrong (Ed.), *Instructional design in the real world: A view from the trenches* (pp. 119-135). Hershey, PA: Idea Group Publishing.

Clark, R.E. (1991). When teaching kills learning: Research on mathemathantics. In H. Mandl, E. De Corte, N. Bennett, & H.F. Friedrich (Eds.), *European research in an interantional context: Volume 2. Learning and Instruction* (pp. 1-22). Oxford, NY: Pergamon Press.

Console, L., Molino, G., Ripa di Meana, V., & Torasso, P. (1992). LIED-liver: Information, education and diagnosis. *Methods of Information in Medicine, 31*, 284-297.

Cromie, W.J. (1988). Expert systems and medical education. *Educational Researcher*, *17*(3), 10-12.

Crooks, S.M., Klein, J.D., Jones, E.E., & Dwyer, H. (1996). Effects of cooperative learning and learner-control modes in computer-based instruction. *Journal of Research in Computing in Education*, *29*, 223-244.

Elstein, A.S., & Rabinowitz, M. (1993). Medical cognition: Research and evaluation. In M. Rabinowitz (Ed.), *Cognitive Science Foundation of Instruction* (pp. 189-201). Hillsdale, NJ: Lawrence Erlbaum Associates.

Friedman, C.P., France, C.L., & Drossmann, D.D. (1991). A randomized comparison of alternative formats for clinical simulations. *Medical Decision Making, 11*(4), 265-271.

Frohlich, M.W., Miller, P.L., & Morrow, J.S. (1990). PATHMASTER: Modelling differential diagnosis as "Dynamic Competition" between systematic analysis and disease-directed deduction. *Computers and BiomedicalResearch, 23*, 499-513.

Hannafin, M.J., Land, S., & Oliver, K. (1999). Open learning environments: Foundations, methods and models. In C.M. Reigeluth (Ed.)¸ *Instructional design theories and models. A new paradigm of Instructional Theory* (Vol. 2, pp. 115-140). Mahwah, NJ: Lawrence Erlbaum Associates.

Jacobson, M.J., & Spiro, R.J. (1995). Hypertext learning environments, cognitive flexibility and the transfer of complex knowledge. *Journal of Educational Computing Research, 12*(4), 301-333.

Jonassen, D.H. (1997). Instructional design models for well-structured and ill-structured problem-solving learning outcomes. *Educational Technology Research and Development, 45*(1), 65-91.

Kennedy, G., Petrovic, T., & Keppell, M. (1998). The development of multimedia evaluation criteria and a program of evaluation for computer aided learning. In R.M. Cordeory, (Ed.), *Proceedings of the 15th Annual Conference of the Australian Society for Computers in Tertiary Education (ASCILITE)* (pp. 407-415.). Wollongong, Australia: University of Wollongong.

Land, S.M. (2000). Cognitive requirements for learning with open-learning environments. *Educational Technology Research and Development, 48*(3), 61-78.

Moreno, R. (2004, April). Agent-based methods for multimedia learning environments: What works and why? Paper presented at the *annual meeting of the American Educational Research Assocation*, San Diego, CA.

Myers, J.H., & Dorsey, J.K. (1994). Using diagnostic reasoning (DxR) to teach and evaluate clinical reasoning skills. *Academic Medicine, 69*, 429.

Shaw, E., Johnson, W.L., & Ganeshan, R. (1999). Pedagogical agents on the Web. In *Proceedings of the Third Int'l Conf. on Autonomous Agents* (pp. 283-290).

Van den Ende, J., Blot, K., Kestens, L., Van Gompel, A., & Van den Enden, E. (1997). KABISA: An interactive computer-assisted training program for tropical diseases. *Medical Education, 31*, 202-209.

Winne, P.H., & Marx, R.W. (1982). Students' and teachers' view of thinking processes for classroom learning. *The Elementary School Journal, 82, 493-518.*

KEY TERMS

Animated Pedagogical Agents: Animated figures operating in a learning environment and aiming at supporting learners in their learning process and capable of adapting their support to the learners' paths.

Criterion Path: A representation of an "ideal path" to go through a specific learning environment. It specifies for each possible step in the program what the most ideal subsequent steps are.

Embedded Support Devices: Support devices integrated in the learning environment. Learners cannot but use these devices (e.g., structure in a text).

Instructional Conceptions: Conceptions about the functionalities of (elements of) a learning environment. These conceptions can relate to the effectiveness or efficiency of specific features in a learning environment (e.g., tools) or to the environment as a whole (e.g., KABISA as a learning environment).

Non-Embedded Support Devices (synonym: Tools): Support devices that are put to the disposal of learners. Learners decide themselves when and how to use these tools.

Open Ended Learning Environments: A learning environment that aims at fostering complex problem solving skills by confronting learning with a realistic or authentic problem in a learning environment with a large amount of learner control and different tools.

Perceptions: Students' perceptions relate to how they perceive a specific environment (c.q., KABISA), they are the results of an interaction between students' instructional conceptions and a specific learning environment.

ENDNOTES

[1] A more extended version of this manuscript was published in Armstrong, A. (Ed.). (2004). *Instructional design in the real world. A view from the trenches.* Hershey, PA: Idea Group Inc.

[2] The authors express their gratitude to Stefano Laganà who spend a lot of effort in the adaptation of KABISA and in the development of a log file system.

[3] The term 'characteristic' refers to either a symptom or disease characteristics, either a request for results of a physical examination, laboratory test, or imaging. An example of a characteristic can be 'the patient has severe headache' or 'the palpation of the abdomen is negative.'

Evolution of ERP Systems

Mohammad A. Rashid
Massey University–Albany, New Zealand

INTRODUCTION

The functional units of today's complex business environment require more and more interfunctional data flow for decision making, timely and efficient procurement of product parts, management of inventory, accounting, human resources, and distribution of goods and services. Management of such organizations need efficient information systems to improve competitiveness by cost reduction and better logistics. Enterprise resource-planning systems (ERP), or enterprise systems (Brady, Monk, & Wagner, 2001; Grant, 2003; Hamilton, 2002; Hossain, Patrick, & Rashid, 2002; O'Leary, 2000), are such software systems for business management encompassing modules supporting functional areas such as planning, manufacturing, sales, marketing, distribution, accounting, finances, human resource management, project management, inventory management, service and maintenance, transportation, and e-business. The architecture of the software facilitates transparent integration of modules providing flow of information between all functions within the enterprise in a consistently visible manner. Corporate computing with ERPs allows companies to implement a single integrated system by replacing or reengineering their mostly incompatible legacy information systems.

The concept of the ERP system as "one database, one application and a unified interface across the entire enterprise" (Tadjer, 1998) can be illustrated, following Davenport (1998), with the diagram in Figure 1. The American Production and Inventory Control Society (APICS, 2001) has defined ERP systems as "a method for the effective planning and controlling of all the resources needed to take, make, ship and account for customer orders in a manufacturing, distribution or service company." An ERP system "comprises of a commercial software package that promises the seamless integration of all the information flowing through the company—financial, accounting, human resources, supply chain and customer information" (Davenport, 1998). They are "configurable information systems packages that integrate information and information-based processes within and across functional areas in an organisation" (Kumar & Van Hillsgersberg, 2000).

ERP systems surfaced in the market in the late 1980s and the beginning of the 1990s, targeting mainly large complex business organizations. During the 1960s, most organisations designed, developed, and implemented centralised computing systems, mostly automating their inventory control systems using inventory control (IC) packages. Material requirements planning (MRP) systems were developed in the 1970s and involved mainly planning the product or parts requirements according to the master production schedule. Following this route, new software systems called manufacturing resources

Figure1. ERP systems concept

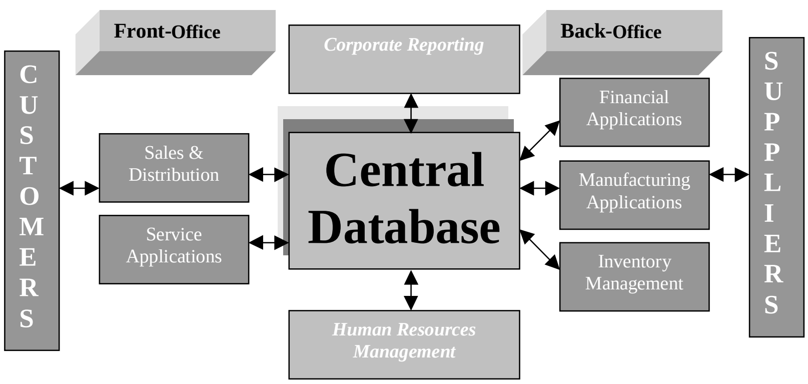
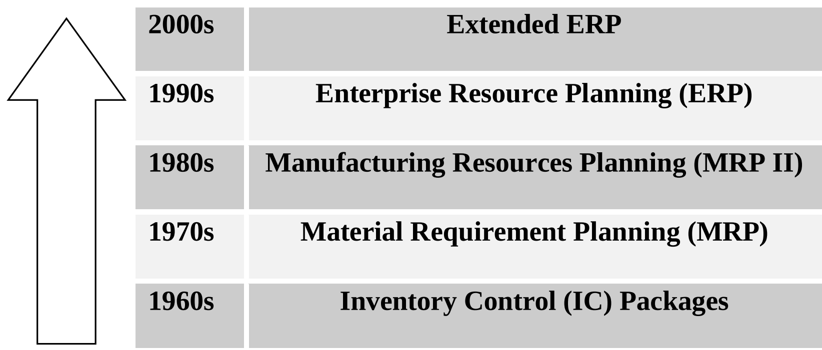


Figure 2. ERP evolution



planning (MRP II) were introduced in the 1980s with an emphasis on optimizing manufacturing processes by synchronising the materials with production requirements. MRP II included areas such as shop floor and distribution management, project management, finance, human resource, and engineering. Based on the technological foundations of MRP and MRP II, ERP systems integrate business processes including manufacturing, distribution, accounting, finances, human resource management, project management, inventory management, service and maintenance, transportation providing accessibility, visibility, and consistency across the enterprise.

During the 1990s, ERP vendors added more modules and functions as "add-ons" to the core modules, giving birth to the "extended ERPs." These ERP extensions include advanced planning and scheduling (APS), e-business solutions such as customer relationship management (CRM), and supply chain management (SCM). Figure 2 summarises the historical events related with ERP.

Organisations choose and deploy ERP systems for many tangible and intangible benefits and strategic reasons. In many cases, the calculation of return on investment (ROI) is weighted against the many intangible and strategic benefits. The benefits that an ERP system may bring to organizations are shown in Table 1 while Table 2 shows the problems and disadvantages organisations need to overcome to reap the benefits.

Table 1. Advantages of ERP systems

What Benefit	How
Reliable information access	Common DBMS (database management system), consistent and accurate data, improved reports
Avoid data and operations redundancy	Modules access same data from the central database, avoid multiple data input, and update operations
Delivery- and cycle-time reduction	Minimizes retrieving and reporting delays
Cost reduction	Time savings, improved control by enterprise-wide analysis of organisational decisions
Easy adaptability	Changes in business processes, easy to adapt and restructure
Improved scalability	Structured and modular design with add-ons
Improved maintenance	Vendor–supported, long-term contract as part of the system procurement
Global outreach	Extended modules such as CRM and SCM
E-commerce, e-business	Internet commerce, collaborative culture

Table 2. Disadvantages of ERP systems

Disadvantage	How to Overcome
Time consuming	Minimize sensitive issues, internal politics, and raise general consensus
Expensive	Cost may vary from thousands of dollars to millions, business process reengineering cost may be extremely high
Conformity of the modules	The architecture and components of the selected system should conform to the business processes, culture, and strategic goals of the organisation
Vendor dependence	Single vendor versus multivendor consideration, options for "best of breeds," long-term committed support
Feature and complexity	ERP system may have too many features and modules that the user needs to consider carefully and implement the needful only
Scalability and global outreach	Look for vendor investment in R&D, long-term commitment to product and services, consider Internet-enabled systems
Extended ERP capability	Consider middleware add-on facilities and extended modules such as CRM and SCM

An ERP system is required to have the following characterisitics:

- Modular design comprising many distinct business modules.
- Use centralized common database management system (DBMS).
- The modules are integrated and provide seamless dataflow.
- They are generally complex systems involving high cost.
- They are flexible and offer best business practices.
- They require time-consuming tailoring and configuration setups.
- The modules work in real-time: on-line and batch processing capabilities.
- They are or soon they will be Internet-enabled.

Different ERP vendors provide ERP systems with some degree of specialty, but the core modules are almost the same for all of them. Some of the core ERP modules found in the successful ERP systems is the following.

- Accounting management
- Financial management
- Manufacturing management
- Production management
- Transportation management
- Sales and distribution management
- Human resources management
- Supply chain management
- Customer relationship management
- E-business

COMMERCIAL ERP SYSTEMS

The dominating ERP software suppliers in 2003 were SAP, PeopleSoft, Oracle, SSA Global, Microsoft, and Sage. According to IDC (International Data Corp), the top-ten vendors in 2003 had a market share of 46% of the $25 billion ERP market. Each vendor, due to historic reasons, has a specialty in one particular module area such as SAP in logistics, PeopleSoft in human resources management, Oracle in financials, and SSA Global's SSA Baan ERP in manufacturing. There are also about 50 established and a few more newly emerging smaller and midsize ERP vendors and third-party developers competing for the ERP market. The result is stiff competition and feature-overlapping products difficult to differentiate. Long-term vision, commitment to service and support, module features, specialty, experience, and financial strength for R&D are the major vendor qualities for product selection and turnkey implementation. Table 3 shows the brief profile of these dominating ERP vendors.

Table 3. Brief profile of the dominating ERP vendors

Vendor Name	*Founding Year*	*Flagship Product*
SAP AG	Germany, 1972	R/3, mySAP.COM
PeopleSoft Inc.	USA, 1987	PeopleSoft8
Oracle Corporation	USA, 1977	*Oracle Applications* Oracle11i
SSA Global	USA, 1981	*SSA Baan ERP*
Sage	UK, 1980	Sage Line 500
Microsoft (Microsoft Business Solutions)	USA, 1981	Great Plains, Navision, Axapta, and Solomon

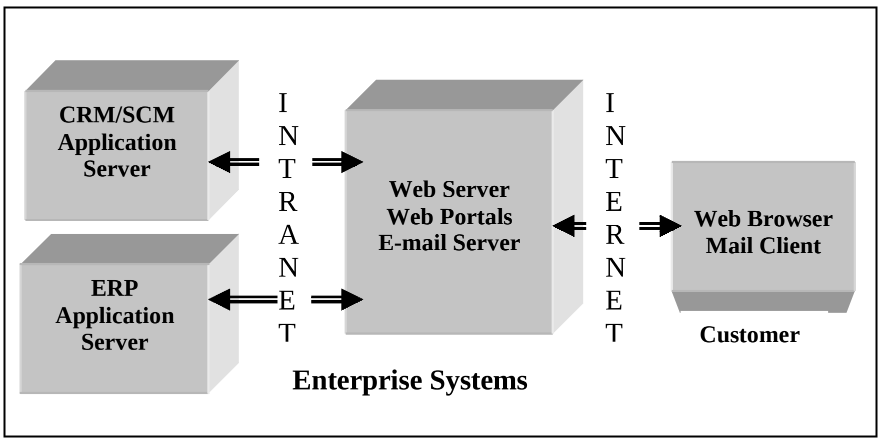
The continued growth of the ERP systems market is attributed to the fact that the vendors are adding applications such as supply chain management, customer relationship management, and the integration of Internet-enabled applications for e-business. ERP vendors are targeting the untapped SME (small and medium-sized enterprise) market with supposedly scaled-back systems suitable for smaller firms by offering simple, cheaper, and preconfigured, easy-to-install solutions within budget and time constraints. For some vendors, this may lead to offering centrally managed, Internet-enabled, ERP-system-based services for SMEs to access and use anytime from anywhere.

ERP systems are becoming more and more Internet-enabled (Lawton, 2000), extending legacy ERP systems to integrate with newer external business modules such as supply chain management, customer relationship management, sales-force automation (SFA), advanced planning and scheduling, business intelligence (BI), and e-business capabilities. In fact, ERP is becoming the e-business backbone for organisations doing online business transactions over the Internet. Adoption of e-commerce and e-business solutions, especially business-to-business (B2B) solutions, is seen by many as the wave of current and future extensions of traditional ERP systems of most small, medium, and large vendors. The front-end, Web-based Internet-business applications are integrated with the back-office, ERP-based applications enabling business transactions such as order placement, purchasing, inventory updates, employee benefits, and so forth to take place between the customers, suppliers, and the enterprise based on reliable, relevant data and applications instantly in a borderless domain. Examples of such Internet-enabled, extended ERPs are available from most of the ERP vendors. The concept of the Internet-enabled, extended ERP system is shown in Figure 3.

CONCLUSION

The unprecedented growth of computing power and the proliferation of the Internet brings new challenges to ERP vendors for redesigning their products, breaking the barrier of proprietorship and customisation, and embrac-

Figure 3. Web-enabled extended ERP system



ing the collaborative business over intranet, extranet, and the Internet in a seamless manner. A sign of acceptance of these challenges is the appearance of new add-on modules that follow open architecture, provide interchangeable modules, and allow easy customisation and user interfacing.

There are scopes for emerging ERP vendors to contribute to the field, especially in the aerospace industry, the finance industry, and the logistics industry. Analysis of the market penetration of ERP systems shows clearly that the current players have to downsize their products and offerings to be attractive to SMEs. Changes in ERP software architecture taking advantage of the smaller but powerful hardware platforms can provide significant opportunities to smaller players to bring new small, robust, easy-to-use systems into the market suitable for SMEs. Future successful vendors will capture large markets of smaller businesses by providing a more consistent and enduring income stream.

REFERENCES

APICS. (2001). American Production and Inventory Control Society (APICS). Retrieved from *http://www.apics.org*

Bakos, Y. (1998). The emerging role of electronic marketplaces on the Internet. *Communications of the ACM, 41*(8), 35-42.

Bingi, P., Sharma, M. K., & Godla, J. K. (1999). Critical issues affecting an ERP implementation. *Information Systems Management, 16*(3), 7-14.

Brady, J., Monk, E. F., & Wagner, B. J. (2001). *Concepts in enter resource planning.* Course Technology.

Broatch, M. (2001, July). Making the ERP connection. *Computerworld*.

Davenport, T. H. (1998). Putting the enterprise into the enterprise system. *Harvard Business Review, 76*(4), 121-131.

Davenport, T. H (2000). *Mission critical: Realizing the promise of enterprise systems.* Boston: Harvard Business School Press.

Grant, G. G. (2003). *ERP and data warehousing in organizations: Issues and challenges.* Hershey, PA: IRM Press.

Hamilton, S. (2002). *Maximizing your ERP system: A practical guide for managers.* McGraw-Hill Trade.

Holland, C., & Light, B. (1999, May/June). A critical success factors model for ERP implementation. *IEEE Software*, 30-36.

Hossain, L., Patrick, J. D., & Rashid, M. A. (2002). *Enterprise resource planning: Global opportunities & challenges.* Hershey, PA: Idea Group Publishing.

Kalakota, R., & Robinson, M. (1999). *E-business roadmap for success.* Addison-Wesley.

Kumar, K., & Van Hillsgersberg, J. (2000). ERP experiences and evolution. *Communications of the ACM, 43*(4), 23-26.

Lawton, G. (2000). *Integrating ERP and CRM via the Web.* SW Expert.

Li, C. (1999). ERP packages: What's next? *Information Systems Management, 16*(3), 31-35.

Norris, G., Dunleavy, J., Hurley, J. R., Ballis, D., & Hartley, K. M. (2000). *E-business and ERP: Transforming the enterprise.* John Wiley & Sons.

O'Leary, D. E. (2000). *Enterprise resource planning systems: Systems, life cycle, electronic commerce, and risk.* United Kingdom: Cambridge University Press.

Shields, M. G. (2001). *E-business and ERP: Rapid implementation and project planning.* John Wiley & Sons.

Tadjer, R. (1998, April 13). Enterprise resource planning. *Internetweek*.

Wallace, T. F., & Kremzar, M. H. (2001). *ERP—Making it happen: The implementers' guide to success with enterprise resource planning.* John Wiley & Sons.

KEY TERMS

APICS: American Production and Inventory Control Society. A not-for-profit international, educational organization founded in 1957, respected throughout the world for its education and professional certification programs.

Customer Relationship Management (CRM): Software systems that help companies to acquire knowledge about customers and deploy strategic information systems to optimize revenue, profitability, and customer satisfaction

Extended ERP: Extends the foundation ERP system's functionalities such as finances, distribution, manufacturing, human resources, and payroll to customer relationship management, supply chain management, sales-force automation, and Internet-enabled integrated e-commerce and e-business

Fat Client/Server Architecture (C/S): A client/server network architecture where the client (PCs or workstations on which users run applications) performs the bulk of the data-processing operations while the data itself is stored on the server

Graphical User Interface (GUI): A software interface based on the computer's graphics capabilities with pointing devices such as the mouse that free the user from learning complex commands for using programs

Return On Investment (ROI): An estimate expressed as a ratio comparing the net benefits (the "return") to its total cost (the "investment") for measuring operating performance and efficiency in utilizing the assets of the company

Small and Medium-Sized Enterprise (SME): A business enterprise independently owned by contributing most of the operating capital and managed by the owners or managers, having fewer than 250 employees and a small-to-medium market share

Supply Chain Management (SCM): Software systems for procurement of materials, transformation of the materials into products, and distribution of products to customers, allowing the enterprise to anticipate demand and deliver the right product to the right place at the right time at the lowest possible cost to satisfy its customers

Exception Rules in Data Mining

Olena Daly
Monash University, Australia

David Taniar
Monash University, Australia

INTRODUCTION

Data mining is a process of discovering new, unexpected, valuable patterns from existing databases (Frawley, Piatetsky-Shapiro, & Matheus, 1991). Though data mining is the evolution of a field with a long history, the term itself was only introduced relatively recently, in the 1990s. Data mining is best described as the union of historical and recent developments in statistics, artificial intelligence, and machine learning. These techniques are then used together to study data and find previously hidden trends or patterns within.

Data mining is finding increasing acceptance in science and business areas that need to analyze large amounts of data to discover trends, in which they could not otherwise find. Different applications may require different data mining techniques. The main kinds of knowledge that could be discovered from a database are categorized into association rules mining, sequential patterns mining, classification and clustering.

In this article, we concentrate on exception rules mining.

BACKGROUND

Exception rules mining has attracted a lot of research interest (Déjean, 2002; Grosof & Poon, 2002, 2003; Hellerstein, Ma, & Perng, 2002; Hussain, Liu, Suzuki, & Lu, 2000; Keogh, Lonardi, & Chiu, 2002; Liu, Hsu, Mun, & Lee, 1999; Padmanabhan & Tuzhilin, 2000; Suzuki, 2002a, 2002b; Yamada & Suzuki, 2002; Zhang, Zhang, Yan, & Qin, 2002). Exception rules have been defined as rules with low support and high confidence (Hussain et al., 2000). A traditional example of exception rules is the rule Champagne=>Caviar. The rule may not have high support but it has high confidence. The items are expensive so they are not frequent in the database, but they are always brought together so the rule has high confidence. Exception rules provide valuable knowledge about database patterns.

Exception rules discovery can be classified as either directed or undirected. A directed search obtains a set of exception rules each of which contradicts to a user-specified belief (Liu et al., 1999; Padmanabhan, 2000). An undirected search obtains a set of pairs of an exception rule and a general rule (Hussain et al., 2000; Suzuki, 2002a, 2002b; Yamada & Suzuki, 2002).

Directed search of exception rules will be described next. User-specified beliefs are obtained first. Each of the discovered exception rules contradicts to user-supplied beliefs.

In Liu et al. (1999), post-analysis of the discovered database patterns is performed to identify the most interesting patterns. The technique is characterized by asking the user to specify a set of patterns according to his/her previous knowledge or intuitive feelings. This specified set of patterns is then used by a fuzzy matching algorithm to match and rank the discovered patterns. The assumption of this technique is that some amount of domain knowledge and the user's interests are implicitly embedded in his/her specified patterns. In general, the discovered patterns are ranked according to their conformities to the user's knowledge or their unexpectedness, or their actionabilities.

In terms of unexpectedness, patterns are interesting if they are unexpected or previously unknown to users. In terms of actionability, patterns are interesting if users can do something with them to their advantage. With such rankings, a user can simply check the few patterns on the top of the list to confirm his/her intuitions (or previous knowledge), or to find those patterns that are against his/her expectation, or to discover those patterns that are actionable.

Padmanabhan and Tuzhilin (2000) focus on discovering unexpected patterns and propose methods for discovering a minimal set of unexpected patterns that discover orders of magnitude fewer patterns and retain most of the interesting ones. The approach has been experimentally tested using a case study application in a marketing domain.

The rule A=>B is defined in Padmanabhan and Tuzhilin (2000) to be unexpected with respect to the belief X=>Y on the dataset D if B and Y logically contradict each other, the antecedents of the belief and the rule hold on the same statistically large subset of D, and the rule A, X=>B holds.

Now the undirected method of searching exception rules will be explained. Exception rules will be obtained based on general rules or common sense rules.

In Hussain et al. (2000), a method for mining exception rules is presented based on a novel measure which estimates interestingness relative to its corresponding common sense rule and reference rule. Common sense rules are rules with high support and high confidence. Reference rules are rules with low support and low confidence. Exception rules are defined as rules with low support and high confidence.

The formula for the relative interestingness measure RI in Hussain et al. (2000) is derived based on information theory and statistics. The measure has two components, which are interestingness based on the rule's support and interestingness based on the rule's confidence.

Suzuki (2002a) introduces undirected discovery of exception rules, in which a pattern represents a pair of an exception rule and its corresponding strong rule. Proposed scheduled discovery and exception rule discovery guided by a meta-pattern are described and tested on data sets.

Suzuki (2002b) presents an algorithm for discovering exception rules from a data set without domain-specific information. The method is based on sound pruning and probabilistic estimation. The normal approximations of the multinomial distributions are employed as the method for evaluating reliability of a rule pair. The method has been validated using two medical data sets and two benchmark data sets in the machine learning community.

The main contribution of Yamada and Suzuki (2002) is the formalization of spiral discovery for interesting exception rules and a method that employs initial knowledge, MDL-based discretization and reduction of the number of discovered rule pairs. The experimental evaluation was performed on meningitis data set.

EXCEPTION RULES MINING

A new approach to mine exception rules will be proposed in this section. The approach belongs to the category of directed search. An interconnection between exception rules and strong association rules will be considered. As opposed to the research work described in the previous section, both strong positive and negative association rules are considered.

Based on the knowledge about positive and negative association rules in the database, the candidate exception rules will be generated. A novel exceptionality measure will be proposed to evaluate the candidate exception rules. The candidate exceptions with high exceptionality will form the final set of exception rules.

In order to formulate the proposed approach, a few data mining terms have to be defined. Itemset is a set of database items. For example, itemset XY means a set of two items X and Y. Association rule is an implication of the form X=>Y, where X and Y are database itemsets. An example of an association rule could be supermarket items Chips=>Coke purchased together frequently.

The rule X =>Y has support s, if s% of all transactions contain both X and Y. The rule X =>Y has confidence c, if c% of transactions that contain X, also contain Y. In association rules mining user-specified minimum support (minsup) and minimum confidence (minconf) are given.

Association rules with support greater or equal to minsup and confidence greater or equal to minconf are referred to as strong rules (Agrawal, Imielinski, & Swami, 1993; Agrawal & Srikant, 1994).

Itemsets that have support at least equal to minsup are called frequent itemsets. Negative itemsets are itemsets that contain both items and their negations. For example, consider the negative itemset X~Y. In this itemset ~Y means negation of item Y (absence of item Y in the database record).

Negative association rule is an implication of the form X=>~Y, ~X=>Y, ~X=>~Y, where X and Y are database items, ~X, ~Y are negations of database items. An example of a negative association rule is Coke=>~Pepsi, which means that people do not buy Coke and Pepsi together.

In our approach, the search for exception rules will be based on the knowledge about strong association rules in the database. An example: we discover a strong association rule in the database, for instance, shares of companies X and Y most times go up together X=>Y. Then those cases when shares of the companies X and Y do not go up together, X=>~Y or ~X=>Y, we call *exceptions* when satisfying the proposed *exceptionality* measure explained next. An algorithm for mining exception rules based on the knowledge about association rules will be proposed in as well.

We explain a few proposed definitions first. For exception rules mining instead of minsup we employ *lower* and *upper bounds*, satisfying the conditions: 0<lower bound<upper bound<minsup;

Low support belongs to the range [lower bound; upper bound]. *Infrequent itemsets* have low support. Note that the lower bound is always greater than 0, as we are not interested in rules with 0 support or close to 0. Upper bound is lower than minsup. The lower and upper bounds are chosen specifically for each data mining application.

Exception Rules are rules with *low* support and high *exceptionality* values. Infrequent itemsets with high exceptionality are called *exceptional* itemsets.

In the proposed exception rules mining the confidence measure is not applicable to evaluate the exception rules.

$$\text{Exceptionality(CandExc/AssosRule)} = \text{FuzzySup(CandExc)} + \text{FuzzyFraction(CandExc/AssosRule)} + \\ + \text{Neglect(CandExc/AssosRule)}$$

For example, we obtain a strong rule A=>B and would like to evaluate a potential exception rule A=>Not B. The strong rule A=>B has high confidence, implying that A=>Not B cannot have high confidence. Let us say the minimum confidence is 60%. The strong rule A=>B satisfies the minimum confidence constraint, so at least 60% of database records containing A also contain B. It means that maximum 40% records containing A do not contain B. The exception rule A=>Not B has maximum 40% confidence. As confidence is not applicable for evaluating exception rules, we propose a special measure *exceptionality* to evaluate the exception rules.

Exceptionality of a candidate exception rule given the corresponding association rule is defined by the above formula.

Refer to Daly and Taniar (2004) for the details of our proposed exceptionality measure.

We propose a novel exception rules classification and explain the premises of mining exceptions based on the negative association rules in data bases. We suggest two general types of exception rules, which are exceptions in positive sense and exceptions in negative sense. Exceptions in negative sense will be descried next.

After basic mining for positive and negative association rules in a database we obtain steady patterns of database items that occur together frequently. Let us say X and Y are database items and

$$\left.\begin{array}{l} X \\ Y \end{array}\right\} \text{frequent} \qquad XY = \text{fre} \qquad (1)$$

$$X => Y \qquad \text{high confidence}$$

Also we obtain that

$$\left.\begin{array}{l} X\sim Y \\ \text{or} \\ \sim XY \end{array}\right\} \text{infrequent} \qquad (2)$$

So we have a strong association rule (1), and we make sure that (2) are infrequent. Rules (1) and (2) are our premises to check if one of the rules (3) has a high exceptionality, which would prove it is an exception in negative sense.

$$\left.\begin{array}{l} X=>\sim Y \\ \text{or} \\ \sim X=>Y \end{array}\right\} \text{high exceptionality — (Exception)} \qquad (3)$$

For example, consider two oil companies X and Y. Their stock normally goes up at the same time: X=>Y. In the case when their shares do not go up at the same time X=>~Y, we call the rule X=>~Y an exception if X~Y is infrequent and has high exceptionality measure.

Exceptions in positive sense will be descried next. After basic mining for positive and negative association in a database, we obtain a steady pattern of database items. Let us say X and Y are database items and

$$\left.\begin{array}{l} X \\ Y \end{array}\right\} \text{frequent} \qquad X\sim Y = \text{frequent} \qquad (4)$$

$$\left.\begin{array}{l} X=>\sim Y \\ \text{or} \\ \sim Y=>X \end{array}\right\} \text{high confidence}$$

Also we obtain that

$$XY = \text{infrequent} \qquad (5)$$

We have a strong negative association rule (4), and we make sure that (5) is infrequent. (4) and (5) are our premises to check if one of the rules (6) has a high exceptionality, which would prove it is an exception rule in positive sense.

$$\left.\begin{array}{l} X=>Y \\ \text{or} \\ Y=>X \end{array}\right\} \text{high exceptionality — (Exception)} \qquad (6)$$

For example, consider two oil companies X and Y. Their stock never goes up at the same time: X=>~Y. In the case when their shares do go up at the same time X=>Y, we call the rule X=>Y an exception if XY is infrequent and has high exceptionality measure.

In Figure 1, we present an algorithm for mining exception rules based on strong association rules. Association rules are generated from frequent itemsets with high confidence. The confidence calculation is a straightforward procedure after all frequent itemsets have been generated. We do not consider the confidence calculation as it is easy and conceptually proven correct. The input of the exception rules mining algorithm are frequent 1-itemsets. The output of the algorithm is exceptional itemsets. Exceptional itemsets will become exception rules after the confidence of association rules has been checked.

Figure 1. Exceptional itemsets generation algorithm

```
k=1
1-freq_itemsets                 // generate frequent 1-itemsets
k=2
2_candidate_itemsets            // generate candidate 2-itemsets
forEach c in 2_candidate_itemsets
 if (c frequent)                // verify the condition (1)
        if (negative_sets infrequent)   // verify the condition (2)
                {generate_2_Exc_cand_negative
                 check_Exceptionality: true: //condition(3): if high Exceptionality
                          ExceptionalItemsets.Add()  //to array of exceptional itemsets
             }
     else
         if (negative_sets frequent)     // verify the condition (4)
                {generate_2_Exc_cand_positive
                check_Exceptionality: true://condition (6): if high Exceptionality
                         ExceptionalItemsets.Add() //to array of exceptional itemsets

              }
k++
```

We generate frequent itemsets and on each step k (k = length of the itemset) we check the conditions (1), (2) or (4), (5) described previously and if they hold true, we check the exceptionality values for candidate exceptions.

Refer to Daly and Taniar (2004) for details of our proposed algorithm and performance evaluation on a data set.

FUTURE TRENDS

The overview of the future trends or open research questions in exception rules mining is presented in this section.

Future Trends:

- Develop a novel approach in addition to directed/ undirected search
  Directed search of exception rules has been recognized as a subjective search method. It is up to the user to define the system of the beliefs and expected patterns. Based on the obtained belief system, the contradicting patterns will become exception rules in the given domain. Therefore, the generated exception rules may be contradictive given different user input.
  On the other hand, undirected search generates exception rules independently from domain knowledge or user experience. Exception rules are based on the common sense rules (strong patterns).
  There could be another way of mining exception rules, independently of user beliefs or strong patterns in the database. One of open research areas is to identify such a novel method of mining exceptions rules in databases.

- In directed search
  One of the research issues is to create a uniform user belief system for the given domain. It may involve a lot of data collecting from a number of experienced users/decision makers. Obviously, the uniform belief system would have to be updated on regular basis. Besides, it may be possible to develop a limited system applicable to a number of domains.

- In undirected search
  Develop novel interest measures for exception rules evaluations. Develop new resource-efficient algorithms for mining exception rules

CONCLUSION

Exception rules mining has attracted a lot of research interest. A lot of interesting works have been published and a lot to appear. This is an open research area.

The main approaches in exception rules mining have been highlighted and potential future research directions have been discussed. There are two main approaches in mining exception rules. The first approach is directed discovery of exception rules relying on user-specified beliefs. The second approach generates pairs of common sense rules and corresponding exception rules.

This article is a review of the research that has been done in the exception rules mining and a good starting point for people who desire to contribute in research in the area.

REFERENCES

Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *Proceedings of International Conference Management of Data*, (pp.207-216).

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules in large databases. *Proceedings of the 20th International Conference on Very Large Data Bases*, (pp.487-499).

Daly, O., & Taniar, D. (2004). Exception rules mining based on negative association rules. *Computational Science and Its Applications-ICCSA 2004, Part IV, Lecture Notes in Computer Science LNCS.* Springer-Verlag, (pp. 543-552).

Déjean, H. (2002). Learning rules and their exceptions. *Journal of Machine Learning Research, 2*, 669-693.

Frawley, W., Piatetsky-Shapiro, G., & Matheus, C. (1991). Knowledge discovery in databases: An overview. In G. Piatetsky & W. Frawley (Eds.), *Knowledge discovery in databases*, (pp.1-30). American Association for Artificial Intelligence Press.

Grosof, B., & Poon, T. (2002). Representing agent contracts with exceptions using XML rules, ontologies, and process descriptions. *Proceedings of 2002 RuleML Conference*.

Grosof, B., & Poon, T. (2003). SweetDeal: Representing agent contracts with exceptions using XML rules, ontologies, and process descriptions. *Proceedings of 2003 World Wide Web Conference*, (pp.340-349).

Hellerstein, J., Ma, S., & Perng, C. (2002). Discovering actionable patterns in event data. *IBM Systems Journal, 41*(3), 475-493.

Hussain, F., Liu, H., Suzuki, E., & Lu, H. (2000). Exception rule mining with a relative interestingness measure. *Proceedings of the 4th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, (pp.86-97).

Keogh, E., Lonardi, S., & Chiu, B. (2002). Finding surprising patterns in a time series database in linear time and space. *Proceedings of 2002 Knowledge Discovery in Databases Conference*, (pp.550-556).

Liu, B., Hsu, W., Mun, L., & Lee, H. (1999). Finding interesting patterns using user expectations. *Institute of Electrical and Electronics Engineers, Transactions on Knowledge and Data Engineering, 11*(6), 817-832.

Padmanabhan, B., & Tuzhilin, A. (2000). Small is beautiful: Discovering the minimal set of unexpected patterns. *Proceedings of the 6th International Conference on Knowledge Discovery and Data Mining*, (pp.54-63).

Suzuki, E. (2002a). In pursuit of interesting patterns with undirected discovery of exception rules. *Progress in Discovery Science, Lecture Notes in Computer Science 2281*, State-of-the-Art Surveys, Springer-Verlag, (pp.504-517).

Suzuki, E. (2002b). Undirected discovery of interesting exception rules. *International Journal of Pattern Recognition and Artificial Intelligence, 16*(8), 1065-1086.

Yamada, Y., & Suzuki, E. (2002). Toward knowledge-driven spiral discovery of exception rules. *Proceedings of the 2002 Institute of Electrical and Electronics Engineers International Conference on Fuzzy Systems, 2*, 12-17, pp. 872-877.

Zhang, S., Zhang, C., Yan, X., & Qin, Z. (2002). Identifying exceptional patterns in multi-databases. *Proceedings of the 1st International Conference on Fuzzy Systems and Knowledge Discovery*, (pp.146-150).

KEY TERMS

Association Rules: An implication of the form A=>B, where A and B are database itemsets. Association rules have to satisfy the pre-set minimum support (minsup) and minimum confidence (minconf) constraints.

Confidence: The rule A=>B has confidence *c*, if *c%* of transactions that contain A, also contain B.

Exception Rules: Rules with low support and high confidence.

Frequent Itemsets: Itemsets that have support at least equal to minsup.

Itemset: A set of database items.

Negative Association Rules: An implication of the form X=>~Y, ~X=>Y, ~X=>~Y, where X and Y are database items, ~X, ~Y are negations of database items.

Negative Itemsets: Itemsets that contain both items and their negations.

Support: The rule A=>B has support *s*, if *s%* of all transactions contain both A and B.

Executive Judgment in E-Business Strategy

Valerie Baker
University of Wollongong, Australia

Tim Coltman
University of Wollongong, Australia

INTRODUCTION

One of the main strategic challenges for organizations today is to effectively manage change and stay competitive in the future. Change appears to be the only constant in contemporary business and is present in every industry and in every country (Brown & Eisenhardt, 1998). Moreover, the key area of importance, current within many organizations, is how to effectively leverage technology within such a complex and dynamic business environment (Sauer & Willcocks, 2003). The alignment or fit approach, which has its roots in contingency theory, has long been promoted as the way to get high returns from technology investment. However, the realization of advantage from the Internet and related e-business technology investment has long been a source of frustration for corporate executives. Impressive performance returns by companies such as Dell Computers, Cisco Systems and General Electric illustrate that returns can be achieved by linking the Internet and related e-business technologies to firm strategy. These companies have shown that successful management of their IT investments can generate returns as much as 40% higher than those of their competitors (Ross & Weill, 2002). Yet, many executives view the Internet and related e-business technologies with intense frustration. They recollect investment in the great speculative bubble of the 1990s and excessive expenditure on year 2000 (Y2K) compliant systems (Keen, 2002). They recall high profile examples of botched enterprise resource planning (ERP) systems that have consistently run over time and budget and report that customer-relationship management (CRM) initiatives were largely a flop (Reinartz & Chugh, 2002). Unfortunately, it is not yet clear how firms should go about capturing the potential that exists in e-business, as few normative frameworks exist to guide practitioner investment.

BACKGROUND

One area of scholarly activity where consistent advances have been made regarding the determinants of firm performance is in structural contingency theory. Here, the contingency factor (i.e., environment-structure) has enabled predictions to be made in a relatively unambiguous manner (Donaldson, 1995). Applied to an e-business setting, contingency theory argues that performance increases can be expected whenever information technology is applied in an appropriate and timely way, in harmony with business, environmental and organizational conditions. Consider a typical scenario where an executive wants to make a strategic investment in information systems. They have two choices: (1) a system to support backend operations using ERP technology, and (2) a CRM support system. How do they prioritise between these competing investments? Contingency literature would argue that it depends upon the organization's strategy and decision-making information requirements (Chandler, 1962; Child, 1972; Galbraith & Kazanjian, 1986). Manufacturing excellence strategies associated with companies like Carrefour or Ford Motor Company would get greater value from ERP systems. Customer intimacy strategies at companies like CitiBank or IBM Global Services would benefit most by customer feedback systems.

As simple as this observation may appear, the application of alignment has proven elusive. Despite 20 years of effort and investment in consulting advice, CIOs are still struggling with the same set of alignment problems. A recent survey by CIO Insight (Patterson, 2001) highlights the point that only 34% of organizations considered the link between their IT priorities and their enterprise strategy to be "strong." While these statistics reflect the difficulties of coordinating complex organizations, they provide evidence that most managers are not using the basic tools of alignment that have been developed over several decades of research.

Priem and Cycyota (2000) equate the process of alignment between IT strategies and business goals with executive judgment. The literature regarding judgment theory argues that firm success can be explained by the judgments executives make concerning the current state of the environment and the vision of the organization. In uncertain times, where market pressures and time constraints dominate the business landscape senior manager's perceptions, skill and vision often form the basis on which strategic choices regarding IT investments are made. For example, it takes little more than a browsing of the management section of the local bookstore—blazoned with titles such as Inside the Minds: Leading CEOs—or a visit to the local news agent to pick up a recent copy of Forbes, Fortune or Business Week to recognize the importance that publishers and managers place on the philosophies and actions of even some of the least successful or most unlikely of management leaders. Perhaps more relevant is that often the appointment of "higher quality CEOs" leads to immediate stock market reactions and greater long term performance. One such example was the reappointment of Steve Jobs as CEO of Apple Computer. Jobs has been widely praised for his skill in judging the commercial potential of convergent Internet technologies and his return to the company was considered instrumental in its reversal of bad fortunes (Stevens, 1997).

The corollary here is that judgment is an essential skill for setting the overall direction of the organization. In turbulent environments, often the context of e-business, quick trade offs need to take place, as the strategic direction of the firm enables it for the future. This being the case management discretion becomes increasingly important, as decisions are made "on the fly" with little information or understanding of the decision problem. Management play a vital role in "trading off" elements of organization control, that is, structure for better adaptation, a view supported by complexity theory (Brown & Eisenhardt, 1998). This theory views strategy as a process which constantly changes, and thus needs a type of structure or execution method that is dynamic and will allow the organization to be ready for the future.

Thus, although judgment appears to be important to organizational success, scholars have largely ignored executive intentions and no empirical link between executive choices and firm outcomes has been established. Instead, strategic outcomes are presumed to be due to strategic choice (Preim & Harrison, 1994). This omission may account in some part, for why practitioners continue to pay little attention to the large amount of published work concerning the antecedents of strategy and performance. This concern provided the motivation for a special issue of the Academy of Management Journal (AMJ, 1998, p.746) that sought greater understanding of the way knowledge is transferred between academics and practitioners. The issue again surfaced in a recent issue of the Academy of Management Executive, providing evidence that practitioners still typically turn to sources of information other than academics or the scientific literature when searching for ways to improve performance (Ford, Duncan et al., 2003).

FUTURE TRENDS

Clearly, we need greater understanding of the conditions which lead executives to make strategic choices if we are to develop research that has an impact on practitioners.

Existing research into the change process and the implementation of e-business related technology is limited because it fails to measure the link between strategic choices and firm outcomes. As we have suggested, the judgments that executives make provides important insight into how IT strategic change or e-business change is approached given different situations and organization contexts.

Peterson (2002) suggests that it is the processing of information and the judgments that are made by top management that leads to critical decisions being made about how firms deal with IT-related strategic change. As the business environment rapidly changes, the variance in possible outcomes ranges from failure to unparalleled success. These differences can largely be explained by the "mythical relationship between technology ecology, human nature, decision cycles, IT and the speed and veracity of their interactions" (Peterson, 2002, p.485). Executives process information about these relationships and form critical strategic judgments regarding the future direction of their organization through its e-business strategy.

Managers face conditions such as dynamic markets, casual ambiguity and path dependence that make it extremely difficult to predict the outcomes of their IT strategic investments. As this illustration suggests, it is imperative that managers have in place strategies to cope with changes as they occur. Faced with external environmental changes (e.g., new rates of Internet adoption, killer mobile commerce applications, etc.), managers need to be able to adjust their strategic choices accordingly "just as water shapes itself according to the ground, an army should manage its victory in accordance with the enemy. Just as water has no constant shape, so in warfare there are no fixed rules and regulations" (Sun Tzu in Hussey, 1996, p.208)

What Sun Tzu highlights is the requirement that strategies be flexible in order to manage strategic change. Mintzberg, Ahlstrand, and Lampel (1998) describe this as an emergent strategy, where rather than pursuing a strategy, an organization makes decisions based on the situation, effectively testing the market as they go.

Thus strategic decisions regarding IT management need to be a mixture of both deliberate and emergent strategies. "Real-world strategies need to mix these in some way: to exercise control while fostering learning" (Mintzberg et al., 1998, p.11). The importance of strategic alignment between the organization and its environment becomes even more critical given recent environmental turbulence and the evolving importance of technology and e-business to competitive advantage.

Executive Judgment and Strategic Alignment

Priem and Cycyota (2000) state that understanding judgments by strategic leaders is essential to determine the role of mental processes in strategy development and how these strategies and processes affect firm performance. They suggest that a number of theoretical platforms commonly found in the strategy literature provide a solid platform from which we can examine strategic judgement. The "fit" or "alignment" paradigm is perhaps one of the most pervasive in strategy. Good strategy requires at a minimum alignment with changing external conditions. In simple terms, the proposition is that there is an organizational structure that fits the level of contingency factor whether it is environmental uncertainty, organizational characteristics, technological characteristics or strategy design interdependence so that an organization in fit creates significant and positive implications for performance. This idea that fit between organization structure and contingency factor leads to superior performance has been empirically supported in both qualitative and quantitative studies (Donaldson, 1995). Given this distinguished history, it might reasonably be expected that executives would frequently make decisions based on the principles of organization congruence (Priem, 1994).

Most early theories of structural contingency focused on how the fit between bivariate variables (i.e., structure-environment alignment, strategy-structure alignment or strategy-environment alignment) are associated with increased firm performance. However, information technology (IT) is becoming an important substitute for organization structure in modern organisation (Sauer & Willcocks, 2003). For example, Oracle's ability to transform itself into an Internet-enabled business would not have been possible without an appropriate technology base. The wrong technology base would have made such an initiative a massive technological and organizational challenge because of the custom integration required. Emerging evidence indicates that structure and technology are complementary. Where structures create boundaries for management control, technology permits those boundaries to be traversed thereby enabling more complex commercial activities to be effectively integrated and managed.

EMPIRICAL INVESTIGATION OF STRATEGIC JUDGMENT

The "integrative framework" developed by Lee (1989, 1991) in a series of papers regarding the management of information systems provides a suitable approach to the study of judgment. Lee's integrative framework formally presented in his 1994 paper combines three levels of "understandings": the subjective, the interpretive and the positivist. According to Lee, the three understandings are "far from being mutually exclusive and irreconcilable"; in fact, "they may be utilised as mutually supportive and reinforcing steps in organisational research". Priem and Cycyota (2000) also support this view by claiming that both qualitative and quantitative studies are necessary to increase our understanding of strategic judgement.

In the case of e-business, qualitative work can be useful in exploratory investigation that may highlight issues more formal approaches may miss. For example, case studies of the most spectacular strategic information systems initiatives Baxter Healthcare and American Airline's SABRE indicate that these IT/e-business systems were largely accidental success stories (Clemons, 1986). However, these subjective and interpretive studies cannot test hypotheses adequately, because of the close contact needed with research subjects and the resulting small sample size. Quantitative studies provide the crucial positivist link that complements exploratory work in a way that can generate more widely generalizable insights. Notable examples, include the study of IT's contribution to performance in the retail industry (Powell & Dent-Micallef, 1997), and the investigation of organizational antecedents to e-business adoption (Srinivasan, Lilien, & Rangaswamy, 2002).

The following sections focus on measurement techniques, which can be used to examine individual judgment. These techniques can be grouped into two categories: (1) composition methods, and (2) decomposition methods (Priem & Harrison, 1994).

Composition Methods

Composition methods focus on the processes that underlie individual judgments. Composition involves methods such as verbal protocol analysis, information searches and cause mapping to gather interpretive information from executives about the processes that lead them to make certain judgments. These types of techniques would be useful in identifying the variables that executives use in their strategic decision making, but which are not included in current management theory (Priem & Harrison, 1994).

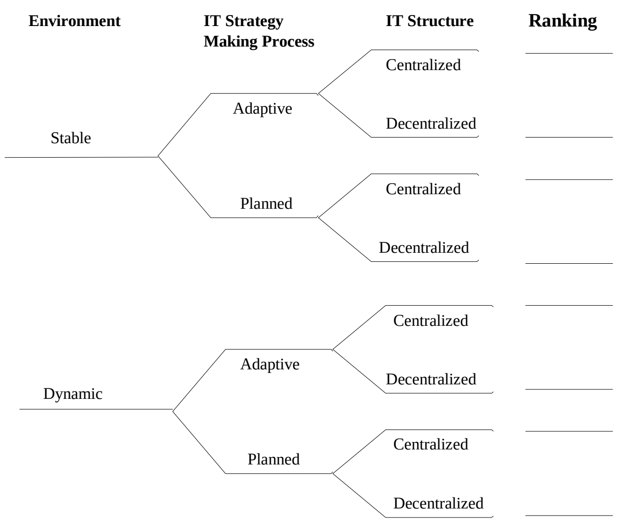
Decomposition Methods

Decomposition techniques focus on the interactions that take place surrounding the judgment itself. The technique requires that the variables or judgment attributes be known a priori. The substantive nature of those variables must come from existing strategy theory (Priem & Harrison, 1994), and contingency theory provides an excellent starting point.

In this case, decomposition methods are required to focus on executive choices in response to a series of decision scenarios (i.e., behavioural simulations regarding the environment, firm structure and the strategy making process). The variance in executive choice is evaluated against these factors of interest, which can be manipulated across scenarios, using conjoint or choice analysis techniques.1 Figure 2 shows the way we can manipulate the important choices outlined in structural contingency theory. Paired comparisons (or stated preferences) are collected and then used to evaluate direct and interaction effects. In this way, we reveal the direction and strength of the three factors considered central to contingency theory. For example, an executive faced with a stable environment would lean towards a planned strategy and decentralized structure according to contingency prescriptions. Respondent rankings on each path reflect the perceived utility respondents have for each combination of variables.

This combination of composition and decomposition techniques enables one to test whether the prescriptions of at least one well-known theory (i.e., contingency theory) influences executive judgment. The extent to which these prescriptions are already "obvious" to, or widely known by, practising executives will shed new light on the role of judgment in IT strategy and change.

Figure 2. Judgment evaluation survey



CONCLUSION

We have outlined the importance of executive judgment to the strategic choice process and its particular relevance to the study of e-business, where environmental turbulence increases the relevance of managerial discretion. By separating the outcomes from the actual decision choice, we can begin to more fully understand how these strategic choices influence firm performance. Until now the process of strategic choice has largely been treated as a "black box" where it is assumed that measured outcomes are the result of deliberate choices.

One of the reasons contingency theory has become so popular is that it provides managers with prescriptive advice regarding which configurations lead to higher performance. Further examination of executive judgments to ascertain whether executives are making decisions based on the idea of alignment or fit is required. This will help in our understanding of whether material taught by academics in business schools is actually being used by students in industry. Are executives using the ideas of alignment or fit as prescribed in theory or are they making judgments based on other factors? The answer to this question has important implications for relevance and improving the linkage between theory and practice.

Understanding the processes that occur in strategy development will lead to greater knowledge of the decisions that executives make in uncertain environments and hyper-turbulent contexts. This understanding is important if we want to develop e-business related research that is applicable to practitioners. It is this type of research that will guide executives in the strategic management of change and allow them to gain advantage from leverage their investments in e-business technology.

REFERENCES

A Special Research Forum Call for Papers: Knowledge Transfer between Academics and Practitioners. (1998). *Academy of Management Journal, 41*(6), 746.

Bharadwaj, A. (2000). A resource based perspective on information technology capability and firm performance: An empirical investigation. *MIS Quarterly, 24*(1), 169-196.

Brown, E., & Eisenhardt, K.M. (1998). *Competing on the edge strategy as structured chaos*. Boston: Harvard Business School Press.

Chandler, A. (1962). *Strategy and structure*. Cambridge: M.I.T. Press.

Child, J. (1972). Organisation structure, environment and performance. *Sociology*, *6*, 1-21.

Clemons, E.K. (1986). Information systems for sustainable competitive advantage. *Information & Management, 11*(3), 131-137.

Donaldson, L. (1995). *Contingency theory*. Aldershot, England: Dartmouth Publishing Company

Ford, E.W., Duncan, J.W. et al. (2003). Mitigating risks, visible hands, inevitable disasters, and soft variables: Management research that matters to managers. *Academy of Management Executive, 17*(1), 46.

Galbraith, J.R., & Kazanjian, R.K. (1986). Organizing to implement strategies of diversity and globalization: The role of matrix designs. *Human Resource Management, 25*(1), 37.

Henderson, C., & Venkatraman, N. (1999). Strategic alignment: Leveraging information technology for transforming organizations. *IBM Systems Journal*, *38*(2&3), 472-482.

Hussey, D. (1996). A framework for implementation. *The Implementation Challenge*, New York: John Wiley & Sons

Keen, P., (2002). Getting value from IT. *Sydney University*, 19 August.

Lee, A.S. (1989). A scientific methodology for MIS case studies. *MIS Quarterly 13(1)*, 33-50.

Lee, A.S. (1991). Integrating positivist and interpretative approaches to organisational research. *Organization Science, 2*(4), 342-365.

Mintzberg, H. (1994). *The rise and fall of strategic planning*. Prentice Hall.

Mintzberg, H., Ahlstrand, B., & Lampel, J. (1998). *Strategy safari: A guided tour through the wilds of strategic management*. New York: The Free Press.

Patterson, S. (2001). The truth about CRM, CIO Magazine, May 1st *http://www.cio.com/archive/050101/truth.content.html*

Peterson, J.W. (2002). Leveraging technology foresight to create temporal advantage. *Technological Forecasting and Social Change, 69*, 485-494.

Powell, T.C., & Dent-Micallef, A. (1997). Information technology as competitive advantage: The role of human, business, and technology resources. *Strategic Management Journal*, *18*(5), 375-405.

Priem, R.L. (1994). Executive judgment, organizational congruence, and firm performance. *Organization Science*, *5*(3), 421-437.

Priem, R.L., & Cycyota, C. (2000). On strategic judgement. In M. Hitt, R. Freeman & J. Harrison (Eds.), *Handbook of strategic management*, Blackwell.

Priem, R.L., & Harrison, D.A. (1994). Exploring strategic judgment: Methods for testing the assumptions of prescriptive contingency theories. *Strategic Management Journal, 15*(4), 311-324.

Reinartz, W.J., & Chugh, P. (2002). Learning from experience: Making CRM a success at last. *International Journal of Call Centre Management*, April, 207-219.

Ross, J.W., & Weill, P. (2002). Six decisions your IT people shouldn't make. *Harvard Business Review*, *80*(11), 84.

Sauer, C., & Willcocks, L. (2003). Establishing the business of the future. *European Management Journal, 21*(4), 497-508.

Srinivasan, R., Lilien, G.L, & Rangaswamy, A. (2002). Technological opportunism and radical technology adoption: An application to e-business. *Journal of Marketing 66*(3), 47-61.

Stevens, A. (1997). Deja blue. *Industry Week*, *246*(21), 82-88.

Teece, D. J., Pissano, G., & Sheun, A. (1997). Capabilities and strategic management. *Strategic Management Journal, 18*(7), 509-533.

KEY TERMS

Contingency Theory: A meta-theory, which argues that firm performance is defined by the environment-strategy-structure relationship, where the organization's strategy is contingent on the external environment and the organization structure is contingent on the firm's strategy.

E-Business Technology: Any technology, which enables an organization to conduct business electronically, with the overall aim of improving firm performance.

Executive Judgment: A decision that an executive makes, when they do not have a full understanding of the decision problem, based on their mental models of the environment and their vision for the organization.

External Environment: Factors which are external to an organization, such as new technology or product developments, changing rates of market growth, which an organization must respond to.

Fit/Alignment: Terms used to explain the relationship between IT and strategy. The IT strategy should work in synergy with the organizations strategy. These terms have their roots in the meta-theory contingency theory.

Strategic Choice: The choices that executives make which impact on the strategic direction of the organization. These choices exist as the intended strategies of the organization.

Strategic Decision-Making: The process of making important decisions (usually made by the top management team) to put executive choices into action by implementing strategies.

ENDNOTE

[1] There are a number of conjoint analysis methods, which can be used to test executive judgments. Each of the methods uses a variation of regression to decompose an executive's judgment. The most appropriate for evaluating executive judgments is metric conjoint analysis. For example, Priem (1994) used metric conjoint analysis to examine the judgments of CEO's in manufacturing firms. The outcome of this research was that the executives in manufacturing firms often make contingent judgments, regarding key strategy variables.

Expanding Data Mining Power with System Dynamics

Edilberto Casado
Gerens Escuela de Gestión y Economía, Peru

INTRODUCTION

Data Mining

Business intelligence (BI) is a key topic in business today, since it is focused on strategic decision making and on the search of value from business activities through empowering a "forward-thinking" view of the world. From this perspective, one of the most valuable concepts within BI is the "knowledge discovery in databases" or "data mining," defined as "the process of discovering meaningful new correlations, patterns, and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques" (SPSS, 1997).

The usage of data mining as we currently know dates back to 1995 (Pyle, 2003). Since then, many applications were developed, and now it is a critical discipline to gain business insight. Table 1 shows a list of current data mining applications.

Table 1. A short list of data mining applications

- Credit scoring
- Customer behavior prediction
- Customer segmentation
- Database marketing
- Drug discovery
- Fraud prevention
- Non-compliant taxpayer detection
- Targeted marketing
- Web mining (analysis on Internet-generated data from a Web site)

Table 2. A short list of fields where system dynamics is applied

- Environmental sciences
- Social sciences
- Economics
- Software development
- Project management
- Enterprise management
- Supply chain management
- Health care
- Customer behavior

System Dynamics

System dynamics was created by Jay W. Forrester, Germeshausen Professor Emeritus of Massachusetts Institute of Technology (MIT), in 1956. It is defined by the System Dynamics Society (www.systemdynamics.org) as "a methodology for studying and managing complex feedback systems, such as one finds in business and other social systems."

System dynamics evolved from prior work in feedback control systems, and progressively its application was extended to fields other than engineering. Though its primary application is focused in the understanding of complex systems, it is also used as a predictive tool (Sterman 2000; An, Uhm, Kim & Kwak, 2002; Forrest, 1998). Table 2 shows a list of some fields where system dynamics has found applications.

LIMITATIONS OF DATA MINING FOR PREDICTIVE APPLICATIONS

Regarding data mining techniques and algorithms, Table 3 summarizes the most commonly accepted classification (Berson & Smith, 1997; Thearling, 2003; The Pilot Software's Data Intelligence Group, 1995).

As predictive tools, the data mining techniques listed in Table 3 have the following shortcomings:

1) *The statistical foundation of data mining:* Most of the current BI methods and tools, such as rule induction, decision trees, neural networks, and so forth, are extensively used to develop predictive models; and their conceptual foundation are a combination of mathematical, statistical, and artificial intelligence techniques. It is here where we find a source of limitation for a wider set of real-world applications, since statistics works with historical data and there is no full guarantee about predictions based on such data. A change in the characteristics

Table 3. Main categories of data mining techniques and algorithms

Decision Trees
Building of tree-shaped structures that represent sets of decisions.

Rule Induction
Extraction of useful if-then rules from data, based on statistical significance.

Nearest Neighbor
Classification of each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset (where $k > 1$).

Neural Networks
Statistical analysis tools that, through a "learning" process, build a model of the dependencies between the descriptive variables and the behavior to be explained.

Genetic Algorithms
Building of models using processes such as genetic combination, mutation, and natural selection in a design based on the concepts of evolution.

of a market from a developing to a mature phase, for example, is enough to invalidate the use of statistical methods to forecast customer behavior (An et al., 2002).

2) *Explanation of results:* Another limitation of some data mining methods like neural networks, genetic algorithms, and so forth is their inability to provide an adequate explanation of their results, because they are not easily mapped into human terms (Moxon, 1996), or they are seen as black boxes and no explanation of the results is given, which inhibits confidence, acceptance, and application of results (Parallel Computer Centre at The Queen's University of Belfast, 1997).

3) *The thinking paradigm under the application of data mining methods:* As consequence of their statistical foundation, data mining methods provide information about trends and patterns in the context of a "straight-line thinking" paradigm (i.e., an outcome is expressed as a function of one or more independent variables). However, real world is a feedback system made of interacting elements in a closed-loop context, where an action leads to a result that affects current conditions, and the changed conditions become the basis for future action (Forrester, 1991). A growth strategy executed only relying in trends provided by data mining methods, for example, can lead to failures in achieving the desired goals, because the feedback effects in the business system were ignored (Avila, Mass & Turchan, 1995).

TOOLS OF SYSTEM DYNAMICS: BASIC CONCEPTS

There are two fundamental tools in system dynamics to represent the structure and behavior of a system: causal loop diagrams and stock-and-flow diagrams.

Causal Loop Diagrams

A causal loop diagram is a tool to represent the feedback structure of a system. It is a closed representation of the links between causes and effects involved in it, and also contains an identification of their most important feedback loops (Kirkwood, 1998; Sterman, 2000). Figure 1 shows an example of a causal loop diagram.

Stock-and-Flow Diagrams

A stock-and-flow diagram is another way to represent the feedback structure of a system. But, unlike a causal loop diagram, it can include more precise and detailed information about the nature of the system it represents. Figure 2 shows the following basic elements in a stock-and-flow diagram (Kirkwood, 1998; Sterman, 2000). While causal loop diagrams are more oriented to a qualitative description of a system, stock-and-flow diagrams allow relating mathematical functions to their elements and, therefore, they represent a quantitative description.

Figure 1. Causal loop diagram and its elements

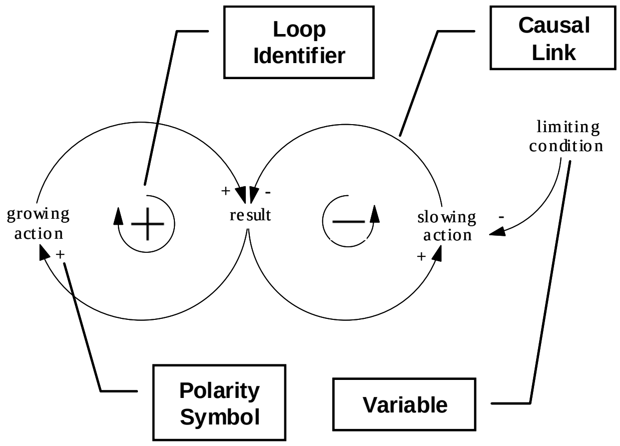
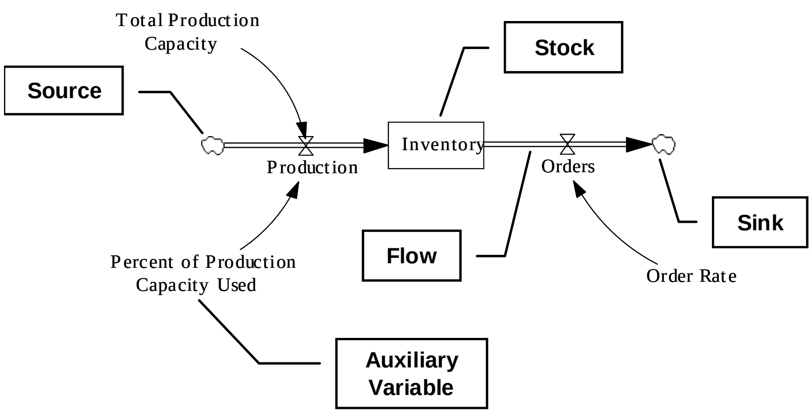


Figure 2. Stock-and-flow diagram and its elements

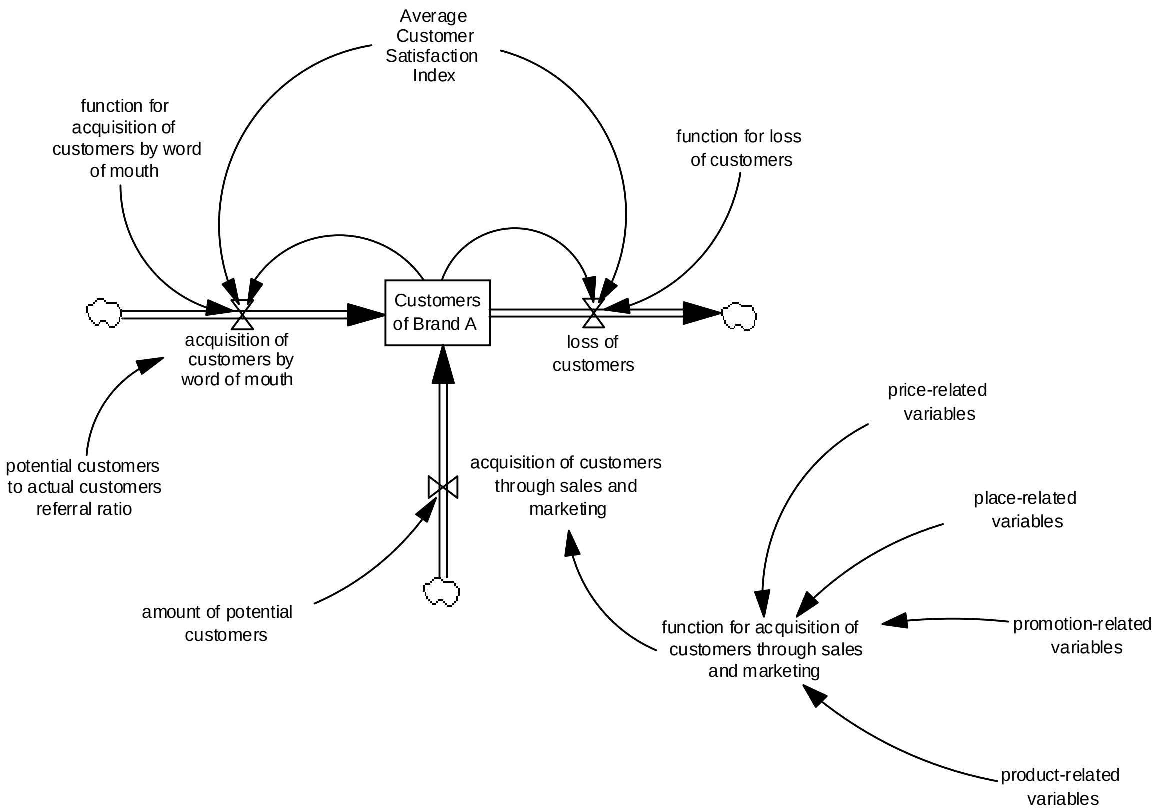


LIMITATIONS OF SYSTEM DYNAMICS

The limitations of system dynamics come from its primary application as modeling tool, and this fact is reflected in the expression "any model is only as good as its assumptions." The following shortcomings of simulation models are cited by Sterman (1991):

- *Accuracy of the decision rules:* Every system dynamics model must be constructed beginning with two processes: the problem articulation or purpose, which must be oriented to solve a specific problem; and the proposal of a dynamic hypothesis, which is a supposition about the behavior of the system to model (Sterman, 2000). The main problem resides in the decision rules under the dynamic hypothesis, which cannot be determined directly from statistical data. Decision rules must be discovered from observation, and this task is limited by the perception of the modeler about the system to study.
- *Soft variables:* Reputation, customer satisfaction, employee motivation, and other soft or intangible variables are fundamental to understand complex systems. Unfortunately, in most cases there is not measures of these data, and under this circumstance these variables are usually discarded from the model. At present, the development of fields like Intellectual Capital and Knowledge Management are contributing to a better estimation of soft variables; in the meantime, the best approach is to include at least reasonable estimates of such variables instead of ignoring them.

Figure 3. Simplified system dynamics model of a CRM application



- *Model boundary:* The model boundary is an important factor that affects its validity. This concept involves which variables will be treated as exogenous (outside the boundary) or endogenous (inside the boundary), and what feedback will be incorporated. As with decision rules, the definition of the boundary will be limited by the perception of the modeler about the system.

DATA MINING AND SYSTEM DYNAMICS TOOLS WORKING TOGETHER

With the current availability of computer power and software tools, there is a great opportunity to take the best of data mining and system dynamics to improve decision-making processes. This approach involves the creation of a system dynamic model to represent the target system. Later, those relationships without a mathematical expression available (because there is not any theory or formula available for them) will be represented by rules or equations obtained from data mining processes. The advantage of this approach is that, once validated, the system dynamics model is a powerful tool for decision making, since it reflects the nature of the target system and provides better insights about its behavior, without the limitations of data mining tools described earlier. In contrast, this approach is affected by the limitations of system dynamics tools, and the direct inclusion of rules or equations into the model could hide important causal relationships that need to be discovered.

AN APPLICATION EXAMPLE

Figure 3 shows a simplified system dynamics model of a Customer Relationship Management (CRM) application related to a brand A, where the processes of customer acquisition and retention are represented. Acquisition is made through sales and marketing or through the phenomena of "word of mouth," and retention is related to a loss of customers' rate.

The average customer satisfaction index for the brand A is calculated through surveys and other data (like frequency of repeated purchases, for example); then, data mining tools provide a set of rules or equations that define mathematical relationships between customer satisfaction and acquisition (or loss) of customers in the form of functions that could be integrated in expressions like the following examples:

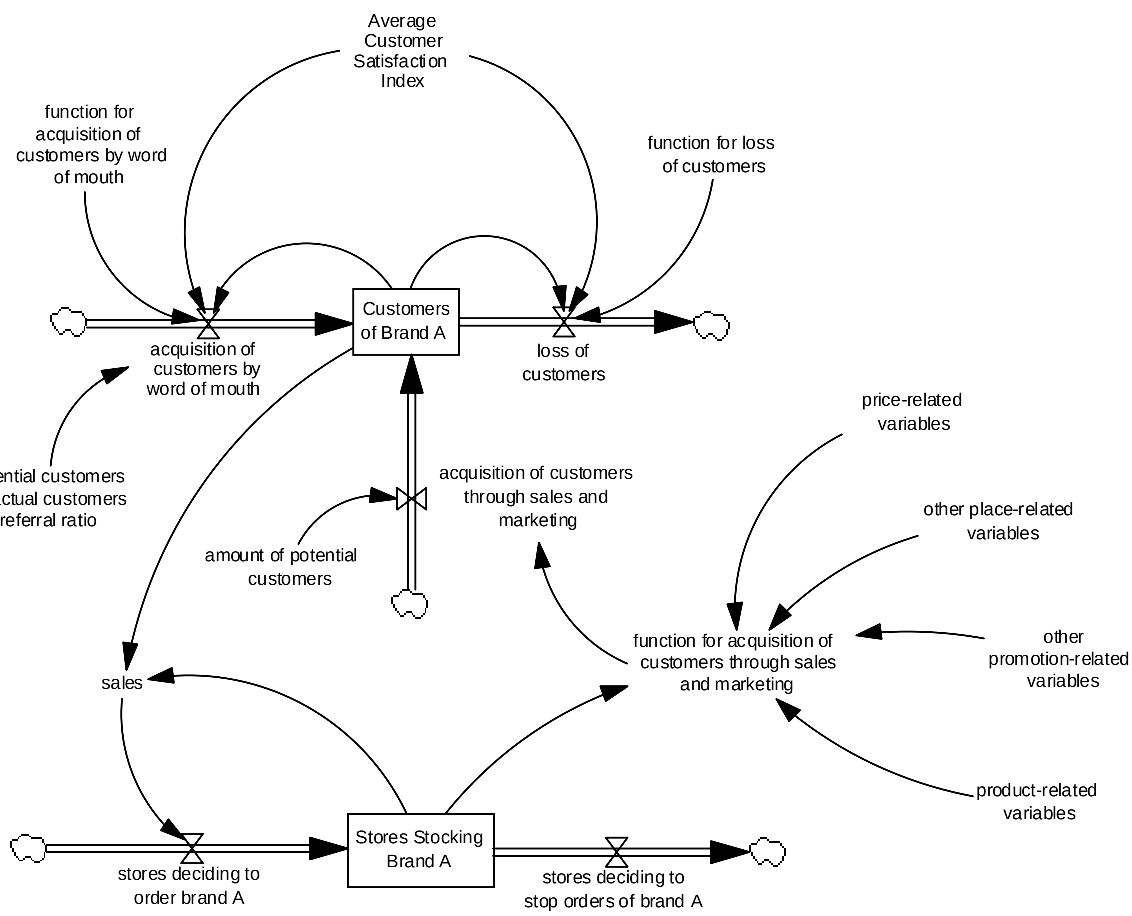
1) acquisition of customers by word of mouth = customers of brand A * potential customers to actual customers referral ratio * function for acquisition of customers by word of mouth (average customer satisfaction index)
2) loss of customers = customers of brand A * function for loss of customers (average customer satisfaction index)

In the equations (1) and (2), "customers of brand A" is the amount of actual customers of brand A, and "potential customers to actual customers referral ratio" is the estimated proportion of potential customers that could be referred by actual customers. The "function for acquisition of customers by word of mouth" and the "function for loss of customers" are derived from data mining processes.

In a similar fashion, another function for the acquisition of customers through sales and marketing is obtained incorporating variables related to the well-known components of the "marketing mix": price, product, promotion, and place (Kotler, 1999, p. 31). As a better understanding of the cause-and-effect relationships of the CRM process is achieved, such function can be reformulated as a new set of stock-and-flow diagrams which include, if needed, a new set of functions obtained via data mining methods. To clarify these concepts, a simplified example is described below.

We suppose that the "function for acquisition of customers through sales and marketing" involves, in addition to other marketing-mix related variables, the amount of money spent in advertising (X) and the amount of stores having the brand A in stock (Y). Through a

Figure 4. Redesign of the model in Figure 3 to incorporate concepts of the dynamic brand value management framework



neural network algorithm, a function of the form $Z = f(X, Y)$ is obtained, and thus we know a mathematical relationship between such variables. However, this expression does not explain conceptually the real effect and impact of X and Y on Z.

To overcome this issue, we could incorporate the principles of an approach known as "dynamic brand value management" or DBVM (Desmet et al., 1998), which takes into account the factors that influence brand value and also dynamic elements such as time delays and feedback loops. Figure 4 shows the redesigned model for the CRM application, where a "virtuous reinforcing loop" is visible: when a product sells well (see variable "sales"), more retailers are motivated to order it (see variable "stores deciding to order brand A"), and the product sells even better.

As more valid cause-and-effect relationships are added, the model will provide more insights for the CRM processes, and it will be possible to apply the intrinsic knowledge contained here to formulate new strategies when a change in the underlying conditions occurs, and when there is not enough available data to mine.

FUTURE TRENDS

The cooperation between data mining and system dynamics tools will progress with a framework that facilitates the model creation process and the integration of their respective elements. A starting point could be the concept of "complex analytic layer" (Oguz 2002), where system dynamics tools would be integrated.

Another way to improve the cooperation between data mining and system dynamics tools is the definition of a common methodology to create and evaluate models that incorporate elements from both approaches. The "Catalyst" methodology (Pyle, 2003), which explicitly incorporates system dynamics with data mining techniques, is an example of this concept.

CONCLUSION

System dynamics is a powerful complement for business intelligence in general. With a clear vision of the features, advantages, and limitations of these tools, we can expect important improvements in decision making and consequently in organizational performance.

The application of system dynamics implies the need to adopt the "systems thinking" paradigm as a new way to see business dynamics. A gradual process is suggested for this purpose, combining skills development exercises with the support of modeling software, if this is available (Casado, 2002). Another concern is the need to trace continuously the developments in knowledge management and intellectual capital, because it is necessary to improve the quantification and representation of intangible variables to be included in system dynamics models.

In the near future, the use of BI tools will be so wide that organizations will not find significant advantages with respect to their competitors. At this point, only a clear understanding of the behavior of systems, far away from discovery of trends and patterns only, will bring back competitive advantage in the business world.

REFERENCES

An, S., Uhm, K., Kim, K. & Kwak, S. (2002, July 28-August 1). System dynamics model for forecasting demand of each automobile type in the Korean automobile market. *Proceedings of the 20th International Conference of the System Dynamics Society*, Palermo, Italy.

Avila, J., Mass, N. & Turchan, M. (1998). Is your growth strategy your worst enemy? *The McKinsey Quarterly*, (2nd Quarter).

Berson, A. & Smith, S. (1997). *Data warehousing, data mining & OLAP*. New York: McGraw-Hill.

Casado, E. (2002). Thinking outside the lines. *Intelligent Enterprise*, (June 28). Retrieved January 3, 2003, from *www.iemagazine.com/020628/511feat3_1.shtml*.

Desmet, D., Finskud, L., Glucksman, M., Marshall, N., Reyner, M. & Warren, K. (1998). The end of voodoo brand management? *The McKinsey Quarterly*, (2nd Quarter).

Forrest, J. (1998, July 20-23). System dynamics, alternative futures, and scenario planning. *Proceedings of the 16th International Conference of the System Dynamics Society*, Quebec, Canada.

Forrester, J. (1991, April 29). *System dynamics and the lessons of 35 years*. Retrieved January 3, 2003, from *sysdyn.mit.edu/sdep/papers/D-4224-4.pdf*.

Kirwood, C. (1998). *System dynamics methods: A quick introduction*. College of Business, Arizona State University, USA.

Kotler, P. (1999). *Kotler on marketing: How to create, win, and dominate markets*. New York: The Free Press.

Moxon, B. (1996). Defining data mining. *DBMS Online*. Retrieved December 28, 1997, from *www.dbmsmag.com/9608d53.html*.

Oguz, M. (2002). Strategic intelligence: Complex analytics part I—The next step in business intelligence. *DM Review*. April. Retrieved January 3, 2003, from *www.dmreview.com/master.cfm?NavID=198&EdID=5069*.

Parallel Computer Centre at The Queen's University of Belfast. (1997). *Data mining techniques*. Retrieved December 12, 1997, from *www-pcc.qub.ac.uk/tec/courses/datamining/stu_notes/dm_book_4.html*.

Pilot Software's Data Intelligence Group. (1995). *An overview of data mining at Dun & Bradstreet*. Retrieved April 28, 2003, from *www.thearling.com/text/wp9501/wp9501.htm*.

Pyle, D. (2003). *Business modeling and data mining*. San Francisco: Morgan Kaufmann.

SPSS, Inc. (1997). *What is data mining?* Retrieved December 28, 1997, from *www.spss.com/datamine/define.htm*.

Sterman, J. (1991). *A skeptic's guide to computer models*. Retrieved January 3, 2003, from *sysdyn.mit.edu/sdep/Roadmaps/RM9/D-4101-1.pdf*.

Sterman, J. (2000). *Business dynamics: Systems thinking and modeling for a complex world*. Boston: Irwin McGraw-Hill.

Thearling, K. (2003). *An introduction to data mining: Discovering hidden value in your data warehouse*. Retrieved April 28, 2003, from *www.thearling.com/text/dmwhite/dmwhite.htm*.

KEY TERMS

Causal Link: An arrow that denotes the cause-effect relationships between two variables in a system dynamics model.

Flow: A representation of the rate of change in the condition of a variable in a system dynamics model.

Intellectual Capital: The set of intangible assets that enable a company to function.

Knowledge Management: The set of processes to manage the creation, dissemination, and utilization of knowledge.

Loop Identifier: A symbol that denotes the nature of each important loop in a system dynamics model. It can be "reinforcing" (positive loop) or "balancing" (negative loop).

Polarity: A positive or negative sign that denotes the sense of the effect between variables connected by a causal link.

Stock: A representation of the state of a variable at a given time in a system dynamics model.

Sink: A representation of a stock into which flows are leaving the boundary of the system.

Source: A representation of a stock from which flows are originated outside the boundary of the system.

Experiential Perspective on Knowledge Management

V.P. Kochikar
Infosys Technologies Ltd., India

J.K. Suresh
Infosys Technologies Ltd., India

INTRODUCTION

The last few decades have seen a growing proportion of organizational wealth being represented by intangible assets, i.e., assets with value that cannot be measured in terms of any physical attribute. Management thinking, conditioned over centuries to extract the greatest value out of physical assets, has had to bring within its ambit the leveraging of these intangible assets in building the capabilities required to deliver superior products and solutions. The discipline of knowledge management (KM) was born and came to encompass the gamut of organizational processes, responsibilities, and systems directed toward the assimilation, dissemination, harvest, and reuse of knowledge. In simpler terms, KM is the answer to the question, "How can the organization update and use its knowledge more effectively?" (Kochikar, 2000).

Some of the world's most successful organizations, be they corporate, academic, or government, invest considerably in KM. McKinsey & Co. spends at least 10% of revenues on managing knowledge. The World Bank's annual KM budget is $50 million. IBM has one of the oldest formal KM initiatives, dating back to 1994.

Substantial benefits have been reported across industries. Johnson & Johnson has implemented KM for speeding up the FDA application process and reported savings of $30 million on one product alone (Berkman, 2001). British Petroleum has estimated savings of $400 million a year, while Chevron has discovered operational cost savings of $2.5 billion over 8 years (Infosys, 2002). Tufts University's school of medicine has used KM to integrate its curricula and has been hailed as a national model for medical education (Genusa, 2001). KM is, however, not an unmixed blessing, as Storey and Barnett (2000) noted. Each organization must fashion a KM strategy that takes cognizance of its unique competencies, aspirations, and business context.

Infosys Technologies (*NASDAQ: INFY*) has conceived, developed, and deployed internally an elaborate architecture for KM that aims to empower every employee with the knowledge of every other employee. The company's success on the knowledge-sharing front has been affirmed by the fact that the company has been a Global MAKE (most admired knowledge enterprises) winner in 2003 (Chase, 2003) and Asia MAKE winner for 2002, 2003. Key elements of the KM architecture include the *Knowledge Currency Unit* scheme, a comprehensive mechanism for reward, recognition, and measurement of KM benefits; *KShop*, the corporate knowledge portal built in-house; and the *knowledge hierarchy*, a four-level taxonomy of 1800 subject areas that constitute knowledge in the Infosys context.

Along the KM journey, we also accumulated a sizeable body of thought on what organizations need to do in order to implement KM successfully, and it is the intention of this article to communicate some of that thought.

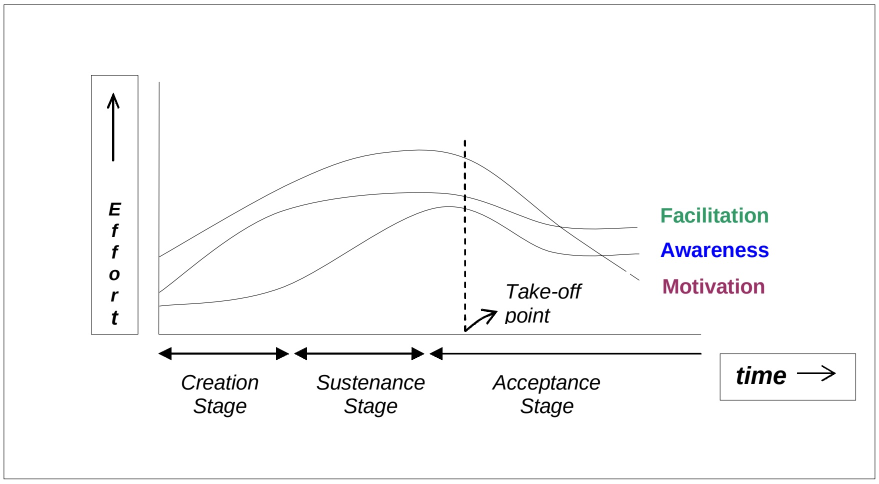
BACKGROUND—THE BUSINESS CASE FOR KM

In achieving its goal, KM needs to percolate into every corner of the organizational mind and create a culture of sharing within the organization. The following definitive statement of Lew Platt, Hewlett-Packard's former CEO, sums up the case for KM: "If HP knew what HP knows, we would be three times as profitable." A strong focus on KM has paid undeniable dividends to leading organizations wordwide. This year's Global MAKE winners have delivered a total return to shareholders of 19.6%, twice the Fortune 500 median of 9.1% (Chase, 2003). Similarly, these leading KM practitioners have shown a return on capital employed of 30.4% versus a Fortune 500 median of 18.5%. These winners also figure prominently in other honor lists, such as *Fortune* magazine's Most Admired Companies list (Hjelt, 2003) and *Business Week* magazine's list of the world's top brands (*Business Week*, 2003).

Chard (1997) and Bartlett (1998) have identified the following drivers for KM: the *pace of change* in a knowledge-driven age, which makes constant learning an imperative; *globalization*, which means acquiring knowledge about new environments and cultural and economic

E

Figure 1. Relative organization-wide effort devoted to motivation, facilitation, and awareness, as a KM initiative evolves over time



Source: Infosys Research

issues; the emergence of *new technologies* that offer new leverage if used well; the increase in *virtual work*, which needs much better knowledge sharing; *rising expectations* from all stakeholders, to meet the companies that need to be proactive and agile; and *growth*, which accentuates the challenge of leveraging the knowledge of individuals for corporate advantage.KM Review magazine's survey of 400 global corporations revealed that the following are key objectives of KM programs (KM Review, 2002):

1. Increasing organizational communication
2. Gaining competitive advantage
3. Increasing collaboration among employees
4. Improving customer relationships
5. Becoming more efficient
6. Innovating
7. Learning from previous mistakes and successes
8. Capturing and retaining tacit knowledge

Using the framework of Nahapiet and Ghoshal (1998), the above objectives can be classified as improving *financial capital* (2, 5); improving *social capital* (1, 3, 4), and improving *intellectual capital* (6, 7, 8). While KM activity, as enumerated below, focuses strongly on the social and intellectual capital aspects, the success of KM must necessarily be measured in terms of improving financial capital. As our extracts from Chase (2003) above demonstrate, successful KM adopters have found that this is the case.

BUILDING AN ORGANIZATIONAL KM ARCHITECTURE: CHALLENGES

An organizational architecture for KM must exhibit the following properties:

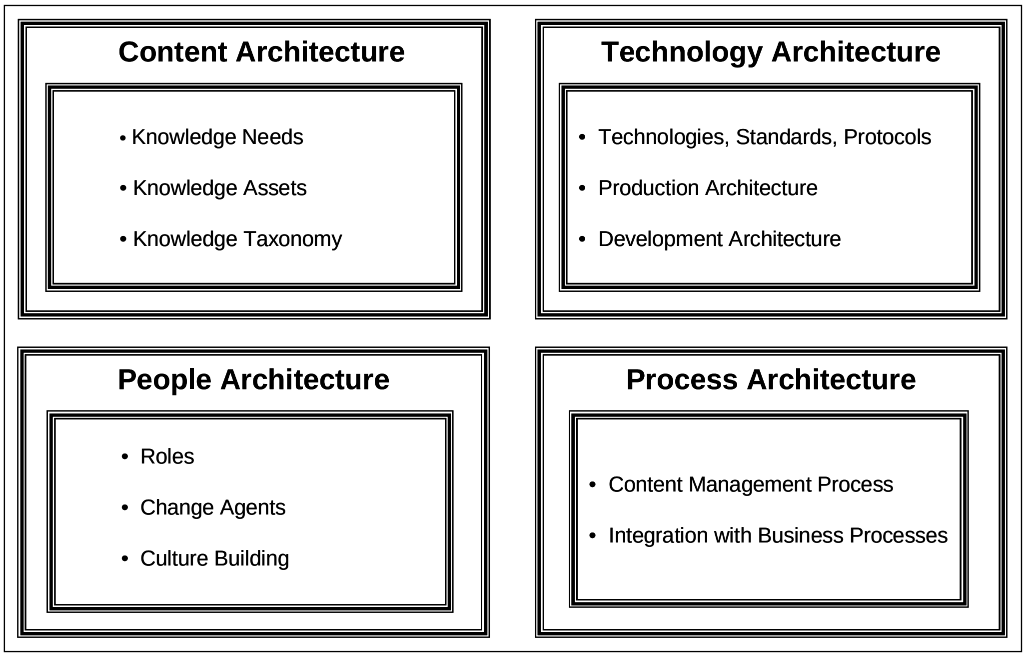
- *Motivation*: It must make people *want* to share knowledge.
- *Facilitation*: It must make it *easier* for them to do so.
- *Awareness*: It must make people *aware* of the KM architecture that has been created, and their roles in using it.

Figure 1 depicts conceptually the relative organization-wide effort required to be devoted to each of the above as the KM initiative evolves over time.

KM success hinges on an architecture that is designed specifically to suit each organization's business and cultural context. Such an architecture must address four key dimensions: *people*, *process*, *content*, and *technology*. Figure 2 outlines the key considerations to be addressed by a KM architecture along these four dimensions.

Enumerated below are a few key challenges on the people, process, content, and technology dimensions that must be addressed when creating a KM architecture.

Figure 2. Sample people, process, content and technology considerations to be addressed by a KM architecture



Deciding How Much Effort Should be Devoted to Technology

As its name implies, information technology (IT) is a mechanism for classifying, storing, and presenting information, as well as ensuring its speedy and effective flow across the organization. Because these capabilities are so critical for KM, it is difficult to conceive of any successful KM implementation that does not lay significant emphasis on the use of IT.

It is, however, necessary to guard against excessive reliance on IT—a tendency that has been liberally fostered by vendors of various hues. Clearly, the use of IT must be judicious, appropriate, preferably ROI-driven, and most importantly, backed by a clear strategy and ownership within the organization (Kochikar, 2000).

A popular yardstick is that not more than a third of the effort and investment of implementing KM should go into technology (Davenport & Prusak, 1998)—any excess makes it highly likely that the "softer" issues related to people, content, and process are being neglected. The extent of IT focus is governed by the target audience's familiarity with, and willingness to use, IT tools.

In the Infosys experience, because *KShop* has been built internally by the central KM group's technology team without the use of any special hardware or software products, the monetary investment in IT for KM has been extremely small. In terms of effort, the investment hovers around the one-third mark.

Table 1 provides a classification typology of technologies for KM and summarizes some of the key technologies.

Deciding the Degree of Centralization of Responsibility

A key question related to the people architecture for KM is, whose responsibility is it to make KM happen? Although every individual and group in the organization must work to make KM successful, there must clearly be an identified group of some form that will "own" KM in the organization. At the same time, this argument must not be taken to the absurd extreme of saying that this central group will have complete responsibility for all aspects of making KM happen. The question thus reduces to, what is the most appropriate position to occupy on the centralization–decentralization spectrum?

The answer to this question is, of course, dependent on issues deeply embedded in the organization's specific context. The extent of senior management's willingness to commit resources for KM is a key determinant. Expectations of KM in the organization also strongly determine the degree of centralization; in general, as more "guarantees" are expected, the more the onus shifts onto the central group.

E

Table 1. Typology of technologies for KM

Collaboration Technologies
Stable Collaboration Technologies:
Electronic mail
Newsgroups/discussion forums/list servers
Chat and instant messaging
Expertise management systems
Audio, video, and desktop conferencing
E-learning technologies
Mobile devices
Emerging Collaboration Technologies:
Whiteboarding and application sharing
Virtual workspaces
Enabling Technologies for Collaboration
XML (extensible markup language)
WAP (wireless application protocol) and WML (wireless markup language)
WebDAV
LDAP
DSL (digital subscriber line)
IP Telephony (VoIP)
Streaming media
Enterprise Technologies
Application servers
Enterprise portals
Business intelligence
Emerging enterprise technologies
Voice portals:
Peer-to-peer (P2P) technology
Presentation Technologies
Portlets
Visualization technologies
Personalization
Dashboards and Web parts
Enabling Technologies for Presentation
Dynamic HTML (hypertext markup language):
Applets
Cookies

Source: Infosys Research

In our experience, a "facilitated decentralized approach" works best: the technology architecture management for KM—development, deployment, and maintenance—is done by a central KM group. All stages of the content management process are anchored by the KM group; however, creation of internal content must happen in the field and is *facilitated* by the KM group. The conception and implementation of the KM strategy is also anchored by the central KM group.

This approach entails considerable investment in training and development of staff in the central KM group and in orienting all other roles in the organization to their knowledge-sharing aspect. One practice that we have found useful has been to "catch them young"—sensitize every fresh employee joining the organization to the mechanics of using the organizational KM infrastructure, within the first 2 weeks. In this manner, they learn the KM ropes at a time when their need to acquire knowledge about the organization is very high.

The "Buy versus Build" Decision

There is an abundance of packaged IT solutions for KM on the market. Some of these have been conceived by their vendors with the KM paradigm in mind, and thus support organizational KM requirements to a fair degree. Others are existing products repackaged with a "KM" label, and thus, they meet only a subset of an organization's requirements.

Organizations interested in reaping the full benefits of KM must, as said earlier, invest in designing a customized KM architecture that fits their context. For the technology dimension then, the "build" route becomes inevitable, as the technology solution must fit well into the people, process, and content dimensions of the KM architecture. Clearly, the economics of the buy versus build decision are also influenced in significant degree by relative costs.

In the Infosys experience, the decision to build was influenced largely by the fact that our proprietary knowledge hierarchy referred to earlier was a central mechanism for accessing content and expertise, and it needed to be supported by the technology system. Similarly, the unique knowledge currency unit (KCU) mechanism needed to be supported. The system was required to compute and use *composite KCU ratings* (a quality measure for documents) and provide interfaces for users of documents to award KCUs and manage their individual KCU accounts. Further, the system needed to exchange data with a number of existing corporate databases that store employee and project data in disparate schemas. A key requirement, the ability to control access based on organizational role, needed to be supported. Another strong need was for the technology architecture to evolve in response to changes in the other KM architectural elements, and an extraordinary degree of flexibility was a must.

Deciding the Degree of Centralization of Content

From a practical angle, the degree of distribution of content is another challenge to be surmounted. Does all content need to be centralized in one repository, or does it live and evolve in different pockets across the enterprise? The former choice appears attractive, because it offers features such as simplicity of architecture, administration, usage, change management, etc. However, the fact that specialized knowledge continues to develop in pockets—and that the responsibility (and pride) associated with ownership is the best driving force for its continued currency and relevance—indicates a distributed content architecture.

Using Synchronous versus Asynchronous Mechanisms for Knowledge Sharing

Another important aspect is the relative focus on direct, people-to-people collaborative sharing vis-à-vis sharing through codification of knowledge. Collaboration is characterized by its reliance on bringing people "in touch" for knowledge transfer through synchronous mechanisms. The latter approach (also called content management) is asynchronous, depends on repositories, and emphasizes defined ways of explicating knowledge. Conventional KM wisdom associates collaboration with the potential for tacit knowledge transfer and content management with explicit knowledge transfer. However, our experience suggests that technology, used innovatively, can help blur the boundaries between the two. For example, collaboration through technology systems is largely synchronous, and yet, as for example in the open source movement, is capable of creating exciting possibilities for knowledge transfer in virtual teams. The knowledge represented by the content can also be viewed as providing the live context for the exchange of implicit knowledge. It may, therefore, be said that the conversion of the implicit knowledge of the organization into content is a property of mature KM programs. Technology systems that sit atop the e-mail system, and content mining agents on collaborative forums, are further examples indicating the blurring distinctions between collaboration and content management.

Measuring the Benefits of KM

As KM implementations mature, it becomes imperative that mechanisms to evaluate the benefits must emerge. A complete treatment of KM metrics is beyond the scope of this article, and the interested reader may refer to Skyrme (2003).

In our experience, there are three forms of evidence that can be used in assessing the benefits of a KM program, in increasing order of reliability and difficulty: anecdotal, survey-based, and metrics-based (Kochikar & Suresh, 2003). Anecdotal evidence—"*Thanks to using good KM, my project saved 12 weeks and saved 15% of cost to the customer*"—although it may represent the viewpoint of a few opinion leaders, can be used to great effect. Surveys are generally more convincing, as they represent the view of a sample carefully chosen to represent the views of the entire target population. For example, an internal survey at Infosys found that more than 99% of the respondents believed KM is essential for the company, and 70% said good knowledge-sharing practices had helped in delivering tangible benefit to customers.

The metrics-based approach is the most powerful, and yet the most difficult to implement, demanding a high degree of maturity of the KM implementation. Our approach has been to measure in quantitative terms the impact of knowledge sharing on traditional indicators of project performance, using the KCU mechanism. Several projects have been able to quantify savings in dollar terms, and also report various other benefits, such as improved quality, faster turnaround times, etc.

FUTURE TRENDS

A heartening trend observed over the past few years is the increasing geographical uptake of KM. Of the 1998 MAKE winners, 73.3% were from North America, 17.7% from Europe, and 4.4% from Asia. In the 2003 MAKE study, North America's representation had dropped to 55%, Europe's had risen to 22.4%, while Asia's had surged to 20.4%. Clearly, Europe's and Asia's acceptance of KM has grown at the expense of North America, and this is likely to be a continuing trend.

Over the coming years, KM will continue to be imperative for organizational effectiveness. There is likely to be greater focus on tacit knowledge sharing, evolution of standards for KM solutions, and sharing through communities of practice. These developments will continue to be supported by evolving technologies: the integration of content management and collaboration products, development of agent technologies to provide proactive services, increased role for multimedia content, automated knowledge capture in collaboration products, automatic content classification, and enterprise-level support for mobile application access (reducing the reliance on speech and voice recognition systems).

CONCLUSION

Creating a culture of sharing is governed by principles that have much in common with Metcalfe's law—as more people grow convinced of the benefits of participating in the knowledge-sharing movement, it becomes easier to convince still more people to buy in. Thus, as long as steady progress is made on the road toward achieving greater sharing, the pace of adoption accelerates with time. Once a "critical mass" of users has been reached, the movement reaches a *take-off stage* (Figure 1), beyond which it becomes self-sustaining, without significant effort being devoted toward building up the motivation and awareness properties. Until this point is reached, it is crucial to understand and manage KM adoption as a process of organizational change.

REFERENCES

Bartlett, C. A. (1998). *McKinsey & Company: Managing knowledge and learning*. Case, Harvard Business School.

Berkman, E. (2001). When bad things happen to good ideas. Darwin Magazine, April. Received from *www.darwinmag.com/read/040101/badthings_ content.html*

Bontis, N., Crossan, M., & Hulland, J. (2002). Managing an organizational learning system by aligning stocks and flows. *Journal of Management Studies, 39*(4), 437–469.

Business Week (BW). (2003, August 4). The top 100 brands. *Business Week International*, pp. 72–76.

Chard, A. -M. (1997). Knowledge management at Ernst & Young. Case, Graduate School of Business, Stanford University.

Chase, R. (2003). Global most admired knowledge enterprises (MAKE) report. *Teleos*. Retrieved from http://knowledgebusiness.com

Concordia University. (2002). Survey on knowledge leaders' critical issues.

Davenport, T. H. (1996). Hewlett-Packard promotes knowledge management initiatives. Retrieved from *www.webcom.com/quantera/HP.html*

Davenport, T. H., & Prusak, L. (1998). *Working knowledge: How organizations manage what they know*. Cambridge, MA: Harvard Business School Press.

Genusa. (2001). Rx for learning. Retrieved from *http://www.cio.com/archive/020101/tufts.html*

Gurteen. (2003). The Gurteen Knowledge website. Retrieved from *http://www.gurteen.com/gurteen/gurteen.nsf/0/17B666B9EE45086B80256 CD500474AF0/*

Hjelt, P. (2003, March 3). The world's most admired companies. *Fortune*, pp. 24–33.

KM Review. (2002). *KM Review* Industry Survey 2002. Melcrum Publishing.

Kochikar, V. P. (2000, September). Learn once, use anywhere. *Knowledge Management Magazine*, *4*(1).

Kochikar, V. P. (2002). Creating the KM infrastructure at Infosys: The technology challenge. *IIMB Management Review, 13*(4), 104–110.

Kochikar, V. P., & Raghavan, S. (2000). Managing knowledge in the e-organization: The IT perspective. In *Proceedings of ICEIS 2000, The International Conference* *on Enterprise Information Systems*, Stafford, UK. July 4–7, 2000.

Kochikar, V. P., & Suresh, J. K. (2003). The Infosys KM experience. In M. Rao (Ed.), *Leading with Knowledge*. New York: McGraw-Hill.

Metcalfe, R. (1996). The Internet after the fad. Retrieved from *http://www.americanhistory.si.edu/csr/comphist/montic/metcalfe.htm*

Nahapiet, J., & Ghoshal, S. (1998, April). Social capital, intellectual capital and the organizational advantage. *Academy of Management Review, 23*(2), 243.

Skyrme, D. J. (2003). Measuring knowledge and intellectual capital. Business Intelligence.

Storey, J., & Barnett, E. (2000). Knowledge management initiatives: Learning from failure. *Journal of Knowledge Management, 4*(2), 145-156.

Sveiby, K. -E. (1997). *The new organizational wealth: Managing and measuring knowledge-based assets*. San Francisco: Berret-Koehler.

KEY TERMS

Intangible Assets: Organizational assets that do not have any physical manifestation, or that have physical measures with no bearing on their value. The following is a typical list of intangible assets that an organization may have:

- Fragmented knowledge residing with individuals, or encapsulated in artifacts such as documentation and software code
- Codified and classified knowledge residing in repositories
- Unique systems, processes, methodologies, and frameworks that the organization follows
- "Formalized" intellectual property, such as patents, trademarks, and brands
- Relationships and alliances that the organization may have shaped

Intellectual Capital (IC): The "stock" of knowledge that exists in an organization at a particular point in time that can be used to generate value for its stakeholders (Bontis, Crossan, & Hulland, 2002).

Knowledge Currency Units (KCU): A mechanism defined at Infosys Technologies to convert all knowledge-sharing activities to a common denominator, in order to enable their measurement in quantitative terms.

Knowledge Management (KM): The gamut of organizational processes, responsibilities, and systems directed toward the assimilation, dissemination, harvest, and reuse of knowledge.

Metcalfe's Law: States that the utility of a network rises in proportion to the square of the number of its users. This means that as more users get connected into a network, the marginal utility perceived by new users increases dramatically.

Social Capital: The resources available through, and derived from, the network of relationships possessed by an individual or social unit within an organization (Nahapiet & Ghoshal, 1998).

Extensions to UML Using Stereotypes

Daniel Riesco
Universidad Nacional de San Luis and Universidad Nacional de Rio Cuarto, Argentina

Marcela Daniele
Universidad Nacional de Rio Cuarto, Argentina

Daniel Romero
Universidad Nacional de Rio Cuarto, Argentina

German Montejano
Universidad Nacional de San Luis, Argentina

INTRODUCTION

The Unified Modeling Language (UML) allows to visualize, to specify, to build and to document the devices of a system that involves a great quantity of software. It provides a standard form for writing the models of a system, covering so much of the conceptual aspects (such as processes of the business and functions of the system) as the concrete ones (such as the classes written in a specific programming language, schemas of databases and software components).

In 1997, UML 1.1 was approved by the OMG becoming the standard notation for the analysis and the design oriented to objects. UML is the first language of modelling in which a metamodel in its own notation has been published. It is a strict subset called Core. It is a self-referential metamodel.

It is a very expressive language that covers all of the necessary views to develop and to deploy systems. UML is a language that provides three extension mechanisms (Booch, Rumbaugh, & Jacobson, 1999): stereotypes, tag values, and constrains. The stereotypes allow to create new types of elements of model based on the elements that form the metamodel UML extending the semantics of the same one, the tag values are an extension of the properties of an element of UML, allowing to add new information to the specification of the same one, and the constrains are an extension of the semantics of UML that allow to add new rules or to modify the existent ones.

The organization of this overview is given in the following way: first, we present the stereotypes according to the standard of OMG; second, we expose the analysis of works that extend UML using stereotypes in diverse real domains; third, we make an analysis of the stereotypes of UML; and we finish giving a general conclusion where we focus ourselves in the distinction of the works according to their inclusion or not of the created stereotypes in the metamodel of UML.

STEREOTYPE ACCORDING TO THE STANDARD OF OMG

A stereotype provides a form of classifying elements in such a way that they work in some aspects as if they were instances of a new constructor of the "virtual"

Attributes	
BaseClass	This specifies the names of one or more elements from UML modeling to which the stereotype is applied, such as classes, associations.
Icon	This is the geometric description of the icon that will be used to present an image of the element of the marked model with the stereotype.

metamodel. A stereotype could also be used to indicate a meaning or different use between two elements with identical structure. A stereotype can also specify a geometric icon to be used to present elements with the stereotype.

USING STEREOTYPES IN DIVERSE REAL DOMAINS

UML adapts to any technique, because it has extension mechanisms that don't need to redefine the nucleus UML, allowing to obtain a modeling more appropriate to the different particular domains. All of the extensions should follow the standard proposed by the OMG (2001).

Modeling of Business with UML

UML was initially designed to describe aspects of a software system. For the modeling of business, UML needed to be extended to identify and to visualize resources, processes, objectives and rules more clearly. These are the primary concepts that define a business system. The Eriksson-Penker Business Extensions (Eriksson & Penker, 1999) provide new stereotypes for their model. In a diagram of class of UML, they represent a process through a specific symbol that corresponds to an activity stereotyped in an activity diagram. The resources used by the process are modeled with a stereotyped dependence *«supply»* and the resources controlled by the process are modeled with a stereotyped dependence *«control»*.

Modeling of Web Applications with UML

In Baresi, Garzotto, and Paoloni (2001), they propose a framework denominated W2000, for the design of Web applications. They combine the use of UML and HDM (Hypermedia Model). The Web applications require the integration of two different, but interrelated, activities: the design hypermedia that is focused in the navigation way and the structure of the information, and the functional design that is focused in the operations. Among their main purposes is the extension to the standard of UML of the dynamic diagrams.

It uses *«node type»* like stereotype of UML to define a node type that allows to reach different structures of information and defines the symbol "@" to indicate that a node stereotyped with *«node type»* will be the node for defect where all the users begin to navigate.

They propose a symbol called "index" that allows the users of a navigation to select one of the elements from a list of indexes.

The "collection links" define how the users can navigate between the core and the members of a collection, and they add a symbol to graphically represent the pattern "Index + Guided Tour".

For the functional design, they define the diagrams of scenarios, and these are represented through an extended sequence of diagrams of UML. The extensions refer as much to objects as to the step of messages. The objects are organized in entities (components and nodes), semantic associations and collect. The "free navigation" is represented with dotted lines and the "constrained navigation" is represented with a line with a diamond in it.

In 1999, Jim Conallen, Principal Consultant of Conallen Inc., Object Oriented Application Development in Conallen Inc., presented in their paper an extension to UML, in a formal way, to model applications Web.

The extension was presented at several other conferences in 1999, including the Rational Users Conference in Seattle (July 1999), and two Wrox Press ASP conferences in Washington, DC (September 1999) and in London (November 1999).

Various summaries and introductions to the extension have or will appear in the *Communications of the ACM* (ASPToday, *http://www.asptoday.com/articles/19990517.htm*), and in the UML Resources Web site at Rational Software. A full explanation of this work is currently being prepared for the book, *Building Web Applications with UML* (Conallen, 2002), published in the Object Technology Series of Addison-Wesley Longman.

This article presents an extension of UML for Web application designs. Part of the extension mechanism of UML is the ability to assign different icons to stereotyped classes. A list of prototype icons for the most common class stereotypes can be found as an appendix. It defines two new stereotypes to model the difference between the executed methods in the server and the executed functions in the client. In a page, a method that executes on the server will be stereotyped as *«server method»* and functions that run on the client *«client method»*. This solves the problem of distinguishing attributes and methods of a page object. It proposes the modeling of a page with two stereotyped classes, *«server page»* and *«client page»*. They define several stereotypes to represent the associations, such as: *«builds»* that is modeled with an unidirectional association from the server page to client page, *«redirects»* to model the redirection to other *«server page»*, *«links»* for defined associations between pages clients and other pages (client or server). Also, they define stereotypes to model components, *«server component»* and *«client component»*, for Forms *«form»*, for Framesets *«frameset»*. Other defined stereotypes, *«scriplet»* for cached client page, and *«xml»* for a hierarchical data object that can be passed to and from a Web server and client browser.

In Gorshkova and Novikov (2001), a UML extension capable to refine the design of the client part of Web application is defined. Several new diagrams are specified which provide a precise definition of the content of Web pages and navigation between them. The composition diagram is a special case of class diagram. We use it to express the structure of the Web pages and identify their content: how they are connected together and what data is carried from one page to another. The main notion of the composition diagram is the page, defined as an autonomous block of screen. Each screen in the navigation diagram is mapped into several pages in the composition diagram. The tool may provide links from pages to screens and vice versa to show their relationship. A page is modeled in the composition diagram as a class stereotyped *«page»*. A page may play the role of container for other pages. Nested pages are modeled as aggregated classes. The page has elements like buttons, links and input fields. They are modeled as attributes of the corresponding page. The *«form»* stereotype is a child of *«page»*. It is used to model HTML forms. The navigable association between source and target pages is stereotyped *«link»*. Each *«link»* has a tag context with expression as value.

In Koch, Baumeister, and Mandel (2000), the authors define a set of stereotypes that are used in the construction of intuitive analysis and design models in the development of Web applications. These models are the navigation space model, the navigation structure model, and the static presentation models.

The basis of the navigation design is the conceptual model, and the outcome is a navigational model, which can be seen as a view over the conceptual model. The navigational model is defined in a two-step process. In the first step, the navigational space model is defined, and in the second, the navigational structure model is built. The navigation space model defines a view on the conceptual model showing which classes of the conceptual model can be visited through navigation in the Web application. The navigational structure model defines the navigation of the application. It is based on the navigation space model, but also additional model elements are included in the class diagram to perform the navigation between navigational objects: menus, indexes, guided tours, queries, external nodes, and navigational contexts.

The static presentational model is represented by UML composition diagrams that describe how the user interfaces are built. They define stereotypes to be able to build these diagrams.

The set of defined stereotypes is the *«Navigational Class»*. It represents a class whose instances are visited during navigation. The *«Direct Navigability»* are associations in the navigation model. These associations are interpreted as direct navigability from the source navigation class to the target navigation class. An *«Index»* is modeled by a composite object which contains an arbitrary number of index items. A *«Guided Tour»* is an object which provides sequential access to the instances of a navigational class. A *«Query»* is represented by an object which has a query string as an attribute. This string may be given, for instance, by an OCL select operation. A *«Menu»* is a composite object which contains a fixed number of menu items. A *«Presentational class»* models the presentation of a navigational class or an access primitive, such as an index, a guided tour, query, or menu. A *«Frameset»* is a top-level element which is modeled by a composite that contains (lower level) presentational objects but may also contain an arbitrary number of nested framesets. An area of the frameset is assigned to each lower level element, so called *«frame»*, the same stereotype is also used in Eriksson and Penker (1999). A *«window»* is the area of the user interface where framesets or presentational objects are displayed, and also defined are *«text»*, *«anchor»*, *«button»*, *«image»*, *«audio»*, *«video»* and *«form»*. A *«collection»* is a list of text elements that is introduced as a stereotype to provide a convenient representation of composites. An *«anchored collection»* a list of anchors.

Real-Time Systems Modeling with UML

In Selic and Rumbaugh (1998), a set of constructs that facilitate the design of software architectures in the domain of real-time software systems is described. The constructs, derived from field-proven concepts originally defined in the ROOM methodology (http://wwweb.org/smo/bmc/mb/mb35.html) are specified using the UML standard. In particular, it showed how these architectural constructs can be derived from more general UML modeling concepts by using the powerful extensibility mechanisms of UML. The following stereotypes are defined as UML extension: *«protocol»*, *«protocolRole»*, *«port»*, *«capsule»*, and *«chainState»*.

A protocol role is modeled in UML by the *«protocolRole»* stereotype of Metamodel Class ClassifierRole. A protocol is modeled in UML by the *«protocol»* stereotype of Metamodel Class Collaboration with a composition relationship to each of its protocol roles representing the standard relationship that a collaboration has with its “owned elements”. A port object is modeled by the *«port»* stereotype, which is a stereotype of the UML Class concept. A capsule is represented by the *«capsule»* stereotype of Class. The capsule is in a composition relationship with its ports, sub-capsules (except for plug-in sub-capsules), and internal connectors. A state whose only purpose is to “chain” further automatic (triggerless) transitions onto an input transition is defined as a stereotype *«chainState»* of the UML State concept.

In Toetenel, Roubtsova, and Katwijk (2001), the paper shows an extension UML with mechanisms for specifying temporal constrains and properties. It defines schemes for the translation of UML specifications into semantically equivalent XTG-based (eXtended Timed Graphs) specifications such as the properties given on UML specification can be proved on the XTG specification. The realization of the approach uses the extensibility interface of the Rational Rose UML Tool (Rational Rose 98i, 2000).

ANALYSIS OF UML STEREOTYPES

In Gogolla and Henderson-Sellers (2002), the paper takes up ideas from Gogolla (2001) where stereotypes have been introduced for relational database design. The expressiveness of UML stereotypes has been analyzed, and some concrete suggestions have been made for the improvement of the UML metamodel. Use OCL to define precise stereotypes.

In Riesco, Martellottto, and Montejano (2003) and Riesco, Grumelli, Maccio, and Martellotto (2002), proposals of "evolutionary stereotypes" are presented. These are incorporated into the modeling tool in such a way that they can extend the UML metamodel, including the new elements with their corresponding semantics. In this way, the environment of a tool can dynamically change its appearance and functionality to allow software engineers to use the stereotypes previously defined in the diagrams.

The Clark, Evans, Kent, Brodsky, and Cook study[1] proposes a new metamodelling facility (MMF) containing: Metamodelling Language (MML); Metamodelling Tools (MMT); a satisfaction checker (for instance, does X satisfy constraint C from model M?); to check that a model satisfies its metamodel; to check that a metamodel satisfies the MML rules; to check that MML satisfies the MML rules.

CONCLUSION

UML is a universal language adopted for the modeling of applications in a wide range of domains. It is an open language, which provides extension mechanisms in order to extend the metamodel. The UML extension mechanisms include: Tag values, Constrains and Stereotypes.

In particular, the stereotypes should be carefully declared and should only be used when the message to be communicated could not be expressed in any other UML terms. In order to achieve a better use of stereotypes, there are certain necessary conditions. The UML metamodel should be adjusted and support tools should be provided for the complete part of the UML metamodel dealing with stereotypes.

The papers analyzed in this work propose a varied number of stereotypes to extend UML, with the purpose of using this language to model particular domains, such as Web applications, real-time systems, business modeling, XML, and so forth.

After the reading and the analysis of these papers, it has been detected that only in three of them (Gogolla, 2001; Riesco, Grumelli, Maccio, & Martellotto, 2002; Selic & Rumbaugh, 1998) the addition of stereotypes is defined as a UML complete extension. This occurs through the incorporation of new elements to the metamodel with new semantic that assures the consistency of the metamodel UML.

In the other analyzed papers, the stereotypes are presented being used in their specific context, but without extending the metamodel UML. This is a clear deficiency in their definitions, since the maintenance of the metamodel UML in a consistent way is very difficult, or even impossible, as their stereotypes are not explicitly added to the metamodel UML. Besides, this means that the stereotypes won't be available to be used in generic solutions to problems belonging to the selected particular domain.

REFERENCES

ASPToday. Available at *http://www.asptoday.com/articles/19990517.htm*

Baresi, L., Garzotto, F., & Paoloni, P. (2001). *Extending UML for modeling Web applications.*

Booch, G., Rumbaugh, J., & Jacobson I. (1999). *The Unified Modeling Language user guide.* Addison-Wesley.

Conallen, J. (1999). *Modeling Web applications with UML.*

Conallen, J. (2002). *Building Web applications with UML* (2nd ed.). Addison-Wesley.

Eriksson, H.-E., & Penker, M. (1999). *Business modeling with UML: Business patterns at work*, Wiley & Sons.

Gogolla, M. (2001). Using OCL for defining precise, domain-specific UML stereotypes. In A. Aurum & R. Jeery (Eds.), *Proc. 6th Australian Workshop on Requirements Engineering (AWRE'2001)* (pp. 51-60). Centre for Advanced Software Engineering Research (CAESER), University of New South Wales, Sydney, 2001.

Gogolla, M., & Henderson-Sellers, B. (2002). Analysis of UML stereotypes within the UML metamodel. In J.-M. Jezequel, H. Hussmann, & S. Cook (Eds.), *Proc. 5th Conf. Unified Modeling Language (UML'2002)*, Springer, Berlin, LNCS.

Gorshkova, E., & Novikov, B. (2001). UML extensibility in the design of Web applications. *Exploiting*.

Koch, N., Baumeister, H., & Mandel, L. (2000). Extending UML to model navigation and presentation in Web applications. In G. Winters & J. Winters (Eds.), *Modeling Web applications*, *Workshop of the UML´2000*. York, England, October.

OMG (2001). OMG Unified Modeling Language specification. Retrieved from the World Wide Web at *http://www.omg.org*

Rational Rose 98i. (2000). Rose Extensibility Reference. Rational Software Corporation, *http://www.rational.comwww.se.fhheilbromm.de/usefulstuff/Rational Rose 98i Documentation*

Rational Software and Miller Freeman, Inc, a United Newa & Media Company. (1999). Business modeling with UML. Retrieved from the World Wide Web at *http://www.therationaledge.com/rosearchitect/mag/archives/fall99/f5.html*

Riesco, D., Grumelli, A., Maccio, A., & Martellotto, P. (2002). Extensions to UML metamodel: Evolutionary stereotypes. *3rd ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel / Distributed Computing*, Madrid, Spain.

Riesco, D., Martellottto, P., & Montejano, G. (2003). Extension to UML using stereotypes. In L. Favre (Ed.), *UML and the Unified Process* (Chapter XIV, p.40), Hershey, PA: IRM Press.

ROOM Methodology. *http://wwweb.org/smo/bmc/mb/mb35.html*

Selic, B., & Rumbaugh, J. (1998). *Using UML for modeling complex real-time systems.*

Toetenel, H., Roubtsova, E., & Katwijk, J. (2001). A timed automata semantics for real-time UML specifications. *Proceedings of the IEEE Symposia on Human-Centric Computing Languages and Environments (HCC'01).*

KEY TERMS

Class Diagram: Show the classes of the system, their interrelationships, and the collaboration between those classes.

Extension Mechanisms: Specify how model elements are customized and extended with new semantics.

Metamodel: An abstraction which defines the structure for a UML model. A model is an instance of a metamodel. Defines a language to describe an information domain.

Object Constraint Language (OCL): A notational language for analysis and design of software systems. It is a subset of the industry standard UML that allows software developers to write constraints and queries over object models.

OMG: Has been "Setting The Standards For Distributed Computing™ through its mission to promote the theory and practice of object technology for the development of distributed computing systems. The goal is to provide a common architectural framework for object-oriented applications based on widely available interface specifications (OMG, 2001).

Real Domains (Particular or Specific): The different application areas that can require to be modeled with UML. For example: Web applications, real-time system, XML, business modeling, frameworks, communication protocols, workflows, geographical information systems, and so forth.

Stereotype: Allows to create new types of elements of modeling, based on the elements that form the goal-pattern UML, extending its semantics.

ENDNOTE

[1] See www.puml.org for the document "A Feasibility Study in Rearchitecting UML as a Family of Languages using a Precise OO Meta- Modeling Approach" (Clark, Evans, Kent, Brodsky, Cook) and associated tools.

Face Expression and Motion Analysis over Monocular Images

Ana C. Andrés del Valle
VICOMTech Research Center, Spain

INTRODUCTION

As computers evolve towards becoming more human-oriented machines, human-computer interfaces, behavior-learning robots, and disable-adapted computer environments will use face expression analysis to be able to react to human action. Furthermore, faces play an essential role in human communication. Consequently, they have been the first objects whose motion has been studied in order to recreate animation on synthesized models or to interpret motion for a posteriori use. The analysis of motion and expression from monocular (single) images is widely investigated because non-stereoscopic static images and videos are the most affordable and extensively used visual media (i.e., Webcams).

Synthetic faces are classified into two major groups: avatars and clones. Generally, avatars are a rough and symbolic representation of the person, and their animation is speaker independent because it follows generic rules disregarding the individual that they personify. Clones are more realistic and their animation takes into account the nature of the person and his real movements. Whether we want to animate avatars or clones, we face a great challenge: the automatic generation of face animation data. Manually generated animation has long been used to create completely virtual characters and has also been applied to animate avatars. Nevertheless, many computer applications require real-time and easy-to-use face animation parameter generation, which means that the first solutions developed using motion-capture equipment prove to be too tedious for many practical purposes. Most applications utilizing Talking Heads aim at telecommunication uses. In such a context, real-time capabilities and low computing cost for both analysis and synthesis are required. Current trends in research tend to use speech analysis or synthesized speech from text as a source of real-time animation data. Although these techniques are strong enough to generate parameters to be used by avatars, they cannot provide realistic data for face animation.

BACKGROUND

Systems analyzing faces from monocular images are designed to give motion information with the most suitable level of detail, depending on their final application. Image input is analyzed in the search for general facial characteristics: global motion, lighting, and so forth. At this point, some image processing is performed to obtain useful data that can be afterwards interpreted to obtain face animation synthesis (see Figure 1).

We have categorized the most performing techniques for facial analysis on monocular images in three groups: “those that retrieve emotion information,” “those that obtain parameters related to the Face Animation synthesis used,” and “those that use explicit face synthesis during the image analysis.”

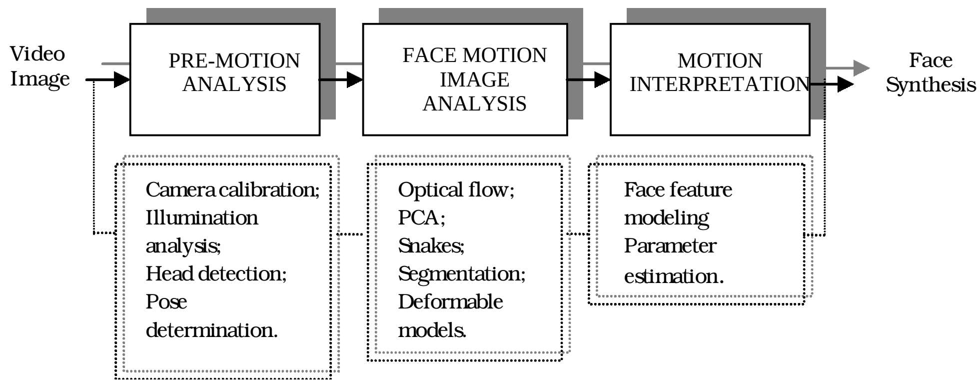
RETRIEVING EMOTION INFORMATION

Humans detect and interpret faces and facial expressions in a scene with little or no effort. The systems we discuss in this section accomplish this task automatically. The main concern of these techniques is to classify the observed facial expressions in terms of generic facial actions or in terms of emotion categories, and not to attempt to understand the face animation that could be involved to synthetically reproduce them.

Summary of the Most Performing Approaches

- The researcher Yacoob (Yacoob & Davis, 1994; Black & Yacoob, 1997) explored the use of local parameterized models of image motion for recognizing the non-rigid and articulated motion of human faces. These models provide a description of the motion in terms of a small number of parameters that are related intuitively to the motion of some facial features under the influence of expressions.
- Huang and Huang (1997) introduce a system developed in two parts: facial feature extraction (for the

Figure 1. Diagram of the general procedure for facial expression analysis



training-learning of expressions) and facial expression recognition. The system applies a point distribution model and a gray-level model to find the facial features. Then, the position variations are described by 10 action parameters (APs).

- Pantic and Rothkrantz (2000) describe another approach, which is the core of the Integrated System for Facial Expression Recognition (ISFER). The system finds the contour of the features with several methods suited to each feature—snakes, binarization, deformable models, and so forth—making it more efficient under uncontrolled conditions: irregular lighting, glasses, facial hair. A neural network (NN) architecture of fuzzy classifiers is designed to analyze complex mouth movements.
- Many systems base their description of face actions on the Facial Action Coding System (FACS) proposed by Ekman and Friesen (1978). The importance granted to FACS is such that two research teams, one at the University of California, San Diego (UCSD) and the Salk Institute, and another at the University of Pittsburgh and Carnegie Mellon University (CMU), were challenged to develop prototype systems for automatic recognition of spontaneous facial expressions.

  The system developed by the UCSD team, described in Bartlett et al. (2001), analyzes face features after having determined the pose of the individual in front of the camera. Features are studied using Gabor filters and afterwards classified using a previously trained hidden Markov model (HMM). In their study, Bartlett et al. claim AU detection accuracy from 80% for eyebrow motion to around 98% for eye blinks.

  CMU opted for another approach, where face features are modeled in multi-state facial components of analysis. They use neural networks to derive the AUs associated with the motion observed. They have developed the facial models for lips, eyes, brows, cheeks, and furrows. In their article, Tian, Kanade, and Cohn (2001) describe this technique, giving details about the models and the double use of NN, one for the upper part of the face and a different one for the lower part. The average recognition rates achieved are around 95.4% for upper face AUs and 95.6% for lower face AUs.
- Piat and Tsapatsoulis (2000) take the challenge of deducing face expression out of images from another perspective, no longer based on FACS. Their technique finds first the action parameters (MPEG-4 FAPs) related to the expression being analyzed, and then they formulate this expression with high-level semantics. To do so, they have related the intensity of the most used expressions to their associated FAPs.
- Other approaches (Chen & Huang, 2000) complement the image analysis with the study of the human voice to extract more emotional information. These studies are oriented to develop the means to create a human-computer interface (HCI) in a completely bimodal way.

Pantic and Rothkrantz (2000) offer overviews and comparative studies of many techniques, including some of those just discussed, analyzed from an HCI perspective.

Table 1. A summary of the most used image processing techniques

PRE-MOTION ANALYSIS	EXPRESSION ANALYSIS	MOTION INTERPRETATION
Camera calibration—Processing techniques that correct the image distortion due to the camera characteristics (lens distortion, projection, etc.). Illumination analysis—Processing techniques that study the illumination of the analyzed scene to understand its influence on the image of the speaker's face. Head detection and pose determination—Image processing algorithms that aim at giving an accurate location and orientation of the speaker's head in the scene.	Optical flow (OF)—The OF between two images is the field of displacements due to local motion. Principal Component Analysis (PCA)—PCA or *Karhunen-Loeve transform* is a statistical tool that permits vector dimension reduction. It is used to analyze images of facial features by contrasting them against a trained database obtained using PCA over images of that feature under different expressions and lighting. Segmentation—Image processing techniques that aim at separating different objects on an image. Snakes—Active contour models, also called snakes, are geometric curves that approximate the contours of an object on an image by minimizing an energy function.	Face feature modeling—Image data processing techniques that relate the results from analysis facial motion into face animation parameters following specific motion models per feature: eye, mouth, eyebrows, etc. Parameter estimation—Image data processing techniques that relate the results from the analysis of facial motion into face animation parameters following a mathematical correspondence. Among others, we find the following estimators: linear, neural networks, and radial basis functions networks.

OBTAINING PARAMETERS RELATED TO THE SYNTHESIS USED

Some face animation systems need action parameters as input to specify how to open the mouth, the position of the eyelids, the orientation of the eyes, and so forth, in terms of parameter magnitudes associated to physical displacements. The analysis methods studied in this section try to measure displacements and feature magnitudes over the images to derive the actions to be performed over the head models.

Summary of the Most Performing Approaches

- Terzopoulos and Waters (1993) developed one of the first solutions of this nature. Their method tracks linear facial features to estimate corresponding parameters of a three-dimensional, wireframe face model, allowing them to reproduce facial expressions. They utilize anatomical-based muscle actions that animate the 3D face wireframe.

- Based on a similar animation system to that of Water's, Essa, Basun, Darrel, and Pentland (1996) define a suitable set of control parameters using vision-based observations. They call their solution FACS+ because it is an extension of the traditional FAC system. They use optical flow analysis along the time of sequences of frontal view faces to get the velocity vectors on 2D, and then these are mapped to the parameters**.**
- Morishima (2001) has been developing a system that succeeds in animating a generic parametric muscle model after having been customized to take the shape and texture of the person the model represents. By means of optical flow image analysis, complemented with speech processing, motion data is generated. These data are translated into motion parameters after passing through a previously trained neural network.
- Sarris and Strintzis (2001, 2002) developed a system for video-phoning for the hearing impaired. The rigid head motion (pose) is obtained by fitting the projection of a 3D wireframe onto the image being analyzed. Then, non-rigid face movements (expressions) are estimated thanks to a feature-based approach adapted from the Kanade, Lucas, and Tomasi algorithm.
- Ahlberg (2002) also exposes in his work a wireframe fitting technique to obtain the rigid head motion. He uses the new parameterized variant of the face model CANDIDE, named CANDIDE-3, which is MPEG-4 compliant. The image analysis techniques include PCA on eigentextures that permits the analysis of more specific features that control the model deformation parameters.
- The MIRALab research team at the University of Geneva (Switzerland) has developed a complete system to animate avatars in a realistic way, in order to use them for telecommunications. Goto, Kshirsagar, and Magnenat-Thalmann (2001) give a more detailed explanation about the image processing involved. Feature motion models for eyes, eyebrows, and the mouth allow them to extract image parameters in the form of 2D point displacements. These displacements represent the change of the feature from the neutral position to the instant of the analysis and are easily converted into FAPs.
- Also aiming at telecom applications, researchers at the Eurecom Institute have developed a system that takes advantage of robust face feature analysis techniques, as well as of the synthesis of the clone of the individual being analyzed (Andrés del Valle, 2003). They use a Kalman filter to recover the head global position and orientation. The pose data predicted by the filter are used to adapt feature analysis algorithms initially designed to work for a frontal point of view under any other head pose. This solution controls face feature analysis during the change of the speaker's pose.

USING EXPLICIT FACIAL SYNTHESIS DURING THE ANALYSIS

Some face motion analysis techniques use the synthesized image of the head model to control or refine the analysis procedure. In general, the systems that use synthesized feedback in their analysis need a very realistic head model of the speaker, high control of the synthesis, and knowledge of the conditions of the face being recorded.

Summary of the Most Performing Approaches

- Li, Roivainen, and Forchheimer (1993) presented one of the first works to use resynthesized feedback. Using a 3D model, CANDIDE, their approach is characterized by a feedback loop, connecting computer vision and computer graphics.
- Ezzat and Poggio (1996a, 1996b) describe image-based modeling techniques that make possible the creation of photo-realistic computer models of real human faces. To generate the motion for this model, they use an analysis-by-synthesis algorithm, which is capable of extracting a set of high-level parameters from an image sequence involving facial movement using embedded image-based models.
- More recent research is able to develop much more realistic results with three-dimensional models. Eisert and Girod (1998), for instance, present a system that estimates 3D motion from image sequences showing head and shoulder scenes for video telephone and teleconferencing applications. They use a very realistic 3D head model of the person in the video. The model constrains the motion and deformation in the face to a set of FAPs defined by the MPEG-4 standard. Combining the 3D information with the optical flow constraint leads to a linear algorithm that estimates the facial animation parameters.
- Pighin, Szeliski, and Salesin (1999) maximize the previous approach by customizing animation and analysis on a person-by-person basis. They use

new techniques to automatically recover the face position and the facial expression from each frame in a video sequence. Their mesh morphing approach is detailed in Pighin, Hecker, Lischinski, Szeliski, and Salesin (1998). Their face motion and expression analysis system fits the 3D model on each frame using a continuous optimization technique. During the fitting process, the parameters are tuned to achieve the most accurate model shape.

FUTURE TRENDS

There exists a trade-off between the degree of motion detail extracted from the images and the level of semantic understanding desired. Very precise analysis techniques that are able to generate information to accurately animate face models often cannot provide meaningful information about the general facial expression. HCI applications seek a high level of comprehension about the speaker's expression; FA applications simply need data to replicate those expressions. Nowadays, applications are complex, and mix FA and HCI requirements; therefore, current research trends try to satisfy both needs: accurate motion analysis and expression understanding. As a result, the different research perspectives of the scientific communities involved in the field of facial animation (CG, CV, IP for Telecommunications, etc.) are starting to converge.

CONCLUSION

Since the importance granted to facial animation is increasing, analysis and synthesis methods developed to generate face animation are under continuous change to meet new application requirements. Following this trend, the analysis of monocular images to extract facial motion data has appeared as a way to simplify and adapt facial animation to current video media. The effort of this research aims at making facial animation technologies and methods available to the general public, and permitting the study of already stored image data.

REFERENCES

Ahlberg, J. (2002). An active model for facial feature tracking. *Eurasip Journal on Applied Signal Processing*, *6*, 566-571.

Andrés del Valle, A.C. (2003). *Facial motion analysis on monocular images for telecom applications: Coupling expression and pose understanding*. Doctoral Thesis, Ecole Nationale Supérieure des Télécommunications, Institut Eurécom, Paris, France. Retrieved October 2004 from *http://pastel.paristech.org/archive/00000578*

Bartlett, M.S. (2001). *Face image analysis by unsupervised learning*. Boston: Kluwer Academic Publishers.

Bartlett, M.S., Braathen, B., LittleWort-Ford, G., Hershey, J., Fasel, I., Marks, T., Smith, E., Sejnowski, T.J. & Movellan, J.R. (2001). *Automatic analysis of spontaneous facial behavior: A final project report*. Technical Report No. 2001.08, University of California, San Diego, MPLab.

Black, M.J. & Yacoob, Y. (1997). Recognizing facial expressions in image sequences using local parameterized models of image motion. *International Journal of Computer Vision*, *25*(1), 23-48.

Chen, L.S. & Huang, T.S. (2000). Emotional expressions in audiovisual human computer interaction. *Proceedings of the International Conference on Multimedia and Expo*.

Eisert, P. & Girod, B. (1998). Analyzing facial expression for virtual conferencing. *Proceedings of the IEEE Computer Graphics & Applications* (pp. 70-78).

Ekman, P. & Friesen, W.V. (1978). *The facial action coding system*. Palo Alto, CA: Consulting Psychologists Press.

Essa, I., Basu, S., Darrel, T. & Pentland, A. (1996). Modeling, tracking and interactive animation of faces and heads using input from video. *Proceedings of Computer Animation*.

Ezzat, T. & Poggio, T. (1996a). Facial analysis and synthesis using image based models. *Proceedings of the 2nd International Conference on Automatic Face and Gesture Recognition*.

Ezzat, T. & Poggio, T. (1996b). Facial analysis and synthesis using image-based models. *Proceedings of the Workshop on the Algorithm Foundations of Robotics*.

Goto, T., Kshirsagar, S. & Magnenat-Thalmann, N. (2001). Automatic face cloning and animation. *IEEE Signal Processing Magazine*, (May), 17-25.

Huang, C.-L. & Huang, Y.-M. (1997). Facial expression recognition using model-based feature extraction and action parameters classification. *Journal of Visual Communication and Image Representation*, *8*(3), 278-290.

Li, H., Roivainen, P. & Forchheimer, R. (1993). 3-D motion estimation in model-based facial image coding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *15*(6), 545-555.

Morishima, S. (2001). Face analysis and synthesis for duplication expression and impression. *IEEE Signal Processing Magazine*, (May), 26-34.

Pantic, M. & Rothkrantz, L.J.M. (2000). Automatic analysis of facial expression: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(12), 1424-1445.

Piat, F. & Tsapatsoulis, N. (2000). Exploring the time course of facial expressions with a fuzzy system. *Proceedings of the International Conference on Multimedia and Expo*.

Pighin, F., Hecker, J., Lischinski, D., Szeliski, R. & Salesin, S. (1998). Synthesizing realistic facial expressions from photographs. *Proceedings of ACM SIGGRAPH 98* (pp. 75-84).

Pighin, F., Szeliski, R. & Salesin, D.H. (1999). Resynthesizing facial animation through 3D model-based tracking. *Proceedings of the International Conference on Computer Vision.*

Sarris, N. & Strintzis, M.G. (2001). Constructing a video phone for the hearing impaired using MPEG-4 tools. *IEEE Multimedia, 8*(3).

Terzopoulos, D. & Waters, K. (1993). Analysis and synthesis of facial image sequences using physical and anatomical models. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 15*(6).

Tian, Y., Kanade, T. & Cohn, J.F. (2001). Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23*(2), 97-115.

Yacoob, Y. & Davis, L. (1994). Computing spatio-temporal representations of human faces. *Proceedings of the Computer Vision and Pattern Recognition Conference* (pp. 70-75).

KEY TERMS

Computer Graphics (CG): Field of science and technology concerned with methods and techniques for converting data to or from visual presentation using computers.

Computer Vision (CV): A branch of artificial intelligence and image processing concerned with computer processing of images from the real world. Computer vision typically requires a combination of low-level image processing to enhance the image quality (e.g., remove noise, increase contrast), and higher-level pattern recognition and image understanding to recognize features present in the image.

Facial Action Coding System (FACS): A comprehensive system that can distinguish the most possible visually distinguishable facial movements. FACS derives from an analysis of the anatomical basis of facial movement. Using FACS, it is possible to analyze any facial movement into anatomically based minimal action units (AUs).

Facial Animation (FA): The set of CG and CV techniques used to recreate facial motion and expression. We can classify animated characters in two groups: avatars and clones. Avatars are a general, standard representation of a human; a clone is the exact replication of an existing individual.

Facial Animation Parameters (FAP): The standard MPEG-4 defines 68 FAPs. Each one is associated with the movement of specific vertices that compose a head mesh so to generate different a facial expression.

Human-Computer Interaction (HCI): A discipline concerned with the design, evaluation, and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them.

Image Processing (IP): Processing technique algorithms designed to enhance and manipulate an image.

MPEG-4: The latest compression standard developed by the Motion Picture Expert Group. MPEG-4 brings higher levels of interaction with content, controlled by the content developers. It also brings multimedia to new types of networks, including those employing relatively low bit-rates, and mobile ones.

Facial and Body Feature Extraction for Emotionally-Rich HCI

Kostas Karpouzis
National Technical University of Athens, Greece

Athanasios Drosopoulos
National Technical University of Athens, Greece

Spiros Ioannou
National Technical University of Athens, Greece

Amaryllis Raouzaiou
National Technical University of Athens, Greece

Nicolas Tsapatsoulis
National Technical University of Athens, Greece

Stefanos Kollias
National Technical University of Athens, Greece

INTRODUCTION

Emotionally-aware Man-Machine Interaction (MMI) systems are presently at the forefront of interest of the computer vision and artificial intelligence communities, since they give the opportunity to less technology-aware people to use computers more efficiently, overcoming fears and preconceptions. Most emotion-related facial and body gestures are considered to be universal, in the sense that they are recognized along different cultures; therefore, the introduction of an "emotional dictionary" that includes descriptions and perceived meanings of facial expressions and body gestures, so as to help infer the likely emotional state of a specific user, can enhance the affective nature of MMI applications (Picard, 2000).

As a general rule, our intuition of what a human expression represents is based on trying to mimic the way the human mind works, while making an effort to recognize such an emotion. This means that even though image or video input is necessary for this task, this process cannot come to robust results without taking into account features like hand gestures or body pose. These features are able to convey messages in a much more expressive and definite manner than mere wording, which can be misleading or ambiguous. Sometimes, a simple hand action, such as placing one's hands over the ears, can pass on the message that you've had enough of what you are hearing more expressively than any spoken phrase.

BACKGROUND

Emotion Representation

Most emotion analysis applications attempt to annotate video information with category labels that relate to emotional states. However, since humans use an overwhelming number of labels to describe emotion, we need to incorporate a higher level and continuous representation that is closer to our conception of how emotions are expressed and perceived.

Activation-emotion space (Cowie, 2001) is a simple representation that is capable of capturing a wide range of significant issues in emotion. It rests on a simplified treatment of two key themes:

- Valence: The clearest common element of emotional states is that the person is influenced by feelings that are "valenced" (i.e., they are centrally concerned with positive or negative evaluations of people, things, or events).
- Activation Level: Research has recognized that emotional states involve dispositions to act in certain ways. Thus, states can be rated in terms of the associated activation level (i.e., the strength of the person's disposition to take some action rather than none).

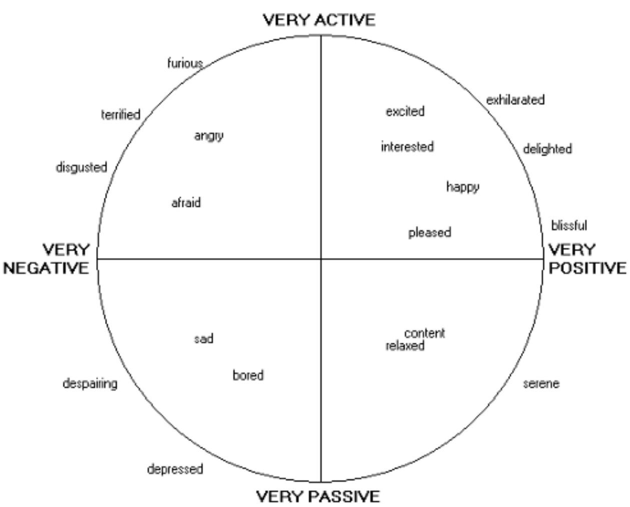
The axes of the activation-evaluation space reflect those themes, with the vertical axis showing activation level, while the horizontal axis represents evaluation. This scheme of describing emotional states is more tractable than using words and still can be translated into and out of verbal descriptions. Translation is possible because emotion-related words can be thought of as positions in activation-emotion space.

A surprising amount of emotional discourse can be captured in terms of activation-emotion space. Perceived full-blown emotions are not evenly distributed in activation-emotion space; instead, they tend to form a roughly circular pattern. In this framework, the center can be thought of as a natural origin, thus making emotional strength at a given point in activation-evaluation space proportional to the distance from the origin. The concept of a full-blown emotion can then be translated roughly as a state where emotional strength has passed a certain limit. An interesting implication is that strong emotions are more sharply distinct from each other than weaker emotions with the same emotional orientation. A related extension is to think of primary or basic emotions as cardinal points on the periphery of an emotion circle. Plutchik (1980) has offered a useful formulation of that idea—the emotion wheel—(see Figure 1).

Facial Expression Analysis

There is a long history of interest in the problem of recognizing emotion from facial expressions (Ekman, 1978), and there have been extensive studies on face perception during the last 20 years (Davis, 1975; Ekman, 1973; Scherer, 1984). The salient issues in emotion recognition from faces are parallel in some respects to the issues associated with voices, but divergent in others. In most cases, these studies attempt to define the facial expression of emotion

Figure 1: The activation-emotion space



in terms of qualitative patterns capable of being displayed in a still image. This usually captures the apex of the expression (i.e., the instant at which the indicators of emotion are most noticeable). More recently, emphasis has switched towards descriptions that emphasize gestures (i.e., significant movements of facial features).

In the context of faces, the task almost always has been to classify examples of the six emotions considered to be universal: joy, sadness, anger, fear, disgust, and surprise (Ekman, 1978). More recently, morphing techniques have been used to probe states that are intermediate between archetypal expressions. They reveal effects that are consistent with a degree of categorical structure in the domain of facial expression, but they are not particularly large, and there may be alternative ways of explaining them—notably by considering how category terms and facial parameters map onto activation-evaluation space (Karpouzis, 2000).

Analysis of the emotional expression of a human face requires a number of pre-processing steps that attempt to detect or track the face; to locate on it characteristic facial regions, such as eyes, mouth, and nose; and to extract and follow the movement of facial features, such as characteristic points in these regions or model facial gestures using anatomic information about the face. Facial features can be viewed (Ekman, 1975) either as static (i.e., skin color), slowly varying (i.e., permanent wrinkles), or rapidly varying (i.e., raising the eyebrows) with respect to time evolution. Detection of the position and shape of the mouth, eyes (particularly eyelids), wrinkles, and extraction of features related to them are the targets of techniques applied to still images of humans. However, in Bassili's (1979) experiments, expressions were recognized at above chance levels when based on image sequences, whereas only happiness and sadness were recognized at above chance levels when based on still images. Techniques that attempt to identify facial gestures for emotional expression characterization face the problems of locating or extracting the facial regions or features, computing the spatio-temporal motion of the face through optical flow estimation, and introducing geometric or physical muscle models describing the facial structure or gestures.

Body Gesture Analysis

The detection and interpretation of hand gestures have become an important part of human computer interaction in recent years (Wu, 2001). To benefit from the use of gestures in MMI, it is necessary to provide the means by which they can be interpreted by computers. The MMI interpretation of gestures requires that dynamic and/or static configurations of the human hand, arm, and other parts of the human body, be measurable by the machine. First attempts to address this problem resulted in mechanical devices that directly measure hand and/or arm joint angles and spatial position. The so-called glove-based devices best represent this solution's group.

Since the processing of visual information provides strong cues in order to infer the states of a moving object through time, vision-based techniques provide at least adequate alternatives to capture and interpret human hand motion. At the same time, applications can benefit from the fact that vision systems can be very cost efficient and do not affect the natural interaction with the user. These facts serve as the motivating forces for research in the modeling, analysis, animation, and recognition of hand gestures. Analyzing hand gestures is a comprehensive task involving motion modeling, motion analysis, pattern recognition, machine learning, and even psycholinguistic studies.

The first phase of the recognition task is choosing a model of the gesture. The mathematical model may consider both the spatial and temporal characteristics of the hand and hand gestures. The approach used for modeling plays a pivotal role in the nature and performance of gesture interpretation. Once the model is decided upon, an analysis stage is used to compute the model parameters from the image features that are extracted from single or multiple video input streams. These parameters constitute some description of the hand pose or trajectory and depend on the modeling approach used. Among the important problems involved in the analysis are those of hand localization, hand tracking, and selection of suitable image features. The computation of model parameters is followed by gesture recognition. Here, the parameters are classified and interpreted in light of the accepted model and perhaps the rules imposed by some grammar. The grammar could reflect not only the internal syntax of gestural commands, but also the possibility of interaction of gestures with other communication modes like speech, gaze, or facial expressions. Evaluation of a particular gesture recognition approach encompasses accuracy, robustness, and speed, as well as the variability in the number of different classes of hand/arm movements it covers.

Gesture Interpretation

Gesture analysis research follows two different approaches that work in parallel. The first approach treats a hand gesture as a two- or three-dimensional signal that is communicated via hand movement on the part of the user; as a result, the whole analysis process merely tries to locate and track that movement, so as to recreate it on an avatar or translate it to specific, predefined input interface (e.g., raising hands to draw attention or indicate presence in a virtual classroom).

The low level results of the approach can be extended, taking into account that hand gestures are a powerful expressive means. The expected result is to understand gestural interaction as a higher-level feature and encapsulate it into an original modal, complementing speech and image analysis in an affective MMI system (Wexelblat, 1995). This transformation of a gesture from a time-varying signal into a symbolic level helps overcome problems such as the proliferation of available gesture representations or failure to notice common features in them. In general, one can classify hand movements with respect to their function as follows (Cadoz, 1994):

- Semiotic: These gestures are used to communicate meaningful information or indications.
- Ergotic: Manipulative gestures that are usually associated with a particular instrument or job.
- Epistemic: Related to specific objects, but also to the reception of tactile feedback.

Semiotic hand gestures are considered to be connected, or even complementary, to speech in order to convey a concept or emotion. Especially two major subcategories—deictic gestures and beats. Gestures that consist of two discrete phases are usually semantically related to the spoken content and used to emphasize or clarify it (McNeill, 1992). This relation provides a positioning of gestures along a continuous space. This space is shown in Figure 2 below.

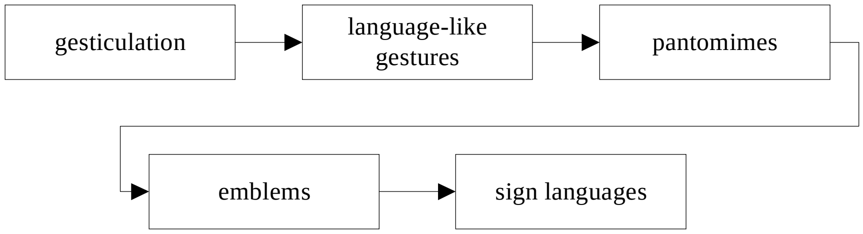
MAIN THRUST OF THE ARTICLE

Facial Feature Extraction

Robust and accurate facial analysis and feature extraction has always been a complex problem that has been dealt with by posing presumptions or restrictions with respect to facial rotation and orientation, occlusion, lighting conditions, and scaling. These restrictions are being eventually revoked in the literature, since authors deal more and more with realistic environments, keeping in mind the pioneering works in the field. A hierarchical, robust scheme coping with large variations in the appearance of diverse subjects, as well as of the same subject in various instances within real video sequences (see Figure 3), has been proposed by Votsis (2003). Soft a priori assumptions are made on the pose of the face or the general location of the features in it. Gradual revelation of information concerning the face is supported under the scope of optimization in each step of the hierarchical scheme, producing a posteriori knowledge about it and leading to a step-by-step visualization of the features in search.

Face detection is performed first through detection of skin segments or blobs, merging of them based on the

Figure 2. Gesture continuum (Cadoz, 1994)



probability of their belonging to a facial area, and identification of the most salient skin color blob or segment. Following this, primary facial features, such as eyes, mouth and nose, are dealt with as major discontinuities on the segmented, arbitrarily rotated face (Figure 4). In the first step of the method, the system performs an optimized segmentation procedure. The initial estimates of the segments, also called seeds, are approximated through min-max analysis and refined through the maximization of a conditional likelihood function. Enhancement is needed so that closed objects will occur, and part of the artifacts will be removed. Seed growing is achieved through expansion, utilizing chromatic and value information of the input image. The enhanced seeds form an object set, which reveals the in-plane facial rotation through the use of active contours applied on all objects of the set, which is restricted to a finer set, where the features and feature points are finally labeled according to an error minimization criterion (Figure 5).

Gesture Tracking and Recognition

The modeling of hand gestures depends primarily on the intended application within the MMI context. For a given application, a very coarse and simple model may be sufficient. However, if the purpose is a natural interaction, a model has to be established that allows many, if not all, natural gestures to be interpreted by the computer.

In general, human hand motion consists of the global hand motion and local finger motion. Hand motion capturing deals with finding the global and local motion of hand movements. Two types of cues are often used in the localization process: color cues (Kjeldsen, 1996) (see Figure 6), and motion cues (Freeman, 1995) (see Figure 7). Alternatively, the fusion of color, motion, and other visual or non-visual cues like speech or gaze are used (Sharma, 1996; Karpouzis, 2004).

To capture articulate hand motion in full degree of freedom, both global hand motion and local finger motion should be determined from video sequences. Different methods have been taken to approach this problem. One possible method is the appearance-based approaches in which 2-D deformable hand shape templates are used to track a moving hand in 2-D (Darrell, 1996). Another possible way is the 3-D model-based approach, which takes the advantages of a priori knowledge built in the 3-D models.

In certain applications, continuous gesture recognition is required; as a result, the temporal aspect of gestures must be investigated. Some temporal gestures are specific or simple and could be captured by low detail dynamic models. However, many high detail activities have to be represented by more complex gesture semantics, so modeling the low-level dynamics is insufficient. The HMM (Hidden Markov Model) technique (Bregler, 1997) and its variations (Darrell, 1996) are often employed in modeling, learning, and recognition of temporal signals. Because many temporal gestures involve motion trajectories and hand postures (see Figure 8), they are more complex than speech signals. Finding a suitable

Figure 3. The original frame from an expressive sequence

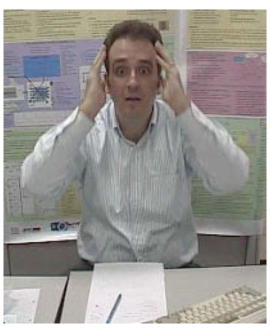


Figure 4. Detected primary facial features

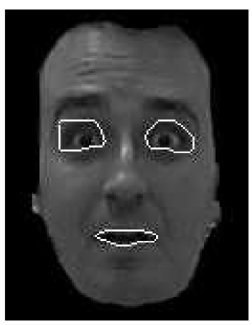
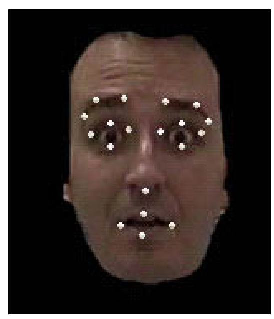


Figure 5. Detected facial features in the apex of an expression



approach to model hand gestures is still an open research problem. Practical large-vocabulary gesture recognition systems by HMM are yet to be developed.

FUTURE TRENDS

General gesture analysis studies consider gestures to be spontaneous, free form movements of the hands during speech (gesticulation), while others, termed emblems, are indicative of a specific emotion or action, such as an insult. An interesting conclusion (McNeill, 1992) is that the alternative use of gestures and speech in order to comprehend the communicated emotion or idea makes the whole concept of body language obsolete. Indeed, the study shows that instead of being "mere embellishments" of spoken content, gestures possess a number of paralinguistic properties. For example, such gestures convey a specific meaning only when considered as a whole, not as mere collections of low-level hand movements. While spoken words are usually unambiguous and can be semantically interpreted only when in a complete sentence or paragraph, gestures are atomic when it comes to conveying an idea, and typically, their actual form depends on the personality and current emotional state of a specific speaker. As a result, gestures cannot be analyzed with the same tools used to process the other modals of human discourse. In the case of gesticulation, we can regard gestures as functions of hand movement over time; the result of this approach is that the quantitative values of this representation (i.e., speed, direction, or repetition) can be associated with emotion-related values such as activation. This essentially means that, in many cases, we do not need to recognize specific gestures to deduce information about the user's emotional state, but merely track the movement of the user's arms through time. This concept can also help us distinguish a specific gesture from a collection of similar hand movements. For example, the "raise hand" gesture in a classroom or discussion and the "go away" or "I've had enough" gestures are similar when it comes to hand movement, since in both cases the hand is raised vertically. The only way to differentiate them is to compare the speed of the upward movement in both cases. In the latter case, the hand is raised in a much more abrupt manner. In our approach, such feedback is invaluable, since we try to analyze the user's emotional state by taking into account a combination of both gesture- and face-related features and not decide based on merely one of the two modals.

CONCLUSION

In this chapter, we presented a holistic approach to emotion modeling and analysis. Beginning from a symbolic representation of human emotions, based on their expression via facial expressions and hand gestures, we described approaches to extracting quantitative and qualitative feature information from video sequences. While these features can be used for simple representation purposes (e.g., animation or task-based interfacing), this methodology is closer to the target of affective computing and can prove useful in providing feedback on users' emotional states. Possible applications include human-like agents that assist everyday chores and react to user emotions, or sensitive artificial listeners that introduce conversation topics and react to specific user cues.

REFERENCES

Bassili, J.N. (1979). Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology, 37*, 2049-2059.

Bregler, C. (1997). Learning and recognition human dynamics in video sequences. *Proceedings of the IEEE Conference in Computer Vision and Pattern Recognition* (pp. 568-574), Washington, DC.

Cadoz, C. (1994). Les réalités virtuelles, Dominos: Flammarion.

Figure 6. Skin color probability

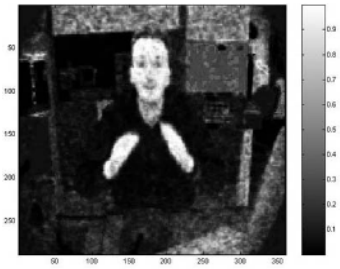
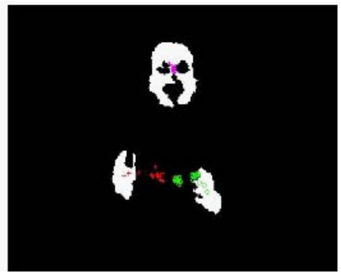


Figure 7. Detected moving hand segments

Figure 8. Hand tracking in a "clapping" sequence



Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J. (2001). Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine,1,* 32-80.

Darrell, T., Essa, I., & Pentland, A. (1996). Task-specific gesture analysis in real-time using interpolated views. *IEEE Transactions in Pattern Analysis and Machine Intelligence,* Washington, DC, *18*(12), 1236-1242.

Darrell, T., & Pentland, A. (1996). Active gesture recognition using partially observable Markov decision processes. *Proceedings of the IEEE International Conference in Pattern Recognition*.

Davis, M., & College, H. (1975). *Recognition of facial expressions*. New York: Arno Press.

Ekman, P. (1973). Darwin and facial expressions. Burlington, MA: Academic Press.

Ekman, P., & Friesen, W. (1975). *Unmasking the face*. Upper Saddle River, NJ: Prentice-Hall.

Ekman, P., & Friesen, W. (1978). The facial action coding system. Consulting Psychologists Press, Palo Alto, CA.

Freeman, W.T., & Weissman, C.D. (1995). Television control by hand gestures. *Proceedings of the International Workshop on Automatic Face and Gesture Recognition,* Zurich, Switzerland.

Karpouzis, K., Raouzaiou, A., Drosopoulos, A., Ioannou, S., Balomenos, T., Tsapatsoulis N., & Kollias, S. (2004). Facial expression and gesture analysis for emotionally-rich man-machine interaction. In N. Sarris, & M. Strintzis, (Eds.), *3D Modeling and Animation: Synthesis and Analysis Techniques*. Hershey, PA: Idea Group Publishers.

Karpouzis, K., Tsapatsoulis, N., & Kollias, S. (2000). Moving to continuous facial expression space using the MPEG-4 facial definition parameter (FDP) set. *Proceedings of SPIE Electronic Imaging,* San Jose, CA.

McNeill, D. (1992). Hand and mind: What gestures reveal about thought. Chicago: University of Chicago Press.

Picard, R.W. (2000). *Affective computing*. Cambridge: MIT Press.

Plutchik, R. (1980). Emotion: A psychoevolutionary synthesis. New York: Harper and Row.

Scherer, K., & Ekman, P. (1984). *Approaches to emotion*. Lawrence Erlbaum Associates, Mahwah, NJ.

Sharma, R., Huang, T.S., & Pavlovic, V.I. (1996). A multimodal gramework for interacting with virtual environments. In C.A. Ntuen, & E.H. Park (Eds.), *Human interaction with complex systems*. New York: Kluwer Academic Publishers.

Votsis, G., Drosopoulos, A., & Kollias, S. (2003). A modular approach to facial feature segmentation on real sequences, signal processing. *Image Communication*, *18*, 67-89.

Wexelblat, A. (1995). An approach to natural gesture in virtual environments. *ACM Transactions on Computer-Human Interaction, 2*(3), 179-200.

Wu, Y., & Huang, T.S., (2001). Hand modeling, analysis, and recognition for vision-based human computer interaction. *IEEE Signal Processing Magazine, 18*(3), 51-60.

KEY TERMS

Activation-Emotion Space: A 2-D representation of the emotion space, with the two axes representing the magnitude and the hue of a specific emotion.

Affective Computing: A recent theory that recognizes that emotions play an essential role in perception and learning by shaping the mechanisms of rational thinking. In order to enhance the process of interaction, we should design systems with the ability to recognize, to understand, and even to have and express emotions.

Hidden Markov Model: Statistical models of sequential data utilized in many machine learning applications (e.g., speech and gesture recognition).

Model-Based Gesture Analysis: In this framework, gestures are modeled as usually finite states of hand shape or position. This approach captures both the spatial and the temporal nature of the gestures, which is essential for analysis and recognition purposes.

Semiotic Hand Gestures: Gestures used to communicate meaningful information or serving as indications; such gestures convey specific emotions in a more expressive manner than vague hand movement.

Skin Color Estimation: A breakthrough in face detection and segmentation was the representation of skin color with a Gaussian model in a subset of the CrCb space, irrespectively of the actual skin color of the subject.

Universal Emotions: Mainly after the influence of Ekman, these six emotions are considered to be universal, in the sense that they are uniformly recognized across different cultures.

Faculty Perceptions and Participation in Distance Education

Kim E. Dooley
Texas A&M University, USA

James R. Lindner
Texas A&M University, USA

Chanda Elbert
Texas A&M University, USA

Timothy H. Murphy
Texas A&M University, USA

Theresa P. Murphrey
Texas A&M University, USA

INTRODUCTION

Research in the field of distance education has recognized the need for a change and modification of the faculty role in teaching at a distance (Jones, Lindner, Murphy & Dooley, 2002; Kanuka, Collett & Caswell, 2002; Miller & Pilcher, 2001). While technological advancements are an important part of the distance-learning environment, basic changes in teaching methods, technique, and motivation are needed to make distance education more effective (Purdy & Wright, 1992). Many studies cite faculty resistance to instructional technology as a primary barrier to the continued growth of distance education programs (Jones et al., 2002; McNeil, 1990). McNeil (1990) noted that attitudinal issues related to how faculty perceive and interact with technology are a bigger barrier to adoption and diffusion of distance education than is technology infrastructure.

BACKGROUND

This chapter addresses perceptions of faculty with respect to barriers to adoption, roles and responsibilities, competencies, and rewards. Barriers stem from the lack of perceived institutional support (faculty rewards, incentives, training, etc.) for course conversion to distance education formats (O'Quinn & Corry, 2002; Perreault et al., 2002). As distance education programs continue to proliferate globally, colleges and universities must commit to address the needs of faculty (McKenzie, Mims, Bennett & Waugh, 2000). Despite the fact that much of the literature in distance education discusses the importance of faculty, this group has been largely neglected by the research.

Dooley and Murphy (2000) found that faculty members lacked experience in teaching learners at a distance and that they were much more confident in their technical competence than they were in their methodological ability to use modern technologies in their teaching. These authors further found that faculty perceived training and assistance in the use of instructional technologies to be less available than equipment and facilities. Additionally, faculty members who had not participated in distance education perceived the level of support as lower than those who had taught classes at a distance. The ability of an organization to adapt to these changes is influenced by the following: competence, or the knowledge, skills, and abilities of its staff; value, or the amount of importance the staff places on the role of these technologies to accomplish teaching and learning; information technology support, or the availability of high quality facilities, equipment, technical support, and training (Dooley & Murphy, 2000).

Lindner, Murphy, and Dooley (2002) extended these conclusions by looking at how these factors affect faculty adoption of distance education. Research revealed that faculty members lacked confidence in their ability to use technology in their teaching, perceived technology to be a valuable addition to the teaching and learning environment, and believed the overall level of support for the use of technology in teaching to be low. Tenure status and academic rank/position for tenure-track faculty were inversely related to overall distance education scores. Non-tenured assistant professors had the highest overall distance education scores and the highest competency scores.

Students learn from competent instructors who have been trained how to communicate effectively through technology. Thomas Cyrs (1997) identified areas of competence important to a distance education environment: course planning and organization, verbal and nonverbal presentation skills, collaborative teamwork, questioning strategies, subject matter expertise, involving students and coordinating their activities at field sites, knowledge of basic learning theory, knowledge of the distance learning field, design of study guides, graphic design and visual thinking (Cyrs, 1997).

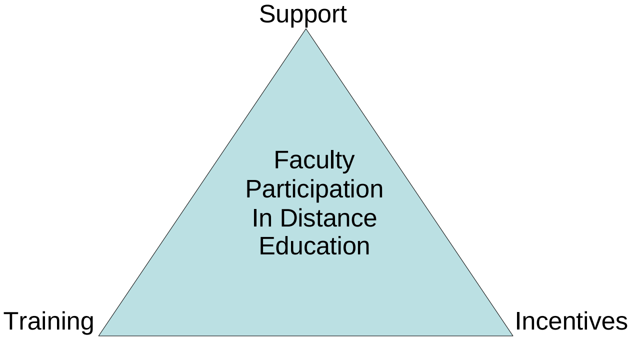
Linda Wolcott (1997) conducted an analysis of the institutional context and dynamics of faculty rewards at research universities. She discovered that 1) distance education occupies a marginal status, 2) distance teaching is neither highly valued nor well rewarded as a scholarly activity, 3) distance teaching is not highly related to promotion and tenure decisions, and 4) rewards for distance teaching are dependent on the academic unit's commitment to distance education.

As indicated by Moore (1997), distance education programs with a commitment to faculty support and training result in higher quality programs. As the complexity continues and the desire to integrate distance education programs expands, attention must be given to faculty training and support.

Enhancing Faculty Participation

Overall, faculty members recognize that distance education technologies are—and will be—an important part of the instructional process. However, they perceive support and training to be less available than equipment. Enhancing faculty participation requires that resources be directed to provide adequate levels of support and training such that these technologies are used for the benefit of students (Howard, Schenk & Discenza, 2004).

Figure 1. Enhancements to increase faculty participation in distance education.



It is the integration of incentives, training, and support that promote the adoption of distance education delivery strategies by university faculty.

While faculty recognize the potential, intervention strategies are necessary to alter how people perceive and react to distance education technologies. It is apparent that steps must be taken to increase faculty training and support. Three major areas require consideration: 1) support, 2) training, and 3) incentives. Support extends beyond "verbal" to providing the support/professional staff to assist faculty. Training should not only include technology exposure, but instructional design, pedagogy/andragogy, and "cook-book" strategies and "how-to" manuals. By providing incentives such as release time, mini-grants, continuing education stipends, and recognition in the promotion and tenure process, faculty will have more than verbal encouragement to continue, or begin, using distance education technologies and will have the reason to do so (Dooley & Murphrey, 2000; Murphrey & Dooley, 2000).

Rockwell et al. (1999) found that the primary incentives for faculty participation were *intrinsic* or personal rewards, including the opportunity to provide innovative instruction and apply new teaching techniques. Other incentives included extending educational opportunities beyond the traditional institutional walls, and release time for faculty preparation.

FUTURE TRENDS AND CONCLUSION

Faculty member participation in distance education requires a competence in using technology, an attitude that distance education is important and valuable, and access to quality infrastructure (Hawkes & Coldeway, 2002). Faculty roles and responsibilities must change to accommodate the use of these technologies and it must be recognized that teaching at a distance requires a different set of competencies (Richards, Dooley & Lindner, 2004). Integration of distance education technologies into the teaching and learning process requires a shift of attitude on the part of the faculty members and the removal of barriers created by the lack of institutional support.

REFERENCES

Cyrs, T. (1997). *Teaching at a distance with merging technologies: An instructional systems approach.* Las Cruces, NM: Center for Educational Development, New Mexico State University.

Dooley, K.E., & Murphy, T.H. (2000). College of Agriculture faculty perceptions of electronic technologies in teaching. *Journal of Agricultural Education, 42*(2), 1-10.

Dooley, K.E., & Murphrey, T.P. (2000). How the perspectives of administrators, faculty, and support units impact the rate of distance education adoption. *The Journal of Distance Learning Administration, 3*(4). *http://www.westga.edu/~distance/jmain11.html*

Hawkes, M., & Coldeway, D.O. (2002). An analysis of team vs. faculty-based online course development: Implications for instructional design. *Quarterly Review of Distance Education, 3*(4), 431-441.

Howard, C., Schenk, K., &Discenza, R. (2004). *Distance learning and university effectiveness: Changing educational paradigms for online learning.* Hershey, PA: Idea Group Publishing.

Jones, E.T., Lindner, J.R., Murphy, T.H., & Dooley, K.E. (2002). Faculty philosophical position towards distance education: Competency, value, and education technology support. *Online Journal of Distance Learning Administration [Electronic Journal], 5*(1). Retrieved from *http://www.westga.edu/~distance/jmain11.html*

Kanuka, H., Collett, D., & Caswell, C. (2002). University instructor perceptions of the use of asynchronous text-based discussion in distance courses. *American Journal of Distance Education, 16*(3), 151-167.

Knowles, M.S. (1990). *The adult learner: A neglected species.* Houston, TX: Gulf Publishing.

Lindner, J.R., Murphy, T.H., & Dooley, K.E. (2002). Factors affecting faculty perceptions of technology-mediated instruction: Competency, value, and educational technology support. *NACTA Journal, 46*(4), 2-7.

McKenzie, B., Mims, N., Bennett, E., & Waugh, M. (2000). Needs, concerns and practices of online instructors. *Online Journal of Distance Learning Administration [Electronic Journal], 3*(3). Retrieved from *http://www.westga.edu/~distance/jmain11.html*

McNeil, D.R. (1990). *Wiring the ivory tower: A round table on technology in higher education.* Washington, DC: Academy for Educational Development.

Miller, G., & Pilcher, C.L. (2001). Levels of cognition researched in agricultural distance education courses in comparison to on-campus courses and to faculty perceptions concerning an appropriate level. *Journal of Agricultural Education, 42*(1), 20-27.

Moore, M.G. (1997). Quality in distance education: Four cases. *The American Journal of Distance Education, 11*(3), 1-7.

Murphrey, T.P., & Dooley, K.E. (2000). Perceived strengths, weaknesses, opportunities, and threats impacting the diffusion of distance education technologies for colleges of agriculture in land grant institutions. *Journal of Agricultural Education, 41*(4), 39-50.

O'Quinn, L., & Corry, M. (2002). Factors that deter faculty from participation in distance education. *Online Journal of Distance Learning Administration [Electronic Version], 5*(4). Retrieved from *http://www.westga.edu/~distance/jmain11.html*

Perreault, H., Waldman, L., Alexander, M. et al. (2002). Overcoming barriers to successful delivery of distance-learning course. *Journal of Education for Business, 77*(6), 313-318.

Purdy, L.N., & Wright, S.J. (1992). Teaching in distance education: A faculty perspective. *The American Journal of Distance Education, 6*(3), 2-4.

Richards, L.J., Dooley, K.E., & Lindner, J.R. (2004). Online course design principles. In C. Howard, K. Schenk & R. Discenza (Eds.), *Distance learning and university effectiveness: Changing education paradigms for online learning.* Hershey, PA: Idea Group Publishing.

Rockwell, S.K., Schauer, J., Fritz, S.M., & Marx, D.B. (1999). Incentives and obstacles influencing higher education faculty and administrators to teach via distance. *Online Journal of Distance Learning Administration, 2*(4). Retrieved from *http://www.westga.edu/~distance/rockwell24.html*

Rogers, E.M. (2003). *Diffusion of innovations* (5th ed.). New York: Freepress.

Wolcott, L.L. (1997). Tenure, promotion, and distance education: Examining the culture of faculty rewards. *The American Journal of Distance Education, 112*, 3-18.

KEY TERMS

Adoption: A decision to make full use of a new idea as a preferred method (Rogers, 2003).

Andragogy: The art and science of teaching adults (Knowles, 1990).

Competence: A measure of perceived level of ability by faculty in the use of electronic technologies often associated with distance education (Jones, Lindner, Murphy & Dooley, 2002).

Distance Education: Any education received by learners that occurs when the instructor and learner are separated by location and/or time.

Incentive: Intrinsic or extrinsic motivational factors that impact faculty decisions to participate in distance education.

Pedagogy: The art and science of teaching children (Knowles, 1990).

Value: A measure of the importance perceived by faculty of the role that technology will have on distance delivery (Jones, Lindner, Murphy & Dooley, 2002).

Fault Tolerance for Distributed and Networked Systems

Wenbing Zhao
University of California, Santa Barbara, USA

Louise E. Moser
University of California, Santa Barbara, USA

P. Michael Melliar-Smith
University of California, Santa Barbara, USA

INTRODUCTION

The services provided by computers and communication networks are becoming more critical to our society. Such services increase the need for computers and their applications to operate reliably, even in the presence of faults. Fault tolerance is particularly important for distributed and networked systems (Mullender, 1993), including telecommunication, power distribution, transportation, manufacturing, and financial systems.

Fault-tolerant computing has been oriented towards custom designs and computer hardware (Siewiorek & Swarz, 1998), towards particular kinds of applications (Wensley et al., 1978), and towards operating systems (Borg, 1989). However, as computer and communication systems have become more complex and as hardware has become cheaper, designs have moved towards commercial off-the-shelf hardware and towards fault tolerance middleware located as a software layer between the application and the operating system.

Traditional proprietary designs are now being challenged by industry standards, such as the Fault Tolerant CORBA standard (Object Management Group, 2000) for distributed object applications based on the Common Object Request Broker Architecture (CORBA), and also the Hardware Platform Interface and Application Interface Specification (Service Availability Forum, 2003) for telecommunication and other embedded systems.

Of particular importance in the development of fault-tolerant computing is the distinction between application-aware and application-transparent fault tolerance. In application-aware fault tolerance, the application is aware of, and explicitly exploits, the mechanisms provided by the fault tolerance infrastructure, using application program interfaces (APIs). The application programmer writes code corresponding to the APIs to perform specific operations, such as to checkpoint an application process and to restore the process from the checkpoint, or to send a message across the network to a process on another processor and to receive the message on that other processor.

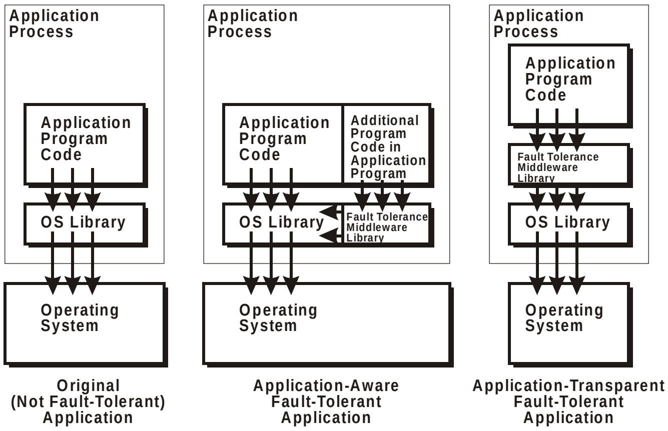
In application-transparent fault tolerance, the application is rendered fault tolerant, without involvement by, or modification to, the application program (Bressoud, 1998; Moser, Melliar-Smith, & Narasimhan, 1998). A fault tolerance middleware library is interposed ahead of the operating system libraries, between the application and the operating system (Narasimhan, Moser, & Melliar-Smith, 2002). When the application invokes standard operating system functions, those invocations are diverted to the fault tolerance middleware library, which modifies the operating system functions to provide additional fault tolerance functionality (Zhao, Moser, & Melliar-Smith, 2004). The application programmer does not need to implement or invoke additional methods for fault tolerance or to include additional fault tolerance code in the application program.

BACKGROUND

The basic terminology of fault-tolerant computing can be found in Laprie (1992). The terms failure, error, and fault were defined originally by Melliar-Smith & Randell (1977) and have become part of the ISO standard.

- A *failure* is the event of a system's generating a result that does not satisfy the system specification, or of the system's not generating a result that is required by the system specification. A failure is defined by the system specification, without reference to any components internal to the system, or to any enclosing system of which the system is a component.
- An *error* is incorrect information, or lack of information, within a system that will, unless detected and corrected, lead to a failure of the system.

Figure 1. The original application is shown at the left. In application-aware fault-tolerance (shown in the middle), the application program includes additional code that invokes the fault tolerance middleware library. In application-transparent fault-tolerance (shown at the right), the fault tolerance middleware library is interposed between the application and the operating system library.



- A *fault* is the original cause of an error, whether hardware or software. Sometimes the cause of the error is strictly objective but sometimes, particularly for software, it is a matter of subjective opinion.

A failure of a system might be a fault within a larger enclosing system. Similarly, a fault might be the failure of one of the components from which the system is constructed.

Faults (Cristian, 1991) are further classified as follows:

- A *crash fault* occurs when a component operates correctly up to some point in time, after which it produces no further results.
- A *timing fault* occurs when a component produces results at the wrong time, either too early or too late.
- An *omission fault* occurs when a component produces some results but not others.
- A *commission fault* occurs when a process or processor generates incorrect results. A *Byzantine* or *malicious fault* is a form of commission fault in which a process or processor generates incorrect results that are intentionally designed to mislead the algorithms or components of the system.

The metrics used in fault-tolerant computing include:

- *Reliability* is a measure of the uptime of a system in the absence of failure, and is given by the Mean Time Between Failure (MTBF).
- *Repairability* is a measure of how quickly a failed component or system can be restored to service, and is given by the Mean Time To Repair (MTTR).
- *Availability* is a measure of the uptime of a system, and is related to MTBF and MTTR by the formula: Availability = MTBF/(MTBF + MTTR). High availability typically means five nines (99.999%) or better, which corresponds to 5.25 minutes of planned and unplanned downtime per year.

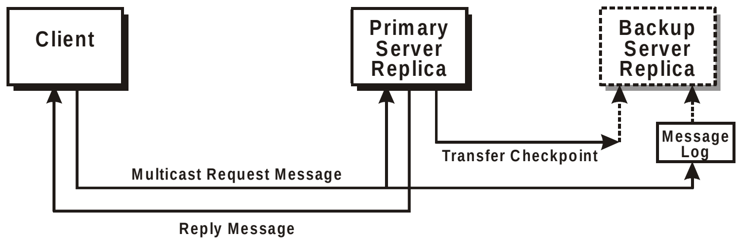
TECHNOLOGY OF FAULT TOLERANCE

Replication

The basic strategy used in fault-tolerant systems to protect an application against faults is *replication*, so that if one replica becomes faulty, another replica is available to provide the service. The unit of replication can be an entire processor, a process, a Java container, a CORBA object or some other component. In this article, we refer to such a unit as a component. The replicas of a component constitute a group of two or more components. To provide fault isolation, the replicas of a component must be independent of each other, with no shared memory or data and with communication controlled by the fault tolerance middleware, so that the failure of one replica does not disable another replica. Several kinds of replication are possible (Powell, 1991).

Passive replication, shown in *Figure 2*, distinguishes one of the server replicas as the primary server replica and the other as the backup server replica. The primary executes methods invoked on the group, and the backup does not execute those methods. In cold passive replica-

Figure 2. Passive replication. A client invokes a method of the replicated server, which is multicast in a request message. The primary server replica executes the method, but the backup server replica does not. A checkpoint containing the state of the primary is transferred to the backup at the end of each method invocation.



tion, the program code is not loaded into memory at the backup. In warm passive replication, the program code is loaded into memory at the backup, and the primary transfers its state periodically to the backup, where it is copied directly into the backup. Warm passive replication results in more rapid recovery from the failure of the primary replica than does cold passive replication.

Active replication, shown in *Figure 3*, regards all of the server replicas as equal and, thus, all of the server replicas execute every method invoked on the group. The replicas execute the methods independently of each other and at approximately, but not necessarily exactly, the same physical time. The first result received from any of the replicas is accepted, and the results from the other replicas are suppressed. Active replication has faster recovery from faults than passive replication because, if one replica fails, the other replicas are already calculating the required results and can supply them without delay.

Active replication with voting is similar to active replication, except that the results from the replicas are subjected to majority voting and the majority value is used.

Semi-active replication is a hybrid of passive and active replication. Both the primary and the backup replicas execute each method invocation. The primary determines the order in which the messages are processed and communicates that order to the backups. A backup follows the directives of the primary, such as the order in which to process messages, and thus lags slightly behind the primary in executing method invocations. Only the primary communicates results and makes further method invocations; the backup does not do so.

Passive replication, active replication, and semi-active replication are used to protect against crash faults. Active replication with voting is used to protect against commission faults.

The most challenging aspect of replication-based fault tolerance is maintaining strong replica consistency, as methods are invoked on the replicas, as the states of the replicas change, and as faults occur. Strong replica consistency simplifies the programming of fault-tolerant applications, particularly for recovery from faults. The basic strategy for maintaining strong replica consistency is to start the replicas in the same state and to ensure that they perform the same operations in the same order. Considerable care is required in implementing the fault tolerance infrastructure to maintain strong replica consistency.

Checkpointing and Restoration from a Checkpoint

To protect an application process against faults, a checkpoint (snapshot) of the process is taken and is trans-

Figure 3. Active replication. All of the server replicas execute every method invoked on the group, and communicate the results of those invocations. Duplicate invocations and duplicate responses are detected and suppressed. A checkpoint is used only to bring up a new active server replica.

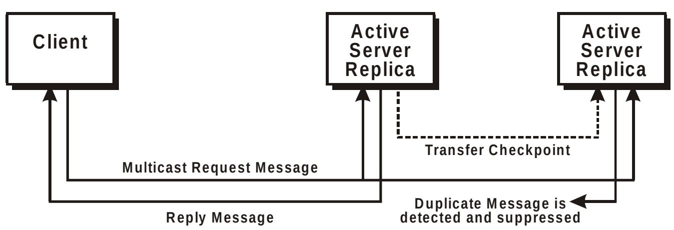
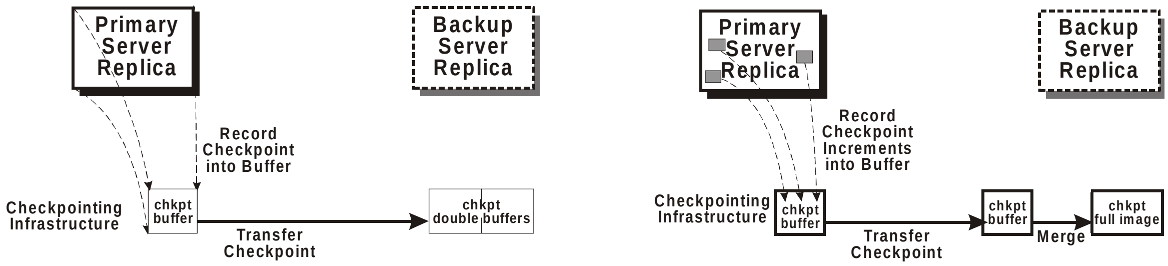


Figure 4. In application-transparent checkpointing, the checkpointing infrastructure captures the state of an application process, either the entire state (full checkpointing), shown at the left, or only the state that has changed since the last checkpoint (incremental checkpointing), shown at the right.



ferred from the primary replica to a backup replica, or from an active replica to a new or recovering active replica, as shown in *Figure 4*.

In application-aware checkpointing, the application programmer implements a getState() method that captures particular parts of the application process state and encodes that state into a byte sequence. The application programmer also implements a corresponding setState() method that decodes the byte sequence and restores the process from the checkpoint. This encoding is called serialization or pickling. The Java Virtual Machine provides serialization for Java objects that are defined to be serializable. For other languages, preprocessor tools can be used to preprocess the source code of a program and synthesize the source code of the required getState() and setState() methods.

In application-transparent checkpointing, the checkpointing infrastructure captures the state of the application process, without the need for the application programmer to implement the getState() and setState() methods.

With full checkpointing, the checkpointing infrastructure captures the entire memory image of the process (including file descriptors, thread stacks, etc.), without knowing about the internal data structures of the application program. It records each checkpoint in a checkpoint buffer and transmits the checkpoint across the network to a designated processor as a background activity. If a fault occurs, the infrastructure restarts the process on the same or a designated processor, using the most recent checkpoint to restore the process to the state it had at the time of the checkpoint.

With incremental checkpointing, the checkpointing infrastructure captures only those pages of the memory image that have changed since the last checkpoint. The infrastructure transmits the incremental checkpoint to the designated processor, where the incremental checkpoint is merged with the most recent full checkpoint.

For applications that involve multiple threads within a process or data structures that contain pointers, it is difficult to implement the getState() and setState() methods of application-aware checkpointing. Thus, application-transparent checkpointing might be preferred. However, application-transparent checkpointing does not produce checkpoints that are portable across hardware architectures, because a checkpoint taken as a binary image contains values of variables that differ for different architectures, such as memory addresses.

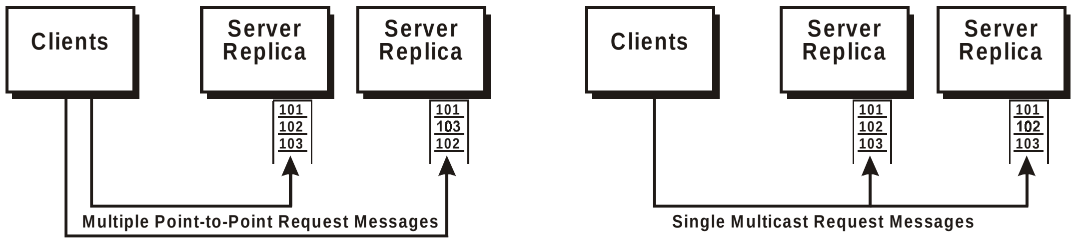
Checkpointing is used by all replication strategies but in different ways. Passive replication uses checkpointing during normal operation. Active replication, active replication with voting, and semi-active replication do not use checkpointing during normal operation, but use it to initialize a new or recovering replica.

Messaging and Multicasting

Fault-tolerant distributed systems typically use multicast protocols to deliver messages to the groups of replicas efficiently, reliably and in the same order, as shown in *Figure 5*. To maintain strong replica consistency, the replicas must receive the messages in the same order, so that they process the invocations and responses in the messages in the same order. For both active and passive replication, the messaging infrastructure must detect duplicate invocations and duplicate responses, and suppress them. Preferably, the duplicates are detected and suppressed at their source, to reduce traffic on the network; otherwise, they must be detected and suppressed at the destinations. For passive replication, the messaging infrastructure logs messages at the backup replica so that, if the primary replica fails, the messages after the checkpoint can be replayed.

Some fault-tolerant systems have exploited multicast group communication protocols that deliver messages reliably and in the same total order (Birman & van Renesse, 1994), which simplifies the maintenance of strong replica consistency. The disadvantage of that approach is that a message cannot be delivered until the total order has been agreed by the processors within the distributed system,

Figure 5. Sending multiple point-to-point request messages to the replicas in a group (shown at the top) is less efficient than sending a single multicast request message to the group (shown at the bottom). Moreover, sending multiple point-to-point request messages to the replicas in a group does not guarantee atomic delivery, with messages from different clients delivered in the same order to all replicas in the group.



with adverse effects on the latency (delay) to message delivery. Moreover, the message order chosen by the multicast protocol might not be appropriate for a real-time application, because it might disregard real-time priorities and deadlines.

Other fault-tolerant systems integrate the multicast protocol more tightly into the replication mechanisms that maintain strong replica consistency. For example, in semi-active replication, messages are delivered to the primary replica when they are received, in whatever order is appropriate, such as that chosen by a real-time scheduler.

Sanitizing Non-Deterministic Operations

Messaging is one source of replica non-determinism in a fault-tolerant distributed system, because messages can be received by the replicas in different orders, due to loss of messages and retransmissions, delays in the network, etc. To maintain strong replica consistency, messages must be delivered to the replicas in the same order.

Another source of replica non-determinism is multi-threading. If two threads within a replica share data, they must claim and release mutexes that protect that shared data. However, the threads in two replicas will most likely run at slightly different speeds. In one replica, one thread might be the first to claim a mutex and, in another replica, a different thread might be the first to claim the mutex. To maintain strong replica consistency, mutexes must be granted to the threads within the replicas in the same order.

Other sources of replica non-determinism include operating system functions that return values local to the processor on which they are executed, such as rand() and gettimeofday(), or inputs for the replicas from different redundant sources, or system exceptions due to, say, lack of memory on one of the processors. These sources of replica non-determinism must be sanitized, so that all of the replicas see the same values of the functions, the same inputs from the redundant sources, and the same system exceptions. Such virtual determinism must be provided for the replicas, regardless of which kind of replication is used.

Fault Detection, Analysis and Notification

Fault detection is essential for reliable operation. Faults can be detected in many different ways and within robust fault-tolerant systems any component can report a fault.

A fault detector (Chandra & Toueg, 1996) monitors the occurrence of faults in a component. Typically, it is based on timeouts and it simply pings a component periodically; failure to respond is regarded as evidence of a fault. A fault detector is unreliable (inaccurate), because it cannot determine whether a component has failed or is merely slow.

A fault notifier allows a component to subscribe for fault reports of particular kinds. It receives fault reports from fault detectors and fault analyzers, and supplies fault reports to the subscribers.

A fault analyzer or filter can be used to reduce the number of fault reports that the fault notifier transmits to a subscriber. It subscribes for fault reports from the fault notifier and aggregates multiple related fault reports into a single fault report.

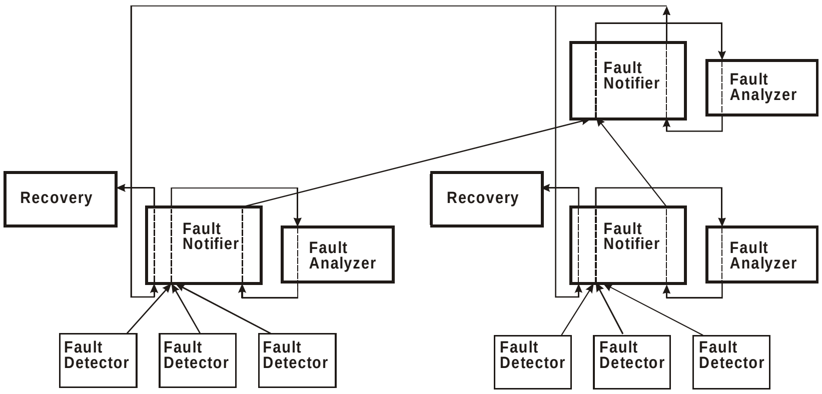
Fault detection, analysis and notification is often arranged hierarchically, as shown in *Figure 6*, so that ping traffic is localized and minimized, and fault reports are disseminated where they are needed quickly.

Recovery and Repair

Recovery ensures that there is a component available that can continue to provide service to the clients. Repair ensures that a faulty component is identified, sometimes

F

Figure 6. Fault detection, analysis and notification are arranged hierarchically to localize and minimize ping traffic, and to disseminate fault reports quickly.



referred to as primary cause analysis, and that steps are taken to test, reinitialize and return to service, or to replace, the faulty component. Repair may take longer than recovery, but is equally essential.

FUTURE TRENDS AND CONCLUSION

Application-aware fault tolerance increases the complexity of the application program, with additional application programming and costs, increased application development and testing timescales, and on-going application maintenance costs.

In contrast, application-transparent fault tolerance incurs little or no increase in the complexity and cost of the application program, but incurs an increase in the complexity of the fault tolerance middleware and in the performance overhead of that middleware.

As applications become more complex, as computer hardware becomes faster and less expensive, and as users demand better quality of service, the costs of application-aware fault tolerance will increase, and the costs of application-transparent fault tolerance will decrease to make application-transparent fault tolerance preferable.

REFERENCES

Birman, K., & van Renesse, R. (1994). *Reliable distributed computing with the Isis Toolkit*. Los Alamitos, CA: IEEE Computer Society Press.

Bressoud, T.C. (1998). TFT: A software system for application-transparent fault tolerance. In *Proceedings of the IEEE 28th Annual International Symposium on Fault-Tolerant Computing* (pp. 128-137). Munich, Germany.

Borg, A., Blau, W., Braetsch, W., & Oberle, W. (1989). Fault tolerance under Unix. *ACM Transactions on Computer Systems*, *7* (1), 1-24.

Chandra, T.D., & Toueg, S. (1996). Unreliable failure detectors for reliable distributed systems. *Journal of the ACM*, *43* (2), 225-267.

Cristian, F. (1991). Understanding fault-tolerant distributed systems. *Communications of the ACM*, *34* (2), 56-78.

Laprie, J.C. (1992). *Dependability: Basic concepts and terminology*. Dependable Computing and Fault-Tolerant Systems Series. Vienna: Springer-Verlag.

Melliar-Smith, P.M., & Randell, B. (1977). Software reliability: The role of programmed exception handling. In *Proceedings of the ACM Conference on Language Design for Reliable Software* (pp. 28-30). Raleigh, NC; *SIGPLAN Notices*, *12* (3), 95-100.

Moser, L.E., Melliar-Smith, P.M., & Narasimhan, P. (1998). Consistent object replication in the Eternal system. *Theory and Practice of Object Systems*, *4* (2), 81-92.

Mullender, S. (1993). *Distributed systems*. Menlo Park, CA: Addison-Wesley.

Narasimhan, P., Moser, L.E., & Melliar-Smith, P.M. (2002). Strongly consistent replication and recovery of fault-tolerant CORBA applications. *Computer Science and Engineering Journal*, *17* (2), 103-114.

Object Management Group. (2000). *Fault Tolerant CORBA*. OMG Technical Committee Document formal/02-06-59. Chapter 23. CORBA/IIOP 3.0. Retrieved from *www.omg.org*

Powell, D. (1991). *Delta-4: A generic architecture for dependable distributed computing*. Research Reports ESPRIT. Berlin: Springer-Verlag.

Service Availability Forum. (2003). *Hardware platform interface and application interface specifications*. Retrieved from www.saforum.org.

Siewiorek, D.P., & Swarz, R.S. (1998). *Reliable computer systems: Design and evaluation*. Natick, MA: A.K. Peters.

Wensley, J.H., Lamport, L., Goldberg, J., Green, M.W., Levitt, K.N., Melliar-Smith, P.M., Shostak, R.E., & Weinstock, C.B. (1978). SIFT: Design and analysis of a fault-tolerant computer for aircraft control. *Proceedings of the IEEE*, *66* (10), 1240-1255.

Zhao, W., Moser, L.E., & Melliar-Smith, P.M. (2004). Design and implementation of a pluggable fault tolerant CORBA infrastructure. *Cluster Computing: The Journal of Networks, Software Tools and Applications, Special Issue on Dependable Distributed Systems*, *7*(4), 317-330.

KEY TERMS

Application Aware vs. Application Transparent: In application-aware fault tolerance, the application programmer writes code for fault tolerance methods that perform specific operations. In application-transparent fault tolerance, the fault tolerance middleware performs those operations automatically, using standard operating system functions and the technique of library interpositioning.

Checkpointing (Full, Incremental): In full checkpointing, all of the state of the process is captured. In incremental checkpointing, only that part of the state that has changed since the last checkpoint in captured.

Failure, Error, Fault: A failure is the event of a system's generating a result that does not satisfy the system specification or of the system's not generating a result that is required by the system specification. An error is incorrect information, or lack of information, within the system that will, unless detected and corrected, lead to failure of the system. A fault is the original cause of the error.

Fault (Crash, Timing, Omission, Commission, Byzantine): A crash fault occurs when a component operates correctly up to some point in time, after which it produces no further results. A timing fault occurs when a component produces results at the wrong time. An omission fault occurs when a component produces some results but not others. A commission fault occurs when a component generates incorrect results. A Byzantine or malicious fault is a form of commission fault in which a component generates incorrect results intentionally to mislead the algorithms or components of the system.

Fault Detector, Analyzer, Notifier: A fault detector monitors the occurrence of faults in a component. A fault analyzer subscribes for fault reports from a fault notifier and aggregates multiple related fault reports into a single fault report. A fault notifier receives fault reports from fault detectors and fault analyzers, and supplies fault reports to subscribers for those reports.

Fault Tolerance: The ability to provide continuous service, even in the presence of faults.

Multicast Protocol: Delivers messages reliably and in the same order to all of the replicas within a group.

Recovery and Repair: Recovery is the restoration of the state of a component on the same or a different processor, so that the component can continue to execute operations and provide service. Repair involves identification of the primary cause of the fault and re-initialization or replacement of the faulty component.

Reliability, Repairability, Availability: Reliability is a measure of the uptime of a system in the absence of failure, and is given by the Mean Time Between Failure (MTBF). Reparability is a measure of how quickly a failed system can be restored to service, and is given by the Mean Time To Repair (MTTR). Availability is a measure of the uptime of a system, and is given by: Availability = MTBF/(MTBF + MTTR).

Replication (Passive, Active, Active with Voting, Semi-active): Multiple copies of an application program execute typically as multiple processes on different processors. In passive replication, only one replica, the primary replica, executes the operations, and one or more backups stand ready to perform the operations if the primary fails. In active replication, all of the replicas execute the operations, and communicate their results. In active replication with voting, all of the replicas execute the operations and the result produced by the majority of the replicas is taken as the result of the operation. In semi-active replication, both the primary and the backup replicas execute the operations, but the backups follow the directives of the primary and only the primary outputs the results.

Flexible Job–Shop Scheduling Problems

Imed Kacem
Université de Technologie de Troyes, France

INTRODUCTION

Planning and scheduling problems in various industrial environments are combinatorial and very difficult. Generally, it is extremely hard to solve these types of problems in their general form. Scheduling can be formulated as a problem of determining the best sequence to execute a set of tasks on a set of resources, respecting specific constraints like precedence or disjunctive constraints (Carlier & Chrétienne, 1988). They consist generally in a simultaneous optimization of a set of non-homogeneous and conflicting goals. Therefore, the exact algorithms such as branch and bound, dynamic programming, and linear programming are not suitable for such problems and need a lot of time and memory to converge.

Because of this difficulty, experts prefer to find not necessary the optimal solution, but a good one to solve the problem. To this end, new search techniques such as genetic algorithms (Dasgupta & Michalewicz, 1997; Sarker, Abbas & Newton, 2001), simulated annealing (Kirkpatrick, Gelatt & Vecchi, 1983), and tabu search (Golver, Taillard & De Werra, 1993) are proposed to reach this aim: construct an approximated solution for a large set of hard optimization problems.

In this article, we are interested in the evolutionary techniques and their application to an important branch of scheduling problems. We aim in particular to present an overview of the recent models proposed to solve flexible job shop scheduling problems using genetic algorithms.

FLEXIBLE JOB SHOP SCHEDULING PROBLEM

The Flexible Job Shop Scheduling Problem is an extension of the classical job shop problem. Indeed, it represents two classical combinatorial optimization problems. The first one is a linear assignment since each task can be carried out on a set of resources according to different processing times. The second problem consists of finding the best sequence of the tasks on the resources by respecting all the problem's constraints (interdiction of pre-emptive execution, disjunctive constraints on the resources, precedence constraints) (Dauzère-Pérès, Roux & Lasserre, 1998; Mastrolilli, 2000).

The structure of the flexible job shop scheduling problem can be described as follows. We have a set of jobs and a set of machines. Each job consists of a certain number of ordered operations. The execution of each operation requires one machine selected from the machines existing in the shop. Therefore, we should find the best assignment of the machines before scheduling. The sequencing of operations is fixed for each job (precedence constraints). All machines are initially available and each job has a release date. For a given machine and a given operation, a processing time is defined as the duration necessary to perform the task on the above machine. The pre-emption condition is not allowed, and each machine can perform operations one after the other (disjunctive resource constraints).

MODELLING FLEXIBLE JOB SHOP

The encoding problem is a main step in the genetic algorithm conception. Such a problem has been intensively studied in the literature (Dasgupta & Michalewicz, 1997). In a simple way, solutions can be represented by binary vectors. Such a model is a judicious method for the problems in which solutions can be reduced to a sequence of zeros and ones. Unfortunately, we are not able to use such a representation for solving real-world problems like planning and scheduling problems (Caux, Pierreval & Portmann, 1995). Corrective procedure should be conceived because some illegal configurations can obviously be observed if we use basic genetic operators (crossover or mutation). The literature presents many different encodings for a large set of problems. We can mainly distinguish two types of codings. The first one is the direct representation in which the chromosome represents the solution itself. Such an approach requires conceiving specific genetic operators. The second one is the indirect representation in which the chromosome does not directly represent a solution, and a transition from the chromosome representation to a feasible solution is needed.

Concerning the application of the genetic algorithms to the planning problems, the literature shows a large set of works. Some of them have been successfully applied for solving the flexible job shop scheduling problem. We aim in this article to present most recent of them.

1. PMR (Parallel Machine Representation) (Mesghouni, 1999)
   This coding represents the extension of Kobayashi's representation initially proposed for the classical job shop problem. Therefore, it offers the possibility of considering the assignment property. In fact, each cell is coded by three elements: the operation index (*i*), the corresponding job *j*, and starting time $t_{i,j}$ of operation $O_{i,j}$ on the corresponding machine. This representation can directly describe legal solutions and give all the information to execute the schedule. Unfortunately, it presents some difficulties concerning the generation of some illegal cyclic configurations. To overcome such an inconvenience, Mesghouni has proposed some corrective procedures. Unfortunately, such procedures imply an important cost in terms of computation time and therefore reduce the coding effectiveness.
2. PJsR (Parallel Jobs Representation) (Mesghouni, 1999)
   Such a representation is an extension of the coding presented by Yamada for the classical job shop problem. Its particularity consists of the fact that resources can be reassigned to the different tasks (resource flexibility property). The chromosome consists of a list of jobs. Each job is coded by a row in which each case is represented by a couple of terms. The first term indicates the machine assigned to the operation. The second term is the corresponding starting time at which the above operation will be started.
   This encoding allows us to obtain feasible solutions without illegal configuration by integrating the precedence constraints. Genetic operators are very simple and can easily be implemented. Unfortunately, such a coding has a reduced search capacity by comparing it to other possible encodings. Crossover and mutation operators have been proposed by Mesghouni for the two preceding representations. Despite their simplicity and fastness, they are completely based on the exchange of assignment choices and are not sufficient to take into account the sequencing property.
3. Ternary Permutation Matrix with Assignment Vector (Portmann, 1996)
   This coding was proposed by Portmann initially to consider the precedence constraints for a large range of scheduling problems (Portmann 1996). In fact such a model consists of describing the sequencing property by a matrix *MT* noted "Permutation matrix" such that $MT(i,j)=1$ if *i* precedes *j*, $MT(i,j)=-1$ if *j* precedes *i*, and $MT(i,i)=0$ for every task *i* (see Figure 2). Such a coding allows us to keep the good properties in the generated schedule. In order to make it efficient, Portmann has proposed well-adapted operators that can generate feasible schedules with any corrective procedure (Portmann, 1996). Unfortunately, such a coding is insufficient to deal with the flexible job shop problem because of the assignment property. Nevertheless, Portmann has proposed to associate an assignment vector to the ternary coding. Such a proposition allows us to handle scheduling problems with assignment property but, unfortunately, dissociates their two independent parts.
4. Operations Machines Coding (OMC) (Kacem, Hammadi & Borne, 2003)
   This is a direct representation of the schedule. It allows us to obtain all the information on the assignment choices and the different starting times. In addition, it allows us to integrate the schemata notion (Kacem et al., 2003). By such a coding, a schedule is represented in a table. Each case of the table can have one of two value possibilities: 0 (if the operation is not assigned to the machine considered) or [$t_{i,j}, tf_{i,j}$] (with $t_{i,j}$ and $tf_{i,j}$ respectively the starting time and the completion time of operation $O_{i,j}$ on the chosen machine). An illustrative example is presented in Figure 3(a).
5. List Operations Coding (LOC) (Kacem et al., 2003)
   This consists of representing the schedule in a table with three columns. The first column is reserved for the operations . The second indicates the assigned machine to execute the corresponding operation, and the third column gives the starting time and the completion time. We can notice a great similarity between OMC and LOC. In fact, exploration assignment and sequencing search spaces have the same size. The only difference consists of the representation form. Such a difference gives more simplicity and more exploration possibilities (vertical crossover) for OMC. An illustrative example is presented in Figure 3(b).
6. Jobs Sequencings List Coding (JSLC) (Kacem et al., 2003)
   Although, it is relatively difficult to be designed and difficult to be implemented, this encoding represents an efficient representation. In fact, it presents the same possibilities of the exploration of the assignment space search and offers more possibilities to explore the sequencing of one compared to the two preceding representations. It enables us to consider jointly or separately the assignment and the scheduling problems, and avoid the limited use of the priority rules. This coding is presented in a list of columns. Each column contains the different

Figure 1. Mesghouni's coding

M_1	$(i, J_j, t_{i,j,1})$	...
M_2	...	$(i', J_{j'}, t_{i',j',2})$
...		
M_m	...	...

(a) PMR coding

	$(M_2, t_{1,1})$	$(M_4, t_{2,1})$	...
J_2	$(M_6, t_{1,2})$	$(M_2, t_{2,2})$	$(M_1, t_{3,2})$
...			
J_n	...	...	

(b) CPOF coding

Figure 2. Portmann's encoding

	Task 1	Task 2	Task 3	Task 4	Task 5
Task 1	0	1	-1	1	-1
Task 2	-1	0	-1	1	-1
Task 3	1	1	0	1	1
Task 4	-1	-1	-1	0	-1
Task 5	1	1	-1	1	0

operations having the same index in the jobs. Each cell represents a quadruplet [j, k, ti, j, tfi, j] with j, k, ti, j and tfi, j are, respectively, the job index of the i-th operation, the machine index, the starting time, and the completion time. An illustrative example is presented in Figure 3(c).

7. Others Models and Special Cases Application
Despite the difficulty of the scheduling field, the literature is abundant in the interesting references dealing with such a problem. In particular, for some special cases of FJSP (like classical job shop problem, flow shop or one-machine problem), researchers have proposed many interesting encodings. As illustration, we cite Yamada, Tamaki, and Kobayashi's encodings for the job shop problem. Some encodings like "permutation coding" or "rank coding" seem interesting for the one-machine problem. For the flow shop problem, authors have essentially proposed linear representation. In addition, we note an intensive effort in order to conceive efficient operators. As example, we can cite the excellent work of Lee and Yamakawa (1998), who have proposed new operators preserving precedence properties. Such operators are very interesting to construct feasible solutions for scheduling problems with precedence constraints with any corrective step.

Figure 3. Kacem's encodings

		M1	M2	M3	M4
J 1	O1,1	0,1	0	0	0
	O2,1	0	0	0	1,3
	O3,1	3,7	0	0	0
J 2	O1,2	0	0	0,1	0
	O2,2	1,3	0	0	0
	O3,2	0	0	3,5	0
J 3	O1,3	0	0,4	0	0
	O2,3	0	0	0	4,6

(a) OMC encoding

$O_{i,j}$	M_k	$t_{i,j}, tf_{i,j}$
O 1,1	1	0, 1
O 2,1	4	1, 3
O 3,1	1	3, 7
O 1,2	3	0, 1
O 2,2	1	1, 3
O 3,2	3	3, 5
O 1,3	2	0, 4
O 2,3	4	4, 6

(b) LOC encoding

Task 1	Task 2	Task i	...	Task nmax
1, 1, 0, 1	1, 4, 1,3			2, 3, 3, 5
3, 2, 0, 4	2, 1, 1,3	$j, k, t_{i,j}, tf_{i,j}$		1, 1, 3, 7
2, 3, 0, 1	3, 4, 4, 6			

(c) JSLC coding

CONCLUSION

As one of most general planning problems, the flexible job shop scheduling one is very hard to solve. Exact methods are limited and are not suitable for such a research field. That is why many researchers have investigated its resolution by using meta-heuristic techniques as genetic algorithms.

The genetic algorithm's literature presents many resolution models based on a direct or an indirect description of possible solutions. Throughout the different encodings presented in this article, we notice the FJSP specificities are very hard to be completely integrated into one model.

In fact, the problem implies two main considerations¾the assignment resource strategy, and how to find the best sequence of tasks on the chosen resources. Nevertheless, the literature shows an intensive and interesting effort to deal with such a problem. Many encodings are proposed and current results are uncourageous and promising.

REFERENCES

Carlier, J. & Chretienne, P. (1988). *Problèmes d'ordonnancement: Modélisation/complexite/algorithmes*. France: Editions Masson.

Caux, C., Pierreval, H. & Portmann, M.C. (1995). Les algorithmes génétiques et leurs application aux problèmes d'ordonnancement. *RAIRO-APII (Automatique Productique et Informatique Industrielle), 29*(4-5), 409-443.

Dasgupta, D. & Michalewicz, Z. (1997). *Evolutionary algorithms in engineering applications*. Berlin: Springer-Verlag.

Dauzère-Pérès, S., Roux, W. & Lasserre, J.B. (1998). Multi-resource shop scheduling problem with resource flexibility. *EJOR, 107*, 289-305.

Golver, F., Taillard, E. & De Werra, D. (1993). A user's guide to taboo search. *Annals of Operations Research, 41*, 3-28.

Kacem, I., Hammadi, S. & Borne, P. (2003). Flexible job shop scheduling problems: Formulation, lower-bounds, encoding and controlled evolutionary approach. *Computational Intelligence in Control* (pp. 233-261). Hershey, PA: Idea Group Publishing.

Kirkpatrick, S., Gelatt, C.D. & Vecchi, M.P. (1983). Optimization by simulated annealing. *Science, 220*(4598), 671-680.

Kobayashi, S., Ono, I. & Yamamura, M. (1995). An efficient genetic algorithm for job shop scheduling problems. *Proceedings of the 6th International Conference on Genetic Algorithms* (pp. 506-511).

Mastrolilli, M. & Gambardella, L.M. (2000). Effective neighborhood functions for the flexible job shop problem. *Journal of Scheduling, 3*(1), 3-20.

Mesghouni, K. (1999). *Application des algorithmes évolutionnistes dans les problèmes d'optimisation en ordonnancement de la production*. PhD Thesis, Lille 1 University, France.

Lee, K.M. & Yamakawa, T. (1998). A genetic algorithm for general machine scheduling problems. *Proceedings of the International Conference on Conventional and Knowledge-Based Electronics Systems* (KES'98) (Volume 2, pp. 60-66), Australia.

Portmann, M-C. (1996, September 9-11). Genetic algorithms and scheduling: A state of the art and some proposition. *Proceedings of the Workshop on Production Planning and Control* (pp. i-xxiv), Mons, Belgium.

Sarker, R., Abbas, H.A. & Newton, C. (2001, July 9-11). Solving multi-objective optimization problems using evolutionary algorithm. *Proceedings of the International CIMCA Conference*, Las Vegas, Nevada, USA.

Tamaki, H. (1992). Maintenance of diversity in a genetic algorithm and an application to the job shop scheduling. *Proceedings of the IMACS/SICE International Symposium* (pp. 869-869).

Yamada, T. & Nakno, R. (1992). A genetic algorithm applicable to large-scale job shop problem. In R. Manner & B. Manderick (Eds.), *Paralelle problem solving from nature* (Volume 2, pp. 281-290). Amsterdam: North Holland.

KEY TERMS

Assignment: A repartition in which we allocate the available resources to achieve the different tasks.

Binary Coding: A basic coding in which a solution of some problems can be represented in a list of zeros and ones.

Coding: A manner in which a solution of a given problem can be represented.

Flexible Job Shop: A shop in which resources are flexible and can execute many types of tasks with variable performances according to the assignment choices.

Genetic Algorithm: A field of algorithms inspired by the evolution of species and applying natural operators like selection, crossover, and mutation.

Scheduling: A research field in which we aim to program the execution of a set of tasks on a set of resources. The schedule obtained should respect some constraints and optimize one or many criteria.

Formal Development of Reactive Agent-Based Systems

P. Kefalas
CITY College, Greece

M. Holcombe
University of Sheffield, UK

G. Eleftherakis
CITY College, Greece

M. Gheorghe
University of Sheffield, UK

INTRODUCTION

Recent advances in both the testing and verification of software based on formal specifications have reached a point where the ideas can be applied in a powerful way in the design of agent-based systems. The software engineering research has highlighted a number of important issues: the importance of the type of modelling technique used; the careful design of the model to enable powerful testing techniques to be used; the automated verification of the behavioural properties of the system; and the need to provide a mechanism for translating the formal models into executable software in a simple and transparent way.

An agent is an encapsulated computer system that is situated in some environment and that is capable of flexible, autonomous action in that environment in order to meet its design objectives (Jennings, 2000). There are two fundamental concepts associated with any dynamic or reactive system (Holcombe & Ipate, 1998): the environment, which could be precisely or ill-specified or even completely unknown and the agent that will be responding to environmental changes by changing its basic parameters and possibly affecting the environment as well. Agents, as highly dynamic systems, are concerned with three essential factors: a set of appropriate environmental stimuli or inputs, a set of internal states of the agent, and a rule that relates the two above and determines what the agent state will change to if a particular input arrives while the agent is in a particular state.

One of the challenges that emerges in intelligent agent engineering is to develop agent models and agent implementations that are "correct." The criteria for "correctness" are (Ipate & Holcombe, 1998): the initial agent model should match the requirements, the agent model should satisfy any necessary properties in order to meet its design objectives, and the implementation should pass all tests constructed using a complete functional test-generation method. All the above criteria are closely related to stages of agent system development, i.e., modelling, validation, verification, and testing.

BACKGROUND: FORMAL METHODS AND AGENT-BASED SYSTEMS

Although agent-oriented software engineering aims to manage the inherent complexity of software systems (Wooldridge & Ciancarini, 2001; Jennings, 2001), there is still no evidence to suggest that any methodology proposed leads toward "correct" systems. In the last few decades, there has been strong debate on whether formal methods can achieve this goal. Software system specification has centred on the use of models of data types, either functional or relational models, such as Z (Spivey, 1989) or VDM (Jones, 1990), or axiomatic ones, such as OBJ (Futatsugi et al., 1985). Although these have led to some considerable advances in software design, they lack the ability to express the dynamics of the system. Also, transforming an implicit formal description into an effective working system is not straightforward. Other formal methods, such as finite state machines (Wulf et al., 1981) or Petri Nets (Reisig, 1985) capture the essential feature, which is "change," but fail to describe the system completely, because there is little or no reference to the internal data and how these data are affected by each operation in the state transition diagram. Other methods, like statecharts (Harel 1987), capture the requirements of dynamic behaviour and modelling of data but are informal with respect to clarity and semantics. So far, little attention has been paid in formal methods that could facilitate all crucial stages of "correct" system development, modelling, verification, and testing.

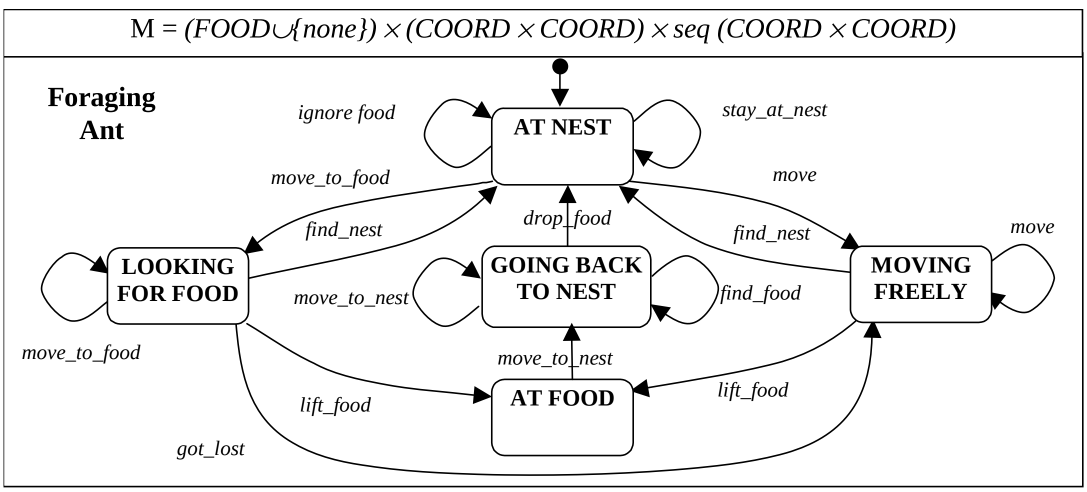
In agent-oriented engineering, there have been several attempts to use formal methods, each one focusing on different aspects of agent systems development. One was to formalise the PRS (procedural reasoning system), a variant of the BDI architecture (Rao & Georgeff, 1995), with the use of Z, in order to understand the architecture in a better way, to be able to move to the implementation through refinement of the specification, and to be able to develop proof theories for the architecture (D'Inverno et al., 1998). Trying to capture the dynamics of an agent system, Rosenschein and Kaebling (1995) viewed an agent as a situated automaton that generates a mapping from inputs to outputs, mediated by its internal state. Brazier et al. (1995) developed the DESIRE framework, which focuses on the specification of the dynamics of the reasoning and acting behaviour of multiagent systems. In an attempt to verify whether properties of agent models are true, work has been done on model checking of multiagent systems with reuse of existing technology and tools (Benerecetti et al., 1999, Rao & Georgeff, 1993). Toward implementation of agent systems, Attoui and Hasbani (1997) focused on program generation of reactive systems through a formal transformation process. A wider approach is taken by Fisher and Wooldridge (1997), who utilised Concurrent METATEM in order to formally specify multiagent systems and then directly execute the specification while verifying important temporal properties of the system. Finally, in a less formal approach, extensions to Unified Modelling Language (UML) to accommodate the distinctive requirements of agents (AUML) were proposed (Odell et al., 2000).

X-MACHINES FOR AGENT-BASED SYSTEM DEVELOPMENT

An X-machine is a general computational machine (Eilenberg, 1974) that resembles a finite state machine but with two significant differences: there is memory attached to the machine, and the transitions are labeled with functions that operate on inputs and memory values. The X-machine formal method forms the basis for a specification language with great potential value to software engineers, because they can facilitate modelling of agents that demand remembering as well as reactiveness. Figure 1 shows the model of an ant-like agent that searches for food but also remembers food positions in order to set up its next goals. Many other biological processes seem to behave like agents, as, for example, a colony of foraging bees, tissue cells, etc. (Kefalas et al., 2003a; Gheorghe et al., 2001; Kefalas et al., 2003b). Formally, the definition of the X-machine requires the complete description of a set of inputs, outputs, and states; a memory tuple with typed elements; a set of functions and transitions; and finally, an initial state and a memory value (Holcombe, 1988).

Having constructed a model of an agent as an X-machine, it is possible to apply existing model-checking techniques to verify its properties. *CTL** is extended with memory quantifier operators: M_x (for all memory instances) and m_x (there exist memory instances) (Eleftherakis & Kefalas, 2001). For example, in the ant-like agent, model checking can verify whether food will eventually be dropped in the nest by the formula: $AG[\neg M_x(m_1 \neq \text{none}) \vee EFM_x(m_1 = \text{none})]$, where m_1 indicates the first element of the memory tuple.

Figure 1. An X-machine that models an ant.



Having ensured that the model is "correct," we need to also ensure that the implementation is "correct," this time with respect to the model. Holcombe and Ipate (1998) presented a testing method that under certain design-for-test conditions can provide a complete test-case set for the implementation. The testing process can be performed automatically by checking whether the output sequences produced by the implementation are identical to the ones expected from the agent model through this test-case set.

A methodology for building complex agent systems by aggregating a set of behaviors of individual agents is available, namely, communicating X-machines. It is demonstrated that they are a powerful extension to the X-machines that also facilitate modelling of multiagent systems (Kefalas et al., 2003a).

FUTURE TRENDS

There are currently three directions to future work. First, there is a need to investigate potential applications. We already identified the area of biology-inspired agent-based systems, such as modelling of biological cells and tissues, as an area that can largely benefit from formal modelling, verification, and testing. Second, current research is underway to extend the testing and verification methods in order to be applicable for communicating asynchronous and possibly nondeterministic systems. Last, there has been an attempt to build tools around the X-machine formal method in order to facilitate the actual agent development process. A markup language, namely X-Machine Definition Language (XMDL), has been defined, and around it, a number of prototype tools, such as a modeller, an automatic translator to Prolog, an animator, and a model checker, have been constructed.

CONCLUSION

Because the X-machine method is fully grounded in the theory of computation, it is fully general and will be applicable to any type of computational task. The paradigm of the X-machine is also convenient when it comes to implementing the models in an imperative programming language. In fact, the translation is more or less automatic. The existence of the powerful testing method described lays the foundation for the method to be used in potentially critical applications. Finally, the model-checking developments will lead to a situation in which one of the key issues in agent software engineering can be solved, namely, how can we guarantee that the agent system constructed will exhibit the desired emergent behavior, or at least substantial progress toward this goal will be achieved.

REFERENCES

Attoui, A., & Hasbani, A. (1997). Reactive systems developing by formal specification transformations. In *Proceedings of the Eighth International Workshop on Database and Expert Systems Applications (DEXA '97)*, (pp.339-344). Washington, DC: IEEE Computer Society.

Benerecetti, M., Giunchiglia, F., & Serafini, L. (1999). A model checking algorithm for multiagent systems. In J. P. Muller, M. P. Singh, & A. S. Rao (Eds.), *Intelligent Agents V* (LNAI Vol. 1555) (pp.163-176). Heidelberg: Springer-Verlag.

Brazier, F., Dunin-Keplicz, B., Jennings, N., & Treur, J. (1995). Formal specification of multi-agent systems: A real-world case. In V. Lesser (Ed.), *Proceedings of International Conference on Multi-Agent Systems (ICMAS'95)* (pp. 25-32), San Francisco, CA, June 12-14. Cambridge, MA: MIT Press.

Eilenberg, S. (1974). *Automata, machines and languages* (Vol. A). New York: Academic Press.

Eleftherakis, G., & Kefalas, P. (2001). Towards model checking of finite state machines extended with memory through refinement. In G. Antoniou, N. Mastorakis, & O. Panfilov (Eds.), *Advances in signal processing and computer technologies* (pp. 321-326). Singapore: World Scientific and Engineering Society Press.

Fisher, M., & Wooldridge, M. (1997). On the formal specification and verification of multi-agent systems. *International Journal of Cooperating Information Systems, 6*(1), 37-65.

Futatsugi, K., Goguen, J., Jouannaud, J. -P., & Meseguer, J. (1985). Principles of OBJ2. In B. Reid (Ed.), *Proceedings, Twelfth ACM Symposium on Principles of Programming Languages* (pp. 52-66). Association for Computing Machinery.

Georghe, M., Holcombe, M., & Kefalas, P. (2001). Computational models for collective foraging. *Biosystems, 61*, 133-141.

Harel, D. (1987). Statecharts: A visual approach to complex systems. *Science of Computer Programming, 8*(3).

Holcombe, M. (1988). X-machines as a basis for dynamic system specification. *Software Engineering Journal, 3*(2), 69-76.

Holcombe, M., & Ipate, F. (1998). *Correct systems: Building a business process solution*. Heidelberg: Springer-Verlag.

Inverno, d' M., Kinny, D., Luck, M., & Wooldridge, M. (1998). A formal specification of dMARS. In M. P. Singh, A. Rao, & M. J. Wooldridge (Eds.), *Intelligent Agents IV* (LNAI Vol. 1365) (pp. 155-176). Heidelberg: Springer-Verlag.

Ipate, F., & Holcombe, M. (1998). Specification and testing using generalised machines: A presentation and a case study. *Software Testing, Verification and Reliability, 8*, 61-81.

Jennings, N. R. (2000). On agent-based software engineering. *Artificial Intelligence, 117*, 277-296.

Jennings, N. R. (2001). An agent-based approach for building complex software systems. *Communications of the ACM, 44*(4), 35-41.

Jones, C. B. (1990). Systematic software development using VDM (2nd ed.). New York: Prentice Hall.

Kefalas, P., Eleftherakis, G., & Kehris, E. (2003a). Communicating X-Machines: From theory to practice. In Y. Manolopoulos, S. Evripidou, & A. Kakas (Eds.), *Lecture notes in computer science* (Vol. 2563) (pp. 316-335). Heidelberg: Springer-Verlag.

Kefalas, P., Eleftherakis, G., Holcombe, M., & Gheorghe, M. (2003b). Simulation and verification of P systems through communicating X-machines. *BioSystems, 70*(2), 135-148.

Odell, J., Parunak, H.V.D., & Bauer, B. (2000). Extending UML for agents. In G. Wagner, Y. Lesperance, & E. Yu (Eds.), *Proceedings of the Agent-Oriented Information Systems Workshop at the 17th National Conference on Artificial Intelligence*, (pp. 3-17), Austin, TX.

Rao, A.S., & Georgeff, M. P. (1993). A model-theoretic approach to the verification of situated reasoning systems. In R. Bajcsy (Ed.), *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence (IJCAI'93)* (pp. 318–324). San Francisco, CA: Morgan Kaufmann.

Rao, A.S., & Georgeff, M. (1995). BDI agents: From theory to practice. In V. Lesser & L. Gasser (Eds.), *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS'95)* (pp. 312-319), MIT Press.

Reisig, W. (1985). Petri nets—An introduction. In *EATCS Monographs on Theoretical Computer Science, Vol. 4*. Heidelberg: Springer-Verlag.

Rosenschein, S. R., & Kaebling, L. P. (1995). A situated view of representation and control. *Artificial Intelligence, 73*(1-2), 149-173.

Spivey, M. (1989). *The Z notation: A reference manual*. New York: Prentice Hall.

Wooldridge, M., & Ciancarini, P. (2001). Agent-oriented software engineering: The state of the art. To appear in the S.K. Chang, *Handbook of Software Engineering and Knowledge Engineering*. Singapore: World Scientific Publishing.

Wulf, W. A., Shaw, M., Hilfinger, P. N., & Flon, L. (1981). *Fundamental structures of computer science*. Reading, MA: Addison-Wesley.

KEY TERMS

Agent: An agent is an encapsulated computer system that is situated in some environment and that is capable of flexible, autonomous action in that environment in order to meet its design objectives. Agents normally exhibit autonomous, reactive, proactive, and social behaviors.

Communicating X-Machine: A communicating X-machine is a set of stream X-machine components that are able to communicate with each other by exchanging messages.

CTL*: A temporal logic formalism used in model checking. *CTL** employs operators, such as *A* (for all paths), *E* (there exists a path), *X* (next time), *F* (eventually), *G* (always), *U* (until), and *R* (release), that facilitate the construction of temporal logic formulas that correspond to desirable properties of a model.

Formal Methods: Formal methods are rigorous techniques based on mathematical notation that can be used to specify and verify software models.

Model Checking: Model checking is a formal verification technique that determines whether given properties of a system are satisfied by a model. A model checker takes a model and a property as inputs, and outputs either a claim that the property is true or a counterexample falsifying the property.

X-Machine: A deterministic stream X-machine is an 8-tuple ($\boldsymbol{\Sigma}$, Γ, Q, M, Φ, F, q_0, m_0), where $\boldsymbol{\Sigma}$, Γ is the input and output finite alphabet, respectively; Q is the finite set of states; M is the (possibly) infinite set called memory; Φ is the type of the machine, that is, a finite set of partial functions φ that map an input and a memory state to an output and a new memory state, $\varphi: \boldsymbol{\Sigma} \times M \rightarrow \Gamma \times M$; F is the next state partial function that, given a state and a function from the type Φ, denotes the next state—*F* is often described as a state transition diagram, $F: Q \times \Phi \rightarrow Q$; q_0 and m_0 are the initial state and memory, respectively.

Formal Methods in Software Engineering

Aristides Dasso
Universidad Nacional de San Luis, Argentina

Ana Funes
Universidad Nacional de San Luis, Argentina

INTRODUCTION

As a general rule, all engineering applications use mathematics or mathematical tools as a basis for their development. However, software engineering is an exception to this rule.

Formal methods (FM) are[1] a collection of methodologies and related tools, geared to the production of software employing a mathematical basis. There are a number of different formal methods each having its own methodology and tools, especially a specification language.

We can say that FM are "mathematically based techniques for the specification, development and verification of software and hardware systems" (retrieved on October 15, 2003, http://foldoc.doc.ic.ac.uk/foldoc).

Most FM are based mainly on specifications – for which they normally have a language to express it. Sometimes, there is also a method to use the language in the software development process.

The aims of FM can vary according to the different methodologies, but they all shared a common goal: the production of software with the utmost quality mainly based on the production of software that is error free. To achieve this, the different FM have developed not only a theory but also different tools to support the formal process.

FM can cover all of the steps of the life cycle of a software system development from requirement specification to deployment and maintenance. However, not all FM have that capacity.

BACKGROUND

Some FM rely on development of a calculus or transformation, where the engineer starts with an expression and then following predefined rules applies them to obtain an equivalent expression. Successive calculations lead to implementation. On the other hand, there are FM that rely on the "invent and verify" technique, where the engineer starts by inventing a new design, which afterwards needs to be verified as correct. From this verified design, implementation follows.

There are several styles of formal specification. Some are mutually compatible while others are not. Table 1 shows a possible classification of the different styles.

Formal languages have formal definitions not only of their syntax but also of their semantics. Table 2 shows a possible classification for the different formal semantic definitions.

Hinchey and Bowen (1995) say that "formal methods allow us to propose properties of the system and to demonstrate that they hold. They make it possible for us to examine system behavior and to convince ourselves that all possibilities have been anticipated. Finally, they enable us to prove the conformance of an implementation with its specification".

In NASA's Langley Research Center site for formal methods, there is a nice definition and also an explanation of different degrees of rigour in FM:

"Traditional engineering disciplines rely heavily on mathematical models and calculation to make judgments

Table 1. Summary of specification language characteristics

Model-oriented. Based on mathematical domains. For example, numbers, functions, sets, etc. Concrete.	**Property-oriented**. Based on axiomatic definitions. Abstract.
Applicative. Does not allow the use of variables.	**Imperative or State-oriented**. Allows the use of variables.
Static. Do not include provisions for handling time.	**Action**. Time can be considered in the specification. There are several ways of doing this: considering time as linear or branching, synchronous, asynchronous, etc.

Table 2. Summary of semantic definitions styles of specification languages

Operational. Concrete, not well suited for proofs.
Denotational. Abstract, well suited for proofs.
Axiomatic. Very abstract, normally only limited to conditional equations.

about designs. For example, aeronautical engineers make extensive use of computational fluid dynamics (CFD) to calculate and predict how particular airframe designs will behave in flight. We use the term 'formal methods' to refer to the variety of mathematical modelling techniques that are applicable to computer system (software and hardware) design. That is, formal methods is the applied mathematics of computer system engineering, and, when properly applied, can serve a role in computer system design analogous to the role CFD serves in aeronautical design.

Formal methods may [be] used to specify and model the behavior of a system and to mathematically verify that the system design and implementation satisfy system functional and safety properties. These specifications, models, and verifications may be done using a variety of techniques and with various degrees of rigour. The following is an imperfect, but useful, taxonomy of the degrees of rigour in formal methods:

Level-1:
Formal specification of all or part of the system.
Level-2:
Formal specification at two or more levels of abstraction and paper and pencil proofs that the detailed specification implies the more abstract specification.
Level-3:
Formal proofs checked by a mechanical theorem prover.

Level 1 represents the use of mathematical logic or a specification language that has a formal semantics to specify the system. This can be done at several levels of abstraction. For example, one level might enumerate the required abstract properties of the system, while another level describes an implementation that is algorithmic in style.

Level 2 formal methods goes beyond Level 1 by developing pencil-and-paper proofs that the more concrete levels logically imply the more abstract-property oriented levels. This is usually done in the manner illustrated below.

Level 3 is the most rigourous application of formal methods. Here one uses a semi-automatic theorem prover to make sure that all of the proofs are valid. The Level 3 process of convincing a mechanical prover is really a process of developing an argument for an ultimate skeptic who must be shown every detail.

Formal methods is not an all-or-nothing approach. The application of formal methods to only the most critical portions of a system is a pragmatic and useful strategy. Although a complete formal verification of a large complex system is impractical at this time, a great increase in confidence in the system can be obtained by the use of formal methods at key locations in the system. (Retrieved October 31, 2003, http://shemesh.larc.nasa.gov/fm/fm-what.html)

NASA as well as other government bodies in the USA, Europe and elsewhere are using FM especially in avionics and systems where the utmost reliability is needed. Some examples from NASA are: Small Aircraft Transportation System (SATS), Formal Analysis of Airborne Information for Lateral Spacing (AILS), also NASA's contractors use FM. For more information on this and other projects, see http://shemesh.larc.nasa.gov/fm/.

As it is stated previously in NASA's definition of the levels of degree of rigour, they are imperfect. Others exist. Most of the FM included in the following text have as an integral part not only a language but also a methodology included, and most of the time this methodology implies different levels of rigour in its use. For example, RAISE – that have its own method – presents three degrees of formality (The RAISE Method Group, 1995):

- *formal specification only*, where formality is only applied to the specification procedure.
- *formal specification and rigourous development*, where formality is applied to the specification procedure as above, and rigour to the development process. This means that the developer starts writing abstract specifications, goes on developing more concrete ones and recording the development relations between them. These relations are then examined, however they are not justified.
- *formal specification and formal development*, it is the extension of the previous degree to do the justification as well.

Here is a not all-inclusive list of FM:

- **ASM** (Abstract State Machines) "methodology for describing simple abstract machines which correspond to algorithms" (Retrieved September 15, 2003, from http://www.eecs.umich.edu/gasm/).
- **B–Method** "B is a formal method for the development of program code from a specification in the Abstract Machine Notation" (Retrieved October 23, 2003, from http://www.afm.lsbu.ac.uk/b/).
- **CSP** (Communicating Sequential Processes) " process algebra originated by C. A. R. Hoare (http://www.afm.lsbu.ac.uk/csp/).

- **Duration Calculus** "modal logic for describing and reasoning about the real-time behavior of dynamic systems" (Retrieved October 20, 2003, from http://www.iist.unu.edu/newrh/II/2/2/page.html).
- **Extended ML** "framework for specification and formal development of Standard ML (SML) programs" (Retrieved October 20, 2003, from http://www.dcs.ed.ac.uk/home/dts/eml).
- **HOL** " automatic theorem proving system based on Higher Order Logic (see http://www.afm.lsbu.ac.uk/hol/).
- **Model Checking** "method for formally verifying finite-state concurrent systems" (Retrieved October 21, 2003, from http://www-2.cs.cmu.edu/~modelcheck/).
- **Petri Nets** "formal, graphical, executable technique for the specification and analysis of concurrent, discrete-event dynamic systems" (Retrieved October 21, 2003, from http://www.petrinets.org/).
- **PVS** (Prototype Verification System) "a specification language integrated with support tools and a theorem prover" (Retrieved October 19, 2003, from http://pvs.csl.sri.com/whatispvs.html).
- **RAISE** (Rigourous Approach to Industrial Software Engineering) "consists of the RAISE development method and RSL, the RAISE Specification Language" (Retrieved November 1, 2003, from http://spd-web.terma.com/Projects/RAISE; See also http://www.iist.unu.edu/raise/).
- **VDM** (Vienna Development Method) "VDM is a notation and set of techniques for modeling computing systems analyzing those models and progressing to detailed design and coding. VDM has its origins in the work of the IBM Vienna Laboratory in the mid-1970s" (Retrieved October 14, 2003, from http://www.csr.ncl.ac.uk/vdm).
- **Z Notation** "is based on Zermelo-Fraenkel set theory and first order predicate logic" (Retrieved October 14, 2003, from http://www.zuser.org/z).

For more information on those and other FM, see http://vl.fmnet.info. This site has links to the sites of many other FM.

Of course, a question to ask oneself is, why are there so many flavors of formal methods? There are undoubtedly several possible answers to the question. One is that since there are many ways to describe a system not everyone agrees on a particular style. Why are there so many different programming languages? The answer to this question could be also the answer to our question. Another answer is that not all of the FM have addressed the same problem. Some of them are geared to system design, others to domain description, some deal with time while others do not included, and so forth.

PROS AND CONS

The issue of whether FM are useful or not was discussed in the literature around ten years ago – see for example, J. Bowen and M. Hinchey´s works (1995) and Luqi and Goguen (1997).

The use of formal specifications has benefits ranging from the possibility of building unambiguous specifications, to the possibility of proving system properties, to automatic code generation. However, they require a high level of expertise in algebra and mathematical logic, and consequently they are more expensive and difficult to use than other traditional methods.

Most opponents give as a main reason the fact that FM make the communication with the end users difficult when requirements must be validated. However, lately there is an ongoing interest in the combined use of semi-formal notations and formal specifications to overcome this problem and the ambiguity and inconsistency inherent in semi-formal notations. This kind of integration is intended to improve understandability – given by graphical notations – and to gain in unambiguity – offered by formal specifications.

According to Pons and Baum (2000), we can carry out integration in four different ways:

1. Supplemental, where the informal notations are enriched with formal concepts.
2. Extension, where the formal notations are extended with concepts from others paradigms, for example, from the object-oriented paradigm.
3. Interface, where formal notations are provided with graphic interfaces to help in the developing of models.
4. Semantics, where the semantics of an accepted semi-formal modelling language is given by a formal language.

There is an important number of theoretical works that deal with the integration of graphical notations and mathematically precise formalisms. Proof of this is the growing interest in providing a more traditional methodology such as UML with formal basis, either through the OCL language or by giving to the UML graphic language a formal semantics – see the works of the 2U consortium (2003) and The precise UML group (2003). The extensive work done on formalizing Java and Java Machine (see Bertelsen, 2003; Hartel & Moreau, 2001). Other good examples can be found in Funes and George (2003), France (1999), DeLoach and Hartrum (2000), Meyer and Souquieres (1999), Goldsack and Kent (1996), Lano (1991), Weber (1996), Kim and Carrington (2000), Amalio et al. (2003), and Reggio and Larosa (1997) among others.

Finally, we should not forget the important role FM have in the development of systems where security and reliability are crucial. Dan Van et al. (2002) cite three factors that were the traditional arguments in favor of FM: first, the growing number of applications with minimal or zero tolerance for errors; second, social, environmental and economic consequences of design errors in hardware and software products, raising social awareness and concern; and third, there are a number of standards and regulatory organizations enforcing quality standards for electronic products.

EXAMPLES

Just so the reader can get the flavor of FM, we present in this section an example of specifications for a stack of natural numbers using the RAISE Specification Language (RSL) (The RAISE Language Group, 1992).

The first specification in Example 1 illustrates the use of the property-oriented style, which is abstract and whose definitions are given by axioms. The second specification given for the stack is model-oriented, based in the use of concrete types (see Example 2). Both specifications use RSL static and applicative style.

The *Stack* type in Example 1 is given by an abstract type since we do not say explicitly how a stack is going to be implemented. The stack is defined by the main operations needed to manage it. To formally specify its behavior, we give the signature of three functions: *pop*, *push* and *top*, the constant *empty* that represents the empty stack, and a set of axioms to express the intended meaning of these operations. The axioms define the es-

Example 1. A stack of natural numbers in RSL (using static applicative property-oriented style)

```
scheme STACK_0 =
class
   type
      Stack
   value
      empty: Stack,
      push: Nat × Stack → Stack,
      pop: Stack ~→ Stack ,
      top: Stack ~→ Nat
   axiom
      e: Nat, s: Stack • pop(push(e, s)) ≡ s,
      e: Nat, s: Stack • top(push(e, s)) ≡ e,
end
```

Example 2. A stack of natural numbers in RSL (using static applicative model-oriented style)

```
scheme STACK_1 =
class
   type
      Stack = Nat *
   value
      empty: Stack,
      push: Nat × Stack → Stack,
      pop: Stack ~→ Stack ,
      top: Stack ~→ Nat
   axiom
      empty ≡ <>,
      e: Nat, s: Stack • push(e, s) ≡ <e> ^ s,
      s: Stack • pop(s) ≡ tl s
                    pre s ≠empty,
      s: Stack • top(s) ≡ hd s
                    pre s ≠empty
end
```

sential properties that must always be true when the operations are applied.

The operations *pop* and *top* are partial functions that are not defined for the empty stack. This fact is reflected by the absence of axioms about *pop(empty)* and *top(empty)*. The operation *push* is a total function.

The first axiom states that pushing an element *e* into a stack *s* and then doing a *pop* returns the former stack *s*. The second axiom states that every time we push an element *e* into a stack, and then we apply *top* to the resulting stack we get the last pushed element *e*.

But, where is the formalism in this specification? We must not forget that as any FM, RSL has defined formally not only its syntax but also its semantic. This allows, for instance, the use of axioms which are truths that must be preserved all along the specification and that in turn can be proved – or not, if the specification has errors or inconsistencies. So the formalism can be found in the fact that using the underlying semantic – in this case, logic and algebra – the specification can be subjected to proof, and therefore, any errors can be found.

In this example, the axioms specify the behavior of the stack; remember that axioms must always be true. If they are not, then there is an error in the specification.

So, besides the abstractness and precision given by a formal specification language, one of the major reasons for expressing specifications in a formal language is the possibility of proving properties of the specifications. Formulating properties and then trying to prove them is a way of ensuring correctness and detecting errors.

The second example – see Example 2 – is a refinement of the first one. Here, we define the *Stack* type not as a sort but as a list of natural numbers – the * besides **Nat** indicates list of **Nat** in RSL.

The axioms that define the operations on the stack are defined using the primitive operations for lists in RSL. The operation *empty* is defined as the empty list. To *push* an element into the stack corresponds to adding the element at the head of the list, returning the resulting augmented stack. To *pop* an element from the stack corresponds to remove the head of the list and return the resulting reduced stack; and to get the *top* of the stack corresponds to get the head of the list and return it. Note that since *pop* and *top* have a pre-condition establishing that they cannot be applied to the empty stack, they are partial functions.

In RAISE – as in many others FM – besides being formal when specifying, we can also be formal in the development process. Software can be developed in a sequence of steps. We can start writing a suitably abstract specification and proceed developing more concrete and detailed ones until the final specification can be automatically translated to a programming language. Each step in the development must conform to the previous and must be proved as a correct development step. For example, we should prove that the specification STACK_1 is a correct development for STACK_0, that is, to prove that STACK_1 implements STACK_0 (i.e., STACK_0 and STACK_1 are in the implementation relation). In RAISE, this means that the specifications have to meet two requirements:

- *property of preservation*: all properties that can be proved about STACK_0 can also be proved for STACK_1.
- *property of substitutivity*: in any specification an instance of STACK_0 can be replaced by an instance of STACK_1, and the resulting specification should implement the former specification.

FUTURE TRENDS

Although traditional methods seem to be the most popular nowadays, the use of FM have grown to occupy more and more a place in software engineering for development of systems where security and reliability are important. However, its use continues being expensive and limited.

Besides its leading role in the development of critical systems, there is also a growing interest in proposing methods and techniques to use formal specifications combined with semi-formal notations to give more rigour to the first phases of the development process.

If formal techniques become, in the future, the foundations of a new generation of CASE tools, it is possible that a bigger portion of software practitioners will adopt them, making formal methods a truly practical tool for all kind of systems.

CONCLUSION

FM are methods based primarily in formal, mathematical notation and principles. They are ideal for defining unambiguously not only the requirements but also every stage in system development including implementation since there are FM that have translators to programming languages.

Because of their profound roots in logic and algebra, they seem harder to understand and learn. However, the FM community aware of this view has been making efforts in producing tools that make their use easier.

REFERENCES

2U consortium (2003). Unambiguous UML. Retrieved November 1, 2003, *http://www.2uworks.org*

Amalio, N. et al. (2003). Modular UML semantics: Interpretations in Z based on Templates and Generics. In *Proceedings of FM 2003*, Sept. 8-14, 2003, Pisa, Italy.

Bertelsen, P. (2003). Semantics of Java byte code. Retrieved November 1, 2003, *ftp://ftp.dina.kvl.dk/pub/Staff/Peter.Bertelsen/jvm-semantics.ps.gz*

Bowen, J., & Hinchey, M. (1995, July). Seven more myths of formal methods. *IEEE Software, 12*(3).

Bowen, J., & Hinchey, M. (1995, April). Ten commandments of formal methods. *IEEE Computer*.

Butler, R. (2003). Singular vs. plural. Retrieved October 15, 2003, from *http://shemesh.larc.nasa.gov/fm/fm-is-vs-are.html*; last modified: September 10, 1998 (15:57:28).

Dan Van, H. et al. (2002). *Specification case studies in RAISE*. Springer.

DeLoach, S., & Hartrum, T. (2000, June). A theory-based representation for object-oriented domain models. *IEEE Transactions on Software Engineering, 6*(6), 500-517.

France, R. (1999). A problem-oriented analysis of basic UML static requirements modeling concepts. In *Proceedings of OOPSLA '99*, Denver, CO, USA.

Funes, A., & George, C. (2003). Chapter 8: Formalizing UML class diagrams. In L. Favre (Ed.), *UML and the Unified Process*. Hershey, PA: Idea Group Publishing.

Goldsack, S., & Kent, S. (1996). Chapter 3: LOTOS in the object-oriented analysis process. *Formal methods and object technology*. Springer.

Hartel, P., & Moreau, L.A.V. (2001, December). Formalizing the safety of Java, the Java virtual machine and Java card. *ACM Computing Surveys, 33*(4), 517-558.

Hinchey, M., & Bowen, J. (1995). *Applications of formal methods*. Prentice Hall International.

Lano, K. (1991). Z++, An object-oriented extension to Z. In J. Nicholls (Ed.), *Z user workshop*, Oxford, 1990, Workshops in Computing, Springer.

Luqi & Goguen, J. (1997, Jan.-Feb.). Formal methods: Promises and problems. *IEEE Software, 14*(1), 73-85.

Kim, S.-K., & Carrington, D. (2000). A formal specification mapping between UML models and object-Z specifications. No. 1878 in LNCS, Springer (pp.2-21).

Meyer, E., & Souquieres, J. (1999). A systematic approach to transform OMT diagrams to a B specification. In *Proceedings of FM '99*, volume I of LNCS (pp.875-895).

Pons, C., & Baum, G. (2000). Formal foundations of object-oriented modeling notations. *Proceedings of ICFEM'00*.

Reggio, G., & Larosa, M. (1997). A graphic notation for formal specification of dynamic systems. In *Proceedings of FME'97*, No. 1313 in LNCS, Springer.

The precise UML group (2003). Retrieved November 1, 2003, *http://www.puml.org/*

The RAISE Language Group (1992). The RAISE specification language. Prentice Hall.

The RAISE Method Group (1995). The RAISE development method. Prentice Hall.

Weber, M. (1996). Combining statecharts and Z for the design of safety-critical control systems. In *Proceedings of the 3rd International Symposium of FME'96*. Oxford, 1996.

KEY TERMS

Action-Oriented Formal Specification Language: Time can be considered in the specification. There are several ways of doing this: considering time as linear or branching, synchronous, asynchronous, and so forth.

Applicative-Oriented Formal Specification Language: Does not allow the use of variables.

Axiomatic Semantics: The meaning is given in terms of conditions, pre and post.

Denotational Semantics: The meaning is given in terms of mathematical functions.

Imperative or State-oriented Formal Specification Language: Allows the use of variables.

Model-Oriented Formal Specification Language: Based on mathematical domains. For example numbers, functions, sets, and so forth. Concrete.

Operational Semantics: The meaning is given in terms of rules that specify how the state of a computer – real or formal – changes while executing a program.

Property-Oriented Formal Specification Language: Based on axiomatic definitions. Abstract.

Semantics: In a language, it is the meaning of a string, as opposed to syntax, which describes how the symbols of the language are combined. Most programming languages have their syntax defined formally (traditionally in BNF), while formal specification languages have also their semantics defined formally.

Specification: A document describing what a system should do, what a problem is or what a domain is all about. In formal methods this document is written in a formal language.

Static-Oriented Formal Specification Language: Do not include provisions for handling time.

Verification: The process of determining whether or not the products of a specification phase fulfill a set of established requirements. Sometimes this is also used to indicate the process of proving that a more concrete specification preserves the properties of a more abstract specification.

ENDNOTE

[1] Some authors use the singular while others employ the plural in referring to FM. We can refer to Ricky W. Butler (2003):
"Some of you are saying, 'Formal Methods is ...? What's wrong with these people, ain't nobody learned them no grammar!' In an age in which few people care about the proper use of language, your concern is commendable; however, in this instance, your concern is also unwarranted."
"In these pages, we are using the term formal methods to refer to a particular collection of knowledge.

Just as the plural-sounding term fluid dynamics is treated as singular, so too may the term formal methods be treated as singular. A legitimate argument can be made as to the acceptability of treating the term as plural, but no legitimate argument can be made as to the necessity of doing so." http://shemesh.larc.nasa.gov/fm/fm-is-vs-are.html; last modified: 10 September 1998 (15:57:28). Retrieved October 15, 2003.

Forward Engineering of UML Static Models [1]

Liliana Favre
CIC, Argentina

Liliana Martinez
Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina

Claudia Pereira
Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina

INTRODUCTION

The Unified Modeling Language (UML) has emerged as a modeling language for specifying, visualizing, constructing, and documenting software-intensive systems. It unifies proven software modeling languages that incorporate the object-oriented community's consensus on core modeling concepts. It also includes additional expressiveness to handle problems that previous visual languages did not fully address (Rumbaugh, Jacobson & Booch, 1999).

UML emerged in response to a call for a standard object-oriented design method by the Object Management Group (OMG) in 1997. In mid-2001, OMG started working on a major upgrade to UML, and its evolution will result in version 2.0 (OMG, 2004).

The UML notation includes diagrams that provide multiple perspectives of the system under development. It defines 12 types of diagrams divided into three categories that represent static application structure, aspects of dynamic behavior, and ways for organizing and managing application modules.

The model's elements are defined in terms of their abstract syntax, well-formed rules, and precise text (OMG, 2004). These well-formed rules are expressed in the Object Constraint Language (OCL). Recently, a new version of OCL, version 2.0, has been formally defined and it has been adopted by OMG (Warmer & Kleppe, 2003).

UML is used in many ways and different domains for expressing different types of concepts such as language-independent software specification, high-level architecture, Web site structure, workflow specification, and business modeling. It has been applied successfully to build systems for different types of applications running on any type and combination of hardware, operating system, programming language, and network.

Although UML does not prescribe any particular development process, the OMG presented the Software Process Engineering Metamodel (SPEM) as a standard in November 2001 (OMG, 2004). This metamodel is used to describe a concrete software development process or a family of related software development processes that use the UML notation. SPEM has a four-layered architecture of modeling for describing performing process, process model, process metamodel, and MetaObject Facility (MOF). Several processes fit SPEM. The most popular is Rational Unified Process, developed and marketed by Rational Software (now a division of IBM). RUP is use-driven, architecture-centered, iterative, and risk-driven. Various industry sectors around the world use RUP in different applications: telecommunications, transportation, aerospace, defense, manufacturing, and financial services (Jacobson, Booch & Rumbaugh, 1999; Krutchen, 2000).

The international standardization of UML leads to improvements in CASE tools, methods, and standard modeling libraries. In the market, there are about 100 UML CASE (computer-aided software engineering) tools that vary widely in functionality, usability, performance, and platforms (CASE, 2004).

A recent OMG initiative is the Model-Driven Architecture (MDA), which promotes the creation of abstract models that are developed independently of a particular implementation technology and automatically transformed by tools into models for specific technologies (Kleppe, Warmer & Bast, 2003).

MDA is emerging as a technical framework to improve productivity, portability, interoperability, and maintenance. It defines how models expressed in one language can be transformed into models in other languages. The MDA process is divided into three main steps:

- Construct a model with a high level of abstraction that is called Platform-Independent Model (PIM).
- Transform the PIM into one or more Platform-Specific Models (PSMs), each one suited for different technologies.
- Transform the PSM to code.

The PIM, PSMs and code describe a system in different levels of abstraction. Using MDA, the business is modeled in Platform-Independent Models, which are transformed into Platform-Specific Models. This is carried out in an automatic manner.

The success of MDA depends on the definition of transformation languages and tools that make a significant impact on full forward engineering processes and partial round-trip engineering processes.

BACKGROUND

UML is having a significant impact on the software development industry. So far, there are about 100 UML CASE tools that vary widely in functionality, usability, performance, and platforms (CASE, 2004). Table 1 shows a taxonomy of the UML CASE tools.

The competing tools can be compared and contrasted by the following requirements: easy interface, modeling productivity, implementation productivity, and extensibility. The main stream object-oriented CASE tools can help with the mechanics of drawing and exporting UML diagrams, eliminating syntactic errors and consistency errors between diagrams, and supporting code generation and reverse engineering.

The current techniques available in the commercial tools are not sufficient for MDA-based forward engineering. A source of problems in the code generation process is that, on the one hand, the UML models contain information that cannot be expressed in object-oriented languages while, on the other hand, the object-oriented languages express implementation characteristics that have no counterpart in the UML models. For instance, languages like Java, C++, and Eiffel do not allow explicit associations. These can be simulated by pointers and references, but then the structure of the system is not apparent. This often leads to problems during forward engineering between the specification and code.

Moreover, the existing CASE tools do not exploit all the information contained in the UML models. For instance, cardinality and constraints of associations and preconditions, postconditions, and class invariants in OCL are only translated as annotations. It is the designer's responsibility to make good use of this information, either selecting an appropriate implementation from a limited repertoire or implementing the association by himself.

UML CASE tools provide limited facilities for refactoring source code through an explicit selection made for the designer. However, it will be worth thinking about refactoring at the design level. The advantage of refactoring at the UML level is that the transformations do not have to be tied to the syntax of a programming language. This is relevant since UML is designed to serve as a basis for code generation with the MDA paradigm (Sunyé, Pollet, Le Traon & Jézéquel, 2001).

Many UML CASE tools support reverse engineering. However, they only use more basic notational features with a direct code representation and produce very large diagrams. Reverse engineering processes are facilitated by inserting annotations in the generated code. These annotations are the link between the model elements and the language. As such, they should be kept intact and not be changed. It is the programmer's responsibility to know what he or she can modify and what he or she cannot modify.

Techniques that currently exist in UML CASE tools provide little support for validating models in the design stages. Reasoning about models of systems is well supported by automated theorem provers and model checkers, however these tools are not integrated into CASE tools environments. Another problem is that as soon as the requirements specifications are handed down, the system architecture begins to deviate from specifications (Kollmann & Gogolla, 2002).

To solve these problems a lot of work has been carried out dealing with the semantics for UML models, advanced metamodeling techniques, and rigorous processes that fit MDA.

The Precise UML Group, pUML, was created in 1997 with the goal of giving precision to UML (Evans, France, Lano & Rumpe, 1998). It is difficult to compare the existing formalizations and to see how to integrate them in order to define a standard semantics since they specify different UML subsets and they are based on different formalisms (Ahrendt et al., 2002; McUmber & Cheng, 2001;

Table 1. UML CASE tools

Main Stream Object-Oriented CASE Tools	Rational Rose, Argo/UML, Poseidon, Together, GDPro, Stp/UML, MagicDraw
Real-Time/Embedded Tools	Tau UML Rhapsody Rational Rose Real Time
Basic Drawing Tools	Visio

Kuske, Gogolla, Kollmann & Kreowski, 2002; Laleau & Polack, 2002; Ziemann & Gogolla, 2003; Favre & Clerici, 2001; Favre, 2003).

Other works describe advanced metamodeling techniques that allow the enhancement of UML. Gogolla and Henderson-Sellers (2002) analyze the UML metamodel part dealing with stereotypes, and make various suggestions for improving the definition and use of stereotypes. Barbier, Henderson-Sellers, Le Parc-Lacayrelle, and Bruel (2003) introduce a formal definition for the semantics of the whole-part relation that can be incorporated into version 2.0 of UML.

UML CASE tools could be enhanced with functionality for formal specification and deductive verification; however, only research tools provide support for advanced analysis. For example, the main task of the USE tool (Ziemann & Gogolla, 2003) is to validate and verify specifications consisting of UML/OCL class diagrams. Key (Ahrendt et al., 2002) is a tool based on Together (CASE, 2004) enhanced with functionality for formal specification and deductive verification.

To date, few tools provide support for the MDA paradigm, for example, AndroMDA, AMEOS, CodagenArchitect, OptimalJ, and ArcStyler (CASE, 2004). The tool market around MDA is still in flux. References to MDA-based tools can be found at the OMG Web site (MDA, 2004).

MAIN THRUST OF FORWARD ENGINEERING AND UML

Formal and semi-formal techniques can play complementary roles in software development processes. This integration is beneficial for both graphical and formal specification techniques. On the one hand, semi-formal techniques have the ability to visualize language constructs, allowing a great difference in the productivity of the specification process, especially when the graphical view is supported by means of good tools. On the other hand, formal specifications allow us to produce a precise and analyzable software specification before implementation and to define semi-automatic forward engineering processes.

Favre, Martinez, and Pereira (2003) propose a rigorous process to forward engineer UML static models. The emphasis is given to the last steps in the road from UML to code. The bases of this approach are the *GSBL^{oo}* algebraic language and the *SpReIm* model for defining structured collections of reusable components.

The GSBL^{oo} language was defined taking into account the structuring mechanisms of UML. This language is relation-centric, which means that it expresses different kinds of UML relations (dependency, association, aggregation, and composition) as primitives to develop specifications. GSBL^{oo} allows us to keep a trace of the structure of UML models in the specification structure that will make it easier to maintain consistency between the various levels when the system evolves.

The *SpReIm* model allows us to define structured collections of reusable components that integrate algebraic specifications and object-oriented code. It takes advantage of the algebraic formalism power to describe behavior in an abstract way, while integrating it with concrete implementations. A component is defined in three levels of abstraction that integrate incomplete algebraic specifications, complete algebraic specifications, and object-oriented code.

A specific reusable component is Association. It describes a taxonomy of associations classified according to kind, degree, navigability, and multiplicity. The first level describes a hierarchy of incomplete specifications of associations using GSBL^{oo} and OCL. Every leaf in this level corresponds to sub-components at the second level. A realization sub-component is a tree of algebraic specifications: the root is the most abstract definition, the internal nodes correspond to different realizations of the root. For example, for a “binary, bi-directional, and many-to-many” association, different realizations through hashing, sequences, or trees could be associated. These sub-components specify realizations starting from algebraic specifications of Eiffel libraries.

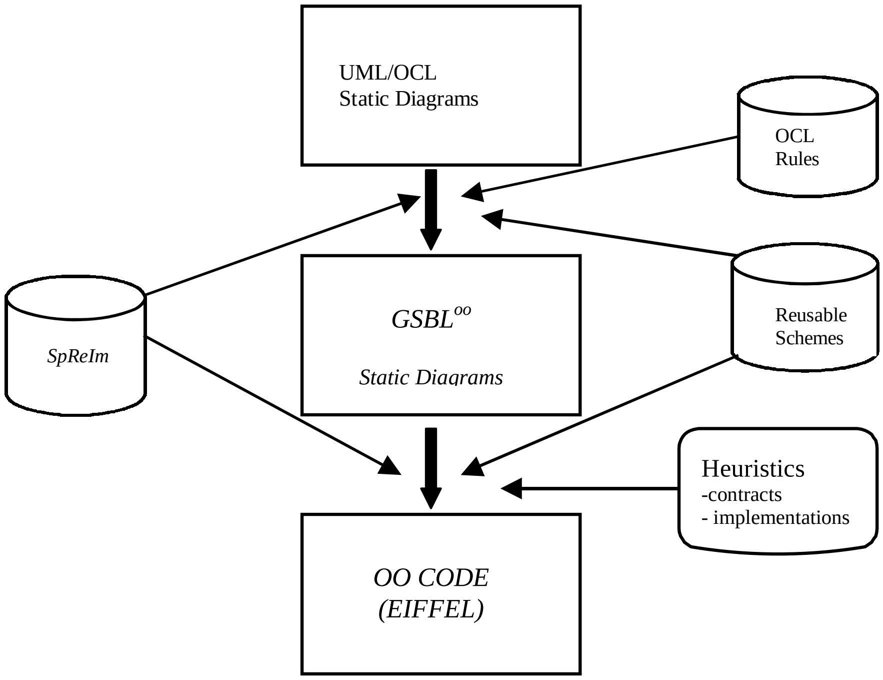
The implementation level associates each leaf of the realization level with different implementations in Eiffel. Implementation sub-components express how to implement associations and aggregations. For example, a bi-directional binary association with multiplicity “one-to-one” will be implemented as an attribute in each associated class containing a reference to the related object. On the contrary, if the association is “many-to-many,” the best approach is to implement the association as a different class in which each instance represents one link and its attributes.

The component reuse is based on the application of reuse operators: *Rename*, *Hide*, *Extend*, and *Combine*. These operators were defined on the three levels of components.

Figure 1 shows the main steps of the transformational process. Starting from UML class diagrams, an incomplete algebraic specification can be built by instantiating reusable schemes and components, which already exist in the GSBL^{oo} predefined library. Analyzing OCL specifications, it is possible to derive axioms that will be included in the GSBL^{oo} specification.

Preconditions written in OCL are used to generate preconditions in GSBL^{oo}. Postconditions and invariants

Figure 1. From UML static models to Eiffel



allow us to generate axioms in GSBL^{oo} (Favre, 2001). Thus, an incomplete specification can be built semi-automatically. The refinement of the GSBL^{oo} specification into a complete algebraic specification and code is based on a library of *SpReIm* components.

Then GSBL^{oo} specifications need to be transformed into object-oriented code. The Eiffel code is constructed gradually. First, associations and operation signature are translated. From OCL and GSBL^{oo} it is possible to construct contracts on Eiffel and/or feature implementation by applying heuristics. Associations are transformed by instantiating schemes that exist in the Association component.

The proposed transformations preserve the integrity between specifications and code. Software developers could perform maintenance and evolution on system specification, but not on the implementations. Modifications at specification levels must be applied again to produce a new and efficient implementation. Most of the transformations can be undone, which provides great flexibility in software development. All of the proposed transformations could be automated; they allow traceability and could be integrated into iterative and incremental software development processes supported by current UML CASE tools.

A detailed description of the forward engineering process may be found in Favre, Martinez, and Pereira (2003).

FUTURE TRENDS

UML is evolving as a result of insights gained through their use, both in industry and academia. The evolution of UML requires a large effort that is guided by user community.

The current UML 1.5 metamodel is insufficient to maintain the consistency between restructured design models, various design views, and implementations. This situation might be changed in the future since UML evolves to version 2.0.

Currently, OMG is promoting a transition from code-oriented to MDA-based software development techniques. Many tools claim to be compliant with MDA. The existing MDA-based tools do not provide sophisticated transformation from PIM to PSM and from PSM to code. To date, they might be able to support forward engineering and partial round-trip engineering between PIM and code. However, it will probably take several years before a full round-trip engineering based on standards occurs (many authors are skeptical about this).

Techniques that currently exist in UML CASE tools provide little support for generating business models. In light of the advances of the MDA paradigm, new types of UML tools that do a more intelligent job might emerge. It is probable that the next generation of tools will be able to describe the behavior of software systems in terms of business models and translate it into executable programs on distributed environment.

CONCLUSION

There is a great number of UML CASE tools in existence that facilitates code generation. Unfortunately, the current techniques available in these tools provide little automation for round-trip engineering. The formalization of UML can help to overcome these problems. In this direction, a rigorous process to transform UML static models into code was defined. It is based on the integration of semiformal notations with algebraic specifications.

Transformations are supported by a library of reusable schemes and by a system of transformation rules that allow translating UML constructs to algebraic specifications and Eiffel step-by-step. All the UML model information (classes, associations, and OCL constraints) are overturned in specifications having implementations implications. In particular, the transformation of different kinds of UML associations to Eiffel have been analyzed.

This research is still evolving, and additional issues will have to be tackled in order to fit new versions of UML and the MDA paradigm.

REFERENCES

Ahrendt, W., Baar, T., Beckert, B., Giese, M., Hähnle, R., Menzel, W., Mostowski, W. & Schmitt, P. (2002). The KeY system: Integrating object-oriented design and formal methods. *Proceedings of FASE 2002* at ETAPS 02, Grenoble, France.

Barbier, F., Henderson-Sellers, B., Le Parc-Lacayrelle, A. & Bruel, J. (2003). Formalization of the whole-part relationship in the Unified Modeling Language. *IEEE Transactions on Software Engineering*, *29*(5).

CASE. (2004). Retrieved from *www.objectsbydesign.com/tools/umltools_byCompany.html*

Evans, A., France, R., Lano, K. & Rumpe, B. (1998). Developing the UML as a formal modeling language. *Proceedings of UML'98—Beyond the Notation* (pp. 336-348). Berlin: Springer-Verlag (LNCS 1618).

Favre, L. (2001). A formal mapping between UML static models and algebraic specifications. Practical UML-based rigorous development methods—countering or integrating the eXtremist. *Lecture Notes in Informatics*, (7, GI Edition), 113-127. Alemania: Konner Kollen-Verlag.

Favre, L. (Ed.). (2003). *UML and the unified process*. Hershey, PA: IRM Press.

Favre, L. & Clérici, S. (2001). A systematic approach to transform UML static models to object-oriented code. In K. Siau & T. Halpin (Eds.), *Unified Modeling Language: System analysis, design and development issues* (Chapter 2, pp. 21-42). Hershey, PA: Idea Group Publishing.

Favre, L., Martínez, L. & Pereira, C. (2003). Forward engineering and UML: From UML static models to Eiffel code. In L. Favre (Ed.), *UML and the unified process* (Chapter IX, pp. 199-217). Hershey, PA: IRM Press.

Gogolla, M. & Henderson-Sellers, B. (2002). Formal analysis of UML stereotypes within the UML metamodel. *Proceedings of the 5th International Conference on the Unified Modeling Language*. Berlin: Springer-Verlag (LNCS 2460).

Jacobson, I., Booch, G. & Rumbaugh, J. (1999). *The unified software development process*. Reading, MA: Addison-Wesley Object Technology Series.

Kleppe, A., Warmer, J. & Bast, W. (2003). *MDA explained. The Model-Driven Architecture: Practice and promise.* Reading, MA: Addison-Wesley Object Technology Series.

Kollmann, R. & Gogolla, M. (2001). Application of the UML associations and their adornments in design recovery. *Proceedings of the 8th Working Conference on Reverse Engineering* (WCRE 2001). Los Alamitos, CA: IEEE Computer Press.

Kruchten, P. (2000). *The rational unified process: An introduction*. Reading, MA: Addison-Wesley Object Technology Series.

Kuske, S., Gogolla, M., Kollmann, R. & Kreowski, H. (2002). An integrated semantics for UML class, object and state diagrams based on graph transformation. *Proceedings of the 3rd International Conference on Integrated Formal Methods* (IFM'02). Berlin: Springer-Verlag.

Laleau, R. & Polack, F. (2002). *Coming and going from UML to B: A proposal to support traceability in rigorous IS development* (pp. 517-534). Berlin: Springer-Verlag (LNCS 2272).

McUmber, W. & Cheng, B. (2001). A general framework for formalizing UML with formal languages. *Proceedings of the IEEE International Conference on Software Engineering* (ICSE01), Toronto, Canada.

MDA. (2004). *Model-Driven Architecture*. Retrieved from *www.omg.org/mda*

OMG. (2004). *Unified Modeling Language Specification, v.1. 4*. Document ad/99-06-08, Object Management Group. Retrieved from *www.omg.org*

Rumbaugh, J., Jacobson, I. & Booch, G. (1999). *The unified modeling language reference manual*. Reading, MA: Addison-Wesley Object Technology Series.

Sunyé, G., Pollet, D., Le Traon, Y. & Jézéquel, J. (2001). Refactoring UML models. *Proceedings of the 2001 Conference on Modeling Languages, Concepts and Tools* (pp. 134-148). Berlin: Springer-Verlag (LNCS 2185).

Warmer, J. & Kleppe, A. (2003). *The object constraint language. Second edition. Getting your models ready for MDA*. Reading, MA: Addison-Wesley Object Technology Series.

Ziemann, P. & Gogolla, M. (2003). Validating OCL specifications with the USE tool—an example based on the BART case study. *Proceedings of the 8th International Workshop on Formal Methods for Industrial Critical Systems (FMICS'03)* (p. 80). Trondheim, Norway: Elsevier (Electronic Notes in Theoretical Computer Science).

KEY TERMS

Class Diagram: A diagram that shows a set of classes and their relationships (association, dependency, generalization/specialization, realization); class diagrams address the static design view of a system.

Computer-Aided Software Engineering (CASE): A tool to aid in the analysis and design of software systems.

Forward Engineering: The process of transforming a model into code through a mapping to a specific implementation language.

Model-Driven Architecture: A framework based on UML and other industry standards for visualizing, storing, and exchanging software design and models.

OCL (Object Constraint Language): A language used to express side effect-free constraints.

Refactoring: A change to a system that leaves its behavior unchanged, but enhances some nonfunctional quality factors such as simplicity, flexibility, understanding, and performance.

Reverse Engineering: The process of transforming code into a model through a mapping from a specific implementation language.

Unified Modeling Language (UML): A language for visualizing, specifying, constructing, and documenting the artifacts of a software-intensive system.

ENDNOTE

[1] This work is partially supported by CIC (Comision de Investigaciones Cientificas de la Provincia de Buenos Aires)

Franchising and Information Technology

Ye-Sho Chen
Louisiana State University, USA

Bin Zhang
Louisiana State University, USA

Bob Justis
Louisiana State University, USA

INTRODUCTION

Franchising is "a business opportunity by which the owner (producer or distributor) of a service or a trademarked product grants exclusive rights to an individual for the local distribution and/or sale of the service or product, and in return receives a payment or royalty and conformance to quality standards. The individual or business granting the business rights is called the *franchisor*, and the individual or business granted the right to operate in accordance with the chosen method to produce or sell the product or service is called the *franchisee*." (Justis & Judd, 2002)

In his best seller, *Business @ the Speed of Thought,* Bill Gates (1999) wrote: "Information Technology and business are becoming inextricably interwoven. I don't think anybody can talk meaningfully about one without talking about the other." Thus, to see how information technology (IT) is used in franchising, one needs to know how franchising really works.

BACKGROUND

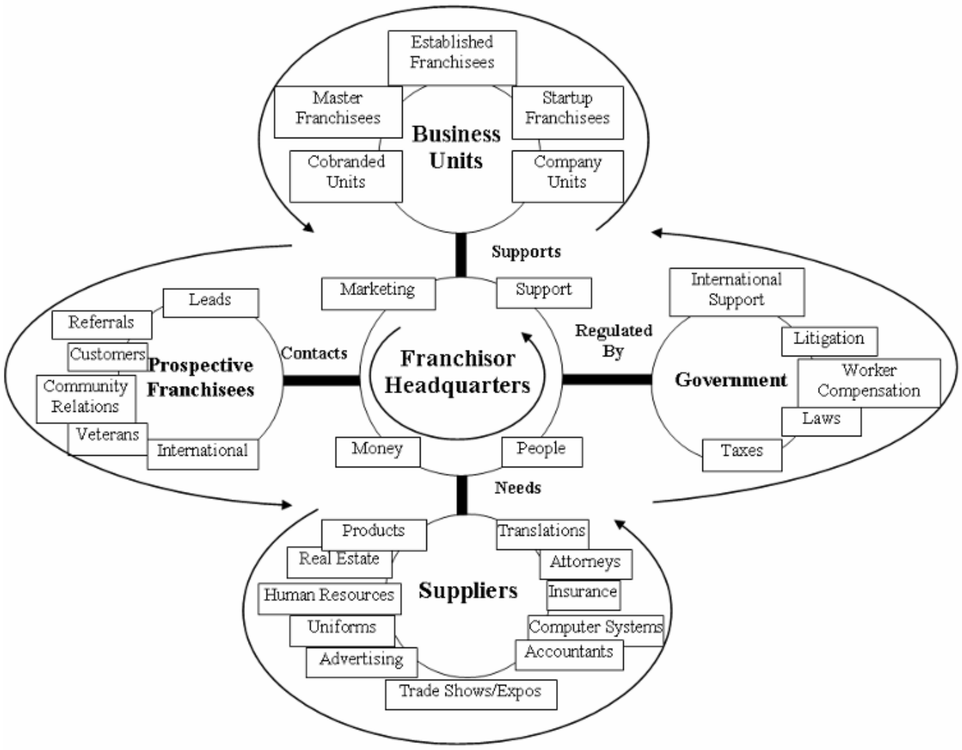
Understanding the Franchisor

The activities of the franchisor can be described using Figure 1 with five categories, having many sub-activities (represented by rectangles) in each of them:

- Franchisor Headquarters, including: (1) Support, helping business units; (2) Marketing, selling to prospective franchisees; (3) People, performing support and development; and (4) Money, dealing with financial/accounting issues. It also supports the following four different categories of activities.
- Business units, including company units, startup franchisees, established franchisees, master franchisees, and cobranded units. Among these five the franchisor should do the best to support startup franchisees, since good beginning is half done. Established and master franchisees are the ones the franchisor needs to provide incentives (e.g., cobranded units) to encourage expansion. Company units are typically used as role models for the franchisees.
- Prospective franchisees, contacted through several ways: (1) leads from marketing channels; (2) referrals from, for example, happy customers; (3) customers who love the products/services so much and would like to own the business themselves; (4) community and media relationships, helping deal with issues such as children's education; (5) public services like recruiting veterans; and (6) international contacts generated from master franchisees.
- Suppliers, including products and goods distributors and business service providers such as real estate agents, human resources providers, uniform vendors, marketing and advertising agents, trade shows/expos organizers, accountants, information systems vendors, insurance providers, attorneys, language translators, and many others.
- Government, including (1) federal, state, and international taxes; (2) local, region, and global laws; (3) insurances such as worker compensation; (4) possibilities of litigations from government, customers, and franchisees; and (5) supports for international expansions.

As we can see from Figure 1, the franchisor has to deal with many issues to grow the system. The learning process of the franchisor is incrementally developed through the following five stages (Justis & Judd, 2002): (1) Beginner – learning how to do it; (2) Novice – practicing doing it; (3) Advanced – doing it; (4) Master – teaching others to do it; and (5) Professional – becoming the best that you can be. At a higher stage of the development, most of the problems in the previous stages have been dealt with.

Figure 1. Understanding how the franchisor works



However, more complicated and challenging questions will arise as the franchise continues the expansion. This is especially true when the system reaches the stage of professional, when many unforeseen and difficult problems could happen all of a sudden. Bud Hadfield (1995), the founder of Kwik Kopy franchise and the International Center of Entrepreneurial Development, said it the best: “The more the company grows, the more it will be tested.” To capture the learning process in Figure 1, a counter-clockwise round arrow is used in each of the five categories. It depicts the increasing intensity of learning in every area of the sub-activities (represented by rectangles) as the franchise continues surviving and thriving. For example, as the system expands, the real estate sub-activity becomes much more complicated since the issue of territory encroachment becomes more significant and harder to deal with.

Understanding the Franchisee

The activities of the franchisee can be described using Figure 2 with five categories, having many sub-activities (represented by rectangles) in each of them:

- Franchisee outlet, consisting of making sales to customers, marketing and advertising, dealing with financial/accounting issues, and managing people who make sales.
- Customers, including five types: potential, infrequent, frequent, online, and cobranded.
- Franchisor headquarters, providing supports to the franchisee, demonstrations from field representatives, training and continued education from the management group, discussion forums, and distance learning.
- Suppliers, similar to the list shown in Figure 1.
- Government, similar to the list shown in Figure 1.

The franchisee also goes through previously mentioned five-stage learning process (i.e., beginner, novice, advanced, master, and professional) represented by a counter-clockwise round arrow used in each of the five categories in Figure 2. Once again, the arrow depicts the increasing intensity of learning as the franchisee continues growing the business and many unforseen problems/issues may rise up to challenge the practices.

Understanding the Franchisor/ Franchisee Relationship

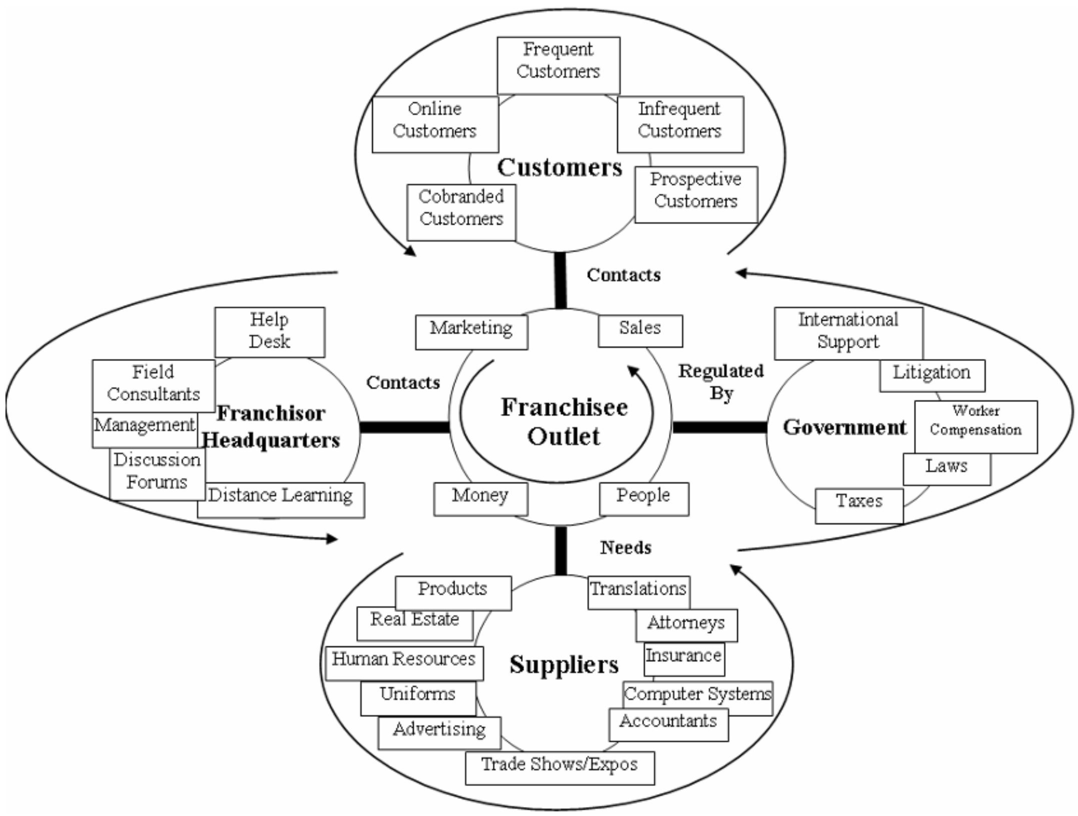
Developing a good “family” relationship between the franchisor and the franchisee is believed to be the most important factor for the success of a franchise (Justis & Judd, 2002). To understand how the relationship is developed, one needs to know the franchisee life cycle (Schreuder, Krige & Parker, 2000):

- The Courting Phase: Both the franchisee and the franchisor are excited with the relationship. This corresponds typically to the beginner stage of the franchisee.
- The "We" Phase: The relationship starts to deteriorate, but the franchisee still values the relationship. This corresponds typically to the novice stage of the franchisee.
- The "Me" Phase: The franchisee starts to question the reasons for payments-related issues with the attitude that the success so far is purely of his/her own work. This corresponds typically to the advanced stage of the franchisee.
- The Rebel Phase: The franchisee starts to challenge the restrictions being placed upon him or her. This corresponds typically to the master stage of the franchisee, since the rebel ones tend to be those who know the system very well and are capable of influencing others to follow them.
- The Renewal Phase: The franchisee realizes the "win-win" solution is to continue teaming up with the franchisor to grow the system. This corresponds typically to the professional stage of the franchisee.

The major challenge for the franchisor is to turn a "rebel" franchisee into the "renewal" one by providing a learning and innovative environment where the franchisee can continue contributing to the growth of the system. Successful collaborative learning and innovations will provide a strong incentive for the professional franchisees to continue their renewal relationship with the firm, which in turn will have positive impact on maintaining the good relationship with other franchisees and recruiting new ones. On the other hand, constant failures in this stage will intensify the rebel franchisees to desert the firm, which in some cases leads to the demise of the franchise. Lying behind the successful collaboration is the working knowledge of the franchise firm. Knowledge is defended as "a justified personal belief that increases an individual's capacity to take effective action" (Alavi & Leidner, 1999). Knowledge becomes "working" when the action produces results. When knowledge produces results, the personal belief becomes much more strengthened, intensified, and justified. As such, the individual's capacity is much more increased and better results are obtained.

This spiral-up cycle of working knowledge development is very important in the context of franchising. Figure

Figure 2. Understanding how the frachisee works



3 shows that the development process is incrementally developed through the five stages of the spiral-up cycle defined earlier: beginner, novice, advanced, master, and professional. The foundation of the learning cycle is the capability of sharing and coaching the working knowledge throughout the franchise system. The process of influencing others for knowledge dissemination consists of five steps (Justis & Vincent, 2001): (1) Knowledge, proven abilities to solve problems in the franchise environment; (2) Attitude, constructive ways of presenting and sharing the working knowledge; (3) Motivation, incentives for learning or teaching the working knowledge; (4) Individual Behavior, the strengths of the participants to learn and enhance the working knowledge; and (5) Group Behavior, collaborative ways to create, dissimilate, and manage the hard-earned working knowledge. The franchisor/franchisee "family" relationship building in Figure 3 is also surrounded with a dashed line, meaning the relationship is enlarged and expanded without limits as the franchisee incrementally learns the working knowledge through the influencing of the franchisor and the fellow franchisees. By going through the processes of learning and influencing, both the franchisor and franchisee gain the working knowledge progressively and mange the franchisee life cycle effectively.

FUTURE TRENDS

An Attention-Based IT Infrastructure in Franchising

In an information-rich world, Herbert Simon (Nobel laureate in Economics in 1978) wrote: "a wealth of information creates a poverty of attention" (Simon, 1971). As such, the "proper aim of a management information system is not to bring the manager all the information he needs, but to

Figure 3. Understanding the franchisor/franchisee family relationship

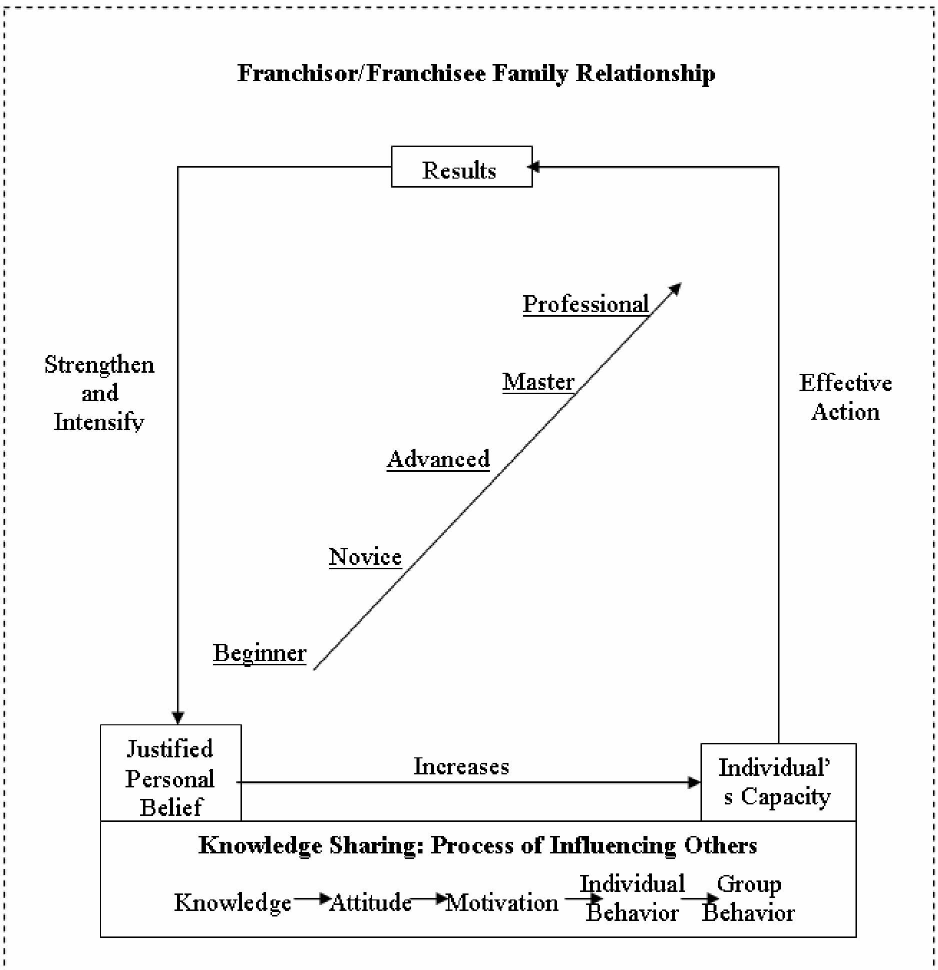
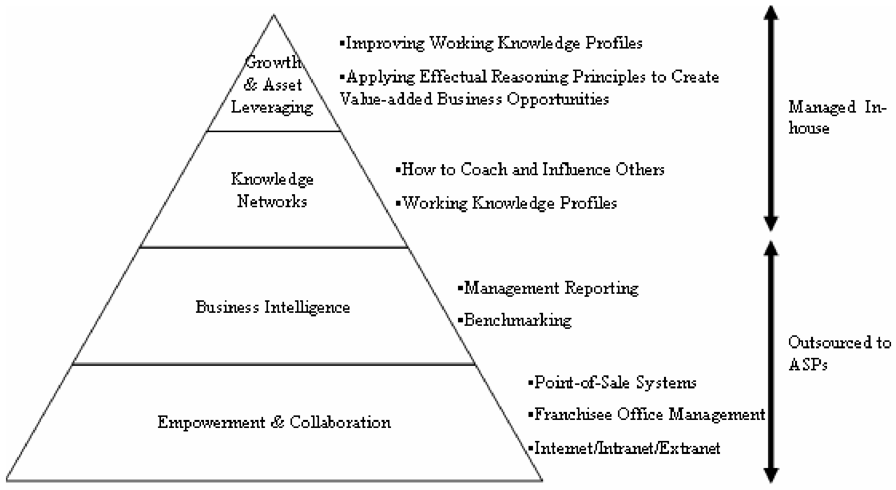


Figure 4. An attention-based IT infrastructure in franchising



recognize the manager's environment of information so as to reduce the amount of time he must devote to receiving it" (Simon, 1971). From the discussions in the last section, it is obvious that an attention-based IT infrastructure in franchising shall be one devoted to enabling the building of a good "family" relationship between the franchisor and the franchisee. In Figure 4, we propose such an IT architecture for franchise organizations to manage the massive information produced by the firms. The architecture, adapted from Gates' Digital Nervous Systems (1999), consists of the following four layers:

- Empowerment and Collaboration: Point-of-sale and office management systems are used to empower the franchisees (Chen, Justis & Chong, 2002). Internets, extranets, and intranets networking the franchisor, the franchisees, customers and suppliers are deployed to improve collaboration (Chen, Chong & Justis, 2002).
- Business Intelligence: Data warehousing and data mining techniques are used to transform volumes of data into reports and benchmarks for management to glean business intelligence (Chen, Justis & Chong, 2004; Chen, Justis & Watson, 2002; Zhang, Chen & Pawlowski, 2003).
- Knowledge Networks: Intranet-based systems consisting of the skills of coaching/influencing others and working knowledge profiles (Chen, Hammerstein & Justis, 2002) are implemented for knowledge sharing and learning within the franchise business. A distance-learning curriculum (Chen, Chong & Justis, 2000) of working knowledge modules can also be deployed.
- Growth and Asset Leveraging: Value networks are developed with the goal of improving working knowledge profiles and applying effectual reasoning principles to create value-added business opportunities (Chen, Justis & Yang, 2004).

To demonstrate how the attention-based IT infrastructure is related to the activities of the franchisor and the franchisee (shown in Figures 1 and 2) and the "family" relationship building (shown in Figure 3), we show in Figure 5 the character of the business process and the information flow in the franchise business and explain why the attention-based IT infrastructure is needed. The foundation of the architecture, adapted from Inmon (1996), consists of four levels: (1) data collected from the empowerment and collaboration activities of the franchisor (Figure 1) and the franchisee (Figure 2); e-business strategy shall be one empowering the franchisor and the franchisees to do their activities (Chen, Chong & Justis, 2002); (2) reconciled data in the franchise data warehouse (Chen, Justis & Watson, 2002); (3) derived data residing in data marts based on various franchisee-centered segmentations such as franchise development and support; and (4) business intelligence reports generated from analytical data analysis; for example, management reporting and benchmarking could be produced by online analytical processing (OLAP) periodically, which may lead to more proactive analysis of top/low performer attributes using the data mining techniques (Chen, Justis & Chong, 2004).

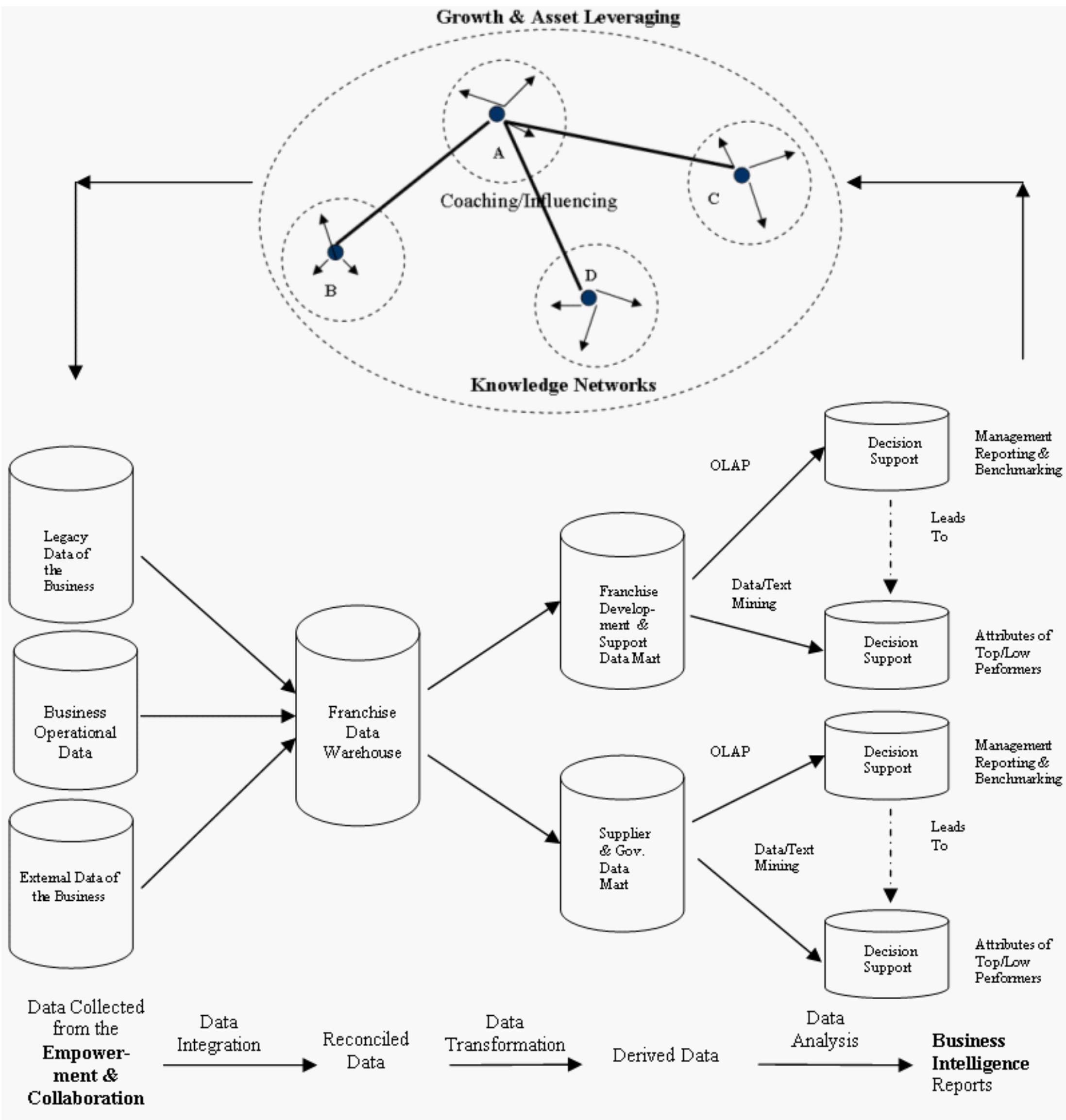
The business intelligence reports will help the franchise system to identify key top performers and their success attributes. With their involvement, working knowledge profiles can be developed and dissimilated throughout the knowledge networks of the business using coaching/influencing techniques (Chen, Hammerstein & Justis, 2002). The top of Figure 5 depicts such a knowledge sharing and dissemination idea described in Figure 3 in detail. There are four franchisees (A - D) in the figure. Each franchisee (a dot) has his/her personal knowledge network (arrows pointing out of the dot). The knowledge network may include the customers' likes and dislikes, the kind of employees to hire, the competitors' and suppliers' pricing strategies, and the social needs in the local community. Each franchisee is surrounded with a circle with dashed lines, meaning there is no limit to the personal knowledge network. Suppose franchisee A is the top performer and is charged and rewarded with coaching/influencing other franchisees to survive and thrive. Thus, clusters (connected dots) of knowledge network are formed and surrounded with a circle with dashed lines, meaning there is no limit for improving and leveraging the assets of the business (Chen, Justis & Yang, 2004).

Using the attention-based IT architecture as a guide, we recommend franchise organizations to outsource IT in the first two layers to application service providers (ASPs) (Chen, Ford, Justis & Chong, 2001). The concept of subscribing IT services through ASPs has special appeal in the franchising industry because an ASP can duplicate success for other similar franchises quickly and inexpensively (Yao, Wohl, Watson & Chen, 2004).

CONCLUSION

Franchising has been popular as a growth strategy for small businesses; it is even more so in today's global and

Figure 5. Implementing the attention-based IT infrastructure in franchising



e-commerce world. Although IT is quite important in franchising, IT researchers have largely ignored this arena. One major reason is that few IT researchers are vested in the knowledge of how franchising functions, and without this intimate knowledge it is difficult to implement effective IT systems. Based on years of academic and consulting experiences in franchising and IT, the authors propose a framework of IT in franchising. At the heart of the framework is to manage working knowledge and use IT to build up the "family" relationship between the franchisor and the franchisee.

REFERENCES

Alavi, M., & Leidner, D.E. (1999, February). Knowledge management systems: Issues, challenges, and benefits. *Communications of the Association for Information Systems, 1*(article 7).

Chen, Y., Chong, P., & Justis, R.T. (2000, February 19-20). Franchising knowledge repository: A structure for learning organizations. *Proceedings of the 14th Annual International Society of Franchising Conference,* San Diego, CA.

Chen, Y., Chong, P.P., & Justis, R.T. (2002, February 8-10). E-business strategy in franchising: A customer-service-life-cycle approach. *Proceedings of the 16th Annual International Society of Franchising Conference,* Orlando, FL.

Chen, Y., Ford, C., Justis, R.T, & Chong, P. (2001, February 24-25). Application service providers (ASP) in franchising: Opportunities and issues. *Proceedings of the 15th Annual International Society of Franchising Conference,* Las Vegas, NV.

Chen, Y., Hammerstein, S., & Justis, R.T. (2002, April 5-6). Knowledge, learning, and capabilities in franchise organizations. *Proceedings of the 3rd European Conference on Organizational Knowledge, Learning, and Capabilities,* Athens, Greece.

Chen, Y., Justis, R., & Watson, E. (2000). Web-enabled data warehousing. In M. Shaw, R. Blanning, T. Strader & A. Whinston (Eds.), *Handbook of electronic commerce* (pp. 501-520). Springer-Verlag.

Chen, Y., Justis, R.T., & Chong, P.P. (2002). Franchising and information technology: A framework. In S. Burgess (Ed.), *Managing information technology in small business: Challenges and solutions* (pp. 118-139). Hershey, PA: Idea Group Publishing.

Chen, Y., Justis, R.T., & Chong, P.P. (2004). Data mining in franchise organizations. In H.R. Nemati & C.D. Barko (Eds.), *Organizational data mining: Leveraging enterprise data resources for optimal performance* (pp. 217-229).

Chen, Y., Justis, R.T., & Yang, H.L. (2004, March 5-7.). Strategic growth of franchise firms in the digital economy. *Proceedings of the 18th Annual International Society of Franchising Conference*, Las Vegas, NV.

Gates, W. (1999). *Business @ the speed of thought.* Warner Books.

Hadfield, B. (1995). *Wealth within reach.* Cypress Publishing.

Inmon, W.H. (1996). *Building the data warehouse.* John Wiley & Sons.

Justis, R.T., & Judd, R.J. (2003). *Franchising* (3rd ed.). DAME Publishing.

Justis, R.T., & Vincent, W.S. (2001). *Achieving wealth through franchising.* Adams Media Corporation.

Schreuder, A.N., Krige, L., & Parker, E. (2000, February 19-20). The franchisee lifecycle concept – A new paradigm in managing the franchisee-franchisor relationship. *Proceedings of the 14th annual International Society of Franchising Conference*, San Diego, CA.

Simon, H.A. (1971). Designing organizations for an information rich world. In M. Greeberger (Ed.), *Computers, communications, and the public interest* (pp. 38-52). Baltimore, MD: The Johns Hopkins Press.

Yao, Y., Wohl, M., Watson, E., & Chen, Y. (2004). Customers' decision to adopt application service provider and applications service providers' business strategy in the hospital industry: A research framework. To appear in *Journal of Information Technology Cases and Applications*.

Zhang, B., Chen, Y., & Pawlowski, S. (2003, August 4-6). Online data mining in franchising supply chain management: A case study in apparel industry. *Proceedings of the Americas Conference of Information Systems,* Tampa, FL.

KEY TERMS

Attention-Based IT Infrastructure: An IT infrastructure that is able to sort through volumes of data and produce the right information at the right time for the right persons to consume.

Franchisee: The individual or business that receives the business rights and pays the royalties for using the rights.

Franchisee Life Cycle: The stages a franchisee goes through in the franchise system: Courting, "We," "Me," Rebel, Renewal.

Franchising: A business opportunity based on granting the business rights and collecting royalties in return.

Franchisor: The individual or business that grants the business rights.

Franchisor/Franchisee Learning Process: The stages of learning, including beginner, novice, advanced, master, and professional.

Franchisor/Franchisee Relationship Management: The vital factor for the success of a franchise, including: Knowledge, Attitude, Motivation, Individual Behavior, and Group Behavior.

From Digital Divide to Digital Dividend

Nikhilesh Dholakia
University of Rhode Island, USA

Nir Kshetri
The University of North Carolina at Greensboro, USA

INTRODUCTION

Despite rapidly falling costs of hardware, software, and telecommunications services, a wide gap persists between rich and poor nations in terms of their capabilities of accessing, delivering, and exchanging information in digital forms (Carter & Grieco, 2000). Developing countries, comprising over 81% of the world population, account for a tiny fraction of global e-commerce. An estimate suggests that 99.9% of business-to-consumer e-commerce in 2003 took place in the developed regions of North America, Europe, and Asia Pacific (Computer Economics, 2000).

Whereas high-income countries have income 63 times that of low-income countries, the respective ratios are 97 for PCs (personal computers), 133 for mobile phones, and over 2,100 for Internet hosts (Dholakia & Kshetri, 2003). While reliable data on e-commerce transactions are not available, the ratio is likely to be even higher for e-commerce transactions since e-commerce is virtually nonexistent in many developing countries. The pattern indicates that the gap between developed and developing countries is wider for more recent technologies such as PCs, mobile phone, and the Internet than for technologies that were introduced earlier.

In the absence of appropriate policy measures, it is likely that the "global digital divide" will become wider (Dholakia & Kshetri, 2003; Economist.com, 2000). Policy measures directed at making appropriate networks available to the digitally excluded populations at reasonable costs could bridge the gap or at least decrease the rate at which it widens.

This article provides an assessment of two computer networks that redefine the conventional definition of market value by allowing developing nations and communities (Brooks, 2001) to reap the benefits of modern ICTs (information and communications technologies): Global Trade Point Network (GTPNet) and Little Intelligent Communities (LINCOS).

BACKGROUND

The global digital divide is the result of the complex interactions among ICTs and various economic, political, and social factors in the environment. First, a large majority of potential users in developing countries are unable to afford a telephone line, a PC, and the telephone and Internet services provider (ISP) access charges. Whereas the cost of a PC is 5% of the per capita GDP (gross domestic product) in high-income countries, it is as high as 289% in low-income countries (ITU, 2001). Furthermore, monthly Internet access charge as a proportion of per capita GDP in the world varies from 1.2% in the US to 614% in Madagascar (UNDP, 2001).

Second, for some consumers in developing countries that are willing to pay for the connection of a telephone line, there is a big gap between demand and supply. For instance, in 2001, 33 million people in the developing world were on the registered waiting lists for telephone connections, the average waiting periods being over 10 years in some countries.

A third problem is related to the lack of skills. A majority of potential users in developing countries lack English language and computer skills, prerequisites for the use of the Internet. For instance, in 1998 about 85% of the text on the Internet was in English (Nunberg, 2000). This proportion decreased to 80% in 1999 and 50% by 2003. Although a shift of Internet content to non-English languages is under way, some knowledge of English is still necessary to use the Internet as the bulk of software used in the Internet is in English (Hedley, 1999) and most of the human-computer interfaces favor English-language users (Goodman, Press, Ruth, & Ruthowski, 1994).

A fourth problem is related to the lack of relevant content. Although there are over 17 billion Web pages[1] in existence, the content remains largely geared to the needs of advanced nations. Edejer (2000) observes the difficulty of finding reliable health-related information relevant to developing countries online:

Few reports of health research from developing countries are published in journals indexed by Western services such as Medline. *Western indexing services cover some 3,000 journals, of which 98% are from the developed world. The whole of Latin America accounted for 0.39% of the total number of articles referenced by Medline in 1996.... Because only a small number of journals from developing countries are indexed by Medline, research from these countries is almost invisible.*

CREATIVE WAYS TO BRIDGE THE DIGITAL DIVIDE: SOME EXAMPLES

The effectiveness of a network in bridging the global digital divide is thus a function of (a) the network's ability to identify priorities of digitally excluded populations and (b) the network's ability to attack the major barriers to Internet and e-commerce adoption. In the following section, we examine two networks designed to enable e-business systems for the global poor: GTPNet (Figure 1) and LINCOS (Figure 2).

Global Trade Point Network of the World Trade Point Federation (http://www.wtpfed.org/)

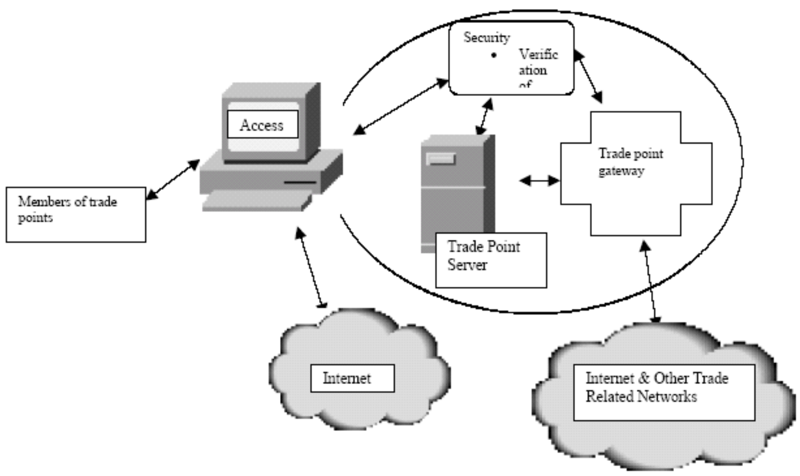
The United Nations Conference on Trade and Development (UNCTAD) launched the Global Trade Point Program in 1992 to facilitate the access to international markets for SMEs. The program was taken over by the World Trade Point Federation in November 2002. In mid-2003, GTPNet had a human network of 121 trade points in over 80 countries on five continents[2].

In a trade point, participants in foreign trade transactions (e.g., customs authorities, foreign trade institutes, banks, chambers of commerce, freight forwarders, transport and insurance companies) are grouped together under a single physical or virtual roof to provide all required services at a reasonable cost. It is a source of trade-related information providing actual and potential traders with data about business and market opportunities, potential clients and suppliers, trade regulations and requirements, and so forth. A survey found that 85.7% of trade-point customers are SMEs and microenterprises (UNCTAD, 1997).

The ETO (electronic trading opportunity) system is probably one of the most important aspects of the GTPNet. It was started by the UN Trade Point Development Center (UNTPDC) in June 1993 and is the world's largest Internet-based business opportunities system. ETOs are offers and demands for products, services, and investment and are distributed point to point and company to company. They are forwarded to the GTPNet system by trade points and third-party information providers. A random survey of ETO users conducted in 1998 revealed that 48% of the ETO users received 1 to 10 responses per posted ETO, an additional 14% receive 10 to 30 responses, and about 7% receive over 100 reactions. About a third of respondents made business deals on the basis of ETOs.

Developing countries also have a much higher share in the ETOs than in overall global e-commerce. A UNTPDC analysis of ETOs posted on the GTPNet during March 1 to July 15, 1998, indicated that 20% of them were posted by US-based companies, followed by companies in China (19%), South Korea (11%), and India (7%; UNCTAD, 1998). It is interesting to note that the US accounted for 74% of the global Internet commerce market in 1998 (Wang, 1999).

Figure 1. UNCTAD GTPN



Password-restricted areas have been added to the GTPNet site, which uses state-of-the-art tools for uploading, downloading, automatic updating, and searching for information. Only trade points and members of trade points can send ETOs and see hot ETOs[3]. Java is used to control access, certify trading partners, and handle payments. The Java-based secure infrastructure ensures integrity and confidentiality of all trade information. Certification is the first step in secure trading. Prospective traders download the UNTPDC's 100%-pure Java-based applet and use it to provide the UN with reference data about their banking, trading, and services. After the UN certifies it, the company uses the Java applet residing in their standard Internet browser to access the network. Similarly, the smart-card project of the UNCTAD is facilitating the payment flow in international trade. As discussed in the next section, the first- and second-level smart cards allow secure ETO, confidentiality, payment information integrity, authentication, and so forth.

The Secure Electronic Authentication Link (SEAL) project and concept were developed by the United Nations Trade Point Development Center. Its smart-card project facilitates payment flows in international trade. The first-level smart card allows users to automatically authenticate their user profile to the SEAL and secure ETO on the GTPNet. The second-level smart card allows confidentiality of information, payment information integrity, cardholder account authentication, merchant authentication, and interoperability with the ETO system on the Internet and the GTPNet.

Most of the trade points are, however, located in big cities, and many least developed countries are still deprived of the services of trade points. GTPNet sites are also vast and not well organized (Lehrer, 2003). A recent international forum on ways to improve the trade point programs for SMEs came up with many suggestions[4] including the necessity of encouraging and assisting nonexporting companies, possibly by even creating special e-commerce programs open to new exporters only; providing guidance for the management; providing translation services so that import-export business can be conducted in multiple languages; and providing convenient online payment mechanisms.

LINCOS: Little Intelligent Communities (http://www.lincos.net/)

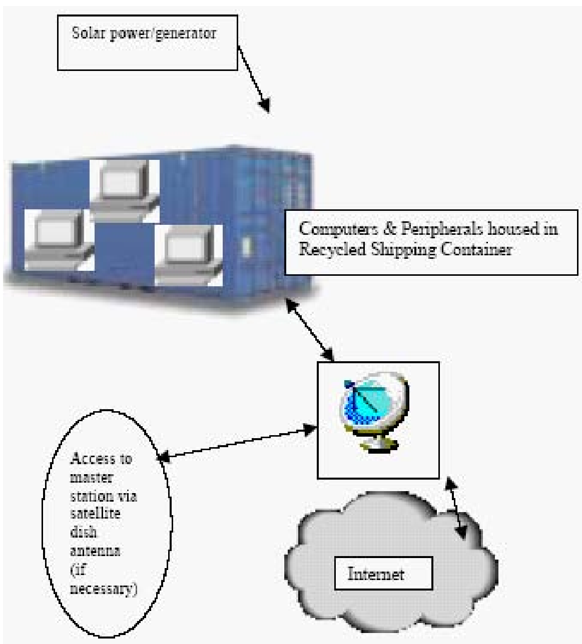
The LINCOS initiative was developed jointly by the Fundación Costa Rica para el Desarrollo Sostenible, the Media Lab at MIT, and the Instituto Tecnológico de Costa Rica in 1998 (Saxe et al., 2000). LINCOS selected Hewlett Packard's E-inclusion as the model. It has alliances with over 10 academic institutions and at least 10 technology companies (United Nations, 2000). Initially, LINCOS used recycled shipping containers to house computers, peripherals, and generators. Each unit comprised of five computers and other facilities to provide a broad range of services including Internet access; health, education, banking, and government services; electronic trading; technical support for SMEs; telecom and information centers; video conferencing and entertainment; forest, soil, and water analysis; and so forth. The ultimate goal of the project is to achieve sustainability by allowing each community within the network to make decisions related to the technologies and advance them independently[5].

Each unit is satellite operated and solar-power enabled, and can operate independently of traditional infrastructures. The satellite dish antennae link them to any telecom network or master station as needed. Eggers and Siefken (2000) comment on LINCOS' performance:

LINCOS units can be taken anywhere—mountain, jungle or village—and make web-browsing, telephony and e-mail available even in the most remote spot. Equipped with their own generators, these units need neither external energy nor communication cables; they connect directly via satellite.

The units are installed in a container equipped with five computers. Additional LINCOS units are to be provided if more computers need to be added. The investment for each unit was as high as $85,000 in the Dominican Republic and $50,000 in San Marcos de Terazu, Costa Rica (Proenza, 2001), but is expected to decrease with the

Figure 2. LINCOS network



increase in production (United Nations, 2000). In 2000, LINCOS won ALCATEL's first place prize for technological innovation in Latin America. Over 30 projects in Latin America participated in the competition[6].

The prototype LINCOS sites, which have already been deployed in Costa Rica, are providing several benefits. For instance, coffee growers use LINCOS sites to find the best prices in the world as well as next week's weather. Thanks to LINCOS' assistance, Costa Rican coffee farmers have also created Web pages and learned how to request budgets for buying equipment and register their trademarks (Amighetti & Reader, 2003).

While a container has had obvious advantages for carrying the equipment to remote communities, the "container concept" was later abandoned because it created the concept of "temporariness"; the target users did not regard it as "rooted in the community" and tended to develop the feeling that the project was brought to the community in a "top-down" manner as a wrapped-up "development package"[7]. A second drawback of LINCOS is that it has attracted relatively rich people instead of helping the poor (Amighetti & Reader, 2003). According to the World Bank's Charles Kenny, LINCOS have had mixed results because of the lack of sufficient interest of the target audience. He argues: "Poor people don't seem to think that the Internet is the answer to all their problems" (Rich, 2003, p. 93).

POLICY IMPLICATIONS: LOCAL, NATIONAL, AND GLOBAL

Rapidly dropping costs of ICTs, developments of user-friendly software and interfaces, and the versatility of the Internet offer the potential for leapfrogging many of the development obstacles. Civil society, governments, and entrepreneurs of developing countries can take actions to bridge the digital divide by targeting highly excluded communities by designing an appropriate combination of new and old technologies, and by setting projects in the context of a longer term plan to extend the benefits more widely.

Although the networks discussed in this article are helping to bridge the digital divide between the developing and developed countries to some extent, most of them are not yet able to reach really excluded populations in developing countries. For instance, mainly rich people are benefiting from LINCOS. Similarly, there are no trade point programs in small villages of developing countries.

There is a huge untapped market for modern ICTs in developing countries if the services are *affordable* and *appropriate* to the target population, and e-business companies need not provide their services in philanthropic ways. Lyle Hurst, director of HP's E-inclusion, a partner of LINCOS, said in 2002 that the mobile digital community centers will be a significant market opportunity for all involved. To exploit the potential, comprehensive research on the needs of the digitally excluded population, and on the most appropriate networks to satisfy such needs by using locally available expertise and resources is needed.

The experiences of LINCOS and GTPNet indicate that lack of awareness and interest with the target audience are the major drawbacks of the networks discussed in this article. National governments, international agencies, and technology marketers are required to work together to educate the target users about the potential benefits of such networks.

Equally important, perhaps, is what Rossman (2004) refers to as the "fluency divide"—the disparity between privileged and unprivileged groups in terms of their capability of finding and applying information to achieve their goals. While it is important to bridge the digital divide of unequal access to devices, equipment, and contents, policy makers are also required to work on bridging the fluency divide between those passively consuming information and those that have the motivation and skills to find and apply information to achieve their goals.

REFERENCES

Amighetti, A., & Reader, N. (2003, July 24) Internet project for poor attracts rich. *The Christian Science Monitor,* 16.

Brooks, K. (2001). Pas de deux: The dance of digital design. *Design Management Journal, 12*(2), 10-15.

Carter, C., & Grieco, M. (2000). New deals, no wheels: Social exclusion, tele-options and electronic ontology. *Urban Studies, 37*(10), 1735-1748.

Computer Economics. (2000). The global economy is not so global. *Internet & E – Business Strategies, 4*(4), 1-3.

Dholakia, N., & Kshetri, N. (2003). Electronic architectures for bridging the global digital divide: A comparative assessment of e-business systems designed to reach the global poor. In S. Nansi (Ed.), *Architectural issues of Web-enabled electronic business* (pp. 23-40). Hershey, PA: Idea Group Publishing.

Economist.com. (2000, September 23). *Falling through the net.*

Edejer, T. T.-T. (2000). Disseminating health information in developing countries: The role of the Internet. *British Medical Journal, 321*, 797-800.

Eggers, I., & Siefken, S. T. (2000). Four countries connect. *UN Chronicle, XXXVII*(2). Retrieved from *http://www.un.org/Pubs/chronicle/2000/issue2/0200p32.htm*

Goodman, S. E., Press, L. I., Ruth, S. R., & Ruthowski, A. M. (1994). The global diffusion of the Internet: Patterns and problems. *Communications of the ACM, 37*(8), 27-31.

Hedley, R. A. (1999). The information age: Apartheid, cultural imperialism, or global village? *Social Science Computer Review, 17*(1), 78-87.

ITU. (2001). *The Internet: Challenges, opportunities and prospects.* Retrieved July 1, 2001, from *http://www.itu.int/newsroom/wtd/2001/ExecutiveSummary.html*

Lehrer, B. (2003). Finding business opportunities on the Internet. Fita.org. Retrieved from *http://www.fita.org/aotm/tbird.html*

Nunberg, G. (2000, March 27-April 10). Will the Internet always speak English? *The American Prospect*, 40-43.

Rich, J. L. (2003, November/December). Not-so-simple solution. *Foreign Policy*, 93

Rossman, P. (2004). COSMOPEDIA. *The Futurist, 38*(3), 26-30.

Saxe, E. B. (2000). *Taskforce on bridging the digital divide through education*. Retrieved from *http://www.worldbank.org/edinvest/lincos.htm*

UNCTAD. (1998). Trade point review. *United Nations Conference on Trade and Development.* Retrieved from *http://www.sdnp.undp.org/mirrors/lc/pan/untpdc/gtpnet/tpreview/#2.%201% 20The%20current% 20status%20 of%20Trade%20Points*

UNDP. (2001). Human development report 2000. *United Nations Development Program*. Retrieved from *http://www.undp.org/hdr2001/completenew.pdf*

United Nations. (2000). Report of the high-level panel of experts on information and communication technology. *General Assembly, Economic and Social Council*. Retrieved from *http://www.un.org/documents/ga/docs/55/a5575.pdf*

Wang, A. (1999, April 26). Verio expands global reach of e-commerce. *E-Commerce Times*. Retrieved from *http://www.ecommercetimes.com/perl/story/143.html*

KEY TERMS

Digital Divide: The fact that people in developing countries use modern ICTs less than those in developed countries.

Gross Domestic Product (GDP): The sum of the total value of consumption expenditure, total value of investment expenditure, and government purchases of goods and services.

High-Income Countries: Have a per capita income of $9,266 or more (in 2001).

Information and Communications Technologies (ICTs): Technologies that facilitate the capturing, processing, storage, and transfer of information.

Internet: The "global information system that (i) is logically linked together by a globally unique address space based on the Internet Protocol (IP) or its subsequent extensions/follow-ons; (ii) is able to support communications using the Transmission Control Protocol/Internet Protocol (TCP/IP) suite or its subsequent extensions/follow-ons, and/or other IP-compatible protocols; and (iii) provides, uses or makes accessible, either publicly or privately, high level services layered on the communications and related infrastructure described herein" (The Federal Networking Council definition).

Low-Income Countries: Have a per capita income of less than $875 (in 2001).

PPP (Purchasing Power Parity): A rate of exchange that accounts for price differences across countries, allowing international comparisons of real output and incomes.

ENDNOTES

1. See "How Big is the Internet?" and "How Fast is the Internet Growing?" at http://www.metamend.com/internet-growth.html
2. See introductory word by the WTPF president (May 23, 2003) at http://www.wtpfed.com/newsite/index1.php#
3. ETOs that are less than 8 days old
4. These suggestions are condensed from UNCTAD/WTO e-mail discussions on SMEs, reported at http://www.intracen.org/e_discuss/sme/welcome.htm (retrieved August 27, 2003)
5. See the details of the project at http://projects.takingitglobal.org/lincos
6. See http://www.lincos.net/webpages/english/acerca/historia.html
7. See "Assessing ICT Efforts in Marginalized Regions...LINCOS-Dominican Republic" at http://www.developmentgateway.org/node/603248/browser/?&sort_by=title

Functional Dependency and Other Related Dependencies

Robert A. Schultz
Woodbury University, USA

INTRODUCTION

The invention and rise of the relational database starting in the 1960s was accompanied by the remarkable development of canonical design techniques, making it possible to avoid many database designs with unintended bad consequences (Codd, 1970, 1971). These techniques, called "normal forms," prevent the occurrence of what are called "anomalies." If a database design contains an anomaly, the implemented database will behave in unintended ways. If there is a deletion anomaly, data will unexpectedly disappear; an insertion anomaly will produce difficulties in adding data; and a modification anomaly will result in extra and unexpected operations in the course of changing data in the database (Kroenke, 2002).

In a relational database, data is organized into tables; tables are composed of records; and records are made up of fields (or data items or attributes). One or more fields or data items, called keys, are used to locate records and the remaining data items in them. A primary key uniquely identifies each record. If more than one field is needed to guarantee unique identification, the primary key is called a concatenated key or composite key. When more than one combination of data items or fields could serve as a primary key, the keys not actively used are called candidate keys.

The following (partial) table contains potential examples of all three anomalies:

- *Deletion anomaly*: If all the instructors who teach a given course are deleted, we will lose information about the course name.
- *Insertion anomaly*: We cannot input information about a course or register a student for the course unless we have an instructor for the course.
- *Modification anomaly*: In order to change the name of a course, we have to go through the records of all the instructors in the table and change the name there.

Intuitively, it is somewhat obvious that Instructor is the wrong field to choose as a key for finding information about classes. The advantage of the normal forms is that they provide a standard procedure for finding a database design that avoids the anomalies.

BACKGROUND

The normal forms are almost all defined in terms of the concept of dependency, and, in particular, the concept of functional dependency. Dependency in almost all cases is between data items (equivalently, fields or attributes); intuitively, one attribute A is functionally dependent on another attribute B if one needs to know the value of B in order to determine the value of A.

These are the most commonly found normal forms:

- **First Normal Form (1NF):** The table designs must have no repeating groups. That is, there are no fields with more than one distinct value within the same record. This normal form is automatically satisfied if a relational database is used, because data must be entered in a table with single values for each field. Primary Key 2 is an example if we were allowed to enter multiple values in a field.

To put this table into First Normal Form, we have to create a second table containing the repeating group and the original key. The two tables are as follows (with primary key underlined):

Course(<u>CourseID</u>,CourseName,Instructor,Rank) and
Course Student<u>(CourseID,StudentID,</u>StudentName,Grade)

(Primary Key 1)

Instructor	InstructorRank	CourseNumber	CourseName	StudentName	Grade
Dr. Wilson	Assistant Prof.	CI 110	Computer Apps	Rod Hudson	C
Mrs. Day	Associate Prof	EN 111	Freshman Comp	Alice Adams	B

(Primary Key 2)

CourseID	**CourseName**	**Instructor**	**Rank**	**Student ID**	**StudentName**	**Grade**
CI110.F03.sec1	Computer Apps	Dr. Wilson	Asst Prof	555-44-3323 430-22-1123 554-55-6689 5454-92-5587	Rod Hudson, Joan Crawford, Andy Bierce, Luigi Nono	A B A C
EN111.S04.sec 3	Freshman Comp	Mrs. Day	Full Prof	490-40-2221 554-55-6689	Alice Adams, Andy Bierce	D B

- **Second Normal Form (2NF):** The table designs are in 1NF, and there are no partial functional dependencies, that is, functional dependencies on part of a primary key.

In the example above, **StudentName** in the second table Course-Student is functionally dependent only on **StudentID**. So, it is a violation of Second Normal Form. We convert this table to 2NF by creating a new table with the partially dependent fields and the key they were partially dependent on. Now we have the following:

Course(CourseID,CourseName,Instructor,Rank)
Course-Student(CourseID,StudentID,Grade)
Student(StudentID,StudentName)

- **Third Normal Form (3NF):** The table designs are in 2NF, and there are no nonkey functional dependencies, that is, no functional dependencies on data items that are not keys.

Continuing the example, in the Course table, **Rank** is functionally dependent on **Instructor**, which is not a key. To produce 3NF, we remove the functionally dependent field(s) to another table and include the original nonkey field as a key. So, we get the following:

Instructor(Instructor,Rank)
Course(CourseID,CourseName,Instructor)
Course-Student(CourseID,StudentID,Grade)
Student(StudentID,StudentName)

- **Boyce–Codd Normal Form (BCNF):** The definition of BCNF is that the table designs are in 3NF and continue to remain so for all candidate keys.

Examples grow more complex for the remaining normal forms. It is not too hard to construct an example of BCNF. If we add fields to the Student table, say, which could also serve as primary keys such as **DriversLicense** or **CreditCardInfo,** then if one of these fields were made key, the Student table would no longer be in 3NF, because **StudentName** would now be dependent on a nonkey field. Basically, the result would be to once again remove **StudentName** to a separate table dependent only on **StudentID**, and to leave all the candidate keys by themselves in a separate table.

- **Fourth Normal Form (4NF):** The table designs are in BCNF, and there are no multivalued dependencies.

- **Fifth Normal Form (5NF):** The table designs are in 4NF, and there are no join dependencies.

- **Domain Key/Normal Form (DK/NF):** All constraints on values can be derived from domain and key dependencies.

As you will note, Second and Third Normal Forms explicitly mention functional dependencies, and the Boyce–Codd Normal Form extends the application of the Third Normal Form to other choices of primary key. Fourth Normal Form uses a more complex kind of dependency defined in terms of functional dependency. The remaining two, Fifth Normal Form and Domain Key Normal Form, introduce somewhat different kinds of dependencies. The last two Normal Forms are not often met with in practice. Near the end, I will comment on the different forms of dependency these Normal Forms employ. For the commonly employed Normal Forms, functional dependency is the key concept.

UNDERSTANDING FUNCTIONAL DEPENDENCY

There are two different ways of defining functional dependency: an intuitive intensional way and a precise extensional way. Intensional definitions use psychological or meaning elements involving dependencies in knowledge—for example, we need to know customer name in order to know customer address. Extensional definitions are based solely on differences in the way data appear in the table—notably, patterns of repetition of field values (Lewis & Langford, 1959).

Thus, intension has to do with the meanings of the terms involved, and extension with what objects are denoted by or referred to by the terms. In general, if two terms have the same intension or meaning, they have the same extension, but two terms can have the same extension but different meanings or intensions (Frege, 1892; Geach & Black, 1952). Intensions are difficult to use in mathematical or scientific or technical contexts, because we do not have clear criteria for the identity of intensions. But, in fact, functional dependency (and the normal forms using it) can be defined extensionally (Martin, 1983):

Data item B of record R is functionally dependent on data item A of record R if, at every instant of time, each value in A has no more than one value of B associated with it in record R.

All one needs to do is examine the pattern of repetitions in the record. If there are multiple values of *B* for one value of *A* in a record at any time, then, according to Martin's definition, *B* is not functionally dependent on *A*.

Many authors use intentional elements in their definitions. For example, Kroenke (2002), in his *Database Processing,* explicitly defines functional dependency using intensional terms:

…attribute Y is functionally dependent on attribute X if the value of X determines the value of Y. Stated differently, if we know the value of X, we can obtain the value of Y.

However, intensional elements may be almost indispensable shortcuts. If we actually had available all the data elements in a database, there would be no question about using extensional methods. But when the full (extensional) database is not available, we may need to turn to intensional elements to determine functional dependency and, hence, database design (Schultz, 2002).

OTHER DEPENDENCIES

Finally, there remain the "higher" normal forms, that is, Boyce–Codd Normal Form, 4th Normal Form, 5th Normal Form, and Domain Key Normal Form.

For Boyce–Codd Normal Form, we need to check whether every field on which other fields are functionally dependent is potentially a key (that is, has unique values for each record) and then check functional dependencies on these candidate keys. There are no new implications about dependency or functional dependency.

Fourth Normal Form adds the concept of multivalued dependency. One definition is as follows:

In a table R with columns A, B, and C, B is multivalued dependent on A if there is more than one value of B for a single value of A.

If *C* is also multivalued dependent on *A*, anomalies arise (mainly update anomalies). A set of tables is in Fourth Normal Form if there are no multivalued dependencies present. Thus, multivalued dependency is also clearly an extensional matter, having to do with whether or not there are certain repetitions of values in a table. So even though multivalued dependency is different from functional dependency, it can be treated in a similar way (Kroenke, 2002).

Typically, one encounters multivalued dependency when two independent attributes depend on the same attribute in the same table. Please see the table below.

This situation produces a modification anomaly. If Joan Crawford switches her major to Theater, we need to update at least three rows to reflect her activities. If she becomes a double English and Theater major, we need to add three rows. This cannot be done automatically by the database—special efforts or programming would be required.

A better solution would be to put the table into 4NF by splitting the two independent attributes into two tables, each with the original key:

Student-Major(<u>StudentName</u>,Major)
Student-AthleticActivity(<u>StudentName,</u> AthlActiv)

StudentName	Major	AthlActiv
Rod Hudson	IT	Weight room
Joan Crawford	English	Weight room
Joan Crawford	English	Tennis
Joan Crawford	English	Swimming

Fifth Normal Form uses the concept of join dependency rather than functional dependency. Join dependency states that any decomposition of tables must be able to be rejoined to recover the original table. Tables in 5NF are join dependent. Failure to achieve 5NF is normally uncovered when data are added to the decomposed tables. Then it is discovered that the rejoined tables contain spurious data. It seems more likely that intensional considerations will help with 5NF than the other normal forms, because join dependency cannot be determined by simply looking at the data. It is not clear whether anything other than an iterative trial and error procedure is possible for detecting join dependency. These considerations are extensional, because they depend only upon patterns of field values in joined and unjoined tables. But that does not make the process of determining Fifth Normal Form any easier (Kroenke, 2002).

There is some question whether 5NF is actually desirable in all circumstances. Examples are also a bit more complex than for the preceding normal forms, so I will refer the interested reader to C. J. Date's treatment (1994).

Domain key normal form (DK/NF) is different. If a database is in DK/NF, it is provable that no anomalies can occur. Whereas the other normal forms (First, Second, Third, Boyce–Codd, Fourth, and Fifth) were designed to avoid certain anomalies, and there is no guarantee that some further anomaly may pop up that is not prevented by these normal forms. Unfortunately, there is no effective procedure for putting a set of tables into DK/NF (Fagin, 1981; Quine, 1966). As David Kroenke (2002) put it, "Finding, or designing, DK/NF relations is more of an art than a science." Although the theoretical claims for DK/NF are probably unassailable, the practical difficulties in achieving DK/NF can probably not be ameliorated (Schultz, 2003).

Again, the complexity of DK/NF examples requires us to refer the interested reader to the sources mentioned in the last paragraph.

CONCLUSION

Database normalization is an accepted part of database design and will continue to be, through whatever enhancements are brought about by integration of more traditional design methods into object-oriented technology and other new technologies. Because functional dependency is a critical part of database normalization, it will continue to be necessary for system designers to be familiar with the concept.

REFERENCES

Codd, E. F. (1970). A relational model of data for large shared data banks. *Communications of the ACM, 13*(6).

Codd, E. F. (1971). Normalized data base structure: A brief tutorial. *IBM Research Report,* San Jose, California RJ935.

Date, C. J. (1994). *An introduction to database systems* (6th ed.). Reading, MA: Addison-Wesley.

Fagin, R. (1981). A normal form for relational databases that is based on domains and keys. *ACM Transactions on Database Systems, 6*(3), 387–415.

Frege, G. (1892). On sense and reference. In P. T. Geach & M. Black (1952) (Eds.), Translation of "Ueber Sinn und Bedeutung" *Zeitschrift fuer Philosophie und philosophische Kritik* 100 (pp. 25–50).

Geach, P. T., & Black, M. (Eds.). (1952). *Translations from the philosophical writings of Gottlob Frege* (pp. 56–78). Oxford: Blackwell.

Kroenke, D. (2002). *Database processing* (pp. 120–134). Upper Saddle River, NJ: Prentice Hall.

Lewis, C. I., & Langford, C. H. (1959). *Symbolic logic*. New York: Dover.

Martin, J. (1983). *Managing the database environment*. Englewood Cliffs, NJ: Prentice Hall.

Quine, W. V. (1966). Church's theorem on the decision problem. In *Selected logic papers* (pp. 212–219). New York: Random House.

Schultz, R. (2002). Understanding functional dependency. In Becker (Ed.), *2002 Proceedings of IRMA. Effective databases for text and document management* (2003; pp. 278–287). Hershey, PA: Idea Group Press.

Schultz, R. (2003). The availability of domain/key normal form. In *2002 Proceedings of IRMA International Conference*. Hershey, PA: Idea Group Publishing.

KEY TERMS

Anomalies: These are unwanted consequences of inadequate database design. Data can be unintentionally lost or be more difficult to add or modify than the designer intended.

Dependencies: These are relations between data items or fields, or, occasionally, tables. Detecting dependencies is the key to putting tables in the various normal forms and, hence, is the key to avoiding anomalies.

Domain Key/Normal Form (DK/NF): Rather than functional dependencies, DK/NF is based on domain dependencies and key dependencies only. Although it is provable that a set of tables in DK/NF avoids all anomalies, it is also provable that there is no procedure for producing this normal form.

Functional Dependencies: Intuitively, one attribute is functionally dependent on a second attribute when you need to know the value of the second in order to find out the value of the first. More precisely, each value of the second attribute has no more than one value of the first associated with it.

Join Dependency: Any decomposition of tables must be able to be rejoined to recover the original table. Tables in 5NF are join dependent.

Key: This is a data item or field used to locate records in a table. A *primary key* uniquely identifies each record. If more than one field is needed to guarantee unique identification, the primary key is called a *concatenated key* or *composite key*. When more than one combination of data items or fields could serve as a primary key, the keys not actively used are called *candidate keys*.

Multivalued Dependencies: These are relations between three attributes. In a table *R* with columns *A*, *B*, and *C*, *B* is multivalued dependent on *A* if there is more than one value of *B* for a single value of *A*. Multivalued dependences must be eliminated to produce Fourth Normal Form.

Normal Forms: These are canonical patterns for table design that enable designers to avoid anomalies. Occasionally, tables are "denormalized" to improve database performance.

Relational Database: Database management system (DBMS) type in which relations between tables are implemented only by means of connections in the data, rather than hard-coded in the database. Its main advantages are flexibility and maintainability. Current popular examples include Oracle and SQL Server. Increased system performance made it the preferred DBMS starting in the 1990s.

Relational Database Structure: Information is organized into *tables*; tables are composed of *records*; and records are made up of *fields* (also known as *data items* or *attributes*).

Functional Integration of Decision Making Support

Guisseppi A. Forgionne
University of Maryland Baltimore County, USA

INTRODUCTION

Because of the importance to individual, group, and organizational success, information systems research has examined ways to improve support for decision making for the last three decades. The research has generated a variety of information systems to provide the necessary support. In the process, there has been an evolution from simple data access and reporting to complex analytical, creative, and artificially intelligent support for decision making (Holsapple & Whinston, 1996).

Various information systems have evolved to support the decision-making process. By studying the systems' characteristics, advantages, and disadvantages, researchers and practitioners can better design, develop, and implement robust decision-making support systems (Forgionne & Kohli, 1995; Kumar, 1999). This article facilitates such study by presenting and illustrating the underlying information system architectures for robust decision-making support.

BACKGROUND

Several frameworks have been developed to describe the human decision-making process. The most popular is Simon's three-phase paradigm of intelligence, design, and choice (Simon, 1960). This paradigm seems to be the most general, implying virtually all other proposed frameworks, and the Simon paradigm appears to have best withstood empirical testing (Borenstein, 1998; Martinsons, Davison & Tse, 1999). Such scrutiny, however, has suggested the expansion of the basic formulation to conclude with an implementation phase.

During the intelligence phase, the decision maker observes reality, gains a fundamental understanding of existing problems or new opportunities, and acquires the general quantitative and qualitative information needed to address the problems or opportunities. In the design phase, the decision maker develops a specific and precise model that can be used to systematically examine the discovered problem or opportunity. This model will consist of decision alternatives, uncontrollable events, criteria, and the symbolic or numerical relationships between these variables. Using the explicit models to logically evaluate the specified alternatives and to generate recommended actions constitutes the ensuing choice phase. During the subsequent implementation phase, the decision maker ponders the analyses and recommendations, weighs the consequences, gains sufficient confidence in the decision, develops an implementation plan, secures needed financial, human, and material resources, and puts the plan into action.

A variety of individual information systems have been offered to support during the phases and steps of the decision-making process (Mirchandi & Pakath, 1999; Sauter, 1997). Much can be learned about this support by examining the support offered by the individual systems.

ISSUES, CONTROVERSIES, AND PROBLEMS

Decision-making support has evolved over time and across disciplines (Mirchandani & Pakath, 1999). Initial support was offered by a decision support system (DSS). In the typical DSS, the decision maker utilizes computer technology to: (a) organize the data into problem parameters, (b) attach the parameters to a model, (c) use the model to simulate (experiment with) alternatives and events, and/or (d) find the best solution to the problem. Results are reported as parameter conditions (status reports), experimental forecasts, and/or recommended actions. Feedback from the user-controlled processing guides the decision maker to a problem solution, and created information and knowledge are stored as additional inputs for future or further processing.

The DSS concept presumes that the problem pertinent data and models have been created and made available to the system prior to user processing (Hooghiemstra, Kroon, Odijk, Salomon & Zwaneveld, 1999). It also assumes that the user can utilize the computer technology to perform the technical processing operations and computations required by the system (Lawrence & Sim, 1999). In fact, DSS users rarely have the technical skill to recognize, capture, and process pertinent data and models or to interpret the results of the models' processing within the problem context (Raghunathan, 1999). In short, the DSS concept offers little direct support for the intelligence, early design, and implementation phases of decision making.

To be useful for decision making, problem pertinent data must be identified, located, captured, stored, accessed, and interpreted (Seely & Targett, 1999). Data warehousing can be used to facilitate access and reporting, while data mining can help with the interpretation function. An executive information system (EIS) can deliver these data access, reporting, and interpretation functions to the decision maker in an intuitive and appealing manner.

In a typical EIS, the decision maker utilizes computer technology to: (a) organize the data into specified broad categories, (b) view (slice and dice) the data from interesting perspectives, (c) generate "warnings" for the decision maker by scanning current trends, and (d) mine the data for less obvious relationships. Results are reported as category summaries (status reports), sliced and diced details (drill down reports), and/or suggested problem parameters (events). Feedback from the user-controlled processing guides the decision maker to a general problem understanding, and the created parameters are stored as additional inputs for future or further processing.

The user should exit EIS processing with a general understanding of the problem or opportunity and with relevant problem information (such as general objectives, range of decision alternatives, and range of pertinent events). Additional decision analysis beyond EIS processing will be required to explicitly formulate the problem and complete the decision-making process. Put another way, an EIS directly supports only the intelligence phase of decision making.

Technical and domain expertise will be needed to recognize, formulate, and solve most complex and significant decision problems or opportunities. Although such expertise will be available within, and outside, an organization, the expertise may be difficult, costly, and time-consuming to locate, access, and utilize. The corresponding knowledge, however, can be acquired and embedded within a Knowledge Based System (KBS), and the system can be used to capture, store, and deliver the expertise to the decision maker (Ayyub, 2001). A typical KBS captures and stores as inputs problem pertinent knowledge, either from experts, cases, or other sources, and the models (inference engine or reasoning mechanisms) needed to draw problem solution inferences from the knowledge. In other words, a KBS directly supports some of the design and most of the choice phases of decision making. Specifically, a KBS facilitates problem structuring and the evaluation and selection of alternatives.

Since decision making is a sequential and continuous process, learning will be essential to the successful completion of the process. Users will learn from their interactions with a KBS (or other individual decision-making support system) and, in the process, gain skills that can be applied to future decision-making tasks. Applying learning to the solution of the current problem, however, often will require system support (Steiger, 1998). Machine learning systems (MLS) can provide such support by mimicking the learning processes of physical systems. In a typical MLS, the decision maker utilizes computer technology to: (a) organize the problem data, (b) structure (operationalize) the learning model, and (c) simulate learning. Results are reported as problem conditions (status reports), forecasted problem outcomes, and/or and an explanation of the learning logic.

Besides learning, creativity often is needed to successfully complete the decision-making process (Keys, 2000). While the previous systems free decision makers to

Table 1. Individual decision-making support systems

System	Type	Support
Decision Support System (DSS)	Individual	Specifying relationships between criteria, alternatives, and events; choice
Executive Information System (EIS)	Individual	Intelligence; developing decision criteria; identifying relevant uncontrollable events
Knowledge-Based System (KBS)	Individual	Develop decision alternatives; choice
Machine Learning System (MLS)	Individual	Logically evaluate decision alternatives
Creativity Enhancing System (CES)	Individual	Design; develop an implementation plan; put implementation plan into action

Table 2. Integrated decision-making support systems

System	Type	Support
Intelligent Decision Support System (IDSS)	Integrates the functions of DSS and KBS (and/or MLS)	Developing decision alternatives; specifying relationships between criteria, alternatives, and events; choice
Executive Support System (ESS)	Integrates the functions of DSS and EIS	Intelligence; developing decision criteria; identifying relevant uncontrollable events; specifying relationships between criteria, alternatives, and events; choice
Whole-Brained Decision Support System (WDSS) and Group Decision Support System (GDSS)	Integrate the functions of DSS and CES	Gain problem/opportunity understanding; design; choice
Management Support System (MSS)	Integrates the functions of DSS, EIS, and KBS (and/or MLS)	Intelligence; design; choice

concentrate on the creative aspects of decision making, they do not provide direct support for the creative process (Savransky, 2001). Since decision makers may not be inherently creative, support for creativity can considerably enhance their decision-making process. A creativity enhancing system (CES) offers such support (Forgionne, Clements & Newman, 1995). In a typical CES, the decision maker utilizes computer technology to: (a) organize (chiefly categorize and classify) the problem ideas and concepts, (b) structure ideas and concepts into problem elements and relationships, and (c) simulate conceptual problem solutions. Results are reported as problem elements (status reports), the problem's conceptual structure (criteria, alternatives, events, and relationships), and/or forecasted outcomes from the conceptual analyses.

The major individual systems, and their primary and direct support, are summarized in Table 1. An examination of this table shows that none of the individual systems offer complete and integrated support for all phases and steps of the decision-making process.

SOLUTIONS AND RECOMMENDATIONS

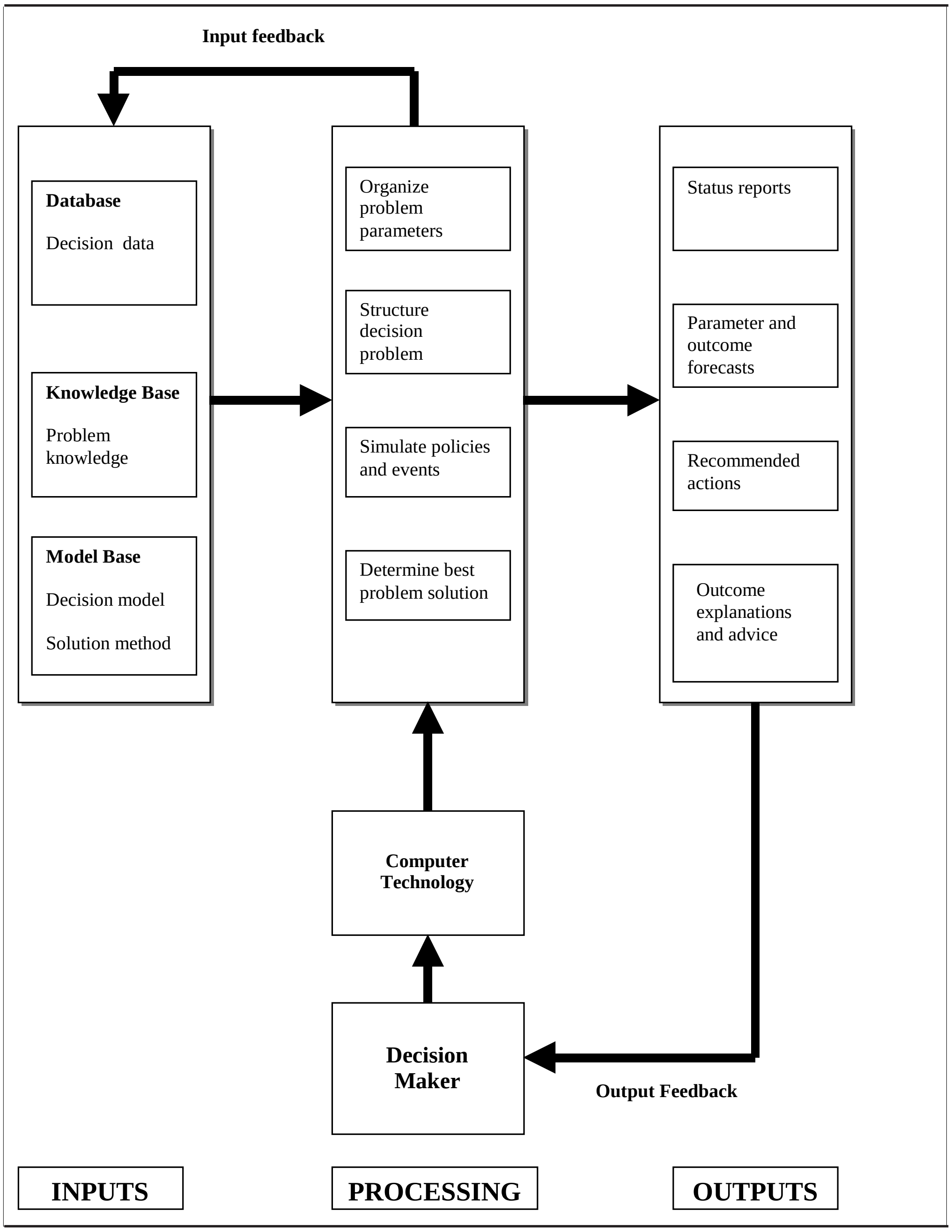
The need for complete and integrated decision-making support has encouraged researchers to seek the synergistic effects that can be achieved by combining the functionalities of the individual systems. The result has been the development of the various integrated systems for decision-making support summarized in Table 2.

As Table 2 indicates, each integrated system, such as an ESS or MSS, integrates the functionality of particular individual systems to provide decision-making support. While the integrated functionality has created more complete and unified decision-making support, the suggested synergies still leave significant gaps in decision-making support. For example, an IDSS still leaves gaps in design support, while an MSS does not provide creativity support. With even more system choices available than previously, the decision maker, and/or staff, are forced to match the relevant functionality with his/her/their decision-making support needs. Decision makers, and/or staff, may be ill equipped to make these selections and design, build, and implement the desired system.

An alternative strategy is to create one decision-making support system that synthesizes the main features and functions for the various decision-making support systems. The decision technology system (DTS), which is presented in Figure 1, has been proposed to support this alternative strategy. As Figure 1 illustrates, a DTS has a database, knowledge base, and model base. The database contains the data directly relevant to the decision problem, including the values for the uncontrollable events, decision alternatives, and decision criteria. The knowledge base holds problem knowledge, such as formulas for converting available data into the problem's parameters, guidance for selecting decision alternatives and problem relationships, or advice in interpreting possible outcomes. The model base is a repository for the formal (tabular, graphic, conceptual, or mathematical) models of the decision problem and the methodology for developing results (simulations and solutions) from the formal models.

Decision makers utilize computer technology (hardware and software) to process the inputs into problem-relevant outputs. Processing will involve:

Figure 1. Decision technology system (DTS)



(a) organizing problem parameters – accessing the database, extracting the decision data, and organizing the information in the form needed by the solution model and methodology;

(b) structuring the decision problem – accessing the model base, retrieving the appropriate decision model, and operationalizing (attaching organized parameters to) the decision model;

(c) simulating policies and events – using the operationalized decision model to perform the computations needed to simulate outcomes from user-specified alternatives and then identifying the alternative (or alternatives) that best meets the decision criterion (or criteria) among those tested;

(d) finding the best problem solution – accessing the model base, retrieving the appropriate solution method, and using the retrieved method to systematically determine the alternative (or alternatives), among all possible alternatives, that best meets the decision criterion (or criteria).

The DTS can use problem ideas, concepts, and knowledge drawn from the knowledge base to assist users in performing these processing tasks.

Processing will generate status reports, forecasts, recommendations, and explanations. The status reports will identify relevant uncontrollable events, decision alternatives, and decision criteria and show the current values for these problem elements. Forecasts will report the events and alternatives specified in the simulations and the resulting projected values of the decision criteria. The recommendations will suggest the values for the decision alternatives that best meet the decision criteria, and the corresponding criteria values, under current and forecasted values for the uncontrollable events. Explanations will justify the recommendations and offer advice on further processing. Such advice may include suggestions on interpreting the output and guidance for examining additional scenarios.

Input feedback from the processing provides additional data, knowledge, and models that may be useful for future decision making. Output feedback (which can include outcomes, cognitive information, task models, and what-if, goal-seeking, and other types of sensitivity analyses) is used to extend or modify the original analyses and evaluations (Sengupta & Abdel-Hamid, 1993).

Figure 1's general DTS architecture can support all phases of the decision-making process in a complete, integrated, and continuous manner. Critical problem data can be captured in a DTS database, and the system can be used to organize this captured information, generate timely focused reports, and project trends. Such processing helps the decision maker to quickly monitor the decision environment, set objectives, and evaluate the processed information for opportunities or problems, thereby supporting the intelligence phase of decision making.

The DTS, augmented by the managers' (or perhaps staff's) insights and judgments, can be used to process captured constructions and frameworks into criteria, events, and alternatives needed to formulate a model of the decision problem. Additional processing with the captured statistical methodologies can estimate the parameters required to operationalize the formulated decision problem model, thereby supporting the design phase of decision making. The formulated models, again augmented by the managers' insights and judgments, are used to evaluate alternatives in a systematic and analytic fashion and to recommend alternatives, thereby supporting the choice phase of decision making.

Decision technology systems (DTS) can provide the analyses in vivid detail with tables, graphs, and other supporting material. Such supporting material will increase the decision maker's confidence in the recommendations, improve the decision maker's perception of support system effectiveness, and enable the decision maker to better explain, justify, and communicate the decisions during implementation, thereby supporting the implementation phase of decision making.

Along with the original analyses and evaluations, DTS feedback loops increase the users' confidence in the recommendations and enable the decision maker to better explain, justify, and communicate the decisions during implementation. The loops also support decision making in a continuous and dynamic manner (Forgionne, 1999, 2000).

FUTURE TRENDS

Realizing the DTS promise presents significant technical and management challenges. Problem pertinent data, models, and knowledge must be identified, located, retrieved, and captured (Balasubramanian, Nochur, Henderson & Kwan, 1999). Intelligent data warehousing and mining can support the data retrieval tasks, and it may be possible to adapt these methodologies for model and knowledge retrieval support. Differences in data, knowledge, and model structures, however, may necessitate the development of new methodologies for knowledge and model retrieval tasks.

Also, it will be challenging to collect and deliver the tools and to manage the design, development, and implementation effort. Agents and object-oriented methods can be used to capture the tools and make them available for system operation (Siskos & Spyridakos, (1999). The resulting system, however, will profoundly challenge the nature of the decision maker's work as well as altering the structure of the organization. By providing complete and integrated decision-making support, a DTS will enable the decision maker to perform technical tasks previously outsourced to specialists. The result may be an organization with fewer hierarchical levels and smaller staff.

CONCLUSION

Over the years, support for decision making has taken a variety of forms. As the forms have evolved, decision-making support has become more comprehensive and integrated. Today, there are many system choices available, and matching the appropriate system to the particular problem or opportunity has created a new task for management.

The evolution has illustrated the synergistic value that can be achieved through higher levels of functional integration. This article has presented a concept, the DTS, which can offer a mechanism to consolidate the advances and promote a revolution in management. The proposed system has also created significant research opportunities – determining the best integration strategy, identifying the best design and development tools to achieve the strategy, and examining the impact of integrated decision support on management, decision making, and organizational structure, among others. It also clarifies the needs to: (a) have effective and efficient information reporting and communication systems in place and (b) to integrate the decision-making support systems with information reporting and communication systems.

REFERENCES

Ayyub, B.M. (2001). *Elicitation of expert opinions for uncertainty and risks.* Andover, UK: CRC Press.

Balasubramanian, P., Nochur, K., Henderson, J.C., & Kwan, M.M. (1999). Managing process knowledge for decision support. *Decision Support Systems, 27*(1-2), 145-162.

Borenstein, D. (1998). Towards a practical method to validate decision support systems. *Decision Support Systems, 23*(3), 227-239.

Forgionne, G.A. (1999). An AHP model of DSS effectiveness. *European Journal of Information Systems, 8,* 95-106.

Forgionne, G.A. (2000). Decision-making support system effectiveness: The process to outcome link. *Information Knowledge Systems Management, 2*(2), 169-188.

Forgionne, G.A., Clements, J.P., & Newman, J. (1995). Qualitative thinking support systems (QTSS). *Journal of Decision Systems, 4*(2), 103-137.

Forgionne, G.A., & Kohli, R. (1995). Integrated MSS effects: An empirical health care investigation. *Information Processing and Management, 31*(6), 879-896.

Holsapple, C.W., & Whinston, A.B. (1996). *Decision support systems: A knowledge-based approach.* New York: ITP.

Hooghiemstra, J.S., Kroon, L.G., Odijk, M.A., Salomon, M., & Zwaneveld, P.J. (1999). Decision support systems support the search for win-win solutions in railway network design. *Interfaces, 29*(2), 15-32.

Keys, P. (2000). Creativity, design and style in MS/OR. *Omega, 28*(3), 303-312.

Kumar, R.L. (1999). Understanding DSS value: An options perspective. *Omega, 27*(3), 295-304.

Lawrence, M., & Sim, W. (1999). Prototyping a financial DSS. *Omega, 27*(4), 445-450.

Martinsons, M., Davison, R., & Tse, D. (1999). The balanced scorecard: A foundation for the strategic management of information systems. *Decision Support Systems, 25*(1), 71-88.

Mirchandani, D., & Pakath, R. (1999). Four models for a decision support system. *Information & Management, 35*(1), 31-42.

Raghunathan, S. (1999). Impact of information quality and decision-maker quality on decision quality: A theoretical model and simulation analysis. *Decision Support Systems, 4*(26), 275-286.

Sauter, V. (1997). *Decision support systems.* New York: Wiley.

Savransky, S.D. (2001). *Engineering of creativity: Introduction to TRIZ methodology of inventive problem solving.* Andover, UK: CRC Press.

Seely, M., & Targett, D. (1999). Patterns of senior executives' personal use of computers. *Information & Management, 35*(6), 315-330.

Simon, H. (1960). *The new science of management decision.* New York: Harper and Row.

Siskos, Y., & Spyridakos, A. (1999). Intelligent multiple criteria decision support: Overview and perspectives. *European Journal of Operational Research, 113*(2), 236-246.

Steiger, D.M. (1998). Enhancing user understanding in a decision support system: A theoretical basis and framework. *Journal of Management Information Systems, 15*(2), 199-220.

KEY TERMS

Artificially Intelligent Systems: Information systems that help users manage data and models by delivering virtual expertise and other forms of artificial intelligence in support of these tasks.

Creativity Enhancing System: Information systems that are designed to offer creative tools that help users formulate problems and perform other creative tasks in decision making.

Decision Making Process: The process of developing a general problem understanding, formulating the problem explicitly, evaluating alternatives systematically, and implementing the choice.

Decision Support Systems: Information systems that interactively support the user's ability to evaluate decision alternatives and develop a recommended decision.

Decision Technology System: An information system that is designed to support all phases of the decision-making process in a complete and integrated manner.

Executive Information Systems: Information systems that access, report, and help users interpret problem pertinent information.

Integrated Decision-Making Support Systems: Information systems that integrate the functions of one or more individual (stand alone) decision-making support systems.

Knowledge-Based Systems: Information systems that capture and deliver problem pertinent knowledge to users.

Machine Learning System: Information systems that mimic the human learning process and deliver the knowledge to users.

F

Functionalities and Position of Manufacturing Execution Systems

Vladimír Modrák
Technical University of Košice, Slovakia

Introduction

Efforts to separate unequivocally substantial signs of versatile tools of manufacturing management are usually marked by a narrowed view of the field of their use. Similarly, it is so also in the case of specifying the functionality and position of MES (Manufacturing Execution Systems) in the hierarchy of information systems. Presentations generalising MES in this field do not always correspond with models that have a generic character. For that reason it appears useful to investigate the mentioned MES characteristics from a number of angles, and particularly in relation with the basic types of manufacturing systems.

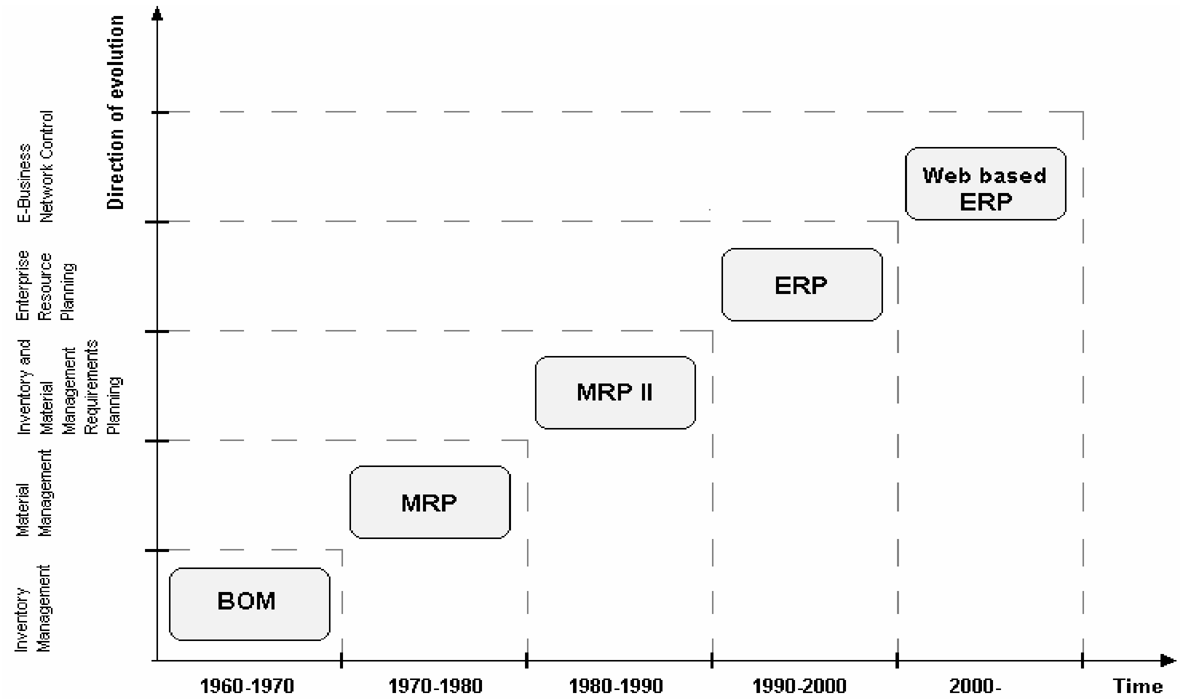
BACKGROUND OF MES EVOLUTION

From a historical point of view, the infiltration of information technology into manufacturing technology was conditioned by the development and advancement of host mainframe computing in the 1950s and '60s. It gave manufacturers the ability to capture, manipulate, and share information and automate calculation and analysis in order to support design of increasingly complex and capable products. Simultaneously in the framework of manufacturing management, an inventory control took on great importance and most of the software in the 1960s was developed for this purpose. Typically, tools called BOM (bill of materials) processors, which were used as a means to represent process plans, handled inventory control. The focus shifted in the 1970s to Material Requirement Planning (MRP) as the complexity of manufacturing operations increased. This managerial instrument enabled financial managers to both view and control their business processes much more closely. The tools to automate business processes were enhanced by adding further functionalities to meet the increased requirements. Subsequently in the 1980s, the term Manufacturing Resources Planning (MRP II) became popular. An MRP II presented extension of MRP functions to integrate all aspects of the planning and control of the personnel, materials and machines (Kimble & McLoughlin, 1995). Following, solutions that are marked by the acronym ERP (Enterprise Resource planning) were performed in the early 1990s. An ERP system can be defined as an integrated information processing system supporting various business processes such as finance, distribution, human resources and manufacturing (Choi & Kim, 2002). The newest version of ERP II has been much publicized by the Gartner Group. Fundamentally, ERP II signals a shift in traditional ERP applications from focusing on internal data gathering and management process information to partners, vendors and customers externally via the Web (Farver, 2002). The overall view on evolution of ERP systems is shown in Figure 1. Initially this concept attained a huge popularity among manufacturers, but as the scope of managed systems increased, the ERP system was not suitable for controlling activities on the shop floor level. For this purpose, new tool of manufacturing management called "Manufacturing Execution Systems" was evolved and utilized during the 1990s. There are more interpretations of MES depending on different manufacturing conditions, but the common characteristic to all is that an MES aims to provide an interface between an ERP system and shop floor controllers by supporting various "execution" activities such as scheduling, order release, quality control, and data acquisition (MESA 6, 1997).

VIEW ON MES FUNCTIONALITIES

A concept of Manufacturing Execution Systems is one of several major information systems types aimed at manufacturing companies. MES can be in simple way also defined as a "tool for manufacturing management". The functions of an MES range from operation scheduling to production genealogy, labor and maintenance management, performance analysis, and to other functions in between. There are several general models of typical MES functions that are principally divided into core and support functions. The core functions deal primarily with actual management of the work orders and the manufacturing resources. Other functional capabilities of MES may be required to cover support aspects of the manufacturing operations. According to McClellan (1997), the function parts pertaining to first group of functions include:

Figure 1: Functionality evolutions of ERP systems



- Planning system interface
- Work order Management
- Workstation management
- Material movement management
- Data collection
- Exception management
- Inventory/materials.

The same author describes support functions as open systems and simultaneously gives a picture of which other functions the MES should include:

- Maintenance management
- Time and attendance
- Statistical Process Control
- Quality assurance
- Process data/performance analysis
- Documentation/product data management
- Genealogy/product trace-ability
- Supplier management

Alike quantity of core and support functions along with some identical components are identified in other concepts (Kisiel, 2001) that are depicted in Figure 2.

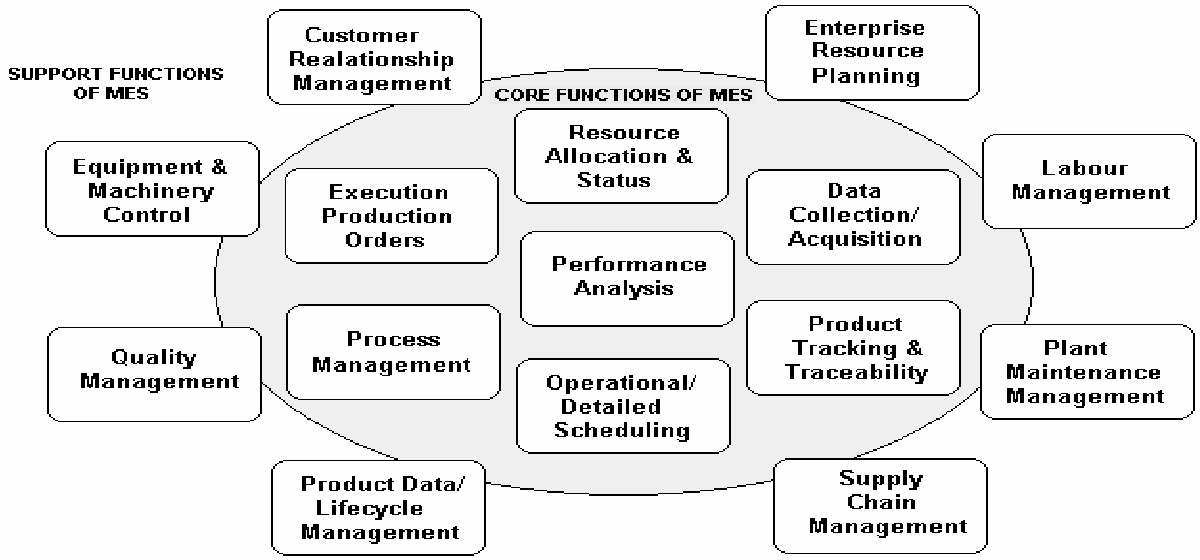
MESA International presents another attitude to MES functionalities that is more-or-less based on the assumption of profitability to begin to deal with wider model of basic elements to ensure incorporating all-important functions into MES (MESA 2,1997). Accordingly, MES would include functionalities such as:

1. Resource Allocation and Status
2. Operations/Details Scheduling
3. Dispatching production Units
4. Document Control
5. Data Collection/Acquisition
6. Labor management
7. Quality Management
8. Process Management
9. Maintenance Management
10. Product Tracking and Genealogy
11. Performance Analysis

A point of debate about MES functionalities also is connected with different types of manufacturing. Commonly, manufacturing can be divided into three types (Grover, 1987):

- Job Shop Production. The manufacturing lot sizes are small, often one of a kind.
- Batch Production. This category involves the manufacture of medium-sized lots of the same item of product. This type is called also discrete manufacturing.
- Mass Production. This is the continuous specialized manufacture of identical products.

Figure 2. A MES functionality model



Understandably, from an automation point of view a discrete manufacturing presents a much more complicated concept comprising of various technologies that are used to integrate manufacturing system to one another. One of specific models of MES functions is aimed to a typical FMS (flexible manufacturing system) consisting of numerical controlled machining centers, automated handling systems for jobs and tools, automated storage/retrieval systems, auxiliary processing facilities, and set-up stations. Choi and Kim (2002) propose a two-tier MES architecture suitable for bridging the gap between an FMS controller and ERP system. The two-tier MES consists of a main-MES in charge of main shop-floor operations and an FMS-MES in charge of FMS operations. The main MES is connected to the ERP system, and the FMS-MES is connected to the FMS controller. Then on the basis of IDEFO (Information Definition for Data Modeling), the overall structure of shop floor operations required for main-MES is grouped into the following functions:

- Customer Inquire Handling
- Received Order Handling
- Load/Process Control
- Engineering Change Orders Handling

and structure of FMS operations pertaining to FMS-MES functions include:

- Retrieve detailed machining-process plans for the jobs
- Generate FMS machining schedule and send out NC-file transfer instructions and tool preparation instructions
- Download NC files and send out set-up instructions
- Tool presetting and installation
- Perform machining operations while collecting data

Outlined structures of shop floor operations and FMS operations, present methodical approaches to specifying MES functionalities. In this connection, functional requirements of MES may be identified without functional redundancy. Obviously, the scope of operations or functions depends on number of subsystems, but the key functions remain unchanging in their essence. Because, there are no reference MES models that can be used for general manufacturing environments, overcoming of this aspect leads through the presentations of sample solutions by types of environment and other criterions. For example, modeling three different management systems for maintenance, quality and production (Brandl, 2002) based on the new S95 standard of ISA (ISA, 2000).

AGGREGATIVE ROLE OF MES

MES has been developed and used as the interface between ERP and automation systems, since it was generally recognized that ERP systems weren't scalable.Because batch production presents the most intricate manufacturing concept and constitutes a significant portion of total manufacturing activities, then discrete manufacturing may be considered to be a substantive area of MES applications. The mentioned and other ISA standards, significantly help in the implementation of integrated manufacturing systems. It aims to model integration of business systems like ERP with control systems like DCS and SCADA. The standard S88.01 (ISA, 1995) has been developed to support batch control level optimization. It provides standard models and terminology for the design and operation of batch control systems. At the control level, the key attribute is the integration of all process information into one place. Ordinarily used for this purpose, are both a programmable logic controllers and SCADA software (Supervisory Control and Data Acquisition) that is positioned on the top of hardware to which it is interfaced, in general via PLCs or other hardware modules (SCADA).

An Important function of MES is to provide a feedback to ERP with aim to adjust their scheduling data and algorithms in more realistic manner. In this connection there is necessary to take in consideration of certain aspects of incompatibility. Originally, ERP systems have been developed more for financial managers then manufacturing managers. Moreover, the data flow frequency in ERP system is lower then at the MES level. By analogy, it would not make sense to feed a quantity of output data available at shop floor level into ERP system and a reversal. An effort to do so often involved large amount of manual data entry. Therefore, one important role of optimization by MES may be seen in reduction of manual data entry on the boundaries of information system layers. Basically, it's relevant in this case when enterprise information and control systems are designed as one complex. Then expected reductions of manual activities may be achieved by the integration of system layers that are specified in the concepts of the reference model of Open System Interconnection (OSI). Management functions by this model usually include activities, which are common for several layers, but they are not supported by the services provided by a specific layer. For modeling layer-to-layer interactions, with aim to integrate the bordering layers, it is also appropriate to analyze of manufacturing systems from the viewpoint its three main categories. These are:

- The physical plant
- The formal control/information system
- Human beings working in the system

The first two are deterministic and mechanistic. They are designed to be controllable and predictable. Human beings working in the systems exhibits humanistic traits that are quite different from the technical systems they interact with. This manufacturing system classification helps to create object-oriented models on the highest modeling level that are enabling to utilize advanced technologies as object-oriented data management and others.

Supplementary integration between MES and ERP is seldom uncomplicated. One manner to get around presumable complications is by execution business process modeling. The sense of this kind of modeling is to make such new organization of work more feasible than in the past. That is why process modeling and improvement becomes a key issue of successful implementation of any MES strategy. Business process improvement especially is the fundamental way to stay competitive in the market. Of particular importance in business process improvement, is the way of the collected measurements are analyzed. Both operational and structural measurements are needed to properly benchmark business processes.

FUTURE TRENDS

In general, the research of MES is still in its beginning stages. The expected role of the next research can be seen in the development of fully integrated systems of manufacturing management. It also includes information integration of manufacturing activities with external entities, what will lead to web-based solutions of MES. From this perspective, the future trend clearly calls for overall information control systems convergence.

Many other research initiatives with diverse objectives can be identified in specific areas of MES development. One of them lies in development of so-called *component-based application framework* (Füricht et al. 2002; Boyle, 2000), which might more easily compose MES application. This direction of improvement is oriented on the development of high-level modeling methods respecting requirements for functional openness of the framework approach.

CONCLUSION

In reality, majority enterprise information and control systems on different levels are developed and operated on incompatible technologies and based on heterogeneous architectures. Additionally, they are ordinarily in different stages of their own lifecycles and have varied levels of openness. A supplementary integration of separate information systems is rather complicated when taking into consideration the lack of standard interfaces. A plant–wide system, built according to the requirements of each facility, is needed in the situation of an absence of enterprise information and control systems. In this sense, a frequent interest of manufacturers concerns a balanced scale of MES functionalities. As mentioned earlier, it depends on more factors. For instance, when an existing ERP system contains factory floor control functionality, then the functionality model of MES has only supplement character. In this sense, one of expected direction of MES evolution that could help to overcome outlined topical issues is "fusion" of MES and ERP systems. Hopefully it will reduce implementation difficulties of information system integration, but it probably will not mean the elimination of MES level.

REFERENCES

Brandl, D. (2002, October). Making sense of the MES at the MES layer. *ISA Technical Conference*. Chicago IL.

Boyle, W. (2000). Distributed manufacturing execution framework. In M. Fayad & R. Johnson (Eds.). *Domain-specific application frameworks*,121-138.

Choi, B.K. & Kim, B.H. (2002). MES architecture for FMS compatible to ERP. *International Journal of Computer Integrated Manufacturing, 15*(3), 274-284.

Daneels, A. & Salter, W. (1999). What is SCADA? *International Conference on Accelerator and Large Experimental Physic Control Systems*, Trieste, Italy, 339-343.

Falco, J., Stouffer, K., Wavering, A. & Proctor, F (2004). IT security for industrial control systems. National Institute of Standards and Technology, Gaithersburg, USA, 1-16. Retrieved from *http://www.isd.mel.nist.gov/documents/falco/ITSecurityProcess.pdf*

Farver, D. (2002). 2 ERP or ERP 2. *Journal of Business Innovation*. Retrieved from *http://www.agilebrain.com*

Füricht, R., Prahofer, H., Hofinger, T. & Altmann, J. (2002). A component-based application framework for manufacturing execution systems in C# and .NET. In *Proceedings of the 40th International Conference on Technology of Object-Oriented Languages and Systems*, Sydney: Australian Computer Society

Groover, M.P. (2001). *Automation, production systems, and computer-integrated manufacturing* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.

International Society for Measurement and Control (ISA). (1995). Batch Control Part 1: Models and terminology, International Society for Measurement and Control (NSI/ ISA S88.01)

International Society for Measurement and Control (ISA), RTP (2000). Enterprise-Control System Integration- Part 3: Models of Manufacturing Operations, Draft 7. (ANS/ ISA-95.00.03). North Carolina, USA.

Kimble, C. & McLoughlin, K. (1995, March). Computer-based information systems and managers' work. *New technology, Work and Employment, 10*(1), 56-67.

Kisiel, T. (2001). Manufacturing Execution Systems: Steps to a Successful Implementation. DAI White Paper, Digital Applications International, London, United Kingdom.

Manufacturing Execution Systems Association (MESA). (1997). *MES explained: A high-level vision*. (White Paper 6, MESA #6). Pittsburgh, P.A.

Manufacturing Execution Systems Association (MESA). (1997). *MES functionalities and MRP to MES data flow possibilities*. (MESA #2, White Paper 2). Pittsburgh, P.A. (Originally published March 1977)

McClellan, M. (1997, August). *Applying manufacturing execution systems*. Saint Lucie Press.

Object Management Group. (1997). *Manufacturing domain task force RFI-3 manufacturing execution system (MES)*. (OMG Document mfg/97- 11-01). Framingham, MA, USA.

Wallace, E., Barkmyer, E., Denno, P., Feng, S., & Jones, A. (1999). NITS response to MES request for information. *NISTIR 6397, National Institute of Standards and Technology, Gaithersburg, USA,* 1-1.

KEY TERMS

Distributed Control System (DCS): A supervisory control system typically controls and monitors set points to sub-controllers distributed geographically throughout a factory (Falco, Stouffer, Wavering & Proctor, 2004).

Enterprise Resources Planning: consists of those systems that provide financial, order management, production and materials planning, and related functions. The modern ERP systems focus on global planning, business processes and execution across the whole enterprise (intra-enterprise systems), with an accrued recent importance of aspects like supply chain planning and the whole supply chain management aspects and extending to include the whole inter-enterprise supply chain (OMG, 1997).

Manufacturing Execution System: A collection of hardware/software components that enables the management and optimization of production activities from order launch to finished goods. While maintaining current and accurate data, an MES guides, initiates, responds to and reports on plant activities as they occur. An MES provides mission-critical information about production activities to decision support processes across the enterprise (Wallace, Barkmyer, Denno, Feng, and Jones,1999).

Manufacturing Resources Planning II: is a suite of sophisticated programs that attempt to integrate all aspects of the planning and control of the personnel, materials and machines required to manufacture a range of products in one comprehensive computer based system (Kimble & McLoughlin, 1995).

Programmable Logic Controller (PLC): A small industrial computer used in factories originally designed to replace relay logic of a process control system and has evolved into a controller having the functionality of a process controller. (Falco, Stouffer, Wavering & Proctor, 2004).

Scalability: is understood as the possibility to extend the SCADA based control system by adding more process variables, more specialized servers (e.g. for alarm handling) or more clients. The products achieve scalability by having multiple data servers connected to multiple controllers (Daneels & Salter,1999).

Supervisory Control and Data Acquisition System (SCADA): It is a part of control system, but rather focused on the supervisory level. As such , it is a purely software package that is positioned on top of hardware to which it is interfaced, in general via Programmable logic Controllers (PLCs) or other commercial hardware modules (Daneels & Salter,1999).

F

Fundamentals of Multirate Systems

Gordana Jovanovic-Dolecek
National Institute for Astrophysics, Optics and Electronics, Puebla, Mexico

INTRODUCTION

Digital signal processing (DSP) is an area of science and engineering that has been rapidly developed over the past years. This rapid development is a result of significant advances in digital computers technology and integrated circuits fabrication (Elali, 2003; Grover & Deller, 1999; Mitra, 2001; Oppenheim & Schafer, 1999; Smith, 2002; Stein, 2000; White, 2000).

Classical digital signal processing structures belong to the class of single-rate systems since the sampling rates at all points of the system are the same.

The process of converting a signal from a given rate to a different rate is called sampling rate conversion. Systems that employ multiple sampling rates in the processing of digital signals are called multirate digital signal processing systems. Sample rate conversion is one of the main operations in a multirate system (Stearns, 2002).

BACKGROUND

Decimation

The reduction of a sampling rate is called decimation, because the original sample set is reduced (decimated). Decimation consists of two stages: filtering and downsampling, as shown in Figure 1. The discrete input signal is *u*(*n*) and the signal after filtering is *x(n)*. Both signals have the same input sampling rate f_i.

Downsampling reduces the input sampling rate f_i by an integer factor *M*, which is known as a downsampling factor. Thus, the output discrete signal *y(m)* has the sampling rate f_i/M. It is customary to use a box with a down-pointing arrow, followed by a downsampling factor as a symbol to represent downsampling, as shown in Figure 2.

The output signal *y(m)* is called a downsampled signal and is obtained by taking only every *M*-th sample of the input signal and discarding all others,

$$y(m) = x(mM) \tag{1}$$

The operation of downsampling is not invertible because it requires setting some of the samples to zero. In other words, we can not recover *x(n*) from *y(m)* exactly, but can only compute an approximate value.

In spectral domain downsampling introduces the repeated replicas of the original spectrum at every $2\pi/M$. If the original signal is not bandlimited to $2\pi/M$, the replicas will overlap. This overlapping effect is called aliasing. In order to avoid aliasing, it is necessary to limit the spectrum of the signal before downsampling to below $2\pi/M$. This is why a lowpass digital filter (from Figure 1) precedes the downsampler. This filter is called a decimation or antialiasing filter.

Three useful identities summarize the important properties associated with downsampling (Dabrowski, 1997; Jovanovic-Dolecek, 2002). The first identity states that the sum of the scaled, individually downsampled signals is the same as the downsampled sum of these signals. This property follows directly from the principle of the superposition (linearity of operation). The second identity establishes that a delay of *M* samples before the downsampler is equivalent to a delay of one sample after the downsampler, where *M* is the downsampling factor. The third identity states that the filtering by the expanded filter followed by downsampling, is equivalent to having downsampling first, followed by the filtering with the original filter, where the expanded filter is obtained by replacing each delay of the original filter with *M* delays. In the time domain, this is equivalent to inserting *M*-1 zeros between the consecutive samples of the impulse response.

Figure 1. Decimation

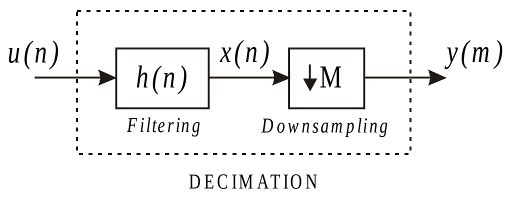
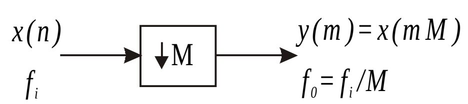


Figure 2. Downsampling



The polyphase decimation, which utilizes polyphase components of a decimation filter, is a preferred structure for decimation, because it enables filtering to be performed at a lower sampling rate (Crochiere & Rabiner, 1996; Diniz, da Silva & Netto, 2002).

Interpolorization

The procedure of increasing the sampling rate is called interpolation, and it consists of two stages: upsampling and filtering (shown in Figure 3).

The upsampler increases the sampling rate by an integer factor L, by inserting L-1 equally spaced zeros between each pair of samples of the input signal *x(n)* as shown by,

$$y(n) = \begin{cases} x(n/L) & for \;\; n = mL \\ 0 & otherwise \end{cases} \tag{2}$$

where L is called interpolation factor.

As Figure 4 illustrates, the symbol for this operation is a box with an upward-pointing arrow, followed by the interpolation factor. We can notice that the input sampling rate f_i is increased L times.

The process of upsampling does not change the content of the input signal, and it only introduces the scaling of the time axis by a factor L. Consequently, the operation of upsampling (unlike downsampling) is invertible, or, in other words, it is possible to recover the input signal *x(m)* from samples of *y(n)* exactly.

The process of upsampling introduces the replicas of the main spectra at every $2\pi/L$. This is called imaging, since there are L-1 replicas (images) in 2π. In order to remove the unwanted image spectra, a lowpass filter must be placed immediately after upsampling (Figure 3). This filter is called an anti-imaging filter. In the time domain, the effect is that the zero-valued samples introduced by upsampler are filled with "interpolated" values. Because of this property, the filter is also called an interpolation filter.

We have already seen three useful identities of the downsampled signals, and now we will state the identities associated with upsampling. The fourth identity asserts that the output signal obtained by upsampling followed by scaling of the input signal will give the same result as if the signal is first scaled and then upsampled. The fifth identity states that a delay of one sample before upsampling is equivalent to the delay of L samples after upsampling. The sixth identity, which is a more general version of the fifth identity, states that filtering followed by upsampling is equivalent to having upsampling first followed by expanded filtering (Dabrowski, 1997; Diniz, da Silva & Netto, 2002; Jovanovic-Dolecek, 2002).

Cascade of Sampling Convertors

An interchange of cascaded sampling converters can often lead to a computationally more efficient realization (Fliedge, 2000; Vaidyanathan, 1993). If upsampling precedes downsampling, where both operations have the same factor, the signal is not changed. However, if downsampling is performed before upsampling, and both operations have the same factor, the resulting signal will be different from the input signal. Rational sampling conversion, that is, changing the sampling rate by a ratio of two integers, L/M can be efficiently performed as a cascade of upsampling and downsampling, where the interpolation and decimation filters are combined into one filter.

FUTURE TRENDS

Applications

Multirate systems have applications in digital radio, speech processing, telecommunications, wavelet transform, digital filtering, A/D converters, spectrum estimation, and so forth.

There are many applications where the signal of a given sampling rate needs to be converted into an equivalent signal with a different sampling rate. For example, in digital radio, three different sampling rates are used: 32 kHz in broadcasting, 44.1 kHz in digital compact disc (CD), and 48 kHz in digital audio tape (DAT) (Fliedge, 2000; Mitra, 2001). Conversion of the sampling rate of audio signals between these three different rates is often necessary. For example, if we wish to play CD music which has a rate of 44.1 kHz in a studio which operates at a 48 kHz rate,

Figure 3. Interpolation

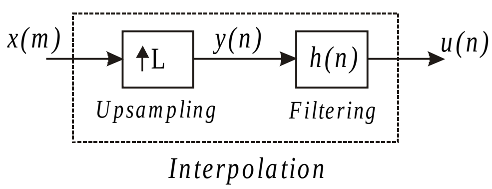
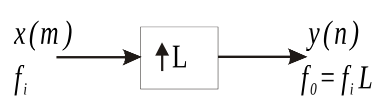


Figure 4. Interpolation



then the CD data rate must be increased to 48 kHz using a multirate technique.

In speech processing, multirate techniques are used to reduce the storage space or the transmission rate of speech data (Ifeachor & Jervis, 2001). In the past years, multirate speech and audio signal processing has been a research topic that has produced several efficient algorithms for echo and noise cancellation, active noise cancellation, speech enhancement, and so forth (Diniz, da Silva & Netto, 2002; Jovanovic-Dolecek, 2002; Smith, 2001).

An example of an application of multirate signal processing in telecommunications is the translation between two multiplexing formats, time division multiplexing (TDM) and frequency division multiplexing (FDM) (Fliedge, 2000; Harris, 2004; Smith, 2001). The filter banks are used to separate a signal into two or more signals (analysis filter bank) or to compose two or more signals into a single signal (synthesis filter bank). Multirate systems and filter banks have found many applications in source and channel coding, thereby providing a bridge between communication system design/analysis and signal processing tools (Vaidyanathan, 1993).

One of the most fascinating developments in the field of multirate signal processing has been the establishment of its link to the discrete wavelet transform (Diniz, da Silva, & Netto, 2002; Rao, 2002; Suter, 1998). This link has been responsible for the rapid application of wavelets in fields such as image compression.

Multirate processing has found important application in the efficient implementation of DSP functions (Ifeachor & Jervis, 2001). The multirate approach increases the computational speed, decreases the overall filter order, reduces word-length effects, decreases power consumption, making it vital for efficient filtering. Basic multirate techniques used to satisfy the filter requirements are the polyphase decomposition, multistage filtering and frequency masking approach.

The need for inexpensive, high resolution analog/digital (A/D) converters has led to the use of oversampling techniques in the design of such converters, that is, to sample the analog signal at a rate much higher than the Nyquist rate, one uses a fast low-resolution A/D converter, and then decimates the digital output of the converter to the Nyquist rate. Such A/D converter relaxes the sharp cutoff requirements of the analog anti-aliasing filter, resulting in a more efficient structure, (Jovanovic-Dolecek, 2002; Mitra, 2001).

Another application of multirate signal processing is in the area of spectrum estimation. The computational requirements for narrowband spectrum estimation based on discrete Fourier transform can be significantly reduced by using the sampling rate conversion (Vaidyanathan, 1993).

It is expected that in future the applications of multirate systems will play a very important role for resolving problems in communications, control, wavelet analysis, analog/digital converters and so forth. From the other side, the different applications will give an impulse to develop a new algorithms and design methods for multirate systems.

CONCLUSION

There are many applications where the signal of a given sampling rate needs to be converted into an equivalent signal with a different sampling rate. The main reasons could be to increase efficiency or simply to match digital signals that have different rates. During the past several years, the multirate processing of digital signals has been attracted by many researchers and "the utilization of multirate techniques is becoming an indispensable tool of the electrical engineering profession" (Suter, 1998).

Changing the sampling rate can reduce the computation necessary to complete some DSP operations, and thus reduce the overall system cost. Consequently, the main advantage of multirate systems lies in their high computational efficiency.

REFERENCES

Crochiere, R.E., & Rabiner, L.R. (1996). *Multirate digital signal processing*. Pearson Education POD.

Dabrowski, A. (1997). *Multirate and multiphase switched-capacitor circuits*. London: Chapman & Hall.

Diniz, P.S.R., da Silva, E.A.B., & Netto, S.L. (2002). *Digital signal processing, system analysis and design*. Cambridge: Cambridge University Press.

Elali T.S. (2003). *Discrete systems and digital signal processing with MATLAB.* CRC Press.

Fliedge, N.J. (2000). *Multirate digital signal processing*. New York: John Willey & Sons.

Grover, D., & Deller, J.R. (1999). *Digital signal processing and the microcontroller*. New Jersey: Prentice Hall, Inc.

Harris F., (2004). *Multirate signal processing for communication systems*. Prentice Hall PTR.

Ifeachor, E.C., & Jervis, B.E. (2001). *Digital signal processing: A practical approach* (2nd ed.). Prentice Hall.

Jovanovic-Dolecek, G. (Ed.) (2002). *Multirate systems: Design & applications.* Hershey, PA: Idea Group Publishing.

Mitra, S.K. (2001). *Digital signal processing: A computer-based approach*. New York: The McGraw-Hill Companies.

Oppenheim, A.V., & Schafer, R.W. (1999). *Discrete-time signal processing* (2nd ed.). New Jersey: Prentice-Hall, Inc.

Rao, R. (2002). Wavelet transforms and multirate filtering. In G. Jovanovic-Dolecek (Ed.), *Multirate systems: Design & applications* (pp.86-104). Hershey, PA: Idea Group Publishing.

Smith, D. (2001). *Digital signal processing technology: Essentials of the communications revolution*. Amer Radio Relay League.

Smith, S. (2002). *Digital signal processing: A practical guide for engineers and scientists*. Newnes.

Stearns, S.D. (2002). *Digital signal processing with examples in MATLAB*. CRC Press.

Stein, J. (2000). *Digital signal processing: A computer science perspective*. New York: Wiley- Interscience.

Suter, B.W. (1998). *Multirate and wavelet signal processing*. San Diego: Academic Press.

Vaidyanathan, P.P. (1993). *Multirate systems and filter banks*. New Jersey: Prentice Hall.

White, S. (2000). *Digital signal processing: A filtering approach*. Delmar Learning.

KEY TERMS

Analysis Filter Bank: Decomposes the input signal into a set of subband signals with each subband signal occupying a portion of the original frequency band.

Continuous-Time Signals: Defined along a continuum of time t and thus are represented by continuous independent variables, for example $x_c(t)$. Continuous-time signals are often referred to as analog signals.

Decimation: The process of decreasing the sampling rate. It consists of filtering and downsampling.

Decimation Filter: The filter used in decimation to avoid aliasing caused by downsampling.

Digital Filter: The filter is a discrete-time system, which changes the characteristics of the input discrete signal in a desired manner to obtain the discrete output signal.

Digital Filter Bank: Set of digital bandpass filters with the common input or a common output.

Discrete-Time Signals: Defined at discrete time values and thus the independent variable has discrete values n, as for example x(n).

Interpolation: The process of increasing the sampling rate. It consists of upsampling and filtering.

Interpolation Filter: The filter used in interpolation to remove the unwanted images in the spectra of the upsampled signal.

Multirate System: Discrete–time systems with unequal sampling rates at various parts of the system.

Sampling: The generation of a discrete-time signal x(n) from a continuous signal $x_c(t)$ is called sampling, where $x(n)= x_c(nT)$. T is called the sampling period and its inverse $1/T$ is the sampling frequency or the sampling rate.

Subband Coding (SBC) Filter Bank: Consists of an analysis filter bank followed by synthesis filter bank. This type of filter bank is used for partitioning signals into subbands for coding purposes, and vice versa.

Synthesis Filter Bank: Combines subband signals into one signal.

Transmultiplexer Filter Bank (TMUX): Consists of a synthesis filter bank followed by an analysis filter bank. This type of filter bank is used for converting time-multiplexed signals (TDM) into frequency-multiplexed signals (FDM), and vice versa.

F

Future of Small Business E-Commerce

Simpson Poon
IVE, Hong Kong and Charles Sturt University, Australia

INTRODUCTION

The use of the Internet for business purposes among small businesses started quite early in the e-commerce evolution. In the beginning, innovative and entrepreneurial owners of small businesses attempted to use rudimentary Internet tools such as electronic mail (e-mail) and file transfer protocol (FTP) to exchange messages and documents. While primitive, it fulfilled much of the business needs at the time. Even to date, e-mail and document exchange, according to some of the latest research findings, are still the most commonly used tools despite the fact that tools themselves have become more sophisticated.

There has been a plethora of research on small-business use of the Internet and e-commerce in general. Some of the earlier research can be traced back to the mid-1990s, when small-business e-commerce was in its infancy (e.g., Abell & Lim, 1996; Cronin, Overfelt, Fouchereaux, Manzvanzvike, Cha, & Sona, 1994; Fuller & Jenkins, 1995; Poon & Swatman, 1995), and when this chapter was written, articles were still being published in this area around the world. While the findings are many and varied, there are a number of highlights. Essentially, four groups of factors had been commonly found to influence the adoption of e-commerce among small businesses (see Figure 1).

(1) Organizational factors: Demographic factors such as size and the industry sector had been found to have an influence on the level of adoption (Poon & Swatman, 1999; Thong, 1999). The information intensity of products was also found to be a determining factor influencing adoption (Hui & Chau, 2002). While there was evidence that the nature of a product did not necessarily determine e-commerce success (Poon, 2000), anecdotal evidence suggested that small businesses offering digital products or products with high information intensity seem to benefit more from e-commerce.

(2) Technological factors: The level of IT adoption within an organization as well as along the supply chain of an organization was suggested to be good predictors of success in small-business c-commerce adoption (Mehrtens, Cragg, & Mills, 2001). Research has shown that the technical competence of key personnel within a small firm also related to the perceived success when adopting IT (Thong, 1999).

(3) Management factors: Studies into IT and e-commerce adoption have repeatedly identified the importance of management's involvement in facilitating adoption. Findings generally point out that the proactive attitude and actions of management, as well as the innovativeness of their strategy toward e-commerce, had led to success (Poon & Swatman, 1998; Thong, 1999). In the case of a small business, the dominant role of the owner or director means management is often the determining factor of e-commerce success.

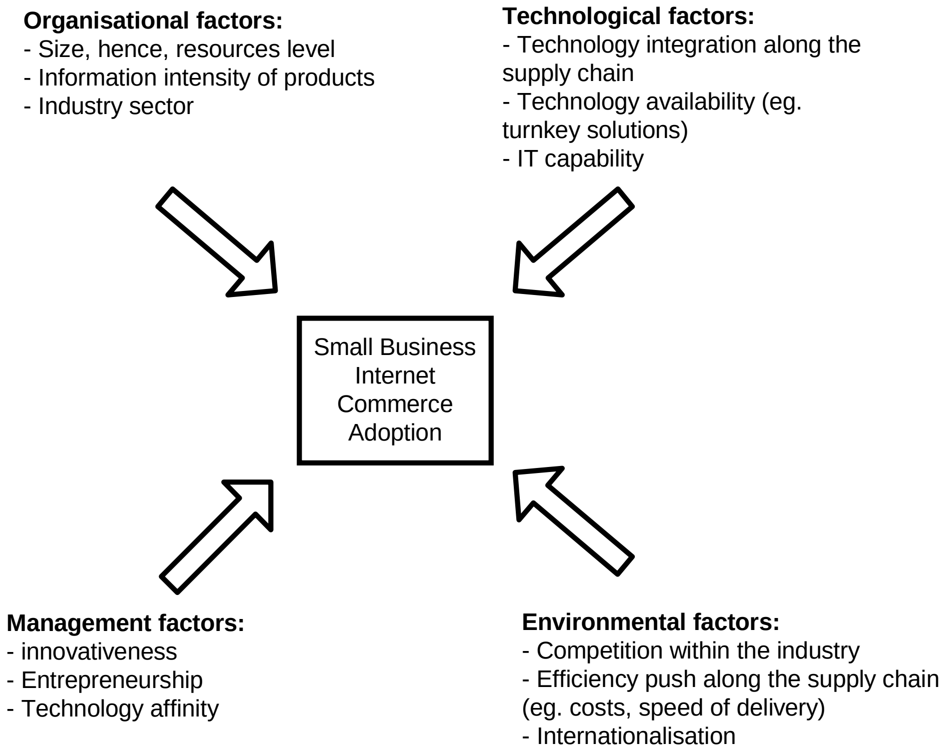
(4) Business-sector factors: The level of IT adoption and the expectation to use IT for interfirm activities within the industry sector a small business is in also seem to influence the level and extent of e-commerce adoption (Premkumar & Roberts, 1999; Raymond & Blili, 2001). If there is a lesson from electronic-data-interchange (EDI) adoption to be learned, it is that small businesses were often under the influence of more powerful business partners along the supply chain to adopt interorganizational technologies such as EDI. Similar patterns have been observed in c-commerce adoption cases.

BACKGROUND

Besides the well-mentioned business-to-consumer (B2C) and business-to-business (B2B) e-commerce setup, the Internet also enables small businesses to form business networks known as e-clusters. Compared to the traditional "hub-and-spokes" topology of networked organizations, involving a large organization and a number of small businesses during EDI adoption, the Internet sees the formation of e-clusters. Small businesses are known to form business clusters: a group of firms working closely together for business advantages. This phenomenon has been observed both in the East (Akizawa & Kijima, 1997), the West (Brown & Lockett, 2001), and around the world (UNIDO, 2001).

Brown and Lockett (2001) highlighted how small businesses empower such clusters using e-commerce technologies. The e-clusters formed can leverage both the synergy of the group and the technological advantages

Figure 1. Four groups of factors commonly found to influence Internet adoption among small businesses



to further enhance the agility of the whole cluster. Governments are also keen to investigate how industries, particularly those whose key products are of digital nature, can benefit from e-clustering (Sakai, 2002). E-clusters will be a key economic drive for nations that have a high percentage of small and medium-sized enterprises (SMEs). The productivity and operational efficiency among SMEs within such clusters can be enhanced through the economy of information exchange.

FUTURE TRENDS

From Competitive Advantage to Necessity

As more small businesses are using the Internet to support business activities, it is no longer an option but a key to competitiveness. For small businesses that have suppliers and/or customers who have adopted e-commerce, not being able to interact using Internet technologies is a competitive disadvantage. As such, small businesses are expecting to be e-commerce ready or risk losing out in the future. As e-commerce software and services are maturing and becoming affordable, widespread adoption is expected.

From Fragments to Clusters

For a few years, e-commerce adoption has been a fragmented process, meaning individual small businesses tried to establish a Web presence or install e-commerce software onto their systems. As efforts such as application service providers, B2B marketplaces, and industry-wide platforms are maturing, small businesses are becoming part of a supply web, a member of a B2B marketplace, or are tightly integrated into a large organization's extranet. The result is that small businesses will become part of one or more e-clusters, reflecting the nature and flows of business activities.

From Paper Based to Paperless

While clichés about a paperless office have been around for over a decade, it becomes closer to such a reality when B2C, B2B, and B2G (business to government [or public establishments]) activities are increasingly popular. In some countries, submission of information such as taxation assessments is now carried out as part of an e-government setup. Small businesses are under pressure to carry out business transactions not only with other businesses, but also with government and public organizations over the Internet.

From Driven to Driving

Many small businesses in the 1990s were driven to adopt e-commerce, sometimes unwillingly. This might be due to the lack of knowledge and human resources to address related technical and business issues. As technologies are becoming more user friendly and affordable, together with readily available technical specialists, adoption of some kind of e-commerce applications and/or systems has risen dramatically. The complexity of e-commerce systems from a user's perspective is not that much different from office software. Most importantly, the e-commerce strategy should not be someone else's agenda, but the driving solution of the small firm.

CONCLUSION

The percentage of small businesses adopting and using e-commerce is increasing every year. Small businesses that used to be slow to adopt IT are now either already pursuing or actively using e-commerce for their business activities. There are many reasons for this, but the availability of easy-to-use and affordable solutions, lower entry barriers, and driving forces from business partners are the key influencing factors. Small businesses are now left with few options except to adopt e-commerce to streamline their operations and stay competitive. As business activities and transactions are increasingly carried out over the Internet, small businesses that have scant resources will find e-commerce is now part of their solution rather than part of the problem as in the early days of the dot-com era.

REFERENCES

Abell, W., & Lim, L. (1996). *Business use of the Internet in New Zealand: An exploratory study*. Retrieved from *http://ausweb.scu.edu.au/aw96/business/abell/paper.htm*

Akizawa, H., & Kijima, K. (1997). *Internet-based entrepreneurial networking to evolve the Japanese industrial structure*. Retrieved from *http://www.caravan.net/246c/hikari/pac.pdf*

Brown, D. H., & Lockett, N. J. (2001). Engaging SMEs in e-commerce: The role of intermediaries within eClusters. *Electronic Markets, 11*(1), 52-58.

Cronin, B., Overfelt, K., Fouchereaux, K., Manzvanzvike, T., Cha, M., & Sona, E. (1994). The Internet and competitive intelligence: A survey of current practice. *International Journal of Information Management, 14*, 204-222.

Fuller, T., & Jenkins, A. (1995). Public intervention in entrepreneurial innovation and opportunism: Short cuts or detours to the information superhighway? *Proceedings of the 15th Annual Entrepreneurship Research Conference*. Retrieved from *http://www.babson.edu/entrep/fer/papers95/*

Hui, K. L., & Chau, P. Y. K. (2002). Classifying digital products. *Communications of the ACM, 45*(6), 73-79.

Mehrtens, J., Cragg, P. B., & Mills, A. M. (2001). A model of Internet adoption by SMEs. *Information and Management, 39*, 165-176.

Poon, S. (2000). Product characteristics and Internet commerce benefit among small businesses. *Journal of Product and Brand Management, 9*(1), 21-34.

Poon, S., & Swatman, P. (1998). A combined-method study of small business Internet commerce. International Journal of Electronic Commerce, *2*(3), 31-46.

Poon, S., & Swatman, P. (1999). An exploratory study of small business Internet commerce issues. *Information & Management, 35*, 9-18.

Poon, S., & Swatman, P. W. C. (1995). The Internet for small businesses: An enabling infrastructure for competitiveness. Retrieved from *http://inet.nttam.com*

Premkumar, G., & Roberts, M. (1999). Adoption of new information technologies in rural small businesses. *The International Journal of Management Science (OMEGA), 27*, 467-484.

Raymond, L., & Blili, S. (2001). Organizational learning as a foundation of electronic commerce in networked organizations. *International Journal of Electronic Commerce, 5*(2), 29-45.

Sakai, K. (2002). Global industrial restructuring: Implications for small firms. *STI working paper series, DSTI/DOC, 4*. Retrieved from *http://www.olis.oecd.org/olis/2002doc.nsf/LinkTo/DSTI-DOC(2002)4*

Thong, J. (1999). An integrated model of information systems adoption in small business. *Journal of Management Information Systems, 15*(4), 187-214.

UNIDO. (2001). *Development of clusters and networks of SMEs*. Retrieved from *http://www.unido.org/en/doc/view/4927*

KEY TERMS

E-Cluster: An e-cluster is a group of firms; in this article they are small businesses, which engage in various kinds of business relationships and are using e-commerce to support such business relationships.

E-Commerce: E-commerce in this article means using the Internet and related technologies and software to support business activities in general. It ranges from simple setup such as e-mail and file transfer to complex systems such as supply chain management and enterprise resources-planning systems.

Electronic Data Interchange (EDI): EDI in this article means a set of document-exchange standards used to implement interorganizational systems exchanges to achieve automated computer-to-computer document exchange regardless of the communication technologies used.

Industry-Wide Platform: In this article it refers to using a mixture of information and communications technologies set up over the Internet to set up a multipurpose platform that enables firms within a sector, along a supply chain, or as part of a consortium to exchange business information and transact.

Small Business: The commonly adopted definition of a small business is that the firm size should be less than 500 employees. However, there are other interpretations of how small a business should be to be considered a small business including firm sizes of less than 100 (the Netherlands), 50 (Ireland), or even 20 (Australia). Business turnover, management and organizational structure, as well as other supplementary criteria, are also used by some authors.

Supply Web: An elaborated web of supply chains intersecting each other. A small business can be on one or more of the supply chains within this web. Supply chains within this web may or may not be in the same industry.

Gender and Computer Anxiety

Sue E. Kase
The Pennsylvania State University, USA

Frank E. Ritter
The Pennsylvania State University, USA

INTRODUCTION

Because of their ability to enhance productivity, computers have become ubiquitous in the workplace. By the early 1990s the use of computers in the workplace reached a per capita penetration that the telephone took 75 years to achieve (Webster & Martocchio, 1992). During the past several decades, there has been both speculation and hard research related to the psychological effects of computer technology. More recently the role of attitudes towards computers in influencing the acceptance and use of computer-based management information systems (MIS) has been highlighted by a growing number of MIS researchers. Generally, these studies focus on the negative attitudes towards computers and concerns about the impact of MIS on individual performance in the workplace.

Computer anxiety has been reported to be associated with negative attitudes towards computers. As computers play a pervasive role in MIS and decision support systems, these findings emphasize the need for additional empirical research on the determinants of computer anxiety and attitudes towards computers. Furthermore, with the increasing participation of women in information technology professions, important questions are whether men and women differ with regard to computer anxiety and attitudes towards computers, and what factors explain such differences where they exist, and how to ameliorate anxiety where it occurs.

The Concept and Correlates of Computer Anxiety

Much has been speculated about computer anxiety, both what it is and what to do about it. Computer anxiety is context specific and covers a wide variety of situations in which people interact with computers. Context-specific anxiety tests ask the question: "How do you feel when a specific type of situation occurs?" Commonly, the relationship between a measure of computer anxiety and other variables is examined. For example, the relationship of computer anxiety to computer-related experience has historically been a hotly contested question in MIS research, human-computer interaction (HCI), and educational psychology. Demographic variables posited or found to be related to computer anxiety include gender, age, organizational level, and academic major (Dambrot, Watkins-Malek, Silling, Marshall & Garver, 1985; Gutek & Bikson, 1985; Zmud, 1979). Personality variables examined as potential determinants of computer anxiety include trait anxiety, math anxiety, cognitive style, and locus of control (Howard & Smith, 1986; Igbaria & Parasuraman, 1989; Morrow, Prell & McElroy, 1986). Additionally, several studies have examined the relationship between computer anxiety and academic achievement. For example, Hayek and Stephens (1989) and Marcoulides (1988) reported significantly lower computer anxiety being associated with higher academic achievement.

BACKGROUND

Initially, computer anxiety became of interest during the technological revolution. In 1963 a social psychologist at IBM completed a nationwide study to examine popular beliefs and attitudes about one of the prime symbols of our rapidly changing technology—the electronic computer. Lee's (1970) findings concluded that the American public viewed computers on two independent dimensions. The first dimension, the "Beneficial Tool of Mankind Perspective," described a positively toned set of beliefs that computers are beneficial in science, industry, and business. The second dimension, the "Awesome Thinking Machine Perspective," connoted fear of an incomprehensibly complex machine with capabilities far exceeding those of a human. This perspective, which reflects ignorance about the capabilities and limitations of computers, is one of the generic origins of computer anxiety.

Later, during the 1980s, much of the writing about computer anxiety and attitudes towards computers was concentrated in trade and business publications (e.g., Howard, 1986; Igbaria & Parasuraman, 1989). During this time period uncertainty was often considered the primary predictor of computer anxiety. This uncertainty referred to an individual's ability to learn to use the computer or to the potential the machine had to rearrange traditional office functions and power structures. Sabotage and hostility were sometimes responses to these uncertainties, especially when they were accompanied by fear of replacement by the machine. This particular concern was often voiced by middle managers who viewed their jobs as information conduits or as a mosaic of clerical tasks, all of which could be performed more efficiently by a computer. Managers with longer tenure with a company and those who felt they were currently utilizing their time quite effectively were likely to resist computer adoption and use. Additionally, computer usage required typing skills. Those persons who did not know how to type or considered typing a low-status skill were reluctant to adapt to the new technology.

Collectively two groups displayed the most susceptibility to computer anxiety: individuals without computer experience overestimated the difficulties involved in learning and interacting with computers; and individuals whose jobs appeared threatened resisted adaptation to technological improvements (Gilroy & Desai, 1986). It has been well documented that among individuals demonstrating computer anxiety are significant numbers of women, as examined in the next section.

The Role of Gender in Computer Anxiety

According to feminist technology studies, computers are widely perceived as belonging to the "male domain" of mathematics, science, electronics, and machinery (Beyer, 1999; Cockburn & Ormrod, 1993; Faulkner, 2001). This, coupled with reports of greater prevalence of math anxiety among women than men (e.g., Brown & Josephs, 1999; Chipman, Krantz & Silver, 1992), suggests that women are likely to have a more negative view of computer use than men. It is not surprising that men have been found to display lower computer anxiety, higher computer aptitude, and more positive attitudes towards computers in general than women (Chua, Chen & Wong, 1999; Coffin & Machintyre, 2000; Colley, Gale & Harris, 1994; Whitely, 1997).

The limited empirical research on gender differences in computer anxiety, attitudes towards computers, and computer experiences among working adults reveals conflicting results, however. By the early 1990s, only 25 studies presented sufficient statistical information that could be converted to correlations, and an additional 13 qualitative research reports supported only slight differences between men and women in computer anxiety (Rosen & Maguire, 1990). In contrast, other studies reported no gender differences associated with computer use in the workplace.

More specific studies found stronger correlations and gender differences. One of the most frequently cited studies on computer anxiety is by Rosen, Sears, and Weil (1987). They examined the relationship between computer anxiety and gender role as measured by the Bem Sex Role Inventory (Bem, 1974). This instrument identified individuals as belonging to one of four identity groups: masculine, feminine, androgynous, or undifferentiated. They found that feminine-identity individuals had more computer anxiety and more negative computer attitudes than did masculine-identity individuals, regardless of gender. Another influential instrument employed extensively in early studies on microcomputer anxiety in management is the Computer Anxiety Rating Scale (CARS) developed by Raub (1981). Raub investigated math anxiety, gender, age, trait anxiety, and knowledge of computers as possible correlates of computer anxiety. She suspected a gender effect based on the negative socialization of women toward mathematics, science, and technology and on the resulting production of anxieties. Raub found the relationship between computer anxiety and gender so strong that she ran separate regressions for males and females.

CARS has been used in more recent research as well. Anderson (1996) utilized CARS to determine whether or not perceived knowledge of software, computer experience, overall knowledge of computers, programming experience, and gender were predictors of computer anxiety. Table 1 displays the CARS portion of the Anderson questionnaire. Collaborating Raub's results, Anderson's study showed higher computer anxiety is accompanied by less experience and less perceived knowledge of computers, and that at higher levels of computer anxiety, women are over-represented.

Indirectly, the use of CARS led to the introduction of statistical modeling as a more formal investigation of the psychological mechanisms that trigger computer anxiety and the remedies for it. For example, the work of Howard (1986) is distinctive during the 1980s in its similarity to more recent research on computer anxiety. Howard developed a sequence of models addressing the predictors of computer anxiety and the use of computers in management. In schematic form, Figure 1 shows the possible relationships between psychological variables and the attitudes of managers toward the usefulness of microcomputers as management tools. Howard's study confirmed that computer anxiety is a significant inverse correlate of managers' positive attitudes toward microcomputers.

Similar to Raub, Howard speculated that gender may correlate with math anxiety and possibly with computer anxiety. Gender as a math anxiety correlate reflects psychological differences between men and women with

Table 1. The Computer Anxiety Rating Scale (CARS) is based on research by Raub (1981)

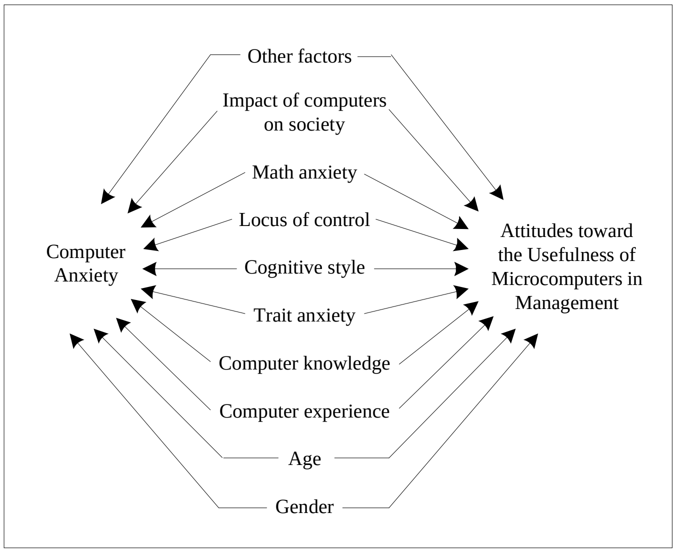
The responses are scaled as follows:

1 = Strongly Agree
2 = Agree
3 = Unsure
4 = Disagree
5 = Strongly Disagree

Items 2 through 10 are reverse scored so that high scores indicate high levels of computer anxiety.

1. I am confident that I could learn computer skills.
2. I am unsure of my ability to learn a computer programming language.
3. I will be able to keep up with the important technological advances of computers.
4. I feel apprehensive about using the computer.
5. If given the opportunity to use a computer, I'm afraid that I might damage it in some way.
6. I have avoided computers because they are unfamiliar to me.
7. I hesitate to use the computer for fear of making mistakes that I cannot correct.
8. I am unsure of my ability to interpret a computer printout.
9. I have difficulty understanding most technological matters.
10. Computer terminology sounds like confusing jargon to me.

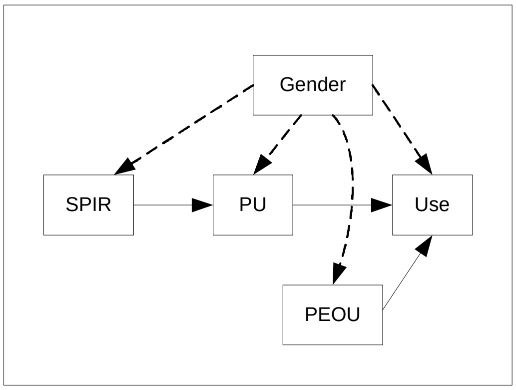
Figure 1. Computer anxiety and other possible correlates of managers' attitudes towards microcomputers adapted from Howard (1986)



regard to mathematics that may result from early socialization of females away from scientific and technical endeavors. If math anxiety and computer anxiety are similar phenomena, then they are likely to have common psychological roots. Thus, for both Raub and Howard, the prime psychological root of computer anxiety appeared to be that certain people simply did not see themselves as technological types. Math anxious types see mathematics and computers and all the paraphernalia of technology as for someone else and when required to use it experience stress.

A current perspective of Howard's usefulness of microcomputers research is several recent studies examining the diffusion of information technology (IT) in the workplace. Many of these studies have demonstrated a strong link between self-efficacy and individual reactions

Figure 2. Gender effects on TAM variables (Gefen & Straub, 1997)



to computing technology, both in terms of adoption and use of computers, and in terms of learning to use computer systems and applications. Beliefs about our capabilities to use technology successfully are related to our decisions about whether and how much to use technology, and the degree to which we are able to learn from training (Compeau, Higgins & Huff, 1999). While inconsistencies in group differences exist, overall findings of IT diffusion research suggest that women and men exhibit different perceptual tendencies and usage patterns in computer-related circumstances (Chou, 2001; Gefen & Straub, 1997; Hackbarth, Grover & Yi, 2003; Venkatesh, Morris & Ackerman, 2000).

Gefen and Straub (1997), and Venkatesh, Morris, and Ackerman (2000) are two prevalent examples of IT diffusion research predicated on influential theories evaluating information technology perception and use: the Technology Acceptance Model (TAM) and the Theory of Planned Behavior (TPB). Gefen and Straub (1997) modeled IT diffusion by extending TAM and the Social Presence/Information Richness (SPIR) factor addendum to include gender in the context of e-mail system usage. Figure 2 shows the effects of gender on TAM and cultural extensions, specifically the perceived attributes of SPIR, perceived ease of use (PEOU), and perceived usefulness (PU). Gefen and Straub sampled 392 female and male knowledge workers using e-mail systems in the airline industry in North America, Asia, and Europe. Study findings indicated that women's perceptions of e-mail are different from male co-workers. Covariates gender and culture accounted for 37% of the variance in SPIR, 53% of the variance in PEOU, and 59% of the variance in SPIR, culture, and PU combined. In this study gender definitely had an impact on the IT diffusion process.

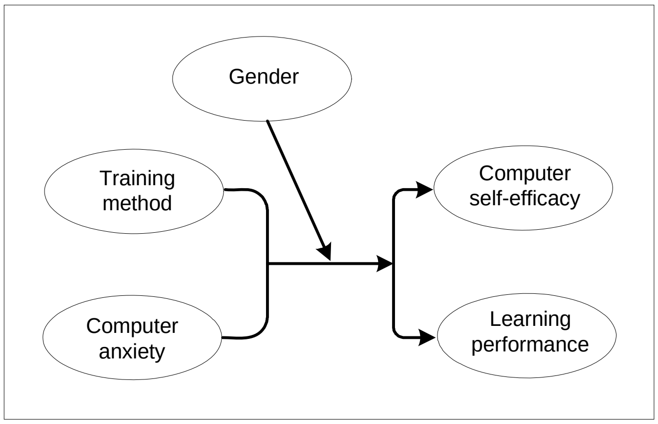
Using TPB as a framework, Venkatesh, Morris, and Ackerman (2000) investigated gender differences in the context of adoption and sustained usage of technology in the workplace. User reactions and technology usage behavior were studied over a five-month period among 355 workers being introduced to a new software application. Men's attitude toward using the new technology was strongly motivated by achievement needs and task-oriented or instrumental behavior. In contrast, women's attitude toward using the new technology was more strongly influenced by subjective norm and perceived behavioral control. Sustained technology usage behavior in both men and women was driven by early usage behavior, emphasizing the importance of gender-based early evaluations of a new technology.

The above studies suggest several implications: when marketing IT and considering its effects, the gender of the users should be considered; and when training users on particular information technology such as e-mail systems, groups composed primarily of women should be addressed in a different manner than mixed or mainly masculine groups. When training mostly female groups, user friendliness and the ability of the system to convey the presence of the communicator should be emphasized.

Computer training has been widely researched and considered an essential contributor to the success of organizational computing. Computer anxiety and attitudes toward computers have often been identified as the critical factors influencing computer learning performance and training methodology. From the late 1980s up to today, the relationship of training techniques and personal characteristics has played a key role in training end users of information systems.

Characteristic of such training studies, Chou (2001) developed a conceptual model to evaluate how training method, an individual's gender, and computer anxiety level affect learning performance and computer self-efficacy. Chou used two types of training methods: an instruction-based method and a behavior-modeling method. The instruction-based method is a traditional approach that teaches primarily by lecture and follows a deductive way to learning, where learners proceed from general rules to specific examples (Davis & Davis, 1990; Simon, Grover, Teng & Witcomb, 1996). On the other hand, the behavior-modeling approach is a task-focused method involving a visual observation of the behaviors of a user performing a task. Learners then imitate and extend the demonstrated behavior in practice and experimentation to master the task. The behavior-modeling method employs an inductive approach that teaches by hands-on demonstrations first, followed by complimentary lectures (Compeau & Higgins, 1995; Gist, Schwoerer & Rosen, 1989). Figure 3 shows Chou's research model. In this model, gender was

Figure 3. Moderating effects of gender on computer self-efficacy and learning performance (Chou, 2001)



proposed as a moderating variable that moderates the effects of training method and computer anxiety on both learning performance and computer self-efficacy. When the training methods were tested on students, gender effects in general were found to be significant: male subjects performed better then female subjects. Male students had better learning performance, higher computer self-efficacy, and lower computer anxiety. Female students had significantly lower self-efficacy and a lower self-image about their computer learning capabilities. The behavior-modeling method appeared to enhance the computer self-efficacy of male students, whereas the instruction-based method benefited female students, suggesting that instruction technique preference varies with gender.

Studies on the relationships between computer anxiety, learning performance, and gender contribute solid knowledge about end-user training potential. This knowledge aids educators and trainers in the development of more personalized and therefore effective training programs for groups and individuals (Bostrom, 1998; Chou, 2001; Davis & Davis, 1990; Santhanam & Sein, 1994).

FUTURE TRENDS

In the developing research area of profiling Internet users, extrapolation of the computer anxiety literature offers parallels between gender differences in Internet use and differences in expertise and attitudes towards computers. It is widely assumed that women's participation in the Internet is hampered by their attitudes towards computers, which in turn is reflective of their attitudes towards new technology (Durndell & Haag, 2002; Gackenback, 1998; Jackson, Ervin, Gardner & Schmitt, 2001; Kraut et al., 1998; Schumacher & Morahan-Martin, 2001; Weiser, 2000).

Supporting this claim, Durndell and Haag (2002) utilized a Computer Self-Efficacy Scale, a Computer Anxiety Scale, and an Attitude to the Internet Scale in obtaining information from 74 female and 76 male university students on their Internet usage. Durndell and Haag found significant gender effects throughout, with males tending to report greater computer self-efficacy, lower computer anxiety, more positive attitudes towards the Internet, and longer use of the Internet than females. Similarly, Schumacher and Morahan-Martin (2001) reported males feeling more comfortable and competent with computers and the Internet. One explanation is that males, from childhood on, have more experience with computers than females, especially with games and programming that enhance technological sophistication and increase overall levels of competence and comfort with computers. Sussman and Tyson (2000) suggested the nature of communication on the Internet may vary by gender, and Balka and Smith (2000) proposed gender differences in Web navigation strategies.

Employing psychological methodology, DeYoung and Spence (2004) designed an instrument, the Technology Profile Inventory (TPI), to profile information technology users for dynamic personalization of software interfaces. Their approach was to generate a broad range of items for assessing responses to information technology, to examine their factor structure in a normal population, and to investigate potential associations between the emergent factors and variables that have, in the past, been associated with responses to computers, including gender, age, experience with information technology, and use of information technology. In application, a Web page could conform itself to suit the technology profile of each user who encounters it, for instance, a program could display all of its options for someone high in "computer interest" while displaying only the most functional options for someone who just wants to accomplish a task as simply as possible.

CONCLUSION

Computers are a vital asset in today's business and education world. The emergence of computers and information systems has been perhaps the single largest factor influencing organizations during the past three decades. Despite the increasing dispersal of computers, there is significant evidence that individual computer usage is affected by the computer anxiety or fear of computers that is widespread, and negative attitudes towards computers in general. This suggests that the potential benefits of computers as aids to professionals may not be fully realized, and the success of using a computer is dependent on the user's acceptance and commitment. The presence of computer anxious individuals in the workplace can lead to performance problems, decline in motivation, work quality, and moral, and can increase errors, absenteeism, interpersonal conflicts, and turnover (Brosnan, 1998; Mikkelsen, Ogaard, Lindoe & Olsen, 2002).

Historically, a commonly held stereotype of computer anxiety is that of a frightened female secretary, struggling to learn the new word processor that her (male) boss is making her use as a replacement for her old, trusty IBM Selectric typewriter. The majority of gender and computer anxiety research does not support this description—both the secretary and the boss in this story may be anxious, but for different reasons and with different implications. In this particular case role has a greater effect than gender. However, the confusing continuum of disagreement characterizing the role of gender as it relates to attitudes toward computers, and by extension, IT diffusion and learning performance in the workplace, carries on today. While men may still represent a majority of the IT workforce, the number of women in technology-oriented areas continues to rise. As a result, the implementation of new technology requires an understanding of the factors that are likely to lead to user acceptance and sustained usage across gender and experience levels.

Does gender matter when examining attitudes and anxiety toward computers? The only valid conclusion that can be drawn from the existing body of literature is: it is important, but we do not fully understand gender as a moderating variable in the context of computer anxiety. One standard recommendation is to do more research and bring to the forefront the need to be cognizant of sex differences. The hypothesis that some users will have less overall experience with computers and are therefore more likely to have negative attitudes towards computers should be kept in mind when creating, adopting, and using information technology systems. Also, if cognitive style is distributed differently across genders, which it currently appears to be, it too may be a relevant factor in IT diffusion. One way to ameliorate these differences is for new technology introductions to be accompanied by user involvement, training, and active practical use. Special attention should be paid to the user's sex and their experience level. Because of the ramifications in society, education, and the workplace, educators and managers need to know how to recognize computer anxiety and the strategies to help alleviate or eliminate it. It is important that future research has focus and direction, and answers larger questions, such as: understanding how high levels of computer anxiety develop; what role such anxiety plays in career choices; and constructing methods to reduce computer anxiety within technological environments.

REFERENCES

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*, 179-211.

Anderson, A.A. (1996). Predictors of computer anxiety and performance in information systems. *Computers in Human Behavior*, *12*(1), 61-77.

Balka, E. & Smith, R. (Eds.). (2000). *Women work and computerization*. Boston: Kluwer.

Bem, S.L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, *42*, 155-162.

Beyer, S. (1999). The accuracy of academic gender stereotypes. *Sex Roles*, *40*, 787-813.

Brosnan, M.J. (1998). The impact of computer anxiety and self-efficacy upon performance. *Journal of Computer Assisted Learning*, *14*, 223-234.

Brown, R.P. & Josephs, R.A. (1999). A burden of proof: Stereotype relevance and gender differences in math performance. *Journal of Personality and Social Psychology*, *76*(2), 246-257.

Chipman, S.F., Krantz, D.H. & Silver, R. (1992). Mathematics anxiety and science careers among able college women. *Psychological Science*, *3*(5), 292-295.

Chou, H.W. (2001). Effects of training method and computer anxiety on learning performance and self-efficacy. *Computers in Human Behavior*, *17*, 51-69.

Chua, S., Chen, D. & Wong, P. (1999). Computer anxiety and its correlates: A meta analysis. *Computers in Human Behavior*, *15*(5), 609-623.

Cockburn, C. & Ormrod, S. (1993). *Gender and technology in the making*. London: Sage Publications.

Coffin, R. & Machintyre, P. (2000). Cognitive motivation and affective processes associated with computer-related performance: A path analysis. *Computers in Human Behavior*, *16*(2), 199-222.

Colley, A.M., Gale, M.T. & Harris, T.A. (1994). Effects of gender role identity and experience on computer attitude components. *Journal of Educational Computing Research*, *10*(2), 129-137.

Compeau, D.R. & Higgins, C.A. (1995). Application of social cognitive theory to training for computer skills. *Information Systems Research*, *6*(2), 118-143.

Compeau, D., Higgins, C.A. & Huff, S. (1999). Social cognitive theory and individual reactions to computing technology: A longitudinal study. *MIS Quarterly*, *23*(2), 145-158.

Dambrot, F.H., Watkins-Malek, M.A., Silling, M.S., Marshall, R.S. & Garver, J.A. (1985). Correlates of sex differences in attitudes toward and involvement with computers. *Journal of Vocational Behavior*, *27*, 71-86.

Davis, F. (1989). Perceived usefulness, perceived ease of use and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-340.

Davis, D.L. & Davis, D.F. (1990). The effect of training techniques and personal characteristics on training end users of information systems. *Journal of Management Information System*, *7*(2), 93-110.

DeYoung, C.G. & Spence, I. (2004). Profiling information technology users: En route to dynamic personalization. *Computers in Human Behavior*, *20*, 55-65.

Durndell, A. & Haag, Z. (2002). Computer self-efficacy, computer anxiety, attitudes towards the Internet and reported experience with the Internet, by gender, in an East European sample. *Computers in Human Behavior*, *18*, 521-535.

Faulkner, W. (2001). The technology question in feminism: A view from feminist technology studies. *Women's Studies International Forum*, *24*(1), 79-95.

Gachenback, J. (Ed.). (1998). *Psychology and the Internet: Intrapersonal, interpersonal and transpersonal implications*. New York: Academic Press.

Gefen, D. & Straub, D.W. (1997). Gender differences in the perception and use of e-mail: An extension of the technology acceptance model. *MIS Quarterly*, *21*(4), 389-400.

Gilroy, F.D. & Desai, H.B. (1986). Computer anxiety: Sex, race and age. *International Journal of Man- Machine Studies*, *25*, 711-719.

Gist, M.E., Schwoerer, C. & Rosen, B. (1989). Effects of alternative training methods on self-efficacy and performance in computer software training. *Journal of Applied Psychology*, *74*, 884-891.

Gutek, B.A. & Bikson, T.K. (1985). Differential experience of men and women in computerized offices. *Sex Roles*, *13*(3/4), 123-136.

Hackbarth, G., Grover, V. & Yi, M.Y. (2003). Computer playfulness and anxiety: Positive and negative mediators of the system experience effect on perceived ease of use. *Information & Management*, *40*, 221-232.

Hayek, L.M. & Stephens, L. (1989). Factors affecting computer anxiety in high school computer science students. *Journal of Computers in Mathematics and Science Teaching*, *8*(4), 73-76.

Hofstede, G. (1980). *Culture's consequences: International differences in work related values*. London: Sage Publications.

Howard, G.S. (1986). *Computer anxiety and the use of microcomputers in management*. Ann Arbor, MI: UMI Research Press.

Howard, G.S. & Smith, R. (1986). Computer anxiety in management: Myth or reality? *Communications of the ACM, 29*(7), 611-615.

Igbaria, M. & Parasuraman, S. (1989). A path analytic study of individual characteristics, computer anxiety and attitudes toward microcomputers. *Journal of Management*, *15*, 373-388.

Jackson, L., Ervin, K., Gardner, P. & Schmitt, N. (2001). Gender and the Internet: Women communicating and men searching. *Sex Roles*, *44*(5/6), 363-379.

Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukopadhyay, T. & Scherlis, W. (1989). Internet paradox: A social technology that reduces social involvement and psychological well-being? *American Psychologist*, *53*(9), 1017-1031.

Lee, R.S. (1970). Social attitudes and the computer revolution. *Public Opinion Quarterly, 34*, 53-59.

Marcoulides, G.A. (1988). The relationship between computer anxiety and computer achievement. *Journal of Educational Computing Research*, *4*, 151-158.

Mikkelsen, A., Ogaard, T., Lindoe, P. & Olsen, O. (2002). Job characteristics and computer anxiety in the production industry. *Computers in Human Behavior*, *18*, 223-239.

Morrow, P.C., Prell, E.R. & McElroy, J.C. (1986). Attitudinal and behavioral correlates of computer anxiety. *Psychological Reports*, *59*, 1199-1204.

Raub, A.C. (1981). *Correlates of computer anxiety in college students*. Unpublished doctoral dissertation, University of Pennsylvania, USA.

Riding, R.J. & Rayner, S. (1998). *Cognitive styles and learning strategies: Understanding style differences in learning and behavior*. London: D. Fulton Publishers.

Rosen, L.D. & Maguire, P. (1990). Myths and realities of computer phobia: A meta-analysis. *Anxiety Research, 3*, 175-191.

Rosen, L.D., Sears, D.C. & Weil, M.M. (1987). Computerphobia. *Behavior Research Methods, Instruments & Computers*, *19*, 167-179.

Santhanam, R. & Sein, M.K. (1994). Improving end user proficiency effects of conceptual training and nature of interaction. *Information Systems Research*, *5*, 378-399.

Schumacher, P. & Morahan-Martin, J. (2001). Gender, Internet, and computer attitudes and experiences. *Computers in Human Behavior*, *17*(1), 95-110.

Simon, S.J., Grover, V., Teng, J.T. & Whitcomb, K. (1996). The relationship of information system training methods and cognitive ability to end-user satisfaction, comprehension, and skill transfer: A longitudinal field study. *Information Systems Research*, *7*(4), 466-490.

Straub, D.W. (1994). The effect of culture on IT diffusion: E-mail and FAX in Japan and the U.S. *Information Systems Research*, *5*(1), 23-47.

Sussman, N. & Tyson, D. (2000). Sex and power: Gender differences in computer mediated interactions. *Computers in Human Behavior*, *16*(4), 381-394.

Venkatesh, V., Morris, M.G. & Ackerman, P.L. (2000). A longitudinal field investigation of gender differences in individual technology adoption decision-making processes. *Organizational Behavior and Human Decision Processes*, *83*(1), 33-60.

Webster, J. & Martocchio, J.J. (1992). Microcomputer playfulness: Development of a measure with workplace implications. *MIS Quarterly*, *16*, 201-226.

Weiser, E. (2000). Gender differences in Internet use patterns and Internet application preferences: A two-sample comparison. *CyberPsychology and Behavior*, *3*(2), 167-178.

Whitely, B. (1997). Gender differences in computer related attitudes and behavior: A meta analysis. *Computers in Human Behavior*, *13*(1), 1-22.

Witkin, H., Moore, C., Goodenough, C. & Cox, P. (1977). Field dependent and field cognitive styles and their educational implications. *Review of Educational Research*, *47*, 1-64.

Zmud, R.W. (1979). Individual differences and MIS success: A review of the empirical literature. *Management Science*, *25*(10), 966-979.

KEY TERMS

Cognitive Style: Information processing habits that represent an individual's typical modes of perceiving, thinking, remembering, and problem solving. Various cognitive styles have been identified, measured, and shown to affect the manner in which individuals perceive their environments. As just one example, two such styles are field-independence and field-dependence. Field-independent individuals perceive objects as separate from the field, impose personal structures on the environment, set self-defined goals, work alone, choose to deal with abstract subject matter, are socially detached and rely on their own values, and are self-reinforcing. In contrast, field-dependent individuals tend to rely on the environment for clues about an object, prefer a structure provided by the environment, experience the environment more globally, are interested in people, use externally defined goals, receive reinforcement from others, focus on socially oriented subject matter, and prefer to work with others (Riding & Rayner, 1998; Witkin, Moore, Goodenough & Cox, 1977).

Computer Anxiety: The tendency of a particular individual to experience a level of uneasiness over his or her impending use of a computer, which is disproportionate to the actual threat presented by the computer. Computer anxiety, defined by Raub (1981), is "the complex emotional reactions that are evoked in individuals who interpret computers as personally threatening."

Computer Anxiety Rating Scale (CARS): A self-report inventory consisting of 10 statements designed to measure computer anxiety. The scale comprises a mix of anxiety-specific statements (e.g., "I feel apprehensive about using the computer") and positive statements (e.g., "I am confident that I could learn computer skills") (Raub, 1981).

Computer Self-Efficacy: Computer self-confidence or perceptions of ability. Beliefs about one's ability to perform a specific behavior or task on a computer.

Locus of Control: Individuals' perceptions of whether they themselves influence events and outcomes in their lives (internal control), or that events and outcomes are influenced by factors such as luck, fate, chance, or powerful others (external control). Locus of control is considered a trait characteristic that is unlikely to change significantly in an individual's lifetime.

Math Anxiety: The psychological fear or anxiety associated with engaging in mathematical activity. Characteristics of math anxiety are an above-average number of negative attitudes (e.g., nervousness, solitude, uneasiness, and low confidence) and/or intense emotional reactions to math based on past experiences. Math anxiety and test anxiety are generally significant correlates and somewhat resemble computer anxiety as a situational manifestation of a general anxiety construct.

Perceived Ease of Use (PEOU): The degree to which an individual believes that using a particular information technology system would be free of effort. An application perceived to be easier to use than another is more likely to be accepted by users (Davis, 1989).

Perceived Usefulness (PU): The degree to which an individual believes that using a particular information technology system would enhance his or her job performance. A system high in perceived usefulness is one that a user believes has a positive usage to performance relationship (Davis, 1989).

Social Presence/Information Richness Factor (SPIR): A factor appended to TAM derived from Hofstede's (1980) work on dimensions of cultural differences among countries that include a disposition toward masculine attitudes and other behavioral indexes. The extension combines perceived social presence and the sense of human contact embodied in a medium with the information richness of the medium (Straub**,** 1994).

Technology Acceptance Model (TAM): A causal model hypothesizing that actual information technology system use is affected by behavioral intentions that themselves are affected by attitudes toward use. Beliefs about the system, perceived usefulness, and perceived ease of use in TAM directly affect attitudes toward use (Davis, 1989).

Technology Profile Inventory (TPI): A psychological instrument that generates technology profiles to predict how individuals are likely to respond to various aspects of information technology. The ability to profile information technology users facilitates the design of software capable of dynamic personalization (DeYoung & Spence, 2004).

Theory of Planned Behavior (TPB): Defines relationships among beliefs, attitude toward a behavior, subjective norm, perceived behavioral control, behavioral intention, and behavior. The theory has been widely applied across a range of disciplines such as marketing, consumer and leisure behavior, medicine, and information technology. When applied in technology adoption and usage contexts, TPB explains an individual's adoption of new technologies (Ajzen**,** 1991).

Trait Anxiety: Traits are properties of individuals that dispose them to react in certain ways in given classes of situations. Trait anxiety is a chronic predisposition to be anxious and nervous that may be based on feelings of inadequacy, usually due to poor past performances, low-self image, or low-self esteem.

Generic Framework for Defining Domain-Specific Models

Arnor Solberg
SINTEF ICT, Norway

Jon Oldevik
SINTEF ICT, Norway

Audun Jensvoll
EDB Telesciences, Norway

INTRODUCTION

As a result of the widespread popularity of the Unified Modeling Language (UML) (OMG, 2003-1), many companies have invested in introducing a UML-based methodology. There are many general purpose UML-based methodologies on the market today; among the most popular are UP (Jacobson, Booch & Rumbaugh, 1999), RUP (Kruchten, 2000), Catalysis (D'Souza & Wills, 1998), Select Perspective (Allen & Frost, 1998), and KOBRA (Atkinson et al., 2001). Typically, these general purpose software system development methodologies do not immediately fulfill a company's need. Aiming to provide methodologies that may be applied in many domains and for many purposes, these general purpose methodologies typically become extensive and are perceived as overwhelming. At the same time they typically lack support for the more exclusive needs that the companies and domains encounter. Thereby, introducing a general purpose methodology in an organization commonly implies two particular challenges that at first sight seems to be contradictory. On one hand there is a problem that the general purpose methodology provides/prescribes far too much and encounters too many situations. On the other hand the general purpose methodology does not support specific modeling concepts, mechanisms, and techniques wanted by the particular company or development group. Thus, in that respect the general purpose methodology actually covers too little. This state of affairs is why lots of consultants, researchers, and others are in the business of helping companies to introduce these methodologies, as well as customizing general purpose methodologies to be appropriate for the actual company and purpose. The customization is typically tuned based on different criteria such as domain, kind of customers, quality demands, size of the company, and size of the software development teams. A common way of customizing a general purpose methodology is by removing, adding, and/or merging prescribed tasks, phases, roles, and models/artifacts of the methodology. However, even if introduction of a general purpose methodology almost always requires a customization effort, there does not seem to be any standard and formalized way of doing it.

BACKGROUND

Our research group has for quite some time worked with customizing methodologies to satisfy specific needs. Our customization has been accomplished by taking a set of different general purpose methodologies (e.g., RUP, UP, OOram (Reenskaug & Wold, 1996)), methodology expertise, and experience as input into a collaborative process together with architects and super-users. By massaging this input through an iterative and incremental process in which we have analyzed the company's need and existing methodology (or practice) in use within the company, company culture, particularities of the domain, customers, market, and so forth. The output has been a tailored methodology.

Some results of this work have been the COMBINE methodology (COMBINE, 2003, 2000), the DAIM methodology (Hallsteinsen, Solberg, Fægri, Oldevik & Syrstad, 2003; DAIM, 2001), TeMOD (Solberg & Oldevik, 2001) and the Configurable LIght-Weight Method (CLIMB) (Solberg, Oldevik & Jensvoll, 2002a). What we have discovered during our work was that even if we gained substantial benefits from tailoring general purpose methodologies to the needs of the company, the company itself is quite diverse. Thereby, a need was expressed of even more tailoring to fit the purpose of different domains and product families within the company. For instance when developing TeMOD for Telenor[1] and later CLIMB for EDB Telesciences[2], a main request was to deliver a methodology that was tailored to capture and utilize existing domain knowledge. However, one of the goals of making

TeMOD and CLIMB was to provide a common methodology to be used throughout the company, in order to achieve a common way of developing and specifying systems. Thus, we were not supposed to end up with a set of proprietary special purpose methodologies, one for each domain and system development group. Our challenge became to keep TeMOD and CLIMB as the common methodologies for the respective company, enforcing standardized processes and specifications, and at the same time get the methodology to support specific needs of different domains and utilize the existing domain knowledge possessed within the company.

The most popular general purpose UML-based software engineering methodologies have both diversities and commonalties. One frequent commonality is that they are model driven. A model-driven methodology signifies that the methodology prescribes a set of models as the artifacts to be produced during the system development process. Model-driven methodologies have gained increasing popularity, even more so after the Model-Driven Architecture (MDA) (OMG, 2001; Frankel, 2003) initiative was launched. Our approach to the above described challenge was to exploit this model-driven aspect to develop a generic framework that provides utilities for tailoring model-driven methodologies in a formal and standardized way. Using the framework, the tailoring will only affect the expression of the models prescribed by the general purpose methodology.

FRAMEWORK DESCRIPTION

By applying the tailoring framework, a domain-specific reference model is produced. The reference model describes the extensions of the actual general purpose methodology made for a specific domain. It consists of UML-profiles, existing (reusable) models, and patterns.

The set of UML-profiles, existing models, and patterns defined in a reference model are developed, structured, and aligned according to the chosen general purpose software engineering methodology. UML-profiles are used for defining domain concepts and reference architectures. Existing models are prepared for reuse, and patterns describe standard solutions of recurring problems within the domain. Thus, tailoring a software engineering methodology using the framework constitutes a leveraging of the methodology in an environment of domain concepts, defined reference architectures, existing models, and patterns.

Figure 1. Example of framework usage

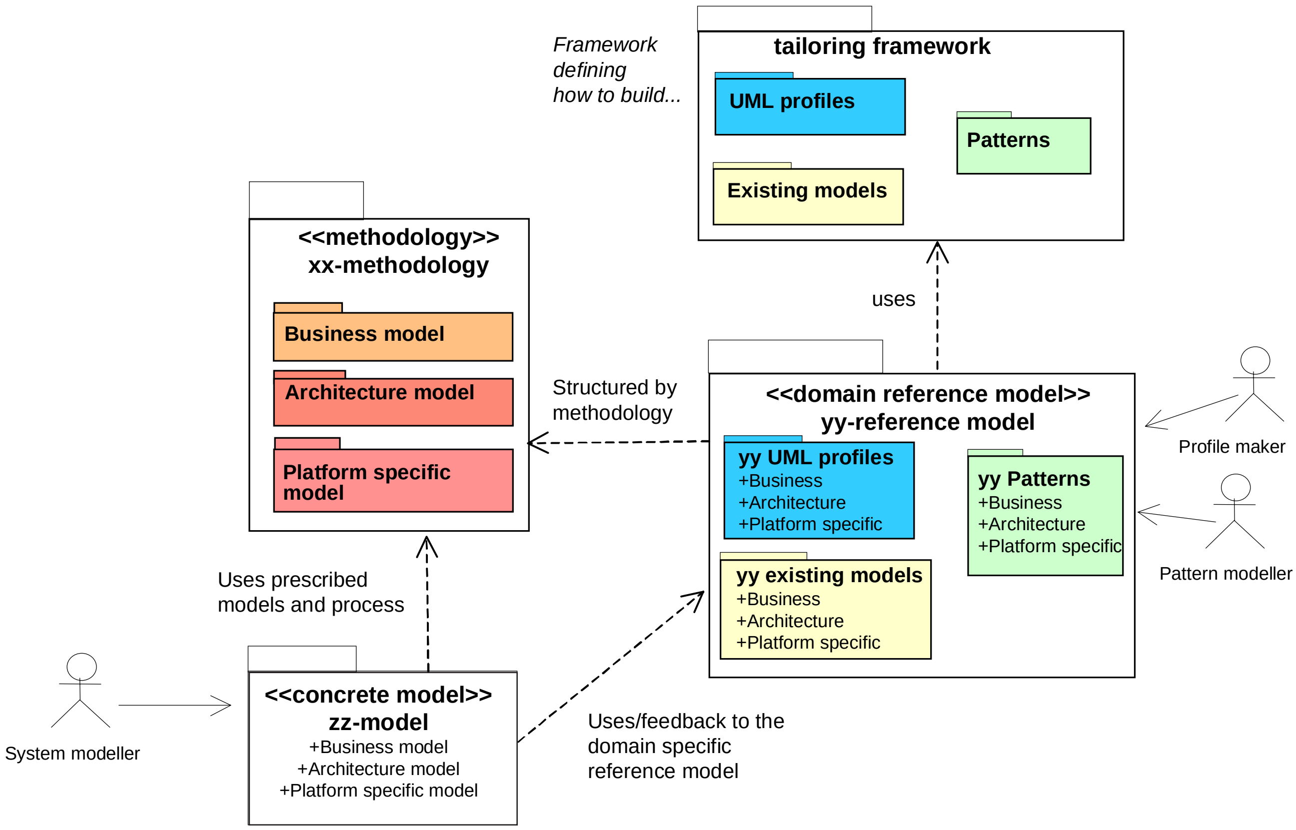


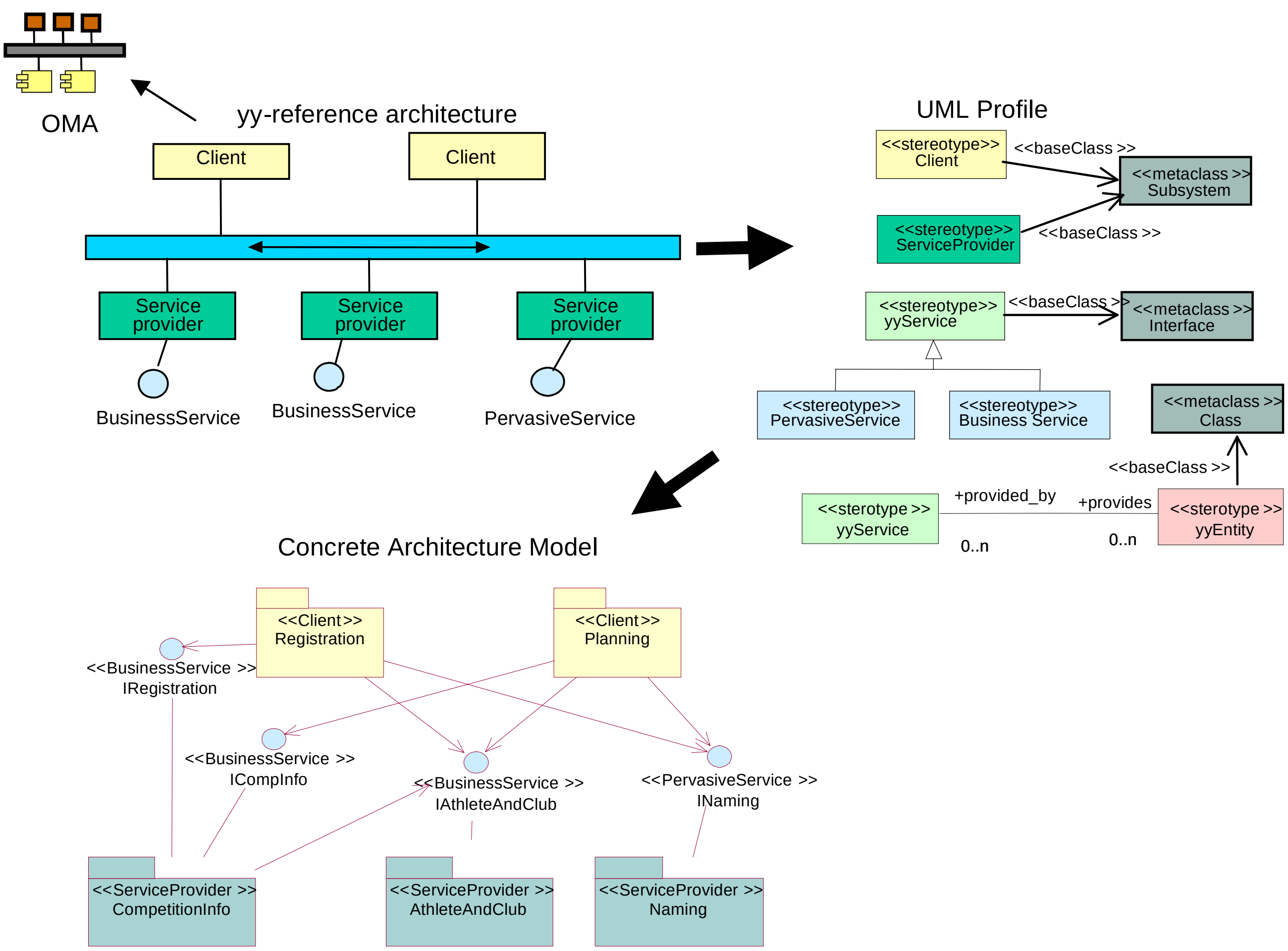
Figure 1 shows the structure of use of the framework. The tailoring framework defines how to build appropriate UML-profiles, patterns, and existing models, and it defines how to use the reference-model in correspondence with the chosen general purpose methodology. The domain-specific reference model is built using the tailoring framework and is structured according to the chosen methodology. The profile maker and pattern modeler have the main responsibilities of building the domain-specific reference model. The system modeler uses the chosen methodology's prescribed models and process, as well as the reference model for the actual domain, to build concrete models. The system modeler might also feed back reusable models to be part of the reference model. Those models will then be categorized according to the chosen methodology, and become an existing model of the domain-specific reference-model.

Figure 1 indicates use of a general purpose methodology, xx-methodology, which specifies three main models as outcomes of the development process: a business model, an architecture model, and a platform-specific model. The yy-reference-model (supporting the yy-domain or yy-product-family) is structured according to the xx-methodology and specifies a set of UML-profiles, patterns, and existing models in accordance with the model architecture of the xx-methodology (business, architecture, and platform-specific). The concrete zz-model consists of a business model, an architecture model, and a platform-specific model produced according to the xx-methodology and the yy-reference-model.

In the example of Figure 1, the domain reference model includes UML-profiles, patterns, and existing models at all model levels (business, architecture, and platform-specific). This is not required. However, the tailoring framework requires that the constituents of the domain reference model should be structured according to the model architecture of the chosen general purpose methodology. This will ensure commonality and standardization across domain-specific reference models.

The tailoring framework is generic and might be used to customize all UML-based, model-driven methodolo-

Figure 2. Reference architecture, UML-profile, and usage example



gies (including UP, RUP, TeMOD, and CLIMB). In principle, an infinite set of domain reference models supporting a specific domain or product family might be developed as customizations of a particular methodology (thus, a one-to-many relationship between methodology and reference model). There are also one-to-many relationships between methodology and concrete model, as well as between domain reference model and concrete model.

It is also plausible to use the tailoring framework to extend or specialize a specific domain reference model, for instance to support a sub-domain. In the following we describe how the three constituents of the reference model are applied. A more thorough example can found in Solberg, Oldevik, and Jensvoll (2002b).

UML-Profiles

The framework prescribes the common techniques for defining UML-profiles, using the UML extension mechanisms, stereotypes and tagged values, both of which extend the concepts we can work with in the UML world.

The UML-profiles of the framework describe the essential concepts from the domain in question. These profiles - that is, the concepts defined in these profiles (the stereotypes) - can be used as first-class modeling concepts when defining concrete models. Thus, customizing for an organization or a development group, the wanted domain modeling concepts are defined using UML profiles.

The framework recommends always developing a profile defining the reference-architecture used within the domain or product family. Such a profile will give essential support when modeling concrete system architectures. Figure 2 shows an example of defining the reference architecture for the yy-domain by a UML-profile. The architecture model prescribed by the general purpose methodology is then customized to support the reference architecture of the domain or product family. The figure shows the yy-reference-architecture for the yy-domain or the yy-product-family. The figure also indicates that the yy-reference-architecture is related to a standard bus architecture like the OMA.[3] The UML-profile extends the UML with the appropriate architectural concepts, which then become employed as first-class modeling concepts in the concrete architecture model.

Detailing of the UML-profile might be done in a tabular form as shown below (for the yyService and the yyEntity).

Similarly, UML-profiles might be made for all the model levels defined by the chosen methodology. For example, for the xx-methodology indicated in Figure 1, we might have domain profiles for the business, architecture, and platform-specific level defining the vocabulary to be used for modeling each of these levels respectively. An example of a platform-specific UML profile can be found in JSR (2001).

Patterns

Patterns represent special kinds of existing models that describe a recurring problem and suggest a general solution to that problem. The tailoring framework is used to define patterns and categorize them according the actual model architecture of the chosen methodology. For the xx-methodology in Figure 1, there might be business model patterns, architecture model patterns, and platform-specific model patterns. A pattern is employed by instantiating it (as a template) into a concrete model; the concrete model defines who/what is fulfilling the responsibilities defined by the roles in the pattern.

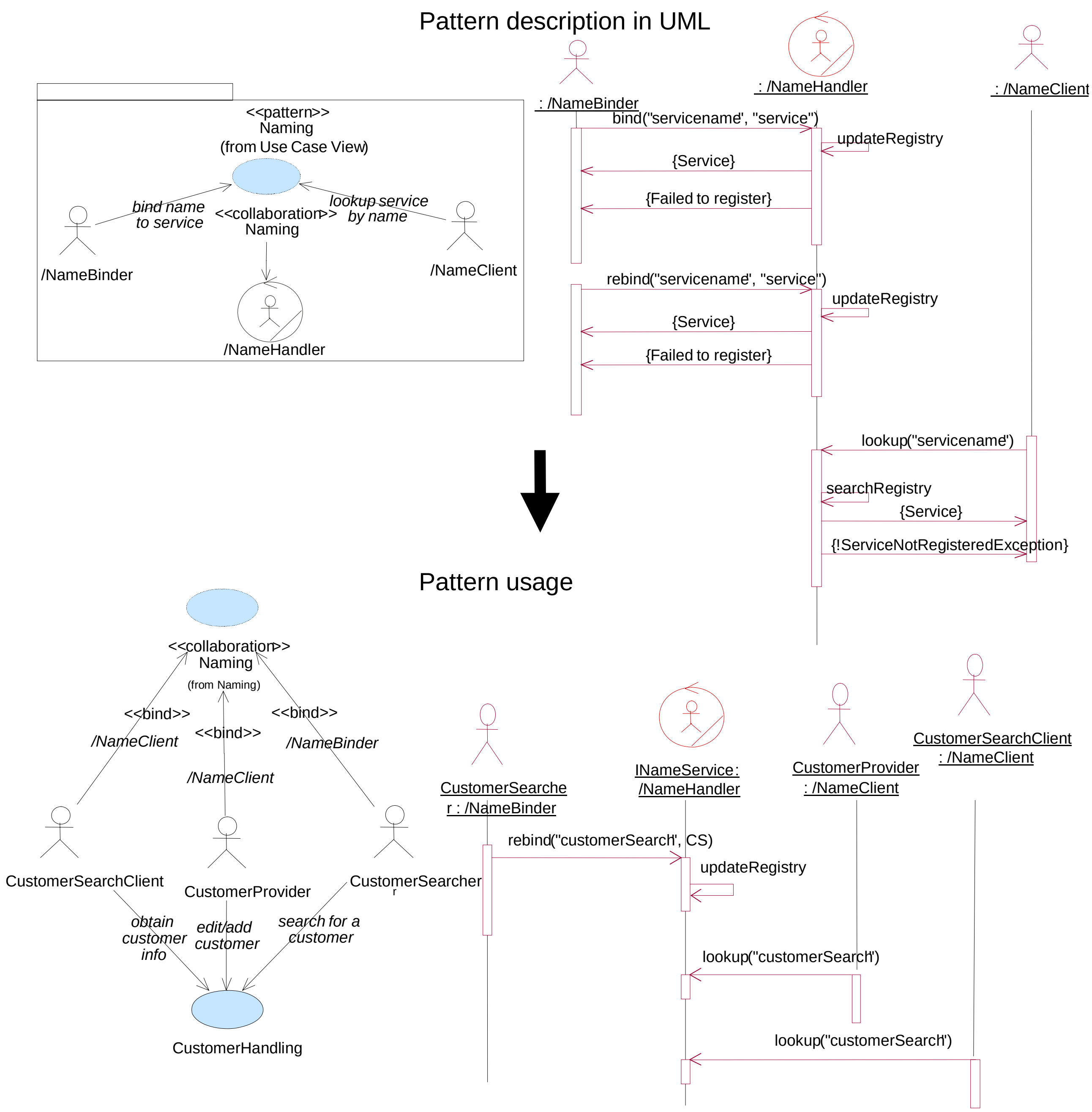
The tailoring framework includes a framework for pattern definition and use. This pattern framework includes both some special notation (defined in a UML-profile) and a template for pattern description.

The pattern structure technique of UML collaboration is used as the basis to define the pattern. The pattern structure technique is used for viewing the collaboration as a single entity from the outside. The collaboration details are further defined using UML activity diagram or UML sequence diagram. A simple example describing a pattern for a naming service and the usage of the pattern is shown in Figure 3.

UML has defined a naming convention denoting roles which allows a simple way of indicating a role. The general syntax is:

Stereotype	**Metamodel base**	**TaggedValues**	**Description**	**Constraints**
yyService	Interface	Transactional network-accessible	A service access interface, corresponds to a yyservice from the architecture profile.	self.allOppositeAssociationEnds -> forAll (a | a.type.oclIsTypeOf (yyService) or a.type.oclIsTypeOf (yyEntity))
yyEntity	Class	Persistent	Represents information provided through an yyservice.	self.allOppositeAssociationEnds -> forAll(a | a.type.oclIsTypeOf(yyEntity))

Figure 3. Pattern description and usage



ObjectName '/' ClassifierRoleName ':' ClassifierName [',' ClassifierName]*

The top left of Figure 3 shows the pattern structure in terms of roles collaborating to fulfill a mission defined by the unified responsibilities of the roles in the collaboration. The naming pattern defined includes three roles: NameBinder, NameClient, and NameHandler.

A collaboration is modeled as a use case stereotyped with <<collaboration>>. The roles are modeled as UML actors with role-names. These roles can either be external or internal to the pattern. It is the external roles that are parameterized when the pattern is used. The worker stereotype is used to denote internal actors. The NameHandler is an internal role in the example. The semantics of the collaboration 'use-case' is the same as a UML collaboration pattern structure (a use-case realization).

- The sequence diagram of the top right defines the behavior of the pattern by specifying interactions between roles.

In order to use a pattern, the desired external roles of the pattern must be instantiated. This is done by using a specialized '*binding*' relationship from the pattern collaboration source to the roles that instantiate the designated roles of the pattern. The role parameters are bound by the role-name specified on the binding relation, for example, '/NameClient'. The lower left of Figure 3 shows how a pattern can be instantiated by binding the roles from the pattern. There are no limits as to how many roles can be bound to another role or actor.

The sequence diagram at the lower right shows an example of synthesizing the pattern onto a specific architecture.

Existing Models

Existing models represent already defined concrete models that can be reused by package import and referencing. In the same way as for the UML-profiles and patterns, the tailoring framework prescribes that the existing models should be categorized and structured according to the model architecture of the chosen general purpose methodology. Typical usage of an existing model is to reuse, for example, an existing interface by inheritance or reference. Thus, the existing model part of the framework is to provide support for reuse also at the model level.

An existing model is reused in terms of package import, where the namespace defined by the package becomes available. All public artifacts (use-cases, actors, interfaces, classes, etc.) become available to the importing package. Alternatively, elements from an existing model can be used simply by scoped naming (UML pathnames), that is,

nnCore::PervasiveServices::NamingService::INaming,

which refers to the INaming interface contained in the package structure implied.

The general mechanism for reuse of a model, being it a pattern or a standard model, is by import of the package that defines the model. This makes the model elements defined in that package available (by model elements, meaning interfaces, classes, etc.). In principle, we can then reference interfaces from that package.

Package import is straightforward in UML, done with the <<import>> dependency stereotype between packages.

FUTURE TRENDS

A future trend will still be to configure model-driven, general purpose methodologies by means of selecting those parts of the general purpose methodology which seem relevant - for example, the set of roles, tasks, models, and process steps. This is basically how Rational Unified Process is configured to fit the actual purpose (Kroll & Kruchten, 2003). However getting better tool support for this configuration activity will be more in focus in the future. The RUP platform already provides some tool support for configuration (see *www-306.ibm.com/software/rational/*). Also, preconfigured methodologies for different domains and purposes will be provided, such as the dX process (Martin, 2004), which more or less is RUP made agile. However, these are still at a general purpose level and will typically need to be further customized to be appropriate for a specific company or development group.

Another trend in this regard is the development of domain languages in the context of the actual methodology and to utilize these in an MDA-based system development. The idea is to have your modeling concept formally defined and categorized in alignment with the set of main models you develop (e.g., Business Model, Architecture Model, and Platform-Specific Model), and to perform model transformation and code generation. This is typically accomplished by means of UML's extension mechanisms and the MOF (OMG, 2003b). There exists already many domain languages in the form of UML profiles; some are also standardized (e.g., JSR, 2001; OMG 2002). Defining your metamodel and syntax more or less from scratch, like in MetaCase *(www.metacase.com)* is another possible future trend in this respect.

The scoping for your language and your methodology is notified to be important as to get efficient methodologies providing efficient modeling concepts. This is for instance an important aspect of the System Family approach (Atkinson et al., 2001; Families, 2003), where the right level of scoping is said to be the System Family typically covering a set of tidy, related products where the commonality are substantial.

Also, formalizing and combining some of these ideas, like selecting parts of general purpose methodologies, appropriate scoping and making extensions of your modeling language to fit your domain or system family, and providing utilities for supporting reuse we believe will be a future trend. Our framework is one potential approach.

CONCLUSION

In this article, we have described a generic framework for customizing general purpose methodologies for the purpose of providing better support to specific needs within a domain or product family. An important aspect of this has been to ensure utilization of the existing knowledge possessed within the actual domain.

The tailoring framework introduces the need for new roles responsible for developing and maintaining reference models. The developers must also learn to use the actual reference model appropriately. Thus, successful introduction of the tailoring framework requires a well-defined process and careful consideration of the relevant risks.

The framework has already been used within the Telenor group and the EDB Telesciences, and we have seen several benefits from its application. It assists in:

- Establishing and maintaining models representing knowledge possessed within the domain.
- Model reuse: The reference-model advocates model-level reuse, leveraging existing models and patterns describing best practice for solving recurring problems within the domain. Current tool support is rather immature, so there is still potential for gaining substantial improvement of the efficiency of the model development with sufficient tool support.
- Reuse at the right level: It has been proven that efficient reuse is easier to gain within a product family community or a fairly small-scoped domain, as opposed to general purpose reuse in widely scoped domains. It is then easier to build a reusable asset library within the reach of the users.
- Ensuring consistency of a set of models: Customizing a general purpose methodology with stereotypes, common domain models, or common patterns will help make the models more consistent.
- Standardization: The reference model functions as the "standard" for the specific domain, without contradicting the prescriptions of the general purpose methodology which function as the standard for a set of domains (e.g., the enterprise as a whole).
- Adding more semantics to the models: The use of stereotypes and the description of patterns can help make the models more powerful and expressive for the readers and the modelers.
- Preparing for code generation: Code generation can be made more powerful when defining reference models, and can utilize the defined UML-profiles, patterns, and existing models.

REFERENCES

Allen, P. & Frost, S. (1998). *Component-based development for enterprise systems, applying the SELECT perspective.* SIGS Book and Multimedia.

Atkinson, C., Bayer, J., Bunse, C., Kamsties, E., Laitenberger, O., Laqua, R., Muthig, D., Paech, B., Wust, J. & Zettel, J. (2001). *Component-based product line engineering with UML, KOBRA.* Addison-Wesley.

COMBINE. (2000). *COMponent-Based INteroperable Enterprise system development.* ESPRIT IST-1999-20893. Retrieved from *www.opengroup.org/combine/overview.htm.*

COMBINE. (2003). *COMBINE methodology.* COMBINE ESPRIT IST-1999-20893 deliverable D23.

DAIM. (2001). *Distributed Architecture, Internet, and Multimedia, a research project sponsored by the Norwegian Research Council and the Software Industry Association of Norway.* Retrieved from *www.ikt-norge.no.*

D'Souza, D. & Wills, A.C. (1998). *The catalsysis approach.* Available from *www.catalysis.org.*

Families. (2003). *Eureka! 2023 programme.* ITEA Project ip02009. Retrieved from *www.esi.es/en/Projects/Families/.*

Frankel, D.S. (2003). *Model-driven architecture.* OMG Press.

Hallsteinsen, S.O, Solberg, A., Fægri, T.E., Oldevik, J. & Syrstad, M. (2003). *The DAIM software engineering handbook, version 1.0.*

Jacobson, I., Booch, G. & Rumbaugh J. (1999). *The Unified Software Development process.* Addison-Wesley.

JSR. (2001). Java community process. *UML profile for EJB.* JSR 26. Retrieved from *www.jcp.org/jsr/detail/26.jsp.*

Kruchten, P. (2000). *The Rational Unified Process: An introduction.* Addison-Wesley.

Kroll, P. & Kructhen, P. (2003). *The Rational Unified Process made easy: A practitioner's guide to the RUP.* Addison-Wesley (Object Technology Series).

Martin, R.C. (2004). Preliminary chapter. In G. Booch, R.C. Martin &J.W. Newkirk (Eds.), *Object-oriented analysis and design with applications* (3rd ed.). Addison-Wesley. Retrieved from *www.objectmentor.com/publications/RUPvsXP.pdf.*

OMG. (2001). *MDA™ Guide v1.0.1.* Retrieved from *www.omg.org/docs/omg/03-06-01.pdf.*

OMG. (2002). *UML profile for schedulability, performance, and time specification.* Retrieved from *www.omg.org.*

OMG. (2003a). *Unified Modelling Language (UML™) 1.5 specification.* Object Management Group.

OMG. (2003b). Meta-Object Facility (MOF) 1.4. Object Management Group.

Reenskaug, T. & Wold, L. (1996). *Working with objects. The OOram software engineering method.* Manning/ Prentice-Hall.

Solberg, A. & Oldevik, J. (2001, August). *Telenor methodology for interface modelling with UML, version 3.02.*

Solberg, A., Oldevik, J. & Jensvoll, A. (2002a). *Configurable LIght-Weight Method (CLIMB), version 1.0.*

Solberg, A., Oldevik, J. & Jensvoll, A. (2002b, May 19-22). A generic framework for defining domain-specific models. *Proceedings of the IRMA 2002 International Conference,* Seattle, WA.

KEY TERMS

Framework: The underlying structure supporting or containing something.

General Purpose Methodology: Methodology that attempts to be applicable across domains and for different kinds of systems and purposes.

Model-Driven Architecture (MDA): An approach under construction within the OMG trying to define and standardize model-driven software system development.

Model-Driven System Development: A software system development approach where the output of the development activities is a set of models. The code itself is seen as a model that is specified using programming language. The process is driven by the development of the set of prescribed models.

Pattern: Describes a recurring problem and suggest a general solution to that problem.

Software System Development Methodology: Defines techniques and mechanisms, as well as the process steps to go about developing software systems.

UML Profile: An extension of the UML metamodel typically used to define further concepts, e.g., domain-specific concepts. A set of UML profiles is standardized within the OMG (e.g., JSR, 2001; OMG, 2002). A UML profile is developed using the UML's extension mechanisms: stereotypes, tagged values, and constraints.

Unified Modeling Language (UML): A graphical specification language, standardized by Object Management Group (OMG).

ENDNOTES

[1] The major Norwegian telecom company *(www.telenor.com).*

[2] A company within the EDB, a major Norwegian software house *(www.edb.com; www.edbtele sciences.com).*

[3] Object Management Architecture defined by OMG (Object Management Group, *www.omg.org*).

Geographic Information Systems as Decision Tools

Martin D. Crossland
Oklahoma State University, USA

INTRODUCTION

Geographic information systems (GISs) as a technology have been studied and reported extensively and, not unexpectedly, in the field of geography. The various ways of capturing spatial data, arranging attribute data into appropriate database structures, and making the resulting large data sets efficient to store and query have been extensively researched and reported (Densham, 1991). However, the geographic research community has only recently noted the need to study how GISs are used as decision tools, especially with regard to how such decision making might be related to a decision maker's cognitive style (Mennecke, Crossland, et al., 2000). As an example, the University Consortium for Geographic Information Science called for research examining how geographic knowledge is acquired through different media and by users with different levels of experience and training (University Consortium for Geographic Information Science, 1996).

Researchers in the fields of decision sciences and information systems have more recently begun to make contributions in the area of decision making with GISs. When a GIS is employed as a decision support system, in these studies the resultant system is often referred to as a *spatial decision support system*, or *SDSS* (see Crossland, 1992; Crossland, Perkins, et al., 1995; Mennecke et al., 2000).

A *geographic information system* in its simplest form is a marriage of accurately scaled digital maps with a database. The digital maps comprise spatially referenced details such as natural elements (lakes, rivers, topographic elevation contours, etc.), manmade objects (buildings, roads, pipelines, etc.), and political boundaries (city limits, state and county lines, international boundaries, etc.). These natural elements are typically referenced, with varying degrees of precision, to latitude/longitude coordinates on the earth's surface. It must be noted here that the degree of precision and, more importantly, differences in degrees of precision for the various elements are the subjects of much research and user consternation in applications of GISs to solving problems. The database, in turn, catalogs information about the various spatial elements (e.g., the names of rivers, names of buildings, building owner, operator of a pipeline, etc.). These descriptive entries in the database are often referred to as *attributes* of the various spatial elements.

A GIS may be paired with the *global positioning system* (*GPS*), from which real-time, satellite-derived location information may be derived, as provided by an appropriate GPS receiver.

BACKGROUND

With regard to the effectiveness of decision making when using information tools, there is a relatively long history of researchers emphasizing that tools which provide graphical presentations and graphical representations of information are deserving of special note and study. For example, Ives (1982) discussed at great length the role of graphics in business information systems. He even went so far as to state, "The map, perhaps more than any other chart form, gains the most from the availability of computer graphics" (p. 16).

Several more recent studies have drawn from Image theory (Bertin, 1983) to help explain why decision makers using GISs may experience greater effectiveness in decision making. Image theory states that one graphical representation of information may be considered more efficient than another for a particular question, if that question can be answered in the mind of the decision maker in a lesser amount of time. In his *Semiology of Graphics*, Bertin defined image theory and put forth the constructs of images and figurations. An *image* is a meaningful visual form, perceptible in a minimum instant of vision. A *figuration* is a more complex construction comprising multiple images. Figurations are inherently less efficient than images, according to image theory. This is because the viewer is able to grasp the full informational content of an image in a brief moment of viewing it. Figurations, on the other hand, comprise multiple images which must be mentally extracted, processed, and related in the viewer's perception. Although the informational content may be richer in a figuration, it is inherently less efficient for quick extraction of specific information.

The more recent studies propose that one role of GISs is to collapse more complex figurations into simpler figurations or even to simple images. This has the net effect of increasing a decision maker's efficiency in extracting relevant information for the purpose of evaluating and making a decision. For examples the reader is encouraged to review Crossland (1992), Crossland, Herschel, et al. (2000), Crossland et al. (1995), and Mennecke et al. (2000).

Although there seems to be a common assumption that GISs improve decision making (Morrison, 1994), only a few studies to date have performed controlled experiments to actually test this assumption. Those that have been accomplished typically used dependent variables of decision time and decision accuracy to measure decision-making effectiveness. These include Crossland (1992), Dennis and Carte (1998), Mennecke et al. (2000), Smelcer and Carmel (1997), and Swink and Speier (1999). All of these studies found that the addition of a GIS to a spatially referenced decision-making task had a positive effect on decision outcomes.

THE ROLE OF COGNITIVE STYLE IN DECISION MAKING WITH GISS

With respect to decision making, the term *cognitive style* has been used to refer to enduring patterns of an individual's cognitive functioning that remain stable across varied situations. Various elements of cognitive style have been speculated upon and studied in various disciplines. With respect to decision making using GISs, two elements have been studied in some depth, *field dependence* and *need for cognition*.

Field dependence (FD) measures a person's ability to separate an item from an organized field or to overcome an embedded context (Witkin, Lewis, et al., 1954). Zmud and Moffie (1983) proposed that people with lower field dependence tend to outperform those with higher field dependence in structured decision tasks and that they tend to make more effective use of transformed information (e.g., aggregated values and graphical formats, such as are typically found in a GIS). FD can be measured using commercially available testing instruments. Because making decisions using a GIS, by its nature, involves mentally extracting relevant information from a potentially complex field of information, studies have hypothesized that low field dependence should predict better decision making with a GIS or other spatially referenced tool. In particular, field dependence is seen as an inverse proxy for an individual's level of spatial cognition—the ability of an individual to grasp and analyze information within a spatial context.

Need for cognition (NFC) was proposed by Caccioppo and Petty (1982) as a measure of a person's internal motivation to pursue and enjoy cognitive tasks and activities. They developed a questionnaire which can be used to measure this cognitive-style attribute. People who score high on the need for cognition scale tend to enjoy the engagement of thought activity in a task as much or more than even the result of a task. The studies named below hypothesized that this tendency to engage more fully in a task should lead to more effective decision making, as measured by the dependent variables of decision time and decision accuracy.

Studies that looked at FD, NFC, or both as independent variables of decision-making performance using GISs include Crossland (1992), Crossland et al. (1995), and Mennecke et al. (2000). In general, the findings may be summarized as follows:

- Field dependence exhibits an inverse main effect on decision time, but not on decision accuracy. That is, subjects with lower field dependence tend to solve spatially referenced problems more quickly, but not more accurately. It may be that the efficiency predicted by image theory does contribute to faster decision making, but not to more accurate decisions.
- Need for cognition exhibits a positive main effect on decision accuracy, but not on decision time. That is, higher-NFC subjects tend to solve spatially referenced problems more accurately, but not more quickly. This last finding was noted as unexpected by Crossland (1992). He speculated that perhaps an individual with a high NFC might tend to spend longer in thinking about the problem and its solution, thereby extending the decision time. It would seem, however, that this extra thinking effort may have contributed to a more accurate solution.

FUTURE TRENDS

Some questions and issues in this area of research that remain to be addressed include:

- How do other important measures of cognitive style affect a decision maker's ability to solve spatially referenced problems accurately and quickly?
- How does problem complexity factor in or even interact with the decision maker's task? Several studies also examined problem complexity as an independent variable (Crossland, 1992; Crossland et al., 1995; Mennecke et al., 2000). Crossland et al. (1995) reported an observed interaction of field dependence with problem complexity that would be interesting to explore further.
- To what extent are SDSSs/GISs effective in collapsing figurations (as defined by image theory) into

images or into simpler figurations? Are there certain levels of complexity beyond which it becomes impractical or ineffective to combine or collapse displays into simpler decision tools? How does the cognitive style of the decision maker factor into this consideration?

- Are SDSSs/GISs even necessary for certain types of problems? Perhaps a series of static, hard-copy outputs are sufficient for some decisions by some decision makers, and the combined or flattened displays are not necessary. The cognitive style of the decision maker may be an important factor in this.
- Studies in this area may be useful in understanding more generally how technology is useful in supporting decision makers in other contexts. For example, Vessey (1991) and Vessey and Galletta (1991) suggested that three variables would influence the mental representation that the decision maker develops: (1) the problem representation, (2) the problem-solving task, and (3) the decision maker's problem-solving skills.
- It may be useful to apply cognitive fit theory to these types of problem-solving tasks, as proposed by Mennecke et al. (2000), to better understand the factors that appear to be important in influencing a user's formation of a mental representation of spatial tasks.

CONCLUSION

The roles of components of cognitive style of a decision maker are important factors in how well he can solve problems and carry out spatially referenced tasks when using a GIS. Although current research has shown that simply using a GIS can enhance decision-making effectiveness, the cognitive style of the subject is a less well understood element of the process. More research is needed in this area to better define and understand it.

REFERENCES

Bertin, J. (1983). *Semiology of graphics: Diagrams, networks, maps*. University of Wisconsin Press.

Caccioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *4*(1), 116-131.

Crossland, M. D. (1992). *Individual decision-maker performance with and without a geographic information system: An empirical investigation*. Unpublished doctoral dissertation, , Indiana University, Bloomington.

Crossland, M. D., Herschel, R. T., et al. (2000). The impact of task and cognitive style on decision-making effectiveness using a geographic information system. *Journal of End User Computing*, *12*(1), 14-23.

Crossland, M. D., Perkins, W. C., et al. (1995). Spatial decision support systems: An overview of technology and a test of efficacy. *Decision Support Systems*, *14*, 219-235.

Dennis, A. R., & Carte, T. (1998). Using geographical information systems for decision making: Extending cognitive fit theory to map-based presentations. *Information Systems Research*, *9*(2), 194-203.

Densham, P.J. (1991). Spatial decision support systems. In P. J. Densham, M. F. Goodchild, & D. W. Rhind (Eds.), *Geographical information systems: Principles and applications* (Vol. 2, pp. 403-412). London: Longman Scientific & Technical,

Ives, B. (1982). Graphical user interfaces for business information systems. *MIS Quarterly*, 15-42.

Mennecke, B. E., Crossland, M. D., et al. (2000). Is a map more than a picture? The role of SDSS technology, subject characteristics, and problem complexity on map reading, and problem solving. *MIS Quarterly*, *24*(4), 601-629.

Morrison, J. L. (1994). The paradigm shift in cartography: The use of electronic technology, digital spatial data, and future needs. In T. C. Waugh & R. G. Healey (Eds.), *Advances in GIS research* (pp. 1-15). London: Taylor and Francis.

Smelcer, J. B., & Carmel, E. (1997). The effectiveness of difference representations for managerial problem solving: Comparing tables and maps. *Decision Sciences*, *28*, 391-420.

Swink, M., & Speier, C. (1999). Presenting geographic information: Effects of data aggregation, dispersion, and users' spatial orientation. *Decision Sciences*, *30*(1), 169-195.

University Consortium for Geographic Information Science. (1996). Research priorities for geographic information science. *Cartography and Geographic Information Systems*, *23*(3), 1-18.

Vessey, I. (1991). Cognitive fit: Theory-based analysis of the graphs vs. tables literature. *Decision Sciences*, *22*(1), 219-241.

Vessey, I., & Galletta, D. (1991). Cognitive Fit: An empirical study of information acquisition. *Information Systems Research*, *2*(1), 63-84.

Witkin, H. A., Lewis, H. B., et al. (1954). *Personality through perception*. New York: Harper.

Zmud, R. W., & Moffie, R. P. (1983). The impact of color graphic report formats on decision performance and learning. *International Conference on Information Systems*.

KEY TERMS

Attributes: are the pieces of information contained in a GIS database that describe or detail a spatially referenced element.

Cognitive Style: refers to enduring patterns of an individual's cognitive functioning that remain stable across varied situations.

Digital Map: any form of geographic boundaries or spatially referenced drawings that have been captured, or "digitized," into an electronic form. Each element of the map is or may be linked to various descriptive or identifying types of information in a database.

Field Dependence (FD): measures a person's ability to separate an item from an organized field or to overcome an embedded context.

Figuration: as defined in image theory, is a complex construction comprising multiple images. Figurations are inherently less efficient for extracting information than images, according to image theory.

Geographic Information System (GIS): a marriage of accurately scaled digital maps with a database. The digital maps comprise spatially referenced details such as natural elements (lakes, rivers, topographic elevation contours, etc.), manmade objects (buildings, roads, pipelines, etc.), and political boundaries (city limits, state and county lines, international boundaries, etc.). These natural elements are typically referenced, with varying degrees of precision, to latitude/longitude coordinates on the earth's surface.

Global Positioning System (GPS): provides real-time, satellite-derived location information based on information received by an appropriate GPS receiver. GPS is funded by and controlled by the U.S. Department of Defense (DOD). While there are many thousands of civil users of GPS worldwide, the system was designed for and is operated by the U.S. military. A GPS may be employed in the original construction of the digital map information to be stored in a GIS. Or, if the GIS is already constructed, the GPS may be employed to accurately render the position of new elements to be added to the GIS or the current position of a mobile element to be referenced against the information stored in the GIS. A good example might be a freight truck moving on a highway. The GPS receiver on the truck can derive its current latitude and longitude and then send that information to the GIS system in the truck cab, to a GIS in a central control center via radio, or to both for subsequent reporting and analysis.

Image: as defined in image theory, is a meaningful visual form, perceptible in a minimum instant of vision.

Need for Cognition (NFC): a measure of a person's internal motivation to pursue and enjoy cognitive tasks and activities.

Spatial Decision Support System (SDSS): typically, a geographic information system (GIS) that has been extended to provide knowledge workers with decision-making tools and support data.

Spatially Referenced Data: entities, typically recorded as records in a database, which have some notion of a definite location in space.

Geospatial Information Systems and Enterprise Collaboration

Donald R. Morris-Jones
SRA, USA

Dedric A. Carter
Nova Southeastern University, USA

INTRODUCTION

Organizations and teams are becoming increasingly more distributed as groups work to expand their global presence while rationalizing team members across skill sets and areas of expertise instead of geographies. With this expansion comes the need for a robust and comprehensive language for pinpointing locations of globally distributed information systems and knowledge workers. Geospatial information systems (GISs) provide a common framework for jointly visualizing the world. This shared understanding of the world provides a powerful mechanism for collaborative dialogue in describing an environment, its assets, and procedures. The collaborative framework that GIS provides can help facilitate productive dialogue while constraining impulses of extreme positions. Collaboration and GIS intersections take many forms. Under a collaborative work-flow model, individuals use GIS to perform their job and post data back to the central database (e.g., engineering designs and as-built construction).

This article addresses the increasing role of GIS in emerging architectures and information systems in a number of applications (e.g., land planning, military command and control, homeland security, utility-facilities management, etc.). Real-time applications, mobile access to data, GPS (global positioning satellite) tracking of assets, and other recent developments all play a role in extending the scope and utility of the GIS-enabled enterprise. The impact of new GIS Web services standards and open geospatial-data archives are also addressed as areas of increased potential for remote GIS collaboration in global organizations. The expansion of enterprise GIS within organizations increases the opportunity and necessity of using GIS collaboratively to improve business processes and efficiency, make better decisions, respond more quickly to customers and events, and so forth.

BACKGROUND

The term *geospatial* is increasingly used to describe digital data about the earth in GIS, image, or GPS formats. The related technologies of GIS, remote sensing image-processing systems, and GPS data collection are all components of geospatial information systems. Geospatial-technologies use continues to expand in a great variety of applications ranging from land planning to utility-engineering design and military command and control. Those applications, which were once relegated to discrete groups of specialists, have now begun to take a more prominent role in the enterprise. Duffy (2002) describes the transition of GIS from a specialist technology to a more mainstream environment in the industry information-systems department from the end of the last decade into 2002.

The essence of collaboration is people and organizations working together to accomplish a common goal. Information–technology- (IT) enabled collaboration has improved business processes in many organizations and contributed to more functional and profitable operations. Collaboration technologies are characterized by three major generic attributes: communication, information sharing, and coordination (Munkvold, 2003). These characteristics can be further refined into available channels such as synchronous or asynchronous, the medium of sharing information through repositories or real-time interaction, and work-flow management to coordinate steps in a decision process or protocol. Geospatial technologies and systems extend collaboration in unique ways for problems that are related to location.

GIS provides a geographic dimension to enterprise collaboration, which helps solve a variety of problems that are difficult to address by any other means. For example, vehicle-routing and dispatching applications make it possible for Sears to deliver goods to customers more efficiently within tighter time windows. As a result, Sears is more profitable and customers are more satisfied.

This example of distributed-network optimization using efficient queuing mechanisms based on location information is a simple illustration of the impact that GIS data may have on existing business processes. In fact, most aspects of business-process automation initiatives at present require some element of collaboration either between networked systems or dispersed individuals.

Collaboration utilizing GIS and geospatial frameworks continues to be a focus of research both in the United States and abroad (Boettcher, 2000; Songnian, 2004; Stasik, 2000).

EMERGING GEOSPATIAL INFORMATION-SYSTEMS ARCHITECTURES AND COLLABORATIVE ENTERPRISE APPLICATIONS

As organizations become more dispersed in an effort to rationalize across areas of expertise in lieu of geographies, complex infrastructures for location analysis and coordination may emerge (Munkvold, 2003). In recent years, GIS software companies have developed an expanding and increasingly capable enterprise suite of tools. Early generations of GIS were used by GIS specialists only; these systems were available in stand-alone or project-systems configurations. GIS product options have improved and now provide a sound basis for supporting casual users as well as specialists with desktop, distributed client-server, and Internet solutions.

Geospatial data standards and interoperability have greatly improved the ease of using data in different formats or geographic projections. Geospatial Web-services standards provide Internet access to geospatial data stored in geospatial-data archives. Federal-government initiatives (e.g., Geospatial One Stop, the National Map, Homeland Infrastructure Foundation Level Database, etc.) will increase data standardization and access, and reduce expensive, redundant data collection.

GIS-enabled collaboration can now involve a broad range of different types of users within and outside of a particular organization. These users can be expert or casual as well as stationary or mobile. Medeiros, de Souza, Strauch, and Pinto (2001) present an analysis of aspects of coordination in a collaborative system for spatial group-decision support that resulted in a prototype system for a distributed GIS.

Geospatial Information-Systems Products and Architectures

GIS-product vendors continue to innovate and expand the solution set available to the user (Atkinson & Martin, 2000).

Enterprise suites of GIS include the following different types of products.

- Desktop GIS with varying levels of functionality
- Spatial analysis extensions
- Internet GIS with limited functionality or full functionality
- Mobile GIS
- Geospatial-data middleware
- Software to embed geospatial functionality in business applications
- 3D GIS
- Geospatial-data visualization software
- GPS tracking software
- Remote sensing image-processing software
- Geospatial Web-services software
- Location services

GIS products are available to support stand-alone users, and distributed client-server and centralized Internet architectures. GPS tracking units and mobile GIS on Personal Digital Assistants (PDAs) and pocket PCs extend the range of the technology into the field. Wireless communication of data is improved through the use of data compression and area-of-interest extraction techniques.

Geospatial-data management functionality is improving but is less capable than business-data management functionality. While Oracle states that their products now provide equivalent data-management functionality for spatial and business data, experience is limited for enterprise replication of geospatial data. ESRI, the GIS-software market leader, promises to add geospatial-data replication to its ArcSDE product with the release of ArcSDE 9.1, which is projected for 2005. The large size of geospatial-data files means that substantial bandwidth is needed to move data through a communications network.

Location services refer to mobile geospatial services that will primarily be delivered to location-aware smart phones. The E-911 legislation mandates that cell phones must become location aware so emergency vehicles can locate 911 callers who use cell phones. Cell-phone operators and partners are and will offer an increasing array of location services to provide users with directions for driving, the nearest services of different types, and the location of buddies (i.e., those who have authorized sharing this information, etc.). An example system is presented in the location-based tourist-guide application of Simcock, Hillenbrand, and Thomas (2003). This tool combines a mobile PDA device and GPS technology to provide the user with location tours that are self-guided.

Collaborative Land-Use Planning

Geospatial data provides a common view or abstraction of an area. GIS has been used extensively in land-use planning. While GIS initially served and continues to serve as a tool for planners, it is increasingly used to facilitate collaboration with the public. The ability to show land and its characteristics to groups of people with divergent views provides a common frame of reference, which can make it easier to develop consensus or agreement on difficult issues. The common GIS data framework serves to constrain more extreme positions, which are more likely to be presented without geospatial data. The implementation of the collaborative system in Medeiros et al. (2001) described earlier was targeted as a land-use and planning application for a distributed GIS.

The Urban and Regional Information Systems Association (URISA) is in the third year of hosting an annual conference on the topic of public participation in GIS (URISA, 2004b; Voss et. al., in press). A discussion of presentations at the second conference in the series mentioned that Internet GIS technology was used to present and solicit public comment regarding alternative proposed designs for the World Trade Center buildings and parks (URISA, 2004a). From the number of questions and issues that were identified in the conference, it is clear that there is not yet a commonly accepted model for how GIS should be used to facilitate public participation in planning. As the role and power of planners, politicians, GIS specialists, developers, and the public could potentially shift with more publicly available geospatial data and analytic capabilities, there is no simple answer to the question of how collaboration between government and the public should be enhanced using GIS. Clearly, the technology increases options for public participation in land planning.

The Orton Family Foundation, a Vermont-based non-profit organization dedicated to better decision making by communities, has developed 2D and 3D GIS software to help people develop, visualize, and analyze the implications of alternative approaches to land- and growth-management planning (Orton Family Foundation, 2004). Their CommunityViz GIS software has been used extensively by smaller communities to help involve the public in land planning. In Eureka Township, Minnesota, GIS was used to help the Eureka Township Envisioning Project understand the impacts of alternative growth-management scenarios on septic-system placement and water quality (Orton Family Foundation).

Utilities-Engineering Design and Construction

Utilities use geospatial-information systems to plan and manage their assets and design new components of their outside plant infrastructure (e.g., poles, wires, transformers, pipes, etc.). While most of the collaboration occurs within the utility, collaboration occurs to some degree with developers and government officials. For example, a developer will submit their design for a new subdivision, usually in CAD (computer-aided design) format. Field surveys are often conducted at the site and these surveys are increasingly conducted using GPS technology. The utility designer will import this CAD into either a CAD or GIS tool, which will be used to design the utility infrastructure that is needed to support the proposed development. Typically, the proposed design is analyzed as if it were connected to the network, so a proposed version of the design is developed. Alternative designs may be developed and costs of the alternatives may be determined so that the most cost-effective option can be chosen. Designs may be reviewed by other designers or a design supervisor to select the preferred option, which may also be with the developer and government officials. When approved for construction, the design will be attached to a work order and assigned to a crew which will be dispatched to the construction site. The design may be modified in the field if logistical difficulties are encountered. Construction crews will record as-built conditions for the project, and this data will be reviewed and posted to the database, which represents the utility assets that are constructed.

This utility design and construction process occurs over an extended period of time, which makes it a long transaction. Engineers will typically extract a version of the database to use to develop their designs. Usually, a utility will use optimistic logic and not formally lock the portion of the database that is included in the versioned database. Conflicts between alternative designs in the same general area are flagged for resolution by an engineering supervisor.

This CAD- and GIS-based process represents a major increase in productivity in design and maintenance of utility maps and records. Utility GIS practitioners often network through an organization called the Geospatial Information & Technology Association (GITA, 2004).

C4ISR and Defense Geospatial Applications

Command, control, communications, computers, intelligence, surveillance, and reconnaissance (C4ISR) is a military application that makes extensive use of geospatial technologies. The National Geospatial-Intelligence Agency, the National Reconnaissance Office, and other agencies collect imagery and produce different types of geospatial data for military and intelligence purposes. For example, cruise missiles use topographic data as a model of the terrain. The terrain is sensed by the cruise missile and pattern-matching techniques are used to follow the landscape to reach a target. High-resolution imagery is interpreted to help select targets.

C4ISR provides the military forces with geospatially enabled command and control capabilities. Trends are to make C4ISR more widely available to help cut through the fog of war and provide a better coordinated understanding of the battle space. The Geo-Intel 2003 conference included discussions of the importance of geospatial data and technologies and the role played by them in the rapid and successful movement of U.S. troops in the Iraqi War.

In the future, geospatial intelligence and C4ISR will be more ubiquitous. Currently, the military is still in the process of adopting geospatial data for battle-space operations. The Commercial Joint Mapping Toolkit initiative could place GIS-enabled C4ISR in the hands of up to one million U.S. military forces.

Homeland-Security Command Centers

Homeland–security command centers are making effective use of GIS to provide planning and response capabilities. The emphasis on situational awareness and command and control is similar to DOD C4ISR applications. 3D GIS applications developed by IT Spatial are playing a major role in command centers in Washington, DC, and elsewhere. Plume modeling software combined with demographic data, 3D GIS data, and real-time weather data provide a capability to predict terrorist acts, like a dirty nuclear explosion, and their consequences. There is no shortage of investigation of 3D GIS technologies (Manoharan, Taylor, & Gardiner, 2002). The better the consequences are understood, the more effective the response can be from emergency first-responder forces. First responders in the field can access data and also feed back information regarding emergencies and events to provide command and control personnel and other responders with better understanding. This impact of geospatial information in emergency situations has been the subject of research by several in the wake of the attacks of September 11, 2001 (Kevany, 2003; Kwan & Lee, in press; Rauschert, Agrawal, Sharma, Fuhrmann, Brewer, & MacEachren, 2002), in addition to some early research examining this area prior to the attacks (Kumar, Bugacov, Coutinho, & Neches, 1999). GPS tracking can be used to understand the location of police cars, ambulances, and other emergency vehicles. Integration of GIS with live video data feeds also provides a useful real-time picture of ground conditions and events. There are some, however, that continue to point out the numerous impediments to effectively using GIS in disaster-control decision situations (Zerger & Smith, 2003).

GIS additionally continues to play an important role in the containment of medical outbreaks and the analysis of systems mapped to locations, such as in the work of Cockings, Dunn, Bhopal, and Walker (2004).

Presence Awareness in Instant Messaging and RFID-Based Asset Tracking

In an effort to better display the impact of emerging technologies such as radio-frequency identification (RFID), Web services on GIS, and collaboration for real-time presence management, AMS, through its Center for Advanced Technologies, customized the BuddySpace instant messaging (IM) product to allow IM users the ability to display the location of other users. This information was automatically captured through the implementation of a GIS Web service and an RFID network (Del Vecchio & Carter, 2004). This location or presence awareness adds an additional useful dimension to dialogue between IM users. There are alternative methods of collecting the location data and these options include

- entering and geocoding addresses,
- GPS data collection, and
- RFID-tag readings.

The RFID-tag approach is interesting due to future expectations for expansion in the use of this technology. RFID tags must be read by a stationary reading device, and the location of this reader can be determined by address geocoding or GPS. RFID tags that are read can be related to their location, and this could prove useful for supply chain applications. This experiment builds on the assertions of Mitchell (2003), which state that Web services will further propel GIS into the spotlight.

FUTURE TRENDS

Geospatial technologies have begun to play mission-critical roles on a large scale in a variety of different types of organizations. Early adopters of geospatial technologies included many organizations that traditionally have maintained maps and records to manage their assets and operations. Geospatial functionality will increasingly be embedded in various business applications (e.g., SAP ERP) that might support almost any organization with simple geospatial data visualization, queries, and analysis. In addition, spatial extensions to commercial RDBMS products will permit organizations to visualize the geospatial dimensions of their business data without a major investment.

The ability to deliver data in real time to mobile users extends GIS to the field. The number of mobile GIS users and the wireless communication of data to and from these mobile users will continue to expand. Location services will deliver focused geospatial-information services (e.g., directions, nearest services, etc.) to cell-phone users willing to pay for these supplemental options. According to Nellis (2004), these innovations are "…now at the heart of a vast array of real-time interactive mobile computing, geo-location applications and asset management, along with wireless geographic services that are revolutionizing the role of geography and geospatial information analysis in meeting the needs of everyday society."

The high cost of geospatial data has been a major impediment to its adoption. As geospatial data are increasingly available from geospatial Web-services-compliant sites, costs to access data will decline and users will be able to develop and use applications more rapidly. There will continue to be some significant policy concerns and debate with regard to the privacy implications of GIS technology as outlined in Balough (2001).

CONCLUSION

Geospatial information systems provide unique capabilities for geospatial-data visualization and analysis. While the large data sizes associated with this technology make real-time and mobile uses more challenging, it is now possible, and increasingly practical and necessary, to use GIS across an enterprise for real-time, mobile geospatial applications. Military and homeland-security markets are currently driving the market for collaborative GIS enterprise applications. Other markets have also emerged and more will surely follow as the advantages of location awareness and geospatial analysis are better appreciated.

REFERENCES

Atkinson, P., & Martin, D. (2000). Innovation in GIS application. *Computers, Environment, and Urban Systems, 24*, 61-64.

Balough, R. C. (2001). *Global Positioning System and the Internet: A combination with privacy risks* (Excerpt from the Chicago Bar Association's CBA record). Retrieved January 20, 2004, from *http://www.isoc.org/internet/issues/privacy/balough.shtml*

Boettcher, R. L. (2000). Collaborative GIS in a distributed work environment. *Master's Abstracts International, 38*(4), 1097-1190. (UMI No. AAT MQ46234).

Cockings, S., Dunn, C. E., Bhopal, R. S., & Walker, D. R. (2004). Users' perspectives on epidemiological, GIS, and point pattern approaches to analyzing environment and health data. *Health & Place, 10*, 169-182.

Del Vecchio & Carter, D. A. (2004). Enhanced presence management in real-time instant messaging systems. *Proceedings of the 14th Information Resource Management International Conference.*

Duffy, D. (2002, August 1). GIS goes worldwide. *CIO Magazine*. Retrieved August 20, 2002, from *http://www.cio.com*

Geospatial Information & Technology Association (GITA). (2004). GITA Web site. Retrieved March 15, 2004, from *http://www.gita.org*

Kevany, M. J. (2003). GIS in the World Trade Center attack: Trial by fire. *Computers, Environment, and Urban Systems, 27*, 571-583.

Kumar, V., Bugacov, A., Coutinho, M., & Neches, R. (1999). Integrating geographic information systems, spatial digital libraries and information spaces for conducting humanitarian assistance and disaster relief operations in urban environments. *Proceedings of the Seventh ACM International Symposium on Advanced Geographic Information Systems*, 146-151.

Kwan, M. P., & Lee, J. (in press). Emergency response after 9/11: The potential of real-time 3D GIS for quick emergency response in micro-spatial environments. *Computers, Environment, and Urban Systems*.

Manoharan, T., Taylor, H., & Gardiner, P. (2002). A collaborative analysis tool for visualization and interaction with spatial data. *Proceedings of the Seventh International Conference on 3D Web Technology*, 75-83.

Medeiros, S. P. J., de Souza, J. M, Strauch, J. C. M., & Pinto, G. R. B. (2001). Coordination aspects in a spatial group decision support collaborative system. *Proceedings of the 2001 ACM Symposium on Applied Computing*, 182-186.

Mitchell, R. L. (2003, December 15). Web services put GIS on the map. *Computerworld*. Retrieved December 15, 2003, from *http://www.computerworld.com*

Munkvold, B. E. (2003). *Implementing collaboration technologies in industry: Case examples and lessons learned*. London: Springer-Verlag.

Nellis, D. M. (2004, February 20). Geospatial information, cybergeography, and future worlds. *Directions Magazine*. Retrieved March 15, 2004, from *http://www.directionsmag.com*

Orton Family Foundation. (2004). CommunityViz Web site. Retrieved March 15, 2004, from *http://www.communityviz.org*

Rauschert, I., Agrawal, P., Sharma, R., Fuhrmann, S., Brewer, I., & MacEachren, A. (2002). Designing a human-centered, multimodal GIS interface to support emergency management. *Proceedings of the 10th ACM International Symposium on Advanced Geographic Information Systems*, 119-124.

Simcock, T., Hillenbrand, S. P., & Thomas, B. H. (2003). Developing a location based tourist guide application. *Proceedings of the Australian Information Security Workshop Conference on ACSW Frontiers 2003, 21*, 177-183.

Songnian, L. (2004). Design and development of an Internet collaboration system to support GIS data production management. *Dissertation Abstracts International*, *64*(7), 3147-3487. (UMI No. AAT NQ82574)

Stasik, M. I. (2000). Collaborative planning and decision-making under distributed space and time conditions. *Dissertation Abstracts International, 60*(7), 2629-2758. (UMI No. AAT 9938995)

Urban and Regional Information Systems Association. (2004a). *Hotbutton questions/issues from 2002 conference*. URISA Public Participation GIS Web site. Retrieved March 15, 2004, from *http://www.urisa.org/ppgis.html*

Urban and Regional Information Systems Association. (2004b). URISA Public Participation GIS Web site. Retrieved March 15, 2004, from *http://www.urisa.org/ppgis.html*

Voss, A., Denisovich, I., Gatalsky, P., Gavouchidis, K., Klotz, A., Roeder, S., et al. (in press). Evolution of a participatory GIS. *Computers, Environment, and Urban Systems*.

Zerger, A., & Smith, D. I. (2003). Impediments to using GIS for real-time disaster decision support. *Computers, Environment, and Urban Systems, 27*, 123-141.

KEY TERMS

C4ISR: Command, control, communications, computers, intelligence, surveillance, and reconnaissance, a military application framework that makes extensive use of GIS technologies.

CAD: Computer-aided design.

Enterprise Collaboration: Application of systems and communications technologies at the enterprise level to foster the collaboration of people and organizations to overcome varying levels of dispersion to accomplish a common goal. This term, when applied to the use of technologies, is also known as e-collaboration or distributed collaboration.

Geospatial Information Systems (GISs): Systems that provide a common framework for jointly visualizing the world.

Global Positioning Satellite (GPS): A format of presenting geospatial data for tracking purposes used often in location-based services.

RFID: Radio-frequency identification. This technology uses the electromagnetic spectrum radio signals to transmit information from a transponder (tag) to a receiver for purposes of identifying items. This technology has been in development for a standard to replace the Universal Product Code (UPC) symbol with the Electronic Product Code (ePC) symbol through the Auto ID Center, formerly of MIT.

GIS-Based Accessibility Measures and Application

Fahui Wang
Northern Illinois University, USA

Wei Luo
Northern Illinois University, USA

INTRODUCTION

Accessibility refers to the relative ease by which the locations of activities, such as work, shopping and healthcare, can be reached from a given location. Access varies across space because of uneven distributions of supply and demand (*spatial factors*), and also varies among population groups because of their different socioeconomic and demographic characteristics (*nonspatial factors*). Taking healthcare access for example, spatial access emphasizes the importance of geographic barrier (distance or time) between consumer and provider, whereas nonspatial access stresses non-geographic barriers or facilitators such as social class, income, ethnicity, age, sex, and so forth. Since the 1960s, health policymakers in the United States have attempted to improve health care for the citizenry by considering aspects of both spatial and nonspatial factors. Such efforts are exemplified in designations of Health Professional Shortage Areas (HPSA) and Medically Underserved Areas or Populations (MUA/P) by the U.S. Department of Health and Human Services (DHHS), for the purpose of determining eligibility for certain federal health care resources. The DHHS is considering consolidating the HPSA and MUA/P designations into one system because of their overlapping criteria (U.S. DHHS, 1998). See guidelines at http://bphc.hrsa.gov/dsd (last accessed April 1, 2004).

BACKGROUND

Measuring Spatial Accessibility

According to Joseph and Phillips (1984), measures of spatial accessibility include *regional availability* and *regional accessibility*. The former is expressed as a demand (population) to supply (i.e., practitioner in the case of healthcare access) ratio within a region, and it is simple and easy to implement. The latter considers complex interaction between supply and demand in different regions based on a gravity kernel, and it is less intuitive and requires more computation.

The regional availability approach has two problems: interaction across regional boundaries is generally not adequately accounted for and spatial variability within a region is not revealed (Wing & Reynolds, 1988). Several methods have been developed to mitigate the problems. For example, Luo (2004) uses a *floating catchment area (FCA) method* for assessing physician accessibility. Assuming a threshold travel distance of 15 miles for primary health care, a 15-mile circle is drawn around a residential tract as its catchment area. The circle with the same radius (i.e., catchment area) "floats" from the centroid of one tract to another, and the physician-to-population ratio within each catchment defines the accessibility there. The underlying assumption is that services that fall within the circle are fully available to any residents within that catchment. However, not all physicians within the catchment are reachable within the threshold distance by every resident in the catchment, and physicians on the periphery of the catchment may also serve nearby residents outside the catchment and thus may not be fully available to residents within the catchment.

A method developed by Radke and Mu (2000) overcomes the above fallacies. It repeats the process of "floating catchment" twice (cone on physician locations and cone on population locations), and can be easily implemented in a geographic information system (GIS) (Wang & Luo, 2004).

First, for each physician location j, search all population locations (k) that are within a threshold travel time (d_0) from location j (i.e., catchment area j), and compute the physician to population ratio R_j within the catchment area:

$$R_j = \frac{S_j}{\sum_{k \in \{d_{kj} \le d_0\}} P_k}, \qquad (1)$$

where P_k is the population of tract k whose centroid falls within the catchment (i.e., $d_{kj} \le d_0$), S_j is the number of physicians at location j, and d_{kj} is the travel time between k and j.

Next, for each population location i, search all physician locations (j) that are within the threshold travel time

(d_0) from location i (i.e., catchment area i), and sum up the physician to population ratios R_j at these locations:

$$A_i^F = \sum_{j \in \{d_{ij} \le d_0\}} R_j = \sum_{j \in \{d_{ij} \le d_0\}} \left(\frac{S_j}{\sum_{k \in \{d_{kj} \le d_0\}} P_k} \right), \qquad (2)$$

where A_i^F represents the accessibility at resident location i based on this two-step FCA method, R_j is the physician to population ratio at physician location j whose centroid falls within the catchment centered at i (i.e., $d_{ij} \le d_0$), and d_{ij} is the travel time between i and j.

One may notice that it draws an artificial line (say, 30 minutes) between an accessible and inaccessible physician. Physicians within that range are counted equally regardless of the actual travel time. *A gravity model* such as the one in Joseph and Bantock (1982) can be used to weight a nearby physician higher than a remote one. The gravity-based accessibility A_i^G at location i can be written as:

$$A_i^G = \sum_{j=1}^{n} \frac{S_j d_{ij}^{-\beta}}{V_j}, \text{ where } V_j = \sum_{k=1}^{m} P_k \, d_{kj}^{-\beta}, \qquad (3)$$

where n and m indicate the total numbers of physician and population locations, respectively, and all other variables are the same as in Equation (2).

Luo and Wang (2003, p.874) have proven that the two measures A_i^F and A_i^G are equivalent. The only difference is except that travel time impedance is dichotomous in Equation (2) but continuous in Equation (3). The measure A_i^F by the two-step FCA method may be a more favorable choice for practical uses. First, it is simple and can be easily adopted by state health departments. Secondly, it is intuitive as it compares supply vs. demand and does not need to define the travel friction coefficient b in the gravity model. Defining bð is particularly troublesome since its value varies from place to place and also over time. Finally, the FCA method is particularly suitable for identifying areas with low accessibility, as the gravity-based method tends to conceal local pockets of poor accessibility (Luo & Wang, 2003, p.876).

Analyzing Nonspatial Factors

Population subgroups differ in terms of needs and accessibility according to their age, sex, social class, ethnicity, and other nonspatial characteristics. For example, Field (2000) compiled a list of factors affecting healthcare access, standardized all indicators according to a normal distribution, and then combined them to produce a final composite index of relative advantage.

Possible nonspatial factors for healthcare access include:

(1) Demographic variables (such as age and sex). For example, populations with high needs include seniors with ages above 65, children with ages 0-4 and women with ages 15-44.
(2) Socioeconomic status. Low socioeconomic status may incur important barriers to health access and lead to ill health. Variables may include population in poverty, female-headed households, home ownership and median income.
(3) Environment. Overcrowding or poor living conditions may contribute to higher levels of ill health (e.g., Field, 2000, p.315). Variables may include households with an average of more than one person per room and housing units lack of basic amenities (lacking complete plumbing or kitchen facilities).
(4) Linguistic barrier and service awareness. Minorities or lower educational attainment may be associated with lower service awareness (e.g., Field, 2000, p.317), and linguistic isolation may create an important barrier to healthcare access (e.g., U.S. DHHS, 1998). Variables may include non-white minorities, population without a high-school diploma and households linguistically isolated.
(5) Transportation mobility. People dependent solely on public transit may have less mobility and their accessibility to physicians is diminished to a great degree (e.g., Field, 2000). Variables may include households without vehicles.

As these variables are often correlated, a simple aggregation of the indicators may not be appropriate. Wang and Luo (2004) use the factor analysis to consolidate nonspatial factors, and identify three major factors (i.e., socioeconomic disadvantages, socio-cultural barriers, and high healthcare needs).

FUTURE TRENDS

The spatial accessibility measure and nonspatial factors need to be integrated together for assessing accessibility. In evaluating healthcare access, one may assign larger weights to population subgroups with high healthcare needs and directly incorporate this factor into the spatial accessibility measure such as Equation (2). The spatial accessibility measure can be used to identify the first type of physician shortage areas (i.e., *geographic areas* as in the official HPSA designation guidelines). Nonspatial factors (e.g., the "socioeconomic disadvantages" and "socio-cultural barriers" factors) can be used to identify the second type of physician shortage areas (i.e., *popu-*

lation groups as in the official HPSA designation guidelines). Figure 1 shows the result of an integrated approach to defining physician shortage areas in Illinois, U.S., based on the data in 2000.

CONCLUSION

Both spatial and nonspatial factors need to be considered in evaluating accessibility. Spatial access emphasizes the importance of geographic barrier between supply and demand, and nonspatial factors include non-geographic barriers or facilitators such as age, sex, ethnicity, income, social class, education and language ability. A GIS-supported two-step floating catchment area method has proven to be appropriate in measuring spatial accessibility. Nonspatial variables can be consolidated by factor analysis into a few independent factors. Applying the methodology in assessing healthcare access, spatial and nonspatial factors can be integrated together to identify two types of physician shortage areas (i.e., geographic areas and population groups).

REFERENCES

Field, K. (2000). Measuring the need for primary health care: An index of relative disadvantage. *Applied Geography, 20*, 305-332.

Joseph, A.E., & Bantock, P.R. (1982). Measuring potential physical accessibility to general practitioners in rural areas: A method and case study. *Social Science and Medicine, 16*, 85-90.

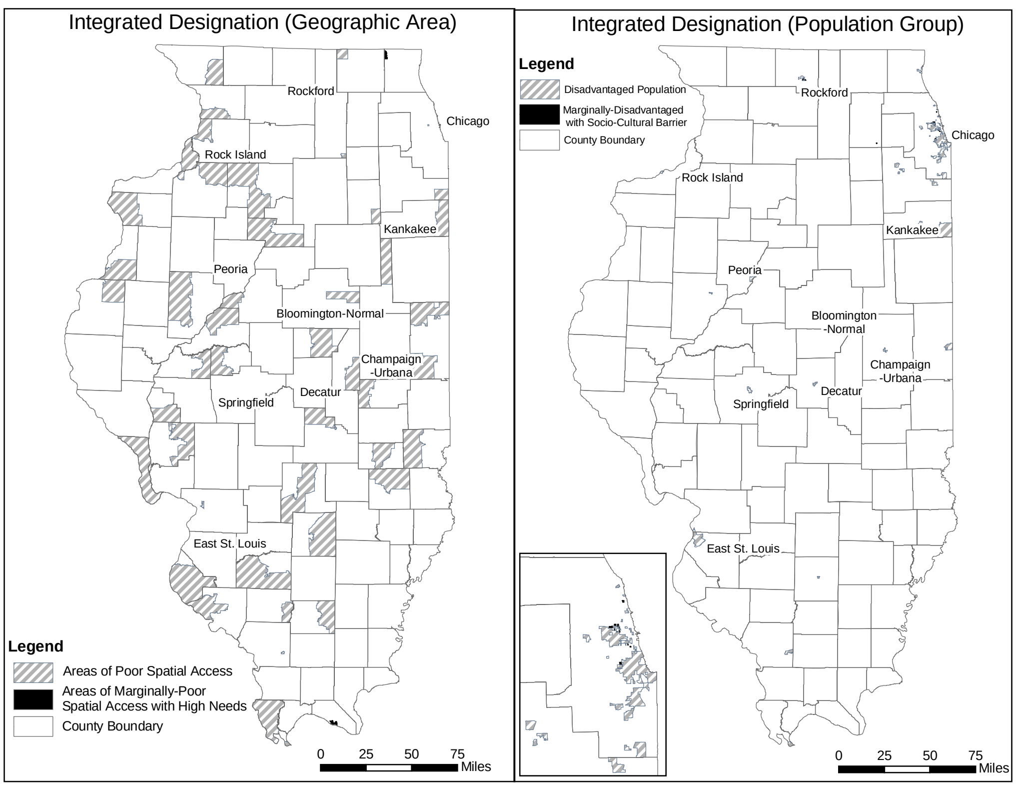
Joseph, A.E. & Phillips, D.R. (1984). *Accessibility and utilization – Geographical perspectives on health care delivery*. New York: Harper & Row Publishers.

Luo, W. (2004). Using a GIS-based floating catchment method to assess areas with shortage of physicians. *Health and Place*, *10*, 1-11.

Luo, W., & Wang, F. (2003). Measures of spatial accessibility to healthcare in a GIS environment: Synthesis and a case study in Chicago region. *Environment and Planning B: Planning and Design, 30,* 865-884.

Radke, J., & Mu, L. (2000). Spatial decomposition, modeling and mapping service regions to predict access to

Figure 1. Identified physician shortage areas in Illinois, U.S. (2000) [Inset shows an enlargement of Chicago area]



social programs. *Geographic Information Sciences*, *6*, 105-112.

U.S. Department of Health and Human Services (DHHS). (1998). Designation of medically underserved populations and health professional shortage areas: proposed rule, *Federal Register*, *63*, 46537-46555.

Wang, F., & Luo, W. (2004). Assessing spatial and nonspatial factors in healthcare access in Illinois: Towards an integrated approach to defining health professional shortage areas. Forthcoming in *Health and Place*.

Wing, P., & Reynolds, C. (1988). The availability of physician services: a geographic analysis. *Health Services Research*, *23*, 649-667.

KEY TERMS

Accessibility: The relative ease by which the locations of activities, such as work, school, shopping and healthcare, can be reached from a given location.

Floating Catchment Area Method: A geographic information systems (GIS) supported method for assessing the scarcity of supply versus demand. For example, assuming a threshold travel distance of 15 miles for primary health care, a 15-mile circle is drawn around a residential location as its catchment area. The circle "floats" from one location to another throughout a study area, and the physician-to-population ratio within each catchment indicates whether an area is medically underserved.

Gravity Models: Models that measure the interaction between social/economic objects, similar to the gravity model in physics. In the models, the intensity of interaction (e.g., trips, migration or communication) is positively related to the sizes of objects (e.g., population) but inversely related to the distance or travel time between them.

Health Professional Shortage Area (HPSA): Geographic areas or population groups designated by the U.S. Department of Health and Human Services (DHHS) that are underserved in healthcare.

Nonspatial Accessibility: Non-geographic barriers or facilitators such as social class, income, ethnicity, age, sex, and so forth, that affect one's accessibility to a service.

Spatial Accessibility: Consumers' ability to overcome geographic barriers (distance or time) in order to gain access to service providers.

Global Implications of E-Commerce Tool and Artefact Creation

Deborah Bunker
SISTM, University of New South Wales, Australia

INTRODUCTION

It is argued that Electronic Commerce (EC) platforms can be seen as artefacts—tools that are made, used, inherited and studied within a cultural context. This context encompasses economic, historical, technical, and social values and assumptions that are focused on particular ideas and definitions (relating to the example of B2C activities and processes contained within Electronic Grocery Systems (EGS)). The issue we should face as makers, users, inheritors, and scholars of these tools, however, is that the tool context and inherent in-built values on which this context is based, particularly relating to matters of effective use of EC tools in a B2C marketplace, may not be in evidence across all cultures. This would make the successful use of EC, in a global sense, a difficult and complex undertaking.

BACKGROUND

Understanding the process and role of tool creation and use in relation to the information technology and systems (ITS) discipline/paradigm (and EC in particular) is fundamental to understanding the cultural bias inherent in the process. The definitions used in this chapter are those generally used by the research community in the area of technology transfer (TT) (Robinson, 1988). These definitions are loaded with such terms as "artefacts," "technology," and "tools." If we accept that ITS is a tool-focused discipline, then we must look at the context of the creation of those tools in order to better understand how they can be used more appropriately and effectively in different cultural contexts.

Bunker and Dean (1996) discuss Kuhn's (1970) theory that suggests that the underlying structure of a discipline arises from a set of assumptions generally accepted by practitioners, teachers, and disciplinary constituents. Techniques and tools defined by the discipline are created from a common understanding. They may be part of an evolutionary creation process or equally be created through a "paradigm shift"; however, they do become indicative of the generally held underlying assumptions of the discipline. Schein (1984) suggests a three-stage ascent from basic assumptions to the artefacts and creations that drive the evolution of paradigms. If we look at the tools in current use by a discipline, generally accepted underlying assumptions may be deduced.

The model of a discipline is in four parts (see Figure 1). By extending the idea of tools as the visible sign of paradigmatic assumptions that underpin a discipline, we find that four entities play a role: the tool Maker, the tool User, the Scholar, and the Inheritor of the discipline. The tools that reflect the ITS discipline are created and used in a multitude of contexts. What are the implications for the discipline and the diffusion and use of these tools?

ITS Tools in Context

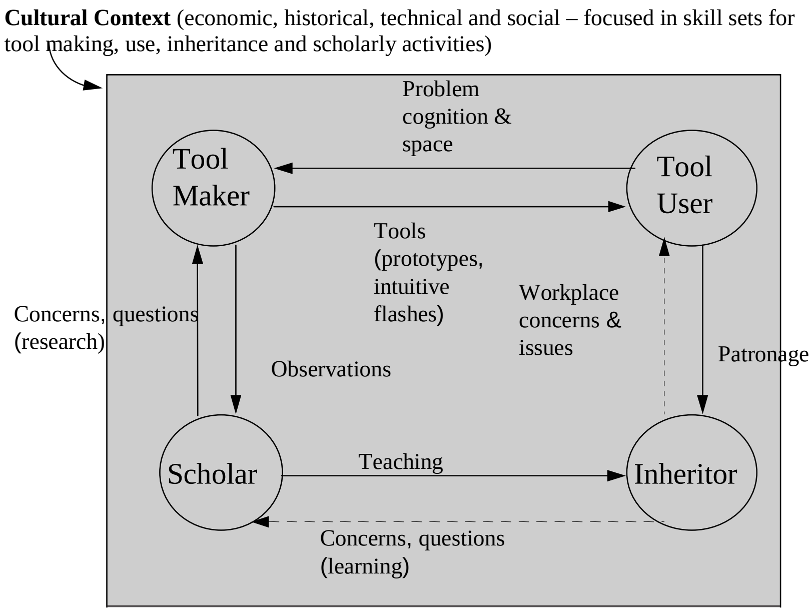
Young (1971) in his *An Introduction to the Study of Man* defines tools in the following ways: tool making assumes skills and ways of life that are transmitted by interpersonal communication and tradition rather than genetics; tool making involves foresight as to the use of the tool; tools are made by a technique that is learned from others and involves symbolic communication (language); tools are made according to an evolutionary convention (gradual) and are made and used in a cultural context. For example, the ITS literature is full of stories about ITS failure (Sauer, 1993). Are these really stories about the lack of contextual acceptance of these tools within organisations or target cultures?

Techniques and tools are created from a common understanding or cultural context. They become an expression of generally held underlying assumptions (Schein, 1984). The cultural context underpinning tool creation and use is created and evolves over time. The assumptions, which in turn underpin cultural context, are deep-seated, taken-for-granted, invisible, and preconscious, and are patterned into what Schein (1984) terms cultural "paradigms."

Cultural paradigms can be expressed in a number of ways. Research conducted by Kluckhorn and Strodtbeck (1961) and Hofstede (1980) are typical examples of the different assumptions upon which various cultures are based. Schein (1984) expresses the assumptions of cultural paradigms as the: organisation's relationship to its environment; the nature of reality and truth; the nature of human nature; the nature of human activity; and the nature of human relationships.

If we accept that there are vast differences in cultural paradigms around the world, and that ITS tool creation

Figure 1. Disciplinary model (Bunker & Dean, 1996)



and use is underpinned by these cultural paradigms, what then, are the implications for the development and use of an ITS tool in one culture for diffusion and use in another? In this fast-paced, highly volatile and mobile global society, many organisations find themselves in a situation where the transfer, diffusion, and use of ITS is pushed at an extraordinary rate. Cardwell (1994) argues the importance of ITS as being a strategic technology of which the applications and scope exceeds all others. Richard Baskerville (1995) explains that the use and management of ITS are an expression of an organisation's structure and culture, and that as we computerise the symbolic universe of the organisation, we can view the system and its information as artefacts.

This has far reaching implications for organisations, nations, and the global economy. Robinson (1988) has likened the process of developing technology in one society and implanting it in another without adaptation as an organ transplant from one body to another without any testing for compatibility.

Electronic Commerce (EC) Tools in Context: Electronic Grocery Systems (EGS)

The Internet has played a key role in the establishment of EC activities (especially in B2C relationships), and EGS systems have been dependent on the development of Internet technologies and their uptake speed and scale by potential consumers. Consumers in the West have begun to demand increases in the quality of service being offered by their grocery provider. Shoppers want a product selection adapted to their needs and the ability to shop at their convenience at any time (Choudrie et al., 1998).

The introduction of the Internet and the development of EGS systems have allowed consumers to shop at their convenience rather than being bound to retail opening trading times. The subsequent growth of the Internet and the increasing demand for value-added grocery shopping has created demand-pull for EGS services. Consumer demand for EGS services, combined with supplier demands for shopper profile information, has created a large uptake of EGS systems from a business and consumer perspective (Deutsch et al., 2000).

As discussed earlier, however, we have also seen the need for contextual value or cultural compatibility between the creator and the recipient of a technology in order for the technology to be transferred and utilised successfully. A certain level of cultural homogeneity is necessary in order for the tools and artefacts of one culture to be successfully utilised within another (Bunker, 1998). Electronic grocery shopping (EGS) systems utilise the tools and artefacts of ITS that have been created from a predominantly Western and particularly North American cultural viewpoint (Simon & Doran, 1998). This viewpoint assumes that the shopping culture of the consumer accepts multiple distribution channels for consumer items, but it neglects resources and skills required by the consumer to best utilise these multiple channels, as well as any essential social activity associated with the shopping task and associated processes. These types of systems may be acceptable in a culture that has had a history and experience (and thus development of suitable associated technology and skills) of multiple purchase and distribution channels in B2C transactions such as the retail, catalogue, and home shopping B2C channels of North America. We can see that EGS systems, in their current form, generally assume:

- Computer ownership or access
- Internet access (Internet service providers, as well as reliable and secure telecommunications infrastructure)
- Basic computing skills
- Product and outlet brand awareness (knowing exactly what you want and who you want it from)
- Product quality awareness (where it is manufactured and which company manufactures it)
- Basic purchasing strategies (little opportunity for special savings)
- Particular lifestyle benefits to consumers in terms of time savings, convenience, and prestige of using EGS systems
- Acceptance of additional distribution costs to the purchase decision
- Population density for overall cost effectiveness of storage and distribution for the e-business
- A limited range of goods
- A "type" of customer behaviour (pattern buying, rational purchase decision, and utility)
- A trust in the security and reliability of the system

(Brandtweiner, 1998; Heikkila et al., 1998; Pereira, 1998, Strader & Hendrickson, 1998)

A certain lifestyle, skill set, and access to resources is assumed, and the more social element of the shopping experience is missing. Diverse cultural dimensions are not expressed within the definition of the tool.

FUTURE TRENDS

In viewing EGS systems as tools or artefacts that have been created with in-built values and assumptions, we find that the cultural context of the user of an EGS system is important in understanding how these systems might be best implemented and utilised (Simon & Doran, 1998). Johnson (1997) discusses four types of value meanings in technology: moral/metaphysical (past intended use of the technology); support (present intended use of the technology); material (in-built characteristics which influence the use of the technology); and expressive values (motivation for use of the technology). All of these contribute to the assumed skill sets required to utilise the technology. It is these expressive values that encapsulate the more diverse elements of the shopping experience within many non-Western cultural contexts. Value meanings link back to the definitions of tools as outlined by Young (1971). Tool making is not only technical in nature, but is bound by cultural values and an understanding of how the tool has been created for use in an acceptable manner. This understanding reflects a certain level and mixture of skills. Ayres (1978), in discussing his ideas on the theory of economic progress of civilisation, states that "…absolute mutual contingency of skills and tools is of supreme importance for an understanding of technology as a function of human behaviour" (Ayres, 1978, p. 108).

Skills and technology are bound together as technology is created and used as a result of an assumed skill set that is heavily influenced by core assumptions and values (culture). It is evident that EGS systems (in their current state) assume limited technical and purchase decision skills within a culture and the B2C experience, while totally neglecting social skills and their important skills.

EGS systems are engendered with the cultural attributes of Western society (Pargman, 1998). For non-Western cultures that may have different resource availability and different social marketplaces, these systems may not be as effective as more traditional means of transacting business. Many non-Western cultures, for instance, are considered developing economies, a value-laden term, to be sure, which is generally used to describe a local village-based economic structure. These economies are reliant on intricate social interactions and customs to ensure effective B2C trading relationships and the maintenance of a robust local economy. If you have ever visited a local marketplace in China, Egypt, or suburban Australia, you are able to see these relationships in action and their effect in a B2C transaction. The utility and logistical benefits associated with EGS systems may be of limited benefit to the types of economic structures where personal interaction is important, as they are a product of a different cultural context and, therefore, may be better suited to a Western mass market style of B2C transaction.

There are distinctly different philosophical positions in particular extremes of culture. The skills required to design, make, understand, and use tools, as well as the languages we use to define and represent them, would be significantly different from one type of culture to another. An EGS system that lacks consideration and design for a diverse resource base and social attributes of a shopping experience may not necessarily fit within the shopping experience of other cultures and contexts.

For instance, in many cultures in a B2C transaction, the process of bargaining over a particular price for a product is an important ceremonial and symbolic act that reflects on a vendor and a customer's perception of intent within the B2C transaction process. This affects the understanding by the customer of the product quality, integrity of the vendor, product range, and so forth, and also affects the vendor's understanding of his or her potential customers, the marketing strategies, and the range of services or products to offer. In this context, the nature of information gathering and the transaction process takes on a significantly richer meaning, which may not be facilitated or captured by an EGS system.

CONCLUSION

If Global Technology Transfer (GTT) is to be fully understood, then further research is necessary. If we better understand the cultural context in which ITS is created, as well as the cultural context of the recipient, we may have to admit that some ITS may not be appropriate for transfer under certain circumstances, it may have to be radically redeveloped to accommodate differences in context, or the culture of the recipient may have to change in order to accommodate the technology. We may need to consider the development of information technology and systems as artefacts that should be designed to be adaptive as a matter of priority. Within this design, there is a real need to consider the stakeholders' expectations and the resultant "evaluative dispositions" that are derived from their values (Lyytinen & Hirschheim, 1987 in Lycett et al., 1997).

These values are cultural paradigms and, as such, are not generally articulated (Claver et al., 2001). Lycett et al. (1997) discuss the stakeholder's "intentionality" or the representation of values in structures, practices, and conventions of an organisation. Schein (1984) articulates much the same argument when he speaks of the artefacts and creations of an organisation such as documents and charters. ITS are also artefacts or tools which are a direct expression of the cultural paradigm of their creators.

In researching within our discipline, this question needs to be considered: Can an ITS only be reflective of the basic assumptions and values (cultural paradigms) of its creators and immediate user population? Is the temporal "snapshot" that Lycett et al. (1997) discuss that provides "the picture of reality" of the system all that we can ever hope to have, or can we assume that we can (and should) develop ITS tools (physical tools) and techniques and methods (conceptual tools) that can be adapted and changed to suit contextual differences?

The new age of globalisation in which information is shared and ITS has had a significant impact in the world marketplace, would appear to be only taking place where there is a similar cultural, historical, and economic context of development (Bunker & MacGregor, 2002). It is the development context that facilitates the building of these ITS tools, as well as their transfer and use within other contexts. The context of ITS tool creation, use, inheritance, and knowledge may, therefore, be a limiting factor in its potential usefulness and applicability on a global scale.

REFERENCES

Ayres, C.E. (1978). *The theory of economic progress*. Kalamazoo, MI: New Issues Press.

Baskerville, R. (1995, October 14-17). Structural entropy model of technology transfer. In K. Kautz, J. Pries-Heje, T. Larsen, & P. Sorgaard (Eds.). *Proceedings of the First IFIP TC8 Working Conference on Diffusion and Adoption of Information Technology*, Oslo, Norway.

Brantweiner. R. (1998, August 14-16). Risks and challenges for retailers: The value chain transformation – A European perspective. *Proceedings of the 4th Association of Information Systems Conference*, Baltimore, Maryland.

Bronowski, J. (1977). *A sense of the future*. Cambridge, MA: MIT Press.

Bunker, D.J. (1998). A philosophy of information technology and systems (IT & S) as tools: Tool development context, associated skills and the global technical transfer (GTT) process. *Proceedings of IFIP WG 8.7 Working Conference*, Helsinki, Finland.

Bunker, D.J., & Dean, R.G. (1996, August 16-18). Towards an information systems paradigmatic framework. *Proceedings of the Americas Conference on Information Systems*, Phoenix, Arizona.

Bunker, D.J., & MacGregor, R.C. (2002, August 9-11). The context of information technology (IT) and electronic commerce (EC) adoption in small/medium enterprises (SMEs) – A global perspective. *Proceedings of the Americas Conference on Information Systems*, Dallas Texas.

Cardwell, D. (1994). *The fontana history of technology*. London: Fontana Press.

Choudrie, J., Hlupic, V., & O'Keefe, B. (1998, June 8-10). A framework for using information technology within the grocery retail sector. *Proceedings of the 11th Bled Electronic Commerce Conference*, Bled, Slovenia.

Claver, E., Llopis, J., Reyes-Gonzales, M., & Gasco, J.L. (2001). The performance of information systems through organisational culture. *Information, Technology & People*, *14*(3), 247-260.

Davis, W.S., & Benamati, J. (2003). *E-commerce basics: Technology foundations & e-business applications*. USA: Addison Wesley.

Deutsch, J., Bunker, D.J., & MacGregor, R. (2000, December 13) Electronic grocery shopping (EGS) – Ordinary success and spectacular failure: A tale of two models. *Proceedings of the Third Electronic Grocery Shopping (EGS) Workshop*, Brisbane, Australia.

Heikkila, J., Kallio, J., Saarinen, T., & Tuunainen, V.K. (1998, November). Analysis of expectations on electronic grocery shopping for potential customer segments. *AJIS Special Edition - Electronic Commerce*, 56-69.

Hofstede, G. (1980). *Culture's consequences: International differences in work related values*. Beverly Hills, CA: Sage Publications.

Johnson, D.G. (1997, September). Is the global information technology infrastructure a democratic technology? *Computers and Society*, 20-26.

Kluckhohn, F.R., & Strodtbeck, F.L. (1961). *Variations in value orientations*. Evaston, IL: Row, Peterson.

Kuhn, T.S. (1970). *The structure of scientific revolution* (2nd ed.). Chicago, IL: University of Chicago Press.

Lien, L. (1994, October). Transferring technologies from developed to developing industrial and commercial environments. *Proceedings of the IFIP TC8 Working Conference on Diffusion, Transfer and Implementation of Information Technology*, Pittsburgh, PA.

Lyytinen, K. & Hirschheim, R. (1987). Information systems failures: A survey and classification of the empirical literature. *Oxford Surveys in Information Technology*, *4*, 257-309.

Lycett, M., Kanellis, P., & Paul R.J. (1997, August 23-27). Philosophical directions for information systems development. *Proceedings of the Americas Conference on Information Systems*, Minneapolis, MN.

Oakley, K.P. (1957). Tools makyth man. *Antiquity*, *31*, 199-209.

Pargman, D. (1998). Reflections on cultural bias and adaptation. *Proceedings of Cultural Attitudes Towards Communication and Technology '98*, Sydney, Australia.

Periera, R.E. (1998, August 14-16). Factors influencing consumer purchasing behavior in electronic commerce. *Proceedings of the 4th Association of Information Systems Conference*, Baltimore, MD.

Reisman, A. & Zhao, L. (1991, Spring). A taxonomy of technology transfer transaction types. *Journal of Technology Transfer*, *16* (2), 38-42.

Robinson, R (1988) *The International Transfer of Technology: Issues, Theory and Practice*. Ballinger Publishing Company (Harper and Row, Publishers Inc. Cambridge).

Sauer, C. (1993). *Why information systems fail: A case study approach*. Alfred Waller, Henley-On-Thames. London: McGraw-Hill International.

Schein, E.H. (1984, Winter). Coming to a new awareness of organisational culture. *Sloan Management Review*, 3-16.

Simon, S.J., & Doran, P.M. (1998, August 14-16). Global Web page design: Issues of culture and gender. *Proceedings of the 4th Association of Information Systems Conference*, Baltimore, MD.

Strader, T.J., & Hendrickson, A.R. (1998, August 14-16). A framework for the analysis of electronic market success. *Proceedings of the 4th Association of Information Systems Conference*, Baltimore, MD.

Young, J.Z. (1971). *An introduction to the study of man*. Oxford: Clarendon Press.

KEY TERMS

Context: "The client as the receiver of the technology, must create the context (technology, transfer, context, talent, modification, management, resource and contribution) and verify its capacity to receive and apply the technology and information in the new environment" (Lien, 1994).

Electronic Commerce: "(1) Transactions conducted over the Internet, either by consumers purchasing goods and services or directly between businesses. (2) Conducting business online" (Davis & Benamati, 2003).

Electronic Grocery Shopping: "By extension of this concept (Electronic Commerce) Electronic Grocery Shopping (EGS) involves the specific purchase of grocery items which is facilitated via computer networks" (Deutsch et al., 2001).

Technology: "Technology is broadly defined as the documented or licensed technology, the embodying artefacts (equipment, software tooling), and the wherewithal to know how to successfully deploy the technology" (Robinson, 1988).

Technology Transfer: "The conveyance or shift of tools, techniques, procedures and/or the legal titles thereto used to accomplish some human purpose" (Reisman & Zhao, 1991).

Globalization of Consumer E–Commerce

Daniel Brandon, Jr.
Christian Brothers University, USA

INTRODUCTION

This article reviews globalization aspects of "business to consumer" (B2C) electronic commerce. According to *Computerworld*, "Globalization is the marketing and selling of a product outside a company's home country. To successfully do that on the Internet, a company needs to *localize* – make its Web site linguistically, culturally, and in all other ways accessible to customers outside its home territory" (Brandon, 2001). This overview describes the key issues in the globalization of electronic commerce; for more detail, see the full book chapter (Brandon, 2002).

BACKGROUND

"Ever since the end of the Cold War, the world has been rushing toward ever-higher levels of national convergence, with capital markets, business regulation, trade policies, and the like becoming similar" (Moschella, 2000). The value of cross-border mergers grew six-fold from 1991 to 1998 from U.S. $85 billion to $558 billion. The world has not witnessed such a dramatic change in business since the Industrial Revolution (Korper & Ellis, 2000). More than 95% of the world population lives outside of the U.S., and for most countries, the majority of their potential market for goods and services is outside of their borders. Over 60% of the world's *online* population resides outside of the United States (IW, 2000).

Today, the majority of Fortune's 100's Web sites are available only in English (Betts, 2000). In our rush to get on the WWW, we sometimes forget that WW is for "World Wide" (Giebel, 1999). Today's average Web site gets 30% of its traffic from foreign visitors, yet only 1% of small and mid-size American businesses export overseas (Grossman, 2000b).

KEY ISSUES

"Localization" (shortened to L12N in Internet terms) considers five global dimensions: geographic, functional, regulatory, cultural, and economic (Bean, 2000). We shall overview each of these somewhat overlapping and interrelated issues in these groupings: language, cultural, legal, payment/currency, dates/units, and logistics.

Language

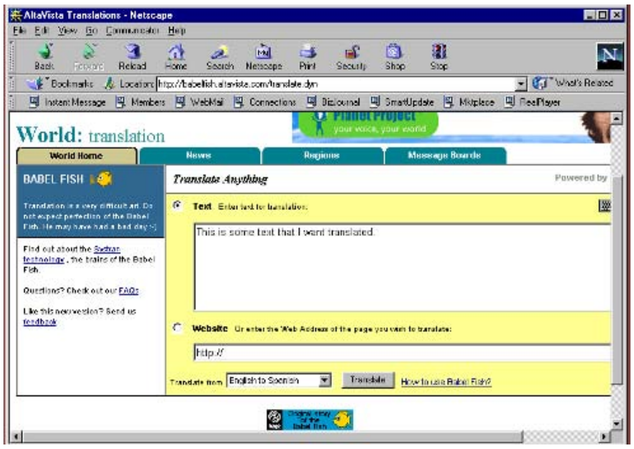
According to IDC, by 2005, more than 70% of the one billion Web users around the world will be non-English speakers (Wonnacott, 2001). For the immediate future, most of the Internet community will still understand English, but overall English is the native language to only 8% of the world. Most users in foreign countries prefer content in their own language; for example, 75% of users in China and Korea have such a preference (Ferranti, 1999). It was found that visitors spend twice as long, and are three times more likely to buy from a site presented in their native language (Schwartz, 2000). We also have to take into account differing dialects that are used across various countries speaking a specific language. The combination of language and dialect is called a "locale".

One can convert Web pages by hiring a translator or using a computer-based translation product or service. Hiring a translator will provide the best localization but is more costly than the automatic methods. Translators can easily be found in the Aquarius directory (http://aquarius.net) or Glen's Guide (www.gleensguide.com). It is best to use a translator that "lives" in the local region; if a translator has not lived in a region for a decade, he has missed 10 years of the local culture. There are also many companies that provide translation services such as: Aradco, VSI, eTranslate, Idiom, iLanguage, WorldPoint, and others. The cost of these services is about 25 cents per word per language (Brandon, 2002). Automatic translation software is another option, but it is still in its infancy (Reed, 2000). Some popular software products for translation are: www.e-ling.com, www.lhs.com, and www.systransoft.com. The automatically-translated text typically does not convey the meaning of the original text.

There are several Web sites which provide free translation services such as: http://babelfish.altavista.com, *http://translator.go.com*, and www.freetranslation.com. For example, Figure 1 shows the "BabelFish" Web site where we are requesting a translation of an English sentence into Spanish. Figure 2 shows the translation results.

Another alternative, although certainly not optimal, is to provide a link on your English Web page for these free services so that visitors can translate your content themselves. Figure 3 shows a portion of the CBU School of Business English version Web site.

Figure 1



The automatic Spanish translated version (using BabelFish) is shown in Figure 4. Note that automatic version, while syntactically and grammatically correct, does not convey the exact intended meaning to most of the titles and phrases.

Figure 5 is the version converted by a translator manually, and even though you may not speak Spanish, you can see the extent of the differences (Brandon, 2000). Shown in Figure 6 is the home page for FedEx (www.fedex.com). One can select from over 200 countries for specific language and content.

Cultural

Creating an effective foreign Web site involves much more than just a good language translation. Not only do languages differ in other countries but semantics (the meaning of words and phrases) and cultural persuasions in a number of key areas are different. "Sensitivity to culture and national distinction will separate success from failure" (Sawhney & Mandai, 2000). To be effective, a Web site has not only to be understandable and efficient, but has to be culturally pleasing and inoffensive. To accomplish that, it may be necessary that not only language be localized, but that content, layout, navigation, color, graphics, text/symbol size, and style may be different. Many companies have put forth global Web sites simply by translating the English into the targeted language, but then had to pull back and redesign the localized site due to cultural offenses.

A country's humor, symbols, idioms, and marketing concepts may not send the same messages to other countries in the world. Some areas of global disagreement to avoid are: equality of the sexes or races, body parts and sexuality, abortion, child labor and majority age, animal rights, nudity, guns, work hours and ethic, capital punishment, scientific theories, and religious particulars (Brandon, 2002).

Colors have symbolic and special meaning in most locals. Purple is a problem in many places; it symbolizes death in catholic Europe and prostitution in the Middle East. Euro Disney had to rework its European sites after the first version used too much purple. Overall blue is the most culturally accepted color (Brandon, 2001). It is also very important to respect other cultures "symbols" (heroes, icons, etc.) both positive and negative (swastika). One guide site is *Merriam Webster's Guide to International Business* (*www.bspage.com/address.html*).

Legal

Recently French court's ruling that Yahoo must make auctions of Nazi memorabilia unavailable in France indicates how uncertain and risky international e-business can be. "The troubling aspect of this case is that different countries can say that content not even targeted at their population breaks the law" (Perrotta, 2000). With the Internet, it is not possible to know for sure where a user is logged in due to "IP tunneling" possibilities.

"Freedom" laws (such as the U.S. First Amendment) are not universal, and saying/printing some things can be illegal in some parts of the world. In the U.S., you can say what you like about "public figures" but not so in most of the rest of the world. Another legal issue concerns the privacy of personal data collected online. Many parts of the world have stricter laws than does the U.S., and U.S. companies have had judgments rendered against them in foreign courts. There are other areas that could cause legal problems, too. One is foreign advertising restrictions; for example, in Germany, one cannot directly compare your product with that of a competitor. In some other countries this comparison may not be illegal but may leave a bad taste.

Figure 2

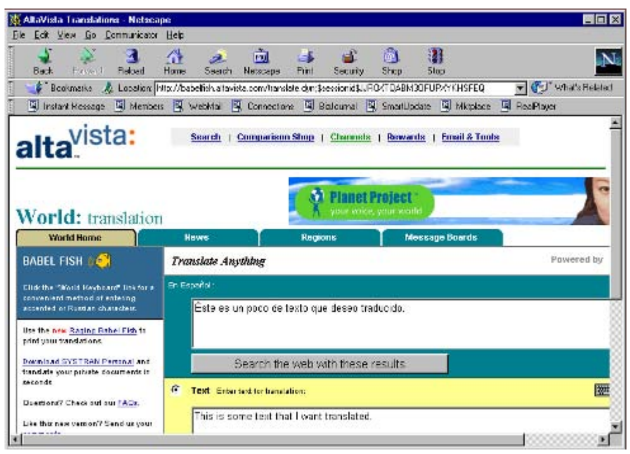
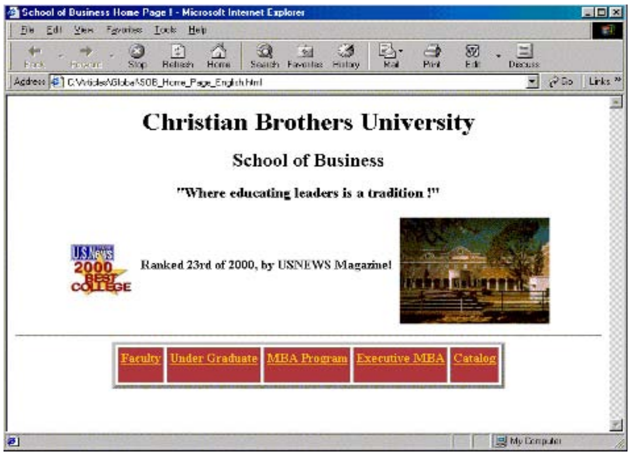


Figure 3



Payment and Currency

Nearly half of the U.S. Web sites refuse international orders because they are unable to process them (Grossman, 2000a). Foreign exchange rates vary daily so indicating that your prices are in your country's funds (exclusive of local taxes and custom duties) and using credit cards (so the credit card company does the conversion) is one way to deal with that issue. One can also link to a converter site (www.xe.net/ucc, www.oanda.com) or place a calculator on your page (www.xe.net/currency) as a utility for your customers or do your own conversions (see Figure 8, later).

However, credit cards are rare in Japan as is the use of checks. There, postal workers collect cash on delivery (CODs), and some companies send goods to brick and mortar places for consumers to pick up. In Germany, only 5% of Web users (second to U.S. in overall net usage) use credit cards. Eighty-eight percent of European merchants use invoice billing (with a long net payment due time). So while credit cards are a convenient and popular mechanism in the U.S., it is not so in the rest of the world. To complicate matters even further, there are many (and always changing) international sales taxes, value added taxes (VAT) in Europe, with different exempt items in each country. One approach to avoid all these problems is to use an escrow service such as Paymentech (www.paymentech.com) which now handles about three billion transactions a year.

Time-Date and Units of Measure

Dates are very important in e-commerce when being used for events such as: delivery dates, credit card expiration dates, product expire dates, etc. There is an international standard on dates (ISO 8601 Date Format), and even though you may not use it internally in your programs (for database operations and calculations), your Web display should be in the localized format. For example, the common U.S. format of 10/6/2000 is not uniformly understood; instead use Oct-6-2000. Major databases (i.e., Oracle) allow you to switch date formats per session or connection so the way a date is input (inserted into a table) or output (selected from a table) is automatically converted to the internal table representation of the date.

In the U.S., a 12-hour clock is common, except in U.S. military establishments. The rest of the world uses mostly a 24-hour clock, so it is best to display time in the 24-hour format. Of course, time zones will be different, so include your time zone along with the phone numbers for personal customer support. It is best to spell out the time zone in the native language. You could instead give your support time in GMT (Greenwich 2000 Standard) and use or link to www.timeanddate.com for a customizable world clock and calendar. In addition to dates and times, other units of measure will be different also. Only the U.S. and Canada still use the "English System"; the rest of the world is on the metric system now, even Britain.

"Addressing" a customer may be more involved; some foreign addresses may have longer and more address fields. There is a universal standard of sorts here called the "UPU" (universal address formats). Generally, it is of good advice including a country code (for validation of remaining fields), at least three address lines (40 characters each), city field (30 characters), a "state/province/region" field (20 characters), a postal code/zip field (10 characters), and a contact phone number (20 characters). Figure 7 shows an order form using these specifications.

Logistics

Logistics involve both getting your products to the customer, as well as allowing the customer to return un-

Figure 4

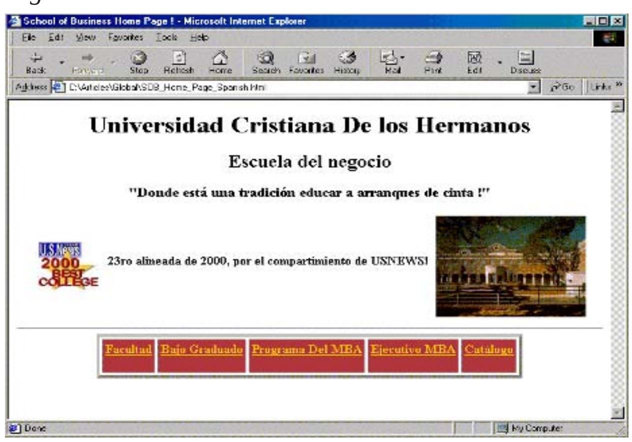
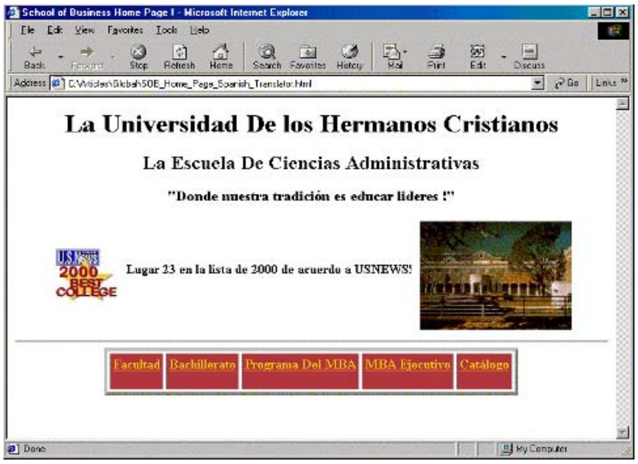


Figure 5



wanted goods. Some parts of the world have relatively primitive transportation networks. In China, villages do not have postal service. Also, each locale typically has a set of customs and tariffs that you may need to add to the price of your goods. This "landed cost" of an order is the sum of the price of goods, shipping charges, insurance, duties/customs, VAT, and any import or export fees. You may need a "Shippers Export Declaration" depending on value and mode of transportation (www.census.gov/foreign-trade/www/correct.way.html) or other documents depending on countries and goods. As well as normal shipping insurances, you may need to consider export insurance (www.exim.gov). Of course, the language as well as logistic terminology varies; however, there is a standard set of international logistic acronyms ("incoterms" - *www.schenkerusa.com/incoterms.html*).

Many countries have *foreign import restrictions* and/or quotas. In addition, many countries have certain *export restrictions*. Japan has more than 200 trade laws and 17,000 regulations on imports (Pfenning, 2001). Today, 85% of U.S. companies do not ship to customers seeking delivery abroad, and the 15% that do ship ignore these compliance issues and push the responsibility of customs, restrictions, and payment onto their customers (Shen, 2000).

There are several ways to handle all these logistics issues. One is to use shipping companies that handle all these problems for you (at a nominal charge) such as FedEx (www.fedex.com) or UPS (www.ups.com). Another alternative is to use software or services that handle all these payment, custom, and restrictions issues by preparing the paperwork and calculating "landed costs"; one example can be found at www.mycustoms.com.

Still another alternative is to use a centralized distribution center in foreign regions to reduce shipping costs and eliminate some import taxes and tariffs (Tapper, 2000), either directly or with a partner. There are also total fulfillment providers such as: National Fulfillment Services, DupliSoft, Fill It, SubmitOrder, Equire, FedexLogistics, and so forth. These organizations not only handle delivery but also inventory, returns, customer service, and in some cases, Web ordering and payment.

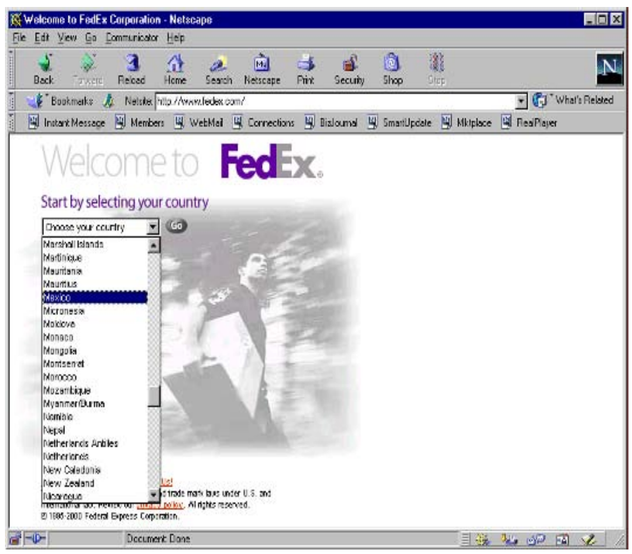
FUTURE TRENDS

In the not too distant future, the Web will be everywhere, and by "everywhere", we mean not only in all of our electronic devices but everywhere in the world. Future trends in B2C e-commerce will involve this convergence of Web connectivity both in geographics, demographics, and electronics. For example in the future, automobiles are likely to have satellite and wi-fi Internet connections not just for navigation and information, but for marketing and sales also (in several languages). As one drives down a street, they will be able to receive live commercial messages from businesses as they physically approach the retail outlets, and these retail outlets could be outlets for international products as well.

CONCLUSION

Is globalization right for an organization? It can be very costly to build and maintain a foreign presence. We have identified and briefly discussed the key issues in this article. But there are many other issues that may affect your global e-commerce. "Building a global e-business calls for hosts of strategies that include partnering with or acquiring foreign companies, assembling sales and

Figure 6



support operations, understanding new laws, languages, cultures, and implementing technology that can sustain a global endeavor" (Bacheldor, 2000). Many organizations are successful by using foreign partners such as: E-Steel, GlobalFoodExchange, and Office Depot. It has been said that the "Net brutally punishes latecomers" (Sawhney & Mandai, 2000), so it is essential to start planning the internationalization and localization of e-commerce now. Also remember the Web is a two-way street; foreign corporations will be coming after your customers soon!

REFERENCES

Bacheldor, B. (2000, May). Worldwide e-commerce: It's more than a Web site. *Information Week*.

Bean, J. (2000, March). A framework for globalization. *Enterprise Development*.

Betts, M. (2000, August). Global Web sites prove challenging. *Computerworld*.

Brandon, D. (2001). Localization of Web content. *15th Southeastern Small College, Computing Conference, 17*(1), Nashville, TN, November.

Brandon, D. (2002). Issues in the globalization of electronic commerce, In *Architectural issues of Web-enabled electronic business*. Hershey, PA: Idea Group Publishing.

Ferranti, M. (1999, October). From global to local. *Infoworld*.

Ferranti, M. (2000, November). Globalization tidal wave. *Infoworld*.

Giebel, T. (1999, November). Globalize your Web site. *PC Magazine*.

Grossman, W. (2000a, July). The outsiders. *Smart Business*.

Grossman, W. (2000b, October). Go global. *Smart Business*.

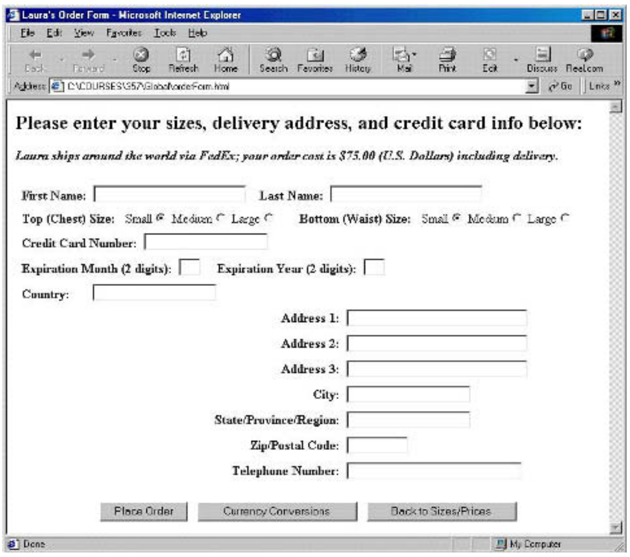
IW (staff). (2000, November 20). Weekly stats. *InternetWeek*.

Kiplinger, K. (2000, November). Globalization – alive & well. *Fidelity Outlook*.

Klee, K. (2001, March). Going global: Out ten tests can help you get started. *Forbes Small Business*.

Korper, S., & Ellis, J. (2000). *The e-commerce book, building the e-empire*. Academic Press

Figure 7



Moschella, D. (2000, December). Ten key IT challenges for the next 20 years. *Computerworld*.

Perrotta, T. (2000, July). Yahoo ruling exposes risks of being global. *InternetWorld*.

Pfenning, A. (2001, March 19). E-biz must chart international path. *InternetWeek*.

Reed, S. (2000, August). Want to limit the audience for your Web site? Keep it English only. *Infoworld*.

Sawhney, M., & Mandai, S. (2000, May). Go global. *Business*.

Schwartz, H. (2000, September). Going global. *WebTechniques*.

Shen, J. (2000, November). The commerce diplomats. *WebTechniques*.

Tapper, S. (2000, September). Is globalization right for you. *WebTechniques*.

Wonnacott, L. (2001, April). Going global may bring new opportunities for existing customers. *InfoWorld*.

KEY TERMS

Character Set: The set of symbols used to represent a language (alphabet, numerals, special symbols).

Encoding: The bit pattern to use for each symbol in a character set.

Export Restrictions: Restrictions on the type, quantity, or destination of goods that can be exported out of a country.

Globalization: The marketing and selling of a product outside a company's home country.

Import Restrictions: Restrictions on the type, quantity, or origin of goods that can be imported into a country.

Incoterms: A standard set of international logistic acronyms.

Landed Cost: The cost of an order including the price of goods, shipping charges, insurance, duties/customs, value added tax (VAT), and any import or export fees.

Locale: The combination of language and dialect.

Localize: Make a Web site linguistically, culturally, and in all other ways accessible to customers outside ones home territory.

Shippers Export Declaration: Documentation necessary to export goods outside one's home country.

UPU: Universal address formats.

VAT: Value added tax.

G

Governance in IT Outsourcing Partnerships

Erik Beulen
Atos Origin / Tilburg University, The Netherlands

INTRODUCTION

The literature first devoted attention to IT outsourcing partnerships in 1990 (Gantz, 1990; Rochester & Douglas, 1990). An IT outsourcing partnership consists of a service recipient and one or more external service providers and the relationship between them. The service recipient hands over the responsibility for the execution of the IT services to the service provider but remains the responsibility for managing the IT outsourcing partnership. The relationship between the service recipient and the service provider is defined in contracts describing the mutual obligations. The contract value of IT outsourcing partnerships is substantial, meaning over 10 million US$ or Euros. These contracts are also long-term contracts. The average duration of contracts is over 36 months (Cox, 2002). This description is based on the work of Lacity and Hirschheim (1993), Willcocks, Fitzgerald and Feeny (1995), and Currie and Willcocks (1998). The IT services outsourcing market is still growing every year, approximately 10% (Cox, 2002). Therefore, it is essential that sufficient attention be devoted to the governance of IT outsourcing partnerships.

BACKGROUND

Yet, not all organizations that have outsourced their IT services are satisfied. The reasons why outsourcing does not always result in success are legion. On the basis of both the literature, as well as actual practice, three important factors that influence the success of an IT outsourcing partnership have been identified: maturity of the service recipient, the degree of flexibility inherent in the contracts and the ability of the service providers to integrate the IT services taken over into their own organization.

Maturity

Service recipients with prior outsourcing experience are better able to implement the governance structure required for IT outsourcing partnerships (McFarland & Nolan, 1995). However, most organizations have limited experience with outsourcing (Willcocks & Fitzgerald, 1994), particularly outsourcing relationships that involve the management of multiple IT outsourcing suppliers (Currie & Willcocks, 1998).

Flexibility

Most outsourcing contracts are of long duration, however, and are not structured to allow changes in information needs to be easily operationalized (Klepper, 1995). This requires procedures for dealing with changes that lead to situations that are not covered by the contract: the Liaison Model (Burnett, 1998). This procedure is used as a basis for formulating agreements between both parties that complement the signed IT outsourcing contract: the concept of ex-post negotiations (Hart, 1995).

Integration

The service providers must integrate the IT services, including the IT professionals taken over, into their own organization. Objectives can only be realized through integration (Lacity & Willcocks, 2001). Many service providers have difficulty with this integration, however. In addition to management attention, initial investments constitute an important issue related to integration and result in an increase in costs (Beulen & Ribbers, 2002).

FRAMEWORK

In line with the definition of IT outsourcing partnerships, the descriptive framework for the governance of IT outsourcing partnerships contains three dimensions. The framework includes governance factors for the service recipient, the relationship between the service recipient and the service providers, and the service providers.

Table 1 contains a summary of the governance factors. These are further explained in the next three sections. The governance factors are based on the case studies analyzed – see Table 2 – the interviews with the experts, professional judgment and a study of the literature.

GOVERNANCE FACTORS RELATED TO THE SERVICE RECIPIENT

There are four management factors related to the service recipient: attention to IT within business units, a clear IT strategy, information management as the link between business functions and service providers and a properly

Table 1. Governance factors by dimension

Dimensions	Governance Factors
The service recipient	1.1 Attention to IT within business units=
	1.2 A clear IT strategy
	1.3 Information management as the link between business units and service providers
	1.4 A properly functioning Chief Information Officer
The relationship between the service recipient and the service providers	2.1 Mutual trust between the service recipient and the service provider
	2.2 Experience in establishing and maintaining IT outsourcing relationships
	2.3 Efficient and effective IT outsourcing contracts
	2.4 An audit & benchmark process in place
The service providers	3.1 Adequate contract and account management
	3.2 Adequate service delivery processes
	3.3 The availability of human resources to service providers

Table 2. IT outsourcing partnerships analyzed

Firm	Sector	# of Employees	Region(s)	Total Contract Value in US $	Contract Start Date	Contract Duration (Years)
1	Discrete Manufacturing	6,2	Europe	30 million	1992 (already renewed)	5
2	Utilities	1,8	Europe	23 million	1996	5
3	Discrete Manufacturing	200	Asia	21 million	1998	5
4	Services	200	Europe	40 million	1997	5
5	Process Industry	68	Europe/Asia/ North	Yearly revenues 90 million (750 employees)	1999 onwards	Purchase of internal IT division
6	Process Industry	68	Asia	0.4 million	1999	3
7	Tele-communications	100	Europe/Middle East/Asia	Confidential (> 20 million)	1997	5
8	Media	8	Europe	4 million	1995 (already renewed)	5
9	Discrete Manufacturing	2	Asia	1 million	2000	2
10	Utilities	2	Europe	100 million	2000	5
11	Discrete Manufacturing	200	Europe/Asia/ Americas	Yearly revenues 550 million (initially 1,500 employees)	1990 onwards	Purchase of internal IT division (various contracts)

functioning chief information officer (CIO). These factors are further explained in this section.

Attention to IT within Business Units

The outsourcing objectives of service recipients are often focused on cost savings (Outsourcingproject, 2002). This is not necessarily in line with the information needs of the business units of the service recipient. Building on this, business units should no longer judge IT on the basis of costs, but on the basis of its added value. A shift must take place from the minimization of costs to the maximization of business impact (Kotwica and Field,1999). When business management pays proper attention to IT, the organization is able to anticipate needs in a timely fashion. In view of the ever decreasing time-to-market, it is not an excessive luxury to involve IT proactively in the development of new products and services.

A Clear IT Strategy

IT strategy is defined as the strategy of the service recipient in relation to its information technology and IT services, and the role these play, or will play, within the service recipient (King, 1978). The development and implementation of the IT strategy is the responsibility of the CIO. An IT strategy is essential because organizations are able to implement new technologies within their organization on the basis of this strategy, and will subsequently be able to derive strategic benefits from this (Earl, 1987). From a governance perspective, it is also important to create alignment between the business and IT. The IT strategy can be a facilitating factor in accomplishing this (Henderson & Venkatraman, 1993).

Information Management as the Link between Business Units and Service providers

The information manager is responsible for the alignment of the demand for IT services by the service recipient's business units with the services provided by the service providers (Quinn, Doorley, & Paquette, 1990). The information manager is furthermore responsible for the IT outsourcing partnership and supports the CIO in the implementation of the IT strategy (McFarland & Nolan, 1995). The cost of the effort expended by the information management function is between 2% and 10% of the contract value. In order to be able to properly carry out these tasks, information managers require knowledge of the business operations, as well as IT (Willcocks & Fitzgerald, 1994).

A Properly Functioning Chief Information Officer

The CIO is responsible for the development and implementation of the IT strategy and carries final responsibility for the IT outsourcing partnership relationships (Lacity & Willcocks, 2001). A CIO could also play an important role in the alignment of business operations with IT. Many organizations set up an IT board for this purpose (Dreyfuss, 2002). This is an organization-wide steering committee in which all business units and information management is represented and which is chaired by the CIO. The IT board is able to explore the political field of influence and decisions concerning the IT services to be provided can be prepared and discussed here.

GOVERNANCE FACTORS FOR THE RELATIONSHIP

There are four factors that govern the relationship: mutual trust between the service recipient and the service provider; experience in establishing and maintaining IT outsourcing relationships; efficient and effective outsourcing contracts; and an audit & benchmark process in place. These governance factors are described in further detail in this section.

Mutual Trust between the Service recipient and the Service provider

Trust is a particularly important criterion for the selection process used by service recipients (Apte, 1990; Lacity & Hirschheim, 1993). But trust is also necessary for maintaining a relationship. This is not easy in actual practice. There does not always exist equality, a true partnership, between the service recipient and the service provider. This issue is not specific to the IT services industry. Kraljic (1983) has been pleading for the creation of partnerships since the eighties: "Purchasing must become Supplier Management". The culture factor is furthermore also important for creating trust (Hofstede, 1980). Grönroos (1990) states that the culture of a service provider is focused on creating service-oriented attitudes and proactive behaviors.

Experience Establishing and Maintaining IT Outsourcing Relationships

Experience is an important selection criterion for choosing a service provider (McFarland & Nolan, 1995). In addition, mature service providers use a standard methodology for setting up and managing IT outsourcing partnerships. Aside from the experiences of the service provider, it is also important for the service recipient to build up experience in outsourcing. Lacity and Hirschheim (1993) state that the lack of experience with outsourcing is an important argument for deciding not to outsource: "customers' inexperience with IS outsourcing".

Efficient and Effective Outsourcing Contracts

Many authors, such as Lacity and Hirschheim (1993), and Kern and Willcocks emphasize the importance of a contract "…the centrality of the contract" (Kern and Willcocks, 2002). Aside from a precise description of the IT services to be provided and the agreed upon service levels, it is important for contracts to be flexible (Klepper, 1995). One way of achieving flexibility is to manage IT as a portfolio and to enter into contract with multiple service providers.

An audit & Benchmark Process in Place

It is important for the service recipient to formulate agreements about the conduct of audits (Allen, 1975; Willcocks & Fitzgerald, 1994): a verification of the service provider's processes. In addition, agreements must be made about establishing regular benchmarks [Lac'98]: Is the price/ quality ratio of the IT services provided in conformance with the marketplace? Service providers that deliver their services in a process-oriented manner, for example, through means of ISO certification, the implementation of ITIL or CMM, are generally in a better position to provide the services contracted for.

GOVERNANCE FACTORS FOR SERVICE PROVIDERS

There are three governance factors related to the service provider: adequate contract and account management, adequate service delivery processes and the availability of human resources. These factors are explained in further detail in the following subsections.

Adequate Contract and Account Management

Contract and account management make up the front office of the service provider and are denoted as the "customer outsourcing interface" by MacFarland and Nolan (1995). It is important in this regard to make a clear distinction between contract management and account management (Beulen, 2000). Contract management is responsible for the operational management and therefore the direction of the service delivery processes. Contract management is therefore focused on the effectiveness and efficiency of the agreed upon contractual commitments. Account management is responsible for maintaining the relationship with the service recipient and is focused on obtaining an extension of existing contracts and on expanding the services provided through means of new contracts.

Adequate Service Delivery Processes

The existence of adequate service delivery processes are considered to be a "core capability" by Feeny, Earl and Edwards (1997): "delivery of IS services" and using Mintzberg's (1979) terminology, the "operating core". When the service delivery processes are set up, it is essential to make a distinction between the service provider's service delivery units and competence centers. Service delivery units are involved in the delivery of the IT services that have been contractually agreed upon with the service provider's customers. The service delivery units must be assessed in terms of their degree of effectiveness and efficiency and are directly controlled by the contract managers. Competence centers are involved in researching the potential of new technological developments and which are responsible for building up knowledge about these new technologies (Cash, McFarlan & McKenney, 1988).

The Availability of Human Resources to Service providers

From the mid-nineties to the year 2000, the shortage of IT professionals was seen as a factor that would limit the growth of the IT services industry. Due to economic developments, there has been a decrease in this shortage in the labor market (Hirschheim & Lacity, 2000). In spite of this, service providers must continue to pay attention to the availability of IT professionals for delivering their IT services.

FUTURE TRENDS

Four trends related to IT outsourcing partnerships can be identified: business process outsourcing (BPO), multiple sourcing, offshore outsourcing and insourcing. Except for insourcing, these trends complicate the governance of IT outsourcing partnerships both for the service recipients and the service providers.

Business Process Outsourcing

In addition to outsourcing the IT services for their supporting processes, an increasing number of organizations are also deciding to outsource complete business processes. Therefore, BPO requires different skills from the IT professionals. IT professionals must have greater knowledge of the business processes, and they must possess business knowledge. On top of that, the service providers have to also employ non-IT staff in order to handle the administrative activities related to the offered services.

Table 3. Trends in IT outsourcing partnerships

Business Process Outsourcing
Multiple Sourcing
Offshore Outsourcing
Insourcing

Multiple Sourcing

Many organizations outsource their IT services to more than one service provider: multiple sourcing. This heavily impacts the required service recipient's management attention. Also the complexity of the contracts increases by multiple sourcing. Not only the responsibilities between the service recipient and the service provider but also the responsibilities between service providers have to be detailed in the contracts.

Offshore Outsourcing

In offshore outsourcing, the service providers execute the services out of a low cost country to their customer, the service recipient, which is located in a high cost country. Due to time difference, language barriers, and cultural differences, managing offshore outsourcing partnerships requires extra management attention from both the service recipient and the service provider. In the contractual part of the partnership, jurisdiction and settlement need to be addressed properly.

Insourcing

Insourcing means that not all of the IT services are performed by one or more service providers and that a portion of the IT services are provided by the internal IT division. This implies that some aspects will continue to be looked after by the internal IT division itself. The internal IT division also requires attention. Some service recipients will choose the insourcing option to avoid outsourcing debacles and to ensure business continuity. This remains the question if the internal IT division is capable in executing the services properly.

CONCLUSION

Achieving governance in IT outsourcing partnerships is not easy. It requires a lot of management attention from both the service recipient and the service provider. Also, implementing governance factors for the relationship between the service recipient and the service provider is key. However, time does not stand still. Business process outsourcing is expected to become increasingly dominant and increasing numbers of multiple outsourcing relationships will come into being. Also, offshore outsourcing is emerging. In addition, service recipients will also selectively repatriate certain components of their IT services: insourcing. This will have an impact and will lead to an increase of complexity in managing IT outsourcing partnerships.

REFERENCES

Allen, B. (1975). Guide to computers. *Harvard Business Review*, July-August.

Apte, U. (1990). Global outsourcing of information systems and processing services. *The Information Society*, *7*, 287-303.

Beulen, E. (2000). Beheersing van IT-outsourcingsrelaties. PhD Thesis, Tilburg University, The Netherlands (in Dutch).

Beulen, E., & Ribbers P. (2002). Managing complex IT outsourcing – partnerships. *Proceedings of Hawaii International Conference on Systems Sciences 2002*, 0-7695-1435-9/02.

Burnett, R. (1998). *Outsourcing IT - The legal aspects.* Aldershot: Gower.

Cash, J., McFarlan, F. & McKenney, J. (1988). *Corporate information systems management, the issues facing senior executives.* Irwin.

Cox, R. (2002). Gartner services and sourcing scenario. *Gartner Symposium ITXPO 2002*, Florence, Italy, April 8-10.

Currie, W., & Willocks, L. (1998). New strategies in IT-outsourcing: major trends and global best practices. *Business Intelligence.*

Dreyfuss, C. (2002). Sourcing governance: What, where and how. *Gartner IT Services and Sourcing Summit 2002*, 15-17 May 2002, Nevada USA, C1, STD5.

Earl, M. (1987). Information systems strategy formulation. In R. Boland & R. Hirschheim (Ed.), *Critical issues in information system research*, Wiley.

Feeny, D., Earl, M., & Edwards, B. (1997). Information systems organization: The role of users and specialists. Chapter 7 in L. Willcocks, D. Feeny, & G. Islei (Eds.), *Managing IT as a strategic resource*. McGraw Hill.

Gantz, J. (1990). Outsourcing: Treat or salvation? *Networking Management*, *10*.

Grönroos, C. (1990). *Service management and marketing, managing the moment of truth in service competition*. Lexington books.

Hart, O. (1995). *Contracts and financial structure*. Oxford University Press.

Henderson, J., & Venkatraman, N. (1993). Strategic alignment: Leveraging information technology for transforming organizations. *IBM Systems Journal*, *32*(1).

Hirschheim, R. & Lacity, M. (2000). The myths and realities of information technology insourcing. *Communication of the ACM*. 43(2).

Hofstede, G. (1980). *Culture's consequences*. Sage Publications

Kern, T. and Willcocks, L. (2002). Exploring relationships in information technology outsourcing: the interaction approach. *European Journal of Information Systems*, 11, 3-19.

King, W. (1978, June). Strategic planning for management information systems, *MIS Quarterly*, *2*(2), 27-37.

Klepper, R. (1995). The management of partnering development in I/S outsourcing. *Journal of Technology*, 10.

Kotwica, K. & Field, T. The changing role of the chief information officer, *www.CIO.com*, last update 11 January 1999.

Kraljic, W. (1983, September-October). Purchasing must become supply management. *Harvard Business Review*, 109-117.

Lacity, M., & Hirschheim, R. (1993). *Information systems outsourcing*. Wiley.

Lacity, M., & Willcocks, L. (2001). *Global information technology outsourcing: In search of business advantage*. John Wiley & Sons.

McFarland, W., & Nolan, R. (1995). How to manage an IT-outsourcing alliance. *Sloan Management Review*, *36*, 9-23.

Mintzberg, H. (1979). *The structuring of organizations*. Englewood Cliffs, NJ: Prentice Hall.

Outsourcingproject (2002). Montgomery Research Europe, ISSN 1476-2064.

Quinn, J., Doorley, L., & Paquette, P. (1990). Technology in services: Rethinking strategic focus. *Sloan Management Review*, 79-87.

Rochester, J., & Douglas, D. (Eds.) (1990). Taking an objective look at outsourcing, *I/S Analyzer*, *28*, 9.

Willcocks, L., & Fitzgerald, G. (1994). A business guide to outsourcing IT. *Business Intelligence*.

Willcocks, L., Fitzgerald, G., & Feeny, D. (1995). Outsourcing the strategic implications. *Long Range Planning*, *28*(5).

KEY TERMS

Business Process Outsourcing: Service recipients handover the responsibility for the execution of complete business processes to service providers. Most of the business processes in business process outsourcing are IT related processes.

Insourcing: The internal IT division of service recipients executes the IT services to the departments of the service recipient.

IT Outsourcing Partnership: The relationship between the service recipient and the service provider is defined in contracts describing the mutual obligations. The contract value of IT outsourcing partnerships is substantial, meaning over 10 million US$ or Euros. These contracts are also long-term contracts. The average duration of contracts is over 36 months.

Multiple Sourcing: Service recipients that have outsourced the responsibility to execute the services to more than one service provider.

Offshore Outsourcing: Service providers execute the IT services out of a low cost country to their customer, the service recipient, which is located in a high cost country.

Outsourcing Contracts: Describe long-term mutual obligations between the service provider(s) and the service recipient. The contract value is substantial, meaning over 10 million US$ or Euros. The average duration of outsourcing contracts is over 36 months.

Service Provider: Responsible for the execution of the IT services to the service recipient in IT outsourcing relationships.

Service Recipient: Hands over the responsibility for the execution of the IT services to one or more service providers but remains the responsibility for managing the IT outsourcing partnership.

Governance Structures for IT in the Health Care Industry

Reima Suomi
Turku School of Economics and Business Administration, Finland

INTRODUCTION

The pressures for the health care industry are well known and very similar in all developed countries (i.e., altering population, shortage of resources for staff and from taxpayers, higher sensitivity of the population for health issues, new and emerging diseases, etc.). Underdeveloped countries experience different problems, but they have the advantage of learning from the lessons and actions that developed countries underwent perhaps decades ago. On the other hand, many solutions also exist, but they all make the environment even more difficult to manage (i.e., possibilities of networking, booming medical and health-related research and knowledge produced by it, alternative caretaking solutions, new and expensive treatments and medicines, promises of biotechnology, etc.).

From the public authorities' points of view, the solution might be easy—outsource as much as you can out of this mess. Usually, the first services to go are marginal operational activities, such as laundry, cleaning, and catering services. It is easy to add information systems to this list, but we believe this is often done without a careful enough consideration. Outsourcing is often seen as a trendy, obvious, and easy solution, which has been supported by financial facts on the short run. Many examples show that even in the case of operational information systems, outsourcing can become a costly option, not to mention lost possibilities for organizational learning and competitive positioning through mastering of information technology.

BACKGROUND

We have found the following reasons for the late adoption of modern information technology in the health care sector (Suomi, 2000):

- Fragmented industry structure
- Considerable national differences in processes
- Strong professional culture of medical care personnel
- One-sided education
- Handcrafting traditions
- Weak customers
- Hierarchical organization structures

ICT and governance structures meet in two ways. On one side, ICT enables new governance structures for the health care industry. On the other, it is an object in need of governing. As both sectors offer a multitude of new possibilities, innovations are called for in the industry (Christensen, Bohmer, & Kenagy, 2000).

IT governance thinking matures in organizations as any other discipline. Van Grembergen, De Haes, and Guldentops (2003) have defined the following stages in their IT Governance Maturity Model:

- Non-existent
- Initial/ad-hoc
- Repeatable but intuitive
- Defined process
- Managed and measurable
- Optimized

Needless to say, in the health care industry, IT Governance thinking is non-existent or initial/ad hoc in the best situation.

THE MEANING OF ICT GOVERNANCE STRUCTURE IN HEALTH CARE

IT is an old acronym for information technology. Nowadays, it is replaced often with the term ICT, referring to information and communication technology. This emphasizes the communication services that are developing very quickly, such as the Internet and mobile services. The letter *C* is often upgraded to the second dimension: alongside communication it can refer to contents. IT or ICT governance is defined (IT Governance Institute, 2001) as follows:

IT governance is the responsibility of the board of directors and executive management. It is an integral *part of enterprise governance and consists of the leadership and organizational structures and processes that ensure that the organization's IT sustains and extends the organization's strategies and objectives.*

For many, there is a temptation to understand governance as just a synonym for management. This is an oversimplification. Management is a goal-oriented activity, whereas governance is often given from the outside, and organizations just have to live with it. This is not to say that all governance structures would be beyond management control; management can influence most governance structures, at least in the long run. The long run is a key term in many aspects. We talk about structures that are semi-permanent and not changed very frequently. Structure is a term closer to architecture than to infrastructure; governance structures are architectural terms and are then implemented into infrastructures through different organizational forms. The terms *organization form* and *governance structure* are not synonyms. Organizational forms are more formal and touch upon one organization, whereas governance structures are found in a richer selection of forms and organize themselves over a number of organizations. Table 1 summarizes our discussion.

Governance structures are present in almost any human decision-making situation. In Table 2, we have a collection of key aspects of governance structure issues in health care.

FUTURE TRENDS

One of the biggest changes in the industry is that information related to health, sickness and medicines is not scarce. Internet is a rich source of such information, at different levels of expertise and at different languages. The gap between what information is available and what a health-care professional should know is growing very fast (Weaver 2002). This will shift the power balance between health professionals and patients: increasingly often the patients are the best experts on their disease. Different electronic forums or Virtual Communities (Rheingold 1993) related to health are born on the Internet. They have different services and values to offer to the healthy ones and to the chronic and acute sick (Utbult, 2000). Similarly, the interaction between the patients and health care professionals is going to change: electronic means are going to take share from face-to-face meetings (Cain, Sarasohn-Kahn & Wayne 2000; Gibson 2003).

For organizing patient flows through the health care system modern ICT offers many possibilities. Should patient data be all the time available anywhere though electronic means, would the Healthcare Supply Chains be much more effective (More & McGrath 2001). Effectiveness means that patients are taken care of in the best and most effective places, be they public or private, and of right level of expertise. As patient data can be electronically cumulated into huge databases, these databases can be used for different statistical, research and other purposes. This calls for care and proper legislation giving the principles.

Managing and building governance structures for ICT in health care organizations is not that much different from other organizations. Even in health care organizations, the scope and status of information resource management has to be decided. Issues such as sourcing decisions, charging arrangements, data privacy, and security issues all deserve their attention. There are certain problems that need to be solved in this area:

Table 1. *Comparison of terms management, organizational form, and governance structure*

	Management	Organizational Form	Governance Structure
Time perspective	Short	Medium	Long
Focus	Action	Internal organization	Inter-organizational structures
Management Control	In action	Easy	Difficult
Metaphor	Communication channels	Infrastructure	Architecture
Character	Concrete	Formal	Abstract

***Table 2.** ICT governance structure issues in health care*

ICT as an enabler
• Health-related information on the web
• Private-public sector co-operation
• Allocation of patients to different levels of care
• Customer contacts distribution between electronic and classical means
• Ownership, structure and allocation of patient, population-level and other critical data
• Electronic forums for patients to interact
• Electronic prescription systems
ICT as an object to be governed
• New legislation needs because of the new data processing possibilities
• Data privacy and security
• Structure and status of the information resource management in health care units
• ICT-general management partnership
• Sourcing decisions of ICT
• Charging arrangements on ICT-services

- Data privacy and security needs are extremely important and might sometimes conflict with optimal care.
- As the area is new, legislation is often lagging behind.
- The field is a meeting place for two strong professional cultures—medical doctors and ICT-professionals—that might bring along difficulties.

CONCLUSION

Health is undoubtedly among the most important issues for all of us. In a modern society, the threats towards health are changing all the time, but at the same time, the possibilities to maintain health and cure illnesses grow exponentially. The task is to make needs and solutions meet in an effective way. This is about information and communication technologies and governance structures.

Modern ICT allows health care organizations to structure themselves in new, innovative ways, and simultaneously to empower the customers to interact with the organizations, with fellow patients, and with information sources in revolutionary new ways. In this environment, health care professionals also have to adjust their roles.

REFERENCES

Cain, Mary M., Sarasohn-Kahn, Jane, & Wayne, Jennifer C. (2000). *Health e-people: The online consumer experience*. California Health Care Foundation, Oakland, CA.

Christensen, Clayton M., Bohmer, Richard, & Kenagy, John (2000, September-October). Will distruptive innovations cure health care? *Harvard Business Review*, 102-112.

Gilson, L. (2003). Trustnext term and the development of health care as a social institution. *Social Science & Medicine, 56*(7), 1453-1468.

IT Governance Institute (2001). IT governance executive summary. Retrieved January 10, 2004, from www.itgi.org

More, Elizabeth, & McGrath, G. Mike (2001). Reengineering the healthcare supply chain in Australia: The PeCC initiative. In Robert Stegwee & Ton Spil (Eds.), *Strategies for healthcare information systems* (pp. 114-125). Hershey, PA: Idea Group Publishing.

Rheingold, Howard (1993). *The virtual community – Homesteading on the electronic frontier*. Addison-Wesley, New York.

Suomi, Reima (2000). Leapfrogging for modern ICT usage in the health care sector. *Proceedings of the 8th ECIS conference*, Vienna, Australia.

Utbult, Mats (2000). Näthälsä. *Internetpatienter möter surfande doktorer – uppstår confrontation eller samarbete*. TELDOK Rapport 138. Stockholm.

Van Grenbergen, Wim, De Haes, Steven, & Guldentops, Erik (2003). Structures, processes and relational mechanisms for IT governance. In Wim Van Grembergen (Ed.), *Strategies for information technology governance* (pp. 1-36). Hershey, PA: Idea Group Publishing.

Weawer, Robert R. (2002, March). Resistance to computer innovation: Knowledge coupling in clinical practice. *Computers and Society*, 16-21.

KEY TERMS

Electronic Patient Record: All health-related information related to a patient in electronic form, assembled as a single entity.

Electronic Prescription: Prescriptions created and handled in electronic form in an integrated information system.

Healthcare Supply Chain: A managed set of activities related to the health care activity of a patient, organized so that all necessary information is available all the time and the participants in the chain have a complete picture of the total process.

Sourcing Decision: Decision whether to buy goods/services from the market or to make them self. Sourcing decision can be made as an independent decision unit or as a part of a bigger group.

Virtual Community: A social aggregation on the Internet when people interact long enough to form personal relationships.

Graph Encoding and Recursion Computation

Yangjun Chen
University of Winnipeg, Canada

INTRODUCTION

It is a general opinion that relational database systems are inadequate for manipulating composite objects that arise in novel applications such as Web and document databases (Abiteboul, Cluet, Christophides, Milo, Moerkotte & Simon, 1997; Chen & Aberer, 1998, 1999; Mendelzon, Mihaila & Milo, 1997; Zhang, Naughton, Dewitt, Luo & Lohman, 2001), CAD/ CAM, CASE, office systems and software management. Especially, when recursive relationships are involved, it is cumbersome to handle them in relational databases, which sets current relational systems far behind the navigational ones (Kuno & Rundensteiner, 1998; Lee & Lee, 1998). To overcome this problem, a lot of interesting graph encoding methods have been developed to mitigate the difficulty to some extent. In this article, we give a brief description of some important methods, including analysis and comparison of their space and time complexities.

BACKGROUND

A composite object can be generally represented as a directed graph (digraph). For example, in a CAD database, a composite object corresponds to a complex design, which is composed of several subdesigns. Often, subdesigns are shared by more than one higher-level design, and a set of design hierarchies thus forms a directed acyclic graph (DAG). As another example, the citation index of scientific literature, recording reference relationships between authors, constructs a directed cyclic graph. As a third example, we consider the traditional organization of a company, with a variable number of manager-subordinate levels, which can be represented as a tree hierarchy.

In a relational system, composite objects must be fragmented across many relations, requiring joins to gather all the parts. A typical approach to improving join efficiency is to equip relations with hidden pointer fields for coupling the tuples to be joined. The so-called *join index* is another auxiliary access path to mitigate this difficulty. Also, several advanced join algorithms have been suggested, based on hashing and a large main memory. In addition, a different kind of attempt to attain a compromise solution is to extend relational databases with new features, such as *clustering* of composite objects, by which the concatenated foreign keys of ancestor paths are stored in a primary key. Another extension to relational system is *nested relations* (or NF^2 relations). Although it can be used to represent composite objects without sacrificing the relational theory, it suffers from the problem that subrelations cannot be shared. Moreover, recursive relationships cannot be represented by simple nesting because the depth is not fixed. Finally, *deductive databases* and *object-relational databases* can be considered as two quite different extensions to handle this problem (Chen, 2003; Ramakrishnan & Ullman, 1995).

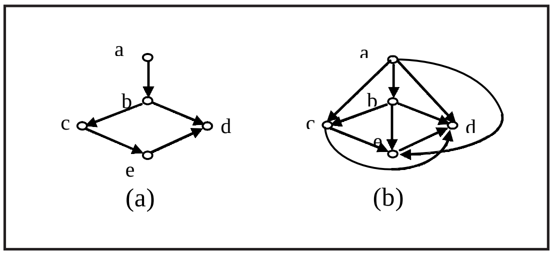
In the past decade, a quite different kind of research has also been done to avoid *join* operations based on *graph encoding*. In this article, we provide an overview on most important techniques in this area and discuss a new encoding approach to pack "ancestor paths" in a relational environment. It needs only $O(e \cdot b)$ time and $O(n \cdot b)$ space, where b is the breadth of the graph, defined to be the least number of disjoined paths that cover all the nodes of a graph. This computational complexity is better than any existing method for this problem, including the graph-based algorithms (Schmitz, 1983), the graph encoding (Abdeddaim, 1997; Bommel & Beck; Zibin & Gil, 2001) and the matrix-based algorithms (La Poutre & Leeuwen, 1988).

RECURSION COMPUTATION IN RELATIONAL DATABASES

We consider composite objects represented by a digraph, where nodes stand for objects and edges for parent-child relationships, stored in a binary relation. In many applications, the transitive closure of a digraph needs to be computed, which is defined to be all ancestor-descendant pairs. For instance, the transitive closure of the graph in Figure 1(a) is shown in Figure 1(b).

In this article, we mainly overview the graph encoding in a relational environment. The following is a typical structure to accommodate part-subpart relationships (Cattell & Skeen, 1992):

Figure 1. A graph and its transitive closure



- Part(Part-id, ...)
- Connection(Parent-id, Child-id, ...)

where Parent-id and Child-id are both foreign keys, referring to Part-id. In order to speed up the recursion evaluation, we will associate each node with a pair of integers, which helps to recognize ancestor-descendant relationships.

In the rest of the article, the following three types of digraphs will be discussed.

(i) Tree hierarchy, in which the parent-child relationship is of one-to-many type; that is, each node has at most one parent.
(ii) Directed acyclic graph (DAG), which occurs when the relationship is of many-to-many type, with the restriction that a part cannot be sub/superpart of itself (directly or indirectly).
(iii) Directed cyclic graph, which contains cycles.

Later we will use the term *graph* to refer to the *directed graph*, since we do not discuss non-directed ones at all.

RECURSION WITH RESPECT TO TREES

Perhaps the most elegant algorithm for encoding is *relative numbering* (Schmitz, 1983), which guarantees both optimal encoding length of log*n* bits and constant time recursion tests for trees. Consider a tree *T*. By traversing *T* in *postorder*, each node *v* will obtain a number (it can be integer or a real number) *post*(*v*) to record the order in which the nodes of the tree are visited. A basic property of postorder traversal is

$$post(v) = \max\{\, post(u) \mid u \in \text{descendant}(v)\}.$$

Let $l(v)$ be defined by

$$l(v) = \min\{\, post(u) \mid u \in \text{descendant}(v)\}.$$

Then, a node u is a descendant of v if $l(v) \leq post(u) \leq post(v)$.

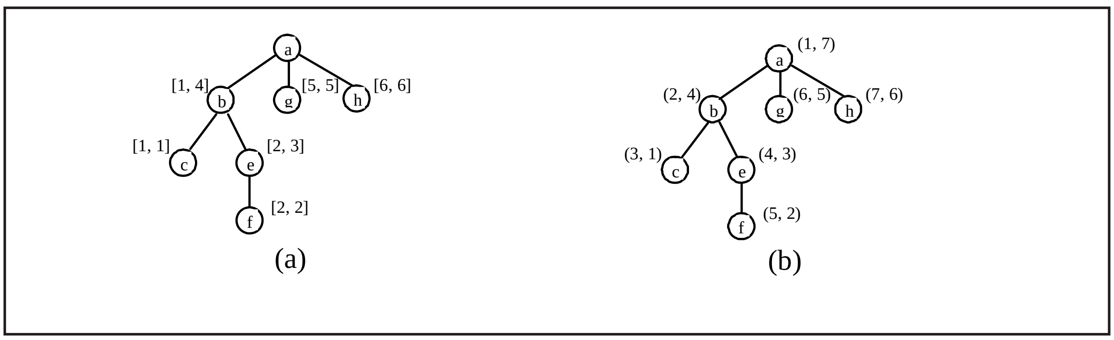
In terms of this relationship, each node v can be encoded by an interval $[l(v), post(v)]$ as exemplified by Figure 2(a).

Another interesting graph encoding method is discussed in Knuth (2003), by which each node is associated with a pair $(pre(v), post(v))$, where $pre(v)$ represents the preorder number of v and can be obtained by traversing *T* in a preorder. The pair can be used to characterize the ancestor-descendant relationship as follows.

Proposition 1 - Let v and v' be two nodes of a tree *T*. Then, v' is a descendant of v if $pre(v') > pre(v)$ and $post(v') < post(v)$.

Proof. See Exercise 2.3.2-20 in Knuth (1969).

Figure 2. Labeling a tree



If *v*' is a descendant of *v*, then we know that *pre*(*v*') > *pre*(*v*) according to the preorder search. Now we assume that *post*(*v*') > *post*(*v*). Then, according to the postorder search, either *v*' is in some subtree on the right side of *v*, or *v* is in the subtree rooted at *v*', which contradicts the fact that *v*' is a descendant of *v*. Therefore, *post*(*v*') must be less than *post*(*v*).

Figure 2(b) helps for illustration. The first element of each pair is the preorder number of the corresponding node and the second is its postorder number. With such labels, the ancestor-descendant relationships can be easily checked. For instance, by checking the label associated with b against the label for f, we see that b is an ancestor of f in terms of Proposition 1. Note that b's label is (2, 4) and f's label is (5, 2), and we have 2 < 5 and 4 > 2. We also see that since the pairs associated with g and c do not satisfy the condition given in Proposition 1, g must not be an ancestor of c and vice versa.

Definition 1 - (*label pair subsumption*) Let (p, q) and (p', q') be two pairs associated with nodes *u* and *v*. We say that (p, q) is subsumed by (p', q'), denoted $(p, q) \prec (p', q')$, if $p > p'$ and $q < q'$. Then, *u* is a descendant of *v* if (p, q) is subsumed by (p', q').

According to the tree labeling discussed previously, the relational schema to handle recursion can consist of only one relation of the following form:

Node(Node_id, *label_pair*, ...),

where label_pair is used to accommodate the preorder and the postorder numbers of the nodes of a graph, denoted label_pair.*preorder* and label_pair.*postorder*, respectively. Then, to retrieve the descendants of node *x*, we issue two queries as below.

```
Q1: SELECT  label_pair
    FROM    Node
    WHERE   Node_id = x
```

Let the label pair obtained by evaluating Q_1 be *y*. Then, the second query is of the following form:

```
Q2: SELECT  *
    FROM    Node
    WHERE   label_pair.preorder > y.preorder
    and     label_pair.postorder < y.postorder
```

From this, we can see that with the tree labeling, the recursion w.r.t. a tree can be handled conveniently in a relational environment. If a tree is stored in a preorder, we can control the search so that it stops when the first pair that is not subsumed by *y* is encountered.

RECURSION WITH RESPECT TO DGAS

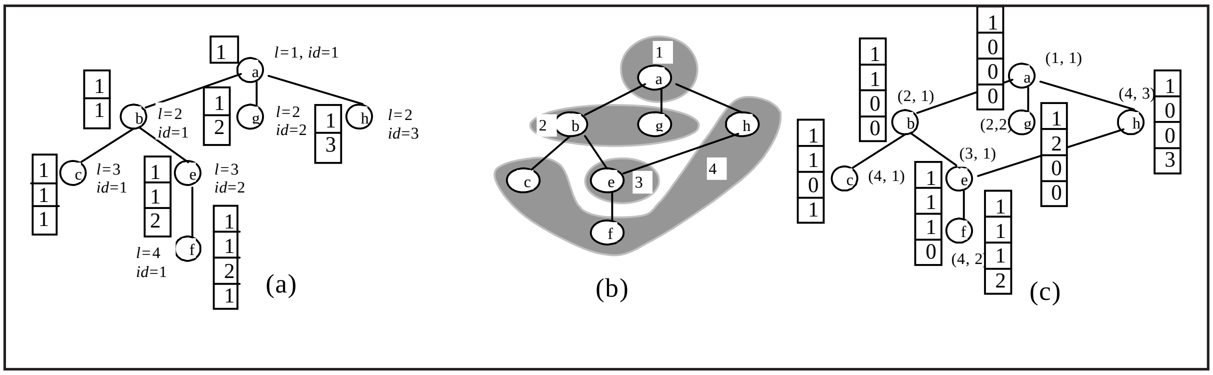
In this section, we consider the recursion computation w.r.t. DAGs.

- *Range-compression*
  The relative numbering was extended to handle DAGs by (Agrawal, Borgida & Jagadish, 1989). Their method is called *range-compression* encoding. It encodes each node *v* in a DAG as an integer *id*(*v*), obtained by scanning a certain spanning forest of the DAG in postorder. In addition, each node *v* is associated with an array *A* of length k ($1 \le k \le n$), in which the *i*th entry is an interval $[l_i, r_i]$ with the property that a node *u* is a descendant of *v* if there exists an integer *j* such that $l_i \le id(u) \le r_j$. Since an array contains *n* entries in the worst case, this encoding method needs $O(n^2)$ space and time.
- *Cohen's encoding and EP encoding*
  In the language research, the graph encoding is also extensively explored for type-subtype checking that is in essence a recursion computation. Thus, the methods proposed in that area can be employed in a relational environment. Perhaps the most interesting method is Cohen's encoding for tree structures (Cohen, 1991). It was generalized to DAGs by Krall, Vitek and Horspool (1997) into what is called *packed encoding* (PE).

Cohen' encoding stores with each node *v* its level l_v, its unique identifier id_v, as well as an array A_v such that for each node $u \in$ ancestor(*v*), $A_v[l_u] = id_u$. The test $v \in$ descendant(*u*) is then carried out by checking whether both $l_v \ge l_u$ and $A_v[l_u] = id_u$ hold. An example of the actual encoding is given in Figure 3(a).

EP encoding partitions a DAG into a number of slices: $S_1, ..., S_k$ for some *k* such that no two ancestors of a node can be on the same slice. In addition, each node *v* is assigned a unique identifier id_v within its slice *S*. Thus, *v* is identified by the pair (s_v, id_v), where s_v is the number for *S*. Furthermore, each node *v* is associated with an array A_v such that for all $u \in$ ancestor(*v*), $A_v[s_u] = id_v$. The DAG shown in Figure 3(b) is partitioned into four slices numbered 1, 2, 3 and 4, respectively. Accordingly, the DAG can be encoded as shown in Figure 3(c). We note that in EP encoding slices play a role similar to that of levels in Cohen's encoding. In fact, Cohen's algorithm partitions

Figure 3. Cohen's encoding and EP encoding



a tree into levels while EP encoding partitions a DAG into slices. According to Fall (1995), it is NP-hard to find a minimal partition of slices. Moreover, if the sizes of slices is bounded by a constant, the array associated with a node is of length O(n) at average. So the space and time overhead of EP encoding is on the order of O(n^2).

- *Pre-postorder encoding*
  Now we discuss a new encoding method, which needs only O($e \cdot b$) time and O($n \cdot b$) space, where b is the breadth of the graph, defined to be the least number of disjoined paths that cover all the nodes of a graph.

What we want is to apply the pre-postorder encoding discussed previously to a DAG. To this end, we establish a *branching* of the DAG as follows (Tarjan, 1977).

Definition 2 - (*branching*) A subgraph $B = (V, E')$ of a digraph $G = (V, E)$ is called a branching if it is cycle-free and $d_{indegree}(v) \le 1$ for every $v \in V$.

Clearly, if for only one node r, $d_{indegree}(r) = 0$, and for all the rest of the nodes, v, $d_{indegree}(v) = 1$, then the branching is a directed tree with root r. Normally, a branching is a set of directed trees. Now, we assign every edge e a same cost (e.g., let cost $c(e) = 1$ for every edge e). We will find a branching for which the sum of the edge costs, $\sum_{e \in E} c(e)$, is maximum.

For example, the trees shown in Figure 4(b) are a maximal branching of the graph shown in Figure 4(a) if each edge has the same cost.

Assume that the maximal branching for $G = (V, E)$ is a set of trees T_i with root r_i (i = 1, ..., m). We introduce a *virtual root* r for the branching and an edge $r \rightarrow r_i$ for each T_i, obtaining a tree G_r, called the representation of G. For instance, the tree shown in Figure 4(c) is the representation of the graph shown in Figure 2(a). Using Tarjan's algorithm for finding optimum branchings (Tarjan, 1977), we can always find a maximal branching for a directed graph in O($|E|$) time if the cost for every edge is equal to each other. Therefore, the representative tree for a DAG can be constructed in linear time.

Figure 4. A DAG and its branching

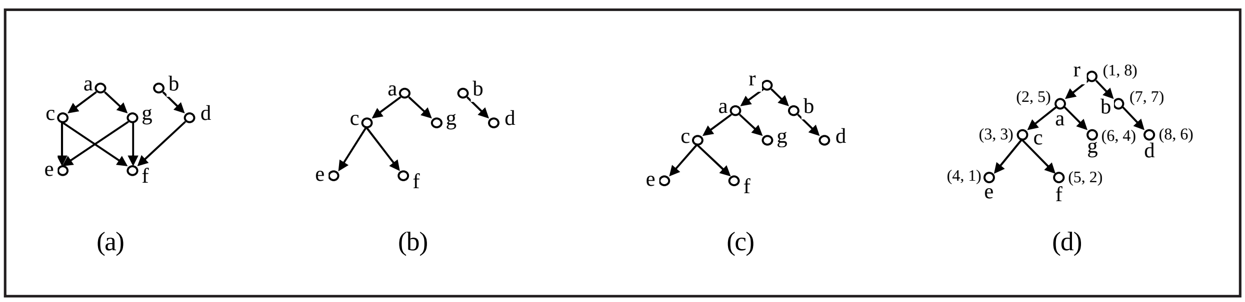
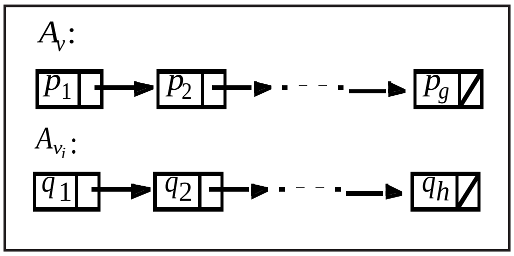


Figure 5. Linked lists associated with nodes in G



We can label G_r in the same way as shown in the previous subsection. See Figure 4(d).

In a G_r (for some G), a node v can be considered as a representation of the subtree rooted at v, denoted $T_{sub}(v)$; and the pair (*pre*, *post*) associated with v can be considered as a pointer to v, and thus to $T_{sub}(v)$. (In practice, we can associate a pointer with such a pair to point to the corresponding node in G_r.) In the following, what we want is to construct a pair sequence: $(pre_1, post_1)$, ..., $(pre_k, post_k)$ for each node v in G, representing the union of the subtrees (in G_r) rooted respectively at $(pre_j, post_j)$ (j = 1, ..., k), which contains all the descendants of v. In this way, the space overhead for storing the descendants of a node is dramatically reduced. Later we will show that a pair sequence contains at most O(b) pairs, where b is the breadth of G. (The breadth of a digraph is defined to be the least number of the disjoint paths that cover all the nodes of the graph.)

The question is how to construct such a pair sequence for each node v so that it corresponds to a union of some subtrees in G_r, which contains all the descendants of v in G. For this purpose, we sort the nodes of G topologically; that is, $(v_i, v_j) \in E$ implies that v_j appears before v_i in the sequence of the nodes. The pairs to be generated for a node v are simply stored in a linked list A_v. Initially, each A_v contains only one pair produced by labeling G_r.

We scan the topological sequence of the nodes from the beginning to the end and at each step we do the following:

Let v be the node being considered. Let v_1, ..., v_k be the children of v. Merge A_v with each for the child node v_l (l = 1, ..., k) as follows. Assume $A_v = p_1 \rightarrow p_2 \rightarrow \ldots p_g$ and $A_{v_i} = q_1 \rightarrow q_2 \rightarrow \ldots\ q_h$, as shown in Figure 5. Assume that both A_v and A_{v_i} are increasingly ordered. (We say a pair p is larger than another pair p', denoted $p > p'$ if $p.pre > p'.pre$ and $p.post > p'.post$.)

We step through both A_v and A_{v_i} from left to right. Let p_i and q_j be the pairs encountered. We will make the following checkings.

Algorithm *pair-sequence-merge*$(A_i, A_{v_i}$ **);**

(1) If $p_i.pre > q_j.pre$ and $p_i.post > q_j.post$, insert q_j into A_v after p_{i-1} and before p_i and move to q_{j+1}.

(2) If $p_i.pre > q_j.pre$ and $p_i.post < q_j.post$, remove p_i from A_v and move to p_{i+1}. (*p_i is subsumed by q_j.*)

(3) If $p_i.pre < q_j.pre$ and $p_i.post > q_j.post$, ignore q_j and move to q_{j+1}. (*q_j is subsumed by p_i; but it should not be removed from A_{v_i} .*)

(4) If $p_i.pre < q_j.pre$ and $p_i.post < q_j.post$, ignore p_i and move to p_{i+1}.

(5) If $p_i = p_j'$ and $q_i = q_j'$, ignore both (p_i, q_i) and (p_j', q_j'), and move to (p_{i+1}, q_{i+1}) and (p_{j+1}', q_{j+1}'), respectively.

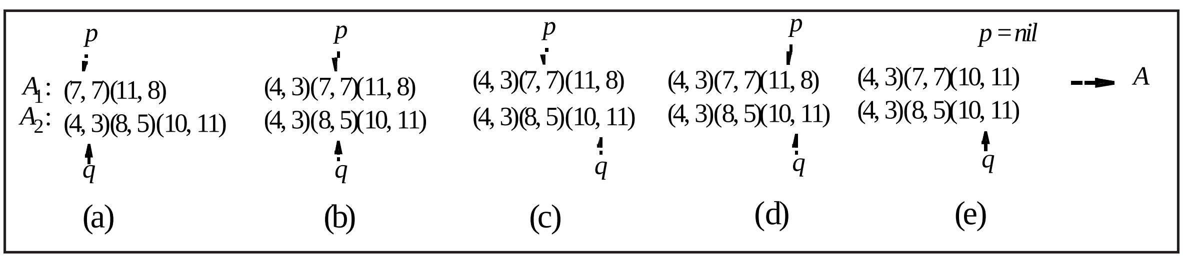
(6) If $p_i = nil$ and $q_j \neq nil$, attach the rest of A_{v_i} to the end of A_v.

The following example helps for illustration.

Example 1 - Assume that A_1 = (7, 7)(11, 8) and A_2 = (4, 3)(8, 5)(10, 11). Then, the result of merging A_1 and A_2 is (4, 3)(7, 7)(10, 11). Figure 6 shows the entire merging process.

In each step, the A_1-pair pointed to by p and the A_2-pair pointed to by q are compared. In the first step, (7, 7) in A_1 will be checked against (4, 3) in A_2 (see Figure 6(a)). Since (4, 3) is smaller than (7, 7), it will be inserted into A_1 before (7, 7) (see Figure 6(b)). In the second step, (7, 7) in A_1 will be checked against (8, 5) in A_2. Since (8, 5) is subsumed by (7, 7), we move to (10, 11) in A_2 (see Figure 6(c)). In the third

Figure 6. An entire merging process



step, (7, 7) is smaller than (10, 11) and we move to (11, 8) in A_1 (see Figure 6(d)). In the fourth step, (11, 8) in A_1 is checked against (10, 11) in A_2. Since (11, 8) is subsumed by (10, 11), it will be removed from A_1 and p becomes *nil* (see Figure 6(e)). In this case, (10, 11) will be attached to A_1, forming the result $A = (4, 3)(7, 7)(10, 11)$ (see Figure 6(e)).

We can store physically the label pair for each node, as well as its label pair. Concretely, the relational schema to handle recursion w.r.t., a DAG can be established in the following form:

Node(Node_id, *label*, *label_sequence*, ...),

where *label* and *label_sequence* are used to accommodate the label pairs and the label pair sequences associated with the nodes of a graph, respectively. Then, to retrieve the descendants of node *x*, we issue two queries. The first query is similar to Q_1:

Q_3: SELECT label_sequence
FROM Node
WHERE Node_id = *x*

Let the label sequence obtained by evaluating Q_3 be *y*. Then, the second query will be of the following form:

•Q_4: SELECT *
FROM Node
WHERE ϕ(label, *y*)

where $\phi(p, s)$ is a boolean function with the input: *p* and *s*, where *p* is a pair and *s* a pair sequence. If there exists a pair *p*' in *s* such that $p \prec p'$ (i.e., $p.pre > p'.pre$ and $p.post < p'.post$), then $\phi(p, s)$ returns *true;* otherwise *false*.

Based on the method discussed in the previous subsection, we can easily develop an algorithm to compute recursion for cyclic graphs. First, we use Tarjan's algorithm for identifying *strongly connected components* (*SCCs*) to find the cycles of a cyclic graph (Tarjan, 1972) (which needs only $O(n + e)$ time). Then, we take each SCC as a single node (i.e., condense each SCC to a node) and transform a cyclic graph into a DAG. Next, we handle the DAG as discussed earlier. In this way, however, all nodes in an SCC will be assigned the same pair (and the same pair sequence). For this reason, the method for computing the recursion at some node *x* should be slightly changed.

FUTURE TRENDS

The computation of transitive closures and recursive relationships is a classic problem in the graph theory and has a variety of applications in data engineering, such as CAD/CAM, office systems, databases, programming languages and so on. For all these applications, the problems can be represented as a directed graph with the edges being not labelled, and can be solved using the techniques described in this article. In practice, however, there exists another kind of problem, which can be represented only by using the so-called weighted directed graphs. For them, the edges are associated with labels or distances and the shortest (or longest) paths between two given nodes are often asked. Obviously, these techniques are not able to solve such problems. They have to be extended to encode path information in the data structure to speed up query evaluation. For this, an interesting issue is how to maintain minimum information but get high efficiency, which is more challenging than transitive closures and provides an important research topic in the near future.

CONCLUSION

In this article, we provide an overview on the recursion computation in a relational environment and present a new encoding method to label a digraph, which is compared with a variety of traditional strategies as well as the methods proposed in the database community. Our method is based on a tree labeling method and the concept of branchings that are used in graph theory for finding the shortest connection networks. A branching is a subgraph of a given digraph that is in fact a forest, but covers all the nodes of the graph. On the one hand, the proposed encoding scheme achieves the smallest space requirements among all previously published strategies for recognizing recursive relationships. On the other hand, it leads to a new algorithm for computing transitive closures for DAGs in $O(e \cdot b)$ time and $O(n \cdot b)$ space, where *n* represents the number of the nodes of a DAG, *e* the numbers of the edges, and *b* the DAG's breadth. In addition, this method can be extended to cyclic digraphs and is especially suitable for a relational environment.

REFERENCES

Abdeddaim, S. (1997). On incremental computation of transitive closure and greedy alignment. In A. Apostolico & J. Hein (Eds.), *Proceedings of 8th Symp. Combinatorial Pattern Matching* (pp. 167-179).

Abiteboul, S., Cluet, S., Christophides, V., Milo, T., Moerkotte, G., & Simon, J. (1997, April). Querying documents in object databases. *International Journal of Digital Libraries, 1*(1), 5-19.

Agrawal, R., Borgida, A., & Jagadish, J.V. (1989, June). Efficient management of transitive relationships in large data and knowledge bases. *Proceedings of the ACM SIGMOD Intl. Conf. on the Management of Data* (pp. 253-262).

Booth, K.S., & Leuker, G.S. (1976, December). Testing for the consecutive ones property, interval graphs, and graph palanity using PQ-tree algorithms. *Journal of Computer Sys. Sci., 13*(3), 335-379.

Cattell, R.G.G., & Skeen, J. (1992). Object operations benchmark. *ACM Trans. Database Systems, 17*(1), 1 -31.

Chen, Y. (2003, May). On the graph traversal and linear binary-chain programs. *IEEE Transactions on Knowledge and Data Engineering, 15*(3), 573-596.

Chen, Y., & Aberer, K. (1998). Layered index structures in document database systems. *Proceedings of 7th Int. Conference on Information and Knowledge Management (CIKM)*, Bethesda, MD (pp. 406-413). ACM.

Chen, Y., & Aberer, K. (1999, September). Combining pat-trees and signature files for query evaluation in document databases. *Proceedings of 10th Int. DEXA Conf. on Database and Expert Systems Application*, Florence, Italy (pp. 473–484). Springer Verlag.

Cohen, N.H. (1991). Type-extension tests can be performed in constant time. *ACM Transactions on Programming Languages and Systems, 13,* 626-629.

Fall, A. (1995). Sparse term encoding for dynamical taxonomies. *Proceedings of 4th International Conf. On Conceptual Structures (ICCS-96): Knowledge Representation as Interlingua,* Berlin (pp. 277-292).

Knuth, D.E. (1969). *The art of computer programming* (vol. 1). Reading, MA: Addison-Wesley.

Krall, A., Vitek, J., & Horspool, R.N. (1997). Near optimal hierarchical encoding of types. In M. Aksit & S. Matsuoka (Eds.), *Proceedings of 11th European Conf. on Object-Oriented Programming,* Jyvaskyla, Finland (pp. 128-145).

Kuno, H.A., & Rundensteiner, E.A. (1998). Incremental maintenance of materialized object-oriented views in MultiView: Strategies and performance evaluation. *IEEE Transactions on Knowledge and Data Engineering, 10*(5), 768-792.

La Poutre, J.A., & van Leeuwen, J. (1988). Maintenance of transitive closure and transitive reduction of graphs. *Proceedings of Workshop on Graph-Theoretic Concepts in Computer Science*, *Lecture Notes in Computer Science, 314,* 106-120. Springer-Verlag.

Lee, W.C., & Lee, D.L. (1998). Path dictionary: A new access method for query processing in object-oriented databases. *IEEE Transactions on Knowledge and Data Engineering, 10*(3), 371-388.

Mendelzon, A.O, Mihaila, G.A., & Milo, T. (1997, April). Querying the World Wide Web. *International Journal of Digital Libraries, 1*(1), 54-67.

Ramakrishnan, R., & Ullman, J.D. (1995, May). A survey of research in deductive database systems. *Journal of Logic Programming*, 125-149.

Schmitz, L. (1983). An improved transitive closure algorithm. *Computing, 30,* 359 - 371.

Stonebraker, M., Rowe, L., & Hirohama, M. (1990). The implementation of POSTGRES. *IEEE Trans. Knowledge and Data Eng., 2*(1), 125-142.

Tarjan, R. (1972, June). Depth-first search and linear graph algorithms. *SIAM J. Compt., 1*(2), 146-140.

van Bommel, M.F., & Beck, T.J. (2000). Incremental encoding of multiple inheritance hierarchies supporting lattice operations. *Linkoping Electronic Articles in Computer and Information Science, http://www.ep.liu.se/ea/cis/2000/001*

Zhang, C., Naughton, J., DeWitt, D., Luo, Q., &. Lohman, G. (2001). On supporting containment queries in relational database management systems. *Proceedings of ACM SIGMOD Intl. Conf. on Management of Data*, California.

Zibin, Y., & Gil, J. (2001, October 14-18). Efficient subtyping tests with PQ-encoding. *Proceedings of the 2001 ACM SIGPLAN Conf. on Object-Oriented Programming Systems, Languages and Application,* Florida (pp. 96-107).

KEY TERMS

Branching: A branching is a subgraph of a directed graph, in which there are no cycles and the indegree of each node is 1 or 0.

Cyclic Graph: A cyclic graph is a directed graph that contains at least one cycle.

DAG: A DAG is a directed graph that does not contain a cycle.

Graph Encoding: Graph encoding is a method to assign the nodes of a directed graph a number or a bit string, which reflects some properties of that graph and can be used to facilitate computation.

Strongly Connected Component (SCC): An SCC is a subgraph of a directed graph, in which between each pair of nodes there exists a path.

Transitive Closure: The transitive closure of a directed graph G is a graph G^*, in which there is an edge from node a to node b if there exists a path from a to b in G.

Tree: A tree is a graph with a root, in which the indegree of each node is equal to 1.

Harmonizing IT and Business Strategies

Kevin Johnston
University of Cape Town, South Africa

INTRODUCTION

The alignment of business strategy with IT strategy has been a concern of chief information officers (CIOs) (Berkman, 2000; Croteau & Bergeron, 2001; Crowley, 2001), chief executive officers (CEOs) (Armstrong, Chamberlain, Moore & Hart, 2002; Mesoy, 1999), academic researchers (Henderson & Venkatraman, 1999; Reich & Benbasat, 2000; Tallon & Kraemer, 2000), and research companies (Broadbent, 2000; Croteau & Bergeron, 2001; Meta Group, 2001) since the age of vacuum tubes. In surveys (Mesoy, 1999) of CIO concerns, alignment has consistently been rated as a major issue. A Cutter study reported that business-IT alignment was "the number one problem facing IT" (Crowley, 2001).

Many authors (Burgleman, Maidique & Wheelwright, 2001; Croteau & Bergeron, 2001; Hartman, Sifonis, & Kador, 2000; Reich & Benbasat, 2000) agree that it is important to align IT strategy with the organization's business strategy. Although the importance of strategic alignment of IT is acknowledged and widely accepted, it remains an issue within many organizations (Armstrong et al., 2002).

Burgleman et al. (2001) stated, "Technological issues only occasionally are included explicitly in typical corporate strategy reviews, and only rarely are they among the regular inputs to corporate planning and development." Gates (1999) wrote: "It is impossible to align IT strategy with business strategy if the CIO is out of the business loop."

Management's attitude towards IT and IT strategy need to be changed (Cosgrove Ware, 2001). A "model of attitude change", described by Nash, Gwilt, Ludwig, and Shaw (2001), listed five steps to affect attitude change:

- Attention and awareness - subject managers to advertising, publications, and word of mouth about IT;
- Comprehension and knowledge - teach managers about IT and its uses;
- Yielding - managers try IT;
- Use of IT - managers use IT; and finally,
- Reinvention - where IT is improved or its use is varied for maximum benefit.

The lack of IT alignment with business can result in late market entry, lost market opportunities, or an unsustainable market advantage (Conarty, 1998).

Some (Bocij, Chaffey, Greasley & Hickie, 1999) view IT strategy and business strategy as two distinct strategies, with IT strategy either supporting or influencing business strategy. Other authors, such as Pukszta (1999), stress that IT strategy must be completely and seamlessly integrated with business strategy at all organizational levels.

IT and business strategies should not be aligned but should, in fact, be one harmonious strategy. IT strategy has to lose its distinctness; in this way, it will gain prominence and exert greater influence within organizations (Pukszta, 1999). Each organization should have a single harmonized strategy.

IT planning should be in harmony, not merely aligned with business strategy.

BACKGROUND

Organizations need to recognize changing business climates, fluctuating resources, and the need to expand or grow (Heske, 2001). Organizations that plan and then move in the right direction at the right time survive. All factors including IT have to be considered and taken into account holistically.

Key tasks of managers within an organization are to acquire, develop and allocate organizations' resources, and to develop and exploit the organizations' capacity for innovation (Burgleman et al., 2001). The acquisition, development, allocation and exploitation of IT should be part of any business strategy. Many new products and services such as online banking have been based on IT (Laudon & Laudon, 2002). IT can contribute to the overall performance of the organization by tying operations of various business units together harmoniously so the organization can act as whole (Laudon & Laudon, 2002). This could lower costs, increase customer access, and speed up the marketing process of new products and services. It is unlikely that these contributions and possible innovations will occur by chance; they need to be planned.

Tallon and Kraemer (2000) found that organizations with focused IT goals achieved higher payoffs from IT. Their results indicated that unfocused organizations achieved consistently lower payoffs at each point along the value chain than focused organizations.

Linear planning is useless; organizations must plan holistically (Hartman et al., 2000) or harmoniously in order to survive.

WHERE, WHY, WHAT, HOW?

Where should an organization do business ": locally, nationally or internationally? Organizations should be thinking in terms of where it makes good business sense and where the organization will survive. A vision of where the organization is going needs to be developed and communicated, and everyone in the organization needs to understand the vision and implement it (Rollins, Bognanno & Lockwood, n.d.).

A business strategy needs to be defined, including all the capabilities (forces/tools/resources) of an organization, so that approved plans may be executed as effectively as possible (Henderson & Venkatraman, 1999). Strategy articulates ways in which opportunities can be exploited using the organizations' capabilities (Burgleman et al., 2001). Strategy without capabilities is meaningless (Burgleman et al., 2001), and excluding the IT capability from the organization's strategy renders the strategy less effective at best. Similarly, having capabilities without strategy makes them aimless (Burgleman et al., 2001).

The IT strategy and capability must, therefore, be part of the overall strategy, or IT will become an aimless capability of the organization and, at best, will be run according to the CIO's aims. Managing the IT resource is a basic business function (Burgleman et al., 2001), which should be the responsibility of all managers within an organization.

Organizations plan in order to anticipate change beyond the control of the organization, so changes within the organization (such as changes to business processes and organizational structure) can be initiated and controlled (Ivancevich & Matteson, 1999).

A number of elements must fit together in a balanced way in order for an organization to function effectively. Sawy (2001) uses the Leavitt Diamond framework to illustrate the balance. This framework has four sets of organizational variables: IT use, organizational form, people skills, and business processes. When any one of these is changed, the other three need to be adjusted to maintain "functional harmony" (Sawy, 2001).

Ward and Peppard (2002) suggest guidelines to align business and IT strategies. They state that, at the very least, the following four domains need to be aligned: business strategy, organizational infrastructure, IT strategy, and finally, IT infrastructure. The business strategy should decide where the business is headed and why, what the businesses competitive advantages are, as well as how the business will be governed. The organizational infrastructure and business processes need to be designed to support the business strategy. The IT strategy should decide what IT is important and required by the business, and how it will used and managed or governed. The IT infrastructure details how IT can be delivered •the architecture, processes and skills required. Strategy needs to be formulated concurrently in all four of these domains (Ward & Peppard, 2002).

Organizations need to define exactly where they are going, why they wish to go there, what resources they will use to get there, and how they will utilize those resources.. Management needs to ensure that all stakeholders are aware of the plan.

STRUCTURE

Organizations need to decide on internal structures that support the plan. Structure and order are necessary for the survival of the organization.

"A gap has developed between the power and choice enjoyed by individuals as consumers and citizens on the one hand, and that available to them in the workspace on the other" (Chowdhury, 2000). This gap has to be reduced. Employees will have to be included in decisions regarding the structure of organizations.

Organizational design and structure have always been important factors that influence the behavior of groups and individuals. It is through structure that management establishes expectations of achievements for individual employees and departments, and decides how the organization's strategy is to be accomplished. The purpose of structure is to regulate, or reduce, uncertainty in the behavior of employees (Ivancevich & Matteson, 1999). Where and how IT is placed in the organizational structure determine the role and influence of IT.

Each organization requires a structure, and when that is ignored, the organization will not be able to crawl, much less fly! No organization will last if everyone acts independently; a structure needs to be developed preferably with the employees who will support the strategy and vision of the organization.

Employees cannot survive alone within an organization; they need to be in some sort of formation. Organizations only "fly" when all the employees are in formation.

Organizations need tradition, ritual, and structure to retain their identity. A department in which the author worked had a daily meeting where employees met each other and offered encouragement, support, guidance and feedback. The meetings were stand-up, 15-30 minute affairs with a fixed, regular program. On Mondays, projects and work for the week ahead were discussed; Tuesdays were "thinking" days, and employees had to solve puzzles in groups; Wednesdays were "learning" days, where one employee had to teach the others something; Thursdays were to announce and discuss change; and Fridays were for external focus or external speakers. Each employee had an opportunity to lead a Tuesday and Wednesday meeting. Rollins et al. (n.d.) claim that 50% of all employees receive no formal feedback on their performance, and of the 50% who do receive feedback, only 20% believe that it is effective. In groups where there is encouragement, the production is much greater.

Organizations need to fly with employees who want to be in the organization, who know where the organization is headed, and who want to fly in formation.

TIMING

Organizations understand that they cannot change or fight the seasons. Organizations need to develop strategies anticipating "winter" and unseasonable changes. Organizations must respond to external demands quickly, adapt their organizations, and redirect their employees and IT to achieve competitive advantage (Rollins et al., n.d.).

Organizations need to understand that IT change is one external factor, which affects competitive position. In order to survive, organizations need to adapt timely to change. Therefore, strategic planning must be a dynamic process, and IT and IT change must form part of the process. Successful organizations balance a well-defined business focus with the willingness and the will to undertake major and rapid change.

Organizations need to be ready to respond to continual changes. Prerequisites are leadership, governance, competencies and technology (Hartman et al., 2000). The first prerequisite is leadership; outstanding companies are associated with their leaders (Welch, Gates & Bezos). Leaders create a vision that is shared and accepted within the organization. Governance is the operating model that defines the organization. The formation or structure of the organization must be clear. People roles, responsibilities, and authority levels must be defined. Organizations need to have methods for assessing, selecting, allocating and monitoring resources. Competencies are the ways in which the organization responds to change, exploits available resources and opportunities, and accommodates reality. Technology needs to be robust and comprehensive.

"The reality of a strategy lies in its enactment, not in those pronouncements that appear to assert it" (Burgleman et al., 2001). Strategic intent needs to be converted into strategic action to be meaningful.

Organizations need strategies to deal with obstacles and may, in some cases, need to revise their strategies in order to be successful.

SUCCESS

The organization has to expect some casualties, but if the organization survives and looks likely to continue surviving, the strategy is successful. To succeed, the energy, creativity and resources of the organization must be used. If there are parts of the organization that are not being used, the question must arise as to why they are in the organization.

Burgelman et al. (2001) ask and answer the question, "What strategies, policies, practices, and decisions result in successful management of high-technology enterprises?" Six themes of success are listed:

- business focus,
- adaptability,
- organizational cohesion,
- entrepreneurial culture,
- sense of integrity, and
- hands-on top management.

All are controlled or influenced by the organizational strategy. Unless each and every one of the organization's resources is in harmony, an organization cannot succeed. If IT strategy (and all other strategies) is harmonized into a single business strategy, then the organization can claim to have a holistic business focus, to have organizational cohesion, and to have a sense of integrity. A disharmonized strategy certainly cannot be regarded as being honest, fair or open, or having other attributes of integrity.

FUTURE TRENDS

Organizations will develop single evolving harmonized strategies, to which all employees have contributed. Organizational strategy will be a continual, changing, probing neverending cycle involving all in the organization.

CONCLUSION

Organizational success must include the ability to harmonize the organization and mobilize the work force (Reich & Benbasat, 2000). Organizations and individuals can only realize their potential for greatness and goodness when they join the flock, fly in formation, and contribute something for the common good (Nash et al., 2001; Sawy, 2001). Business and IT strategies will be in harmony when IT is seen as contributing positively to the organizations business strategy. IT strategy must form a harmonized part of the organizations overall strategy (Ward & Peppard, 2002).

REFERENCES

Armstrong, T., Chamberlain, G., Moore, B., & Hart, M. (2002). *Key information systems management issues for CEOs and other executives in South Africa 2002*. Unpublished Honours Empirical Research. University of Cape Town.

Berkman, E. (2000). Why we're still talking about alignment. Retrieved March 2002, from *http://www.cio.com/archive/010101_talking.html*

Bocij, P., Chaffey, D., Greasley, A., & Hickie, S. (1999). *Business information systems, technology, development and management*. Harlow: Pearson Education.

Broadbent, M. (2000). Today's CIO energizes, enables, executes and exploits. Retrieved March 2002, from: *http://www4.gartner.com/UnrecognizedUserHomePage.jsp*

Burgleman, R.A., Maidique, M.A., & Wheelwright, S.C. (2001). *Strategic management of technology and innovation*. Singapore: McGraw-Hill.

Chowdhury, S. (Ed.). (2000). *Management 21C*. London: Prentice Hall.

Conarty, T.J. (1998). Alignment for success: Information technology and business strategy. Presentation at IISI-32. Retrieved March 2002, from *http://www.worldsteel.org/events/proceed/IISI-32_1998/PR_conarty1.html*

Cosgrove Ware, L. (2001). Measuring IT alignment. Retrieved March 2002, from *http://www.cio.com/archive/040101/tl_numbers.html*

Croteau, A., & Bergeron, F. (2001, June). An information technology trilogy: Business strategy, technological deployment and organisational performance. *Journal of Strategic Information Systems*, *10*(2), 77-99.

Crowley, D. (2001, September 18). Business-IT alignment – A turn for the better. *The Cutter Edge*. Cutter Information Corp.

Gates, B. (1999). *Business @ the speed of thought*. London: Penguin.

Hartman, A., Sifonis, J., & Kador, J. (2000). *Net ready*. New York: McGraw-Hill.

Henderson, J.C., & Venkatraman, N. (1999). Strategic alignment: Leveraging information technology for transforming organizations. *IBM Systems Journal*, *38*(2-3), 472-484.

Heske, P. (2001, September 17). Leaders, lemmings or laggards. *Computing SA*, 10.

Ivancevich, J.M., & Matteson, M.T. (1999). *Organisational behaviour and management*. Singapore: McGraw-Hill.

Laudon, K.C., & Laudon, J.P. (2002). *Management information systems*. Upper Saddle River, NJ: Prentice Hall.

Mesoy, T. (1999). Transforming IT to position the IT function as a valued and respected business partner. *IT Alliance Forum*, Sun City, South Africa.

Meta Group. (2001). Top CIO issues for 2001. *http://www.metagroup.com/cgi-bin/inetcgi/search/displayArticle.jsp?oid=23211*

Nash, J., Gwilt, D., Ludwig, A., & Shaw, K. (2001). *The use of technology to support decision-making in South Africa*. Unpublished. University of Cape Town.

Pukszta, H. (1999, January 11). Don't split IT strategy from business strategy. *Computerworld*, *33*(2), 35.

Reich, B., & Benbasat, I. (2000, March). Factors that influence the social dimension of alignment between business and information technology objectives. MIS Quarterly, 24(1), 81-113.

Rollins, T., Bognanno, M.A., & Lockwood, M. (n.d.). The strategy-focused workforce: Using the Balanced Scorecard to align people with strategy. Retrieved March 2002, from *http://rm.haygroup.com/balanced.htm*

Sawy, O. (2001). *Redesigning enterprise processes for e-business*. Singapore: McGraw-Hill.

Tallon, P.P., & Kraemer, K.L. (2000). *Executives' perspectives on IT: Unravelling the link between business strategy, management practices and IT business value*. Center for Research on Information Technology and Organizations, University of California.

Ward, J., & Peppard, J. (2002). *Strategic planning for information systems*. England: John Wiley & Sons.

KEY TERMS

Alignment: The arrangement or position of different separate elements (strategies) in relation to each other.

Business Process: A collection of business activities which take several inputs and create one or more outputs.

Business Strategy: A description of the plans, actions or steps an organization intends to take in order to strengthen and grow itself.

CIO: Chief Information Officer, the head of the IS department in an organization.

Harmony: A pleasing combination of elements in a whole. The combination of elements intended to form a connected whole, as opposed to alignment where the elements remain separate.

IT: Information technology, the collection of all systems in an organization

IT Strategy: A description of the plans, actions or steps an organization intends to take in order to make the best use of IT within itself.

Strategy: A plan to create and manage change, and exploit opportunities.

Heuristics in Medical Data Mining

Susan E. George
University of South Australia, Australia

HISTORICAL PERSPECTIVE

Deriving—or discovering—information from data has come to be known as data mining. Within health care, the knowledge from medical mining has been used in tasks as diverse as patient diagnosis (Brameier et al., 2000; Mani et al., 1999; Cao et al., 1998; Henson et al., 1996), inventory stock control (Bansal et al., 2000), and intelligent interfaces for patient record systems (George at al., 2000). It has also been a tool of medical discovery itself (Steven et al., 1996). Yet, it remains true that medicine is one of the last areas of society to be "automated," with a relatively recent increase in the volume of electronic data, many paper-based clinical record systems in use, a lack of standardisation (for example, among coding schemes), and still some reluctance among health-care providers to use computer technology. Nevertheless, the rapidly increasing volume of electronic medical data is perhaps one of the domain's current distinguishing characteristics, as one of the last components of society to be "automated."

Data mining presents many challenges, as "knowledge" is automatically extracted from data sets, especially when data are complex in nature, with many hundreds of variables and relationships among those variables that vary in time, space, or both, often with a measure of uncertainty, as is common within medicine. Cios and Moore (2001) identified a number of unique features of medical data mining, including the use of imaging and need for visualisation techniques, the large amounts of unstructured nature of free text within records, data ownership and the distributed nature of data, the legal implications for medical providers, the privacy and security concerns of patients requiring anonymous data used, where possible, together with the difficulty in making a mathematical characterisation of the domain.

Strictly speaking, many ventures within medical data mining are better described as exercises in "machine learning," where the main issues are, for example, discovering the complexity of relationships among data items, or making predictions in light of uncertainty, rather than "data mining," in large, possibly distributed, volumes of data that are also highly complex. Large data sets mean not only increased algorithmic complexity but also often the need to employ special-purpose methods to isolate trends and extract "knowledge" from data. However, medical data frequently provide just such a combination of vast (often distributed) complex data sets.

Heuristic methods are one way in which the vastness, complexity, and uncertainty of data may be addressed in the mining process. A heuristic is something that aids discovery of a solution. Artificial intelligence (AI) popularised the heuristic as something that captures, in a computational way, the knowledge that people use to solve everyday problems. AI has a classic graph search algorithm known as A* (Hart et al., 1968), which is a heuristic search (under the right conditions). Increasingly, heuristics refer to techniques that are inspired by nature, biology, and physics. The genetic search algorithm (Holland, 1975) may be regarded as a heuristic technique. More recent population-based approaches have been demonstrated in the Memetic Algorithm (Moscato, 1989), and specific modifications of such heuristic methods in a medical mining context can be noted (Brameier et al., 2000).

Aside from the complexity of data with which the medical domain is faced, there are some additional challenges. Data security, accuracy, and privacy are issues within many domains, not just the medical (Walhstrom et al., 2000). Also, while ethical responsibility is an issue in other contexts, it is faced by the medical world in a unique way, especially when heuristic methods are employed. One of the biggest ethical issues concerns what is done with the knowledge derived combined with a "forward-looking responsibility" (Johnson et al., 1995). Forward-looking responsibility is accountable for high-quality products and methods and requires appropriate evaluation of results and justification of conclusions.

George (2002) first identified and proposed a set of guidelines for heuristic data mining within medical domains. The proposed guidelines relate to the evaluation and justification of data-mining results (so important when heuristic "aids to discovery" are utilised that "may" benefit a solution) and extend to both where and how the conclusions may be utilised and where heuristic techniques are relevant in this field. The remainder of this article summarises some heuristic data-mining applications in medicine and clarifies those proposed guidelines.

BACKGROUND

First, we will explain some of the heuristic methods that have been employed in medical data mining, examining a range of application areas. We broadly categorise applications as clinical, administrative, and research, according to whether they are used (or potentially used) in a clinical context, are infrastructure related, or are exploratory, in essence. We also note that with the exception of some medical imaging applications and mining of electronic medical records, the databases are small.

There is a wide variety of automated systems that have been designed for diagnosis—systems that detect a problem, classify it, and monitor change. Brameier and Banzhauf (2000) described the application of linear genetic programming to several diagnosis problems in medicine, including tests for cancer, diabetes, heart conditions, and thyroid conditions. Their focus was upon an efficient algorithm that operates with a range of complex data sets, providing a population-based heuristic method that is based upon biological principles. Their heuristic method is based on an inspiration from nature about how "introns" (denoting DNA segments with information removed before proteins are synthesised) are used in generating new strings. They suggest that introns may help to reduce the number of destructive recombinations between chromosomes by protecting the advantageous building blocks from being destroyed by crossover. Massive efficiency improvements in the algorithm are reported.

An interesting administrative application of data mining in a medical context comes in the area of interfaces for electronic medical records systems that are appropriate for speedy, accurate, complete entry of clinical data. At the University of South Australia, George et al. (2000) reported on the use of a data-mining model underlying an adaptive interface for clinical data entry. As records are entered, a database is established from which predictive Bayesian models are derived from the diagnosis and treatment patterns. This heuristic is used to predict the treatment from new diagnoses that are entered, producing intelligent anticipation. The predictive model is also potentially incremental and may be re-derived according to physician practice. This application addresses issues in incremental mining, temporal data, and highly complex data with duplication, error, and nonstandard nomenclatures.

One interesting ongoing database mining project at Rutgers is the development of efficient algorithms for query-based rule induction, where users have tools to query, store, and manage rules generated from data. An important component of the research is a facility to remember past mining sessions, producing an incremental approach. They are using heuristics for efficiently "re-mining" the same or similar data in the face of updates and modifications. In their trials, a major insurance company was trying to explore anomalies in their medical claims database. The new data-mining techniques aided the isolation of high-cost claims and scenarios in each disease group that would lead to high-cost claims. They also identified characteristics of people who were likely to drop out of their health plans and locations where there were higher dropout rates. This is a general approach to mining, where information from prior mining is utilised in new mining to prevent the need to compute relationships from scratch every time data is added to the database. This is, naturally, a general approach to mining large-scale changing databases that may be considered in a variety of fields.

Medical data mining is a natural method of performing medical research, where new relationships and insights are discovered in human health. The University of Aberdeen address the problem of mammographic image analysis using neural nets together with conventional image analysis techniques to assist in the automated recognition of pathology in mammograms (Undrill, 1996). The group also addresses the use of genetic algorithms for image analysis, applying this powerful general optimisation technique to a variety of problems in texture segmentation and shape analysis in two-dimensional and three-dimensional images (Delibassis, 1996). Mining information from the data in these tasks must address many of the problems of finding patterns within large volumes of highly complex data.

Banerjee et al. (1998) described the use of data mining in medical discovery. They reported on a data-mining tool that uncovered some important connections between diseases from mining medical literature. The data-mining tool compared the article titles in various medical journals. Medical discoveries were made, such as the connection between estrogen and Alzheimer's disease, and the relationship between migraine headaches and magnesium deficiency. Ngan et al. (1999) reported on medical discovery using data mining based upon an evolutionary computation search for learning Bayesian networks and rules. They were able to discover new information regarding the classification and treatment of scoliosis as well as knowledge about the effect of age on fracture, diagnoses, and operations and length of hospital stays.

Kargupta and colleagues (1999) were interested in an epidemiological study that involved combining data from distributed sources. Their study investigated what affects the incidence of disease in a population, focusing upon hepatitis and weather. They illustrated the collective data-mining approach, emphasising the importance within medicine of merging data from heterogeneous sites. Their solution minimises data communication using decision-tree learning and polynomial regression. As more hospitals and general practitioners, pharmacists, and other health-care-related professions utilise electronic media, mining ventures are going to have to cope with mining across data sources. They will have to address issues such as those addressed by this study, such as minimising data exchange and adopting suitable heuristic approaches.

GUIDELINES FOR HEURISTIC MEDICAL DATA MINING

Responsibility is clearly an issue in medical data mining given the unique human arena in which the conclusions are outworked. If medical data-mining products are ever produced by "professionals" or are ever exploited "commercially," there may be serious legal consequences for their creators in the wake of harmful consequences from information produced. In the context of software engineering, the computer field seeks to promote high-quality software products, so too, should data miners seek to guarantee high-quality data-mining techniques.

Johnson and Nissenbaum (1995) distinguished "backward-looking" responsibility from "forward-looking" responsibility. A "backward-looking" responsibility asks questions in the wake of a harmful event and seeks to discover who is to blame for the harm and who should be punished. It is often conducted in the context of discovering legal liability. The Therac-25 computer-controlled radiation treatment is a shocking example of a malfunction disaster that resulted in loss of life for people who were receiving computer-controlled radiation treatments for cancer. In contrast a "forward-looking" responsibility addresses the particular responsibilities in advance. It defines guidelines for creating quality products; measures the quality of the product; defines the method of evaluation, and the limitations and scope of the operation in advance of harmful incidents.

One of the biggest ethical issues in medical mining concerns what is done with the knowledge derived. There is tremendous potential for good in improving quality of human life, managing disease effectively, efficiently administering programs, and preserving life, but the same knowledge can also be put to less-constructive ends, or benefit only a few, or conform to the contemporary political agendas influenced by the philosophy of the age. Forward-looking responsibility requires not only making ethical uses of data but also ensuring the quality of automated techniques and knowing the limitations and scope of methods in advance of harmful consequences.

Crucial to forward-looking responsibility is a way to evaluate products. This is not as pertinent as when heuristic methods are utilised to derive that knowledge. Whatever is ultimately done with the conclusions, we know that heuristics do not guarantee "optimality" or even the "accuracy" or "validity" of the conclusion. Forward-looking responsibility within medical data mining will address, among other things, how knowledge is evaluated, how conclusions are justified, what is the scope of validity, and the limitations of "products." One of the best forms of evaluation for clinical data-mining solutions is a clinical trial. Another approach to evaluation makes use of benchmark data sets, where various techniques (heuristic and other) could be compared to assess quality and efficiency of solutions. Additionally, some types of specialist data may be invaluable resources for data-mining researchers. It is also important to define the scope, including justification of explanations, and limitations of systems, from technical to clinical applicability of algorithms (in terms of patient populations and other), especially under heuristic conditions.

CONCLUSION

This article has reviewed heuristic medical data mining and some of the applications of medical mining, identifying administrative, clinical, and medical areas of applicability, focusing on the guidelines for use of heuristics in such a field. Forward-looking responsibility is vital, focusing upon appropriate use of knowledge, a means to evaluate the heuristic solutions and assess the scope and limitations of the system, including explanations of the behaviour.

REFERENCES

Bansal, K., Vadhavkar, S., & Gupta, A. (2000). Neural networks based data mining applications for medical inventory problems. Retrieved September 21, 2000, from *http://scanner-group.mit.edu/htdocs/DATAMINING/Papers/paper.html*

Brameier, M., & Banzhaf, W. (2001). A comparison of linear genetic programming and neural networks in medical data mining. *IEEE Transactions on Evolutionary Computation*, 5(1), 17-26. Retrieved September 22, 2000, from *http://ls11-www.cs.uni-dortmund.de/people/banzhaf/ieee_taec.pdf*

Cao, C., Leong, T. Y., Leong, A. P. K., & Seow, F. C. (1998). Dynamic decision analysis in medicine: A data driven approach. *International Journal of Medical Informatics*, *51*(1), 13–28.

Cios, K., & Moore. (2001). Medical data mining and knowledge discovery: Overview of key issues. In K. Cios (Ed.), *Medical data mining and knowledge discovery*. Heidelberg: Springer-Verlag.

Delibassis, K., & Undrill, P. E. (1996). Genetic algorithm implementation of stack filter design for image restoration. *IEE Proc. Vision, Image & Signal Processing, 143*(3), 177-183.

George, S. E. (2002). Heuristics and medical datamining (Chap. 13). In H. A. Abbass, R. A. Sarker, & C. S. Newton (Eds.), *Heuristics and optimisation for knowledge discovery* (pp. 226-240). Hershey, PA: Idea Group Publishing.

George, S. E., & Warren, J. R. (2000). Statistical modelling of general practice medicine for computer assisted data entry in electronic medical record systems. *International Journal of Medical Informatics, 57*(2–3), 77-89.

Hart, P. E., Nilsson, N. & Raphael, B.(1968). A formal basis for the heuristic determination of minimum cost paths. *IEE Transactions on SSC, 4*, 100-107.

Holland, J. (1975). *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control and artificial intelligence*. Ann Arbor, MI: University of Michigan Press.

Johnson, D., & Nissenbaum, H. (1995). *Computers, ethics and social values*. Englewood Cliffs, NJ: Prentice Hall.

Kargupta, H., Park, B., Hershberger, D., & Johnson, E. (1999). Collective data mining: A new perspective toward distributed data mining. In H. Kargupta & P. Chan (Eds.), *Advances in distributed data mining*. Cambridge, MA: AAAI/MIT Press.

Mani, S., Shankle, W., Dick, M., & Pazzani, M. (1999). Two-stage machine learning model for guideline development. *Artificial Intelligence in Medicine, 16*, 51-71.

Moscato, P. (1989). On evolution, search, optimization, genetic algorithms and martial arts: Towards Memetic algorithms. Caltech Concurrent Computation Program, C3P Report 826.

Ngan, P. S., Wong, M. L., Lam, W., Leung, K. S., & Cheng, J. C. Y. (1999). Medical data mining using evolutionary computation. *Artificial Intelligence in Medicine, 16*, 73-96.

Undrill, P. E., Gupta, R., Henry, S., Downing, M., & Cameron, G. G. (1996). Outlining suspicious lesions in mammography by texture focussing and boundary refinement. In M.H. Loew & K.M. Hanson (Eds.) *Proceedings SPIE Medical Imaging: Image Processing*, 2710 (pp. 301-310).SPIE, Newport Beach, CA.

Vajdic, S. M., Brooks, M. J., Downing, A., & Katz, H. E. (1996). AI and medical imagery: Strategy and evaluation of inexact relational matching. Retrieved September 21, 2000, from *http://www.csu.edu.au/ci/vol2/vajdic/vajdic.html*

Walhstrom, K., Roddick, J. F., & Sarre, R. (2000). *On the ethics of data mining*. Research Report ACRC-00-003, January 2000, School of Computer and Information Science, University of South Australia.

H

KEY TERMS

Backward-Looking Responsibility: When backward-looking, we seek to discover who is to blame in wake of a harmful event. There are frequently connotations of punishment, legal intervention, and determination of guilt.

Data Mining: Analysis of data using methods that look for patterns in the data, frequently operating without knowledge of the meaning of the data. Typically, the term is applied to exploration of large-scale databases in contrast to machine-learning methods that are applied to smaller data sets.

Data-Mining Guidelines: A set of standards by which medical data mining, in particular, might be conducted. This is a framework that adopts a forward-looking responsibility in the evaluation of methods and explanation of conclusions, especially in the context of heuristic methods (with outcomes that may be ill-defined). This extends not only to the methods of the data-mining procedure, the security and privacy aspects of data, but also to where and how the results of data mining are utilised, requiring that an ethical reference be made to the final purpose of the mining.

Forward-Looking Responsibility: Addresses the particular responsibilities of individuals, groups, and partners in advance of a product's use or a system's implementation; it defines guidelines for creating quality products, measures the quality of the products, and defines the method of evaluation, the limitations, and the scope of the operation in advance of harmful incidents.

Heuristic: From the Greek "heuriskein," meaning "to discover." A heuristic aids discovery, particularly the search for solutions in domains that are difficult and poorly understood. It is commonly known as a "rule of thumb." Unlike algorithms, heuristics do not guarantee optimal or even feasible solutions and frequently do not have a theoretical guarantee.

Medical data: Frequently demonstrate increased complexity (e.g., uncertainty) and occurrence in large volumes, possibly distributed over many sources. There may be images involved and a high frequency of nonstandardisation (especially in the context of electronic medical records) in coding schemes and other medical concepts utilised. Data security, accuracy, and privacy are particular issues with medical data, as are the ethical issues involved in what is done with the data.

Medical Data Mining: This is the application of data-mining methods to medical data, typically for clinical or administrative and medical research investigation use, particularly in epidemiological studies.

Hierarchies in Multidimensional Databases

Elaheh Pourabbas
National Research Council, Rome, Italy

INTRODUCTION

Hierarchies play a fundamental role in knowledge representation and reasoning. They have been considered as the structures created by *abstraction processes*. According to Smith and Smith (1977), an abstraction process is an instinctively known human activity, and abstraction processes and their properties are generally used for multilevel object representation in information systems. An abstraction can be understood as a selection of a set of attributes, objects, or actions from a much larger set of attributes, objects, or actions according to certain criteria. Repeating this selection several times, that is, continuing to choose from each subset of objects, another subset of objects with even more abstract properties, we create other levels of (semantic) details of objects. The complete structure created by the abstraction process is a *hierarchy* and the type of hierarchy depends on the operation used for the abstraction process and the relations. As for the relations, the best known in the literature are *classification*, *generalization*, *association* (or *grouping*), and *aggregation*. Their main characteristics are briefly listed in the following.

Classification is a simple form of data abstraction in which an object type is defined as a set of instances. It introduces *an instance-of* relationship between an object type in a schema and its instances in the database (Brodie, Mylopoulos, & Schmidt, 1984).

Generalization is a form of abstraction in which similar objects are related to a higher level generic object. It forms a new concept by leaving out the properties of an existing concept. With such an abstraction, the similar constituent objects are specializations of the generic objects. At the level of the generic object, the similarities of the specializations are emphasized, while the differences are suppressed (Brodie, et al., 1984).

This introduces an *is-a* relationship between objects. This relation covers a wide range of categories that are used in other frameworks, such as inheritance, implication, and inclusion. It is the most frequent relation resulting from subdividing concepts, called *taxonomies* in lexical semantics. The inverse of the generalization relation, called *specialization*, forms a new concept by adding properties to an existing concept (Borgida, Mylopoulos, & Wong, 1984).

A particular type of generalization hierarchy, named *filter hierarchy*, is defined by the so-called filtering operation. This operation applies a *filter function* to a set of objects on one level and generates a subset of these objects on a higher level. The main difference from the generalization hierarchy is that the objects that do not pass the filter will be suppressed at the higher level (Timpf, 1999).

Association or grouping is a form of abstraction in which a relationship between member objects is considered as a higher level set of objects. With this relationship, the details of member objects are suppressed and properties of the set object are emphasized. This introduces the *member-of* relationship between a member object and a set of objects (Brodie, 1981).

Aggregation is a form of abstraction in which a relationship between objects is considered as a higher level aggregate object (Brodie et al., 1984). Each instance of an aggregate object can be decomposed into instances of the component objects. This introduces a *part-of* relationship between objects. The type of hierarchy constructed by this abstraction is called an *aggregation hierarchy*.

Like data warehousing and OLAP (online analytical processing), the above-mentioned aggregation hierarchies are widely used to support data aggregation (Lenz & Shoshani, 1997). In a simple form, such a hierarchy shows the relationships between domains of values. Each operation on a hierarchy can be viewed as a mapping from one domain to a smaller domain. In the OLAP environment, hierarchies are used to conceptualize the process of generalizing data as a transformation of values from one domain to values of another smaller or bigger domain by means of drill-down or roll-up operators. In the next sections, the roles of aggregation hierarchies in analysis dimensions of a data cube will be analyzed.

BACKGROUND

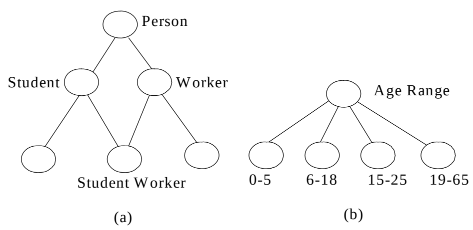
The core of the aggregation hierarchy revolves around the partial order, a simple and powerful mathematical concept to which a lot of attention has been devoted (see Davey & Priestley, 1993). The partial ordering can be represented as a tree with the vertices denoting the elements of the domains and the edges representing the ordering function between elements. The notion of levels has been introduced through the idea that vertices at the same depth in the tree belong to the same level of the hierarchy. Thus, the number of levels in the hierarchy corresponds to the depth of the tree. The highest level is the most abstract of the hierarchy and the lowest level is the most detailed.

As in data warehousing and OLAP, the notion of partial ordering is widely used to organize the hierarchy of different levels of data aggregation along a dimension. Sometimes, hierarchies have been perceived structurally as trees, that is, no generic object is the immediate descendant of two or more generic objects, and where the immediate descendants of any node (supposing any hierarchy is represented by a graph) have classes which are mutually exclusive. A class with a mutually exclusive group of generic objects sharing a common parent is called a *cluster*. Generally speaking, many real cases cannot be modeled by these types of hierarchies (see Figure 1). For this reason, usually, a dimension hierarchy is represented as a directed acyclic graph (DAG). Sometimes, it can be defined with a unique bottom level and a unique top level, denoted by ALL (see Gray, Bosworth, Layman, & Pirahesh, 1996).

One of the most important issues related to the aggregation hierarchy is the *correct aggregation* of data (see Lenz & Shoshani, 1997; Rafanelli & Shoshani, 1990). It is known as *summarizability*, which intuitively means that individual aggregate results can be combined directly to produce new aggregate results.

As subsequently discussed in Lenz and Shoshani (1997), summarizability conditions are the conditions upon which the summarization operation produces the correct result. The authors affirm that three necessary conditions of summarizability have to be satisfied. They are disjointness of levels (or category attributes) in hierarchies, completeness in hierarchies, and correct use of measure (summary attributes) with statistical functions. Disjointness implies that instances of levels in dimensions form disjoint subsets of the elements of a level. Completeness in hierarchies means that all the elements

Figure 1. Two typical structures of hierarchies



occur in one of the dimensions and every element is assigned to some category on the level above it in the hierarchy. Correct use of measures with statistical functions depends on the type of the measure and the statistical function.

More recently, the problem of heterogeneity in aggregation hierarchy structures and its effect on data aggregation has attracted the attention of the OLAP database community. The term heterogeneity, as introduced by Kimball (1996), refers to the situation where several dimensions representing the same conceptual entity, but with different categories and attributes, are modeled as a single dimension. According to this description, which has also been called *multiple hierarchy* and recalled in the next section (see Agrawal, Gupta, & Sarawagi, 1997; Pourabbas & Rafanelli, 2003), dimension modeling may require every pair of elements of a given category to have parents in the same set of categories. In other words, the roll-up function between adjacent levels is a total function. The hierarchies with this property are known to be regular or *homogeneous*. For instance, in a homogeneous hierarchy, we cannot have some cities that roll-up to provinces and some to states, that is, the roll-up function between City and State is a partial function.

In order to model these irregular cases, some authors introduced *heterogeneous* dimensions and tackled the summarizability issue by proposing several solutions.

The proposal of Lehner, Albrecht, & Wedekind (1998) consists of transforming heterogeneous dimensions into homogeneous dimensions in order to be in *dimensional normal form* (DNF). This transformation is actually performed by considering categories, which cause the heterogeneity, as attributes for tables outside the hierarchy. On the flattened child-parent relation, summarizability is achieved for dimension instances.

Pederson and Jensen (1999) considered a particular class of heterogeneous hierarchies, for which they proposed their transformation into homogeneous hierarchies by adding null members to represent missing parents. In their opinion, summarizability occurs when the mappings in the dimension hierarchies are *onto* (all paths from the root to a leaf in the hierarchy have equal lengths), *covering* (only immediate parent and child values can be related), and *strict* (each child in a hierarchy has only one parent). The proposed solutions consider a restricted class of heterogeneous dimensions, and null members may cause a waste of memory and increase the computational effort due to the sparsity of the cube views.

Hurtado and Mendelzon (2001) extended the notion of summarizability for homogeneous dimensions in order to tackle summarizability for heterogeneous dimensions. They classified five classes of dimension schemas, which are

- *homogeneous*, if, along a dimension, the roll-up function from the member set of a level to the member set of its immediate higher level is a total function;
- *strictly homogeneous*, if it is homogeneous and it has a single bottom level,
- *heterogeneous*, if it is not homogeneous;
- *hierarchical*, which allows heterogeneity but keeps a notion of ordering between each pair of levels and their instances; and
- *strictly hierarchical*, if the ordering relationship among different levels is transitive.

Then, a class of constraints on dimension instances is introduced. These constraints are statements about possible categories to which members of a given category may roll up. In Hurtado and Mendelzon (2002), the authors reexamined inferring summarizability in general heterogeneous dimensions. They introduced the notion of *frozen* dimensions, which are minimal homogeneous instances representing different structures that are implicitly combined in a heterogeneous dimension. This notion is used in an algorithm for testing implications of dimension constraints. In a recent work, Hurtado and Guitérrez (2004) proposed a notion of hierarchical equivalence that allows the comparing of dimension schemas with respect to their information contents. In order to achieve this equivalency, an algorithm that transforms heterogeneous dimension schemas into homogeneous schemas is defined.

CHARACTERIZATION

In this section, we discuss the characterization of classification hierarchies from two different perspectives with respect to the above-mentioned proposals: The first is the mapping between domain values, that is, total and partial classification hierarchies, and the second is the hierarchical structure. In relation to them, we represent the characterization of two main OLAP operators, which refer to hierarchies in order to maintain data cube consistency.

Classification Hierarchies

Dimensions have often been associated with different hierarchically organized levels. The name of each level is expressed by the corresponding *variable* name. Generally, the shift from a lower (more detailed) level to a higher (more aggregate) level is carried out by a mapping. A mapping between two variables can be complete or incomplete.

Definition: A mapping between two variables of a hierarchy defines a *containment function* if each variable instance of a lower level corresponds to only one variable instance of a higher level, and each variable instance of a higher level corresponds to at least one variable instance of a lower level. In such a case, it is called *full mapping*.

Note that the containment function satisfies the summarizability conditions discussed in the previous section.

Definition: A *total classification hierarchy* on a given dimension indicates that there is a full mapping between each adjacent pair of variables. If there is no full mapping between at least one adjacent pair of variables, then the hierarchy is called a *partial classification hierarchy*.

In the context of classification hierarchies, we can introduce two specializations given by the following definition.

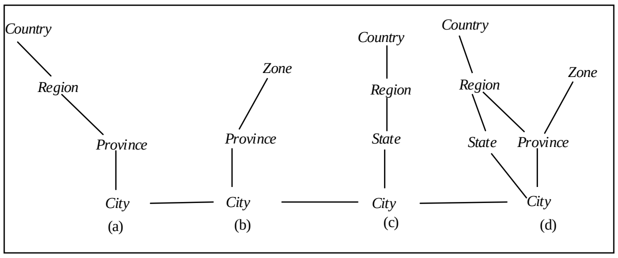
Definition: Let H and H' be two hierarchies. H' is a *multiplicity* of H if they are defined by the same-level domains, and the variable name associated with each level of H' is a specialization of the variable name associated with the corresponding level in H. They form a *multiple hierarchy* if they have at least one variable in common.

In Figure 2, the first three hierarchies represent different partitions of the same location dimension. A possible multiplicity of the hierarchy shown in Figure 2a is City of residence → Province of residence → Region of residence → Country of residence. Figure 2d shows a multiple hierarchy, where the level Province is common in a and b, as well as Region in a and c.

OLAP Operators

The OLAP operators that are involved principally in manipulating dimensions are *roll up* and *slice* (see Agrawal et al., 1997; Cabibbo & Torlone, 1998; Olap Council, 1997). The roll-up operator decreases the detail of the measure, aggregating it along the dimension hierarchy. A problem arises in the case of a partial classifi-

Figure 2. Example of different partitionings of location dimension (a, b, c), and a multiple hierarchy (d)



cation hierarchy, where the domain values of a given level of a dimension hierarchy is a subset of all the possible values that this level can assume in the database, called the *primitive domain*. In particular, when the roll-up operator is applied to change such a level to a higher level, no information about the noncompleteness of its domain is stored. Therefore, the result of this operator incorrectly will be referred to as the primitive domain.

A solution to this problem is to save the information about the domain values that cause the incompleteness of the hierarchy. This can be achieved in two different ways. The first is to add a *note* (the clause “where < variable name > is-a subset of the primitive domain”) to the title of the cube. The second is to add the same note to each variable of the hierarchy whose level is higher with respect to the variable with the incomplete domain.

Example: Let us consider a nation-wide drink company that owns chain stores located in all cities. Let us assume that all stores in the chain sell the same beverages. Figure 3 shows part of the total number of sales in 1997 and 1998.

Let us refer to the classification hierarchy shown in Figure 2c. We suppose the domain of City contains only three instances, San Francisco, San Jose, and San Diego. It is a subset of the primitive domain of City, in which all the cities of California are stored. Let us consider the following query:

“Select vendors for which the sales is >10,000 units in the West.”

It is solved as follows:

Roll-up from City to Region, Dice Region = “West,” Dice Sales > “10000,” Dice Vendor.

The title of the cube obtained from the above query becomes “Vendors with sales > 10,000 units in the West, *where city of California is-a subset of the primitive domain*,” or we have to add the same clause to the variables State, Region, and Country.

The slice operator reduces the dimensions of a cube by eliminating one dimension through its multidimensional space. If the domain of this dimension is a subset of the primitive domain, we lose information about that and the result seems to refer to the whole primitive domain of the omitted dimension. To overcome this problem, a specialization of the slice operator, called *implicit slice* (or *I-Slice*), is introduced, which is based on the concept of *implicit dimension*. An implicit dimension is any dimension of a data cube that has only one instance in its domain. This instance can be single value or multivalued. The I-Slice operator keeps implicit the dimension on which it is applied.

Similarly to the roll-up operator, the same note is added to the title of the cube. For the symmetric reasoning of terminology, we use the term *total slice* (or *T-Slice*) when the domain of the omitted dimension coincides with its primitive domain.

Example: Let us consider the cube in Figure 3, where the primitive domain of Year is defined by the values <1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998>. Instead, the domain value of Year in the considered cube is <1997, 1998>. Let us consider the following query: “Give the sales in all cities by Class and Vendor.” It is solved by Slice Year.

In this case, because the slice operator omits the dimension Year, we lose the exact information on the real period to which the result should refer. In accordance with the above definition, the query can be solved by I-Slice Year. The resulting cube is defined by the same set of dimensions of the primary one, where the title becomes “Sales by Class, Vendor, and Year *where Year is a subset of the primitive domain*.”

FUTURE TRENDS

One of the principal characteristics of hierarchies is their structural changing over time. The role of this feature is investigated in literature. Vaisman and Mendelzon (2003) address mainly only the temporal changing of the schema

Figure 3. Example of a data cube

Sales						
	Class	*City*	*Vendor*	*Year*	1997	1998
	Alcoholic	Los Angles	Smith		10,000	12,000
		New York	Wong		20,000	16,000
		…	…	...	…	…
		Detroit	Clifford		90,000	18,000
	Nonalcoholic	Los Angles	Smith		20,900	14,500
		New York	Wong		12,300	32,009
		...	...		...	...
		Detroit	Clifford		21,000	30,000

of dimensions. The problem of temporal changing of the levels and the relative domains in the context of different types of hierarchies forms an interesting issue to be considered in future work.

The hierarchical structures are extensively used in ontology databases (such as the pathology database which is defined from a large hierarchy of pathology terms). Such databases are primarily hierarchically structured, but can also have DAG structures. They are used in various application areas such as health care, banking, engineering design, and so on. In this framework, representing complex DAG structures as well as their changing over time form interesting issues to be examined.

CONCLUSION

In this paper, we focused on the rules of aggregation hierarchies in analysis dimensions of a cube. We then discussed the hierarchies from two different points of views: mapping between domain values and hierarchical structures. In relation to them, we introduced the characterization of some OLAP operators involved in the hierarchy manipulation.

REFERENCES

Agrawal, R., Gupta, A., & Sarawagi, S. (1997). Modelling multidimensional databases. *Proceedings of the 13th International Conference on Data Engineering*, 232-243.

Borgida, A., Mylopoulos, J., & Wong, H. K. T. (1984). Generalization/specialization as a basis for software specifications. In M. L. Brodie, J. Mylopoulos, & J. W. Schmidt (Eds.), *On conceptual modelling, perspectives from artificial intelligence, databases, and programming languages* (pp. 87-117). New York: Springer Verlag.

Brodie, M. L. (1981). Association: A database abstraction for semantic modelling. *Proceedings of Second International Entity-Relationship Conference*, Washington DC, USA.

Brodie, M. L., Mylopoulos, J., & Schmidt, J. W. (1984). *On conceptual modelling, perspectives from artificial intelligence, databases, and programming languages*. New York: Springer Verlag.

Cabibbo, L., & Torlone, R. (1998). From a procedural to a visual query language for OLAP. *Proceedings of the 10th International Conference on Scientific and Statistical Database Management*, 74-83.

Davey, B. A., & Priestley, H. A. (1993). *Introduction to lattice and order*. Cambridge, England: Cambridge University Press.

Gray, J., Bosworth, A., Layman, A., & Pirahesh, H. (1996). Data cube: A relational aggregation operator generalizing group-by, cross-tab, and roll-up. *Proceedings of the International Conference on Data Engineering*, ICDE'96 (pp. 152-159). New Orleans, LA: IEEE Press.

Hurtado, C., & Guitérrez, C. (2004). *Equivalence of OLAP dimension schemas. Proceedings of Third International Symposium on Foundations of Information and Knowledge Systems (FoIKS)*, Vienna, Austria.

Hurtado, C. A., & Mendelzon, A. O. (2001). Reasoning about summarizability in heterogeneous multidimensional schemas. *Lecture notes in computer science* (Vol. 1973, pp. 375-389). Springer-Verlag.

Hurtado, C. A., & Mendelzon, A. O. (2002). OLAP dimension constraints. *Proceedings of the 21st ACM SIGACT-SIGMOD-SIGArT Symposium on Principles of Database Systems*, 169-179.

Kimball, R. (1996). *The data warehouse toolkit*. J. Wiley and Sons, Inc, New York.

Lehner, W., Albrecht, H., & Wedekind, H. (1998). Normal forms for multidimensional databases. *Proceedings of Eighth IEEE International Conference on Scientific and Statistical Database Management*, 63-72.

Lenz, H.-J., & Shoshani, A. (1997). Summarizability in OLAP and statistical databases. *Proceedings of Ninth International Conference on Scientific and Statistical Data Management*, 132-143.

OLAP Council. (1997). *The OLAP glossary*. Retrieved 1997 from *http://www.olapcouncil.org*

Pedersen, T. B., & Jensen, C. S. (1999). Multidimensional data modelling for complex data. *Proceedings of the 15th International Conference on Data Engineering*, 336-345.

Pourabbas, E., & Rafanelli, M. (2003). Hierarchies. In M. Raffanelli (Ed.), *Multidimensional databases: Problems and solutions* (chap. 4, pp. 91-115). Hershey, PA: Idea Group Publishing.

Rafanelli, M., & Shoshani, A. (1990). STORM: A statistical object representation model. *Lecture notes in computer science* (Vol. 420, pp. 14-29). Charlotte, NC: Springer-Verlag Pub.

Smith, J. M., & Smith, D. C. P. (1977). Database abstractions: Aggregation and generalization. *Journal of the ACM Transactions on Database Systems*, *2*(2), 105-133.

Timpf, S. (1999). Abstraction, levels of detail, and hierarchies in map series. In C. Freksa & D. Mark (Eds.), *Lecture notes in computer science* (Vol. 661, pp. 125-140). Springer Verlag.

Vaisman, A., & Mendelzon, A. (2003). Time in multidimensional databases. In M. Rafanelli (Ed.), *Multidimensional databases: Problems and solutions* (pp. 166-199). Hershey, PA: Idea Group Publishing.

KEY TERMS

Cube: A group of data cells arranged by the dimensions of the data. Assigning a value to each dimension of a cube, the measure is obtained by a mapping from this assignment.

Data Warehouse: A system for storing, retrieving, and managing large amounts of data using some sophisticated techniques of cleaning, filtering, hashing, and compression.

Data Warehousing: It refers to the process of extraction of data from different information sources (e.g., databases, files) and their integration in a single data warehouse.

Directed Acyclic Graph (DAG): A graph with one-way edges containing no cycles.

Fact Table: A member of the star schema data model, which records data about a set of phenomena of interest.

OLAP (Online Analytical Processing**):** A category of software technology that allows users to transform raw data according to predefined or user-defined functions, and quickly and interactively performs slice, dice, and roll-up operations in various dimensions.

Partial Order: A partial order on a set is a binary relation that is reflexive, antisymmetric, and transitive.

Star Schema: It represents a dimensional model that is composed of a central fact table and a set of constituent dimension tables.

Histogram Generation from the HSV Color Space

Shamik Sural
Indian Institute of Technology, Kharagpur, India

A. Vadivel
Indian Institute of Technology, Kharagpur, India

A. K. Majumdar
Indian Institute of Technology, Kharagpur, India

INTRODUCTION

Digital image databases have seen an enormous growth over the last few years. However, since many image collections are poorly indexed or annotated, there is a great need for developing automated, content-based methods that would help users to retrieve images from these databases. In recent times, a lot of attention has been paid to the management of an overwhelming accumulation of rich digital images to support various search strategies. In order to improve the traditional text-based or SQL (Structured Query Language)-based database searches, research has been focused on efficient access to large image databases by the contents of images, such as color, shape, and texture. Content-based image retrieval (CBIR) has become an important research topic that covers a large number of domains like image processing, computer vision, very large databases, and human computer interaction (Smeulders, Worring, Santini, Gupta & Jain, 2000). Several content-based image retrieval systems and methods have recently been developed.

QBIC (Query By Image Content) is one of the first image retrieval systems developed at IBM (Niblack et al., 1993). Color, texture, and shape features are combined to represent each image in this system. The VisualSeek system, developed at the Columbia University, is an image retrieval system based on visual features (Chang, Smith, Mandis & Benitez, 1997). The NeTra system is a prototype image retrieval system, which uses color, texture, shape, and spatial location information as features to retrieve similar images (Ma & Manjunath, 1997). Some of the other popular CBIR systems are MARS (Ortega et al., 1998), Blobworld (Carson, Thomas, Belongie, Hellerstein & Malik, 1999), PicToSeek (Gevers & Smeulders, 2000), and SIMPLIcity (Wang, Li & Wiederhold, 2001).

An analysis of these systems reveals that all of them give a lot of importance on the image color for retrieval. In fact, color is always considered to be an important attribute, not only in content-based image retrieval systems, but also in a number of other applications like segmentation and video shot analysis. In color-based image retrieval, there are primarily two methods: one based on color layout (Smith & Chang, 1996) and the other based on color histogram (Swain & Ballard, 1991; Wang, 2001). In the color layout approach, two images are matched by their exact color distribution. This means that two images are considered close if they not only have similar color content, but also if they have similar color in approximately the same positions. In the second approach, each image is represented by its color histogram. A histogram is a vector whose components represent a count of the number of pixels having similar colors in the image. Thus, a color histogram may be considered to be a signature extracted from a complete image. Color histograms extracted from different images are indexed and stored in a database. During retrieval, the histogram of a query image is compared with the histogram of each database image using a standard distance metric like the Euclidean distance or the Manhattan distance. Since color histogram is a global feature of an image, the approaches based on color histogram are invariant to translation and rotation, and scale invariant with normalization.

Color histograms may be generated using properties of the different color spaces like RGB (Red, Green, and Blue), HSV (Hue, Saturation, and Intensity Value), and others. In this article, we give an overview of the different histogram generation methods using the HSV color space. We first present a brief background of the HSV color space and its characteristics, followed by the histogram generation techniques for various applications.

BACKGROUND

A color space or a color model is a specification of a coordinate system and a subspace within that system where a single point represents a distinct color value. There are several well-known color spaces that are used to represent the pixels of an image. This representation is used for image analysis like extraction of color histograms. Each color space has its own merits and demerits depending on the application and hardware specification where it is going to be used. RGB, CMY, CMYK, and HSV are some of the popular color spaces. The RGB color space contains three color components, namely red, green and blue, each of which appears in its primary spectral components. Devices that deposit colored pigments on paper use CMY color space, and the representation of this color space is with the secondary colors of light, which are Cyan, Magenta, and Yellow. CMYK (Cyan, Magenta, Yellow, Black) color space is similar to CMY but is used to produce true black color, which is muddy-black in the CMY color space.

The HSV (Hue, Saturation, Value) color space, on the other hand, closely corresponds to the human visual perception of color. The HSV color space can be represented as a three-dimensional hexacone, where the central vertical axis represents intensity which takes a value between 0 and 255 (Shapiro & Stockman, 2001). Hue is defined as an angle in the range $[0,2\pi]$ relative to the red axis with red at angle 0, green at $2\pi/3$, blue at $4\pi/3$, and red again at 2π. Saturation is the depth or purity of color and is measured as a radial distance from the central axis to the outer surface. For zero saturation, as we move higher along the intensity axis, we go from black to white through various shades of gray. On the other hand, for a given intensity and hue, if the saturation is changed from zero to one, the perceived color changes from a shade of gray to the most pure form of the color represented by its hue. When saturation is near zero, all pixels, even with different hues, look alike and as we increase the saturation towards one, they tend to get separated out and are visually perceived as the true colors represented by their hues. Thus, the effect of saturation may be considered as that of introducing visual shadows on the image for any given value of hue and intensity.

The HSV model is an ideal tool for developing image and video processing algorithms based on color descriptions. A number of histogram generation methods from the HSV color space have recently been proposed for different applications. We next describe some of these approaches.

HISTOGRAM GENERATION FROM THE HSV COLOR SPACE

The HSV color space in general, and the HSV color histogram in particular, plays an important role in image analysis. A color histogram can be used in image retrieval, segmentation, video shot detection, color and intensity-based clustering, place recognition for topological localization, person identification and authentication using biometric techniques, as well as in many other applications.

For image retrieval applications, an HSV color histogram can be generated using an approach similar to the RGB color space. The hue scale is divided into eight groups, saturation scale is divided into two groups, and the intensity scale is divided into four groups. By combining each of these groups, we get a total of 64 cells to represent a 64-component HSV color histogram. The reason for having a different number of groups for the three scales is that, of the three axes, hue is considered to be the most important, followed by intensity, and finally, saturation. For the H, S, and V combination of values, the corresponding histogram component is determined. The respective histogram component is updated by one for each pixel having the corresponding color combination. An efficient indexing of the histograms can enhance the performance of a CBIR application to a great extent. Smith and Chang (1996) exploit this idea in their color set approach. This method extracts spatially localized color information and provides efficient indexing of the color regions. The large single color regions are extracted first, followed by multiple color regions. They utilize binary color sets to represent the color content as a color histogram. The H and S dimensions are divided into N and M bins, respectively, for a total of N×M bins (Ortega et al., 1998). Each bin contains the percentage of pixels in the image that have corresponding H and S colors for that bin. Intersection similarity is used as a measure to capture the amount of overlap between two histograms.

From the properties of the HSV color space, it is observed that for low values of saturation, a color is approximated by a gray value specified by the intensity level while for higher saturation, the color is approximated by its hue. This captures the human visual properties effectively and can be used to generate a histogram for image retrieval applications (Sural, 2003; Sural, Qian & Pramanik, 2002). The saturation threshold that determines this transition is once again dependent on the intensity. Thus, the value of saturation projected onto the hue and intensity plane is useful for the extraction of color information. A threshold function can be used to determine if a pixel should be represented by its hue or by its intensity in the color histogram. For an intensity value of zero, all the colors are considered as black, whatever their hue or saturation may be. On the other hand, with increasing values of intensity, the saturation threshold that separates hue dominance from intensity dominance goes down. This approach treats the pixels as a distribution of "colors" in an image where a pixel may be of a "gray color" or of a "true color." The histogram is a logical combination of two independent histograms—one for the true colors and one for the gray colors. One drawback of this approach is that for saturation values near the threshold, a pixel is neither a true color pixel nor a gray color pixel. In order to capture the fuzzy nature of human visual perception of color, there is a need for using a soft threshold to determine the dominant property of a pixel. In the soft threshold approach, two components of the histogram are updated for each pixel in an image, namely, a gray color component and a true color component. The quantum of update is determined both by the saturation and the intensity of the pixel, and the sum of the weights of the two contributions equals unity. Also, for the same saturation, the weight varies with intensity. For a lower intensity value, the same saturation gives a lower weight on the true color component and vice versa. This histogram has a high recall and precision of retrieval, and is effectively used in content-based image retrieval systems (Vadivel, Majumdar & Sural, 2003). The soft threshold approach can also be used for video shot detection (Sural, Mohan & Majumdar, 2004).

An input image in a content-based retrieval system may be textured or non-textured. Similarly, the images stored in a database can also be classified as textured or non-textured. When the query image is compared with the database images, the search should be restricted to the relevant portion of the database. An HSV histogram can be effectively used to classify an image into a textured or a non-textured class (Li, Wang & Wiederhold, 2000). In this approach, an image is first segmented into 4X4 pixel regions. HSV color histogram is extracted for each such region. When an image is a color-rich image, hue plays an important role in its representation. Although, hue can represent millions of colors, the human visual system cannot distinctly recognize all of them. The visually similar colors can be combined together in the same color band, as suggested by Gong, Proietti, and Faloutsos (1998). After combining the similar colors, the image is segmented and indexed for content-based retrieval. It is observed that human visual system can perceive distinct colors depending on the NBS (National Bureau of Standards) color distance. Colors with NBS distance below 3.0 are indistinguishable to the human eye. Besides hue, if the intensity axis is also divided into a number of bands, then segmentation, clustering, indexing, and retrieval of images can be done even more effectively (Zhang & Wang, 2000). In this approach, the histograms are generated separately—one from the hue component and the other from the intensity component. K-means clustering algorithm is then applied on these two histograms to obtain the center for each class for indexing.

It should be noted that the histogram of an image does not keep track of the spatial information of the pixels in an image. Two images with the same number of color pixels but at different locations would have the same histogram. This results in higher false-positives. If the spatial relationship is also captured during histogram generation, then the retrieval performance can be enhanced. Color correlogram is a type of histogram generated from the HSV color space that retains spatial information (Ojala, Rautiainen, Matinmikko & Aittola, 2001).

In addition to content-based image retrieval, the HSV histogram is used for a large number of other applications. One important class of applications is the domain of topological localization in robotics. Ulrich and Nourbakhsh (2000) use the HSV histogram for appearance-based place recognition. They first determine the candidate locations based on the current belief of a robot's location and transform the input image into six one-dimensional histograms. Then for each candidate location and for each of the two color bands, RGB and HSV, the reference histogram is determined that matches the input histogram most closely. In order to reduce the resource requirements for storage and transmission of image and video, various data compression techniques are used in practice. A fundamental goal of data compression is to obtain the best possible fidelity for a given data rate or, equivalently, to minimize the rate required for a given fidelity. Matching Pursuit (MP) image representation has proven to give good compression results. It is found that the MP coefficients are interestingly distributed along the diagonal of the color cube. Coding of the MP coefficients is done in the HSV color space, where V becomes the projection of RGB coefficients on the diagonal of the cube, S is the distance of the coefficient to the diagonal, and H is the direction perpendicular to the diagonal where the RGB coefficient is placed. MP coefficients are quantized and a histogram of the quantized coefficients is generated in the HSV space for efficient data compression (Rosa, Ventura & Vandergheynst, 2003).

FUTURE TRENDS

The HSV color space provides a close representation of human visual perception of color. New and effective algorithms are being developed to extract more useful information from the HSV color space. Recent research shows that the HSV color space can even be used to represent both color and texture information in a single histogram. An extension of this work could be in the domain of fuzzy feature extraction for segmentation and retrieval. Further theoretical and experimental comparisons should be made between the HSV and other color spaces.

CONCLUSION

We have discussed a number of histogram generation techniques using the HSV color space and made a critical assessment of their merits and demerits. Color histograms generated from the HSV color space have been used in a variety of image and video processing applications. Content-based image retrieval is one of the most popular domains in which the HSV histogram has been used effectively for high recall and precision of retrieval. The histogram may also be combined with other features for representing semantic contents in image and video.

ACKNOWLEDGEMENT

The work done by Shamik Sural is supported by research grants from the Department of Science and Technology, India, under Grant No. SR/FTP/ETA-20/2003 and by a grant from IIT Kharagpur under ISIRD scheme No. IIT/SRIC/ISIRD/2002-2003.

REFERENCES

Carson, C., Thomas, M., Belongie, S., Hellerstein, J.M., & Malik, J. (1999). Blobworld: A system for region-based image indexing and retrieval. *Proceedings of the Third International Conference on Visual Information Systems* (pp. 509-516).

Chang, S-F., Smith, J.R., Mandis, B., & Benitez, A. (1997). Visual information retrieval from large distributed online repositories. *Communications of the ACM, 40,* 63-71.

Gevers, T., & Smeulders, A.W.M. (2000). PicToSeek: Combining color and shape invariant features for image retrieval. *IEEE Transactions on Image Processing, 9,* 102-119.

Gong, Y., Proietti ,G., & Faloutsos, C. (1998). Image indexing and retrieval based on human perceptual color clustering. *Computer Vision and Patten Recognition,* 578-583.

Li, J., Wang, J.Z., & Wiederhold, G. (2000). Classification of textured and non-textured images using region segmentation. *Proceedings of the Seventh International Conference in Image Processing* (pp. 754-757).

Ma, W.Y., & Manjunath, B.S. (1997). NeTra: A toolbox for navigating large image databases. *Proceedings of the IEEE International Conference on Image Processing* (pp. 568-571).

Niblack, W., Barber, R., Equitz, W., Flickner, M., Glasman, E., Pektovic, D., Yanker, P., Faloutsos, C., & Taubin, G. (1993). The QBIC project: Querying images by content using color texture and shape. *Storage and Retrieval for Image and Video Databases, 1908,* 173-187.

Ojala, T., Rautiainen, M., Matinmikko, E., & Aittola, M. (2001). Semantic image retrieval with HSV correlograms. *Proceedings of the Scandinavian Conference on Image Analysis* (pp. 621-627).

Ortega, M., Rui, Y., Chakrabarti, K., Porkaew, K., Meharotra, S., & Huang, T.S. (1998). Supporting ranked Boolean similarity queries in MARS. *IEEE Transactions on Knowledge and Data Engineering, 10*(6), 905-925.

Rosa, M., Ventura, F., & Vandergheynst, P. (2003). *Scalable color image coding with Matching Pursuit.* Technical Report, ITS-TR-05.03. Signal Processing Institute.

Shapiro, L., & Stockman, G. (2001). *Computer vision.* Englewood Cliffs, NJ: Prentice-Hall.

Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(12), 1349-1380.

Smith, J.R., & Chang, S-F. (1996). Tools and techniques for color image retrieval. *SPIE Storage and Retrieval for Image and Video Databases,* 426-437.

Sural, S. (2003). Histogram generation from the HSV color space using saturation projection. In S. Deb (Ed.), *Multimedia systems and content-based image retrieval*. Hershey, PA: Idea Group Publishing.

Sural, S., Mohan, M., & Majumdar, A.K. (2004). A soft-decision histogram from the HSV color space for video shot detection. In S. Deb (Ed.), *Video data management and information retrieval*. Hershey, PA: Idea Group Publishing.

Sural, S., Qian, G., & Pramanik, S. (2002). Segmentation and histogram generation using the HSV color space for content-based image retrieval. *Proceedings of the IEEE International Conference on Image Processing* (pp. 589-592).

Swain, M.J., & Ballard, D.H. (1991). Color indexing. *International Journal of Computer Vision, 7*(1), 11-32.

Ulrich, I., & Nourbakhsh, I. (2000). Appearance-based place recognition for topological localization. *IEEE International Conference on Robotics and Automation, 2,* 1023-1029.

Vadivel, A., Majumdar, A.K., & Sural, S. (2003). Perceptually smooth histogram generation from the HSV color space for content-based image retrieval. *Proceedings of the International Conference on Advances in Pattern Recognition* (pp. 248-251).

Wang, J.Z., Li, J., & Wiederhold, G. (2001). SIMPLIcity: Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23,* 947-963.

Wang, S. (2001). *A robust CBIR approach using local color histograms.* Masters thesis. Department of Computing Science, University of Alberta, Canada.

Zhang, C., & Wang, P. (2000). A new method of color image segmentation based on intensity and hue clustering. *Proceedings of the International Conference on Pattern Recognition, 3,* 3-8.

KEY TERMS

Content-Based Image Retrieval: Retrieval of images similar to a given image based only on features present in the image and not any external information.

Histogram: A vector whose components represent similar colors in an image. The value of a component is the number of image pixels having that color.

HSV Color Space: A color space consisting of hue, saturation, and intensity value. It is a popular way of representing color content of an image.

Precision: The number of relevant images retrieved as a percentage of the total number of images retrieved.

Recall: The number of relevant images retrieved as a percentage of the total number of relevant images in the database.

Soft Threshold: A fuzzy approach to decide the importance of a feature. This is in contrast to a hard threshold where a yes/no decision is made.

History and Future Development of Group Support Systems

Ta-Tao Chuang
Gonzaga University, USA

Wayne W. Huang
Ohio Unversity, USA

Yue "Jeff" Zhang
California State University, USA

INTRODUCTION

A group support system (GSS) is created with information technology (IT) and decision support techniques for assisting problem formulation and evaluation of alternative solutions in a group meeting (DeSanctis & Gallupe, 1987). The idea of GSS dated back to the 1970s; however, not until late 1980s did GSS take the form as we know it now. In 1987, two GSS systems were developed by researchers in different universities: Software Aided Meeting Management system at University of Minnesota, and GroupSystems at University of Arizona (Wagner, Wynne, & Mennecke, 1993). Since then, much research on GSS has been conducted and many organizations such as IBM, the Department of Defense, and the Internal Revenue Services, have used SAMM and GroupSystems in solving organizational problems.

GSS is an inter-disciplinary area that involves management sciences, organizational behavior, IT, and social psychology. The number of publications and presentations attests to the rapid pace of GSS research and development. As of mid-1998, more than 230 papers were published on the topic (Fjermestad & Hiltz, 1998 – 99). Major international conferences, such as ICIS, AMCIS, and HICSS, provide a GSS track on a regular basis. IBM's CEO claimed that groupware (a commercialized version of GSS) would be a direction of software development in the IT industry for the 21st century.

The rest of the chapter is organized as follows: First, the historical development of GSS research is presented. Next, relevant theories and findings of prior research are discussed. Current key issues are identified in section four. Then, future trends of GSS research are described. A recapitulation concludes this article.

DEVELOPMENT OF GSS RESEARCH

In the 1980s, researchers started to explore how computer, communication, and decision support technologies could be put together to improve group meetings. Those applications had been called Group Decision Support Systems, Electronic Meeting Systems, Computer Supported Cooperative Work, and Groupware. In 1993, the standardized term GSS was used to refer to the above-mentioned applications (Jessup & Valacich, 1993).

The GSS research has grown over six distinct phases (Dennis & Gallupe, 1993; Saunders, 2000). The first phase, "Roots," was the early research into computer messaging and individual support systems in the 1970s and formed the basis for the GSS work. Phase two, called "Initial Explorations," occurred during the early 1980s and focused on the impact of rudimentary GSS on group outcomes and processes. The third phase, from mid- to late-1980s, was known as the "Early Experiments". A series of experimental studies was conducted to compare groups supported by a GSS with unsupported groups. The fourth phase, from late 1980s to mid 1993, was called "Field Studies." Research focus was on the use of GSS technology in organizational settings and on the impact of GSS on organizations. Phase five, the "In-Depth Studies," started in mid-1993. In 1993, a group of researchers summarized lessons and experiences of the first decade GSS research in the book "Group Support Systems: New Perspective" (Jessup & Valacich, 1993). This book summarized studies on technical, behavioral, organizational, and social psychological aspects of GSS uses. One major conclusion from the review of the research was that while promising, the effect of GSS was mixed due to such factors as research settings, technologies employed, and group characteristics. This phase featured two developments

Table 1. Theories relevant to GSS

	Theories	Main thesis
Decision Theorist School	Media-Richness Theory (Daft & Lengel, 1986)	• The task performance depends on the extent to which the information richness requirement of the task matches the richness of the media used by the group.
	Task-Medium Fitness Theory (McGrath & Hollingshead, 1993)	• The optimal richness for each type of task is unique and the task performance depends on the extent to which the task fits the communication environment
	Information Exchange Theory (DeSanctis & Gallupe, 1987)	• A GSS may change the process of interpersonal information exchange and thus change the task performance.
	Task/Technology Fit Theory (Zigurs & Buckland, 1998)	• The task performance depends on the ideal fit between complexity level of the task and the GSS elements.
Institutionalist school	Social Information Processing Theory (Fulk, Schmitz, & Steinfield, 1990)	• Media characteristics as well as the attitude, statements, and behaviors of co-workers influence relational developmental perceptions.
	Information Technology Intervention Framework (Clapper & McLean, 1990)	• Group outcome depends on the informational and normative influence processes operating within the group
	Time, Interaction, and Performance (McGrath, 1991).	• Group processes are not necessarily linear or sequential process. The difficulty level of the task may force the group shift gear from one mode to another; thus, a richer medium required.
	Adaptative Structuration Theory (DeSanctis & Poole, 1994)	• Group outcomes are the results of group's attitude toward the technology and faithfulness of appropriation of social structures of the technology in the context of its use.

that responded to the above conclusion: First, more sophisticated research design was employed to examine particular aspects of GSS applications. Second, attempts were made to propose theories to explain the disparities in findings.

Furthermore, since mid-1990s, web technologies provide new opportunities for the deployment of GSS. The evolution of GSS research thus entered its sixth phase, being labeled as "New Frontier of GSS Research," which can be characterized by two main trends: First, the two developments in phase five continue (Chidambaram, Bostrom, & Wynne, 1990-91; Dennis, Wixom, & Vandenberg, 2001). Second, GSS research extends to the design of web-based GSS (Wheeler, Dennis, & Press, 1999), e-collaboration/collaborative commerce (Burke, 2001-02; Johnson & Whang, 2002; Chuang & Nakatani, 2004), and virtual teams (Saunders, 2000). These trends are shaping the future of GSS research.

MAJOR THEORIES ON GSS AND PRIOR RESEARCH FINDINGS

The review of first decade GSS research showed that there were no conclusive findings in prior research (Pinsonneault & Kraemer, 1990; Gray, Vogel, & Beauclair, 1990; Nunamaker, Dennis, Valacchi, Vogel, & George, 1993). Although the inconsistence could be attributed to the difference in technologies or research methods employed, many researchers attempted to explain the disparity with theories. Those theories can roughly be classified into two schools: decision theorist school and institutionalist school (Dennis, Wixom, & Vandenberg, 2001). The decision theorist school holds the view that the decision quality can be improved with aides of techniques or tools. In this sense, GSS is considered as an instrument that is capable of assisting decision makers to manage the complexity of decision. In contrast, the institutionalist school considers GSS as an opportunity for organizational change. The provision of GSS to a group of people engaged in teamwork does not necessarily improve the quality of group task. Instead, it is the way the technology is used that may help improve task quality. Theories in these two schools and their theses are summarized in Table 1.

While early research placed emphasis on the effect of GSS, lately, much research has been done to validate the theories. The effects of independent and intervening variables on group processes and outcomes are examined. Fjermestad & Hiltz (1998-99) presented a classification scheme of constructs that were studied in 200 experiments. The key factors in the scheme are shown Table 2.

Although numerous studies have been done, there are more unanswered questions than answered (Briggs, Nunamaker & Sprague, 1997-98). This section discusses critical issues of GSS in three areas: Traditional GSS, emerging technologies, and virtual team and e-collaboration (Table 3). First, many issues in traditional GSS need to be addressed. Those issues are either theoretical or methodological. A critical one is to validate and compare competing theories (Table 1). Also, recent research (Dennis, Wixom & Vandenberg, 2001) shows that an integrated approach might be more explanatory than a single theory from either the decision theorist school or the institutionalist school alone. Thus, what and how to integrate as a research foundation is another issue. Methodological

Table 2. Main factors of GSS research prior to 1998 (Adapted from Figure 1, Fjermestad & Hiltz, 1998-99)

INPUT	PROCESS		OUTPUT
CONTEXTUAL FACTORS	INTERVENING FACTORS	ADAPTATION FACTORS	OUTCOME FACTORS
1. Technology 2. Group 3. Task 4. Context	1. Research methods 2. Summary variables • Resultant communication dimensions • Group members perception & problem-solving • Organizing concepts • Operating conditions	1. Group adaptation process: • Structuration • Process variables • Process issues 2. Process gains/losses: • Process gains • Process losses 3. Intermediate role outcomes	1. Efficiency measures 2. Effectiveness measures 3. Satisfaction measures 4. Consensus 5. Usability measures

issues include the choice of technology, task types, and group characteristics in research design. Those issues are directly related to the external validity of GSS research. The major concern is that the task type or technology examined does not reflect what managers truly perform or use (Fjermestad & Hiltz, 1998-99). Future research also needs to address the issue of impact of GSS and task types on social and task interaction processes (Huang & Wei, 2000).

Second, from the viewpoint of developmental research (Nunamaker et al., 1993), the development of web-based GSS raises several issues. Due to the shift of processing activities to the server side and the use of a uniform user interface, one issue is how to assure the availability and the security of process and information processing supports (Zigurs & Buckland, 1998). Furthermore, rapid changes in web technologies make upgrades or re-development of GSS with new technologies a formidable task.

Third, coming along with the web-based GSS is the booming of virtual-team research (Sarker & Sahay, 2003). The unique nature of virtual teams entails more studies on distributed group structures and communication, as well as contextual and process factors (Tung & Turban, 1998). The transient life cycle of virtual team may change the interaction of team members (Piccoli & Ives, 2000). Thus, the resolution of technological, structural, and normative relational issues is critical to the success of virtual teams (Saunders, 2000). Another issue is how to incorporate team-building theory into the design of GSS supporting virtual teams (Huang, Wei, & Lim, 2003). Summarized from relevant literature, some key issues are shown in Table 3.

FUTURE DEVELOPMENT

The field of GSS has thrived in the past two decades and will continue to thrive. Issues indicated above will shape the future direction of GSS research. Based on the GSS literature, we identify the following trends of GSS development:

(1) Using GSS to support virtual teams (Saunders, 2000; Huang, Wei & Lim, 2003): Due to such reasons as strategic alliance, globalization, and outsourcing, virtually working together will become a norm rather than an exception (Saunders, 2000). Many web-based GSS are being developed (Woolley, 2004) and incorporated with multimedia capabilities (Burke, 2001-2002). Advantages of those tools (e.g., relatively low cost and more capabilities) will attract more organizations to adopt web-based GSS in virtual teams. Research opportunities exist in addressing issues identified above.

(2) Emergence of e-collaboration/c-commerce (Burke, 2001-2002; Johnson & Whang, 2002): The connectivity offered by the Internet provides a platform on which companies could collaboratively perform commercial activities beyond transactional level. E-collaboration (or c-commerce) could facilitate knowledge sharing, organizational learning, and product lifecycle management, and provide an opportunity for developing different types of partnerships. Research opportunities include structural issues of c-commerce (Chuang & Nakatani, 2004), coordination mechanisms of c-commerce, and the impact of e-collaboration technology on knowledge management and organizational learning.

Table 3. Key issues in GSS research

ISSUES IN TRADITIONAL GSS	
Theoretical Foundation Validate and compare competing theories	**Hybrid Theoretic Approach** What and how to integrate theories into theoretic foundations
External Validity of GSS Research Improve scope and external validity of research	**Effect of Mixed Communication Media** Impact of GSS with multiple media on group outcomes and processes
More Variant of Task Types Impact of complicated and variant task types	**Study on Group Process** Impact of GSS on influence distribution and (social) interaction process
Group Characteristics Impact of large group size of non-students	**GSS Adoption** Critical success factors for the adoption of GSS
ISSUES RELATED TO EMERGING TECHNOLOGIES	
Web-Enabled GSS Take advantage of new technology while maintaining minimum re-development	**Security and Privacy** Maintain a secure meeting environment
Integration of Web-Enabled GSS with Existing Systems How to integrate state-of-the-art systems with deployed GSS	**Business Value of Web-Enabled GSS** Justification for business value of web-enabled GSS
ISSUES RELATED TO VIRTUAL TEAM AND E-COLLABORATION	
Technological Choice and appropriation of technological infrastructure for virtual teams	**Structural** Managerial control mechanisms and role assignment in virtual teams
Normative What and how to develop norms and rules for virtual teams	**Design of Theory-Based Distributed GSS** Theoretical constructs and principles for the design of GSS
Cultural bridging GSS support to virtual teams with different culture	**Authentication** Security and verification of identification of team members

(3) Globalization of GSS research and applications: As GSS proliferates to regions culturally different from its birthplace, implementation issues and cultural impacts on the effect of GSS apparently need to be addressed. The increasing number of articles resulting from studies conducted outside North America attests to this trend.

(4) Multiple, rather than single, theoretical perspectives are adopted to investigate issues surrounding GSS: While numerous theories have been proposed (Table 1), most of them only partially explain the findings of prior GSS research. Thus, adopting multiple theories in research has been proposed and shown promising results. This may occur in the form of an integrated model or comparison of different theories. Thus, research opportunities include testing and comparing existing theories, and developing and empirically validating more integrated models.

CONCLUSION

The GSS research has evolved through various stages of early exploration, experimental and field studies, and theoretical development, and matured with the advent of computing and communication technologies and with the advance of business practices calling for the support for group activities. The GSS research in the past two decades has resulted in a formidable volume of literature; however, many issues remain to be addressed. Methodological issues in traditional GSS in the decision room must be resolved with more sophisticated research design. Various GSS theories need to be validated or modified to explain the disparity in findings of prior research. New technologies and applications of GSS will certainly present plenty of research opportunities in supporting virtual teams with GSS, e-collaboration, and globalization of GSS applications. Research efforts in these regards are shaping the future direction of GSS development.

REFERENCES

Briggs, R.O., Nunamaker, J. F., Jr., & Sprague, R. H. (Winter 1997-98). 1001 unanswered research questions in GSS. *Journal of Management Information Systems, 14*(3), 3 – 21.

Burke, K. (2001-02). E-collaboration: Working together, being apart. *E-Service Journal*, *1*(3).

Chidambaram, L., Bostrom, R.P., & Wynne, B.E. (1990-1991). A longitudinal study of the impact of group decision support systems on group development. *Journal of Management Information Systems*, 7, 7-23.

Chuang, T.T. & Nakatani, K. (2004). Planning and deployment of collaborative commerce: A conceptual framework and empirical evidences. *International Journal of Electronic Business*, *2*(2), 157-173.

Clapper, D L. & McLean, E.R. (1990). Group decision support systems: Towards a theoretical framework for using information technology to support group activity. *Proceedings of the International Federation for Information Processing Working Group 8.2 Conference*.

Daft, R.L., & Lengel, R.H. (1986). Organizational information requirements, media richness and structural design. *Management Sciences*, *32*(5), 554-571.

Dennis, A.R. & Gallupe, R.B. (1993). A history of group support systems empirical research: Lessons learned and future directions. In L. M. Jessup, L. M. & J. S. Valacich (Eds.), *Group support systems: New perspective* (pp. 59 – 76). New York, NY: Macmillan Publishing Company.

Dennis, A.R., Wixom, B.H. & Vandenberg, R. J. (2001). Understanding fit and appropriation effects in group support systems via meta-analysis. *MIS Quarterly*, *25*(2), 167-193.

DeSanctis, G. & Gallupe, R.B. (1987). A foundation for the study of group decision support systems. *Management Science*, *33*(5), 589- 609.

DeSanctis, G. & Poole, M.S. (1994). Capturing the complexity in advanced technology use: Adaptive structuration theory. *Organization Sciences*, *5*(2), 121-147.

Fjermestad, J. & Hiltz, S.R. (1998 - 99). An assessment of group support systems experimental research: Methodology and results. *Journal of Management Information Systems*, *15*(3), 7-150.

Fulk, J., Schmitz, J., & Steinfield, C.W. (1990). A social information model of technology use. In J. Fulk & C. Steinfield (Eds.), *Organization and communication technology* (pp. 117-140). Thousand Oaks, CA: Sage Publications, Inc.

Gray, P. Vogel, D.R., & Beauclair, R. (1990). *Assessing GDSS empirical research. European Journal of Operations Research*, 49, 162 – 176.

Huang, W. & Wei, K.K. (2000). An empirical investigation of the effects of group support systems (GSS) and task type on group interactions from an influence perspective. *Journal of Management Information Systems*, *17*(2), 181 -206.

Huang, W., Wei, K.K. & Lim, J. (2003). Using a GSS to support a virtual teambuilding: A theoretical framework. *Journal of Global Information Management*, *11*(1), 72-89.

Jessup, L.M. & Valacich, J.S. (1993). *Group support systems: New perspectives*. New York, NY: Macmillan Publishing Company.

Johnson, M.E. & Whang, S. (2002). E-Business and supply chain management: An overview and framework. *Production and Operations Management*, *11*(4), 413-423.

McGrath, J.E. (1991). *Time, interaction, and performance (TIP): A theory of group*. Englewood Cliffs, N. J.: Prentice-Hall.

McGrath, J.E., & Hollingshead, A.B. (1993). Putting the group back in group support systems: Some theoretical issues about dynamic processes in groups with technological enhancements. In L.M. Jessup &, J.S. Valacich (Eds.), *Group support systems: New perspectives* (pp. 78 -96). New York, NY: Macmillan Publishing Company.

Nunamaker, J.F., Dennis, A.R., Valacich, J.S., Vogel, D.R. & George, J.F. (1993). Group support systems research: Experience from the lab and field. In L.M. Jessup & J.S. Valacich (Eds.), *Group support systems: New perspectives* (pp. 125-145). New York, NY: Macmillan Publishing Company.

Piccoli, G. & Ives, B. (2000). Virtual teams: Managerial behavior control's impact on team effectiveness. *Proceedings of the Twenty-First International Conference on Information Systems*. Brisbane, Australia.

Pinsonneault, A. & Kraemer, K.I. (1990). The effects of electronic meetings on group processes and outcomes: An assessment of the empirical research. *European Journal of Operation Research*, 46, 143-161.

Sarker, S. & Sahay, S. (2003). Understanding virtual team development: An interpretive study. *Journal of the Association for Information Systems*, 4, 1-38.

Saunders, C.S. (2000). Virtual teams: Piecing together the puzzle. In R. Zmud (Ed.), *Framing the domain of IT management research: Glimpsing the future through the past* (pp. 29-50). Cincinnati, OH: Pinnaflex.

Tung, L.L. & Turban, E. (1998). A proposed research framework for distributed group support systems. *Decision Support Systems*, *23*(2), 175-188.

Wagner, G.R., Wynne, B.E. & Mennecke, B.E. (1993). Group support systems facilities and software. In L.M. Jessup & J.S. Valacich (Eds.). *Group support systems: New perspectives* (pp. 8-55). New York, NY: Macmillan Publishing Company.

Wheeler, B.C., Dennis, A.R., & Press, L.I. (1999). Groupware comes to the Internet: Charting a new world. *Database for Advances in Information Systems*, *30*(3/4), 8-21.

Woolley, D.R. (2004). Conferencing on the Web. Retrieved from *http://thinkofit.com/webconf*

Zigurs, I. & Buckland, B. (1998). A theory of task/technology fit and group support systems effectiveness. *MIS Quarterly*, *22*(3), 313-334.

KEY TERMS

Anonymity of GSS: Communicators do not know owners of expressed ideas in group interaction processes.

E-Collaboration/Collaborative Commerce: IT-enabled joint intellectual efforts between organizations for planning, design, development, production, and delivery of products and services.

Group Process Gains: The improvement in the outcome of group work that benefits from certain aspect of the meeting process (Nunamaker et al., 1993).

Group Process Loss: The impairment in the outcome of group work that is resulted from certain aspect of the meeting process (Nunamaker et al., 1993).

Group Structures: Structures refer to rules and resources (including software tools) that are used in group interaction process (Poole & DeSanctis, 1990).

GSS (Group Support System): Created with computer and communication technologies and decision support techniques for assisting problem formulation and evaluation of alternative solutions in group meeting (DeSanctis & Gallupe, 1987).

Structuration: The process by which rules and resources provided by technology are produced and reproduced in social life (DeSanctis & Poole, 1994).

Virtual Team: A team consists of members who are scatted in different locations that are across geographical boundaries, and who are working together mainly based upon information technology.

How Teachers Use Instructional Design in Real Classrooms

Patricia L. Rogers
Bemidji StateUniversity, USA

INTRODUCTION

"I've learned how to use the [insert new instructional technology here], so now how do I use it in class?"

From filmstrips and mimeographs, to computer-based simulations and virtual reality, technology seems to dominate teachers' lives as they master the new instructional media for use in their classrooms. Good teaching and learning practices tend to take a back seat while the focus on mastery of the technology reduces teaching into basic presentations and lectures, a format most easily controlled by the instructor. While most pre-K-12 and post-secondary instructors do develop effective courses in which students learn, many would be hard pressed to describe *how* they arrive at certain goals and teaching strategies.

BACKGROUND

The field of instructional design provides sound practices and models that, once modified for use by working teachers, can be used to design effective instruction in any content area (Rogers, 2002). The more difficult issue is helping teachers move beyond the tendency to focus on technology rather than instructional goals. Such focus occurs at lower levels of what can be described as a technology adoption hierarchy (summarized in Table 1): familiarization, utilization, integration, reorganization, and evolution (Hooper & Rieber, 1999).

Somewhere at the integration stage, a "magic line" is crossed and the focus is no longer on the technology but on the teaching and learning. A supporting practical design model can help teacher-designers cross this magic line more efficiently and with a high degree of success.

FUTURE TRENDS

A Modified Instructional Design Model

Prescriptive behavioral models in learning would seem, at first encounter, to be inappropriate in light of the more constructivist practices of current educators. However, most constructivists would concur that one must have solid building blocks or elements before construction of new knowledge can be achieved. Dick and Carey's (1990) original systems design model and subsequent modifications by Gagné, Briggs and Wager (1992) and others offer examples of all of the elements necessary for designing and evaluating effective instruction. What the models lacked, however, was a connection to real classroom teachers: those of us who are really teacher-designers and who must create and develop our courses without benefit of design teams and lengthy pilot tests with target audiences.

Figure 1 is a modification based on several interpretations of the most typical instructional design model (Dick

Table 1. A summary of the technology adoption hierarchy

EVOLUTION	Highest level: is most able to cope with change and has skills to adapt newer technologies as needed or desired in teaching and learning environment.
REORGANIZATION	Re-designs teaching strategies with focus on learning and goals of instruction. Students become more involved in the learning environment.
INTEGRATION	Beginning to accept the technology. Focus soon shifts from learning the technology (and fearing its breakdown) to effective use of the technology in teaching.
UTILIZATION	Basic trial of the new technology. Focus is on finding a use for the technology that may or may not continue, particularly if the technology breaks down.
FAMILIARIZATION	Lowest level of exposure to a technology.

& Carey, 1990). Notice that the five phases of design: analyze, design, develop, implement, and evaluate, are focused not on designing teacher-proof curricula but rather on teacher-designers staying focused on their own environment and learners.

The model helps teachers begin designing with the constraints, issues, community demands, and state and federal mandates in mind before thinking about instructional media or "activities". Once parameters are identified, teacher-designers move into the design phase as they document the overall goals of their course (or, in the case of primary teachers, their school year) while simultaneously considering their learners. What does it mean to be a 3^{rd} grade person? What skills should learners have as they move into 4^{th} grade? What new knowledge is gained in 4^{th} grade to allow learners to become 5^{th} grade students? And so on.

Within this phase, assessments are also considered. Effective design, as well as effective teaching, requires teacher-designers to carefully match goals and objectives to appropriate assessments. Desired types of learning, from basic verbal information to higher order thinking skills (Gagné, Briggs & Wager, 1992) must have matched assessments that allow learners to demonstrate their new skills and abilities. Mismatched goals and assessments are common errors in designing instruction.

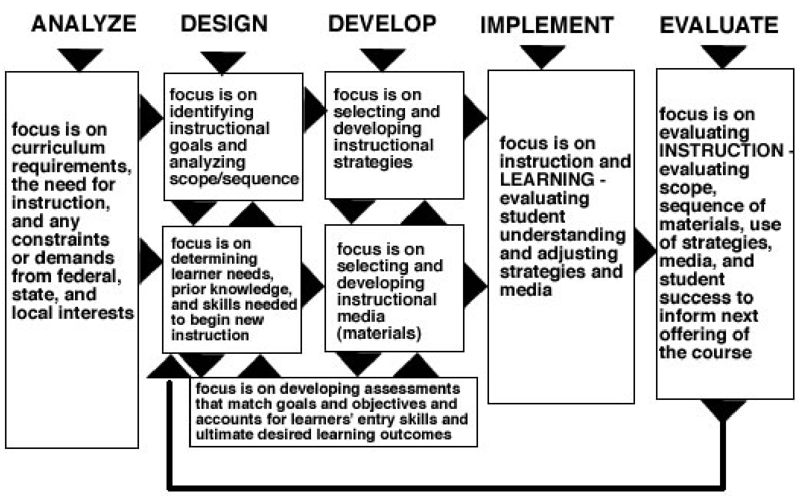
Using this model essentially forces us to wait until the development phase to select teaching strategies and instructional media. For those teachers who are struggling to leave the lower levels of the technology adoption hierarchy, this placement will seem uncomfortable. However, starting with the technology and trying to build an instructional environment is, as should be apparent, in essence turning the design process inside out! Once the focus is away from the goals and objectives and the *learners*, any further course development will likely result in a design that falls far short of the intended learning:

I am elated that I had the opportunity to work on curriculum design for the first time the right way and with a group of faculty members who supported my learning. I have watch[ed] part-time faculty members and even seasoned classroom teachers jump into material they are not familiar with, plan day by day, never really having clear objectives and methods of evaluation [in mind]. (A. Vidovic, personal communication, July 30, 2003)

Notice that the development of assessments also crosses this phase of the design. It is critical to select strategies and media that support the goals and objectives as well as allow students to demonstrate their understanding. Using strategies and media that are similar to the assessment situation strengthens the learning. For example, if students were learning to write poetry, a true-false test would be a very inadequate measure of their skills.

Implementation, *teaching*, is the phase of a teacher-designer's true test. It is here that this model is quite different from traditional instructional design models in that teacher-designers rarely have a chance to "try out" a course on a sample of students. Rather, they often have to simply try things and hope it all works well. However,

Figure 1. Modified instructional design model for teacher-designers. Modifications first introduced in Designing Instruction for Technology-Enhanced Learning*, Rogers, 2002, Idea Group Publishing. Further modifications by Patricia L. Rogers and Catherine E. McCartney, Bemidji State University, for the Online Graduate Program, 2002-2003).*



by following the model thus far, teacher-designers have an advantage over others who do not have clear goals and objectives in mind. During this phase, student achievement and perhaps student evaluations of the course should be examined as evidence that all elements of the design thus far actually form a cohesive course that meets the goals of the instruction. Teacher-designers should take notes on a daily basis regarding which strategies are working with learners, which activities supported new learning, and which instructional medium was appropriate for certain types of learning.

The evaluation phase in this model relies heavily on the evidence from the previous phase and includes a critical look at any notes from the teaching experience, comparison to a previous experience teaching the course, and so on:

In designing and developing this online class using the first couple assignments (objectives, goals, subgoals, etc.), I really feel like [my] course's material fits together much better than it has when I taught it in the past. Though this [instructional design] process took a fair amount of time, I know I would never tackle another class design without using this process first. It does seem to speed up the material/content piece considerably by doing this first. (N. Gregg, personal communication, July 28, 2003)

Barriers to Designing Effective Instruction in Distance Learning

By following a model that is based in practical, real-world experiences of teachers, teacher-designers are able to develop effective and well-documented instruction. However, we should note that there are many reasons good instructional design practices are not followed, and that most are out of the teacher-designer's control. Table 2 is a summary of some of the issues and barriers faced by teacher-designers.

Table 2. A summary of barriers to designing effective instruction

Fear of change Changing teaching methods (strategies) to accommodate newer technologies, different modes of delivery, and the reality of managing a larger student market carries a certain amount of risk and challenge. The human tendency to want things to remain the same introduces a fear factor in designing and delivering instruction in the 21st century (Dublin, June 2003). **Unfamiliarity with newer technologies** The introduction of newer technologies in teaching usually results in teachers defaulting to presentations and lectures. Once the "magic line" is crossed, teaching and learning with technology refocuses from the technology to learning (Dublin, June 2003; Hooper & Rieber, 1999; Strauss, June 2003). **Correspondence, Lecture, and Interactive Learning** Real classrooms rely on interactions among students and the instructor. Some online courses are actually stand-alone correspondence courses that are self-paced and lack high interactivity levels. Lecture courses tend to be one-way communications while other strategies emphasize interactivity. There is a critical need to be clear about levels of interactivity in learning environments (Cavalier, June 2003).	**Ill-defined goals and objectives** Defining goals and objectives is often a new experience for many faculty. Goals and objectives may not match teaching style or adequately address desired learner outcomes. **Unrealistic administrative, policy, or economic pressures** Some teachers have encountered serious constraints when designing instruction. A partial list includes: forced use of traditional "activities" that become the central focus of the instruction, district-wide adoption of specific texts or programs designed to be "teacher-proof" with little flexibility, limited development time for teachers, and a focus on state-wide test scores directly tied to school funding (Rogers, 2000). **Difficulty in translating from one environment to another, such as onground to online** Moving a course from onground delivery to the online environment sets up barriers for inexperienced teachers: some try to limit all transactions to real time and have a felt need to recreate their onground course exactly. Others err on the other side and resort to a type of glorified correspondence approach.

CONCLUSION

A strong case can be made for working with teacher-designers at all levels of education on sound instructional design practices. "Winging it" when it comes to designing effective instruction is ill-advised in the rarified air of the 21st century knowledge and information age, with many demands from learners for high-quality educational experiences. Educational institutions, particularly colleges and universities, are faced with harsh competition for the teaching aspect of their institution from for-profit companies. Such companies outspend higher education in development, maintenance, and marketing of educational offerings, particularly in online learning (Rogers, 2001). Non-profit educational institutions can compete most effectively by providing (a) affordable pricing, (b) greater accessibility to education, and (c) high-quality, personalized educational experiences for their learners. A and B are usually easily attained. High quality education (c) begins with great teachers and support staff and is built and sustained with solid instructional design practices.

REFERENCES

Cavalier, R. (2003, June). Interactions in education: A conversation with Brenda Laurel. *Syllabus. http://www.syllabus.com/article.asp?id=7764*

Dick, W., & Carey, L. (1990). *The systematic design of instruction* (3rd ed.). Glenview, IL: Scott Foresman.

Dublin, L. (2003, June 24). If you only look under street lamps...Or nine e-learning myths. *E-learning Insider, 2003*(1). *http://www.elearningguild.com/pbuild/linkbuilder.cfm?selection=doc.421*

Gagné, R.M., Briggs, L.J., & Wager, W.W. (1992). *Principles of instructional design.* Orlando, FL: Harcourt, Brace, Jovanovich.

Hooper, S., & Rieber, L. (1999). Teaching, instruction, and technology. In A.C. Ornstein & L.S. Behar-Horenstein (Eds.), *Contemporary issues in curriculum* (2nd ed., pp. 252-264). Boston: Allyn and Bacon.

Rogers, P.L. (2000). Barriers to adopting emerging technologies in education. *Journal of Educational Computing Research, 22*(4), 455-472.

Rogers, P.L. (2001). Traditions to transformations: The forced evolution of higher education. *Educational Technology Review, 9*(1). *http://www.aace.org/pubs/etr/issue1/rogers.cfm*

Rogers, P.L. (Ed.). (2002). *Designing instruction for technology-enhanced learning.* Hershey, PA: Idea Group Publishing.

Seels, B.B., & Richey, R.C. (1994). *Instructional technology: The definitions and domains of the field.* Washington, DC: Association for Communications and Technology.

Strauss, H. (2003, June). My dog knows html—Should your faculty? *Syllabus. http://www.syllabus.com/article.asp?id=7774*

KEY TERMS

ADDIE: The five phases of most instructional design models: analyze, design, develop, implement, and evaluate. Some models follow the phases in a linear fashion, while others may approach the phases in a holistic or phenomenologic manner.

E-Learning: A term used to describe learning that takes place usually online, but includes all forms of electronically-enhanced and mediated learning. Computer-aided instruction, just-in-time learning, and intelligent systems can be included in the term "e-learning".

Instructional Design: The field of instructional design includes a range of professions from programmers and graphic artists, to the instructional designer. Designers are able to analyze instruction, learners, environments, strategies, and media to develop effective instruction of training. Designers may or may not be subject matter experts.

Instructional Design Models: Traditional design models are prescriptive step-by-step processes, usually associated with behaviorist instructional strategies. Phenomenological models incorporate constructivist philosophies and practices. In either aspect, design models guide the user in designing effective instruction that takes all aspects of design (see ADDIE) and reminds the user of critical elements and decisions in designing effective instruction.

Instructional (Educational) Technology: Instructional technology is the theory and practice of design, development, utilization, management and evaluation of processes and resources for learning (Seels & Richey, 1994).

Teacher-Designer: "…if you have any experience with instructional design you know that the field and the various models of design associated with it seem most appropriate for teams of people working on the course materials together. Once in a while, some of us are fortunate enough to have instructional designers, subject matter experts, graphic artists, programmers and so on available on our campus or in our school district to assist us with our technology-enhanced course. But most often, it the teacher alone who must rethink and redesign his or her course for technology-enhanced learning. And very often it is the teacher who must also prepare the materials for the Internet, interactive television, or some other delivery medium. They often do not have any background in instructional design theory or practices and have only just mastered the skills for using the delivery medium. These are the people I call 'teacher-designers'" (Rogers, 2002, p. 2).

Technology Adoption Hierarchy: "The model…has five steps or phases: familiarization, utilization, integration, reorientation, and evolution. The full potential of any educational technology can only be realized when educators progress through all five phases; otherwise, the technology will likely be misused or discarded…The *traditional* role of technology in education is necessarily limited to the first three phases, whereas contemporary views hold the promise to reach the evolution phase" (Hooper & Rieber, 1999, p. 253).

Human Body Part Classification and Activity Recognition for Real-Time Systems

Burak Ozer
Verificon Corp., USA

Tiehan Lv
Princeton University, USA

Wayne Wolf
Princeton University, USA

INTRODUCTION

Recent advances in camera and storage systems along with increased algorithmic and computational power of 3D graphics hardware are main factors driving the increased popularity of multicamera applications. Since prices continue to drop on components, cost-effective systems can be developed for a wide range of applications such as teleimmersion, humanoid robots systems, automated video surveillance, and interactive video games.

The increased importance of such applications that require fast, cheap, small, and highly accurate smart cameras necessitates research efforts to provide efficient solutions to the problem of real-time detection of persons and classification of their activities. A great effort has been devoted to three-dimensional human modeling and motion estimation by using multicamera systems in order to overcome the problems due to the occlusion and motion ambiguities related to projection into the image plane. However, introduced computational complexity is the main obstacle for many practical applications. Different applications require different levels of modeling-related performance metrics, that is, accuracy, speed, and robustness, hence different 3D techniques. Therefore, the relationship between the activity recognition algorithms and the architectures required to perform these tasks in real time is a fundamental design issue.

BACKGROUND

Human detection and motion recognition involve interpretation of elementary visual features such as shape, color, texture, and motion. A geometric model is an approximation of the shape and of the deformations of the object. This model can be two-dimensional (modeling the contours of the projections of the object in the images) or three-dimensional (modeling the surfaces of the object). Two-dimensional human shape models are generally made of curves, segments, sticks, snakes, and so forth, where 3D models are either systems of rigid bodies (e.g., spheres, superquadrics, etc.) or deformable surfaces (e.g., mesh). Template matching, chain coding, Fourier transform, boundary signatures, moments, and polygonal approximations are popular shape-matching algorithms. The required properties of a shape description scheme are invariance to translation, scale, rotation, luminance, and robustness to partial occlusion. Color and texture are mostly used to detect skin areas in human-detection tasks. The articulations may be modeled by joints or by the motion of control points (e.g., B-splines). Major approaches for analyzing spatial and temporal patterns include dynamic time warping, neural networks, hidden Markov models (HMMs), time-frequency analysis, template matching, and principal component analysis. The choice between a 2D or 3D model depends on the application, for example, the needed precision, number of cameras, or type of motion to be recognized.

Several researchers work with 2D features to recognize human movement (Gavrila, 1999). Goddard (1994) uses model-based recognition techniques, namely, stick figures. Other researchers who use 2D models are Comaniciu, Ramesh, and Meer (2000), Isard and MacCormick (2001), and Papageorgiu and Poggio (1999). Most of the work in this area is based on the segmentation of different body parts. Wren, Clarkson, and Pentland (2000) proposed a system, Pfinder, to track people by using blobs that represent different body parts. W4 is another real-time human-tracking system (Haritaoglu, Harwood, and Davis, 1998), where the background information should be collected before the system can track foreground objects.

Three-dimensional human and activity recognition systems can be classified in terms of the visual analysis of multiple cameras, that is, the projection of 3D models onto 2D images versus 3D visual reconstruction from stereo images. Aggarwal and Cai (1999), Gavrila (1999), and Moeslund & Granum (2001) present overviews of various methods used for articulated and elastic nonrigid motion detection. More background information on gesture recognition can be found in Wu and Huang (1999).

One of the early works on tracking articulated objects is proposed by O'Rourke and Badler (1980). The authors used a 3D model of a person made of overlapping spheres. Kakadiaris and Metaxas (1998) proposed a method to generate the 3D model of an articulated object from different views. The authors used extended Kalman filter for motion prediction. Gavrila and Davis (1996) used superquadrics to model the human body and dynamic time warping to recognize the human motion. Luck, Debrunner, Hoff, He, and Small (2002) obtained 3D models of the moving human body by extracting the silhouettes from multiple cameras. In this work, 3D information is used for HMM-based activity recognition in real time for different applications without requiring any specific pose or user interaction and using distributed processing. This method is explained in more detail in the next section with an emphasis on issues related to the relationship between the algorithms and design architecture.

Most of the previous work for human detection depends highly on the segmentation results, and mostly motion is used as the cue for segmentation. Most of the activity-recognition techniques rely on the successful feature extraction, and proposed approaches are suitable for a specific application type. There is a need for more robust techniques that can use model-based information feedback and connect low-level features to high-level semantics in real time.

MAIN THRUST OF THE ARTICLE

As previously stated, 3D human- and activity-recognition systems can be classified in terms of the visual analysis of multiple cameras, that is, the projection of 3D models onto 2D images versus 3D visual reconstruction from stereo images. Depending on the selected technique, different steps must be implemented. Visual reconstruction for virtual reality requires high accuracy while real-time activity recognition and trajectory estimation require high-speed techniques. Various aspects and challenges of 3D visual-analysis algorithms include instruction statistics, branch behavior, and memory access behavior of different program parts, for example, stereo matching, disparity map generation, reconstruction, projection, 2D or 3D human body part detection, 2D or 3D tracking, 2D or 3D activity recognition, and so forth.

Algorithmic and Hardware Issues

Algorithmic Issues

In general, we can classify 3D human detection and activity recognition methods into two categories (Cheung, Kanade, JBouguet, & Holler, 2000): off-line methods, where the algorithms focus on detailed model reconstruction, for example, wire-frame generation, and real-time methods with a global 3D human model reconstruction (Delamarre & Faugeras, 2001). The major challenge in many 3D applications is to compute dense range data at high frame rates, since participants cannot easily communicate if the processing cycle or network latencies are long. The works by Kakadiaris and Metaxas (1995) and Mulligan, Isler, and Daniilidis (2001) are examples for non-real-time methods. Most of the real-time methods use a generic 3D human model and fit the projected model to the projected silhouette features (Luck et al., 2002). The speed of the systems depends highly on the voxel resolution.

The algorithmic pipelines perform a wide range of disparate operations:

- pixel-by-pixel operations,
- pixel-region operations,
- mixed operations, and
- nonpixel operations.

Traditionally, the algorithms start with operations that are clearly signal oriented and should move steadily away from the signal representation until the data are very far removed from a traditional signal representation. In general, the volume of data goes down as image processing progresses.

Hardware Issues

Real-time implementation of image- and video-processing algorithms necessitates data- and instruction-level parallelism techniques to achieve the best performance for several application types. Besides the algorithm development, hardware design is one of the most important issues for a real-time system. Watlington and Bove (1997) proposed a data-flow model for parallel media processing. Davis, Borovikov, Cutler, and Horprasert (1999) developed a multiperspective video system. Fritts, Wolf, and Liu (1999) evaluated the characteristics of multimedia applications for media processors. Researchers also pay attention to multiprocessor architecture. Simultaneous multithreading is proposed by Tullsen, Eggers, and Levy (1995). An IMAGINE processor is being developed at Stanford University, which has an explicit, programmable communication structure (Khailany, Dally, Rixner, Kapasi, Mattson, Namkoong, et al., 2001).

The selection of an image and video processor must be based on a number of issues including power, cost, development tools, and performance-related features. Examples include Texas Instruments' DSPs (the TMS320C6201 and C6701) that use a VLIW (very long instruction word) architecture. Another VLIW processor is the Philips TriMedia processors that can execute up to five instructions per cycle. Besides the main CPU (central processing unit), there are other peripherals that can take the load from the main CPU for particular computations. General-purpose processors' (GPP) high power consumption and large size are the main disadvantages for portable image- and video-processing applications. Another important factor is cost. For cost-sensitive applications, Digital Signal Processing (DSP) devices are significantly less expensive than GPPs.

3D Visual Analysis and Synthesis Steps

Disparity Map Generation

One well-known technique to obtain depth information from digital images is the stereo technique. In stereo techniques, the objective is to solve the correspondence problem, that is, to find the corresponding points in the left and right image. The difference in the spatial position of the corresponding points, namely, disparity, is stored in a disparity map. Whenever the corresponding points are determined, the depth can be computed by triangulation. Attempts to solve the correspondence problem have produced many variations, which can be grouped into matching pixels and matching features, for example, edges. The former approach produces dense depth maps while the latter produces sparse depth maps. Even though stereo vision techniques are used in many image-processing applications, the computational complexity of matching stereo images is still the main obstacle for practical applications.

Two main performance evaluation metrics are throughput that is the frame rate times the frame size and the range of disparity search that determines the dynamic range of the distance measurement. There is still a great deal of research devoted to develop stereo systems to achieve the desired performance. Early real-time stereo systems do not provide a complete video-rate output of range as dense as the input image with low latency. Another major problem is that the depth maps obtained by current stereo systems are not very accurate or reliable. Newer methods that address these problems include the multibaseline stereo method (Kanade, Yoshida, Oda, Kano, & Tanaka, 1996) that requires local computation, block-matching approach with parallel algorithms (Koschan & Rodehorst, 1995), area-based stereo matching (Konolige, 1997), and hybrid recursive matching (Schreer, Brandenburg, & Kauff, 2001) with an efficient selection of a small number of candidate vectors.

Reconstruction and Calibration

Reconstruction involves computing for each corresponding point pair in the images a point in space. This requires calibration of the cameras. There are two major parameter sets for cameras, namely, intrinsic and extrinsic parameters. If both of the parameter sets are known, then the cameras are fully calibrated. By using the intrinsic parameters, the 3D depth map can be converted into Cartesian coordinates. The extrinsic parameters are used to convert the camera-centered position into a position in the world coordinates (Narayanan, Rander, & Kanade, 1998). These 3D points are converted into a surface representation via a triangular mesh. The algorithm calculates the correspondence that minimizes the geometric error subject to the epipolar constraint. Detailed information about cameras and camera calibration can be found in Hartley and Zisserman (2000). Although many algorithms are proposed for more accurate and reliable 3D object reconstruction, they are not suitable for practical applications due to their computational complexity. Depending on the application type, algorithmic solutions, for example, using constraints from prior information, and hardware solutions, for example, using a graphics hardware that effectively combines a plane-sweeping algorithm with view synthesis (Yang, Welch, & Bishop, 2002), are proposed.

Rendering

Rendering is the process of producing realistic 3D images. The rendering issues are related to the interaction between light and surface, intersection of viewing rays, and objects sampling of the scene and displaying techniques. There are four main rendering methods used in visualization, that is, ray tracing, volume rendering, radiosity, and polygon rendering. Due to the high computational requirements of traditional computer graphics, general-purpose computers are not efficient in rendering applications. Consequently, special-purpose graphics engines are developed primarily for polygon rendering. Similarly, special-purpose volume-rendering architectures are developed to meet the special needs of volume rendering in order to compute rapidly and repeatedly from a volume data set. A detailed survey on graphics hardware can be found in Thompson, Hahn, and Oskin (2002).

A Data Fusion Algorithm for 3D Human-Activity Recognition

In this subsection, we present our test bed architecture where a single camera node is composed of a standard camera and a TriMedia video-processing board. Multiple TriMedia processing boards can be installed to one host PC (personal computer) to provide multiprocessing ability. A TriMedia board has a TM1300 TriMedia processor with its own dedicated memory. A 32-bit TM1300 TriMedia processor has a five-issue VLIW CPU together with several coprocessors. The CPU in the processor has multiple functional units and 128 registers.

The proposed algorithm detects a wide range of activities for different applications. For this reason, the scheme detects different body parts and their movement in order to combine them at a later stage that connects to high-level semantics. Low-level processing steps include background elimination, skin-area detection, segmentation, finding connected components, ellipse fitting, and graph matching with invariant shape attributes (Ozer, Wolf, and Akansu, 2000). High-level processing is based on HMMs where each body part has its own freedom of motion, and the activity recognition for each part is achieved by using several HMMs in parallel. The Mahalanobis distance classifier is used for combining the activities of different body parts by assigning different weights for each activity. Different activity patterns can have overlapping periods (same or similar patterns for a period) for some body parts. Hence, the detection of start and end times of activities is crucial (Ozer & Wolf, 2002b). A global 3D ellipsoid model of the human body is generated from 2D ellipses obtained from orthogonal views, and the resulting 3D information is used to verify the fit of the real body parts with the actual model. Camera calibration and data synchronization are main issues in data fusion from multiple cameras.

FUTURE TRENDS

In this section, potential architectures, that is, VLIW, the symmetric parallel architecture, and macropipeline architectures, are discussed to exploit the parallelism resulting from the data independencies. Some pixel-level operations on different pixels may be independent. This independence can be converted into a different form of parallelism such as instruction-level parallelism, thread-level parallelism, process-level parallelism, as well as spatial parallelism, which can be utilized by array processors. When the input frame size is not large enough to make those additional costs ignorable, it is more preferable to convert this intraframe data independency into instruction-level parallelism, which can be explored by VLIW or superscalar architecture processors. While a superscalar processor is more efficient for processing dynamic instruction-level parallelism, a VLIW processor can achieve similar performance on the program with explicit parallelism with significantly less hardware effort. The first step is to use loop fusion, a way of combining two similar, adjacent loops for reducing the overhead, and loop unrolling, which partitions the loops to discover loop-carried dependencies that may let several iterations be executed at the same time, to increase the basic block size and thus increase available instruction parallelism. In the second step, we sought two methods to reduce the branches, which limit the basic block size in loops. A solution for this is to use conditional execution instructions, which require hardware support. Another technique to convert control flow dependency to data dependency is using lookup tables.

A coarse-grained data independency in multiple camera systems is the interframe data independency. The corresponding parallelisms are thread- or process-level parallelism. SMT (simultaneous multithreading) and CMP (single chip multiprocessor) architectures can exploit process-level parallelism. However, the SMT architecture seems not a good choice for this parallelism since the almost-identical threads content the same resource and do not increase the functional unit utilization over the single thread model. Therefore, the CMP architecture or even separate chip processors are more preferable to exploit such interframe parallelism.

The pipelined multiprocessor architecture exploits parallelism resulting from the data flow structure. The additional benefit of such an architecture over other parallel architectures is that the processor can be tailored to the requirements of the algorithm stages. Among all the architectures, the symmetric parallel architecture can provide the better speedup, while the pipelining architecture will be able to reduce hardware effort on processors. As we can see through discussion, those different architectures do not exclude each other. Thus, we would expect a better solution by combining them together. Another important point is the parallelism architecture aspect of the multiple-camera system. The interstage and interframe parallelism, communication cost between processors (i.e., amount of data to be transferred; synchronization), and workload balance should be taken into account to distribute the load between processors.

CONCLUSION

Three-dimensional visual analysis overcomes the problems due to the occlusion and motion ambiguities related to projection into the image plane. Three-dimensional human modeling and activity recognition offer new, emerging applications such as teleimmersion, humanoid robots systems, automated video surveillance, and interactive video games. These applications require fast, cheap, small, and highly accurate smart cameras and efficient algorithmic solutions. The issues related to algorithmic and hardware architectures discussed in this article offer many challenges for researchers. Different applications require different levels of modeling-related performance metrics, that is, accuracy, speed, and robustness, hence different 3D techniques. Exploiting the relationship between the activity-recognition algorithms and the architectures required to perform these tasks in real time is a critical issue for the success of 3D visual systems. An example study is presented where the available data independencies are analyzed and potential architectures to exploit the parallelism resulting from these independencies are discussed.

REFERENCES

Aggarwal, J. K., & Cai, Q. (1999). Human motion analysis: A review. *Computer Vision and Image Understanding, 73*(3), 428-440.

Cheung, G.K.M., Kanade, T., JBouguet, J.-Y., & Holler, M. (2000). A real time system for robust 3D voxel reconstruction of human motions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 714-720.

Comaniciu, D., Ramesh, V., & Meer, P. (2000). Real-time tracking of non-rigid objects using mean shift. *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 142-149.

Davis, L. S., Borovikov, E., Cutler, R., & Horprasert, T. (1999). Multi-perspective analysis of human action. *Proceedings of the International Workshop on Cooperative Distributed Vision*, Cambridge, UK.

Delamarre, Q., & Faugeras, O. (2001). 3D articulated models and multi-view tracking with physical forces. *Computer Vision and Image Understanding, 81*(3), 328-357.

Fritts, J., Wolf, W., & Liu, B. (1999). Understanding multimedia application characteristics for designing programmable media processors. *Proceedings of the SPIE Photonics West, Media Processors*, 2-13.

Gavrila, D. M. (1999). The visual analysis of human movement: A survey. *Computer Vision and Image Understanding, 73*(1), 82-98.

Gavrila, D. M., & Davis, L. (1996). 3D model based tracking of humans in action: A multi-view approach. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 73-80.

Goddard, N. (1994). Incremental model-based discrimination of articulated movement direct from motion features. *Proceedings of the IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, 89-94

Haritaoglu, I., Harwood, D., & Davis, L. (1998). W4: A real time system for detecting and tracking people. *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, 222-227.

Hartley, R., & Zisserman, A. (2000). *Multiple view geometry*. Cambridge University Press, Cambridge, UK.

Isard, M., & MacCormick, J. (2001). A Bayesian multiple-blob tracker. *Proceedings of the IEEE International Conference on Computer Vision*, 34-41.

Kakadiaris, I. & Metaxas, D. (1995). 3D human body model acquisition from multiple views. *Proceedings of the IEEE International Conference on Computer Vision*, 618-623.

Kakadiaris, I. A., & Metaxas, D. (1998). Three-dimensional human body model acquisition from multiple views. *International Journal of Computer Vision, 30*(3), 227-230.

Kanade, T., Yoshida, A., Oda, K., Kano, H., & Tanaka, M. (1996). A stereo machine for video rate dense depth mapping and its new applications. *Proceedings of the Conference on Computer Vision and Pattern Recognition*, 196-202.

Khailany, B., Dally, W. J., Rixner, S., Kapasi, U. J., Mattson, P., Namkoong, J., et al. (2001). Imagine: Media processing with streams. *IEEE Micro, 21*(2), 35-46.

Konolige, K. (1997). Small vision systems: Hardware and implementation. *Proceedings of the International Symposium on Robotics Research*, 203-212.

Koschan, A., & Rodehorst, V. (1995). Towards real-time stereo employing parallel algorithms for edge-based and dense stereo matching. *Proceedings of the IEEE Workshop on Computer Architectures for Machine Perception*, 234-241.

Luck, J. P., Debrunner, C., Hoff, W., He, Q., & Small, D. E. (2002). Development and analysis of a real-time human motion tracking system. *Proceedings of the IEEE Workshop on Applications of Computer Vision*, 196-202.

Moeslund, T.B., & Granum, E. (2001). A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding*, 81, 231-268.

Mulligan, J., Isler, V., & Daniilidis, K. (2001). Trinocular stereo: A real-time algorithm and its evaluation. *Proceed-*

ings of the IEEE Workshop on Stereo and Multi-Baseline Vision, 1-8.

Narayanan, P. J., Rander, P. W., & Kanade, T. (1998). Constructing virtual worlds using dense stereo. *Proceedings of the International Conference on Computer Vision*, 3-10.

O'Rourke, J., & Badler, N. I. (1980). Model-based image analysis of human motion using constraint propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 2*(6), 522-536.

Owens, W. H. (1984). The calculation of a best-fit ellipsoid from elliptical sections on arbitrarily oriented planes. *Journal of Structural Geology, 6*, 571-578.

Ozer, I. B., & Wolf, W. (2002a). A hierarchical human detection system in (un)compressed domains. *IEEE Transactions on Multimedia. Special Issues on Multimedia Databases*, *4*(2), 283-300.

Ozer, I. B., & Wolf, W. (2002b). Real-time posture and activity recognition. *Proceedings of the IEEE Workshop on Motion and Video Computing*, 133-138.

Ozer, I. B., Wolf, W., & Akansu, A. N. (2000). Relational graph matching for human detection and posture recognition. *Proceedings of the SPIE, Photonic East Internet Multimedia Management Systems*, Boston.

Papageorgiu, C., & Poggio, T. (1999). Trainable pedestrian detection. *Proceedings of the International Conference on Image Processing*, 25-28.

Schreer, O., Brandenburg, N., & Kauff, P. (2001). Real-time disparity analysis for applications in immersive tele-conference scenarios: A comparative study. *Proceedings of the International Conference on Image Analysis and Processing*.

Thompson, C. J., Hahn, S., & Oskin, M. (2002). Using modern graphics architectures for general-purpose computing: A framework and analysis. *Proceedings of the ACM/IEEE International Symposium on Microarchitecture*, 306-317.

Tullsen, D. M., Eggers, S. J., & Levy, H. M. (1995). Simultaneous multithreading: A platform for next-generation processors. *Proceedings of the International Symposium on Computer Architecture*, 392-403.

Watlington, J. A., & Bove, V. M. (1997). A system for parallel media processing. *Parallel Computing*, *23*(12), 1793-1809.

Wren, C. R., Clarkson, B. P., & Pentland, A. P. (2000). Understanding purposeful human motion. *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, 378-383.

Wu, Y., & Huang, T. S. (1999). Vision-based gesture recognition: A review. *Proceedings of the Gesture Workshop*, 103-115.

Yang, R., Welch, G., & Bishop, G. (2002). Real-time consensus-based scene reconstruction using commodity graphics hardware. *Proceedings of the Pacific Conference on Computer Graphics and Applications*, 225-234.

Zisserman, A., Fitzgibbon, A., & Cross, G. (1999). VHS to VRML: 3D graphical models from video sequences. *Proceedings of the International Conference on Multimedia Systems*, 51-57.

KEY TERMS

Camera Calibration: A process of setting digital imaging components to standardized settings that will produce accurate and predictable results in the output.

Data Fusion: The fully automated method of merging diverse data into a single, coherent representation of the tactical, operational, or strategic situation.

Disparity Map Generation: Solving the correspondence problem, that is, to find the corresponding points in the stereo images by finding the difference in spatial position of the points, namely, disparity.

Gesture/Activity Recognition: Dynamic body configuration that involves spatiotemporal analysis.

Posture Recognition: Static body configuration without a dynamic movement.

Rendering: A process of producing realistic 3D images. There are four main rendering methods used in visualization, that is, ray tracing, volume rendering, radiosity, and polygon rendering.

Visual Reconstruction: Computing for each corresponding point pair in the images a point in space.

Human Motion Tracking and Recognition

Niki Aifanti
Informatics & Telematics Institute, Greece

Angel D. Sappa
Computer Vision Center, Spain

Nikos Grammalidis
Informatics & Telematics Institute, Greece

Sotiris Malassiotis
Informatics & Telematics Institute, Greece

INTRODUCTION

Tracking and recognition of human motion has become an important research area in computer vision. In real-world conditions it constitutes a complicated problem, considering cluttered backgrounds, gross illumination variations, occlusions, self-occlusions, different clothing, and multiple moving objects. These ill-posed problems are usually tackled by simplifying assumptions regarding the scene or by imposing constraints on the motion. Constraints such as that the contrast between the moving people and the background should be high, and that everything in the scene should be static except for the target person, are quite often introduced in order to achieve accurate segmentation. Moreover, the motion of the target person is often confined to simple movements with limited occlusions. In addition, assumptions such as known initial position and posture of the person are usually imposed in tracking processes.

The first step towards human tracking is the segmentation of human figures from the background. This problem is addressed either by exploiting the temporal relation between consecutive frames—that is, by means of background subtraction (Sato & Aggarwal, 2001), optical flow (Okada, Shirai & Miura, 2000), or by modeling the image statistics of human appearance (Wren, Azarbayejani, Darrell & Pentland, 1997). The output of the segmentation, which could be edges, silhouettes, blobs, and so forth, comprises the basis for feature extraction. In tracking, feature correspondence is established in order to locate the subject. Tracking through consecutive frames commonly incorporates prediction of movement, which ensures continuity of motion, especially when some body parts are occluded. For example, when a person is walking, there are some moments when one of the legs occludes the other. So, some techniques try to determine the precise movement of each body part (Sidenbladh, Black & Sigal, 2002), while other techniques focus on tracking the human body as a whole (Okada, Shirai & Miura, 2000). Tracking may be classified as 2D or 3D. 2D tracking consists of following the motion in the image plain either by exploiting low-level image features or by using a 2D human model. 3D tracking aims at obtaining the parameters that describe body motion in three dimensions. The 3D tracking process, which estimates the motion of the body parts, is inherently connected to 3D human pose recovery.

3D pose recovery aims at defining the configuration of the body parts in the 3D space and estimating the orientation of the body with respect to the camera. This work will mainly focus on model-based techniques, since they are usually used for 3D reconstruction. Model-based techniques rely on a mathematical representation of human body structure and motion dynamics. The 3D pose parameters are commonly estimated by iteratively matching a set of image features extracted from the current frame with the projection of the model on the image plane. Thus, 3D pose parameters are determined by means of an energy minimization process.

Instead of obtaining the exact configuration of the human body, human motion recognition consists of identifying the action performed by a moving person. Most of the proposed techniques focus on identifying actions belonging to the same category. For example, the objective could be to recognize several aerobic exercises or tennis strokes or some everyday actions such as sitting down, standing up, walking, running, or skipping.

Next, some of the most recent approaches addressing human motion tracking and 3D pose recovery are presented, while the following subsection introduces some whole-body human motion recognition techniques. Previous surveys of vision-based human motion analysis have been carried out by Cédras and Shah (1995), Aggarwal and Cai (1999), Gavrila (1999), and Moeslund and Granum (2001).

This overview will present some of the techniques proposed in the bibliography, together with their advantages or disadvantages. The outline of this work is as follows. First, a survey about human motion tracking and 3D pose recovery are given. Next, human motion recognition is introduced, followed by a summary of some application works. Finally, a section with future trends and conclusion is introduced.

HUMAN MOTION TRACKING AND 3D POSE RECOVERY

Tracking relies either on monocular or multiple camera image sequences. Using monocular image sequences is quite challenging due to occlusions of body parts and ambiguity in recovering their structure and motion from a single perspective view (different configurations have the same projection). On the other hand, single camera views are more easily obtained and processed than multiple camera views. In Table 1, some recent techniques using only one camera are presented.

In contrast to single-view approaches, multiple camera techniques are able to overcome occlusions and depth ambiguities of the body parts, since useful motion information missing from one view may be recovered from another view. In Table 2, some recent approaches using multiple cameras.

Some currently published papers tackle specifically the pose recovery problem using multiple sensors. A real-time method for 3D posture estimation using trinocular images is introduced in Iwasawa et al. (2000). In each image the human silhouette is extracted and the upper body orientation is detected. Two of the three views are finally selected in order to estimate the 3D coordinates of the representative points and joints. In Rosales, Siddiqui, Alon, and Sclaroff (2001), multiple views are obtained by introducing the concept of "virtual cameras," which is based on the transformation invariance of the Hu moments. One advantage of this approach is that no camera calibration is required.

HUMAN MOTION RECOGNITION

Human motion recognition may be achieved by analyzing the extracted 3D pose parameters. However, because of the extra pre-processing required, recognition of human motion patterns is usually achieved by exploiting low-level features (e.g., silhouettes) obtained during tracking (see Table 3).

APPLICATIONS

Vision-based human body modelling has been used in an extensive spectrum of applications. Some recent works, grouped according to their applications field, are presented in Table 4.

Table 1. Monocular systems

Authors	***Description***
Sminchisescu and Triggs (2001)	A 3D human body model, consisting of tampered superellipsoids, is fitted on the image features by means of an iterative cost function optimization scheme. A multiple-hypothesis approach with the ability of escaping local minima in the cost function is proposed.
Cham and Rehg (1999)	This technique combines a CONDENSATION style sampling with local optimization. A 2D model with underlying 3D kinematics is used. The advantages are its suitability for high-dimensional state-spaces and that the use of discrete features is unnecessary.
Wachter and Nagel (1999)	A 3D model of right-elliptical cones is fitted to consecutive frames by means of an iterated extended Kalman filter. This approach copes with self-occlusions occurring between the legs of a walking person.
Howe, Leventon and Freeman (1999)	Self-occlusions are tackled in this Bayesian system. 3D reconstruction is achieved by establishing correspondence between the training data and the features extracted.
Sidenbladh, Black and Sigal (2002)	A probabilistic approach for modeling 3D human motion for synthesis and tracking, where learning of state transition probabilities is replaced with efficient probabilistic search in a large training set.

Table 2. Multiple-camera systems

Authors	Description
Okada, Shirai and Miura (2000)	3D translation and rotation of the human body is estimated using optical flow, depth, and prediction information. The problem of pose singularities is probabilistically tackled.
Cai and Aggarwal (1999)	Three synchronized cameras are used, but tracking is actually based on one at a time, since the camera providing the best view is always selected. Feature correspondence between consecutive frames is achieved using Bayesian classification schemes associated with motion analysis.
Dockstader and Tekalp (2001)	A distributed real-time platform for tracking multiple interacting people. The features extracted from each camera view are independently processed. The resulting state vectors comprise the input to a Bayesian belief network. Observation fusion follows.
Utsumi, Mori, Ohya and Yachida (1998)	Tracking is decomposed into three sub-tasks (rotation angle estimation, position, and body-side detection) and is based on Kalman filtering. Each sub-task has its own criterion for selecting viewpoints.
Delamarre and Faugeras (2001)	This technique incorporates physical forces to each rigid part of a 3D model consisting of truncated cones. The model's projections are compared with the silhouettes extracted from the image by means of a novel approach. This technique copes with self-occlusions, fast movements, and poor quality images.

Table 3. Human motion recognition

Authors	Description
Ali and Aggarwal (2001)	Continuous human activity (walking, sitting down, bending) in monocular sequences with lateral views of the human body is separated into individual actions and recognized using the angles subtended by the torso, and the upper and lower leg.
Park and Aggarwal (2000)	This technique separates and classifies human interactions (shaking hands, pointing at the opposite person) in indoor monocular grayscale images with limited occlusions. The aim is to interpret interactions by inferring the intentions of the persons.
Sato and Aggarwal (2001)	Classification of nine two-person interactions with motion perpendicular to the camera in outdoor monocular grayscale images. Identification is based on the feature's similarity to an interaction model using the nearest mean method.
Haritaoglu, Harwood and Davis (1998, 2000)	Real-time action/interaction recognition using outdoor monocular grayscale images from a visible or infrared camera. Grayscale textural appearance and shape information of a person are combined to a textural temporal template.
Ayers and Shah (2001)	Action identification in the office (entering the room, using computers, picking up the phone) by using prior knowledge about the layout of the room. Performance is affected if the skin area is occluded, two people get too close, or prior knowledge is insufficient. This approach may be applicable in surveillance systems.

FUTURE TRENDS AND CONCLUSION

The problem of human motion tracking and recognition has become an attractive challenge. The huge amount of articles published during the last few years demonstrates the increasing interest in this topic and its wide range of applications. In spite of this, many issues are still open. Problems such as unconstrained image segmentation, limitations in tracking, development of models including prior knowledge, modeling of multiple person environments, and real-time performance still need to be efficiently solved.

A common limitation in tracking, for example, is that the motion of a person is constrained to simple movements with a few occlusions. Occlusions, which comprise a significant problem yet to be thoroughly solved, may lead to erroneous tracking. The systems should be made robust enough to recover any loss of tracking. Another problem to be totally overcome is the initialization of the pose parameters, which must be automatically performed. Similarly, techniques must be able to automatically self-tune the model's shape parameters even in unconstrained environments.

Table 4. Human body model applications

Application	***Authors***	***Description***
Virtual Reality	Aubel, Boulic and Thalmann (2000)	Models with a high level of detail capable of conveying emotions through facial animation.
	Balcisoy, Torre, Ponedr, Fua and Thalmann (2000)	A combination of virtual reality with computer vision, called augmented reality system; it allows the interaction of real and virtual humans in an augmented reality context.
	Plänkers and Fua (2003)	Flexible framework for video-based modeling using articulated 3D soft objects.
Surveillance Systems	Gavrila and Philomin (1999)	Shape-based pedestrian detection for use in real-time vision on on-board vehicles.
	He and Debrunner (2000)	A robust technique for recognizing individuals from their periodic motions, specifically walking and running. The system can handle individuals independently of their direction, speed, or scene background.
	Yuan, Sun, Varol and Bebis (2003)	Vision-based surveillance system able to combine information from multiple camera units. It can be trained to detect pedestrians, vehicles, pets, and so forth.
User Interface	Weik (2000)	Fully automatic system to create 3D models of persons. By using 16 digital cameras and a shape-from-silhouette approach, 3D models are created. A principal component analysis is utilized to extract the skeleton from the models. The system is conceived for 3D tele-presence applications or multi-player games.
	Cohen, Medioni and Gu (2001)	Real-time 3D body reconstruction, from multiple views, for vision-based perceptual user interface. Highly realistic human body motions can be performed.
	Bobick and Davis (2001)	View-based approach to the representation and recognition of human movement. It is based on a temporal template representation and tested with aerobics exercises sequences.
Medical or Anthropometric Applications	Fouchet (1999)	3D scanner for generating 3D models of the whole human body or part of it. The main purpose of this system is to record and study the evolution of dermatological lesions.
	Durdle, Raso, Hill and Peterson (1997)	Stereovision system consisting of computer graphics and imaging tools for the assessment of human body deformities (e.g., curvature of the spine, axial rotation of individual vertebrae).
	Barron and Kakadiaris (2000)	Four-step technique for estimating humans' anthropometric measurements from a single image.
	Marzani, Calais and Legrand (2001)	System for the analysis of movement disabilities of a human leg during gait.

Future human motion recognition systems must be able to identify human motion even in unconstrained environments. These systems can have many useful applications in areas ranging from everyday life to medicine. Especially, robust real-time systems may offer many benefits. Of course, the reduction of processing time presupposes not only advances on computational techniques, but also improvements on the current technology.

REFERENCES

Aggarwal, J.K. & Cai, Q. (1999). Human motion analysis: A review. *Computer Vision and Image Understanding, 73*(3), 428-440.

Ali, A. & Aggarwal, J.K. (2001). Segmentation and recognition of continuous human activity. *Proceedings of the IEEE Workshop on Detection and Recognition of Events in Video,* Vancouver, Canada.

Aubel, A., Boulic, R. & Thalmann, D. (2000). Real-time display of virtual humans: Levels of details and impostors. *IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on 3D Video Technology, 10*(2), 207-217.

Ayers, D.,& Shah, M. (2001). Monitoring human behavior from video taken in an office environment. *Image and Vision Computing, 19*(12), 833-846.

Balcisoy, S., Torre, R., Ponedr, M., Fua, P. & Thalmann, D. (2000). Augmented reality for real and virtual humans. *Proceedings of the Symposium on Virtual Reality Software Technology,* Geneva, Switzerland.

Barron, C. & Kakadiaris, I. (2000). Estimating anthropometry and pose from a single camera. *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition,* Hilton Head, South Carolina, USA.

Bobick, A. & Davis, J. (20011). The recognition of human movement using temporal templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23*(3), 257-267.

Cai, Q. & Aggarwal, J.K. (1999). Tracking human motion in structured environments using a distributed-camera system. *Transactions on Pattern Analysis and Machine Intelligence*, *21*(12), 1241-1247.

Cédras, C. & Shah, M. (1995). Motion-based recognition: A survey. *Image and Vision Computing, 13*(2), 129-155.

Cham, T.J. & Rehg, J.M. (1999). A multiple hypothesis approach to figure tracking. *Computer Vision and Pattern Recognition, 2*, 239-245.

Cohen, I., Medioni, G. & Gu, H. (2001). Inference of 3D human body posture from multiple cameras for vision-based user interface. *Proceedings of the World Multiconference on Systemics, Cybernetics and Informatics,* USA.

Delamarre, Q. & Faugeras, O. (2001). 3D articulated models and multi-view tracking with physical forces. *Special Issue on Modelling People, Computer Vision and Image Understanding, 81*, 328-357.

Dockstader, S.L. & Tekalp, A.M. (2001). Multiple camera tracking of interacting and occluded human motion. *Proceedings of IEEE, 89*(10), 1441-1455.

Durdle, N., Raso, V., Hill, D. & Peterson, A. (1997). Computer graphics and imaging tools for the assessment and treatment of spinal deformities. *Proceedings of the IEEE Canadian Conference on Engineering Innovation: Voyage of Discovery,* St. Johns, Newfoundland, Canada.

Fouchet, X. (1999). *Body modelling for the follow-up of dermatological lesions*. PhD Thesis. Institut National Polytechnique de Toulouse, France.

Gavrila, D.M. (1999). The visual analysis of human movement: A survey. *Computer Vision and Image Understanding, 73*(1), 82-98.

Gavrila, D.M. & Philomin, V. (1999). Real-time object detection for "smart" vehicles. *Proceedings of the IEEE International Conference on Computer Vision,* Kerkyra, Greece.

Haritaoglu, I., Harwood, D. &. Davis, L.S (1998). W^4: Real-time system for detecting and tracking people. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

Haritaoglu, I., Harwood, D. & Davis, L.S. (2000). W^4: Real-time surveillance of people and their activities. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(8), 809-830.

He, Q. & Debrunner, C. (2000). Individual recognition from periodic activity using hidden Markov models. *Proceedings of the IEEE Workshop on Human Motion 2000,* Los Alamitos, California, USA.

Howe, N., Leventon, M. & Freeman, W. (1999). Bayesian reconstruction of 3D human motion from single-camera video. *Proceedings of the Advances in Neural Information Processing Systems 12 Conference.*

Iwasawa, S., Ohya, J., Takahashi, K., Sakaguchi, T., Ebihara, K. & Morishima, S. (2000). Human body postures from trinocular camera images. *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition,* Grenoble, France.

Marzani, F., Calais, E. & Legrand, L. (2001). A 3-D marker-free system for the analysis of movement disabilities-an application to the legs. *IEEE Transactions on Information Technology in Biomedicine*, *5*(1), 18-26.

Moeslund, T.B. & Granum, E. (2001). A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding, 81*(3), 231-268.

Okada, R., Shirai, Y. & Miura, J. (2000). Tracking a person with 3D motion by integrating optical flow and depth. *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition,* Grenoble, France.

Park, S. & Aggarwal, J.K. (2000). Recognition of human interaction using multiple features in grayscale images. *Proceedings of the 15th International Conference on Pattern Recognition*, Barcelona, Spain.

Plänkers, R. & Fua, P. (2003). Articulated soft objects for multi-view shape and motion capture. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 25*(9), 1182-1188.

Rosales, R., Siddiqui, M., Alon, J. & Sclaroff, S. (2001). Estimating 3D body pose using uncalibrated cameras. *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, Kauai Marriott, Hawaii, USA.

Sato, K. & Aggarwal, J.K. (2001). Tracking and recognizing two-person interactions in outdoor image sequences. *Proceedings of the IEEE Workshop on Multi-Object Tracking,* Vancouver, Canada.

Sidenbladh, H., Black, M.J. & Sigal, L. (2002). Implicit probabilistic models of human motion for synthesis and tracking. *Proceedings of the European Conference on Computer Vision,* Copenhagen, Denmark.

Sminchisescu, C. & Triggs, B. (2001). Covariance scaled sampling for monocular 3D body tracking. *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition,* Kauai Marriott, Hawaii, USA

Utsumi, A., Mori, H., Ohya, J. & Yachida, M. (1998). Multiple-view-based tracking of multiple humans. *Proceedings of the 14th International Conference on Pattern Recognition*, Brisbane, Queensland, Australia.

Wachter, S. & Nagel, H. (1999). Tracking persons in monocular image sequences. *Computer Vision and Image Understanding*, *74*(3), 174-192.

Weik, S. (2000). A passive full body scanner using shape from silhouettes. *Proceedings of the 15th International Conference on Pattern Recognition*, Barcelona, Spain.

Wren, C., Azarbayejani, A., Darrell, T. & Pentland, A. (1997). Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence,19*(7), 780-785.

Yuan, X., Sun, Z., Varol, Y. & Bebis, G. (2003). A distributed visual surveillance system. *Proceedings of the IEEE International Conference on Advanced Video and Signal-Based Surveillance*, Reno, Nevada, USA.

KEY TERMS

Image Feature: A structure in the image with interesting characteristics. Examples of features are single points, curves, edges, lines, surfaces.

Occlusion: When one object is in front of another object in the direction of observation, a portion of the object that is behind cannot be seen. Then, the second object is occluded by the first one.

Real-Time System: A system that processes and updates information always within a given time.

Self-Occlusion: When a part of the object is occluded by another part of itself. For example, when a person is walking, one leg may occlude the other leg or the torso may occlude one arm.

Stereo Vision System: System devised to extract 3D information of a given scene. Allows recovering a 3D property such as surface shape, orientation, or curvature. In binocular stereo systems, two images are taken from different viewpoints, allowing the computation of 3D structure. In trifocal, trinocular, and multiple-view stereo, three or more images are available.

Tracking: The process of estimating the parameters of a dynamic system by means of measurements obtained at successive time instances. An example is the estimation of the position of a moving object based on an image sequence.

Training Set: A set of known, labelled examples used in classification. The training set is representative of the data that will be classified in the application.

Vision Surveillance: Computer vision application area that involves scene monitoring and activity detection.

Hyper Video for Distance Learning

Mario Bochicchio
University of Lecce, Italy

Nicola Fiore
University of Lecce, Italy

INTRODUCTION

In general, the production of hypermedia applications is a complex and expensive task, requiring both technical skills and communicative abilities (Bochicchio, Paiano & Paolini, 1999a, 1999b). Nevertheless, some specific kinds of multimedia production can give good quality results, even without specialized IT skills, at a low cost. We have concentrated on this particular field, with the goal of supplying a valid tool to teachers who want to publish their educational material easily and at a low cost.

It is easy for a good teacher to give a lesson and to explain concepts using images and slides to show objects, to write on the blackboard, and to use his body language to grab and hold the attention of his students.

In our opinion, these kinds of lessons can be effortlessly transformed into very usable and effective multimedia applications based on the video of the lesson, on a simple and regular navigation structure, and on a little set of user-friendly multimedia objects.

BACKGROUND

Various research and commercial tools, such as GRiNS (2001), MTEACH (Montessoro & Caschi, 1999), Video Madeus (Roisin, Tran-Thuong & Villard, 2000), and Real Presenter (PresenterPlus, 2001), are based on this assumption, but their effectiveness is limited by a number of issues:

- their technical complexity makes them unsuitable for a large number of teachers with low technical aptitude;
- in general, they are more data-driven than user-centered;
- the time and the budget needed for a non trivial production (e.g., a course of 10 hours or more) can be remarkable; and
- they are often limited to specific lesson styles (e.g., a frontal lesson based on MS PowerPoint presentations).

Moreover, it is well known that long video sequences (e.g., 1 hour or more) are not compelling and not interactive, and the usual linear cursors and VTR-like controls can be ineffective for navigating video sequences longer than a few minutes.

To solve these problems we created LEZI, an experimental tool oriented to the very easy production of video clips enriched with hierarchical indexes, hyper-textual elements and other multimedia objects (hypervideos).

LEZI PROJECT: REQUIREMENTS

An accurate analysis of both research and commercial tools permitted us to extrapolate the essential requirements of a good development environment based on indexed video.

Starting from these requirements, a LEZI prototype was developed at the Hypermedia Open Center (HOC) of the *Politecnico di Milano*, and a number of real lessons were produced and tested (Bochicchio, Paiano, Paolini, Andreassi & Montanaro, 2000). A project for a more complete prototype, called LEZI II, was then started at the SET-Lab of the University of Lecce, within a large research project focused on the development of innovative educational tools and applications.

The first fundamental requirement for LEZI is that it be very easy to use, so that it can be truly accessible even to users with very basic computer knowledge.

The second, even more important requirement is to keep production times down (ideally to about one hour of work or less for each hour of the lesson). In some cases (e.g., conferences or special events), it may be important to extend this constraint up to the “real time production” limit (i.e., the indexed hypervideo of the event should be available on CD/DVD, and online, by the end of the event itself!).

A third very important requirement is the ability to effectively support the most common “authoring situations”, like those in which a teacher:

- presents his lesson in a classroom, with a blackboard, or outside the classroom (on the field), if this is appropriate for the topic concerned;
- uses gestures to “animate” some concept expressed by “static schema” (typically a slide), so that students need to simultaneously view the two different information sources (the teacher and the schema);
- uses his PC to explain how to use a specific computer program when the attention focus is on the display of the PC, on the voice of the teacher and, optionally, on a blackboard; and
- uses his PC to make a PowerPoint presentation. The attention focus is on the display of the PC and on the voice of the teacher.

The fourth requirement relates to finding the various topics and subtopics in the lesson. The user needs a fast and effective way to find out the contents of the video lesson, so they can easily find and reach the subjects of interest without wasting time on uninteresting or already-known video sequences.

We maintain that the most common video players (Real Player, Microsoft Media Player and QuickTime player) generally do not offer an adequate solution to this problem.

The fifth requirement concerns the technical skills needed in the authoring phase; it is important to have a high-level authoring tool to simplify all technical tasks and to fully support teachers and lecturers, whatever technical knowledge they may have.

A final requirement concerns the possibility of linking suitable comments, bibliographic references, and other teaching materials to the indexed hypervideo. The most common digital document formats (PDF, HTML, PPT, etc.) should be supported.

CONCEPTUAL MODELING

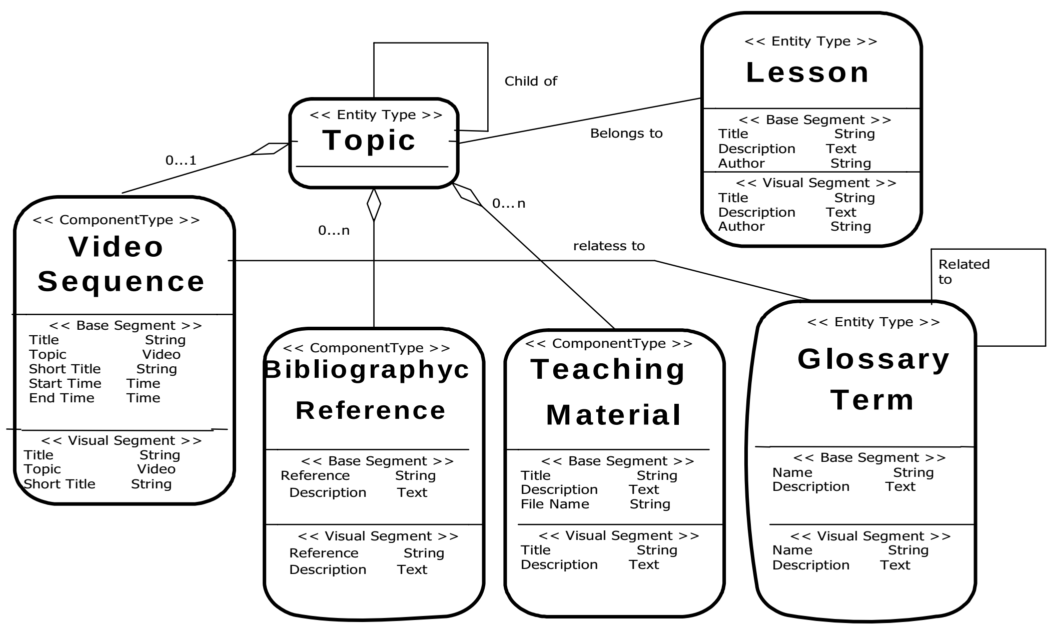
The W2000 (Baresi, Garzotto & Paolini, 2001) methodology has been adopted to refine the informal description presented so far, to obtain a suitable conceptual model for LEZI II, and to derive from it the current LEZI II prototype.

W2000 is a user-centered methodology for conceiving and defining hypermedia applications. It organizes the overall development process into a number of interdependent tasks. Each activity produces a set of related diagrams which describes some aspects of the hypermedia application, and is based on UML. The idea underlying W2000 is a requirements-driven, user-focused approach to design.

In brief, for the LEZI prototype we have identified the following roles:

- *Author:* manages his public/private lessons and related students;
- *Registered Student:* attends public/private lessons and can perform second level authoring (co-authoring) operations (Garzotto, Mainetti & Paolini, 1995).

Figure 1. LEZI II: Hyperbase in the large



- *Unregistered Student:* can only attend public lessons; and
- *LEZI Manager:* manages the system.

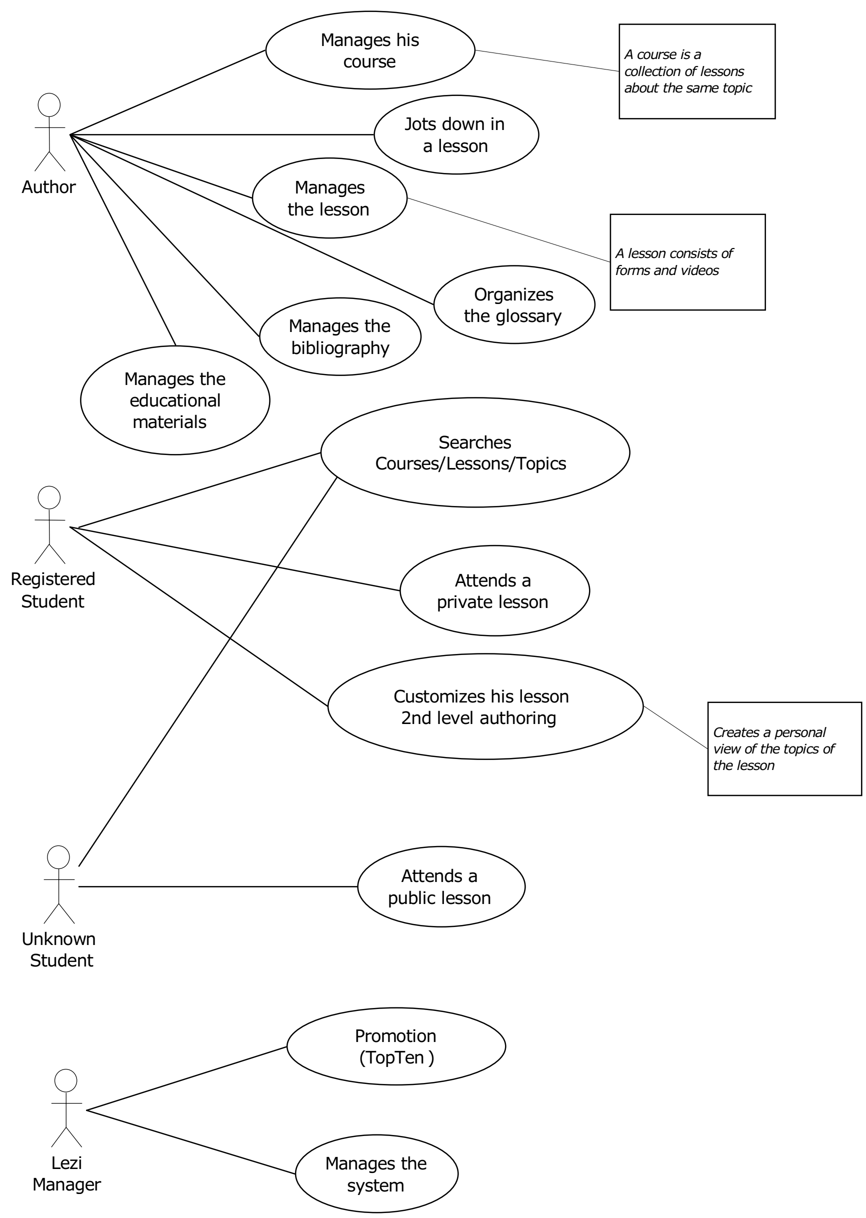
It should be observed that the users of the LEZI II system are not rigidly associated with a single role. A registered student of a given lesson, for example, could also play the role of author for a different lesson. Specifying roles is the best way to make user profiles explicit and to avoid duplicating functionalities.
In Figure 1, the hyperbase diagram of LEZI II is outlined in terms of HDM2000 primitives (Baresi, Garzotto & Paolini, 2001).

The hyperbase schema is adopted to specify:

- the information structures needed by the various classes of users (information design); and
- the navigation paths that allow users to find the piece of information suitable for their task (semantic navigation design).

In Figure 2, we show the main functional use-case diagrams, in which the main functionalities are associated with the previously identified roles.

Figure 2. W2000 functional use-case diagram of LEZI II



THE LEZI II PROTOTYPE

Different LEZI prototypes have been produced since July 2001 at the SET-Lab of the University of Lecce (*http://mb.unile.it/Lezi*).

Referring to the fourth requirement, in all prototypes the main access structure has been implemented as a tree, organized into topic and subtopic nodes. Each topic node corresponds to the sequence of the videos associated with its subtopics, and the root corresponds to the entire lesson. No more than four subtopic levels are allowed, and each leaf of the tree corresponds to 2-5 minutes of video. Each node (both topic and subtopic) of the tree contains a short textual description of the video associated to that node and the indication of its duration. This short description is very effective for finding the interesting topics and skipping the uninteresting (or the already-familiar) ones.

The tree-index acts as a hierarchical table of contents (TOC in the following). It can be generated manually or semi-automatically by creating a sequence of nodes equally-spaced in time. The authors of the lesson can then add/delete/modify the text associated with each node, as well as its duration and its start/end time.

Multiple tree-indexes can be created for a given lesson, so that the same lesson can be easily re-adapted for different purposes and different users.

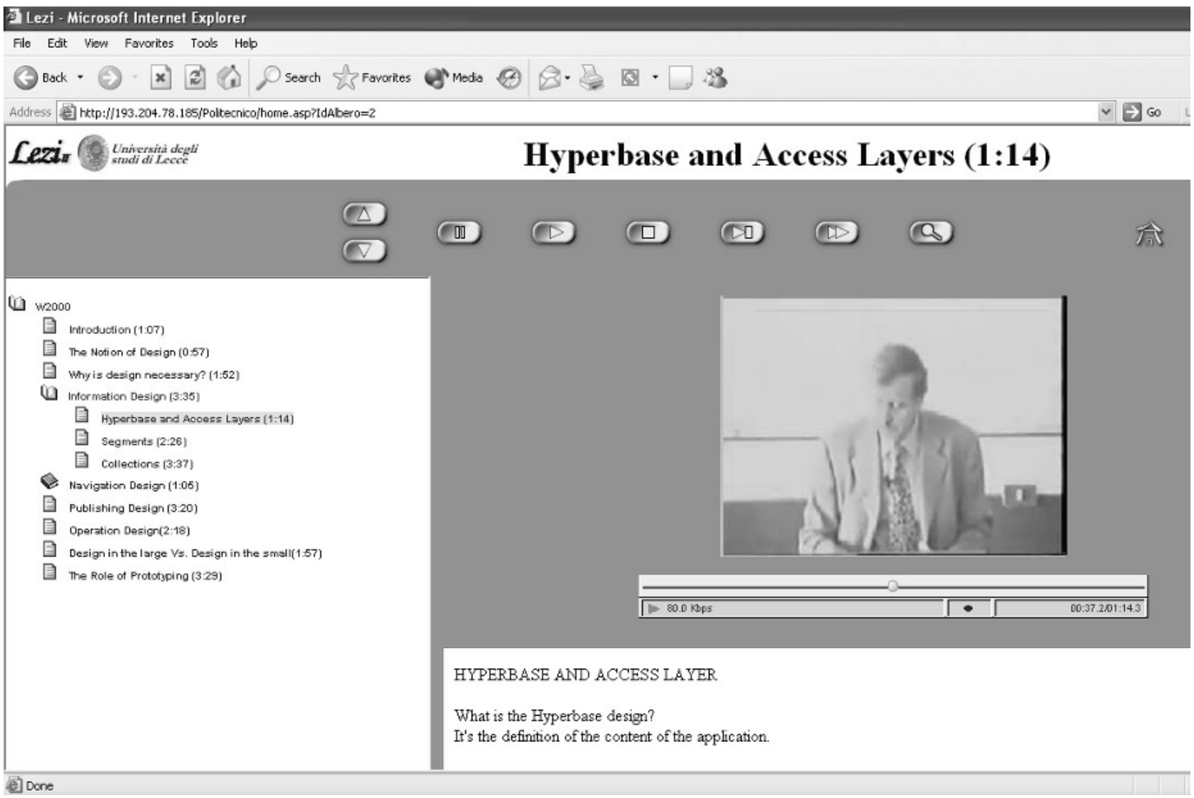
The prototype has two distinct parts: the authoring part, suitable for creating a new LEZI lesson, and the fruition part which may be used to navigate among existing lessons and to select and play the desired one.

From the technical point of view, the online LEZI environment requires a networked workstation, equipped with RealServer (for the video streaming) and Internet Information Server (for the lesson server), while the online LEZI client, suitable for both authoring and fruition, can be executed in any browser supporting JavaScript and equipped with the RealVideo plug-in.

Different user interface styles (multi-skin) and a customizable set of interface objects (background, buttons, colors, fonts, etc.) are supported to better adapt each LEZI II lesson to the expected audience.

From the implementation point of view, the online version is based on the MicroSoft-ASP object model for server-side scripting and on JavaScript and DHTML to implement the visual interface for the client. A SMIL program (SMIL, 2001) has been used to correctly synchronize the tree-index with the video streams.

Figure 3. The “Hyperbase and Access Layers” topic



In comparison with the MicroSoft-ASF format the RealVideo format was more reliable and performed better. In particular, with RealVideo it was very simple and effective producing video clips for multiple bandwidth targets, able to automatically switch to lower/higher bit-rates according to network conditions.

The first step in creating a lesson with the LEZI prototype is to produce the movie in RealVideo format. The lesson can be recorded on a PC or on a notebook equipped with RealProducer by means of a USB video-converter and a video source. If the lesson is already in digital format, they will have to be converted into the RealMedia format before other steps.

Further steps to produce the lesson with LEZI II, are:

- to create a tree index at the end of the recording session;
- to add teaching materials (if available as digital documents) and bibliographical references to the indexed video; and
- to generate the LEZI II lesson (both for CD/DVD and for online use).

VALIDATION

The University of Lecce has produced many applications with LEZI II, to empirically validate the proposed approach.

An example is the computer-graphics class given by Professor Paolo Paolini (Milano-Lecce-Como, 2001). The screen shot in Figure 3 is related to a group of topics on the design/modeling methodology W2000. The class is given to students at the university level.

In order to provide an example of TOC, let us consider the topic "Information Design" that is structured into sub-topics: "Hyperbase and Access Layers", "Segments", and "Collections".

The user is free to "attend" the lesson starting from any point he prefers.

For instance, if the paragraph "Information Design" is chosen, it is not necessary to run the video at higher hierarchical levels, which can be skipped over to go directly to the selected node.

The described LEZI application, that is very cheap and easy to produce, is used to support the normal ("in presence") teaching activity and is considered extremely useful from all involved students.

FUTURE TRENDS

The described idea is very simple: it is possible to publish good educational multimedia applications developed by academic staff with very little technical effort, in a short time, and with limited financial resources.

The future trends in this field come around to support the re-use of existing contents on the net. To better support the LEZI philosophy, it is important the adoption of a content-sharing model and the opening to the standards given by Advance Distance Learning (ADL, 2003) or IEEE (2003) to obtain a common interface for all the learning systems.

CONCLUSION

LEZI philosophy enables teaching staffs without specific technical preparation in multimedia production but with valid content and good teaching skills, to easily prepare good interactive multimedia lessons, both for disk-based (CD/DVD) or online (Web) purposes.

More generally, the widespread use of LEZI or other similar tools can effectively support the development and use of educational multimedia content in universities and schools.

Obviously, this kind of multimedia content is not intended to replace the publications of professional editors.

REFERENCES

Advanced Distributed Learning. (2003). Retrieved August 2003, from *http://www.adlnet.org*

Baresi, L., Garzotto, F., & Paolini P. (2001). Extending UML for modeling Web applications. *Proceedings of the 34th Hawaii International Conference on System Sciences (HICSS'01)*, Maui, USA.

Bochicchio, M.A., Paiano, R., & Paolini, P. (1999a). JWeb: An HDM environment for fast development of Web applications. *Proceedings of the Multimedia Computing and Systems (IEEE ICMCS '99), 2*, 809-813.

Bochicchio, M.A., Paiano, R., & Paolini, P. (1999b). JWeb: An innovative architecture for Web applications. *Proceedings of the IEEE ICSC '99*, Hong Kong.

Bochicchio, M.A, Paiano, R., Paolini, P., Andreassi, E., & Montanaro, T.(2000). LEZI uno strumento per un facile sviluppo di video interattivi a scopo educativo. *Proceedings of the DIDAMATICA 2000*, Cesena, Italy (pp.72-78).

Garzotto, F., Mainetti, L., & Paolini, P. (1995). Hypermedia application design: A structured approach. *Designing user interfaces for hypermedia*. Springer Verlag.

GriNS. (n.d.). Retrieved August 2000, from *www.oratrix.com/GRiNS*

IEEE. (2003). Draft standard learning object metadata. Retrieved August 2003, from *http://www.ieee.org*

JMF. (n.d.). Retrieved August 2000, *from* http://www.javasoft.com/products/java-media/jmf/index.html

Montessoro, P.L., & Caschi, S. (1999). MTEACH: Didactic multimedia production. *Proceedings of the Multimedia Computing and Systems 1999 (IEEE ICMCS '99), 2,* 1017-1019.

PresenterPlus. (n.d.). Retrieved August 2000, from *http://www.realnetworks.com/products/presenterplus*

RealServer Guide. (n.d.). *RealNetworks 1995-2000* (chs. 4-5).

Roisin, C., Tran-Thuong, T., & Villard, L. (2000). A proposal for a video modeling for composing multimedia document. *Proceedings of theMMM2000*, Nagano, Japan.

SMIL. (n.d.). Retrieved August 2000, from *http:// www.w2.org/Audio/Video*

Windows Media Technologies. (n.d.). Retrieved June 2002, from *http://www.microsoft.com/windows/windowsmedia/overview/default.asp*

KEY TERMS

DVD (Digital Versatile Disc): An optical disc technology that is expected to rapidly replace the CD-ROM disc (as well as the audio compact disc) over the next few years. The digital versatile disc (DVD) holds 4.7 gigabyte of information on one of its two sides, or enough for a 133-minute movie.

HDM: The modeling language used by W2000 to describe the information, navigation and presentation aspects of a hypermedia application.

Hypervideo: Indexed video enriched with hypertextual and multimedia elements. It is a fast and effective way to "navigate" in long video clips and to find out the main contents of the video.

USB (Universal Serial Bus): A plug-and-play interface between a computer and add-on devices (such as audio players, joysticks, keyboards, telephones, scanners, and printers). With USB, a new device can be added to a computer without having to add an adapter card or even having to turn the computer off

Video Adapter: An integrated circuit card in a computer or, in some cases, a monitor that provides digital-to-analog conversion and a video controller so that data can be sent to a computer's display (alternate terms include *graphics card*, *display adapter*, *video card*, *video board* and almost any combination of the words in these terms).

Video Streaming: A video sequence that is sent in compressed form over the Internet and displayed by the viewer as they arrive

W2000: A user-centered methodology for conceiving and defining, at conceptual level, hypermedia applications.

Hypothetical Reasoning Over Databases

Lei Dang
University of Manchester, UK

Suzanne M. Embury
University of Manchester, UK

INTRODUCTION

In recent years, the term *agile organisation* has been coined to denote an organisation which is able to change its working practices quickly in order to adapt to changing external pressures or to take advantage of new market opportunities that may arise (Perez-Bustamante, 1999). A key element of such agility is fast and reliable decision making; that is, the ability to determine *what* the new organisational behaviour shall be, as well as *how* the changes to it should be affected. In general, an organisation will be faced with several possible strategies for change, each of which has competing strengths and weaknesses. Management must then evaluate and compare each of these to determine which has the best forecasted outcome. This process is often referred to as "what if?" analysis (WiA), since it tries to answer the question: what will the outcome be if we adopt change X? (Codd, Codd, & Salley, 1993)

BACKGROUND

Perhaps the most familiar WiA tool is the spreadsheet, in which the user can enter different values for the parameters of a mathematical model and observe the changing results. For example, a manager might experiment with different discount rates on forecasted sales, to see how profit levels are affected. The applicability of this technique, however, is limited to situations that can be accurately characterised by a simple set of mathematical equations. A far more common context for such decision making is provided by a company database (or data warehouse), which records historical information about the performance of the organisation.

In order to perform WiA in this context, we would like to be able to update the database, to reflect the proposed changes in organisational behaviour, and then to query it, in order to determine the properties of the new state. Since we obviously cannot risk modifying the live data itself in this way, we need to find alternative methods of making these hypothetical changes to data and querying their composite effect. If the proposed change is sufficiently simple, for example, both the change and the query can be made inside a transaction that is aborted once the query result has been obtained. This will preserve the integrity of the live data, but it can have a performance impact on other business processing and it does not allow easy comparison of different scenarios.

An alternative approach that is commonly adopted is to make a copy of the data involved in the WiA scenario and to make the necessary changes to that, instead of to the live data. This is a much better solution if the change or the queries are large/complex, and it also allows a more *ad hoc* style of analysis, in which changes and queries are interleaved, and determined based on the results of the previous set of queries. If data sets are large, however, or if many different strategies are to be compared, then the disk space required can make this approach impractical. The effect of the data extraction on the performance and availability of the live database must also be considered.

Researchers have therefore considered the possibility of embedding facilities for this form of WiA within the DBMS itself, to provide the capability to perform *hypothetical reasoning* over the contents of the database without impacting the integrity or performance of the database as a whole. Such facilities must support both *hypothetical updates* (HUs), which create a pseudostate (called a *hypothetical state*, or HS) based upon the real database state but modified according to the updates requested, and hypothetical queries, which derive properties from hypothetical states.

A range of possibilities exist for implementing such hypothetical reasoning facilities, depending on how the updates and queries are specified, and to what degree the hypothetical states are materialised or not. In general, the proposals made in the literature so far can be divided into two broad categories: those which support extensional representations of HUs and those which support intensional representations of HUs. The former approach is based around the notion of a hypothetical relation (HR), while the second is focussed on the provision of hypothetical querying facilities. In the remainder of this article, we will present an overview of the principal contributions in both categories, followed by a brief discussion of some additional approaches that might be investigated in the future.

HYPOTHETICAL RELATIONS

The earliest proposal for hypothetical reasoning facilities for databases was based upon the notion of a HR (Stonebraker & Keller, 1980). An HR appears to the query system to be a normal relation, but its contents are in fact defined in terms of the contents of another relation (typically an actual stored relation), with some user-specified additions and deletions. The HR itself does not store data, but instead contains the additions and deletions that have been hypothetically made to the underlying database. Thus, if Add_R is the set of tuples that are to be hypothetically added to a relation R, and Del_R is the set of tuples that are to be hypothetically deleted from R, then the tuples present in the new hypothetical version of R are exactly those given by the query:

$$(R \text{ UNION } Add_R) \text{ DIFFERENCE } Del_R$$

Further queries can be posed against the HR (or a mixture of as many hypothetical and actual relations as are required) in order to understand the ramifications of a particular change in working practices.

Various methods of implementing HRs have been proposed, each with different performance and disk space characteristics. However, the basic concept is to store the details of the HUs in a special auxiliary relation (or relations), for use later in reconstructing the HR itself. In the earliest proposals (Stonebraker & Keller, 1980), a special relation called a *differential file* (DF) is created whenever a new HR is requested by the user (i.e. by making a HU to another relation). The columns of the DF are exactly the same as those of the relation that is being hypothetically updated, with the addition of a special *TupleID* column, which is used to match deleted tuples in the DF with their counterparts in the underlying relation, and a column indicating whether the tuple is an addition to or a deletion from the main relation.

The simplest method of querying an HR represented as a DF is to use a query rewriting approach (Stonebraker, 1981), in which any query involving HRs is transformed into an equivalent query that references only actual relations and DFs. This is done by replacing any references to an HR with a query similar to that given above to describe the contents of an HR. However, this approach (in conjunction with this simple DF design) has several disadvantages. The DF tracks all the updates made to a particular HR, and it may therefore be the case that a tuple in the underlying (actual) relation may appear several times in the DF, if (for example) it has been inserted and deleted many times. This means that the DF can grow much larger than the underlying relation itself, if HUs are numerous and frequent. A second disadvantage is that tuples which are deleted and then re-inserted into the same HR will not be picked up by the query mechanism, because all hypothetical deletions are always enacted after all hypothetical insertions. A later revision of the algorithm was proposed, which uses timestamps on tuples in the DF in order to allow the most recent changes to be visible in the HR (Woodfill & Stonebraker, 1983).

Later, these ideas were developed and extended to produce a more sophisticated form of HR, called an *Independently Updated View* (IUV) (Ramirez, Kulkarni, & Moser, 1991). IUVs are virtual relations defined over a conventional relational view (Kulkarni & Ramirez, 1997). As with HRs, HUs made to an IUV will not affect its *parental views* (i.e. the views or relations from which the IUV is built). The storage mechanism for HUs to IUVs is similar to the DF approach, except that IUVs can be materialised for higher querying performance, or stored in a separate auxiliary relation (called the *differential table*, or DT) for greater flexibility and speed of update. A further difference is that only accumulated net effects of updates are stored in the DT, so that IUVs require much less storage than HRs in the case of frequent repeated updates.

An additional problem which is present in HRs but resolved in IUVs is that of overlapping updates. These occur when the parent relations of an IUV (or HR) are updated normally (i.e. not through the hypothetical update facility). For example, suppose we have hypothetically added a new employee called Fred to a *Workers* relation, and after this the relation is actually updated to contain an employee called Fred. Which of the new Fred tuples should the HR contain? In general, the answer to this question will be application dependent. In some cases, we will wish to give priority to the real updates, while in other cases we will wish to block out any further changes and concentrate only on the HUs. In order to allow this, the designers of the IUV approach provide two options for hypothetical querying, an IUV-prioritised approach, which ignores changes to the parental views, and a Parental-View-prioritised approach, which forces a re-evaluation of the IUV in the light of the updates to the parental view.

HYPOTHETICAL QUERIES

The HR approach concentrates on the problems of representing and processing hypothetical updates in the form of sets of tuples to be added or deleted. However, many forms of WiA are more conveniently expressed using an intensional update specification. For example, rather than adding or deleting specific tuples, we often wish to ask questions such as, “what will be the net effect on profits if we increase the price of all non-fiction books by 2%?” This kind of a question is an example of a common form of hypothetical reasoning, important in many aspects of science and industry, called the *counterfactual condition* (or just *counterfactual*, for short) (Lewis, 2001). A counterfactual is a statement of the form “if A then B”, where A is a truth-valued statement that is contradicted by the current state of the world and B is any truth valued statement. For example, the statement “If Mars were the third planet from the Sun, there would be life on Mars” is a counterfactual. The counterfactual itself expresses the notion that if the world were changed only in ways necessary to make A become true, then B is also true in that updated world. The usual application of such counterfactuals is to determine the value of B for some given starting world, or across all possible worlds, and several languages have been designed for querying with counterfactuals, such as *Hypothetical Datalog* (Bonner, 1990) and *VCU*^{2} (Lewis, 2001).

The Heraclitus system (Ghandeharizadeh, Hull, & Jacobs, 1996) was inspired by previous work on both counterfactuals and the IUV system described earlier. It provides a mechanism for specifying and answering hypothetical queries. A hypothetical query in Heraclitus takes the form “Q when U”, which means “what would be the result of the query Q if it were executed against the current database state *after* it has been updated by the update U?” The key contribution of Heraclitus is that deltas – the descriptions of hypothetical changes required to the database – are elevated to be first class citizens. An extension of the C language is provided for expressing queries and deltas (including the hypothetical *when* operator given earlier), which allows hypothetical queries to be freely mixed with ordinary data access queries, and the compositional specification of deltas from other deltas. Different hypothetical states can be created and compared within a single expression, thanks to the ability to express deltas that exist across multiple transactions (similar to the characteristics of a static variable occurring across many scopes in a traditional programming language) so that different strategies can be easily evaluated against one another.

Two contrasting approaches to the evaluation of hypothetical queries have been proposed (Griffin & Hull, 1997). One of these, the “eager” approach causes hypothetical states to be fully materialized as they are encountered during evaluation of the hypothetical query, so that query operators can be evaluated directly in terms of the materialized state. In the other approach, “lazy” evaluation, materialization of hypothetical states is delayed as far as possible, and queries are transformed relative to the delta so that they can be evaluated directly in terms of the underlying state.

Each approach has its own strengths and weaknesses, and each approach is the most efficient in certain circumstances. For example, if many different hypothetical states are concerned in a relatively simple query then lazy evaluation may be best, since we will not need to materialize many states, most of which are not concerned in the query. However, if many queries are to be evaluated against the same hypothetical state, then it may be better to materialize it once, and use it for all query evaluation, rather than having to go through the complex process of transforming each query relative to the same delta. Because neither approach is best in all circumstances, the authors of Heraclitus have also investigated mechanisms for combining both approaches that allow useful intermediate results to be “cached” (i.e. materialized) while other elements of the query are handled by lazy evaluation.

Although Heraclitus has many advantages over HRs, it shares the problems of the earlier approaches concerning the difficulties of overlapping updates. In general, if the underlying database changes, then it will be necessary to re-evaluate all affected hypothetical queries, whether they have been lazily or eagerly evaluated.

FUTURE TRENDS

As yet, little of the technology described in this article has emerged into commercial database systems. Researchers have (sensibly) concentrated on backend implementation issues, and the attempt to design efficient and straightforward algorithms for query processing in the presence of hypothetical updates. However, there is also need for consideration of the user-interface aspects of hypothetical reasoning, including issues such as correct handling of overlapping updates as well as congenial graphical user interfaces for the creation, management and comparison of multiple hypothetical scenarios.

A number of back-end implementation issues also require further consideration. It is easy to envisage situations in which HRs are a more appropriate supporting technology than hypothetical queries, and *vice versa*, since each takes a slightly different approach to the representation of deltas. HR-based approaches are more suitable when the change to be investigated is best described as additions or deletions of large amounts of extensional data. An example might be when details of sales for some past year are copied forward to represent sales for a future year, without much modification. Hypothetical queries, on the other hand, are more appropriate when the change is most easily described intensionally, as some computation over existing data. For example, if we wish to amend the sales data as we copy it (perhaps increasing all prices to take into account of inflation, or modifying the quantities ordered to take into account forecasted changes in demand for different products), then this may be more easily specified using a hypothetical query approach.

However, the two approaches have much in common, and many real world WiA situations will have elements that are suited to both forms of HU. Ideally, then, we would like some hybrid of the two, which combines the strengths of both approaches, while avoiding their weaknesses. One possible means of achieving such a hybrid is to adapt and exploit new technology in the area of constraint databases (CDBs).

A CDB is a database management system that is extended so that constraint formulae can be used as a basic data type (Brodsky, 1997; Revesz, 2002); that is, the set of tuples in a table can be specified not only extensionally, as in conventional database management systems, but also intensionally, in terms of the constraints that the tuples present in the table must satisfy. As well as allowing very large, or even infinite, data sets to be represented finitely and efficiently, CDBs also allow us to describe new data sets in terms of existing data sets in far richer ways than are allowed by traditional view or derived data mechanisms (Revesz, 2002). This suggests that they may be suitable technology for supporting more advanced forms of hypothetical reasoning for WiA. CDBs have been mainly applied in geographical information systems, but there is an emerging trend in recent years where researchers are attempting to use CDB techniques in other fields, such as genome information systems (Ramanathan & Revesz, 2004), data mining (Leung, 2004) and also the integration of CDBs with traditional relational databases (Cai, 2004). Since several prototype CDB systems, such as MLPQ (P. Revesz et al., 2000), DISCO (Byon & Revesz, 1996) and PReSTO (Cai, Keshwani, & Revesz, 2000) are now available, the time is ripe for an exploration of this area.

CONCLUSION

Hypothetical reasoning over databases greatly depends on the availability of technologies that can maintain a number of hypothetical states simultaneously. With the help of HRs and hypothetical queries, researchers can create tools than can be used by decision makers to act out their imagined scenarios based on various strategies for change and improvement. Though as yet immature, such technology is a natural successor to current work on data warehousing, which collects together the raw material for scenario building in one easy-to-access environment.

REFERENCES

Bonner, A. J. (1990). Hypothetical datalog: Complexity and expressibility. *Theoretical Computer Science, 76*(1), 3-51.

Brodsky, A. (1997). Constraint databases: Promising technology or just intellectual exercise? *Constraints, 2*(1), 35-44.

Byon, J. H., & Revesz, P. Z. (1996). DISCO: A constraint database system with sets. *Constraint Databases and Their Applications, Lecture Notes in Computer Science, 1034*, 68-83.

Cai, M. (2004). Integrating constraint and relational database systems. Paper presented at the *The 1st International Symposium on Applications of Constraint Databases*, Paris, France.

Cai, M., Keshwani, D., & Revesz, P. Z. (2000). Parametric rectangles: A model for querying and animation of spatiotemporal databases. *Lecture Notes in Computer Science, 1777*, 430-446.

Codd, E. F., Codd, S. B., & Salley, C. T. (1993). *Providing OLAP (On-Line Analytical Processing) to user analysts: An IT mandate*. Codd & Date Inc.

Ghandeharizadeh, S., Hull, R., & Jacobs, D. (1996). Heraclitus: Elevating deltas to be first-class citizens in a database programming language. *ACM Transactions on Database Systems, 21*(3), 370-426.

Griffin, T., & Hull, R. (1997). A framework for implementing hypothetical queries. *SIGMOD Record, 26*(2), 231-242.

Kulkarni, U. R., & Ramirez, R. G. (1997). Independently updated views. *IEEE Transactions on Knowledge and Data Engineering, 9*(5), 798-812.

Leung, C. K.-S. (2004). Dynamic FP-Tree based mining of frequent patterns satisfying succinct constraints. Paper presented at *The 1st International Symposium on Applications of Constraint Databases*, Paris, France.

Lewis, D. K. (2001). *Counterfactuals*. Malden, MA ; Oxford: Blackwell Publishers.

Perez-Bustamante, G. (1999). Knowledge management in agile innovative organisations. *Journal of Knowledge Management, 3*(1), 6-17.

Power, D. J. (1997). What is a DSS? *The On-Line Executive Journal for Data-Intensive Decision Support*, *1*(3).

Ramanathan, V., & Revesz, P. Z. (2004). Constraint database solutions to the genome map assembly problem. Paper presented at *The 1st International Symposium on Applications of Constraint Databases*, Paris, France.

Ramirez, R. G., Kulkarni, U. R., & Moser, K. A. (1991). The cost of retrievals in what-if databases. Paper presented at the *Twenty-Fourth Annual Hawaii International Conference on System Sciences*, Maui, Hawaii, USA.

Revesz, P. (2002). *Introduction to constraint databases*. New York: Springer.

Revesz, P., Chen, R., Kanjamala, P., Li, Y., Liu, Y., & Wang, Y. (2000). The MLPQ/GIS constraint database system. *SIGMOD Record, 29*(2), 601.

Revesz, P. Z. (1998). Constraint databases: A survey. *Semantics in Databases, Lecture Notes in Computer Science*, *1358*, 209-246.

Stonebraker, M. (1981, June 1982). Hypothetical databases as views. Paper presented at the *ACM-SIGMOD Conference on Management of Data,* Ann Arbor, MI, US.

Stonebraker, M., & Keller, K. (1980, May 1980). Embedding expert knowledge and hypothetical databases into a data base system. Paper presented at the *ACM-SIGMOD Conference on Management of Data*, Santa Monica, CA, US.

Woodfill, J., & Stonebraker, M. (1983). An implementation of hypothetical relations. Paper presented at the *9th International Conference on Very Large Data Bases*, Florence, Italy.

KEY TERMS

Constraint Databases: A generalization of relational databases by finitely representable infinite relations, parameterized by the type of constraint domains and constraints used (Revesz, 1998).

Decision Support System: An interactive computer-based system intended to help managers make decisions. A DSS helps managers retrieve, summarize and analyze decision relevant data. There are five main classes of DSS: Communications-Driven, Data-Driven, Document-Driven, Knowledge-Driven, Model-Driven. These can be further divided by three secondary dimensions: Inter-Organizational vs. Intra-Organizational; Function-Specific vs. General-Purpose, and Web-Based (Power, 1997).

Hypothetical Queries (HQ): Queries against hypothetical states. They take the form Q when U, where U represents a hypothetical state specified by updates and Q is a query over the hypothetical state (Griffin & Hull, 1997).

Hypothetical Reasoning: The process of creating hypothetical scenarios based on hypotheses and hypothetical assertions, and of exploring the hypothesis space for results of given queries.

Hypothetical Relation (HR): A mechanism to allow users to generate alternate versions of real relations. Each version can be updated as if it were a real relation, but only the differences between the HR and the real relation on which it is defined are actually stored (Woodfill & Stonebraker, 1983).

Hypothetical Updates (HU): Proposed updates that will not be applied on the real database, but the hypothetical effect of which we wish to explore.

ICTs as Participatory Vehicles

Marc Holzer
Rutgers University, Newark, USA

Richard Schwester
Rutgers University, Newark, USA

INTRODUCTION

While citizen participation is central to democratic governance, there is a noticeable disconnect between elected representatives and the citizenry, as evidenced by voter apathy, decreased levels of civic participation, and widespread cynicism toward political institutions (Hudson, 2001; Putnam, 2000; Berman, 1997). Citizen participation advocates, however, are optimistic that information and communications technologies (ICTs) will facilitate direct interactions between citizens and government, thereby altering the dynamics of the policy-making process by affording the citizenry a stronger voice. This article examines the integration of ICTs and digital democratic applications in the context of citizen participation in government.

BACKGROUND

The Notion of Digital Democracy

As ICTs have rapidly developed, the public sector has sought to apply these technologies to digital service delivery, and ICTs have also afforded citizens a more direct means of participating in the public decision-making process. We use the term "digital democracy" instead of "electronic democracy" (Browning, 2002), "cyberdemocracy" (Ogden, 1998; Tsagarousianou, Tambini, & Bryan, 1998) or "teledemocracy" (Cross, 1998; Watson, Akselsen, Evjemo, & Aarsaether, 1999). Digital democracy simply encompasses the use of ICTs in the practice of democracy. Jankowski and van Selm (2000) suggested that while teledemocracy is more related to electronic polling and voting, digital democracy is more generally accepted as including activities related to the democratic process.

Democracy is a form of government in which citizens have a measure of influence over the policies that affect their lives. The relationship between government and citizens is foremost within a democratic system. In a digital democracy, the focus is on the processes and structures that define the relationships between government and citizens, between elected officials and appointed civil servants, and between the legislative and the executive branches of government.

According to Hacker and van Dijk (2000), digital democracy refers to "a collection of attempts to practice democracy without the limits of time, space and other physical conditions, using ICTs or computer-mediated communication instead, as an addition, not a replacement for, traditional 'analogue' political practices" (p. 1). In addition, Nugent (2001) referred to digital democracy as "processes carried out online—communicating with fellow citizens and elected representatives about politics" (p. 223). Digital democracy may be defined as all practices to improve democratic values using ICTs. Central to digital democracy are specific governance issues, which include government openness, citizen participation in governing processes, and digital elections (Arterton, 1988, pp. 620–626).

Government openness is central to digital democracy, and openness is predicated on improving access to government information. That is, well-informed citizens are more capable of playing an active role in government. Citizens can make their voices more powerful with well-informed, active participation in the policy-making process. Thus, citizens may be empowered via e-mails to elected officials, as well as by debating social issues in digital forums.

With digitally available information and advanced ICTs, citizens can participate more fully in the governing process and consult on policies at all levels of government. In addition, citizens can discuss social issues and government policies in digital forums that include public officials. The Public Electronic Network in Santa Monica, California, illustrates a public discussion forum. Launched in 1989, the Network enables citizens to interact with public servants. While initially designed to enhance public access to information, issue forums are now common (Docter & Dutton, 1998; Guthrie & Dutton, 1992; O'Sullivan, 1995; Varley, 1991). Further, the Digital City projects in Amsterdam enable citizens and politicians to interact with each other (Brants, Huizenga, & van Meerten, 1996; Francissen & Brants, 1998).

O'Looney (2002) compared the interaction between citizens and government within the context of traditional and digital democracy. While communications are filtered through representatives and the media in a traditional democracy, direct communications among citizens, public managers, and technical staff are now possible in a digital democracy. Therefore, while public servants' communications with citizens involve a one-message-fits-all approach in a traditional democracy, official communications within a digital democracy may be personalized based on an individual's interests and needs. In a digital democracy, citizens can potentially track and influence decision making at every step in the policy-making process, ranging from agenda setting to a final vote, thereby providing a basis for digital deliberation in government.

Digital Deliberation

Deliberation is the process of thoughtful discussion and consideration regarding an issue or course of action. Deliberative processes comprise discussion and consideration of arguments for and against a proposed measure. According to O'Looney (2002, p. 276), digital deliberation in government is characterized by the following:

- Access to balanced information
- An open agenda
- Time to consider issues expansively
- Freedom from manipulation or coercion
- A rule-based framework for discussion
- Participation by an inclusive sample of citizens
- Broader and freer interaction between participants
- Recognition of differences between participants, but rejection of status-based prejudice

In the context of this discussion, it is important to differentiate between static and dynamic digital deliberation. Static digital deliberation might typically include an online poll without public deliberation, a bulletin board for complaints and recommendations, or citizen participation by mail, fax, or e-mail. Dynamic digital deliberation meets the criteria for the public sphere as suggested by Habermas (1989). It includes applications that are two-way or dialogical, such as digital town hall meetings and digital policy forums. The roles of both public servants and professionals are key elements. True public deliberation includes all participants—citizens, politicians, bureaucrats, interest groups, and the media.

Deliberation is necessary when there is uncertainty, and it proves invaluable when choosing between two equally compelling courses of action (Adams et al., 2002). According to Reich (1990), deliberation is a process of "social learning about public problems and possibilities" (p. 8). Participants in deliberative processes are expected to be open to changes in their attitudes, ideas, and positions, although change is not a required outcome of deliberation. Rather, it is a process of fostering citizen growth both "in the capacity for practical judgment and in the art of living together in a context of disagreement" (Adams, et al., 2002; Roberts, 1997, pp. 124–132; Ryfe, 2002, pp. 359-377; Walters et al., 2000, pp. 349-359; Waugh, 2002, pp. 379-382; Weeks, 2000, pp. 360-372; Zifcak, 1999, pp. 236-272).

Deliberation has long been considered an important element of true democracy, and it is central to public realm theory (London, 1995, pp. 33-55). Scholars such as Arendt (1958) and Habermas (1989) regarded the public sphere as "both a process by which people can deliberate about their common affairs, and as an arena, or space, in which this can happen naturally" (London, 1995, p. 41). According to Habermas (1984, 1989), the public sphere includes requirements for authenticity, including "open access, voluntary participation outside institutional roles, the generation of public judgment through assemblies of citizens who engage in political deliberation, the freedom to express opinions, and the freedom to discuss matters of the state and to criticize the way state power is organized" (London, 1995, p. 42).

Habermas provided an historical and sociological description of European social institutions throughout the 17th and 18th centuries. He conveyed the importance of social institutions as mechanisms by which private individuals passed judgment on public acts. The English coffeehouses, the literary societies of Germany, and the salons of France are examples of such institutions, and they proved extremely egalitarian in the sense that "the

Table 1. Digital deliberation

Type of Deliberation	Characteristics
Dynamic digital deliberation (active)	• Digital town hall meeting • Digital policy forum • Online voting with deliberation
Static digital deliberation (passive)	• Online poll (instant results, presentation of previous polls) • Bulletin board for complaints • Bulletin board for recommendations

bourgeois met here with the socially prestigious but politically uninfluential nobles as 'common' human beings" (Habermas, 1989, p. 35). The salons, literary societies, and coffeehouses brought together diverse intellectuals, creating forums for ideas and opinions regarding the state of society.

Further, in accordance with his theory of communicative action, Habermas (1984) described the cultivation of a public discourse that falls outside of the associational matrices relevant to the state, market, and the bureaucracy. The institutions of deliberation (e.g., coffeehouses, literary societies, and salons) served to revive public opinion as a mechanism for shaping policy or influencing government in a meaningful and reasonable manner. The ideal public arena, in accordance with Habermas' writings, fosters inclusive and voluntary citizen participation within the context of influencing how government power is wielded. In such arenas, ICTs may emerge as mainstream conduits for policy deliberations, whereby elected officials, experts, and citizens can come together and voice their opinions.

Ideally, digital deliberation offers ignored groups a greater voice, thereby narrowing the democratic deficit (i.e., the participatory gap in public policy making). Having been applied in a number of cities throughout the United States and Europe (Tsagarousianou, Tambini, & Bryan, 1998), ICTs allow more citizens to participate in the policy discourse of deliberative governance, and they have the potential to reconnect citizens and decision makers, publicizing views presented by consultative parties and providing greater opportunities for citizens to influence public policy.

ICTs are emerging as more mainstream conduits for policy deliberations. A noteworthy example is Regulations.gov, which is "a one-stop Federal regulatory clearinghouse" that claims to facilitate citizen participation in Federal rule making and the American democratic process (http://www.regulations.gov/help.cfm). Citizens can view a description of proposed and final Federal regulations and read the full texts of the regulations. In addition, citizens can submit their comments to the Federal agencies responsible for the rule-making action through the Regulations.gov Web site.

Through the Regulations.gov Web site, citizens obtain descriptions or full text versions of proposed and final regulations for 75 agencies. Simplification and easy access are central components. According to Mark Forman, former Associate Director for Information Technology and E-Government for the Office of Management and Budget, "the guiding principles for achieving our e-government vision are also about simplifying the process and unifying operations to better serve citizen needs; that is, 'uncomplicating' government" (Forman, 2002, p. 2). Forman further stressed that accessing government information "should not take a citizen more than three 'clicks' of a mouse" (Forman, 2003, p. 2). In particular, through the Regulations.gov home page, citizens can identify regulations open for comment by performing a key word search or by selecting a Federal agency from the menu. For example, selecting "Internal Revenue Service" from the agency menu allows individuals to view proposed IRS regulatory changes (in either HTML or PDF format) that are open for public comment. Regulations.gov represents static digital deliberation, as there is no dialogical component.

Similar to Regulations.gov, the Virginia Regulatory Town Hall is a public space online that allows citizens and interested parties to obtain information regarding state regulations and proposed changes. It offers citizens a means of participating in the rule-making process. Presumably, the Town Hall helps the State of Virginia to manage the administrative rule-making process more efficiently and effectively. Information regarding proposed regulations and changes, agency background discussions, economic impact analyses of proposed regulations, and agency guidance documents, all help citizens interpret agency regulations. Those information services help encourage citizen participation in the rule-making process. Like Regulations.gov, the Virginia Regulatory Town Hall is static in nature.

The 21st Century Town Meeting exemplifies dynamic deliberation using ICTs. Developed by America Speaks, a nonprofit organization dedicated to enhancing citizen participation in the public policy process through the application of ICTs, the 21st Century Town Meeting is a digital democratic model designed to facilitate direct interactions between citizens and government. It affords all citizens a means by which they can impact the policy-making discourse. According to Lukensmeyer and Brigham (2002), "public hearings and typical town hall meetings are not a meaningful way for citizens to engage in governance and to have an impact on decision-making. They are speaker focused, with experts simply delivering information or responding to questions" (p. 351). In contrast, the 21st Century Town Meeting enables thousands of people to come together and voice their opinions through the use of networked computers, electronic keypads, and large video screens. Small group dialogues are a central component of the Meeting. Demographically diverse groups of 10 to 12 people are convened to discuss various issues. Each group is guided by a trained facilitator, who ensures that the dialogue is focused and that all participants are heard. Networked computers are used to record and transmit each group's viewpoints to a central computer. The data are coded into themes, and each individual (within his or her respective group) uses an electronic keypad to vote on each of the themes presented. The most visible application of the

Meeting, known as *Listening to the City*, dealt with rebuilding lower Manhattan following the World Trade Center attacks. In July, 2002, approximately 4,300 individuals from New York City and surrounding areas convened at the Jacobs Javits Center to participate in *Listening to the City*. The forum proved to be a dynamic and worthwhile experience, as a significant majority of participants (73%) expressed either a high or very high level of satisfaction with the process (Report of Proceedings, 2002).

FUTURE TRENDS

While ICTs have emerged as participatory mediums, are they conducive to deliberative democracy? The issues raised by this question are complex, abstract, and much more than matters of judgment. There are two broad viewpoints regarding the impact of ICTs on deliberative democracy. First, there are the technological optimists who believe that ICTs are easier and faster, and offer qualitatively better ways of existing, working, communicating, and participating in public life. According to Dahl (1989), "telecommunications can give every citizen the opportunity to place questions of their own on the public agenda and participate in discussions with experts, policy-makers and fellow citizens" (p. 339). Grossman (1995) further holds that the "big losers in the present-day reshuffling and resurgence of public influence are the traditional institutions that have served as the main intermediaries between government and its citizens: the political parties, labor unions, civic associations, even the commentators and correspondents in the mainstream press" (p. 15).

Cross (1998, pp. 139-143) discussed the relationship between ICTs and democracy, focusing on the following democratic norms: informing voters, representativeness, and participation. ICTs play an important role as a mechanism for disseminating government information to citizens (Charlton, Gittings, Leng, Little, & Neilson, 1997; Korac-Kakabadse & Korac-Kakabadse, 1999, p. 216; Langelier, 1996, pp. 38-45; Lips, 1997 recited in Ranerup, 1999, p. 179). McConaghy (1996) argued that publicizing information used in the development of government policies would allow citizens to be more fully involved in the democratic process. Further, in terms of representativeness, ICTs can alert policy makers as to the needs and preferences of the citizenry regarding potential policies. With respect to participation, McLean (1989, pp. 108-110) maintained that ICTs make direct participation possible because they overcome the problems of large, dispersed populations, while Arterton (1987, p. 189) argued that more citizens can participate because many of the burdens of participation are lowered, which increases equity in public decision making (Arterton, 1987, pp. 50-51; Barber, 1984).

The alternative view is less optimistic and is centered on the premise that bringing about change in institutions and behavior patterns is a slow and problematic process. According to Conte (1995, pp. 33-34), "It's so easy to imagine a scenario in which technology is used to get instant judgments from people. If it is used that way, we haven't seen anything yet when it comes to high-tech lynchings…. Real democracy is slow and deliberative." Unless carefully moderated, digital-based forums can become chaotic. Unmediated forums can potentially become abusive and unfocused. Politicians and other community leaders with whom citizens wish to interact are reluctant to participate in digital forums for fear of being "flamed," which refers to losing the luxury of "leading from behind." Then, there is the problem of dealing with the overload of undifferentiated and uncategorized information. In spite of the increasing amounts of information now available, its wide distribution, and the speed with which it is transferred, there is no evidence to suggest that the quality of decision making has improved or that decisions are more democratic given the integration of ICTs and digital-based applications. The current debate regarding the impact of ICTs on deliberative democracy provides the basis for future research opportunities.

CONCLUSION

Although citizen participation is central to a healthy democracy, our political system is characterized by declining voter turnout, decreased levels of civic participation, and cynicism toward political institutions and elected officials. Technological optimists are hopeful that ICTs will reverse those declines by facilitating direct interactions between citizens and government, providing a more effective means of participating within the policy-making process. Citizen participation via ICTs may foster an increased sense of public engagement as an egalitarian device that helps to engender a measure of societal collectivism necessary for a democratic system to thrive. Given the erosion of civic virtue and the proliferation of political apathy, online public spaces may serve to reenergize the body politic and reaffirm the importance of citizen participation.

Despite the potential benefits, the use of ICTs as participatory vehicles presents specific challenges, in particular, the digital divide. The digital divide is a challenge that democratic societies must address, and it refers to the divide between those with Web access and Web-related skills, and those without such capacities. Even though the online population is increasingly reflective of communities offline, the reality of a digital divide means that certain segments of the population are effectively excluded from online deliberation. This divide undermines the Internet as a mainstream and inclusive participatory medium, as it disproportionately impacts lower socioeconomic individuals who have historically played an insignificant role within the public policy process. For example, according to the Federal Reserve Bank of San Francisco, an Hispanic individual with a high school education and a family income less than $15,000 annually has only a 31% chance of using a computer at home, school, or within the workplace. This compares to a 94% chance for a white individual with a graduate education and a family income that exceeds $75,000 per year (FRBSF Economic Letter, 2003).

The issue of technological determinism complicates the digital divide. According to Warschauer (2003, p. 44), technological determinism holds that the "mere presence of technology leads to familiar and standard applications of that technology," which engenders social change. Consequently, significant capital has been spent providing computers and Internet access to those without such resources. And while on the surface this seems like a logical way of addressing the digital divide, inadequate attention has been paid to the "social contexts in which these technologies might be used" (Warschauer, 2003, p. 44).

Moreover, skeptics may argue that Internet-based applications will merely serve as another avenue of influence and control for the politically efficacious and the power elite. In spite of these criticisms and challenges, on balance, digital citizen participation represents a great potential for democratic renewal.

REFERENCES

Adams, G. B., Baker, D. L., Johnson, T. G., Scott, J. K., Richardson, L. E., Wechsler, B., & Zanetti, L. A. (2002). Deliberative governance: Lessons from theory and practice. Paper presented at the 63rd ASPA National Conference, Phoenix, Arizona, March 23-26.

Arendt, H. (1958). *The human condition*. Chicago: University of Chicago Press.

Arterton, C. F. (1987). *Can technology protect democracy?* Thousand Oaks, CA: Sage Publications.

Arterton, C. F. (1988). Political participation and teledemocracy. *PS: Political Science and Politics, 21*(3), 620-626.

Barber, B. (1984). *Strong democracy: Participatory politics for a new age*. Berkeley: University of California Press.

Berman, E. (1997). Dealing with cynical citizens. *Public Administration Review, 57*(2), 105-112.

Brants, K., Huizenga, M., & van Meerten, R. (1996). The canals of Amsterdam: An exercise in local electronic democracy. *Media, Culture, and Society, 18*(2), 233-247.

Browning, G. (2002). *Electronic democracy: Using the Internet to transform American politics*. Medford: CyberAge Books.

Charlton, C., Gittings, C., Leng, P., Little, J., & Neilson, I. (1997). Diffusion of the Internet: A local perspective on an international issue. In T. McMaster, E. Mumford, E. B. Swanson, B. Warboys, & D. Wastell (Eds.), *Facilitating technology transfer through partnership: Learning from practice and research*. London: Chapman & Hall.

Conte, C. R. (1995). Teledemocracy—For better or worse. *Governing*.

Cross, B. (1998). Teledemocracy: Canadian political parties listening to their constituents. In C. J. Alexander & L. A. Pal (Eds.), *Digital democracy: Policy and politics in the wired world* (pp. 132-148). Oxford: Oxford University Press.

Dahl, R. A. (1989). *Democracy and its critics*. New Haven: Yale University Press.

Docter, S., & Dutton, W. H. (1998). The First Amendment online: Santa Monica's Public Electronic Network. In R. Tsagarousianou, D. Tambini, & C. Bryan (Eds.), *Cyberdemocracy: Technology, cities, and civic networks* (pp. 125-151). London: Routledge.

Federal Reserve Bank of San Francisco. (2003). Is there a digital divide? *FRBSF Economic Letter*, Number 2003-38, December 26.

Forman, M. (2002). Statement of Mark Forman, Associate Director for Information Technology and E-Government, Office of Management and Budget before the Subcommittee on Technology and Procurement Policy of the Committee on Government Reform. March 21. Retrieved January 8, 2004, from *http://www.whitehouse.gov/omb/legislative/testimony/mark_forman_032102.pdf*

Forman, M. (2003). Associate Director for Information Technology and E-Government, Office of Management and Budget before the Subcommittee on Technology, Information Policy, Intergovernmental Relations, and the Census, Committee on Government Reform, United States House of Representatives. March 13. Retrieved November 17, 2003, from *http://www.whitehouse.gov/omb/legislative/testimony/forman031303.html*

Francissen, L., & Brants, K. (1998). Virtually going places: Square-hopping in Amsterdam's digital city. In R. Tsagarousianou, D. Tambini, & C. Bryan (Eds.), *Cyberdemocracy: Technology, cities, and civic networks* (pp. 18–40). London: Routledge.

Grossman, L. K. (1995). *The electronic republic: Reshaping democracy in the information age*. New York: Viking.

Guthrie, K. K., & Dutton, W. H. (1992). The politics of citizen access technology. *Policy Studies Journal, 20*, 574-597.

Habermas, J. (1984). *The theory of communicative action. Vol. 1, Reason and the rationalization of society*. Boston, MA: Beacon Press.

Habermas, J. (1989). *The structural transformation of the public sphere*. Cambridge, MA: Harvard University Press.

Hacker, K. L., & van Dijk, J. (2000). What is digital democracy? In K. L. Hacker & J. van Dijk (Eds.), *Digital democracy: Issues of theory and practice* (pp. 1-9). Thousand Oaks, CA: Sage Publications.

Hudson, W. E. (2001). *American democracy in peril: Seven challenges to America's future*. New York: Chatham House Publishers.

Jankowski, N. W., & van Selm, M. (2000). The promise and practice of public debate in cyberspace. In K. L. Hacker & J. van Dijk (Eds.), *Digital democracy: Issues of theory and practice* (pp. 149-165). Thousand Oaks, CA: Sage Publications.

Korac-Kakabadse, A., & Korac-Kakabadse, N. (1999). Information technology's impact on the quality of democracy. In R. Heeks (Ed.), *Reinventing government in the information age: International practice in IT-enabled public sector reform* (pp. 211-228). London: Routledge.

Langelier, P. (1996). Special Series: Local government on the Internet. Part 3: Local government home pages. *Popular Government*, pp. 38-45.

London, S. (1995). Teledemocracy vs. deliberative democracy: A comparative look at two models of public talk. *Journal of International Computing and Technology, 3*(2), 33–55.

Lukensmeyer, C. J., & Brigham, S. (2002). Taking democracy to scale: Creating a town hall meeting for the twenty-first century. *National Civic Review, 91*(4), 351-366.

McConaghy, D. (1996). The electronic delivery of government services. *Comments on the UK Green Paper* (unpublished).

McLean, I. (1989). *Democracy and the new technology*. Cambridge: Polity Press.

Nugent, J. D. (2001). If e-democracy is the answer, what's the question? *National Civic Review, 90*(3), 221-223.

Ogden, M. R. (1998). Technologies of abstraction: Cyberdemocracy and the changing communications landscape. In C. J. Alexander & L. A. Pal (Eds.), *Digital democracy: Policy and politics in the wired world* (pp. 63-86). Oxford: Oxford University Press.

O'Looney, J. A. (2002). *Wiring governments: Challenges and possibilities for public managers*. Westport: Quorum Books.

O'Sullivan, P. B. (1995). Computer networks and political participation: Santa Monica's teledemocracy project. *Journal of Applied Communication Research, 23*(2), 93-107.

Putnam, R. (2000). *Bowling alone: The collapse and revival of American community*. New York: Simon and Schuster.

Ranerup, A. (1999). Internet-enabled applications for local government democratisation: Contradictions of the Swedish experience. In R. Heeks (Ed.), *Reinventing government in the information age: International practice in IT-enabled public sector reform* (pp. 177-193). London: Routledge.

Reich, R. B. (1990). *Public management in a democratic society*. Englewood Cliffs, NJ: Prentice Hall.

Report of Proceedings. (2002, February 7). Listening to the City. A project of the civic Alliance to Rebuild Downtown New York. Retrieved January 8, 2004, from *http://www.icisnyu.org/admin/files/ListeningtoCity.pdf*

Roberts, N. C. (1997). Public deliberation: An alternative approach to crafting policy and setting direction. *Public Administration Review, 57*(2), 124-132.

Ryfe, D. M. (2002). The practice of deliberative democracy: A study of 16 deliberative organizations. *Political Communication, 19*(3), 359-377.

Tsagarousianou, R., Tambini, D., & Bryan, C. (1998). *Cyberdemocracy: Technology, cities and civic networks*. London: Routledge.

Varley, P. (1991). Electronic democracy. *Technology Review, 94*(8), 42-51.

Walters, L. C., Aydelotte, J., & Miller, J. (2000). Putting more public in policy analysis. *Public Administration Review, 60*(4), 349-359.

Warschauer, M. (2003). Demystifying the digital divide. *Scientific American, 289*(2), 42-47.

Watson, R. T., Akselsen, S., Evjemo, B., & Aarsaether, N. (1999). Teledemocracy in local government. *Communications of the ACM, 42*(12), 58-63.

Waugh, W. L., Jr. (2002). Valuing public participation in policy making. *Public Administration Review, 62*(3), 379-382.

Weeks, E. C. (2000). The practice of deliberative democracy: Results from four large-scale trials. *Public Administration Review, 60*(4), 360-372.

Zifcak, S. (1999). From managerial reform to democratic reformation: Towards a deliberative public administration. *International Public Management Journal, 2*(2), 236-272.

KEY TERMS

Deliberative Democracy: Refers to citizen participation in the context of cultivating a public discourse regarding governmental issues, policies, and courses of action.

Digital Deliberation: The process of thoughtful discussion regarding an issue or course of action through the use of ICTs. Digital deliberation in government is characterized by access to balanced information, an open agenda, time to consider issues expansively, freedom from manipulation or coercion, a rule-based framework for discussion, participation by an inclusive sample of citizens, broader and freer interaction between participants, and the recognition of differences between participants.

Digital Democracy: Encompasses the use of ICTs in the practice of democracy, whereby emphasis is placed on the processes and structures that define the relationships between government and citizens, between elected officials and appointed civil servants, and between the legislative and the executive branches of government.

Digital Divide: Refers to segments of the population lacking Internet access or Internet-related skills.

Dynamic Digital Deliberation: Includes applications that are two-way or dialogical, such as digital town hall meetings and digital policy forums.

Information and Communications Technologies (ICTs): These facilitate citizen participation in government. ICTs may include networked computer systems and Internet-based applications, such as policy or consultation forums, online voting with deliberation, bulletin boards, etc.

Static Digital Deliberation: Digital deliberation that is not dialogical in nature. It typically includes online polls offering instant results and bulletin boards for recommendations or complaints.

ICT–Supported Gaming for Competitive Intelligence

Jan Achterbergh
Radboud University of Nijmegen, The Netherlands

INTRODUCTION

Collecting and processing competitive intelligence for the purpose of strategy formulation are complex activities requiring deep insight in and models of the "organization in its environment." These insights and models need to be not only shared between CI (competitive intelligence) practitioners for the purpose of data direction, collection, analysis, and dissemination, but also between CI practitioners and strategy makers. Sharing these insights and models requires high-quality communication (both regarding content and process) between these parties. In this overview, we examine the potential of (ICT- [information communication technology] enabled) simulation games to improve the quality of communication between CI practitioners and strategy makers.

BACKGROUND

Organizations need to collect and process competitive intelligence about the environment to formulate their strategies. Moreover, the pressure to produce timely, accurate, actionable, and strategically relevant intelligence is growing because the complexity and dynamics of the environment is increasing rapidly (cf. Cook & Cook, 2000; Fuld, 1995; Kahaner, 1997).

CI processes produce this intelligence. Usually, these processes are broken up into four stages: direction, collection, analysis, and dissemination (Fuld & Company, 2002, Kahaner, 1997). In the direction stage, CI professionals establish what data are relevant for the purpose of strategic decision making. In the collection stage, relevant data sources are determined and data are collected from them. In the analysis stage, collected data are transformed into competitive intelligence that can be used by strategic decision makers. In the dissemination stage, the competitive intelligence resulting from analysis is disseminated over strategic decision makers so that they can incorporate it in the process of (re)formulating their strategies.

If CI is to deliver its contribution to the process of strategy formulation, a proactive mode of intelligence gathering seems most appropriate (Ellis, 1993; Hannon, 1997; Tessun, 1997). In this mode, intelligence practitioners try to anticipate environmental developments that may have a strategic impact and assess their consequences. Proactive intelligence requires, in our view, a deep insight into the organization in its environment. For instance, directing the search for information requires an insight into strategic problems the organization in focus has to cope with, and environmental factors having an impact on these problems. To direct the search for data, CI professionals need to construct models of these strategic problems and environmental factors. Analyzing collected information and transforming it into intelligence builds on these constructed models and requires an insight into possible effects of a multitude of states of affairs and events in the environment of the organization, on parties relevant to and the organization in focus itself.

Not only do CI practitioners need a model of the organization in its environment, it is also important that this model is shared among the different parties involved in the intelligence process. The model should be shared among CI professionals so that they have a common orientation toward performing their CI activities. Moreover, it should be shared among CI professionals and strategic decision makers for several reasons. Among these are (a) improving the understanding amongst CI professionals of strategic problems, (b) grounding the model in the strategic orientation of the organization, (c) facilitating the dissemination of the intelligence, (d) ensuring commitment of strategic decision makers to using the intelligence, and (e) improving the process of monitoring and maintaining the model itself.

Sharing the model among the relevant parties in the organization requires high-quality communication (both regarding content and process) between these parties. In this overview, we examine the potential of (ICT-enabled) simulation games to improve this communication process. As Geurts, Caluwé, and Stoppelenburg (2000) assert, simulation games may be a valuable tool contributing to improving the quality of the communication. Simulation games are organized procedures (involving all kinds of paraphernalia) allowing participants to improve communication about complex problems by providing a safe and controlled environment to experiment with different interventions under varying circumstances by means of models representing these complex problems. The element of simulation requires participants to interactively model the organization in its environment, systematically analyzing relevant variables, parties, processes, and their relations. The element of gaming allows participants to interactively experiment with the model in a relatively safe environment. Together, the simulation and the gaming elements can improve both the content and the process of communication required for proactive competitive intelligence. Although the use of games in supporting intelligence activities has been reported (e.g., Allgaier & Powell, 1998; Clark, 1998; Fuld, 1998), the link between simulation games and CI has not been treated thoroughly. In this overview, we examine this link.

To deliver its contribution to the intelligence process, simulation games may be supported by ICT in various ways (e.g., by groupware of various Web-based

Table 1. CI stages and required knowledge

	Description	Required knowledge
Directing	Determine strategic information requirements	Model of organization in its environment; how to use the model to derive required information
Collecting	Identify sources and retrieve data from them	How to select sources; how to approach sources
Analysis	Transform data into intelligence	Model of organization in its environment; how to use the model to assess the impact of specific constellations of environmental variables
Dissemination	Forward intelligence to strategic decision makers	Selection of what users get: what intelligence in what format

Table 2. Contribution of gaming and simulation to improving the quality of communication

		Increasing awareness and motivation	Training skills	Increasing knowledge and insight	Improving communication and cooperation	Integration of learning experiences
Building the simulation game	Model building Transforming model into game Scenario definition	Model building may increase awareness of the importance of knowledge and communication. Motivating to deal with the problem	Training in structuring complex problems in terms of simulations Training in making understandable complex models and making them transferable to other people	Increasing knowledge about the problem under consideration (What is the problem? Why is it a problem?) Increasing knowledge about relevant variables, parameters, events, and relations related to the problem	If participative methods for model building, transformation, and scenario definition are used, chances are created to improve communication and cooperation between parties dealing with the simulated problem.	Participative building allows for pooling knowledge and creating a shared language, improving discussions between parties dealing with the simulated problem.
Using the simulation game	Preparation Introduction Playing the simulation game Analysis and feedback	Awareness of differences between points of departure, differences between events, differences between lines of interventions and their effects Motivation providing a sense of control and security needed to deal with problems Focus is on habitualisation and tacit knowledge	Quickly picking up relevant aspects of a complex problem situation Operationally dealing with complex problems in different circumstances Dealing with unexpected events and interventions	Knowledge and insight in (constellations of related) variables causing certain effects given certain starting conditions, events, and interventions Focus is on analysis and explicit knowledge	Teams of players playing against or with each other need to cooperate, communicate, and get feedback on communication and cooperation and the results. Improvements can be monitored by playing the game more than once.	Shared understanding and awareness of the dynamics of the complex problem given different conditions, events, and interactions

applications). The role of ICT to enable simulation games is also discussed in this overview.

Linking Simulation Games to the CI Process

To explore the possible contribution of simulation games to improve communication in the different stages of the CI –process, we first need to explain what these stages entail and what knowledge is required in these stages. Second, we need to explain in what ways simulation games can improve communication. Given the knowledge requirements in the CI process and the possible contributions of simulation games to improving communication, it becomes possible to examine how communication in the different stages of the CI process can benefit from simulation games.

Tables 1 to 3 provide an answer to these questions. Table 1 provides an overview of the CI stages and of the knowledge required in them. Table 2 provides an overview of possible contributions of simulation games to improving communication. It lists the different functions simulation games can serve. Table 3 links the first two tables. It specifies how simulation games can enhance communication in the different stages of the CI process.

The two left columns of Table 1 present the stages of the CI process. In the right column, the knowledge required in these stages is presented.

Table 2 presents five different functions of simulation games (top row; see Geurts et al., 2000) and links them to the phases of building and using the simulation game (two left columns). In the rest of the table, the contributions of simulation games to enhancing communication are presented.

Table 3 links the knowledge requirements in the stages of the CI process (Table 1) to the possible contributions of building and using simulation games to enhance com-

Table 3. Gaming and simulation contributing to the acquisition of knowledge for the CI stages

	Directing	Collecting	Analysis	Dissemination
Building the simulation game	Participants develop a model of the organization in focus in its relevant environment in terms of parties, variables, and variables relevant for strategic decision making. Making this model can contribute to discussing and getting an awareness of environmental and organizational knowledge needed for strategic purposes, and thus enhances communication about the identification of relevant knowledge domains and data classes needed for directing purposes.	Participants identify possible sources for collecting information about the environment that can be used in the game. Identifying sources forces participants to discuss and evaluate their relevance and accessibility. Knowledge about relevance and accessibility of sources constitutes the core of the collection stage.	Participants model relations between the organization and environment .for gaming purposes. Developing the model enhances communication about variables relevant for strategic decision making and the interpretation of intelligence in terms of these variables in the analysis phase of the CI process.	Participants develop roles in the strategic decision-making process and rules for dissemination of intelligence to persons playing these roles. Developing roles and dissemination rules may focus participants on the importance and structure of different settings for disseminating environmental intelligence and their implications for the strategic decision-making process
Using the simulation-game	Participants identify and vary knowledge domains and data classes for making particular strategic decisions in different scenarios. Identifying knowledge domains and data classes and using them in the game can provide knowledge about the usefulness of these domains and classes in the strategic decision-making process and in varying scenarios.	Participants experiment with different sources of environmental data in different scenarios. Experimenting with these sources may enhance awareness of the usefulness of particular sources for particular problems in the strategic decision-making scenarios.	Participants experiment in different scenarios with giving meaning to environmental and organizational intelligence using the model developed in the building phase as a basis. Experimenting with the model in different scenarios may provide knowledge about the adequacy of the model for analysis purposes. Moreover, participants learn to use the model for analytic purposes in different contexts.	Participants vary who (which persons playing what role) has access to what intelligence in different scenarios. Varying access to intelligence may provide knowledge about the effectiveness and efficiency of different settings for the dissemination of intelligence and its use in the strategic decision-making process (in different scenarios).

Table 4. Examples of ICT applications supporting simulation games to arrive at knowledge for CI (the table builds on the contributions given in Table 3)

	Directing	Collecting	Analysis	Dissemination
Building the simulation-game	To build a model of the environment of the organization, forms of groupware may be used. To identify relevant (values of) variables and parameters, Decision Support Sytem (DSS) or system dynamics software may be used. ICT can also provide a virtual setting for gaming.	To identify possible sources for collecting information about the environment and to find relevant parameters, variables, and their values, ICT applications for monitoring the environment may be used (Internet, online databases, etc.).	Providing a model for analyzing information about the environment. The model used in the direction stage can be used for analysis purposes as well– (for ICT use, see "Directing").	Different forms of ICT may be part of the different settings for distributing environmental information and its use in strategic decisions (e.g., e-mail, electronic conferencing, dedicated software). The adequacy of the different forms may be tested during the game
Using the simulation game	To identify knowledge domains and data classes for making particular strategic decisions in different scenarios, participants may use groupware to support their communication and cooperation. DSS or System Dynamics (SD) software may be used to determine the effect of values of variables in order to direct the search activities. ICT can be used to generate information about autonomous events to the participants.	During the game, the participants may use specific software to monitor the game environment. Also during the game, ICT can be used to store the results about the usefulness of specific sources in different scenarios	ICT should be used to store knowledge about the adequacy of the model for analysis purposes in different scenarios. Groupware can be used to support the communication and cooperation during strategic discussions and the meaning of environmental information. DSS or SD software may be used to determine the effect of values of variables in order to make strategic decisions. Through ICT, common analysis tools can be made available (e.g., an automated Strengths, Weaknesses, Opportunities, and Threats (SWOT).	To store the knowledge about the adequacy of different settings for the distribution of the intelligence and its use (in different scenarios), ICT can be used

munication (Table 2). Table 3 specifies how simulation games can improve knowledge in the CI process for the purpose of strategic decision making.

Linking Gaming and Simulation to ICT

From Table 3, it follows that building and using simulation games can contribute to improving the quality of the content and the process of communication in the different stages of the CI process. Table 4 specifies how ICT may be used to support communication in these stages.

FUTURE TRENDS

As the process of globalization unfolds, the importance of sharing insight in the relations between the organization and parties about its relevant environment by means of high-quality communication between CI practitioners and strategy makers will probably increase. As the complexity and the intensity of these relations (that often include complex feedback mechanisms) increase, demands for model-building techniques and tools will increase as well. Two of the trends allowing for dealing with these increased demands are

- using CI-supported system dynamics instead of econometric models for the purpose of building complex feedback models flexibly handling different simulations and scenarios (Rouwette, 2003), and
- using group model-building for improving the quality of, commitment to, and communication between CI practitioners and strategy makers about the system dynamic models underpinning complex games (Vennix, 1996).

CONCLUSION

The purpose of this overview is to outline the relevance of simulation games for competitive intelligence and to examine the role of ICT to support simulation games for competitive intelligence. Simulation games can be a valuable tool for arriving at knowledge required in intelligence activities. Its main contribution is to build and test different scenarios regarding the impact of the environment on the organization. During building and testing these scenarios, awareness, knowledge, and insight may be gained about the underlying model of the organization in its environment and about the impact of different values of environmental parameters on relevant organizational variables.

Moreover, building and testing different scenarios may also enhance the communication and cooperation needed for intelligence activities. ICT can support delivering the contributions of simulation and games to competitive intelligence in various ways. In this overview, we discussed some possibilities to support the building and using stages, and to facilitate the dissemination and use of the knowledge resulting from the game to intelligence activities.

REFERENCES

Allgaier, C., & Powell, T. (1998). Enhancing sales and marketing effectiveness through competitive intelligence. *Competitive Intelligence Review, 9*(2), 29-41.

Bernhardt, D. C. (1994). "I want it fast, factual, actionable": Tailoring competitive intelligence to executive needs. *Long Range Planning, 27*(1), 12-24.

Clark, B. (1998). Managing competitive interactions. *Marketing Management, 7*(4), 9-20.

Cook, M., & Cook, C. (2000). *Competitive intelligence*. London: Kogan Page.

Ellis, R. J. (1993). Proactive competitive intelligence: Using competitor scenarios to exploit new opportunities. *Competitive Intelligence Review, 4*, 13-24.

Fuld & Company. (2002). *Intelligence software report 2002*. Retrieved from http://www.fuld.com

Fuld, L. M. (1995). *The new competitor intelligence*. Chichester, England: Wiley.

Fuld, L. M. (1998). *The Fuld war room: The ultimate in competitive intelligence*. Cambridge, MA: Fuld & Company.

Geurts, J., Caluwé, L. de, Stoppelenburg, A. (2000). *Changing the organization with gaming/simulations*. Den Haag: Elsevier.

Hannon, J. M. (1997). Leveraging HRM to enrich competitive intelligence. *Human Resource Management, 36*(4), 409-422.

Kahaner, L. (1997). *Competitive intelligence*. New York: Touchstone.

Rouwette, E. A. J. A. (2003). *Group model building as mutual persuasion*. Nijmegen, Netherlands: Wolf Legal Publishers.

Tessun, F. (1997). Scenario analysis and early warning systems at Daimler-Benz aerospace. *Competitive Intelligence Review, 8*(4), 30-40.

Vennix, J. A. M. (1996). *Group model building*. Chichester, England: Wiley.

KEY TERMS

Analysis: The stage of the intelligence cycle in which the strategic significance of environmental data is determined. In this stage, the intelligence is produced. During analysis, intelligence professionals may use different models and techniques to interpret and value environmental data (e.g., SWOT analysis, growth-share matrix, or scenario analysis).

Collection: Stage of the intelligence cycle. In this stage, sources regarding the required environmental data are located and accessed, and the data are retrieved from them.

Competitive Intelligence: In the literature, two definitions are used: a product definition and a process definition. In the product definition, competitive intelligence is defined as information about the environment, relevant for strategic purposes. The process definition highlights producing and processing this environmental information. Process definitions often refer to the intelligence cycle.

Direction: Stage of the intelligence cycle. In the direction stage, one determines the strategic (external) information requirements; that is, one determines what environmental data should be collected.

Dissemination: Stage of the intelligence cycle. In this stage, the intelligence produced in the analysis stage is presented and forwarded to strategic decision makers.

Game Building: Stage in gaming and simulation. During the building stage, game constructors make a model of the problem they want to incorporate in the game. Next, they transform the model into a specific game, and, finally, they define different scenarios that can be played during the game.

Game Using: Stage in gaming and simulation. During the using stage, game facilitators make preparations for playing the game, and participants actually play the game (given a certain scenario).

Gaming and Simulation: The process of building and using simulation games as a means to deal with complex problems (see also "Simulation game").

Simulation Game: An organized procedure involving particular building blocks allowing participants to improve communication about complex problems by providing a safe and controlled environment to experiment with different interventions under varying circumstances by means of models representing these complex problems.

Impacts of Intranet on PDO

Khamis Al-Gharbi
Sultan Qaboos University, Oman

Abdulwahed Mohammed Khalfan
Public Authority for Applied Education & Training, Kuwait

INTRODUCTION

In developing countries such as Oman, the introduction of information technologies is very recent and their use and effect have not been systematically assessed or evaluated. Despite the importance of this subject to researchers and practitioners, the issue of evaluating and assessing the benefits of intranet is still an area that requires further and vigorous research (Blanning & King, 1998). The main aim of this chapter is to evaluate and assess the impact of intranet on one large Omani organisation. The findings are based on an in-depth case study looking at the evolution and the progress of the use of an intranet. The case study basically examines the impact that the use of an intranet has had to date and is likely to have in the next few years on one of the earliest organisations to adopt it in Oman. The impact model presented is a theoretical contribution that can guide other Omani organisations in benefiting from an intranet. Furthermore, the impact model prescribes the factors that need to be considered when implementing an intranet system.

DEVELOPING COUNTRIES CONTEXT

In developing countries such as Oman, the introduction of information technologies is very recent and traditionally, the preferred way of communication and knowledge conversation within a developing country organisational context is the face-to-face meeting. Furthermore, the organisational hierarchical structure is characterised by its rigidity and consequently, information flows from top to bottom (Al-Gharbi, 2001).

A number of studies have reported many obstacles prohibiting developing countries from benefiting from using Internet technologies that need to be considered, such as organisational culture, awareness, education, language, and social and psychological factors (Sharma, Wickramasinghe, & Kitchens, 2002).

The success of intranet will be closely linked to the potentialities of change in a culture to take full advantage of the technology. The culture that is intensively opposed to change, or is hostile toward or suspicious of outsider customers allowed directly to access, interact, and do business transactions with the company systems, must alter their ways and create a new culture in order to reap up the benefits and the rewards of intranet. The challenge of creating such a culture is not an easy job that can be left out to die (Al-Gharbi, 2001). However, management support, training, and organisational culture change are tools to minimise resistance and motivate employees to use the system, as we will see in the case of PDO (Petroleum Development of Oman).

PETROLEUM DEVELOPMENT OF OMAN

Petroleum Development of Oman is one of the largest companies in Oman, with more than 5,000 employees. The Omani government owns 60% of the company and the rest is divided between the Shell Petroleum Company Limited (34%), Total (4%), and the Partex Oman Corporation (2%). The daily oil production of PDO is 800,000 barrels, and the recoverable reserves amount to 5 billion barrels. The company adopted an intranet in early 1996 and currently has 4,500 intranet users.

RESEARCH METHODOLOGY

The qualitative approach is particularly suitable for studying phenomena in which little previous research has been conducted and it is not supported by a strong theoretical base (Benbasat, Goldstein, & Mead, 1987; Walsham, 1995). In addition, Walsham suggests that the case study is the preferred method in investigating the use of information technology in social context, and it can yield rich insight. Furthermore, McBride (1997) stated that to understand the evolution and the progress of information systems, it is far better served by a qualitative approach. The case study offers deep and rich insight, and understanding can be achieved rather than in using a quantitative approach.

In line of the above recommendations, a case-study approach based on frequent visits and face-to-face interviews has been used. The data were collected in two phases: first, through casual conversations and in-depth, informal, and unstructured interviews with PDO staff, which took around 2 weeks; second, through semistructured interviews. Interviews were conducted over a period of 6 months with key participants responsible for intranet adoption and development in the company.

MOTIVATIONS FOR THE INTRANET ADOPTION

One of the main hurdles facing PDO was how to effectively communicate and share information and knowledge with different individuals and groups inside and outside the country without increasing the costs. Thus, PDO has adopted an intranet in order to achieve the following goals: to improve communication and provide a means by which to help knowledgeable employees from different departments and locations to share information and knowledge; to share knowledge, resources, and expertise with the Shell group; to reduce costs; to eliminate duplication and redundancy of information; to enable cross-platform compatibility; and to improve business processes.

THE INTRANET IMPACT MODEL

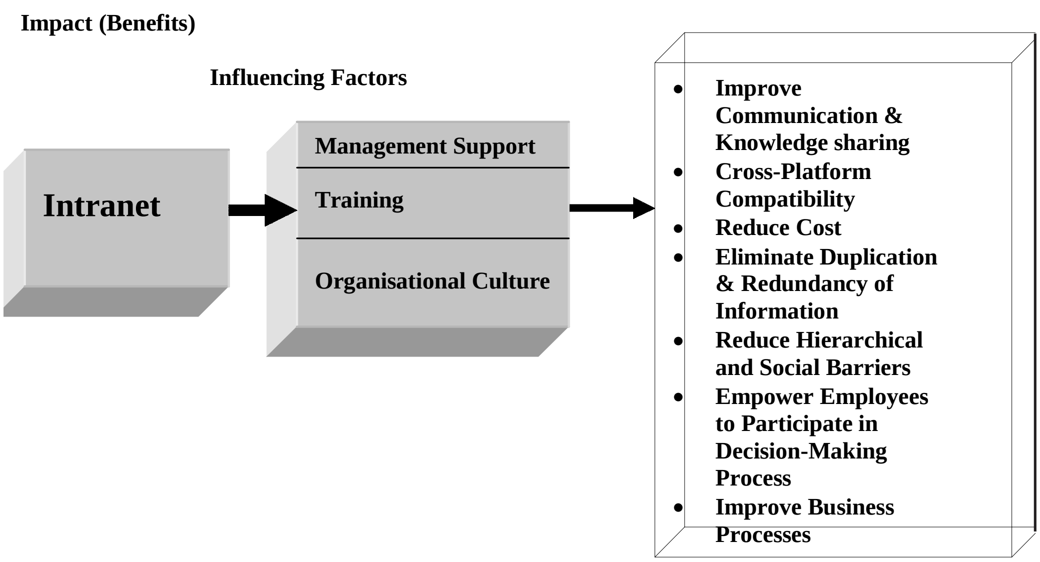
Most of the research on intranet adoption to date concentrates on organisations in the developed countries (Curry & Stancich, 2000; Davies, 2000; Hansen, Nohria, & Tierney, 1999). Among these articles, there is evidence to suggest that intranet has had an impact upon many business aspects and processes. For example, intranet improves access to updated information, improves communication, enables knowledge sharing and IT integration, flattens organisational hierarchies, reduces costs, improves decision making, empowers users, and facilitates organisational knowledge and learning. Furthermore, Callaghan (1998) states that intranets have broken down internal divisional barriers to communication.

The intranet impact model presented in Figure 1 shows that this technology has affected many processes and business aspects. In the following sections, each component of the impact model is discussed.

INTRANET IMPROVES COMMUNICATION AND KNOWLEDGE SHARING

PDO has 4,500 users who use its intranet to share, not only static information, but dynamic knowledge about engineering, oil exploration, and safety procedures. According to a senior manager, "The intranet enables our employees to know where to go to find what they need to know regardless of location and time barriers." The benefits can be accrued when one department is able to take advantage of the experience and understanding of another in the organisation. An organisation would have different entities of knowledge generated over the years by various groups and departments. These various and separate bodies of knowledge can be combined and shared

Figure 1. The intranet impact model



over the intranet and new knowledge can be born. "The intranet enables our employees to access and manage knowledge, and therefore enables them to respond to every situation with the sum total of everything anyone in the organisation has ever learned about situations of a similar nature."

One of the most used discussion groups is called "Speakers Corner," which is an open discussion about all matters concerning the company. Knowledge about oil exploration, engineering, safety rules, and regulations is discussed and shared. One senior manager commented, "Any employee can post anything on the Speakers Corner and this ensures democratic contribution and equal consideration of ideas. There is a big commitment from top management to make the Speakers Corner successful. Top management are participating in the Speakers Corner and providing answers for the issues discussed. Furthermore, top management are listening and benefiting from issues that are discussed via Speakers Corner to find solutions and feedback where it can be used to develop and improve services."

The idea is to make individual knowledge available to others. This concept is one of the most important principles in knowledge management. It is the main and the most important activity of the knowledge-creating company. According to Nonaka (1994), it should take place continuously and includes all levels of the organisation. Top management can take the opportunity and the potential presented by an intranet to communicate the company's mission, vision, and strategy directly and quickly, not only to the employees but also to business partners and customers.

The Speakers Corner is a knowledge- and information-sharing opportunity, which is open not only to PDO employees, but also to Shell personnel as well. Management felt that PDO is a learning organisation and that the intranet should be open to Shell users and vice versa to benefit the company.

One participant stated, "The intranet should be viewed as a learning opportunity for both companies. PDO can learn from the expertise of Shell Group about oil exploration, engineering, and safety procedures, and learn best practice by accessing to the Shell intranet." However, one must be careful and sensitive before importing other cultural approaches to solving local problems or needs without careful consideration for the usefulness of such an approach in a local context.

With top-management support and guidance, the intranet has enabled an appropriate new culture of openness to be accepted. The idea of openness to the outsider (Shell Group) is an important cultural transformation, which was difficult with previous technology. According to Davenport and Prusak (1998), the presence of knowledge-based technologies such as intranet can have a positive effect on organisational culture.

REDUCE HIERARCHICAL AND SOCIAL BARRIERS AND EMPOWER EMPLOYEES

The Speakers Corner became an opportunity to break organisational and social hierarchical barriers and to create knowledge by allowing employees to ask questions, provide (or propose) solutions, and transfer best practices, overcoming the barriers of distance and time.

The intranet helps open organisational boundaries, with many employees from different departments and hierarchical positions participating in the Speakers Corner discussing issues which, before the intranet, were limited to formal meetings that involved senior employees in the company in most cases. The intranet enables a wider audience to participate in the decision-making process and increase informal participation. Consequently, increasing individual autonomy is one important principle in knowledge management (Nonaka, 1994).

The diversity of participants can create an extremely unique opportunity where new ideas can emerge. The idea of wide participation is very important for organisational knowledge to grow, where individual knowledge can be consolidated with group knowledge through technology so that new ideas can be discussed, debated, and even discarded if necessary. In this way new knowledge could be born, which will have a higher possibility of survival and which can then be integrated into the organisation's knowledge (Inkpen, 1998).

The ability of Internet technologies in knowledge capturing and knowledge dissemination to a wider audience, which can transcend organisational boundaries (Earl, 1998), makes these technologies important tools for knowledge management. Indeed, these technologies facilitate knowledge building through the exchange of ideas, papers, hypotheses, data, messages, and even gossip (Earl).

The effect of the intranet on PDO employees is quite noticeable. IT empowers PDO employees to participate in the company strategy regardless of position and location. Before the intranet, information flowed vertically through the PDO hierarchical structure. The request for information had to flow up or down through the chain of command, which was usually rigid and took a long time. The intranet has changed this to a certain extent. Indeed, several empirical studies have reported the assertion that an intranet empowers employees to participate in the decision-making process (Callaghan, 1998; Scott, 1998).

REDUCING COST AND ELIMINATING DUPLICATION

Cost reduction has been achieved on paper printing, telephone directories, forms, policy and safety procedures, and inventory reports. With the intranet, one copy can be posted and any employee can access the document at anytime from anyplace. This helps to eliminate unnecessary duplication of information and costly, unwanted redundancy, while at the same time making it easy and quick to update.

The knowledge-management officer with PDO stated, "We have more than 4,000 different documents (document management, control documents, business records and reports, engineering and building drawings, etc.). These documents represent the company knowledge and information accumulated over the years. They contain vital business information, which now can be accessed by a wider audience. "About 60% of these documents today are on the intranet, which save the company a lot of money. Employees can easily search different topics contained in these documents by selecting a topic or a keyword."

"This is an important cost reduction for our organisation, but more importantly," comments a senior information advisor, "PDO is experiencing a huge cultural change moving from paper documents to almost all materials published on the intranet in a relatively short period of time. The company has achieved a 60% to 70% reduction of hard copy output so far."

The major benefit of intranet technology derives from its role as an online, on-demand information-dissemination tool. Using the intranet, the strategy of PDO has transformed from a calendar-based document-publishing strategy to one based on need. This transformation not only results in cost reduction of information dissemination, but also improves access to a wider audience with updated information.

IMPROVE BUSINESS PROCESSES

Human resources (HR) is one of the important departments using the intranet in PDO. The vision is to provide a quick and easy means of access to and awareness of HR information, activities, and services. PDO is exploiting information technology in order to achieve this goal. The human resources department has developed its own intranet called Infolink; the explicit strategy of Infolink is to "bring previously difficult-to-find information, policies, procedures, and forms directly to employees' computers," as one participant stated.

The designer of Infolink stated that the site promises that "human resources information can be accessed with just two clicks of the mouse." In isolating the company knowledge in one or two functional departments like IT or HR from those who need it, companies consequently risk losing its benefits (Gable, 1994). A better understanding of the interaction between different processes has allowed PDO to save time and money whilst providing better services to employees and customers.

The use of the intranet enables HR to eliminate many of the routine paper-processing tasks. According to a senior information advisor, the so-called "Smart Forms," which will be launched in a few months, "will be circulated in cyberspace, picking up signatures as they go and then, when they return, they are complete, automatically updating records." The Internet technologies have played an important role in changing business processes, which can save organisations and employees time and money.

Another important application of the human resources intranet is the SAPpHiRe online database, which, as part of Infolink, provides access to all PDO employees. SAPpHiRe enables "all the staff to access their own personal information." This self-access service provides the employee with all the information he or she needs. Furthermore, SAPpHiRe provides the employee with the ability to update specific information in his or her file (for example, address, telephone number, e-mail, etc.). SAPpHiRe also provides feedback to the company to enhance certain services and solve problems through the use of surveys. This facility enables the employee to give opinions and raise concerns. The results of the survey can be accessed by all employees through SAPpHiRe. Staff appraisal forms can be found on the intranet and the supervisor(s) can complete the form to evaluate the staff. The relevant staff member can, therefore, access the evaluation and give feedback on his or her performance.

CROSS-PLATFORM COMPATIBILITY

Exploration and production is another department that uses the intranet very heavily. Geologists and geophysicists are able to access and share information from different databases and different platforms (PCs [personal computers], Mainframes, and Macs) via the intranet. The intranet allows cross-platform compatibility, therefore it does not matter where knowledge resides. Tools such as browsers enable users to access and share information and knowledge from and between different platforms that have stored information and knowledge (Kim, 1998; Scott, 1998). Similarly, Davenport and Prusak (1998) stated that "Internet-based technology is ideal for publishing information across multiple types of computer platforms, for multimedia databases, and for displaying knowledge that is linked to other knowledge through hypertext links."

Geologists and geophysicists can easily access different topographical maps, which contain information about the oil fields in the country. Using the intranet, exploration, engineering, production, and economic solutions are easily provided and accessible.

PDO has adopted a holistic view, which consolidates the technology, human, and organisational factors required in using the intranet for knowledge management.

CONCLUSION

The case study has contributed to the literature concerning the impact of intranet by giving a detailed account of one Omani company's experience. The case study has demonstrated that the intranet empowers PDO employees to participate in the company strategy regardless of position and location, and it consequently has flattened the organisational hierarchy. Furthermore, PDO's 4,500 staff members are able to communicate with each other and share knowledge. On the other hand, the intranet has contributed to the elimination of internal divisional barriers. In addition, with top-management support and guidance, the intranet has enabled an appropriate new culture of openness to be accepted. The idea of openness to an outsider, the Shell Group, is an important cultural transformation, which was difficult with previous technology. This is an important cultural change in PDO's history. PDO and other Omani organisations can build on this experience to utilise the technology for future opportunities and collaboration at national and international levels.

The impact model shows that the intranet has affected many processes and business aspects. Furthermore, it shows the importance of top-management support to overcome the human and cultural barriers that stand against better utilisation of the intranet, as well as showing how to develop a culture of shared values that can be a benefit of the adoption of an intranet. Therefore, the model recommends the interaction of management support, training, and organisational culture to minimise resistance and motivate employees to use the system.

REFERENCES

Al-Gharbi, K. (2001). *The impact of the intranet on knowledge management in the Omani private sector*. PhD thesis, De Montfort University, UK.

Benbasat, I, Goldstein, K. D., & Mead, M. (1987, September). The case study research strategy in studies of information systems. *MIS Quarterly*, 369-384.

Blanning, R. W., & King, D. R. (1998). Editorial. *Decision Support Systems, 23*, 1-2.

Callaghan, J. (1998). Corporate intranets: Just another IT system? *Proceedings of the Eighth Annual BIT Conference*, Manchester.

Curry, A., & Stancich, L. (2000). The Intranet: An intrinsic component of strategic information management? *International Journal of Information Management, 20*(4), 249-268.

Davenport, T. H., & Prusak, L. (1998). *Working knowledge*. Boston: Harvard Business School Press.

Davies, N. J. (2000). Knowledge management. *BT Technical Journal, 18*(1), 62-63.

Drew, S. (1999). Building knowledge management into strategy: Making sense of a new perspective. *Long Range Planning, 32*(1), 130-136.

Earl, M. J. (1994). Knowledge as strategy: Reflections on Skandia international and Shorko films. In C. Ciborra & Jelassi (Eds.), *Strategic information systems: A European perspective*. Chichester, England: John Wiley & Sons.

Earl, M. J. (1998). Knowledge strategies: Propositions from two contrasting industries. In M. J. Earl (Ed.), *Information management*. Oxford University Press.

El-Nawawy, M. A., & Ismail, M. M. (1999). Overcoming deterrents and impediments to e-commerce in light of globalization: The case of Egypt. *Proceedings of INET99*, San Jose, CA.

Gable, G. (1994). Integrating case study and survey research methods: An example in information systems. *European Journal of Information Systems, 3*(2), 112-126.

Hansen, T. M., Nohria, N., & Tierney, T. (1999, March-April). What's your strategy for managing knowledge? *Harvard Business Review*, 107-116.

Hills, M. (1998). The definition and dynamics of intranets. In. P. Lloyd & P. Boyle (Eds.), *Web-weaving intranets, extranets and strategic alliances*. Oxford, England: Butteworth Heinemann.

Inkpen, A. (1998). Learning, knowledge acquisition, and strategic alliances. *European Management Journal, 16*(2), 223-229.

Kim, J. (1998). Hierarchical structure of intranet functions and their relative importance: Using the analytic hierarchy process for virtual organisations. *Decision Support Systems, 23*, 59-74.

McBride, N. (1997). The rise and fall of an executive information system: A case study. *Information Systems Journal, 7*, 277-287.

Nonaka, I. (1994). A dynamic theory of organisational knowledge creation. *Organisation Science, 5*(1), 14-37.

Scott, J. E. (1998). Organisational knowledge and the intranet. *Decision Support Systems, 23*, 3-17.

Sharma, S. K., Wickramasinghe, N., & Kitchens, F. (2002). Why India is not positioned well in e-commerce race? *Proceedings of the Third Annual Global Information Technology Management (GITM) World Conference*, New York, NY.

Tan, D. S., & Uijttenbroek, A. A. (1997). Information infrastructure management. *Information Systems Management, 14*(4), 33-41.

Walsham, G. (1995). Interpretive case studies in IS research: Nature and method. *European Journal of Information Systems, 4*(2), 73-83.

Willcocks, L., Graeser, V., & Lester, S. (1998). "Cybernomics" and IT productivity: Not business as usual. *European Management Journal, 16*(3), 272-28.

KEY TERMS

Explicit Knowledge: Knowledge that can be expressed in words and can be acquired from libraries, databases, archives, telephone conversations, meetings, memos, IT, and IS

Infolink: The human resources department Web site on the intranet

SAPpHiRe: An online database that is a part of Infolink

Speaker Corner: The main intranet Web site in PDO

Tacit Knowledge: Knowledge that is difficult to express in words or communicate, and hard to articulate and share

Impediments for Knowledge Sharing in Professional Service Firms

Georg Disterer
University of Applied Sciences and Arts, Hanover, Germany

INTRODUCTION

Professional service firms (PSFs), where professionals (consultants, lawyers, accountants, tax advisors, etc.) work, are interested in knowledge management, because their businesses are heavily dependent on the knowledge of their employees. A core asset is their ability to solve complex problems through creative and innovative solutions, and the basis for this is their employees' knowledge. The "product" that PSFs offer their clients is knowledge (Kay, 2002; Ofek & Sarvary, 2001; Chait, 1999).

Sharing knowledge between colleagues improves the economical benefits a firm can realize from the knowledge of employees. This is especially true for PSFs (Huang, 1998; Quinn, Anderson, & Finkelstein, 1996), where broad ranges of knowledge must be kept to provide intellectual services, and real-life experiences with certain questions and situations are an important asset. The organizations and its members are spread over various offices across the country or the world. The necessity for sharing grows, because the network of professionals in most cases can offer significantly better professional advice than any individual. "We sell knowledge … the most valuable thing we can offer is the collective, institutional knowledge of our firm …" (Roger Siboni, KPMG executive, cit. in Alavi, 1997, p. 1). Working together openly without holding back or protecting vital pieces of knowledge will result in more productivity and innovation than could be reached individually.

BACKGROUND

No professional is denying the worth of using working documents and materials produced by others. All PSFs are trying to set up collections of knowledge acquired in projects in order to share it and conserve it for reuse. Knowledge databases can address what is sometimes called the traditional weakness of PSF: "… narrow specialists who see only their own solutions, self-centered egoists unwilling or unable to collaborate with colleagues" (Liedtka et al., 1997, p. 58). Many authors signal that sharing knowledge seems to be "unnatural" (Quinn, Anderson, & Finkelstein, 1996; Barua & Ravindran, 1996; Holloway, 2000).

However, attempts to use knowledge databases often fail. Only a few databases are accepted as up to date. The special fields of expertise are covered only in fragments. The access is laborious and uncomfortable. Heterogeneous sources (text, internal and external databases, journals, books, comments, codes of law, and so forth) cannot be integrated. The lack of actuality and completeness causes quality risks if dealt with thoughtlessly and if unreflected upon.

People issues are meant to be critical for successful knowledge sharing. According to Ruggles (1998), "In fact, if the people issues do not arise, the effort underway is probably not knowledge management. If technology solves the problem, yours was not a knowledge problem" (p. 88). Therefore, we analyze the reasons why knowledge sharing needs dedicated efforts and describe possible actions to foster knowledge sharing. Through our research (Disterer, 2000, 2001, 2002a) and analyses drawn from literature, we categorize and discuss the various impediments encountered by people sharing knowledge (see Figure1). There are some empirical results that confirm these impediments (APQC, 1996; Ruggles, 1998; KPMG, 2003; Govindarajan & Gupta, 2001). Then we show various approaches to overcome these impediments.

IMPEDIMENTS TO KNOWLEDGE SHARING

Loss of Power

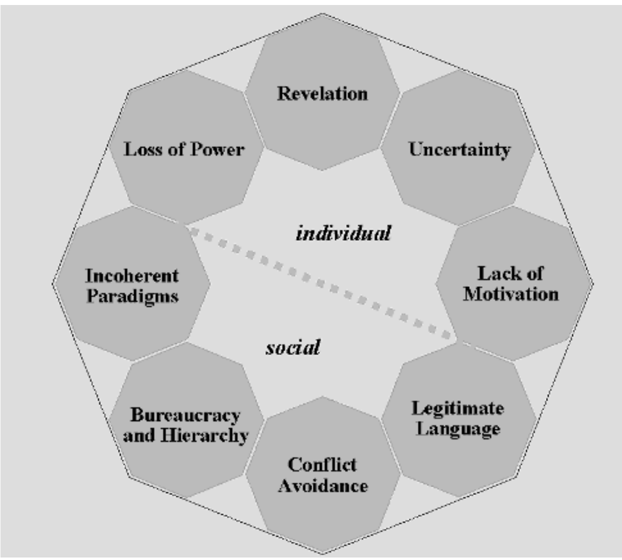
Knowledge can be used to take action and to enforce spheres of influence. Passing knowledge to colleagues might grant some of these potentials. Those who do not have this knowledge are deprived of the capacity to act or to influence. That applies for knowledge about customers, competitors, suppliers, procedures, recipes, methods, formulas, etc. In this sense, someone who passes on knowledge to a colleague loses the exclusiveness of his or her influence, which might have suggested some professional respect and job security. "Knowledge is power" is the well-known citation to describe situations in which experts with rare knowledge have the highest reputation, and monopolies of knowledge causes knowledge hoarding instead of knowledge sharing (Reimus, 1997; Andrews, 2001).

In industries like professional services, employees are competing directly with each other through their special knowledge, gifts, and talents. It might be part of the individual culture of high-performing employees that they voluntarily enter into the competition for scarce seats on their career paths because they like to compete and excel (Quinn, Anderson, & Finkelstein, 1996). But, the drawbacks of competition are obvious: Knowledge workers would be cautious to share their knowledge, because they could possibly give up an individual lead.

Revelation

Passing on knowledge to colleagues or entering working results into a knowledge database may be considered as a revelation, because it proclaims that this knowledge has a certain value and rarity. If this assessment is not shared by others, embarrassment may result (Rodwell & Humphries, 1998). Additionally, hasty colleagues rush to suggest "necessary" improvements to emphasize their expertise. For an individual, knowledge justified as "true belief" is not of particular concern. But in situations of knowledge sharing, more than one individual is involved. At this point, "… justification becomes public. Each individual is faced with the challenge of justifying his true beliefs in presence of other" (Krogh, 1998, p. 35).

Figure 1. Impediments to knowledge sharing



Uncertainty

Less experienced colleagues may feel uncertain, because they cannot judge if their working results and experiences represent valuable knowledge for others. They cannot estimate if their knowledge is too general or too well known or, on the other side, that some results are too specific for a special situation and therefore useless for colleagues in other situations. Positioning on the scale of "general" to "specific" is not trivial and, thus, results in uncertainty.

Lack of Motivation

Sharing knowledge is often seen as additional work because of the time necessary for reflection, documentation, communication, and so forth. Time is scarce, especially if the performance of an organization is measured by billable hours only. Reflection of work and sharing experiences are more an investment for future work than a billable action in the present. As stated in Dixon (2000), "In an organization with a bias for action, the time for reflection may be hard to come by" (p. 18; Hunter, Beaumont, & Lee, 2002).

Some employees do not expect reciprocal benefits from sharing, because they do not believe in these benefits or they did not experience it. Benefits of contributing to a knowledge database are gotten by a different stakeholder later on—the benefits will not be earned by the provider but by others (Nissen, Kamel, & Sengupta, 2000). Therefore, one precondition for contributing is the assumption of an equilibrium—a balanced give and take between colleagues. The insight that knowledge sharing can only be beneficial if everybody provides knowledge unselfishly may have charm only theoretically. In day-to-day practice, the benefit is too uncertain, and payback is not going to be immediate; therefore, the individual's commitment to share knowledge fails.

Legitimate Language

Some organizations lack a legitimate language (Krogh, 1998) that is known and accepted by all colleagues and can carry individual knowledge. This covers the need for a common language to communicate analogies and metaphors to externalize tacit knowledge hidden in individual mental models, viewpoints, working models, schemata, paradigms, and beliefs (Nelson & Cooprider, 1996; Nonaka, 1994; Haldin-Herrgard, 2000).

Conflict Avoidance

Attitudes of conflict avoidance and some conservative habits may prevent knowledge sharing, if the knowledge contains some new thoughts or innovative ideas. If most executives of an organization are not comfortable with change and are not willing to take risks, new ideas may be covered up easily. Different views and perspectives would be hidden, and knowledge not culturally legitimated may be suppressed ("do not rock the boat" attitude).

Bureaucracy and Hierarchy

Bureaucratic and hierarchical organizations show formal and administrative procedures that prevent the sharing of knowledge and new ideas. Strong hierarchical organizations prevent cross-functional communication, cooperation, and knowledge sharing.

Incoherent Paradigms

A lack of alignment between the personal intents of the individuals and the paradigms of the organization (strategic intent, vision, mission, strategies, values, etc.) can cause difficulties in articulating and justifying personal beliefs that do not fit with the ruling paradigms of the firm (Krogh, 1998).

ACTIONS TO FOSTER KNOWLEDGE SHARING

There is no complete methodology and no set of procedures and policies to address systematically all of the above impediments to knowledge sharing. Various approaches should be discussed further. Some of them sound like common sense but need to be emphasized because of their importance.

Concern and Trust

A precondition for knowledge sharing within organizations is an attitude of concern and trust among members of the organizations. Krogh (1998) called this "care" and defined it as serious attention, a feeling of concern and interest within an organization. His concept includes phenomena like trust among the people, interest in different viewpoints and experiences, access to help, lenience in judgment, courage to voice opinions, courage to allow experiments, and courage to take risks.

Organizations have to strive for a culture of accepting mistakes (Soliman & Spooner, 2000). They should not penalize errors to foster a climate of constructive conflicts, giving members the chance to "fall forward." Organizational development processes should build and establish a common set of ethical standards and values for an organization and should achieve a consensus of accepted working practices and habits.

Leadership

Knowledge sharing is based on the consistent, reliable, and plausible behavior of management. Members of management must positively communicate that they are thoroughly convinced that knowledge needs to be "nurtured, supported, enhanced, and cared for" (Nonaka & Konno, 1998, p. 53) and that they financially support knowledge management initiatives. Management must afford time for communication and reflection. There must be organizational slack that permits time for employees to network (Krogh, 1998; Wiig, 1997).

To openly share knowledge, mutual trust is necessary among all organization members. Trust results in common expectations of reliability, consistency, and plausibility. Trust reduces the fear that others will act opportunistically. Likewise, management must act as examples for knowledge sharing. They have to walk-the-talk and give up knowledge hoarding first. Members of a profession or a community accept standards of behavior and working habits from their peers (Quinn, Anderson, & Finkelstein, 1996); therefore, management must act as peers to be an example in knowledge sharing (McDermott & O'Dell, 2001).

Rewards and Incentives

Special rewards and incentive methods can act as extrinsic motivation for employees willing to share knowledge. Organizations are successful with the provision of personal recognition and reputation when people have contributed to knowledge databases or actively participated in knowledge sharing (Hunter, Beaumont, & Lee, 2002).

Some examples for direct rewards and how to provide chances to build reputation and fame include the following: Texas Instruments created an annual award named "Not Invented Here, But I Did It Anyway Award" (Dixon, 2000, p. 57) to reward usage of other employees' knowledge. Buckman Labs rewards the top 150 "knowledge sharers" (judged by knowledge managers) with a laptop and an incentive trip to a resort (Davenport, Long, & Beers, 1998). AMS honors contributors to the knowledge center with a bronze plaque at the headquarters and regularly publishes a top 10 list of most frequently used contributions (King, 1998). Forum, a consultancy in Boston, Massachusetts, holds a "World Cup Capture" to encourage its consultants to make explicit and sharable what they have learned from their latest engagements (Botkin, 1999). An Australian law firm honors individuals who contributed the most by having a star named after them (Robertson, 1999).

Contrary to this, there might be professions with different views: "A major concern of software engineers is the fear of being known as an expert" (Desouza, 2003,

p. 100). They fear being staffed to projects based on their past experience instead of being allocated more challenging tasks with room for learning. In this situation, the brand-like identity of a software engineer works to his or her disadvantage and builds a barrier to individual professional development.

Many organizations incorporate issues of knowledge sharing into their compensation plans and promotion policies. The big consulting and accounting firms commonly base their personal evaluations partly on how many contributions are made to knowledge databases, how many new employees people have tutored, and how many training courses have been designed (Quinn, Anderson, & Finkelstein, 1996; Whiting, 1999).

Tutoring and Mentoring

Administrative actions may define responsibilities for tutoring and mentoring in an organization. Ongoing programs that systematically develop employees can foster common habits and attitudes and can support communication among the members of the organization.

Project Experiences

At the end of bigger projects and transactions, time and effort for explicitly debriefing should be provided to learn systematically by experiences. The lessons learned could be systematically analyzed and stored for access by other employees. In other actions, it can help to use knowledge and experiences gained in projects (Disterer, 2002b).

Communities of Practice

A popular approach for fostering knowledge sharing is to develop communities of practice. These groups of professionals enhance the ability of its members to think together, to stay in touch with each other, and to share ideas with each other. These informal networks, sometimes also called knowledge fairs or clubs, competence centers, or creativity centers, consist of groups of professionals, informally bound to one another through a common class of interests and problems and a common pursuit of solutions. People who are exposed to a common class of interests and problems often develop a common language with which to communicate and develop a sense of mutual obligation to help each other (Manville & Foote, 1996; McDermott, 1999). These phenomena can be used to overcome some of the individual and social barriers to knowledge sharing within communities of practice.

Focus on Codification or Personalization

In PSFs, the knowledge of experts is a core asset, and therefore, careful management of this asset is important. Management is responsible for ensuring that the organization is as independent as possible from individuals. At the same time, these companies are operating in a "people business," where the personal and individual link between clients and professionals is critical (Morris & Empson, 1998). This special situation requires special approaches to manage knowledge. Quite popular are two approaches that consulting firms apply that address cultural issues differently (Hansen, Nohria, & Tierney, 1999).

One strategy ("codification") centers on information technology (IT): The knowledge is carefully codified and stored in knowledge databases and can be accessed and used by others. With the other strategy ("personalization"), knowledge is tied to the person who developed it and is shared mainly through direct person-to-person contact (Hansen, Nohria, & Tierney, 1999). With a codification strategy, knowledge is extracted from the person who developed it, is made independent from the individual, and is stored in the form of interview guides, work schedules, checklists, etc. Knowledge is then searched and retrieved and used by other employees. Personalization focuses on dialogue between individuals; knowledge is shared primarily in personal meetings and in one-on-one conversations.

Individual barriers are significantly lower with a personalization strategy, because professionals keep control through the whole knowledge management cycle. The individual is recognized as an expert and is cared for. In fact, focusing on personalization could be called a communication strategy, because the main objective is to foster personal communication between people. Core IT systems are yellow pages (directories of experts, who-knows-what systems, people finder) that show people with whom they should discuss special topics or problems.

Organizational Design

Some organizational designs can foster intraorganizational collaboration. Partnerships and other forms of ownership by employees can be utilized to produce involvement and commitment (Hildebrand, 1994; Miles et al., 1998). Moreover, these organizational forms address the hesitation of professionals with specialized knowledge to work within strong hierarchies and in working environments with strong regulations (Quinn, Anderson, & Finkelstein, 1996).

Office Design and Construction

To lower the disadvantages of bureaucracy and formal communications, modern office layouts reduce the distance between colleagues to foster ad hoc, informal, and face-to-face communication.

FUTURE TRENDS

Further research will be necessary to understand barriers to knowledge sharing, because cultural barriers are dominant over technical problems while implementing knowledge management initiatives. Today, appropriate reward and incentive systems and well-established communities of practice are the most promising ways to overcome the barriers.

CONCLUSION

Ways to support knowledge management with IT are manifold, but certain cultural aspects of knowledge sharing must be addressed. We describe some possible actions to overcome typical resistance often articulated with phrases like "this is client confidential," "only I know how to use it," "what is in it for me?," and "I have no time for documentation." The descriptions of the impediments and the possible actions make clear that knowledge management could not be seen as a technical field, as it is deeply social in nature.

REFERENCES

Alavi, M. (1997). KPMG Peat Marwick U.S.: One great brain. Harvard Business School Case [No. 9-397-108].

American Productivity and Quality Center (APQC). (1996). Knowledge management—Consortium Benchmarking Study Final Report, APQC.

Andrews, D. (2001). Knowledge management: Are we addressing the right issues? *Managing Partner, 4*(1), 23-25.

Barua, A., & Ravindran, S. (1996). Reengineering information sharing behaviour in organizations. *Journal of Information Technology, 11*(3), 261-272.

Botkin, J. (1999). *Smart business: How knowledge communities can revolutionize your company*. New York: Free Press.

Chait, L. P. (1999). Creating a successful knowledge management system. *Journal of Business Strategy, 20*(2), 23-26.

Davenport, T. H., Long, D. W., & Beers, M. C. (1998). Successful knowledge management projects. *Sloan Management Review, 39*(4), 43-57.

Desouza, K. C. (2003). Barriers to effective use of knowledge management systems in software engineering. *Communications of the ACM, 46*(1), 99-101.

Disterer, G. (2000). Knowledge management—Barriers to knowledge databases in law firms. *Managing Partner, 2*(3), 24-27.

Disterer, G. (2001). Individual and social barriers to knowledge transfer. In R. H. Sprague (Eds.), *Proceedings of the 34th Hawaii International Conference on System Sciences* (pp. 1-7). Los Alamitos, CA: IEEE Computer Society Press.

Disterer, G. (2002a). Social and cultural barriers for knowledge databases in professional service firms. In D. White (Eds.), *Knowledge mapping and management* (pp. 124-130). Hershey, PA: IRM Press.

Disterer, G. (2002b). Management of project knowledge and experiences. *Journal of Knowledge Management, 6*(5), 512-520.

Dixon, N. M. (2000). *Common knowledge: How companies thrive on sharing what they know*. Cambridge, MA: Harvard University Press.

Govindarajan, V., & Gupta, A. K. (2001). Building an effective global business team. *Sloan Management Review, 42*(4), 63-71 .

Haldin-Herrgard, T. (2000). Difficulties in diffusion of tacit knowledge in organisations. *Journal of Intellectual Capital, 1*(4), 357-365.

Hansen, M. T., Nohria, N., & Tierney, T. (1999). What's your strategy for managing knowledge. *Harvard Business Review, 1999*(2), 106-116.

Hildebrand, C. (1994). The greater good. *CIO Magazine, 8*(4), 32-40.

Holloway, P. (2000). Sharing knowledge—And other unnatural acts. *Knowledge Management Magazine*, No. 1.

Huang, K. -T. (1998). Capitalizing on intellectual assets. *IBM Systems Journal, 37*(4), 570-583.

Hunter, L., Beaumont, P., & Lee, M. (2002). Knowledge management practice in Scottish law firms. *Human Resource Management Journal, 12*(2), 4-21.

Kay, S. (2002). Benchmarking knowledge management in U.S. and UK law firms. Retrieved August 15, 2002, from www.llrx.com (Law Library Resource XChange); www.llrx.com/features/benchmarkingkm.htm

King, J. (1998, June 15). Knowledge management promotes sharing. *Computerworld*.

KPMG. (2003). Insights from KPMG's European Knowledge Management Survey 2002/2003, KPMG Consulting (Eds.).

Krogh, G. (1998). Care in knowledge creation. *California Management Review, 40*(3), 133-153.

Liedtka, J. M., Haskins, M. E., Rosenblum, J. W., & Weber, J. (1997). The generative cycle: Linking knowledge and relationships. *Sloan Management Review, 38*(1), 47-58.

Manville, B., & Foote, N. (1996). Harvest your workers' knowledge. *Datamation, 1996*(7), 78-81.

McDermott, R. (1999). Why information technology inspired but cannot deliver knowledge management. *California Management Review, 41*(4), 103-117.

McDermott, R., & O'Dell, C. (2001). Overcoming cultural barriers to sharing knowledge. *Journal of Knowledge Management, 5*(1), 76-85.

Miles, G., Miles, R. E., Perrone, V., & Edvinssen, L. (1998). Some conceptual and research barriers to the utilization of knowledge. *California Management Review, 40*(3), 281-288.

Morris, T., & Empson, L. (1998). Organization and expertise: An exploration of knowledge bases and the management of accounting and consulting firms. *Accounting, Organizations and Society, 23*(5/6), 609-624.

Nelson, K. M., & Cooprider, J. G. (1996). The contribution of shared knowledge to IS group performance. *MIS Quarterly, 20*(4), 409-432.

Nissen, M., Kamel, M., & Sengupta, K. (2000). Integrated analysis and design of knowledge systems and processes. *Information Resources Management Journal, 13*(1), 24-43.

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science, 2,* 14-37.

Nonaka, I., & Konno, N. (1998). The concept of "Ba": Building a foundation for knowledge creation. *California Management Review, 40*(3), 40-54.

Ofek, E., & Sarvary, M. (2001). Leveraging the customer base: Creating competitive advantage through knowledge management. *Management Science, 47*(11), 1441-1456.

Quinn, J. B., Anderson, P., & Finkelstein, S. (1996). Managing professional intellect: Making the most of the best. *Harvard Business Review, 74*(2), 71-80.

Reimus, B. (1997). Knowledge sharing within management consulting firms. Retrieved February 9, 1999, from www.kennedyinfo.com/mc/gware.html

Robertson, G. (1999). The impact of knowledge management on Australian law firms. White paper.

Rodwell, I., & Humphries, J. (1998). The legal face of knowledge management. *Managing Information*, *5*(7), 31-32.

Ruggles, R. (1998). The state of the notion: Knowledge management in practice. *California Management Review, 40*(3), 80–89.

Soliman, F., & Spooner, K. (2000). Strategies for implementing knowledge management: Role of human resources management. *Journal of Knowledge Management, 4*(4), 337-345.

Whiting, R. (1999, November 22). Knowledge management: Myths and realities. *Informationweek Online*, pp. 1-5.

Wiig, K. M. (1997). Knowledge management: Where did it come from and where will it go? *Expert Systems with Applications, 13*(19), 1-14.

KEY TERMS

Community of Practice: A community of practice is a group of people in an organization who are (somehow) held together by common interest in their work topic, purpose, and activities.

Culture: Covers the pattern of basic assumptions accepted and used about behaviors, norms, and values within an organization.

Knowledge Management: The systematic, explicit, and deliberate approach to creating, sharing, and using knowledge in order to enhance organizational performance.

Knowledge Sharing: The processes of transforming and transferring knowledge through an organization are designated by knowledge sharing.

Implementation Management of an E-Commerce-Enabled Enterprise Information System

Joseph Sarkis
Clark University, USA

R.P. Sundarraj
University of Waterloo, USA

INTRODUCTION

The integration of enterprise systems and the supply chain to an organization is becoming more critical in an ever-changing, globally competitive environment. Quick response will require close relationships, especially communications and information sharing among integrated internal functional groups as well as the suppliers and customers of an organization. Texas Instruments (TI), headquartered in Dallas, Texas, has come to realize this requirement for building and maintaining its competitive edge. Thus, it sought to implement an enterprise resource planning (ERP) system with a focus on linking it with a global electronic commerce (e-commerce) setting, an innovative and current issue (Weston, 2003).

There were a number of major players, including project management direction from Andersen Consulting Services, software vendors such as SAP and i2 Technologies, hardware vendors such as Sun Microsystems, and various suppliers and customers of TI.

The purpose of this case is to provide some aspects of implementation of strategic systems that provide valuable lessons for success. We begin and rely on the foundation of a strategic systems implementation model, which is initially described. A description of the case follows, with the various stages as related to strategic systems implementation described. We complete our discussion with implications and conclusions.

BACKGROUND

A process-oriented framework for ERP management is presented to help guide the discussion of this case (see Cliffe, 1998; Davenport, 1999; Miranda, 2002; Sarkis & Sundarraj, 2000).

The elements include the following:

- Strategy formulation and integration—One of the results of this step in the process is determination of an organization's core competencies that need specific technology support.
- Process planning and systems design—Also known as the reengineering phase, three studies are usually undertaken at this stage, and they are named AS-IS, SHOULD-BE, and TO-BE.
- System evaluation and justification—Here, analysis focuses on the economic, technical, and operational feasibility and justification of the system.
- System configuration—As a packaged software system, there are likely to be discrepancies (at the detailed level) between the needs of an organization and the features of the software. Hence, a significant amount of effort can be expected to configure the system or the organizational processes in order to produce an alignment between them.
- System implementation—The implementation stage can be classified into startup, project management, and a migration handing the switch over from the old to the new system.
- Postimplementation audit—This last "feedback" stage, although very important from a continuous-improvement perspective, is one of the more neglected steps.

As can be seen, the process suggested above can be arduous, but this necessary effort must be anticipated for the successful integration of complex and strategic systems into an organization.

IMPLEMENTING A GLOBAL ERP SYSTEM AT TI

Company Background

Texas Instruments Incorporated (TI) is a global semiconductor company and the world's leading designer and supplier of digital signal processing (DSP) solutions and analog technologies (semiconductors represent 84% of TI's revenue base). The company has manufacturing or sales operations in more than 25 countries and, in 1999, derived in excess of 67% of its revenues from sales to locations outside the United States. Prior to the implementation of ERP, TI had a complex suite of stand-alone nonintegrated marketing, sales, logistics, and planning systems consisting of thousands of programs that were based on many independent databases and were running on proprietary mainframe systems.

OVERVIEW

Since the 1980s, TI had used a highly centralized infrastructure utilizing proprietary mainframe computers for meeting its IT requirement. As the first step toward global business processes, certain planning processes and systems were standardized in 1989. Starting in 1996, TI underwent a company-wide reengineering effort that led to the implementation of a 4-year, $250 million ERP system using Sun Microsystems' hardware platform, SAP AG's ERP software, i2's advanced planning tools, and Andersen Consulting's implementation process. In 1998, Texas Instruments implemented the first release of the ERP system, which primarily consisted of a prototype implementation of the i2 system running on a Sun E10000 platform. In early 1999, TI began rolling out the second release. In the middle of 1999, TI completed the i2 Technologies software implementation as part of the third release. Finally, TI turned on the remaining financials, and new field sales, sales, and distribution modules. A high-level architecture of TI's pioneering ERP implementation consists of SAP and the i2 system for advanced planning and optimization. The system is a pioneering large-scale global single-instance implementation of seven modules (finance, procurement and materials management, logistics, planning, field sales, sales, and marketing) for all of TI's divisions, and it is in use by 10,000 TI employees to handle 45,000 semiconductor devices and 120,000 orders per month. This solution also enabled global Web access to information for TI's 3,000 external users at customer, distributor, and supplier sites.

STAGES IN MANAGING THE GLOBAL ERP SYSTEM IMPLEMENTATION

Strategy Formulation

Traditionally, TI was primarily running what was called a "commodity" business, wherein orders were received, manufactured, and shipped as a batch. Mass customization combined with the maturity of TI's business caused it to reexamine its goals and strategies. TI started its shift toward a more customized product environment.

Within this new customized product environment, TI had a number of customer needs that could not be met easily. Thus, the goal was to determine the appropriate processes and information systems that must be put in place in order to support such agile design and manufacturing strategies. Another goal was a move toward supplier-managed inventory and customer-managed orders. Finally, standardizing systems was another integrative corporate goal. TI made extensive use of metrics. Strategic goals are translated into tactical and operational quantifiable objectives.

Process Planning and Systems Design

TI conducted a massive reengineering effort for the whole organization, with the goal of setting standard processes globally. The major result of this effort was to declare that all inventory and manufacturing management be done globally.

TI decided to implement a single-instance ERP system so as to fully leverage the system's capabilities to support the flexibility and standardization demanded by global processes. After site visits by major ERP vendors, TI selected SAP mostly because of its scalability to handle voluminous amounts of data.

System Justification

A budget of approximately $250 million was set for the implementation. The justification of the system was done using a combination of tangible and intangible factors at both the enterprise and business-unit levels. Standard hard-justification measures such as ROI and IRR were used to ensure the financial viability of the project.

Through this business case justification, acceptable financial returns, along with strategic factors such as competing effectively within a given niche market, and operational factors, such as global inventory management, all played roles in ERP's justification at TI.

System Configuration

The goals and processes entailed numerous and significant changes to all aspects of the business process design of the system.

Implementation

In this phase, concepts and goals were translated into tangible action, and as a result, it is perhaps one of most difficult phases of the project. General principles such as global processes and standard systems need to be backed up by convincing and deploying the right people to implement the processes.

Startup

A number of key personnel, along with their families, were expatriated to the United States and stationed in Dallas, Texas, for a few years. About 250 people were transitioned from TI to Andersen Consulting that became the main provisioner of services with respect to the ERP system. IT outsourcing in this case involved Andersen Consulting taking over the employment and management of former TI people.

Project Management

Change management played a large role in this stage. The roles of training, planning, and communicating were of equal importance. All management levels were involved in this process as well as various vendors and suppliers. Some of the practices included the following:

- On-site experts were made available to new users of the system.
- A help desk was set up to handle problems that could not be addressed by these experts.
- A ticketing system for managing and prioritizing problems was also established (e.g., a system stop was a high-priority ticket that would get round-the-clock attention).

Handling Go-Live

To get prepared for "go-live," the key managers who were stationed in Dallas were sent back to their territories for educating the next level of users. Using selected experts, user-acceptance scripts were defined and tested, with problems, if any, being resolved as per one of the schemes outlined above. Daily conference calls were set up for 30 days prior to go-live to obtain status checks on progress and on the tickets.

Based on the results of these checks, a risk analysis was conducted weekly to determine the effects of various potential failures. The implementation plan was to have a few go-live dates one after another, but in relatively quick succession. Except for the planning system, in all the other stages, in this case, a direct conversion was employed. That is, with a downtime of about 2 to 3 hours during a weekend, the old system was turned off and the new one turned on.

Postimplementation Status

The system met most of its goals 9 months after the complete implementation. There are around 13,000 users on the system, with concurrent users ranging from 300 to 1,700. Some of the key performance measures and parameters evaluated were as follows:

- Productivity dip
- Ontime delivery
- Single-instance, global system
- Better response
- Inventory reduction

FUTURE TRENDS—MANAGERIAL IMPLICATIONS

The following lessons are summarized:

- Conduct a thorough strategic plan—The case illustrated how market forces had compelled the company to make radical shifts in its organizational environment and culture.
- Align IT plans with business plans—Conduct reengineering studies and develop strategic IT plans to align key IT needs with those of the business (Barker & Frolick, 2003).
- Get top management support—The prescription of top management support has been made ever since early IT implementations to the present (Mabert et al., 2001).
- Change management—Set realistic user expectations, such as the initial productivity dips. User involvement is critical. Andersen Consulting's process helped to ensure that such was the case. Make sure that the user was supported to help improve user satisfaction (Lee et al., 2003; Legare, 2002).
- Strong champion characteristics (Dean, 1987)—In TI's situation, the manager of the ERP project had over two decades of experience in various levels of the organization. This manager had broad knowledge of corporate operations. Previously, he was a vice president of one of TI's divisions.
- Rationalize business models and processes—Make sure the business models and processes fit within the strategic direction and goals of the organization. Time, mass customization, and flexibility concerns led to a global model (Gardiner et al., 2002). Global cultural issues were also a concern and needed to be managed (Davison, 2002).

- Manage external enterprises—Appropriate and well-planned involvement of consultants is important for keeping the project on a tight schedule. Further, with the advent of e-commerce, companies are more likely to ship and order goods on the basis of Web-based inputs.
- Manage using metrics (Skok et al., 2001; Hitt et al., 2002)—TI and Andersen Consulting have a corporate culture and policy that require the stringent and formal use of metrics in the management and evaluation of projects. They attribute this policy adherence as one of the key reasons for success of the ERP implementation. Key performance indicators included such issues as reduction in inventory, percentage of suppliers linked, and productivity of outputs.

CONCLUSION

TI's ERP implementation with an e-commerce perspective required a significant amount of features that added issues to its management:

- It is a single-instance system, providing access to the same data, irrespective of the geographic location of the user.
- It provides access to 3,000 external users (customers and suppliers), thereby enabling 70% of the transactions to be conducted electronically.

Management saw some problems in this implementation process and tried to address the issues. Some of the major problems included the following:

1. The software for supply chain management (Red Pepper) that was initially chosen did not meet expectations of TI. This system had to be scrapped, and this resulted in a multimillion dollar cost.
2. A productivity dip occurred, and the implementation had to address this issue for all managers throughout the organization who had some stake in the performance of the system. The expectations that this would occur were communicated through newsletters and messages. Consistent and continuous communication helped to mitigate a situation that could have caused a major project failure.
3. Getting buy-in from internal functions not directly associated with the implementation process was difficult. This occurred with the marketing function.

Future extension of developing appropriate infrastructure is another issue that needs to be faced (Kovacs & Paganelli, 2003). Lessons learned here may be appropriate for small or large organizations, but some differences appear in the practices of what is successful and not for ERP implementations in different size organizations. One of the major issues is that small companies get greater operational benefits from ERP, while larger companies get more financial benefits (Mabert et al., 2003).

REFERENCES

Barker, T., & Frolick, M. N. (2003). ERP implementation failure: A case study. *Information Systems Management, 20*(4), 43-49.

Cliffe, S. (1999). ERP implementation. *Harvard Business Review, 77*(1), 16.

Davenport, T. (1998). Putting the enterprise into the enterprise systems. *Harvard Business Review, 77*(4), 121-131.

Davison, R. (2002). Cultural complications of ERP. *Communications of the ACM, 45*(7), 109-111.

Dean, J. (1987). *Deciding to innovate: How firms justify advanced technology*. Cambridge, MA: Ballinger Publishing Company.

Gardiner, S. C., Hanna, J. B., & LaTour, M. S. (2002). ERP and the reengineering of industrial market processes. *Industrial Marketing Management, 31*(4), 357-365.

Hitt, L. M., Wu, D. J., & Zhou, X. (2002). Investment in enterprise resource planning: Business impact and productivity measures. *Journal of Management Information Systems, 19*(1), 71–98.

Kovacs, G. L., & Paganelli, P. (2003). A planning and management infrastructure for large, complex, distributed projects—Beyond ERP and SCM. *Computers in Industry, 51*(2), 165–183.

Lee, J., Siau, K., & Hong, S. (2003). Enterprise integration with ERP and EAI. *Communications of the ACM, 46*(2), 54-60.

Legare, T. L. (2002). The role of organizational factors in realizing ERP benefits. *Information Systems Management, 19*(4), 21-42.

Mabert, V. A., Soni, A., & Venkataramanan, M. A. (2001). Enterprise resource planning: Common myths versus evolving reality. *Business Horizons, 44*(3), 69-76.

Mabert, V. A., Soni, A., & Venkataramanan, M. A. (2003). The impact of organization size on enterprise resource planning (ERP) implementations in the U.S. manufacturing sector. *Omega, 31*(3), 235-236.

Miranda, R. (2002). Needs assessments and business case analysis for technology investment decisions. *Government Finance Review, 18*(5), 12-16.

Reich, B., & Benbasat, I. (2000). Factors that influence the social dimensions of alignment between business and information technology objectives. *MIS Quarterly, 24*(1), 81-113.

Sarkis, J., & Sundarraj, R. P. (2000). Factors for strategic evaluation of enterprise information technologies. *International Journal of Physical Distribution and Logistics Management, 30*(3/4), 196-220.

Skok, W., Kophamel, A., & Richardson, I. (2001). Diagnosing information systems success: Importance-performance maps in the health club industry. *Information and Management, 38*(7), 409.

Weston, F. C. (2003). ERP II: The extended enterprise system. *Business Horizons*, *46*(6), 49-55.

KEY TERMS

Audit: Reviewing and monitoring the performance of a system.

Enterprise Resource Planning (ERP) System: An information system that spans organizational boundaries with various organizational functional modules and systems integrated and managed by one system application.

Go-Live: The actual operation of an information system.

Mass Customization: Producing basically standardized goods but incorporating some degree of differentiation and customization.

Reengineering: Activities that seek to radically change business processes and support systems in an organization.

Single Instance: A one-time full-fledged company-wide initial operation of a system, as opposed to incremental (functionally or organizationally) or modular implementations.

Strategic Justification: The process of evaluating and selecting systems based on tangible (financial) and intangible factors that have implications for long-term and broad management of the organization.

Implementing an Online Academic Evaluation System

William S. Lightfoot
International University of Monaco, Monaco

INTRODUCTION

The explosive growth of networks and the World Wide Web over the past decade has led to increased adoption of new applications by educational institutions. Critical uses of software including courseware such as Blackboard, WebCT, and others have led to an increase in computer literacy amongst faculty and staff. Consistent with this increased literacy, newer applications including the use of the web for tests, surveys, and student evaluation has been increasing. There are many benefits of using web-based surveying and online evaluation techniques (see Table 1). As Mehta and Sivadas (1995) have noted, response times can be reduced from days and weeks to minutes and hours. It is common practice in many universities to have students provide feedback about courses in the form of a questionnaire, at the end of each course. As Felton, Stinson, and Mitchell (2003, p. 2) note, "these surveys are assumed to measure a professor's performance and successfulness in his or her various classes." McGourty, Scoles, and Thorpe (2002, p. 5) cited one of the key benefits as the "immediate availability of data for analysis and reporting and more extensive qualitative responses from students to the open-ended questions." Other studies (see Handwerk, Carson, & Blackwell, 2000; Hmieleski & Champagne, 2000) had similar conclusions. And although not all e-mail users check their e-mail everyday (Kent & Lee, 1999), past research found that e-mail responses generally occurred within the first three days (Comley, 1997; Mehta & Sivadas, 1995).

Strauss (1996) cited an additional benefit to web-based surveying in that the researcher can create surveys that are adaptive – meaning the software can present different questions, as well as audio, video, and pictures based on the users' responses and reactions to questions.

Table 1. Benefits of online educational evaluation systems

- Speed
- Time
- Accuracy
- Adaptability
- Anonymity
- Data Collection
- Data Evaluation
- Reach
- Ability to personalize

Other benefits include the ability for real time interactions with geographically diverse consumer or respondent groups as well as information servers (Kannan, Chang & Whinston, 2000). Results can be tabulated immediately, and interventions can take place to modify questionnaires, add new links, stop the project altogether. Hmieleski and Champagne (2000) suggested that online evaluation systems would enable faculty to adjust courses more frequently, leading to a better overall educational experience for students. Kannan et al. (2000) noted the value of increased bandwidth and the ability to transmit video and audio enables real-time feedback from respondents that can help research firms produce customized services more rapidly and at much lower cost than traditional research projects. Hardwerk, Carson, and Blackwell, (2000) cited the advantage that students could complete the evaluation on their own time, which the authors found led to more comprehensive, qualitative comments by the students.

The process for implementing an online evaluation system has characteristics that are similar to paper-based evaluation systems. Numerous authors have noted that many of the same factors may help increase response rate including advance notification, and personalization of correspondence (Dillman, 2000; Mehta & Sivadas, 1995; Witt & Poytner, 1997). Other variables, including credentials, the e-mail message, subject line, and saliency may also affect response rate. All of these need to be considered when developing an online evaluation system. Additional factors, including the time allocated for completion of the evaluation may also be important. McGorty et al. (2000) noted that at Drexel University, students were given three weeks to complete the evaluation, and that they are sent reminders via email from the system administrator, then the department head, and finally the undergraduate associate dean until they have completed the evaluation. In the same article, the authors noted that Columbia University used a similar approach, over a two-week response window. Online evaluation systems also have a number of potential problems, including "student concerns for privacy and anonymity, the culture to support online student evaluation processes, and insuring that students are aware of the evaluation process." (McGorty et al., 2000, p. 7). McGorty at al. added that initial response rates at Drexel were "dismal" (p. 8) and that after the use of "a combination of technology-mediated communications, incentive packages, and internal marketing strategies," they were still only able to increase the response rate to 50%. They also found that women, and students in their final years were more likely to complete the web-based evaluation than men, or students in their first couple of years of school. Hmieleski and Champagne (2000) added another dimension in noting that faculty buy-in for any evaluation system is critical to student response rate.

This paper examines the implementation of an online evaluation system at a small college in North Carolina. The online surveys used were adapted from paper-based instruments. Therefore the focus of this paper is on the implementation of the system from a technical and organizational perspective. And while the initial impetus was to convert paper-based student evaluations over to an online evaluation system, as the process of implementation moved forward, other applications were identified that were also suitable for an online educational evaluation system.

BACKGROUND ON THE PROJECT

With the rapid evolution of software, many options exist which allow a wide range of users to develop surveys, tests, and evaluation tools without any programming knowledge. Solutions range from open source software available on the web, to web site design tools like Macromedia Dreamweaver and Microsoft FrontPage, to embedded collaborative and communication systems in Microsoft's .net operating system (Sharepoint) and courseware applications (see Blackboard, and WebCT), and finally, to hosted solutions which allow users to create a wide range of tailor made instruments, or to choose a pre-made template designed for specific solutions. (*see Empliant.com or Zoomerang.com*)

ACADEMIC USES OF ONLINE EVALUATION TECHNIQUES

In academia, there are many needs for internal and external assessment of the overall state of the institution and its programs (Table 2). Career services is interested in gathering data on the success of their graduates, as well as information from employers about their needs from the institution; alumni affairs is interested in maintaining a current database of information about alumni; curriculum review committees seek input from alumni, employers, and students alike to help refresh and renew courses, and curricula; the enrollment management and admissions team may use surveys to gather data from prospective students, and their parents; individual departments within the college may also use the tools for specific, departmental assessment and data gathering purposes. Online, blended, and traditional classes may also use surveys for student evaluations of faculty, faculty evaluation of students (i.e. quizzes, tests, and exams), or for other purposes including market research.

Table 2. Academic uses of online survey instruments

- Student evaluation of faculty
- Faculty assessment of student performance (i.e. quizzes, tests, exams)
- Career progress of alumni
- Alumni updates
- Curriculum review
- Market research
- Program assessment

Many colleges struggle to implement any new information system. Finally, government education departments, accrediting bodies, and other stakeholders involved in the institution may require frequent feedback on the overall success of the institution. The biggest challenge lies in the implementation of a system that is easy to use, reliable, and produces the desired results.

The use of computers in research and assessment is well established. The use of online survey and evaluation systems, while frequently used, is still evolving consistent with the rapid changes in technology (Dillman, 2000). Regardless of the method used, there are some common steps that are critical to the proper implementation of any evaluation system:

1. The establishment of goals for the system. This helps insure that the results from the instruments are relevant and useful.
2. Determine the target population(s) and sample size(s) needed for useful results. To ensure external validity, the sample must be representative of the population from which it is drawn.
3. Choose the appropriate methodology to gather the data.

Failure to follow these steps may result in error or provide misleading and useless information.

Table 3. Comparison of response rates

Delivery Method	Sample Size	Response Rate
Online Evaluation	157	73%
In Class Evaluations	2443	86%

AN EXAMPLE: IMPLEMENTATION OF A ONLINE EVALUATION SYSTEM AT A COLLEGE

The first step in implementing an online evaluation system at the college was a challenge from the President to increase the use of technology across all departments. In parallel with this, the President increased funding for information technology initiating a port to pillow campaign for students, which was funded by the Parents Association. Funds were also raised for the complete renovation of computer facilities, and the upgrading of all personal computers (PCs) on campus. This happened within a 6-month time frame.

The second step was the creation of a Faculty Technology Committee that was initially charged with assessing the level of technology use by the Faculty. The committee then developed a web-based survey on the use of technology by faculty using a hosted survey from application service provider (ASP) Zoomerang.com, both to initiate the use of technology with faculty, and to assess their relative level of literacy with specific applications.

This led to two things: first, more faculty became interested in the use of the web for evaluations; and second, the committee identified specific areas of training needed amongst the faculty. A twelve month plan was initiated that included a combination of formal sessions on topics ranging from Microsoft Office applications, to the use of the web in the classroom, to less formal, personal coaching sessions between designated faculty members and their colleagues.

The interest in the use of online evaluations then accelerated greatly, as more and more faculty and staff members became familiar with its potential. This led to several pilot projects during the following academic year. The first involved piloting student evaluations of faculty. Three faculty members from the same academic division agreed to participate in this initial survey. The survey was developed using Microsoft FrontPage 2000 (FrontPage). The purpose of using FrontPage was initially due to limitations in Zoomerang.com, with regards to the number of questions that could be asked. An additional benefit to using FrontPage was that the survey could be customized in appearance and the data could be stored in a Microsoft Access database hosted on a campus server for more complete access to the data.

For this initial pilot test, a total of 157 students from 12 classes were asked to take the evaluations. A mixed delivery mode, including verbal and electronic notification of students, was used to ensure a suitable response rate. Students were given one week to complete the evaluations. The process was as follows:

1. Students were notified in class that an email would be sent out with instructions to complete the evaluation, and with a link to the evaluation.
2. The email was then sent out, with a one-week time frame during which they were expected to complete the evaluation.
3. The professors then reminded students in class, about the email, and the importance of completing the survey.
4. The administrator of the survey checked the results database daily, and provided the professors feedback on the completion rate for their courses.
5. Once the evaluation period was completed, the data was then tabulated, and compared to traditional in-class evaluations for response rate, and response quality. (Table 3)

The quality, as measured by the ratings assigned by the students to specific questions on the evaluation form (related to course content, the Instructor, and the class room facilities) was found to be comparable between the pilot group (Table 4) and the Division and the Institution as a whole. The means were within the first standard deviation, which suggests that the student responses were comparable. Issues of validity, bias, and reliability

Table 4. Comparison of pilot survey results versus institution/division averages

	Course Content	Instructor	Facilities
Institution	4.36 +/- .69	4.49 +/- .66	4.25 +/- .90
Division	4.33 +/- .76	4.44 +/- .73	4.55 +/- .57
Pilot Group	4.59 +/- .64	4.66 +/- .47	4.52 +/- .78

were also evaluated, and as all students were required to take a class in computer literacy, and as there were no students that did not have Internet access, these issues were not considered as major barriers.

While this pilot was going on, two other surveys were being piloted using the same software (Microsoft FrontPage). The first was for a survey of alumni for the alumni association administrator, and the second was for one of the professors who was completing her doctorate in philosophy. These were both identified by the technology committee as important initiatives - on the one had to support the administrative branch of the college, and on the other hand, to encourage external use of web-based surveys by faculty as a part of the professional development process.

After these initial pilot tests, the technology committee reviewed the results, and identified two significant barriers that could limit the effective adoption of web-based surveys at the college:

1. Dependency on a specific professor with FrontPage skills;
2. Lower than desired response rate to surveys.

One professor had led the technical aspects of the implementation with FrontPage. As no other professor had the skills, the concern was that too much support would be expected from this professor, and that it would take a significant effort to educate other faculty members on how to use FrontPage. Since one goal of implementing a web-based system was equal access to all faculty, staff and students, this was considered a major barrier.

The second issue of a lower than expected response rate was also considered a major barrier, as the school considered 80% the lowest acceptable response rate for student evaluations. While extending the time allowed for completion of the evaluations was a possible solution, the general belief was that if students were not asked to complete the evaluation when they were physically in class, then many would not.

THE SOLUTION

As a result of the evaluation of the pilot test, third party software packages, and hosted solutions were evaluated. The technology committee determined that at a minimum, any software had to have the following characteristics:

1. All users needed to be able design and implement surveys;
2. The surveys had to be accessible to all students, regardless of bandwidth;
3. Student anonymity had to be ensured;
4. Results were needed in two formats:
   a. Online for immediate access to faculty and staff;
   b. Downloadable for further analysis;
5. The supplier had to be reliable, with sufficient technical support;

After a thorough review of several options, Empliant was chosen because they met the technology committee's criteria, and also because they agreed to hire a student as an intern to assist in implementation for the college.

The next pilot test using Empliant's hosted solution was expanded to 10 faculty members. In addition to students being given one week to complete the survey, all classes included at least one 15-minute period in which the class was held in a computer lab so that students could complete the survey. The overall response rate increased to over 85%, with only a couple of classes below the required level of 80%. Further comparisons of the overall quality of data, yielded similar results to the first pilot test. This led to a decision to fully implement the new online evaluation system for all courses beginning with the next academic year.

CONCLUSIONS

The successful implementation of an online evaluation system was the result of a lot of time and energy by a few dedicated people. The institution where it was implemented had the right combination of critical success factors (Table 5). Executive support, coupled with the necessary funds, and a collaborative culture that encouraged experimentation helped ensure success.

The use of web-based surveys and online evaluation systems has increased the speed at which information can be processed and analyzed. This enables faculty members to receive feedback from their courses more quickly. It also helps assist other administrative departments in

Table 5. Critical success factors

- Executive Support
- Key investments in IT
  - Upgraded PC's/software
  - Support staff
- Collaborative culture
- Sense of empowerment
- Understanding of Student and Faculty Needs
- Well defined process
  - Pilot
- Faculty initiative
- IT department support
- Developmental funding

surveying alumni and other institutional stakeholders. This helps the college gather meaningful data that supports the institutions mission, and also accreditation and marketing processes.

FUTURE TRENDS

Future efforts that use additional media – including audio and video, plus more adaptive surveys may open up new opportunities for increasing the use of such systems. The key technical point for researchers to remember is that not all stakeholders have Internet access, and therefore the population is a subset of the whole. Total dependency on online systems can lead to response bias, by excluding critical input. Knowing the capabilities and limitations of the institution, and the technology will allow institutions to determine the level of utilization of an online evaluation system, while also increasing the speed at which data is gathered, and feedback provided.

REFERENCES

Comley, J. (1996). The use of the Internet as a data collection method. Retrieved October 12, 2000, from *http://www.sga.co.uk/esomar.html*

Dictionary.com (n.d.) Retrieved July 1, 2004, from *http://dictionary.reference.com*

Dillman, D. A. (1978). *Mail and telephone surveys: The total design method*. New York: John Wiley & Sons.

Dillman, D. A. (2000). *Mail and telephone surveys: The total design method.* New York: John Wiley & Sons.

Empliant.com. (2004). *College seeks low cost, high-tech approach to replace end of semester paper process for student evaluation of faculty*. Retrieved March 15, 2004, from *http://www.empliant.com/AboutUs/Faculty_Evaluations.cfm*

Felton, J., Mitchell, J., & Stinson, M. (2003).*Web-based student evaluations of professors: The relations between perceived quality, easiness, and sexiness*. Retrieved June 13, 2004, from *http://papers.ssrn.com/sol3/papers.cfm?abstract_id=426763*

Handwerk, P., Carson, C., & Blackwell, K. (2000, May). Online vs. paper-and-pencil surveying of students: A case study. Paper presented at the 40th Annual Meeting of the Association of Institutional Research.

Hmieleski, K. & Champagne, M. (2000). Plugging in to course evaluation. *The Technology Source*, September/October. Retrieved June 13, 2004, from *http://ts.mivu.org/default.asp?show=article&id=795*

Kannan, P.K., Change, A. M., & Whinston, A. B. (1998, March) Marketing information on the I-Way. *Communications of the ACM, 41*(3) 35. Retrieved October 1, 2001, from InfoTrac Web: Expanded Academic ASAP.

Kent, R. & Lee, M. (1999). Using the Internet for market research: A study of private trading on the Internet. *Journal of the Market Research Society*, *41*, 4, 377-385.

McCullough, D. (1998). Web-based market research, the dawning of a new era. *Direct Marketing*, *61*, 8, 36-39.

McGourty, J., Scoles, K., & Thorpe, S. (2000, June). Web-based student evaluation of instruction: Promises and pitfalls. Paper presented at the 42nd Annual Forum of the Association for Institutional Research, Toronto, CA. Retrieved June 13, 2004, *from http://www.drexel.edu/provost/ir/conf/webeval.pdf*

Mehta, R., & Sivadas, E. (1995). Comparing response rates and response content in mail versus e-mail surveys. *Journal of the Market Research Society*, *37*(4), 429-439.

Strauss, J. (1996). Early survey research on the Internet: Review, illustration and evaluation. *American Marketing Association Winter Educators' Conference Proceedings*.

Thorpe, Stephen W. (June, 2000). Online student evaluation of instruction: An investigation of non-response bias. Paper presented at the 42nd Annual Forum of the Association for Institutional Research, Toronto, CA. Retrieved June 13, 2004, from *http://www.drexel.edu/provost/ir/conf/bias.pdf*

Witt, K. & Poytner, R. (1997). The do's and don'ts of internet interviewing. Retrieved January 18, 2001 from, *http://www.ds.dial.pipex.com/town/parade/ny03/esomarp1.htm*

KEY TERMS

ASP: Application service provider.

Bandwidth: The amount of data that can be passed along a communications channel in a given period of time. (Dictionary.com)

Courseware: Educational software designed especially for classroom use. (Dictionary.com)

Critical Success Factors: A key area where satisfactory performance is required for the organization to achieve its goals (Austin, 2002).

Open Source: Method and philosophy for software licensing and distribution designed to encourage use and improvement of software written by volunteers by ensuring that anyone can copy the source code and modify it freely.

Pilot Study: Serving as a tentative model for future experiment or development: a pilot project; The pursuit of knowledge, as by reading, observation, or research. (Dictionary.com)

Reach: To succeed in getting in contact with or communicating with. (Dictionary.com)

Web-Host: A computer containing data or programs that another computer can access by means of a network or modem. (Dictionary.com)

Implementing the Shared Event Paradigm

Dirk Trossen
Nokia Research Center, USA

Erik Molenaar
Technical University of Aachen, Germany

INTRODUCTION

For collaboration among users, sharing audio-visual, textual, graphical, or even interface-related information is the essence of *computer-supported collaborative work* (CSCW). Since most applications that are being used in private and work life these days are merely usable on the computer on which they are executed, collaboratively working with a single application is the most challenging part of CSCW. This is not only true because these applications are unaware that they are executed in a distributed environment, but, in particular, because of the numerous possibilities of data to be shared among the distributed users. Thus, the distribution of the application's functionality over the network must be added transparently and, more important, subsequently without changing the application's semantic. The effect has to be created at each remote site that the application is running locally and, therefore, can also be controlled by any remote user with a more or less immediate effect to the application. This problem is referred to as *application sharing* in the remainder of this article.

BACKGROUND

The realization of application sharing faces several challenges to be solved, as explained in more detail in Trossen (2001):

- *Amount of transferred data* is part of the indicator for the generated network load.
- Each technique adds certain *interception points* to the local system to gather required information to be distributed among the session members. The number of these points serves as an indicator of the generated network and processor load (Trossen, 2001).
- *Heterogeneity*: Sharing applications independent from each member's operating system is crucial for wide applicability of the technique.
- *Latecomer's support*: Joining the session later should be supported without leading to inconsistencies of the application's state.
- *Shared data problem*: Using any kind of input data within the shared application should not lead to inconsistencies of the distributed copies of the application.
- *Synchronization*: The shared instances of the application have to be synchronized to ensure consistency of the workspace due to the different processing speed of the sites and the different delays of the transmission lines.

Two different paradigms can be distinguished to tackle the earlier mentioned challenges, namely, *Output Sharing* and *Event Sharing*. In Trossen (2001), a qualitative comparison and detailed presentation of both paradigms is available, including the different application scenarios for both paradigms.

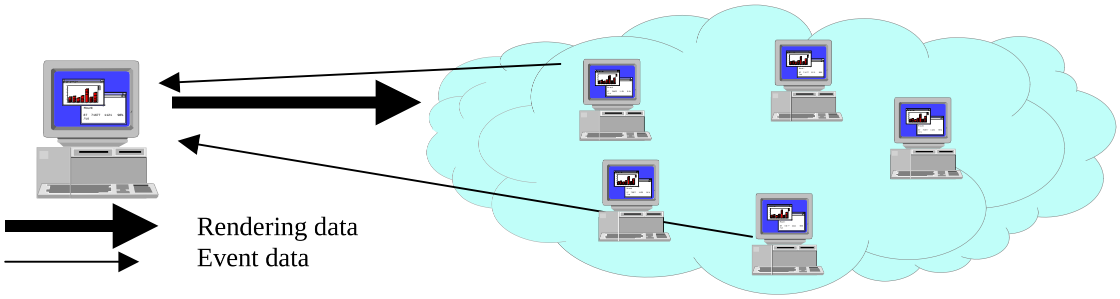
GUI Sharing Technique

The first technique is to share the server's application's output. For feedback from the receivers, any input data like mouse or keyboard events is transferred back to the sender and fed into its local event loop for control, subject to the current floor control policy. Rendering data is transferred from the server to the receiver group, preferably using a reliable multicast protocol.

For that, as indicated in Figure 1, *event data* is sent back to the server to be fed into its local event loop for a remote control of the application. Usually, transferring event data to the server is controlled by means of *floor control* (Hao & Sventek, 1996), that is, the appropriate host is selected based on a *social protocol* (Dommel & Garcia-Luna-Aceves, 1995) with an associated floor, representing the right to control the application.

Latecomer's support is provided by invoking a full refresh of the local GUI, resulting in a transfer of the entire GUI content to the receiver group. Furthermore, the shared GUI approach allows a heterogeneous receiver group,

Figure 1. GUI sharing approach



assuming appropriate rendering engines on the client's side. As shown earlier, the input event data is the only data to be synchronized with the local application, which is realized by means of floor control. Any additional data, like files or local device data, is held locally with the server's host. Hence, there is no *shared data* problem to deal with. However, the different processing speeds of the client rendering engines have to be considered for synchronization of the workspace. For that, *synchronization points* can be used, which have to be acknowledged by each member.

Event Sharing Technique

The assumption being made is that if a set of identical applications is executed with the same start state and evolves using the same sequence of events, its timeline evolution is identical on each site. Hence, the following statement is valid (Proof in Trossen, 2001):

- Theorem 1: A set of application instances that behaves deterministically can be held in a stable state if the starting state and all events can be captured.

In contrast to the shared GUI approach, there is no central server. The initiator of the session is merely used for defining the application's start state. Any input data is transferred from the current floor holder to *all* group members. There is no central entity to which the data is sent first to determine the new output. In this, *homogeneity* of the environment is crucial due to the requirement of having a local application instance.

Applicability of the Techniques

Each approach for tackling the application sharing problem has its specific advantages and weaknesses. The GUI sharing approach is well suited for heterogeneous environments and when using input data which cannot be shared among the other participants.

However, the event sharing approach has also specific advantages, which makes this technique attractive for specific scenarios. Due to the local copy of the application, the additional load on each host is expected to be much lower, which increases the responsiveness of the system and thus improves the user's perception of the system. However, the problem of ensuring the consistency of each user's view when using shared data restricts the applicability of the approach either to not using shared input data or to use the technique in local environments where data sharing is feasible to some extent. Furthermore, this technique is not applicable in heterogeneous scenarios.

Table 1 shows typical scenarios for shared applications and the applicability of both paradigms in these scenarios. It is worth mentioning that the list is only meant to outline sample scenarios; thus, the list is neither exhaustive nor exclusive.

It can be seen that the event sharing technique is not applicable to the last two scenarios due to the heterogeneous character of these situations, while the first three scenarios are fairly good examples as the shared event approach promises to provide a higher responsiveness of the system and, therefore, an improved user's perception. Specifically, the multimedia presentation is hardly conceivable using the shared GUI approach due to the large amount of data to be transferred, which is avoided by the local copy of the application when using the shared event technique. Furthermore, due to the local character of the scenarios, the shared data problem can be handled much easier.

It can be summarized that the event sharing technique is better suited for local environments and high demands on the responsiveness of the shared application, while the GUI sharing approach is preferred in heterogeneous environments and when having problems with data to be shared.

Related Work

In the past, several toolkits have been developed in this problem space, such as the *X Windows* system (Israel & Fortune, 1992), comprised of a central server on which the application is executed. The application's output is redirected to *X Windows clients* for rendering. Extending this system to a multipoint scenario, which occurred in Altenhofen et al. (1993), Minenko and Schweitzer (1994), and Wolf, Froitzheim, and Schulthess (1995), enables a shared application system for cooperative working. However, floor control capabilities have to be added for coordinated control (Altenhofen et al., 1993; Minenko & Schweitzer, 1994).

Despite the wide deployment of X Windows systems, their applicability is mainly restricted to Unix systems. Although X Windows client software is available, the problem remains to share, for example, MS Windows software on other platforms. To tackle this heterogeneity problem, the ITU proposed a multipoint application sharing protocol (ITU-T, 1998a), defining platform-independent rendering and interception functionality. However, the overhead on the server system and the usage of an ineffective transport system, which is defined in the ITU T.120 standard (ITU-T, 1998b), are the main disadvantages of this solution.

The work in Hao and Sventek (1996) replicates the entire data workspace before starting and executing application copies at each computer, as defined earlier as the *event sharing* paradigm. Synchronization among the different copies is ensured for every incoming event, leading to a significant overhead instead of using specific synchronization events for overhead reduction. Moreover, the event mapping and distribution is realized on a central server. Hence, the proposal follows a distributed application but a centralized control approach.

GENERAL REQUIREMENTS FOR A SOLUTION

Apart from the major requirement that the application to be shared must behave deterministically, the following requirements can be defined for an event sharing solution:

- All instances must start in an equivalent state.
- All events changing the application's state must be captured and broadcast to all participants.
- If events access resources, these must be provided to all participants.
- Synchronization of instances must be offered.

In addition, the following minor requirements have to be addressed:

- An interface with the participant must be offered.
- Latecomer's support must be addressed.

Since a shared application service is using resources from existing conferencing systems, such as Trossen (2000), the following requirements for this part of the system can be derived:

- Conference management should be provided.
- Floor control is required to prevent concurrent control of the application.
- Reliable message transport will be provided.
- If possible, multicast capabilities will be utilized.

SAMPLE SOLUTION: THE MULTIPOINT EVENT SHARING SERVICE (MESS)

The earlier mentioned requirements are used as a foundation for the design of the *Multipoint Event Sharing*

Table 1. Scenario examples and paradigms applicability

Scenario Description	Shared GUI	Shared Event
Multimedia presentation in a local environment	--	++
Programming environment in a lecture	-	++
Development environment in a closed user group	-	+
Spreadsheet in an heterogeneous Internet environment	++	-
Accompanying presentation in an Internet lecture	++	-

Service (MESS), which is presented in the next section as a sample solution for implementing the event sharing paradigm.

Architecture

Figure 2 left shows the components of the MESS architecture. The concept of capturing the start state and distributing the evolving state is reflected through the *Starter/StaticReplicator* and the *DynamicReplicator* components. The former takes care of all instances of the application to be in an equivalent state upon startup. It decides what resources to distribute, and it takes care that the local settings for each participant are brought in a consistent state. The *DynamicReplicator* is responsible for keeping all participating applications in an equivalent state after the session has started.

At the currently controlling application side, the *Interceptor* gathers required event information. Two types of events need to be handled, namely *user events* and *system events*. Examples for the former are mouse movement or keyboard keys pressed. The system events are used to keep the application instances in synchronization by monitoring the speed of the end systems. The resources that are used by both events are recognized through the *ResourceGrabber*. The actual sending of needed resources and events is prepared by the *Sender/Synchronizer*. This component also takes care of the synchronization and offers latecomer support. The required conferencing and data transmission functionality is provided by the *Conferencing* component, and the interaction with the participant and coordination of components are performed through the *Controller*.

This component architecture can be transformed in a UML framework, shown in the right part of Figure 2. It is a straightforward mapping of the components onto classes with dedicated methods.

Realization and Lessons Learned

The proposed MESS architecture is implemented as a prototype. Although the current design allows for sharing all types of events and resources, the actual implementation must make certain tradeoffs to keep the realization simple and feasible, but to demonstrate the potential of the proposal. As a consequence, no system events are shared at this time, and there is no synchronization capability. Moreover, resource distribution is not provided. This basic functionality reflects the most important part of the service, namely the evolution of the application's state, and, therefore, demonstrates the ability of the concept to provide application sharing for certain scenarios (see also Trossen, 2001).

As demonstration scenarios, simple text-editing as well as rotation of complex 3D objects are performed. The latter happens in shared engineering scenarios, as described in Trossen, Schueppen, and Wallbaum (2001), and is well suited to demonstrate the potential since it generates heavy graphical output that is difficult to capture appropriately with output sharing systems. However, due to the missing synchronization functionality of our prototype, the computers usually run out of sync after a certain timeframe.

The performance can be evaluated in two dimensions. First, the added overhead to the system due to the interception to gather and distribute event information is a major performance measure. The demonstration application shows that this additional overhead is fairly small; however, adding more system events and resources certainly decreases the overall performance. As a second measure, bandwidth consumption is important. For that, assume a reasonable amount of entered text, for example, 250 characters per minute. Further, assume one sync event after each pressed key as a conservative approach. Thus, the bandwidth consumption would be less than 700 bits per second with an event size of 16 bytes and a synchronization message size of 2 bytes. In the example of rotating 3D objects, this consumed bandwidth is even smaller since user events are generated with a smaller frequency.

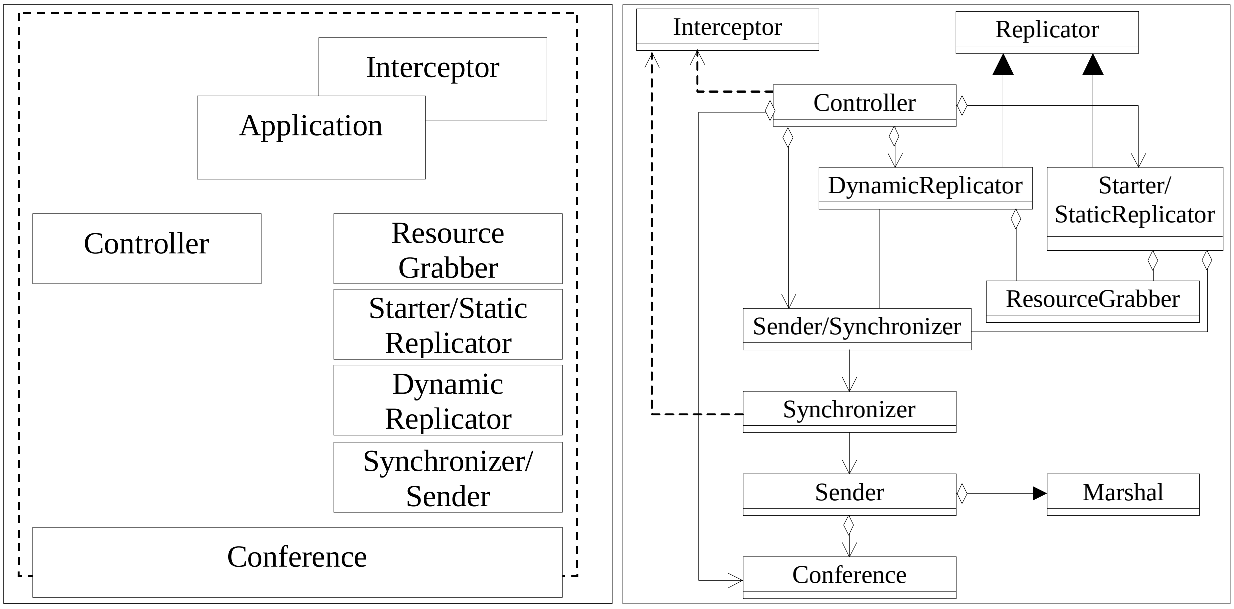
To summarize, the realization of the MESS architecture showed that the basic concept of event sharing works with an impressive speed by leveraging local processing speed for the application functionality. However, the integration of some system events with an additional synchronization to cope with different processing speeds is likely to increase the spectrum of applications that can be used with the system.

FUTURE TRENDS

We see the future trends in collaborative workspaces evolve towards studying integrative approaches of the different collaborative technologies involved. This particularly concerns integration of the different solutions into the Internet Multimedia Architecture, placing the application layer functionality on top of standard signaling and transport protocols.

In our particular future work, the restricted functionality of our demonstrator is to be increased, starting with synchronization functionality to cope with out-of-sync

Figure 2. MESS architecture



effects. Moreover, finding some optima for the applications that can be served by MESS while keeping the used bandwidth to a minimum is a field of future work. In addition to enriching functionality, more systematic evaluation scenarios must be defined to determine a clear view of the added system overhead. In addition, the demonstration system is used within a project, realizing a workspace for shared engineering (Trossen, Schueppen, & Wallbaum, 2001), to show its feasibility in a real-life use case.

CONCLUSION

This article presents issues around implementing the event sharing paradigm as an approach to solving the application sharing problem in the Internet. For that, we have outlined the basic techniques of two different paradigms to tackle the problem of sharing applications remotely, namely sharing an application's output or sharing the application's state via the network. In this, the event sharing paradigm seems to be a promising candidate to increase the efficiency of application sharing in particular for scenarios that have an inherent locality of shared data and a homogeneity of end-systems. For the realization of this paradigm, we have defined a set of general requirements that we see must be fulfilled by an event sharing solution. These requirements are the foundation for our MESS architecture design and implementation, which is presented as a proof-of-concept solution. The components of this architecture are outlined together with the realized objects.

Our prototype shows promising results with respect to its simplicity on the one hand and the increased performance on the other hand, although the simplicity must be re-evaluated when considered in a full-blown implementation. However, it is shown that with a fairly simple set of events, a wide spectrum of applications is enabled. Although there remains future work to be done, we show with our implementation that the concept of sharing application events in order to share the application remotely is a viable one.

REFERENCES

Altenhofen, M. et al. (1993). The BERKOM multimedia collaboration service. *Proceedings ACM Multimedia.*

Dommel, H.-P., & Garcia-Luna-Aceves, J.J. (1995). Floor control for activity coordination in networked multimedia applications. *Proceedings of the 2nd Asian-Pacific Conference on Communications.*

Hao, M.C., & Sventek, J.S. (1996). Collaborative design using your favorite 3D application. *Proceedings of the IEEE Conference on Concurrent Engineering.*

Israel, E., & Fortune, E. (1992). *The X Window system server*. Digital Press.

ITU-T. (1998a). *Multipoint application sharing.* ITU-T Recommendation T.128.

ITU-T. (1998b). *Data protocols for multimedia conferencing.* ITU-T Recommendation T.120.

Minenko, W., & Schweitzer, J. (1994). An advanced application sharing system for synchronous collaboration in heterogeneous environment. *SIGOIS Bulletin, 15*(2), 40-44.

Trossen, D. (2000). Scalable conferencing support for tightly-coupled environments: Services, mechanisms, and implementation design. *Proceedings of the IEEE International Conference on Communications.*

Trossen, D. (2001). Application sharing technology: Sharing the application or its GUI? *Proceedings of the IRMA Conference.*

Trossen, D., & Helbig, T. (1997). The ITU T.120 standard family as basis for conferencing applications. *Proceedings of the SPIE International Symposium Voice, Video, & Data Communications.*

Trossen, D., Schueppen, A., & Wallbaum, M. (2001). Shared workspace for collaborative engineering, *Annals of Cases on Information Technology* (vol. IV).

Wolf, K.H., Froitzheim, K., & Schulthess, P. (1995). Multimedia application sharing in a heterogeneous environment. *Proceedings of the ACM Multimedia.*

KEY TERMS

Application State: Current snapshot of the application itself and all of the resources it addresses.

Deterministic Behavior: If a set of a shared application is started in an equivalent state, and the same set of events is presented to those instances, then if the same state transitions will happen for all instances, the application has deterministic behavior. It is important to note that this definition is more relaxed than other ones, in the sense that resources that an application might need are considered as part of the environment. Where other definitions might assume that an application is no longer behaving deterministically if, for example, the system time of the local machine is used, this definition regards the system time as a part of the environment.

Events: A phenomena that changes the application's state.

Event Sharing Paradigm: Shared applications that are realized based on the event sharing paradigm instantiate a copy of the application on each participating host, distribute an initial start state among all copies and further distribute all future events among those copies that result in a state transition of the applications.

Floor Control: Functionality that allows for resolving concurrent access to common resources, such as mouse pointers in shared applications. A floor is usually assigned to such resource together with a policy for access authorization.

GUI Sharing Paradigm: A shared application follows the GUI sharing paradigm if the output of a centralized application is shared among all participants of the conference. Feedback from the participants is sent to the centralized server to be incorporated in the control of the application, if desired.

Resources: Anything that is not the application itself, but is changed or used by the application to determine its behavior. Examples are files, registry entries, or the system time.

Stable State: A stable state of the shared application is given, if the execution behavior of all instances is equivalent. For example, if a menu entry is selected, the same action belonging to the corresponding menu entry should be performed on all machines. Other terms are *consistent* or *equivalent state*.

Improving Public Sector Service Delivery through Knowledge Sharing

Gillian H. Wright
Manchester Metropolitan University Business School, UK

W. Andrew Taylor
University of Bradford, UK

INTRODUCTION

Since the publication of the first knowledge management article in *Harvard Business Review* (Nonaka, 1991), the world has witnessed a revolution in management practice. While the origins of knowledge management extend further back in history (see Prusak, 2001; Wiig, 1997), it is certainly true that in the last decade the creation, sharing and application of knowledge are increasingly seen as a source of competitive advantage. However, knowledge management is largely a private sector innovation at the present time, although gradually moving towards the public service sector (Bate & Robert, 2002; Hartley & Allison, 2002). The implementation of knowledge management places an emphasis on organizational factors such as learning capability, culture and leadership as well as renewed focus on the importance of information quality (Alavi & Leidner, 2001). The ability to manage the sharing of information (and hence knowledge) effectively remains one of the most important but still least understood activities in modern organizations, no less so in public services.

BACKGROUND: KNOWLEDGE SHARING IN CONTEMPORARY PARTNERSHIP ORGANIZATIONS

Public services represent a significant economic sector in most countries and public demands on services are increasingly consumerist. This has led to escalating scrutiny of the performance of public services. Consequently, the strategic use of information and knowledge to improve service delivery and financial performance has become a key skill for managers in this sector. Partnership working represents a formal departure from the traditional compartmentalized approach to public service delivery. Often referred to as joined-up thinking, partnership working challenges existing hierarchies, encouraging the partner organizations to work together at all levels, including strategy, service planning and service delivery to enhance efficiency and improve user experience and satisfaction. For partners to work effectively together, knowledge of best practices must be shared and utilized towards the common goal of improving the overall quality of service delivery.

Our research has focused on health and social care as an area of public service in which organizations responsible for commissioning and delivering all aspects of care are increasingly expected to work together, to reduce fragmentation of access to the user. Management of the provision of high quality public services continues to be a major social and political issue in many countries. Our research was conducted in the context of UK national policies for performance management (DETR, 2001a), partnership working (DETR, 2000, 2001b; Fordham, 1998), the reduction of health inequalities (DoH, 1998a, 1999), and overall improvements in service quality (DoH, 1998b, 2000). We have concentrated particularly on the issue of making public service partnerships work effectively, to achieve strategic objectives, that is, to improve individual health and personal well being as well as to achieve gains in public health. Specifically, our research questions relate to assessing the readiness of the partners to work together, and to share knowledge that each possesses about their part in the overall service delivery process. By understanding the factors that influence effective knowledge sharing, managers can take practical steps toward improving these antecedent preconditions.

MANAGING THE ANTECEDENTS TO KNOWLEDGE SHARING

The key to partnerships is a focus on the creation of an explicit understanding of what needs to be done to meet strategic objectives –akin to Choo's concept of a "knowing organization" (Choo, 1988). We conceptualize the role of knowledge in the partnership process (Figure 1) in terms of two core aspects, viz:

- The effective management of information to support the vertical deployment of organizational strat-

Figure 1. The public service partnership: The knowing organization

egy in terms of communication and development of meaningful performance measures, and
- The wider organizational culture to support attitudes conducive to new ways of working.

We have identified six key factors that are associated with successful knowledge sharing in public sector partnerships (Wright & Taylor, 2003), namely:

- Innovative culture
- Change readiness
- Information quality
- Clarity of responsibility
- Strategy formulation and deployment
- Accountability

Innovative Culture

An innovative culture is one where people are receptive, rather than resistant to, new ideas, and where they are motivated to embrace and develop these ideas and shape them into improved working practices. Such cultures provide people with time to reflect, to learn from both success and failure, providing supporting systems to facilitate reflection and capture lessons learned. Finally, innovation is focused on the user or customer, whereby people actively search for new ways of improving service delivery. The legal and political constraints on public service managers and persistent demands for strict oversight can lead to rigidity and bureaucracy in public sector organizations that counter the development of an open and inclusive culture (Scott & Falcone, 1998).

Change Readiness

Change and innovation are closely linked. An innovative culture needs to be able to implement changes to working practices and behaviors generated by innovation. This requires a positive attitude to doing things differently, rather than seeking to maintain the status quo. Change requires leadership, to proactively seek opinions and listen to views whilst engendering an atmosphere where ideas are freely expressed and there is no perception of a need for staff to cover their backs to protect themselves from criticism and retribution. Involvement and commitment will decline and the organization's innovative potential will be diminished if there is a culture of reluctance to challenge current ways of working. High levels of media scrutiny of public sector organizations (Perry & Rainey, 1988) and the top-down nature of government-imposed changes (Collier, Fishwick & Johnson, 2001) can reduce public sector employees' receptivity to change (Halachmi & Bovaird, 1997). Being ready to change implies a concomitant sensing of the need to change. Information about performance gaps, that highlights the need for change, must be communicated throughout an organization. If people feel that managers pay little attention to performance statistics, they too will ignore them and continue working in ways that maintain the status quo. The nature of the change and the benefits that it will bring need to be understood.

Quality of Information

Good quality information facilitates performance review, and reflection on service delivery. It supports people in their work tasks and it provides a medium for the capture and dissemination of lessons learned. If timely and meaningful information is not provided, people will find it difficult to know how well they are performing, and they will spend extra time searching for the information they really need. Public sector organizations often place less importance on the quality of information and perceive less need to invest in information systems (Rocheleau & Wu, 2002). Unless there is clarity about the basis of performance measurement, information systems will not be perceived as providing appropriate support.

Clarity of Responsibility

It is important for people to understand their specific roles and responsibilities and to know whom to contact elsewhere in the service value chain. People need to see clearly how their jobs fit into the bigger picture, to have their responsibilities delineated clearly in relation to the organizational strategy, and to see how their roles contribute to its achievement. Thus clarity of responsibility is concerned with the effectiveness of strategy delivery. Managers must ensure that people can grasp the significance of strategy in relation to their own responsibilities, and that the performance measures that derive from the strategic process (see Figure 1) are useable for managing service delivery. All too often, public sector performance measures are regarded by employees as vague and of limited utility (Townley, 2002). To achieve clarity of responsibility in partnerships there needs to be:

- A joint strategy developed, and owned by all partners
- Re-definition of the service value chain and its business processes
- Re-examination of roles and responsibilities within the value chain
- Development of partnership-based performance measures
- Logical and explicit derivation of performance measures from the joint strategy

Strategic Connections

This factor addresses communication gaps in the strategy formulation and deployment process. Public sector managers must ensure that people feel involved in the strategy formulation process and do not feel that strategy is imposed from on high, as is often the case in public services. Staff need to understand the meaning of strategy in their own situations. Strategic plans need to be living documents, owned by all, rather than uninspiring rhetoric that gathers dust on a shelf in the strategist's office. In public service partnerships, strategy must be communicated in terms of improved relationships and outcomes for service users, rather than a structural end in itself.

Accountability

People need to know where the buck stops. In an integrated service delivery chain that crosses organizational boundaries, it must be made clear who is ultimately accountable for performance. Poor redefinition of accountability is symptomatic of, and consistent with, top-down imposition of strategy with little staff involvement, and poor deployment of strategies into meaningful processes and activities.

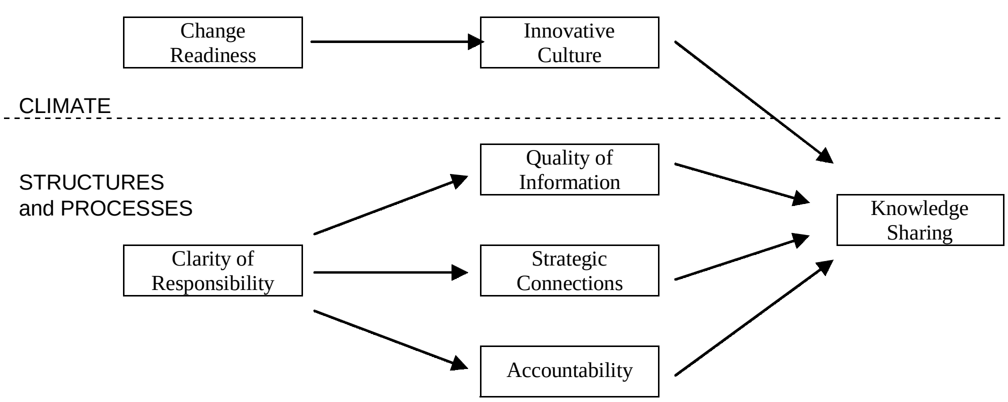
FUTURE TRENDS: MANAGING KNOWLEDGE SHARING

Six antecedents to successful knowledge sharing are illustrated in Figure 2, and these form a basis for the management of knowledge sharing. Increasingly, organizations will move from simply practicing knowledge management and knowledge sharing, toward an emphasis on evaluating its effectiveness. Using this model as a basis for diagnosing an organizational climate that is supportive of knowledge sharing, a targeted, proactive approach can be developed for implementing and maintaining effective knowledge sharing.

This empirically validated model indicates that these six factors are important precursors in the development of an effective knowledge sharing process and can give guidance to managers about where they should focus their efforts. We have found that an innovative culture is the strongest factor in explaining knowledge sharing, with acceptance of new ideas being the strongest element within this factor.

Our research identified the issues of trust and power as being underpinning and tacit impacts on knowledge sharing and our factors include items that embody trust and power (Dirks & Ferrin, 2001). For example, trust can result in higher levels of co-operation, increased willingness to take risks (Mayer, Davis & Schoorman, 1995), increased involvement of employees in decision making (Spreitzer & Mishra, 1999), increased sharing of informa-

Figure 2. Predictors of knowledge sharing



tion and greater satisfaction with organizational change programs (Rousseau & Tijoriwala, 1999). Thus this model incorporates the effects of trust and power in the ways in which they are experienced in a knowledge sharing environment.

CONCLUSION: A MANAGEMENT AGENDA FOR KNOWLEDGE SHARING AND INNOVATION

There are some clear antecedents to the management of effective knowledge sharing in partnerships developed to enhance public services. The change implications of knowledge sharing are very important, and are represented strongly in the innovative culture and change readiness factors. These factors are well established as key strategic management challenges and underline that knowledge sharing is a key strategic issue, rather than an IT-centred initiative, as it is often portrayed. While public policy usually addresses structural and process issues, our research suggests that this may inadvertently be misguided and that the key to improved service delivery lies in changes to the underlying culture, particularly with regard to an innovative orientation and change readiness. Applying structural solutions to behavioral problems is not recommended.

This model of knowledge sharing suggests a management agenda to improve service delivery through knowledge sharing. To develop an innovative culture, there must be a climate where both motivation to innovate and acceptance of new ideas are equally strong, and further, there must be appropriate systems in place to facilitate learning and reflection. To generate a shared understanding of best practices and a sharing of knowledge about lessons learned, staff in such partnerships need to be given time and opportunities to engage in socialisation processes that facilitate learning (Nonaka & Takeuchi, 1995). Organizations need to take on board Garvin's assertion that while much knowledge can be generated from reflecting upon success, there is even more to be gained from reflection on failure (Garvin, 1993). The changes needed to move from a departmental focus to an inter-organizational one point very forcibly to the organizational culture and especially the role of senior management.

Public services are embedded in an environment wherein government requirements for performance reporting and accountability often seem to have little relevance for informing staff about what needs to be improved. The intangibility of public services means that public service managers rely very much on knowledge – insight, understanding and empathy. Extant studies of knowledge have generally focussed on the private sector. However, the prominence of a knowledge sharing culture is clearly a key element in achieving a partnership-based, user-focused public service.

REFERENCES

Alavi, M., & Leidner, D.E. (2001). Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly, 25*(1), 107.

Bate, S.P., & Robert, G. (2002). Knowledge management and communities of practice in the private sector: Lessons for modernizing the National Health Service in England and Wales. *Public Administration, 80*(4), 643-684.

Choo, C.W. (1988). *The knowing organisation*. New York: Oxford University Press.

Collier, N., Fishwick, F., & Johnson, G. (2001). The processes of strategy development in the public sector. In K. Scholes (Ed.), *Exploring public sector strategy* (pp. 17-32). London: Pearson Education.

DETR. (2000). *Local government act*. London: Department of the Environment Transport and the Regions, the Stationary Office.

DETR. (2001a). *Best value and audit commission indicators for 2001/2002: Consultation*. London: Department of the Environment Transport and the Regions, the Stationary Office.

DETR. (2001b). *Local strategic partnerships: Government guidance*. London: Department of the Environment Transport and the Regions, the Stationary Office.

Dirks, K.T., & Ferrin, D.L. (2001). The role of trust on organizational settings. *Organization Science, 12*, 450-467.

DoH. (1998a). *Independent inquiry into inequalities in health report (the Acheson report)*. London: Department of Health, The Stationary Office.

DoH. (1998b). *The new NHS: A national framework for assessing performance: Consultation document*. London: Department of Health, NHS Executive.

DoH. (1999). *Reducing health inequalities: An action report*. London: The Stationary Office.

DoH. (2000). *NHS plan: A plan for investment, a plan for reform*. London: Department of Health, the Stationary Office: Cm 4818.

Fordham, G. (1998). *Building partnerships in the English regions: A guide to good practice*. DETR.

Garvin, D.A. (1993). Building a learning organization. *Harvard Business Review, 71*(4), 78-92.

Halachmi, A., & Bovaird, T. (1997). Process reengineering in the public sector: Learning some private sector lessons. *Technovation, 17*(5), 227-235.

Hartley, J., & Allison, M. (2002). Good, better, best? Inter-organizational learning in a network of local authorities. *Public Management Review, 4*(1), 101-118.

Mayer, R.C., Davis, J.H., & Schoorman, F.D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20*, 709-734.

Nonaka, I. (1991). The knowledge-creating company. *Harvard Business Review, 69*, 96-104.

Nonaka, I., & Takeuchi, H. (1995). *The knowledge creating company: How Japanese companies create the dynamics of innovation*. New York: Oxford University Press.

Perry, J.L., & Rainey, H.G. (1988). The public-private distinction in organization theory: A critique and research strategy. *Academy of Management Review, 13*(2), 182-201.

Prusak, L. (2001). Where did knowledge management come from? *IBM Systems Journal, 40*(4), 1002-1007.

Rocheleau, B., & Wu, L. (2002). Public versus private information systems: Do they differ in important ways? A review and empirical test. *American Review of Public Administration, 32*(4), 379-397.

Rousseau, D., & Tijoriwala, S. (1999). What's a good reason to change? Motivated reasoning and social accounts in promoting organizational change. *Journal of Applied Psychology, 84*, 514-528.

Scott, P.G., & Falcone, S. (1998). Comparing public and private organizations: An exploratory analysis of three frameworks. *American Review of Public Administration, 28*(2), 126-145.

Spreitzer, G., & Mishra, A. (1999). Giving up control without losing control. *Group and Organization Management, 24*, 155-187.

Townley, B. (2002). The role of competing rationalities in institutional change. *Academy of Management Journal, 45*(1), 163-179.

Wiig, K.M. (1997). Knowledge management: Where did it come from and where will it go? *Expert Systems with Applications, 13*(1), 1-14.

Wright, G.H., & Taylor, W. A. (2003). Strategic knowledge sharing for improved public service delivery: Managing an innovative culture for effective partnerships. In E. Coakes (Ed.), *Knowledge management: Current issues and challenges* (ch. XV, pp. 187-211). Hershey, PA: IRM Press.

KEY TERMS

Accountability: Transparency of responsibility for performance, the management of performance and resulting implications for the deployment of future resources.

Change Readiness: An organizational mindset that welcomes challenges to established structures and processes and administrative orthodoxies.

Clarity of Responsibility: An understanding of the roles and responsibilities of individuals and business units working together to deliver a holistic service proposition.

Information Quality: The accuracy, completeness, timeliness and utility of performance related information that is used as the basis of management decision making.

Innovative Culture: An organizational climate that fosters new ideas and reflection on learning from experiences.

Knowledge Sharing: Formal, deliberate and systematic activities of transferring or disseminating knowledge from one person, group or organization to another.

Partnership Working: A network of organizations working together to provide a total service offering to a targeted group of users.

Public Services: Social infrastructure services, delivered wholly or partly with the benefit of public funds and strategically driven through national or regional administrations.

Strategic Connections: The relationships between the strategy formulation process and the deployment of resources to achieve it.

Improving Virtual Teams through Creativity

Teresa Torres-Coronas
Universitat Rovira i Virgili, Spain

Mila Gascó-Hernández
Open University of Catalonia, Spain

INTRODUCTION

Many studies have already shown how a team can become more creative, and therefore more efficient, but only a few researchers have focused on how a virtual team can use creativity techniques to perform better. In this article, we study what differences there are (both in terms of processes and in terms of results) when creativity techniques are used in the management of traditional and virtual teams. To do this, we discuss three main elements: the definition of creativity and its relationships with team performance, the variables that enhance creativity in a virtual team, and the most suitable creativity techniques for a virtual environment.

BACKGROUND

Most researchers and practitioners believe that the key to organizational success lies in developing intellectual capital and acquiring a new set of thinking: the creativity to produce an idea and the innovation to translate the idea into a novel result (Roffe, 1999). Explaining the meaning of creativity is not straightforward; there are thousands of definitions of the term. So, for the purpose of this article, we will understand creativity as the shortest way to search for unconventional wisdom and to produce paradigm-breaking ideas and innovation. This unconventional wisdom through the generation and use of creative knowledge is the key to building sustainable competitive advantages (Carr, 1994).

In order to develop more innovative products, services, or processes, organizations must encourage their employees to become more creative. During the last few decades, several researchers (Andriopoulous, 2001; Nemiro & Runco, 2001; McFadzean, 1998; Amabile, Conti, Coon, Lazenby & Herron, 1996) tried to describe contextual factors largely under the control of managers that influence creativity, though as creativity is a multidimensional concept, there is not a universal theory yet (Walton, 2003). This section focus on how managers and/or team leaders can improve creative climate within virtual structures.

The literature review conducted by Andriopoulous (2001) highlights five major organizational factors that enhance creativity in a traditional work environment: 1) organizational climate, or designing a working atmosphere that fosters participation and freedom of expression; 2) a democratic and participative leadership style; 3) an organizational culture that nourishes innovative ways of solving problems; 4) new resources and skills through the development of human resources creative talent; and 5) a structure and systems that include building flat structures, and rewards, recognition, and career systems that emphasize people creative thinking. Scholars argue that these factors create conditions that enhance creativity both at the team and individual levels.

From a study of the social psychology of creativity, Amabile (1996) cites the three main origins of creative performance as: task motivation, domain-relevant skills, and creativity-relevant skills. She differentiates between intrinsic and extrinsic motivation, proposing that the intrinsic motivation enhances creativity. In Amabile's research, the work team environment is also considered to exert a powerful impact on creativity by influencing the employee's intrinsic motivation. Management practices indicate that performance can be fostered by allowing freedom and autonomy to conduct one's work, matching individuals to work assignments, and building effective work teams that represent a diversity of skills and are made up of individuals who trust and communicate well with each other, challenge each other's ideas, are mutually supportive, and are committed to the work they are doing (Amabile & Gryskiewicz, 1987). Creativity is best achieved in open climates (Feurer, Chaharbaghi & Wargin, 1996).

These studies have not specifically addressed dimensions that may be necessary when groups no longer interact in traditional structures (Nemiro, 2001). In fact, so far, the only research that has been seriously conducted about this issue is that by Nemiro (2001), who identifies several key elements that influence creativity in virtual teams and therefore result in effectiveness and high levels of performance. Table 1 summarizes some of these factors as described by Nemiro (2001, p. 94).

A creativity-based management aimed at fostering virtual team creativity and performance must manage the above environmental variables in order to enhance employees' internal drive to perceive every project as a new creative challenge (Andriopoulos & Lowe, 2000).

A quick analysis of the variables shown in Table 1 gives rise to the conclusion that there are no meaningful differences between the factors that affect creativity in traditional environments and those that affect creativity in virtual contexts. On the other hand, most of the factors that influence creativity (such as work characteristics and situational constraints) are also considered as factors that impact team performance, as the conceptual model of Prasad and Akhilesh (2002) shows. Nevertheless, due to the particular way virtual teams work, there is a need to consider some elements related to the previous variables. Thus, communication and trust become very relevant issues.

In this sense, Henry and Hartzler (1998) find that keeping the synergy and creativity flowing, without frequent face-to-face interaction, is the greatest challenge a virtual team has. Virtual teams lose non-verbal communication and, as has been argued, electronic communication increases the level of social isolation. Schein (1993) points out that most communication workshops emphasize active listening, which means paying attention to the spoken words, the body language, the tone of voice, or the emotional content. Virtual teams that want to communicate successfully cannot actively listen in this sense. Other tools must therefore be explored¾for example, the use of multiple media or several communication technologies (Bal & Teo, 2001). However, as Van der Smagt (2000) showed, it is crucial to ensure that dialogue is the primary form of interaction between team members and that two-way monologues are avoided. Rich media—those that transmit nonverbal cues—are not the solution.

"In a dialogue, the difficult part is to make one's own assumptions manifest, not the exchange of insights with others. The attitude in relation to other actors is one of openness, which makes it relatively easy to get behind the position and possibilities of actors." (Van der Smagt, 2000, p. 155)

Collaborative work also requires a level of personal familiarity and trust. Without trust, building a true team is almost impossible (Duarte & Snyder, 1999). For most newly forming virtual teams, achieving an effective level of trust is not an easy task. Increasingly, virtual teams will form without the advantage of prior face-to-face team building opportunities, but with the added challenges of geographic isolation, time zone differentials, and cultural diversity (Holton, 2001). With virtual team heterogeneity there is a high probability that team members are confronted with mistrust (Prasad & Akhilesh, 2002), though such diversity within a team has the potential to increase opportunities to be innovative and creative (Lipnack & Stamps, 1997), if trust can be established (Dyer, 1995). But how can trust be built? The qualitative research project of Holton (2001) concludes that standard team-building tools

Table 1. A summary of factors that can foster creativity in a team context

Autonomy and Freedom. Allowing individuals responsibility for initiating new ideas and making decisions; a sense of control over one's work.

Challenge. Work that is stimulating, engaging, and meaningful; a sense of having to work hard on challenging and important tasks.

Clear Direction. Goals that facilitate creativity are clear, negotiated, attainable, shared, and valued.

Diversity/Flexibility/Tension. Diversity, both in terms of the work assignments offered and the people one interacts with, and a tolerance of differences. In order to be tolerant of differences, flexibility is needed. Both diversity and flexibility can lead to creative tension.

Support for Creativity. An organizational focus on support for or encouragement of creativity.

Trust and Participative Safety. Especially crucial for group creativity is trust and participative safety. The emphasis is on encouraging participation in a non-threatening, non-evaluative environment.

can be used to enhance collaboration and trust in a virtual team. The book of Simon Priest (2001) is full of examples for virtual team building. But, as with all team building, there is no quick fix for virtual teams.

These difficulties related to communication and trust are only an example that illustrates the need to conduct in-depth studies on the rest of environmental factors that, in non-virtual contexts, have been proven to directly impact teamwork creativity.

TOOLS AND TECHNIQUES TO IMPROVE CREATIVE PERFORMANCE

How can team creativity be encouraged? Until now no serious research has been conducted into which creativity techniques are the most suitable in a virtual environment. In traditional environments, one method of achieving this is to encourage teams to utilize creative problem-solving (CPS) techniques such as synectics, brainwriting, or wishful thinking.

In this context, McFadzean (2000, 1999, 1998) explores creative problem solving and presents a model that helps facilitators and team members choose appropriate techniques. McFadzean (1996a, 1996b) classifies creative problem solving (CPS) techniques into three categories: paradigm preserving, paradigm stretching, and paradigm breaking. Paradigm preserving techniques do not tend to change a participant's perspective. Paradigm stretching techniques encourage users to stretch the boundaries of the problem space. Paradigm breaking techniques allow participants to completely break down the boundaries of the problem space and to look at something entirely new.

In a virtual environment, three variables must also be considered when selecting a technique (Gascó-Hernández & Torres-Coronas, 2004): 1) the effectiveness of the method in finding innovative solutions, considering that quality solutions require the right balance between knowledge of the business issue and novelty (Kim, 1990); 2) the technological context or support system through which the technique can be implemented; and 3) the level of interaction that the technique requires. It is also important that the virtual facilitator has experience in running virtual creative sessions. Team members must also be taught about the dynamics of virtual interactions and about the use of technological tools, such as chats, e-mail, video conferencing, or interactive whiteboards.

Next, McFadzean's creativity continuum (1996a, 2000) will be used to summarize how virtual teams can choose among different techniques, which will be briefly described, to generate ideas. To make valuation accessible, the techniques will be classified according to the above three criteria. They will be rated as well (low, medium, and high). Finally, common technological tools by which each technique can be used will be shown.

It is a fact that there is a remaining need for in-depth research studies that help clarify which is the best creative problem-solving tool in terms of virtual team creative performance. Table 2 intends to be the basis to start evaluating the creativity continuum in a virtual context.

The use of these techniques will only be effective if the organization has a creative culture (McFadzean, 1998). Environmental factors that help managers to build a creative climate within virtual communities and creative problem-solving tools are two sides of the same coin.

FUTURE TRENDS

Virtual teams can use creativity in order to perform better. Nevertheless, there is a need to adapt those tools and techniques to a virtual environment. In this article, we have approached some of the issues that must be considered when studying the relationship between teams that work online and creativity. Nevertheless, several questions still remain unanswered and, in this sense, further research is required. In particular, three important issues need further development. First, the relationship between creativity and virtual team performance needs to be thoroughly explored. Successful studies will determine how structural and environmental factors influence team creativity. Second, although virtual teams are already using idea-generation techniques, their strengths and weaknesses need to be carefully and academically explored. Finally, it is also important to consider the effects of technology on both individual and team creativity. Technology has risks that can sometimes outweigh its benefits. When applying creativity techniques, people need to focus on the creative process, not on the technology being used. Technology must be easy to use, it must be effortless and unsophisticated—the simpler the technology, the better.

CONCLUSION

The emergence of tools based on the new information and communication technologies is currently affecting team creative processes. Nowadays, achieving high levels of creative performance is still an unresolved problem within virtual teams. Only a few researchers have focused on how a virtual team can use creativity techniques to perform better or how to build a creative virtual environment to foster creativity. The critical issues discussed in this article summarize many challenges to implement creative

Table 2. Valuing the creativity continuum in a virtual context

PRESERVING PARADIGM TECHNIQUES	STRETCHING PARADIGM TECHNIQUES	BREAKING PARADIGM TECHNIQUES
• Problem boundaries: unchanged • Creative stimulation: low • Stimuli: related • Expression: verbal/written • Can be used by experienced and inexperienced groups	• Problem boundaries: stretched • Creative stimulation: medium • Stimuli: unrelated • Expression: verbal/written	• Problem boundaries: broken • Creative stimulation: high • Stimuli: unrelated • Expression: unlimited • Should only be used by experienced groups
TECHNIQUES WITHIN EACH GROUP		
BRAINSTORMING	**OBJECT STIMULATION**	**WISHFUL THINKING**
Generation of ideas without criticism	Group members generate ideas using objects unrelated to the problem.	Participants are asked to look at a perfect future, examine each fantasy statement, and look for ideas on how these ideas can be achieved.
Effectiveness: Medium	Effectiveness: Medium	Effectiveness: Medium
Technological context: Online chat rooms Electronic mail Video conferencing	Technological context: Online chat rooms Electronic mail Video conferencing	Technological context: Online chat rooms Electronic mail Video conferencing
Level of interaction: High	Level of interaction: Medium	Level of interaction: Medium
BRAINWRITING	**METAPHORS**	**RICH PICTURES**
Group members write down their ideas on different sheets of paper. They are encouraged to build on others' ideas.	Group members use a metaphor to generate ideas to solve a problem.	Participants draw pictures which can be a metaphor of the problem. The descriptions of the picture help participants to generate ideas.
Effectiveness: High	Effectiveness: Medium	Effectiveness: Medium
Technological context: Online chat rooms Electronic mail Video conferencing	Technological context: Online chat rooms Electronic mail Video conferencing	Technological context: Whiteboard software
Level of interaction: Medium	Level of interaction: Medium	Level of interaction: Low

management within both virtual teams and organizations. With greater emphasis being placed on creative thinking and processes, team creative performance will increase day by day, allowing organizations to succeed and to become more innovative and adaptable.

REFERENCES

Amabile, T.M., Conti, R., Coon, H., Lazenby, J. & Herron, M. (1996). Assessing the work environment for creativity. *Academy of Management Journal, 39*(5), 1154-1184.

Andriopoulos, C. (2001). Determinants of organizational creativity: A literature review. *Management Decision, 39*(10), 834-840.

Andriopoulos, C. & Lowe, A. (2000). Enhancing organizational creativity: The process of perpetual challenging. *Management Decision, 38*(10), 834-840.

Bal, J. & Teo, P.K. (2001). Implementing virtual teamworking: Part 2—a literature review. *Logistics Information Management, 14*(3), 208-222.

Carr, C. (1994). *The competitive power of constant creativity*. New York: AMACOM.

Duarte, D.L. & Snyder, N.T. (1999). *Mastering virtual teams.* San Francisco, CA: Jossey-Bass.

Dyer, W.G. (1995). *Team building: Current issues and new alternatives* (3rd Edition). Reading, MA: Addison-Wesley.

Feurer, R., Chaharbaghi, K. & Wargin, J. (1996). Developing creative teams for operational excellence. *International Journal of Operations & Production Management, 16*(1), 5-18.

Gasco-Hernández, M. & Torres-Coronas, T. (2004). Virtual teams and their search for creativity. In F. Pixis & S. Godar (Eds.), *Virtual and collaborative teams: Process, technologies, and practices* (pp. 213-231). Hershey, PA: Idea Group Publishing.

Henry, J.E. & Hartzler, M. (1998). *Tools for virtual teams.* Milwaukee, WI: ASQC Quality Press.

Holton, J.A. (2001). Building trust and collaboration in a virtual team. *Team Performance Management: An International Journal, 7*(3/4), 36-47.

Kim, S.H. (1990). *Essence of creativity—a guide to tackling difficult problems.* New York: Oxford University Press.

Lipnack, J. &, Stamps, J. (1997). *Virtual teams: Reaching across space, time and organizations with technology.* New York: John Wiley & Sons.

McFadzean, E.S. (1998). Enhancing creative thinking within organizations. *Management Decision, 36*(5), 309-315.

McFadzean, E.S. (1996a). *The classification of creative problem-solving techniques.* Working Paper No. 9632, Henley Management College, Henley-on-Thames, Oxon, UK.

McFadzean, E.S. (1996b). *New ways of thinking: An evaluation of K-Groupware and creative problem solving.* Doctoral Dissertation, Henley Management College/ Brunel University, Henley-on-Thames, Oxon, UK.

McFadzean, E.S. (1999). Encouraging creative thinking. *Leadership & Organization Development Journal, 20*(7), 374-383.

McFadzean, E.S. (2000). Techniques to enhance creative thinking. *Team Performance Management: An International Journal, 6*(3/4), 62-72.

Nemiro, J.E. (2001). Connection in creative virtual teams. *The Journal of Behavioral and Applied Management, 2*(2), 92-112.

Prasad, K. & Akhilesh, G.B. (2002). Global virtual teams: What impacts their design and performance? *Team Performance Management: An International Journal, 8*(5/6), 102-112.

Priest, S. (2001). *100 of the best virtual team-building events.* Tarrack Publication.

Roffe, I. (1999). Innovation and creativity in organizations: A review of the implications for training and development. *Journal of European Industrial Training, 23*(4/5), 224-237.

Schein, E.H. (1993). On dialogue, culture and organizational learning. *Organizational Dynamics, 22*(2), 40-51.

Van der Smagt, T. (2000). Enhancing virtual teams: Social relations v. communication technology. *Industrial Management & Data Systems, 100*(4), 148-156.

Walton, A.P. (2003). The impact of interpersonal factors on creativity. *International Journal of Entrepreneurial Behaviour & Research, 9*(4), 146-162.

Williams, S. (2001). Increasing employees' creativity by training their managers. *Industrial and Commercial Training, 33*(2), 63-68.

KEY TERMS

Autonomy and Freedom: Allowing individuals responsibility for initiating new ideas and making decisions; a sense of control over one's work.

Challenge: Work that is stimulating, engaging, and meaningful; a sense of having to work hard on challenging and important tasks.

Clear Direction: Goals that facilitate creativity are clear, negotiated, attainable, shared, and valued.

Converging Thinking Techniques: Tools used during the convergent phases of the CPS to improve the evaluation and selection of the most relevant ideas, thoughts, or data. Pluses, potentials, and concerns (PPC); highlighting; and the evaluation matrix are some of the most common converging thinking techniques.

Creative Performance: High level of capability in an idea or solution, applied to solve a problem in an imaginative way, resulting in effective action. Environmental factors such as autonomy and freedom, challenge, clear direction, diversity/flexibility/tension, support for creativity, trust, and participative safety directly affect the creative performance within work teams.

Creative Problem Solving (CPS): A systematic process model to solve problems and to harness creativity. Its six steps include objective-finding, data-finding, problem-finding, idea-finding, solution-finding, and acceptance-finding. Each step has a divergent and convergent phase. During the divergent phase, a free flow of ideas is elicited. Convergent phases involve the evaluation and selection of the ideas with the greatest potential or relevancy. The defer-judgment rule separates idea generation from idea evaluation.

Creativity: The production of something new or original that is useful; the act of creating recombining ideas or seeing new relationships among them. Creativity is usually defined in terms of either a process or a product and at times has also been defined in terms of a kind of personality or environmental press. These are four Ps of creativity: process, product, person, and press.

Divergent Thinking Techniques: Tools used during the divergent phases of the CPS to improve the generation of ideas, thoughts, or data without evaluation. These tools are classified according to their primary use of related or unrelated problem stimuli. Brainstorming, brainwriting, forced connections, analogies, and metaphors are some of the most used divergent thinking techniques.

Diversity/Flexibility/Tension: Diversity, both in terms of the work assignments offered and the people one interacts with, and a tolerance of differences. In order to be tolerant of differences, flexibility is needed. Both diversity and flexibility can lead to creative tension.

Support for Creativity: An organizational focus on support for or encouragement of creativity.

Trust and Participative Safety: Especially crucial for group creativity is trust and participative safety. The emphasis is on encouraging participation in a non-threatening, non-evaluative environment.

Virtual Team: A group of people who are geographically separated and who work across boundaries of space and time by utilizing computer-driven technologies such as desktop video conferencing, collaborative software, and Internet/intranet systems. How these teams interact defines them as "virtual."

Inclusion Dependencies

Laura C. Rivero
Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina and Universidad Nacional de La Plata, Buenos Aires, Argentina

INTRODUCTION

Inclusion dependencies support an essential semantics of the standard relational data model. An inclusion dependency is defined as the existence of attributes (the left term) in a table R whose values must be a subset of the values of the corresponding attributes (the right term) in another table S (Abiteboul, Hull & Vianu, 1995; Codd, 1990; Connolly, Begg & Strachan, 1999; Date, 2000). When the right term conforms a unique column or a primary key (K) for the table S, the inclusion dependency is key-based (also named referential integrity restriction, *rir*). In this case, the left term is a foreign key (FK) in R and the restriction is stated as R[FK] <<S[K]. On the contrary, if the right term does not constitute the key of S, the inclusion dependency is non-key-based (simply, an inclusion dependency, *id*). Ids are expressed as R[X]$\subseteq$S[Z], being R[X] and S[Z] the left and right terms respectively. Both, *rir*s and *id*s, are often called referential constraints.

*Rir*s and referential actions are important because they contain basic local semantic aspects, which have been elicited from the Universe of Discourse (UofD). They are sufficient to symbolize many natural semantic links such as the relationships and hierarchies that are captured by semantic models (Abiteboul, Hull & Vianu, 1995). In a different scenario, late changes of the logical design, disregarding the conceptual design, usually promotes some denormalization degree or the presence of complex n-ary relationship constructs. The decomposition and the synthesis coming from that process may add other restrictions, frequently adopting the form of *id*s that misrepresent objects and the corresponding inter-object relationships.

*Rir*s can be declaratively defined via the SQL foreign key clause (SQL:1999-1, 1999, SQL:1999-2, 1999) and are enforced by most current database systems:

FOREIGN KEY (<referencing column list>) REFERENCES <referenced table name> [(<referenced column list>)]
[MATCH <match type>]
[ON UPDATE <update referential action>]
[ON DELETE <delete referential action>]

The *rir*s can be specified with respect to different match types: SIMPLE (implicit if no match option is declared), PARTIAL and FULL. As it has been stated in the SQL:1999 standard document (SQL:1999-2, 1999): If <match type>(SIMPLE) is not specified, then for each row in the referencing table, either the referencing column has at least one null value or its value matches the value of a corresponding row in the referenced table. If PARTIAL is specified, then for each row in the referencing table the value of each foreign key column is null, or it has at least one non-null value that equals the corresponding referenced column value. Finally, if FULL is specified, for each row in the referencing table, either all foreign key columns have been instanced with null or the foreign key value equals the value of the corresponding referenced column.

When an integrity restriction is violated, the usual response of the system is the rollback of the data manipulation intended by the user. In the case of *rir*s, some other alternative actions are possible. These actions, named referential actions or referential rules, specify the behavior of the left and right relations under the deletion or the updating of a referenced row (in the right table), or the insertion in the referencing (left) table. Possible actions are: cascade, restrict, no action, set null, set default (Date, 2000; Markowitz, 1994; SQL:1999_2, 1999; Türker & Gertz, 2001).

With the cascade option, the referencing rows will be deleted (updated) together with the referenced row. With the set null (set default) option, all references to the deleted (updated) row will be set to null (default) values. Restrict and no action rules disallow the deletion (update) of the referenced row, if there exists rows in the left table referencing it. The unique referential rule for insertions is restrict: inserting a row into the referencing table is possible only if the referenced tuple already exists in the right term.

On the other hand, *id*s are usually defined with check statements

CHECK (<referencing column list> IN (SELECT <referenced column list> FROM <referenced table>))

or triggers, thus complicating the development of application programs and integrity maintenance (Connolly, Begg & Strachan, 1999; Date, 2000; Date & Darwen, 1997; Elmasri & Navathe, 2000).

The objective of this entry is to provide an overview of *id*s, summarizing main research on this topic.

BACKGROUND

The comprehension of the syntactic and semantic issues related to referential restrictions is facilitated by the study of the structure of their terms.

Structure

Considering a relation shape, there are five possible placements of a non-empty set of attributes with regard to the key placement. Being W such set of attributes, and K the primary key of R, the five placements are depicted in Figure 1: I) W≡ K (W coincides with K); II) W≡ Z, being Z a subset of non-key attributes (W and K have not common attributes); III) W≡ K_1, being K_1 a proper subset of K, $K_1 \neq \emptyset$ (the set of attributes in W is a proper subset of the set of attributes in K); IV) W≡ K ∪ Z (W attributes include the complete set of attributes of K); and finally V) W≡ $K_1 \cup$ Z, $K_1 \neq \emptyset$ (W and K partially overlap). In all cases, Z $\neq \emptyset$.

Let R and S be the left and right terms of a referential constraint, respectively. Taking into account the placements in Figure 1, 25 pairs <R[W_R], S[W_S]> corresponding to left and right terms of a referential constraint can be derived. The five cases having S[W_S] as the primary key for S (numbered 1 to 5 in Table 1) correspond to *rir*s.

Semantic Perspective

*Rir*s of types I, II and III represent typical relationships in semantic models (Abiteboul, Hull & Vianu, 1995). Type I depicts subtype relationships such as "every salesman is also a person". Type II corresponds to designative relationships such as 1:1, N:1 or n-ary relationships with at least one 1 cardinality. Type III appears in associative relationships such as N:N and n-ary relationships and weak entities (Date, 2000; Elmasri & Navathe, 2000). Types IV and V deserve a different analysis, as they appear as a result of late irregular alterations of logical designs (Rivero, Doorn & Ferraggine, 2004). The remaining cases correspond to *id*s, which cannot be obtained via a semantic model, but as a consequence of similar reasons as types IV and V of *rir*s (Rivero, Doorn & Ferraggine, 2001, 2004).

MAIN CHARACTERISTICS AND APPLICATIONS OF INCLUSION DEPENDENCIES

Following, there is a description of main issues about inclusion dependencies and relevant applications of this concept.

Referential Actions and Global Semantics

Update operations promote the execution of specialized triggers -the referential actions-, for the programmed maintaining of referential integrity. The actions are: *cascade, restrict, no action, set null, set default* (Date, 2000; Markowitz, 1994; SQL:1999_2, 1999; Türker & Gertz, 2001).

Despite the fact that the local effect of such rules is precisely defined, when update operations are executed on the database state, the global effects those interacting actions promote may show ambiguities (Lüdascher & May, 1998; Markowitz, 1994; May & Lüdascher, 2002; Reinert, 1996).This problem has been - and currently is - a matter of profuse research, from the beginning of the relational databases era. While in Markowitz(1994) the

Figure 1. Placements of a set of attributes in correlation with the key

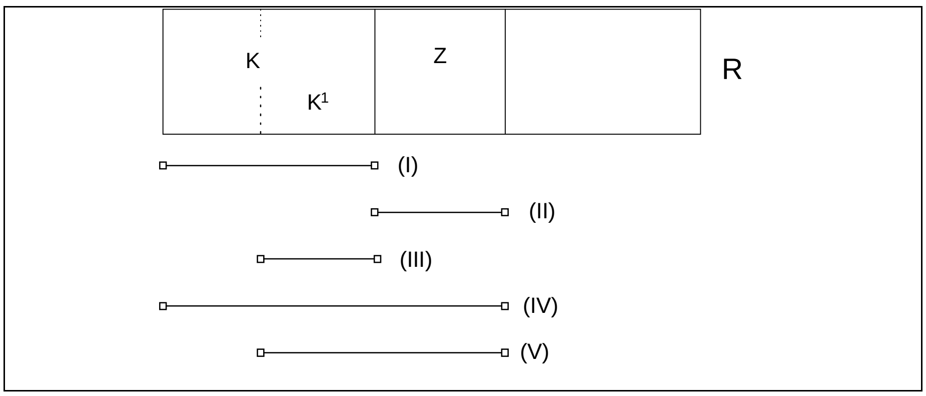


Table 1. Possible structures for referential constraints

W_r \ W_l	**I) Key (K_r)**	**II) Non Key (Z_r)**	**III) Part of a Key (K_r^1)**	**IV) Key + Non Key ($K_r \cup Z_r$)**	**V) Part of a Key + Non Key ($K_r^1 \cup Z_r$)**
I) Key (K_l)	**1.** $K_l << K_r$	**6.** $K_l \subseteq Z_r$	**11.** $K_l \subseteq K_r^1$	**16.** $K_l \subseteq K_r \cup Z_r$	**21.** $K_l \subseteq K_r^1 \cup Z_r$
II) Non Key (Z_l)	**2.** $Z_l << K_r$	**7.** $Z_l \subseteq Z_r$	**12.** $Z_l \subseteq K_r^1$	**17.** $Z_l \subseteq K_r \cup Z_r$	**22.** $Z_l \subseteq K_r^1 \cup Z_r$
III) Part of a Key (K_l^1)	**3.** $K_l^1 << K_r$	**8.** $K_l^1 \subseteq Z_r$	**13.** $K_l^1 \subseteq K_r^1$	**18.** $K_l^1 \subseteq K_r \cup Z_r$	**23.** $K_l^1 \subseteq K_r^1 \cup Z_r$
IV) Key + Non Key ($K_l \cup Z_l$)	**4.** $K_l \cup Z_l << K_r$	**9.** $K_l \cup Z_l \subseteq Z_r$	**14.** $K_l \cup Z_l \subseteq K_r^1$	**19.** $K_l \cup Z_l \subseteq K_r \cup Z_r$	**14.** $K_l \cup Z_l \subseteq K_r^1 \cup Z_r$
V) Part of a Key + Non Key ($K_l^1 \cup Z_l$)	**5.** $K_l^1 \cup Z_l << K_r$	**10.** $K_l^1 \cup Z_l \subseteq Z_r$	**15.** $K_l^1 \cup Z_l \subseteq K_r^1$	**20.** $K_l^1 \cup Z_l \subseteq K_r \cup Z_r$	**25.** $K_l^1 \cup Z_l \subseteq K_r^1 \cup Z_r$

References: K_*(key); Z_* (non-key attributes); K_*^1 (a proper subset of K_*). $_*$ = l, r (left, right)

anomalies caused by the use of referential actions (in some cases interacting with null constraints) are completely described and some solutions for its treatment are presented, in Reinert (1996) this study is revised by showing that the problem of deciding if a relational schema may or may not have an instance leading to ambiguities is undecidable. Recently, the whole understanding of the global behavior of referential actions based on the local characterization of that rules has been formalized in Ludäscher and May (1998) and May and Ludäscher (2002).

Inclusion Dependencies and Functional Dependencies: Interaction, Inference Rules and the Implication Problem

Levene and Loizou (1999) affirm that "the interaction between functional and inclusion dependencies is a complex problem, and there is not a complete and sound system of axioms for functional and inclusion dependencies at all". While some interactions may result in a new (functional or inclusion) dependency being obtained, other cases derive in redundant attributes. Pullback, Collection and Attribute-Introduction rules capture significant interactions between such restrictions (Levene & Loizou, 1999; Levene & Vincent, 2000).In Levene and Loizou (1999) may be found the relational database theory needed to understand the inference rules and the implication problem for *id*s.

Use of Inclusion Dependencies as Domain Constraints

Some relationships have the semantics of *id*s or *rir*s symbolizing, essentially, domain restrictions that indicate the legal values for an attribute. UofD business rules associated to specific domain restrictions over dynamic and voluminous set of values are frequently written as *id*s or *rir*s. For example,

CHECK (LeftAttrList IN (SELECT RightAttrList FROM R))

would indicate that the set of allowable values for LeftAttrList is conformed by the current set of instances of RightAttrList in the relation R (Rivero, Doorn & Ferraggine, 2001, 2004).

Basis for the Database Reengineering

This application is strongly related to the previous one. A deep analysis of the referential constraints helps find hidden objects and misrepresented relationships. While the structure of the left and right terms of *id*s is a key concept for the reengineering of relational database schemas, cases IV and V of *rir*s reveal relationships hiding specific business rules. A careful examination of the structure of these constraints allows the reconstitution of a subject schema into an enhanced form and the subsequent enrichment of the whole system (Rivero, Doorn & Ferraggine, 2001, 2004; Tari et al., 1997).

Referential Integrity in Object-Oriented and Object-Relational Databases

Object-oriented literature typically uses the term "relationship" to mean relationships supported by foreign keys in a relational system. This implementation, the unique available in a SQL-92 system, may be used in an object-relational DBMS (ORDBMS) (Date, 2000; Stonebraker, 1996). Object-relational systems, in compliance with the current standard – SQL:1999 - provide the *reference* as a natural surrogate for primary key-foreign key relationships. In these systems, a column in a table may contain a value that is a *reference* to an instance of a type stored in another table. This implementation is supported by the unique object identifier (OID) of rows (SQL:1999-1, SQL:1999-2).

Triggers for referential integrity are an alternative to the use of foreign key constraints, despite the fact that foreign key constraints are preferable to explicit triggers since they are declarative and then easier to manage.

FUTURE TRENDS

From its beginnings the relational technology has experimented a fantastic process of evolution to finally become a mature environment. Main commercial (and open source) software offer referential constraints and triggers to fully maintain referential integrity. On the other hand, while object-oriented (OO) programming and semantic models are now the standard for best developers and programmers, OO databases constitute just a small portion of the market, albeit they have gained popularity in recent years. One of the reasons why the OO paradigm has been adopted to extend the relational technology is because it allows developers to cope with the growing capacity of current software development. As a consequence, the OO paradigm has been the basis for current research on models that integrate the relational data model and SQL query languages with features coming from the object-oriented world. As a result, a next generation of database systems - object-relational (ORDBMS) - has emerged (Stonebraker, 1996).

This evolution has posed new challenges related to the referential integrity maintenance. Referential actions, defined in the relational context, must be efficiently implemented in the object-relational context. Different strategies suggest implementing integrity rules following the principles of the relational side of ORDBMS. Another, coming from the OO technology, recommends the implementation following the business logic, that is, within the objects. Triggers, specific metadata, and so forth are other ways to implement these concepts. The decision about the choice of the best strategy taking into account the structure of relations, their variability and other performance issues is a novel research field.

The degree of distribution and heterogeneity of the DBMS and the variety of data that current DBMSs can manage are two of the current directions of developments and advances in DBMS technology. With respect to the first issue, the definition and implementation of a consistency checking service is particularly difficult whenever data are stored in a heterogeneous collection that is not controlled by a single DBMS. On the other hand, multimedia databases (Dunckley, 2002), spatio/temporal databases, XML databases (Graves, 2001; Williams et al., 2000) and active databases (Paton & Diaz, 1999; Widom & Ceri, 1996; Zaniolo et al., 1997) are recent examples that illustrate the variety of data information that a DBMS should be able to manage. Efficient implementation of integrity issues, and mainly referential integrity constraints to deal with heterogeneous data sources in centralized or distributed environments, represent a challenge and a promising research area.

The World Wide Web, lacking specific guidelines to enforce referential integrity (Aldana, Yagüé & Gómez, 2002); newer multilevel security strategies posing a different scenario to the integrity maintenance (Lee, Kim & Kim, 2004); and the recent development of applications and DBMSs for mobile databases (Pitoura & Samaras, 1998) have given rise to interesting challenges and research opportunities in the area of database management and particularly in the area of data consistency.

CONCLUSION

This work presents an overview of inclusion dependencies and the particular case of referential integrity restrictions. It sketches some topics related to its structure and semantics and describe relevant issues related to inclusion dependencies interacting with functional dependencies. The version of this concept in third generation databases is briefly described. Atypical behaviors generated as a consequence of design flaws should encourage users to consider referential constraints not only as merely integrity constraints but also as the basis for a good practice on database design and reengineering.

Finally, current and future directions of database evolution are succinctly mentioned to describe the scenarios in which the integrity issues must mature.

REFERENCES

Abiteboul, S., Hull, H., & Vianu, V. (1995). *Foundations on databases.* Addison Wesley.

Aldana, J., Yagüé, M., & Gómez, A. (2002). Integrity issues in the Web: Beyond distributed databases. In J. Doorn & L. Rivero (Eds.), *Database integrity: Challenges and solutions.* Hershey, PA: Idea Group Publishing.

Codd, E. (1990). *The relational model for database management. Version 2.* Addison Wesley.

Connolly, T., Begg, C., & Strachan, A. (1999). *Database systems: A practical approach to design, implementation and management* (2nd ed.). Addison Wesley.

Date, C. (2000). *An introduction to database systems.* Addison Wesley.

Date, C., & Darwen, H. (1997). *The SQL standard* (4th ed.). Addison Wesley.

Dunckley, L. (2002). *Multimedia databases: An object relational approach.* Pearson Education.

Elmasri, R., & Navathe, S. (2000). *Fundamentals of database systems.* Addison Wesley.

Graves, M. (2001). *Designing XML databases.* Prentice Hall.

Lee, S.-W., Kim, Y.-H., & Kim, H.-Y. (2004). The semantics of an extended referential integrity for a multilevel secure relational data model. *Data & Knowledge Engineering, 48*, 129-152.

Levene, M., & Loizou, G. (1999). *A guided tour of relational databases and beyond.* London: Springer-Verlag.

Levene, M., & Vincent, W.M. (2000). Justification for inclusion dependency normal form. *IEEE Transactions on Knowledge and Data Engineering, 12*(2), 281-291.

Ludäscher, B., & May, W. (1998). Referential actions: From logical semantics to implementation. *Proceedings of 6th International Conference on Extending Database Technology (EDBT´98)*, Valencia, Spain (pp. 404-418).

Markowitz, V. (1994). Safe referential integrity and null constraint structures in relational databases. *Information Systems, 19*(4), 359-378.

May, W., & Ludäscher, B. (2002). Understanding the global semantics of referential actions using logic rules. ACM Transactions on Database Systems (TODS), *27*(4), 343-397.

Paton, N.W., & Diaz, O. (1999). Active database systems. *ACM Computing Surveys, 31*(1), 63-103.

Pitoura, E., & Samaras, G. (1998). *Data management for mobile computing.* Kluwer.

Reinert, J. (1996). Ambiguity for referential integrity is undecidable. In G. Kuper & M. Wallace (Eds.), *Constraint databases and applications* (pp. 132-147). Springer.

Rivero, L., Doorn, J., & Ferraggine, V. (2001). Inclusion dependencies. In S.A. Becker (Ed.), *Developing quality complex database systems: Practices, techniques and technologies* (pp. 261-278). Hershey, PA: Idea Group Publishing.

Rivero, L., Doorn, J., & Ferraggine, V. (2004). Enhancing relational schemas through the analysis of inclusion dependencies. *International Journal of Computer Research, 12*(4). Nova Publishers. *To appear*.

SQL99-1. (1999). Database language SQL. *Part 1: SQL Framework Document ISO/IEC 9075-1.*

SQL99-2. (1999). Database language SQL. *Part 2: SQL Foundation Document ISO/IEC 9075-2.*

Stonebraker, M. (1996). *Object-relational DBMSs. The next great wave.* Morgan Kauffman.

Tari, Z. Buhkres, O., Stokes, J., & Hammoudi, S. (1998). The reengineering of relational databases based on key and data correlations. In S. Scappapietra & F. Maryanski (Eds.), *Searching for semantics: Datamining, reverse engineering, etc*. Chapman & Hall.

Türker, C., & Gertz, M. (2001). Semantic integrity support in SQL-99 and commercial (object-)relational database management systems. Retrieved on July 28, 2004 from *http://www.db.cs.ucdavis.edu/papers/TG00.pdf*

Widom, J., & Ceri, S. (1996). *Active database systems: Triggers and rules for advanced database processing.* Morgan Kauffman Publishers.

Williams, K., Brundage, M. et al. (2000). *Professional XML databases.* Wrox Press Inc.

Zaniolo, C., Ceri, S. et al. (1997). *Advanced database systems.* Morgan Kauffman Publishers, Inc.

KEY TERMS

Business Rules: Statements that model the reaction to events that occur in the real world, having tangible side effects on the database content. They respond to application needs.

Database Schema Reengineering: The process of analyzing a subject database schema to recover its components and their relationships. It guides the reconstitution of such system into an enhanced one, with a higher level of abstraction and semantically closer to the Universe of Discourse.

Global Semantics of Referential Actions: The whole effect interacting referential actions lead to under update operations on the database state.

Inclusion Dependency: Defined as the existence of attributes in a table whose values must be a subset of the values of the corresponding attributes in another (or the same). Expressed as $R[X] \subseteq S[Z]$. R and S are relation names (possibly the same); R[X] and S[Z] are named the inclusion dependency's left and right sides respectively. X, Z are compatible attributes.

Inconsistency: Database state that does not reflect a real-world state.

Integrity Constraints: Statements that specify the set of valid values in a database. They must be satisfied to protect the database against inconsistencies.

Referential Action: Specialized rule used to automatically maintain referential integrity. Define specific reactions to compensate referential integrity violations. They may be Cascade, Restrict, No Action, Set Default or Set Null.

Referential Integrity Restriction: A special case of an inclusion dependency, when Z is the primary key K of S. In this case, X constitutes a foreign key FK for R. This restriction can be stated as $R[FK] << S[K]$.

I

Incorporating Data Stream Analysis into Decision Support Systems

Damianos Chatziantoniou
Athens University of Economics and Business, Greece

George Doukidis
Athens University of Economics and Business, Greece

INTRODUCTION

Traditional decision support systems (DSS) and executive information systems (EIS) gather and present information from several sources for business purposes. It is an information technology to help the knowledge worker (executive, manager, analyst) make faster and better decisions. So far, this data was stored statically and persistently in a database, typically in a data warehouse. Data warehouses collect masses of operational data, allowing analysts to extract information by issuing decision support queries on the otherwise discarded data. In a typical scenario, an organization stores a detailed record of its operations in a database, which is then analyzed to improve efficiency, detect sales opportunities, and so on. Performing complex analysis on this data is an essential component of these organizations' businesses. Chaudhuri and Dayal (1997), present an excellent survey on decision-making and on-line analytical processing (OLAP) technologies for traditional database systems.

In many applications, however, it may not be possible to process queries within a database management system (DBMS). These applications involve data items that arrive on-line from multiple sources in a continuous, rapid and time-varying fashion (Babcock, Babu, Datar, Motwani & Widom, 2002). This data may or may not be stored in a database. As a result, a new class of data-intensive applications has recently attracted a lot of attention: applications in which the data is modeled not as persistent relations, but rather as transient *data streams*. Examples include financial applications (streams of transactions or ticks), network monitoring (stream of packets), security, telecommunication data management (stream of calls or call packets), web applications (clickstreams), manufacturing, sensor networks (measurements), and others. In data streams we usually have "continuous" queries (Terry, Goldberg, Nichols & Oki, 1992; Babu & Widom, 2001) rather than "one-time". The answer to a continuous query is produced over time, reflecting the stream data seen so far. Answers may be stored and updated as new data arrives or may be produced as data streams themselves. Continuous queries can be used for monitoring, alerting, security, personalization, etc. Data streams can be either *transactional*, i.e. log interactions between entities, such as credit card purchases; web clickstreams; phone calls; or *measurement*, i.e. monitor evolution of entity states, such as physical phenomena, road traffic, temperature, network.

How to best model, express and evaluate complex queries over data streams is an open and difficult problem. This involves data modeling, rich querying capabilities to support real-time decision support and mining, and novel evaluation and optimization processing techniques. In addition, the kind of decision support over data streams is quite different from "traditional" decision-making: decisions are "tactical" rather "strategic". Research on data streams is currently among the most active areas in database research community. We believe that flexible and efficient stream querying will be a crucial component of any future data management and decision support system.

BACKGROUND

The database research community has responded with an abundance of ideas, prototypes and architectures to address the new issues involved in data stream management systems (DSMS). STREAM is Stanford University's approach for a general-purpose DSMS (Arasu et al., 2003); Telegraph and TelegraphCQ (Madden & Franklin, 2002; Chandrasekaran et al., 2003) are prototypes focused on handling measurements of sensor networks, developed in Berkeley; Aurora is a joint project between Brandeis University, Brown University and MIT (Carney et al., 2002) targeted towards stream monitoring applications; AT&T's Hancock (Cortes, Fisher, Pregibon, Rogers & Smith, 2000) and Gigascope (Cranor, Johnson, Spatscheck & Shkapenyuk, 2003) projects are special-purpose data stream systems for network management; Tribeca (Sullivan, 1996) and NiagaraCQ (Chen, DeWitt, Tian & Wang , 2000) are other well-known projects from Telcordia and University of Wisconsin respectively. The objective of all these projects is to develop systems that can support the challenging analysis requirements of streaming applications.

Furthermore, a plethora of articles, papers and tutorials appeared recently in the research literature. Some of the most well known survey articles follow. Faloutsos (2004), discusses indexing and mining techniques over data streams; Koudas and Srivastava (2003), present the state-of-the-art algorithms on stream query processing; Muthukrishnan (2003), reviews data stream algorithms and applications; Babcock et al. (2002), present an excellent survey on data stream prototypes and issues; Garofalakis, Gehrke and Rastogi (2002), discuss various existing models and mining techniques over data streams.

APPLICATIONS

Stream applications span a wide range of everyday life. Real-time analytics is an essential part of these applications and becomes rapidly more and more critical in decision-making. It is apparent from the list of areas below that efficiently querying and processing data streams is a necessary element of modern decision support systems.

- **Telecommunications.** The telecommunications sector is undoubtedly one of the prime beneficiaries of such data management systems due to the huge amount of data streams that govern voice and data communications over the network infrastructure. Examples of stream analysis include fraud detection, real-time billing, dynamic pricing, network management, traffic monitoring and so on. Streams (calls, packets) have to be mined at real-time to discover outliers and patterns (fraud detection); correlated, joined and aggregated to express complex business rules (billing, dynamic pricing); and monitored – computing averages and min./max. values over periods of time - to uncover unusual traffic patterns (network management).
- **Sensors.** Sensor technology becomes extremely wide-spread and it will probably be the next killer app: large number of cheap, wireless sensors attached to products, cars, computers, even sport players and animals, tracking and digitizing behavior, traffic, location and motion. Examples involve electronic property stickers (super markets, libraries, shopping carts, etc.), vehicle sensors (to pay electronically tolls, route traffic, set speed, etc.) and location-identification sensors (to report location, serve content, detect routes, etc.) A "sensor" world leads to a "stream" world. Millions of input data every few seconds need to be analyzed: aggregate (what is the average traffic speed), correlate (two products sell together), alert (quantity of a product is below a threshold), localize and monitor.
- **Finance.** Financial data streams come in many different forms: stock tickers, news feeds, trades, etc. Financial companies want to analyze these streams at real-time and take "tactical" business decisions (opposed to "strategic" decisions, associated to OLAP or data mining). For example, Charles Schwab wants to compute commission on each trade at real-time; Fidelity would like to route content in trades at real-time. Traderbot (www.traderbot.com) is a web-based financial search engine that evaluates queries (both traditional and continuous) over real-time streaming data (e.g. "find all stocks between •20 and •200 where the spread between the high tick and the low tick over the past 30 minutes is greater than 3% of the last price and in the last 5 minutes the average volume has surged by more than 300%.")
- **Web Management.** Large web sites monitor web logs (clickstreams) online to enable applications such as personalization, performance monitoring, and load balancing. Some web sites served by widely distributed web servers (e.g. Yahoo) may need to coordinate many distributed clickstream analyses, e.g. to track heavily accessed web pages (e.g. CNN, BBC) as part of their real-time performance monitoring. (Babcock et al., 2002}
- **Network Management.** Network traffic management systems monitor a variety of continuous data streams at real-time, such as packet traces, packet flows and performance measurements in order to compute statistics, detect anomalies, adjust routing, etc. The volume of data streams can be humongous and thus, query processing must be done very carefully.
- **Military.** One of the most interesting applications in military is battalion monitoring – where sensors are installed on every vehicle, human, etc. – having thousands of sensors reporting state in real-time. In these applications we want to know each time where vehicles and troops are. Examples include queries such as "tell me when 3 of my 4 tanks have crossed the front line" and "tell me when someone is pointing a gun at me".

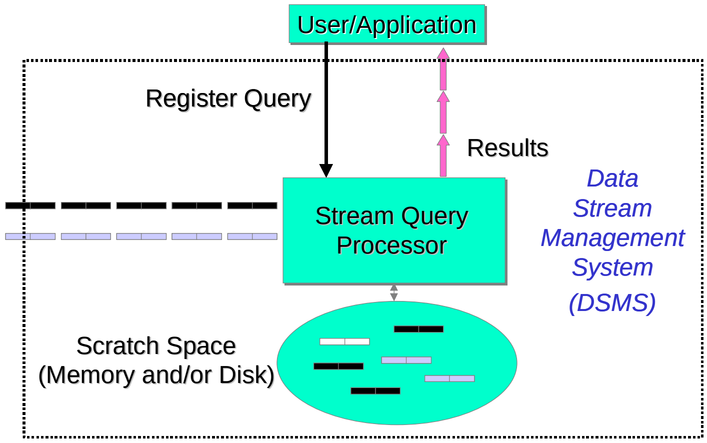
ISSUES AND CHALLENGES

Performing decision-making queries on top of data streams is a major challenge. For example, only one-pass algorithms are allowed (because data can be seen only once) and memory has to be managed very carefully (what to keep and what to discard). The need for a data stream

Table 1. Differences betweens DSMS and DBMS

DBMS	DSMS
Persistent relations	Transient streams
One-time queries	Continuous queries
Random access	Sequential access
“Unbounded” disk store	Bounded main memory
Only current state matters	History/arrival-order is critical
Passive repository	Active stores
Relatively low update rate	Possibly multi-GB arrival rate
No real-time services	Real-time requirements
Assume precise data	Data stale/imprecise

Figure 1. A generalized DSMS architecture



management system comes in two forms: either the volume of data is huge and can not be stored in persistent relations (e.g. packet network traffic) – but still some data analysis has to be carried out, or an answer is required for a report at real-time (e.g. monitoring, alerting, fraud-detection.) As a result, DSMS are quite different in nature from traditional DBMS. Data is transient instead of persistent; queries may be “continuous” instead of “one-time”; processing techniques differ significantly, primarily due to main-memory management requirements. In Table 1 we list the differences between traditional database and data stream management systems (Babcock et al. 2002).

A generalized DSMS architecture is shown in Figure 1 (Babcock et. al.).

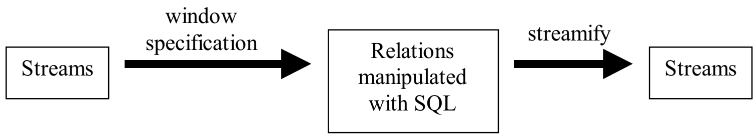
A user or application registers a (continuous) query with the stream system. As new stream data arrives, the answer of the registered query is updated. The stream query processor deals with the incoming data and has to decide on a significant number of important issues: should it shed data to keep up with the rate of the stream (leading to approximate answers)? What is the best way utilize main-memory space? What is the best way to handle multiple registered queries in an optimal way (maximizing resource utilization)? Finally, some scratch space may be required for intermediate computations.

The differences between DSMS and DBMS present several novel challenges for a system handling and analyzing data streams, both in terms of functionality and processing.

In terms of functionality (how to *use* data streams, i.e. reporting):

- Defining complex stream sessions (e.g. a network “flow” or a finance “burst”.)
- Representing nested stream structures (e.g. a sub-session within a session.)
- Defining hierarchical statistical stream functions (similar to roll-up, drill-down in OLAP).
- Allowing multiple sources and/or multiple formats, possibly combine with persistent data.

Figure 2. Turning streams to relations



- Allowing user-defined functions – UDFs (e.g. a non-traditional financial computation).

In terms of evaluation (how to *process* data streams):

- Handling multiple, continuous, rapid, time-varying, ordered streams.
- Having computations only in main-memory.
- Queries may be continuous (not just one-time):
  • Have to be evaluated continuously as stream data arrives.
  • Answer updated over time.
- Queries may be complex:
  • Beyond element-at-a-time processing.
  • Beyond stream-at-a-time processing.
  • Beyond traditional decision making queries (scientific, data mining).
- Queries may involve nested structures.
- Users may select for query answers to be:
  • Exact (worst performance, sometimes not possible due to the stream rate).
  • Approximate (requires load shedding, i.e. discarding data, better performance).

QUERYING AND EVALUATING DATA STREAMS

The challenges mentioned in the previous section lead to a number of major issues in querying and processing data streams. Understanding these aspects is crucial in developing a stream system. Different systems follow different methodologies in processing data streams (as presented in the "Systems Prototypes" section below).

Windows

Being able to formulate queries on top of data streams is a major component of any stream system. Most system prototypes extend SQL in some way in order to make it suitable for stream processing. For example, AT&T's Gigascope defines GSQL and Stanford's STREAM proposes CQL. Most stream query languages reference and produce both relations and streams. Please see the figure below.

It is important however to transform infinite, unbounded streams to finite relations which then can be manipulated by a query processor. Defining windows of streams through some window specification syntax is the mechanism to turn (part of) a stream to a relation. For example, one could define a window as the one-thousand most recent stream tuples. Figure 2 graphically shows this concept.

Windows can be based on ordering attributes (e.g. a sliding or shifting window), tuple counts, explicit markers (called punctuations), etc. An application or a query may define more than one window and "join" them together.

Query Evaluation

The unique characteristics of data streams lead to different evaluation algorithms and query processing: data elements arrive continuously, possibly at variable arrival rate; data streams are potentially unbounded in size; general retrieval of past elements is very difficult, if not impossible. As justly pointed out by Garofalakis et al. (2002), "you only get one look". This does not exclude the presence of conventional data stored in some traditional database systems.

As a result, algorithms operate mainly in main-memory and involve one-pass techniques (or few-passes, if the input data stream can be organized in input blocks). Operators include – besides the traditional ones (selection, projection) – operations specific for combining and aggregating multiple streams, such as sliding window join and merging. Optimization is very different than traditional DBMS. The goal in DSMS is to maximize the tuple output rate for a query, i.e. instead of seeking the least cost plan, seek the plan with the highest tuple output rate. Specialized data structures (synopses) that allow fast modifications and summarize efficiently information are frequently used. Finally, architectures are usually data-

Relations, Streams → ***Stream Query Languages*** → Relations, Streams flow oriented, i.e. operators are assembled together in a workflow-style diagram.

Adaptive Query Processing

Traditional query processors use a request-response paradigm for query answering, where users pose a query, the optimizer finds the "best" query plan and proceeds with query evaluation. However, the introduction of continuous queries along with the possibility of a large number of registered queries with the stream manager changes this model significantly. An aspect of continuous query is the need for adaptivity to change: unbounded queries will run long enough to experience changes in system and data properties as well as system workload during their run. A continuous query engine should adapt gracefully to these changes, in order to ensure efficient processing over time (Madden, Shah, Hellerstein & Raman, 2002). The goal of an adaptive query processing system is to maximize utilization of storage and computation (in terms of operator sharing) of multiple simultaneous queries present in the system.

Exact vs. Approximate Answers

Traditional database systems compute exact answers for queries. However, there are cases that approximate answers are either "good enough" (e.g histograms, sketches, traditional DBMS) or "the only possibility" (e.g. DSMS where main-memory limitations and one-pass algorithms reduce the class of queries with exact answers). High-quality approximate answers are often acceptable instead of exact answers. Several approximation algorithms for data streams applications have been developed in recent years. This work has led to some general techniques for data reduction and synopsis construction, such as sketches, random sampling, histograms, and wavelets. The research on these summarization techniques led to some work on approximate query answering.

Load Shedding

Systems have a limit on how much incoming stream tuples can handle. As a result, when the arrival rate is too high (e.g. during spikes), queues will build up and the system will become overloaded. In this case, some load shedding - i.e. discarding some incoming stream data – is necessary for the system to continue functioning properly (Babcock, Datar & Motwani, 2004, Tatbul, Çetintemel, Zdonik, Cherniack & Stonebraker, 2003). In general, there are two kinds of load shedding, semantic and random. Semantic shedding is based on filtering, i.e. there is a filter (a predicate) that is applied on each stream tuple with selectivity 1-p. Random shedding is based on dropping randomly incoming data with a probability p%, eliminating thus p% of the input load.

A stream system should be able to answer to three questions, when, where and how much load shedding. The first issue - when load shedding should take place – requires constant monitoring of the system. The second one involves deciding the place in a query plan that load shedders will be added. For example, Aurora, a DSMS prototype, defines a stream query as a network (a workflow) of operators. Load shedding can be modeled as another operator, placed within this network. Dropping tuples as early as possible saves resources but there may be a problem later, if streams fan out to multiple queries. The last issue has to do with estimating of how much shedding is required (percent of random drops or filtering appropriately).

Mining Streams

An orthogonal issue to processing streams is mining and indexing streams at real time (Garofalakis et. al., 2002). The goal of this topic is to find patterns in a sequence of measurements (financial, sensor, etc.) Besides the traditional mining patterns (classification, clustering and association rules), there exist many time series patterns that are of special interest in stream analysis (Faloutsos, 2004, Lerner, Shasha, Wang, Zhao & Zhu, 2004).

SYSTEM PROTOTYPES

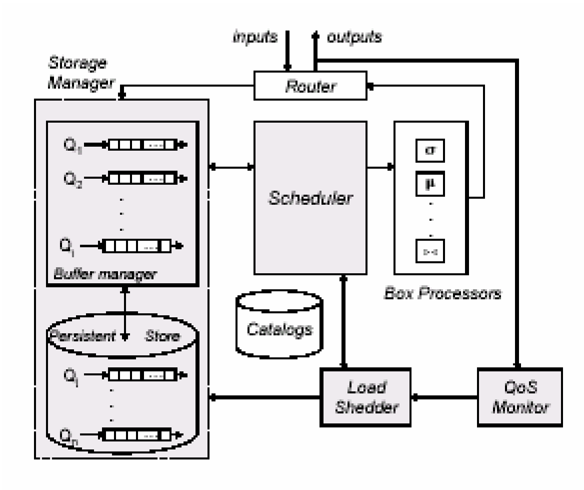
The interest on data streams has led many research organizations (mainly in the U.S.) in the development of data stream prototypes. We list below some of the most well known systems.

- **Aurora.** Aurora is a joint project between Brandeis University, Brown University and MIT targeted towards stream monitoring applications. The processing mechanism is based on the optimization of a set of operators, connected in a data flow graph. Queries (three modes: continuous, ad-hoc, view) are formulated through a graphical "boxes and arrows" interface. Aurora's architecture is shown in Figure 3 (Carney et. al., 2002).

The designers of the Aurora stream system claim five unique features that distinguish it from other stream proposals: a workflow-orientation, a novel collection of operators, efficient scheduling, quality of service concerns and novel optimization methods.

A workflow orientation is necessary due to the frequent preprocessing required for data streams. This way,

Figure 3. Aurora's architecture (Carney et al., 2002)



users design an application (a workflow) with a "boxes" and "arrows" GUI and the system can optimize it more efficiently. The set of operators include "boxes" for filtering, mapping, aggregation (windowed) and join, similar to other stream systems. However, there are some distinct features: windowed operations have a timeout capability, the system deals with "out-of-order" messages (delayed messages), extendability (user-defined functions) and resampling. Scheduling is a major component of Aurora system and it aims on reducing CPU cost of running the scheduler and maximizing the quality of service.

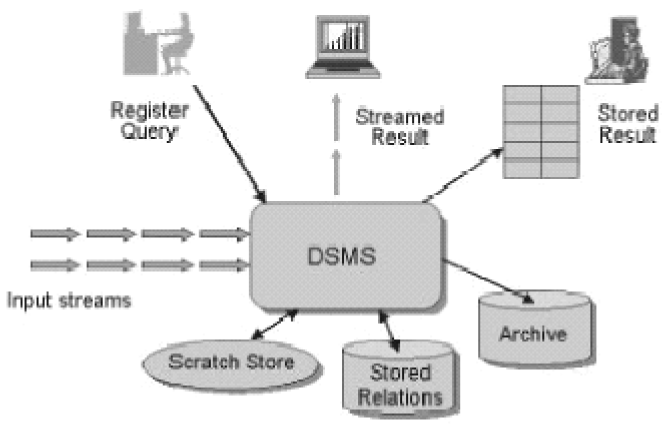
An Aurora workflow is a dynamic object. When new applications get connected to the workflow, the number of boxes and arrows changes. Also ad-hoc queries that are run just once and then discarded have a similar effect. Furthermore, Aurora workflows may be quite large. As a result, Aurora performs only run-time optimization.

- **Gigascope.** AT&T's Gigascope project is a special-purpose data stream system for network applications, such as traffic analysis, intrusion detection, router configuration analysis, network monitoring and performance monitoring and debugging. The central component of Gigascope is a stream manager which tracks the query nodes that can be activated. Query nodes are processes. When they are started, they register themselves with the registry of the stream manager.

Gigascope's query language, GSQL, is an SQL-like language with stream features. All inputs to and outputs of a GSQL query are data streams. Currently GSQL supports selection, join, aggregation and stream merge. A GSQL query is analyzed, translated to C/C++ code, and then executed. Optimization of GSQL queries consists of rearranging the query plan and generated-code optimizations. Another important optimization is pushing selections as far down as possible, even to the network level. To achieve this, the query processor of Gigascope "breaks" queries into high-level and low-level query nodes. Low-level query nodes are separate processes consisting of simple filtering conditions.

- **STREAM.** The STanford stREam datA Manager (STREAM) project at Stanford is developing a general-purpose DSMS for processing continuous

Figure 4. Overview of STREAM Architecture (Arasu et al., 2003)



queries over multiple continuous data streams and stored relations. The architecture of STREAM is shown in Figure 4 (Arasu et al., 2003).

The incoming input streams produce continuous data and drive query processing. Scratch store is used for the intermediate processing of the incoming data and can be stored in memory or on disk. An Archive storage is used for storing some (or all) of the data stream for possible offline processing of expensive mining queries. Users or applications may register continuous queries and get the answers as output data streams or relational results that are updated over time. During the processing of continuous queries, traditional relational tables can be utilized. Currently STREAM offers a Web system interface through direct HTTP, and a Web-based GUI to register queries and view results.

STREAM people developed a declarative query language, CQL (Continuous Query Language) for continuous queries over data streams and relations. They model a stream as an unbounded, append-only bag of (tuple, timestamp) pairs, and a relation as a time-varying bag of tuples supporting updates, deletions and insertions. The key idea is to convert streams into relations using special windowing operations, perform transformations on relations using standard relational operators and then convert back (optionally) the transformed relational data into a streamed answer (Arasu et al., 2003). CQL uses SQL as its relational query language, SQL-99 amendments provide the window specification language, and it includes three relation-to-stream operators. When a continuous query specified in CQL is registered with STREAM, it is compiled into a query plan. The query plan is merged with existing query plans whenever possible, in order to share computation and memory.

- **TelegraphCQ.** TelegraphCQ is a prototype focused on handling measurements of sensor networks. Developed in Berkeley, it is based on adaptive and multi-query processing and uses a mix of SQL and scripting programming languages for expressing complex queries. Its architecture is shown in Figure 5 (Chandrasekaran et al., 2003).

Telegraph, the predecessor of TelegraphCQ, consists of an extensible set of composable dataflow modules or operators that produce and consume records in a manner analogous to the operators used in traditional database query engines. The modules can be composed into multi-step dataflows, exchanging records via an API called Fjords (Madden & Franklin, 2002) that can support communication via either "push" or "pull" modalities.

TelegraphCQ supports continuous queries defined over relational tables and data streams with a rich set of windowing schemes (e.g. landmark and sliding windows).

Other stream projects include Tribeca, developed in Telcordia for network management; NiagaraCQ, a project of University of Wisconsin to monitor internet data (XML-format); Cougar and Amazon of Cornell University, aimed to monitor distributed sensor networks. We present in Table 2 a comparison table between Aurora, Gigascope, STREAM and Telegraph.

Figure 5. TelegraphCQ's architecture (Chandrasekaran et al., 2003)

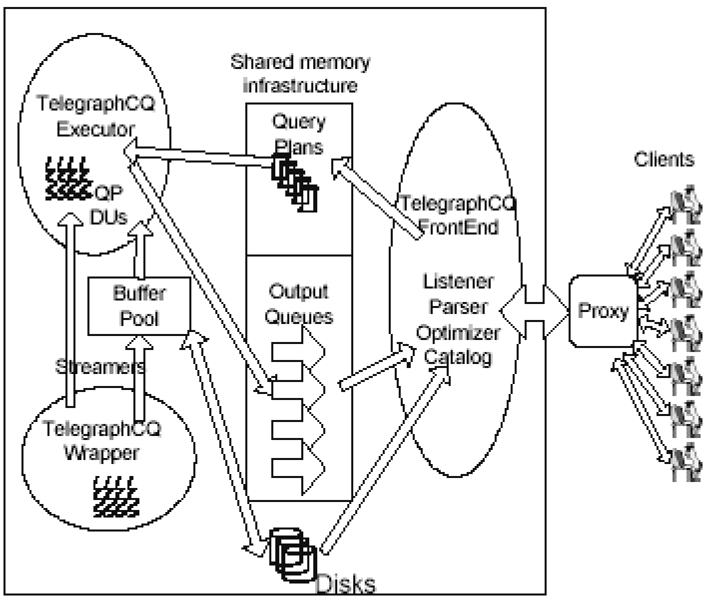


Table 2. Comparing different stream prototypes

Stream System	Input	Output	Query Language	Answers	Evaluation
Aurora	Relations, Streams	Relations, Streams	Operators (boxes and arrows)	Approximate	Run-time optimization
Gigascope	Streams	Streams	GSQL (SQL-like)	Exact	Splitting query to high- and low-level
STREAM	Relations, Streams	Relations, Streams	CQL (SQL-like)	Approximate	Static analysis, relational optimization
Telegraph	Relations, Streams	Relations, Streams	SQL-based, scripting	Exact	Adaptive query processing, multi-query

CONCLUSION

Equipping traditional database management systems with decision support capabilities has been studied extensively in the last decade. Building data warehouses, forming data cubes, creating data marts, carrying out data mining tasks and performing on-line analytical processing and multi-dimensional analysis are well-known subjects in the context of decision-making. However, an emerging trend in data management, data streams, is very different in nature from traditional database applications: queries are different, models and frameworks are special, query processing is peculiar and, subsequently, decision support is defined differently. The keywords in this domain are: real-time, continuous, tactical decisions. A lot of progress has been recently made in this field by several research institutions to build various (either general- or specific-purpose) data stream management systems with rich data analysis capabilities. Given the popularity of sensors and the need for real-time analytics, the ability of data analysis on top of data streams will be a crucial component of future data management and decision support systems.

REFERENCES

Arasu, A., Babcock, B., Babu, S., Datar, M., Ito, K., Motwani, R., et al. (2003). STREAM: The Stanford stream data manager. *IEEE Data Engineering Bulletin*, *26*(1), 19-26.

Babcock, B., Datar, M. & Motwani, R. (2004). Load shedding for aggregation queries over data streams. In *Proceedings of International Conference on Data Engineering (ICDE)*, 350-361.

Babcock, B., Babu, S.,Datar, M., Motwani, R. & Widom, J. (2002). Models and issues in data streams. In *Proc. of the 2002 ACM Symp. on Principles of Database Systems(PODS)*, 1-20.

Babu, S. & Widom, J. (2001). Continuous queries over data streams. *SIGMOD Record*, *30*(3), 109-120.

Carney, D., Cetintemel, U., Cherniack, M., Convey, C., Lee, S. & Seidman, G., et al. (2002). Monitoring streams - A new class of data management applications. In *28th International Conference on Very Large Databases (VLDB)*, 215-226.

Chandrasekaran, S., Cooper, O., Deshpande, A., Franklin, M. J., Hellerstein, J. M., Hong, W., et al. (2003). TelegraphCQ: Continuous dataflow processing for an uncertain world. In *Conference on Innovative Data Systems Research*.

Chaudhuri, S. & Dayal, U. (1997). An overview of data warehousing and OLAP technology. *ACM SIGMOD Record*, *26*(1), 65-74.

Chen, J., DeWitt, D. J., Tian, F. & Wang, Y. (2000). NiagaraCQ: A scalable continuous query system for internet databases. In *ACM SIGMOD, Conference on Management of Data*, 379-390.

Cortes, C., Fisher, K., Pregibon, D., Rogers, A. & Smith, F. (2000). Hancock: A language for extracting signatures from data streams. In *ACM SIGKDD, Conference on Knowledge Discovery and Data Mining*, 9-17.

Cranor, C., Johnson, T., Spatscheck, O. & Shkapenyuk, V. (2003). Gigascope: A stream database for network applications. In *ACM SIGMOD, Conference on Management of Data*, 647-651.

Faloutsos, C. (2004). Indexing and mining streams (Tutorial). In *Proc. of the ACM SIGMOD Intl. Conf. on Management of Data*, 969.

Garofalakis, M., Gehrke, J. & Rastogi, R. (2002). Querying and mining data streams: You only get one look, a tutorial. In *Proc. of the 2002 ACM SIGMOD Intl. Conf. on Management of Data*, 635.

Koudas, N. & Srivastava, D. (2003) Data stream query processing: A tutorial. In *29th Intl. Conf. on Very Large Databases (VLDB)*, 1149.

Lerner, A., Shasha, D., Wang, Z., Zhao, X. & Zhu, Y. (2004). Fast algorithms for time series with applications to finance, physics, music, biology, and other suspects (Tutorial). In *Proc. of the ACM SIGMOD Intl. Conf. on Management of Data*, 965-968.

Madden, S. & Franklin, M. J. (2002). Fjording the stream: An architecture for queries over streaming sensor data. In *Proceedings of the 2002 Intl. Conf. on Data Engineering (ICDE)*, 555-566.

Madden, S., Shah, M. A., Hellerstein, J. M. & Raman, V. (2002). Continuously adaptive continuous queries over streams. In *Proc. of the ACM SIGMOD Intl. Conf. on Management of Data*, 49-60.

Muthukrishnan, S. (2003). Data streams: Algorithms and applications. In *Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 413.

Sullivan, M. (1996). Tribeca: A stream database manager for network traffic analysis. In *22th International Conference on Very Large Databases (VLDB)*, 594.

Tatbul, N., Çetintemel, U., Zdonik, S., Cherniack, M. & Stonebraker, M. (2003). Load shedding in a data stream manager. In *29th Intl. Conf. on Very Large Databases (VLDB)*, 309-320.

Terry, D., Goldberg, D., Nichols, D. & Oki, B. (1992). Continuous queries over append-only databases. In *Proc. of the 1992 ACM SIGMOD Intl. Conf. on Management of Data*, 321-330.

Zdonik, S., Stonebraker, M., Cherniack, M., Çetintemel, U., Balazinska, M., Balakrishnan, H. (2003). The aurora and medusa projects. *IEEE Data Engineering Bulletin 26*(1), 3-10.

KEY TERMS

Continuous Queries: The answer to a continuous query is produced over time, reflecting the stream data seen so far. Answers may be stored and updated as new data arrives or may be produced as data streams themselves.

Data Streams: Data items that arrive on-line from multiple sources in a continuous, rapid, time-varying, possibly unpredictable fashion.

Data Stream Management Systems (DSMS): A data management system providing capabilities to query and process data streams and store a bounded part of it.

Load Shedding: The discarding of input data by the DSMS when the input stream rate exceeds system capacity. It can either be semantic (based on certain semantic rules) or random.

Measurement Data Streams: Data streams representing successive state information of one or more entities, such as sensor, climate or network measurements.

Network Traffic Management: Monitoring a variety of continuous network data streams at real-time, such as packet traces, packet flows and performance measurements in order to compute statistics, detect anomalies and adjust routing.

Stream Window: A mechanism to extract a finite set of records from an infinite stream. This mechanism selects stream items based on time periods, counting, or explicit starting and ending conditions.

Transactional Data Streams: Data streams representing log interactions between entities, such as credit card transactions, phone calls and web click streams.

Incremental Expansion of a Distributed Database System

Amita Goyal Chin
Virginia Commonwealth University, USA

INTRODUCTION

Recent years have witnessed an increasing trend in the implementation of distributed database management systems (DDBMSs) for more effective access to information. An important quality of these systems, consisting of *n* servers loosely connected via a communication network, is to adjust to changes in workloads. To service increases in demand, for example, additional servers may be added to the existing distributed system and new data allocations computed. Conventionally, this requires a system shutdown and an exhaustive data reallocation. Such static methods are not practical for most organizations for these methods result in high costs and in periods of data unavailability.

We present the incremental growth framework to address incremental expansion of distributed database systems. Data is reallocated using one of two data reallocation heuristics—Partial REALLOCATE or Full REALLOCATE. Both heuristics are greedy, hill-climbing algorithms that compute new data allocation based on the specified optimization parameter of the objective cost function. Due to their linear complexity, both heuristics can be used to solve both small and large complex problems, based on organizational needs. The REALLOCATE algorithms in conjunction with the SimDDBMS simulator can be used to answer many practical questions in distributed database systems. For example, in order to improve system response time, a database administrator (DBA) may use SimDDBMS for parametric evaluation. For example, the DBA may analyze the effect of upgrading CPU processing capability, increasing network transfer speed, or adding additional servers into the distributed database system. Furthermore, SimDDBMS may easily be modified to evaluate heterogeneous servers with different CPU processing capabilities. A DBA may also use SimDDBMS to determine the impact and cost-benefit analysis of adding some number, $s \leq 1$, additional servers at one time.

BACKGROUND

Following the pioneering work in Porcar (1982), many researchers have studied the file or data allocation problem (Daudpota, 1998; Ladjel, Karlapalem, & Li, 1998; So, Ahmad, & Karlapalem, 1998; Tamhankar & Ram, 1998;). Since optimal search methods can only be used for small problems, heuristic methods are often used for solving large data allocation problems (Apers, 1988; Blankinship, 1991; Ceri, Navathe, & Wiederhold, 1983; Chin, 2001; (Chin) Goyal, 1994; Du & Maryanski, 1988). Researchers have studied both the static data allocation problem, in which data allocations do not change over time, and the dynamic data allocation problem (Brunstrom, Leutenegger, & Simha, 1995; Theel & Pagnia, 1996; Wolfson, Jajodia, & Huang, 1997), which may be adaptive or nonadaptive. Adaptive models (Babcock, Babu, Motwani, & Datar, 2003; Levin, 1982; Levin & Morgan, 1978; Son, 1988) are implemented when the system senses a substantial deviation in access activities; these models determine a one-time reallocation (for a relatively short period of time) in response to surges in demand. For example, the volume of reservations for a particular airline route may increase during a specific season. Therefore, an airline reservation system may temporarily store additional copies of the files associated with the route at a local server. However, this is a short-term situation, which is resolved by introducing replicated file copies. Nonadaptive models (Levin; Porcar; Segall, 1976) are employed at the initial system design stage or upon system reorganization; these models do not adjust to variations in system activities.

Most previous research on data allocation assumes a fixed number of servers in the distributed database system (Carey & Lu, 1986; Chu, 1969; Laning & Leonard, 1983; Lee & Liu Sheng, 1992; Rivera-Vega, Varadarajan, & Navathe, 1990). Experiments and simulations are designed to test DDBMS factors such as the degree of data replication, workloads per server, and different levels and classes of queries and transactions (Carey & Lu; Ciciani, Dias, & Yu, 1990). Simulation runs vary the number of servers to arbitrary values. However, these values are fixed per run and vary only between runs.

INCREMENTAL GROWTH FRAMEWORK

The incremental growth framework is invoked when system performance, as computed using the objective cost function, is below the acceptable threshold (specified by the DBA). To return to an acceptable state, new servers are introduced incrementally, one at a time, into the distributed database system. With the introduction of each new server, a new data reallocation for the system is computed. This process is iteratively executed until acceptable performance is achieved or the number of servers equals the number of relations in the distributed database system (the latter constraint can easily be relaxed in a distributed database system housing partitioned data). The incremental growth framework, which can easily be adapted for one-server or multiple-server systems, can be used by small, midsize, and large organizations, each having distributed database systems of varying size. In one-server systems, the initial data allocation locates all relations at the server. In multiple-server systems, the current data allocation is required as input into the framework. Additional input information required for the incremental growth framework includes: the database server or servers, including the local processing capacity; the network topology, including transmission capacity; the database relations, including relation sizes and selectivities; the query set; the optimization parameter; and the acceptable threshold for the optimization parameter.

DEFINITIONS

The relational data model is used to describe the data and query processing on the data. Only simple queries are considered. Queries are assumed to be independent and are solved independently. Queries are processed in parallel in the distributed database system. To simplify the estimation of query result sizes, the concept of selectivity (Blankinship, 1991; Chin, 1999; Date, 1991; (Chin) Goyal, 1994) is utilized. Attribute values are assumed to be uniformly distributed, and each attribute in a relation is assumed to be independent of all other attributes in the database. The simple query environment has been chosen because it has a manageable complexity while remaining realistic and interesting.

The parameters describing the simple query environment are (Blankinship, 1991; Chin, 1999; (Chin) Goyal, 1994; Hevner & Yao, 1979):

S_i: Network Servers, i = 1, 2, ..., s, s+1 (S_{s+1} = the new server joining the system),
R_j: Relations, j = 1, 2, ..., r

For each relation R_j, j = 1, 2, ..., r:
n_j: number of tuples,
a_j: number of attributes,
β_j: size (in bytes)

For each attribute d_{jk}, k = 1, 2, ..., a_j of relation R_j:
p_{jk}: attribute density, the number of different values in the current state of the attribute divided by the number of possible attribute values. So, $0 <= p_{jk} <= 1$ (Hevner & Yao, 1979). During join operations the density is used as a selectivity coefficient.
w_{jk}: size (in bytes) of the data item in attribute d_{jk}

For local transaction processing, each server in the distributed database system maintains a queue of incoming requests. Queries are maintained in queue until they are processed using a first in, first out (FIFO) order.

Finally, the distributed database system maintains a centralized data dictionary housing the following information (Blankinship, 1991; (Chin) Goyal, 1994; Hevner & Yao, 1979):

- for each relation R_j, j = 1, 2, ..., r: n_j, a_j, β_j, and S_j (server to which relation R_j is allocated)
- for each attribute d_k, k = 1, 2, ..., a_j of relation R_j: p_{jk}, w_{jk}, and b_{jk} (projected size, in bytes, of attribute $d\,j_k$ with no duplicate values)

Optimizing query strategies is not within the scope of this research. However, since the optimal data allocation is dependent on the implemented query strategy, when computing new data allocations, Algorithm Serial (Hevner & Yao, 1979) for query processing is implemented. Any query optimization algorithm from the research literature, however, can be used in place of Algorithm Serial.

Algorithm Serial (Hevner & Yao, 1979) considers serial strategies to minimize total transmission time in the simple query environment. For each query q accessing ψ relations, there are $\psi!$ possible combinations for processing q. The serial strategy consists of transmitting each relation, starting with R_1, to the next relation in a serial order. The strategy is represented by $R_1 \rightarrow R_2 \rightarrow \ldots \rightarrow R_\sigma$, where σ is the number of relations in the query (Hevner & Yao, 1979).

Consider, for example, a query which accesses relations A, B, and C. Then, the $\psi! = 6$ processing combinations for the query are: $A \rightarrow B \rightarrow C$, $A \rightarrow C \rightarrow B$, $B \rightarrow A \rightarrow C$, $B \rightarrow C \rightarrow A$, $C \rightarrow A \rightarrow B$, $C B \rightarrow A$. Therefore, given four queries—two of which access two relations, one of which accesses three relations, and one of which accesses four relations—the number of possible serial strategy combinations is (2!)(2!)(3!)(4!) = (2)(2)(6)(24) = 576. The serial order is computed so that $\beta_1 \leq \beta_2 \leq \ldots \leq \beta_s$, where β_j is the size of relation R_j, j = 1, ..., r (Hevner & Yao, 1978).

SYSTEM COST EQUATIONS

A fully connected, reliable communication network (Lee & Liu Sheng, 1992) with all servers having equal local storage and processing capacity is assumed. A single instance of each relation is allocated to the distributed database system (Blankinship, 1991; Cornell, 1989; (Chin) Goyal, 1994). (Data partitioning and data replication are not considered in this research.)

When measuring system response time, the objective cost function consists of three cost components: transmission costs, local processing costs, and queuing costs. Costs are measured in terms of the number of CPU cycles or time ticks ((Chin) Goyal, 1994) needed to complete a task. The system response time is equal to the number of CPU cycles needed to process $Q = \{Q_1, \ldots, Q_q\}$ queries.

Transmission cost equations are identical between any two servers (Blankinship, 1991; (Chin) Goyal, 1994), and costs are based on the amount of data transmitted. Local processing cost equations are also identical at any two servers, and costs are based on the amount of data processed. Queuing costs are based on the number of CPU cycles a query spends in queue at each server. Storage costs are considered negligible and are not considered in the cost analysis.

Using the additional notation:

CS_n: cumulative selectivity for Q_n
$n = 1, \ldots, q$
QR_n: query result for Q_n
$n = 1, \ldots, q$
QWT_{ni}: wait time in queue for Q_n at S_i
$n = 1, \ldots, q$; $i = 1, \ldots, s+1$
LPT_{ni}: local processing time for processing Q_n at S_i
$n = 1, \ldots, q$; $i = 1, \ldots, s+1$
NTT_{nij}: network transfer time for transferring Q_n from S_i to S_j
$n = 1, \ldots, q$; $i, j = 1, \ldots, S+1$; $i \neq j$
Θ_{ni}: transmission of Q_n to S_i
$n = 1, \ldots, q$; $i = 1, \ldots, s+1$

ρ: CPU rate per CPU cycle
μ: network transfer rate per CPU cycle

we state our cost equations as follows ((Chin) Goyal, 1994):

$QR_n = (CS_n)(\beta_j)$ $n = 1, \ldots, q$; $j = 1, \ldots, r$

$$LPT_{ni} = \frac{\beta_j + QR_n}{\rho} \quad \text{at } S_i \; n = 1, \ldots, q;\ j = 1, \ldots, r;\ i = 1, \ldots, s+1$$

$$NTT_{nij} = \frac{QR_n}{\mu} \quad \text{from } S_i \text{ to } S_j \; n = 1, \ldots, q;\ i, j = 1, \ldots, s+1;\ i \neq j$$

$$QWT_{ni} = T_{P_{ni}} - T_{Q_{ni}} \quad n = 1, \ldots, q;\ i = 1, \ldots, s+1$$

where,

$T_{Q_{ni}}$ = the CPU cycle S_i places Q_n into its queue
$T_{P_{ni}}$ = the CPU cycle S_i begins processing Q_n

The objective cost function when computing data allocations to minimize response time is to minimize:

$$RT = \sum_{t=1}^{T} C(t) + \varsigma$$

where,

$$C(t) = \begin{cases} 1 & if\,(Q_n \in Q \text{ in queue at } S_i) \\ \vee & (Q_n \in Q \text{ in process at } S_i) \\ \vee & (Q_n \in Q \text{ in transmission at } S_i \rightarrow S_j) \\ 0 & otherwise \end{cases}$$

and ς = system idle time while processing $Q = \{Q_1, \ldots, Q_q\}$.

DATA REALLOCATION HEURISTICS

We present two heuristics for data reallocation: Partial REALLOCATE and Full REALLOCATE. Both algorithms are greedy, iterative, "hill-climbing" heuristics that will not traverse the same path twice. With each iteration, they will find a lower cost solution, or they will terminate. Both algorithms require as input: the current data allocation, the relations, and the queries in the distributed database system.

We define the notation:

$S = \{S_1, \ldots, S_s, S_{s+1}\}$: set of servers ($S_{s+1}$ = the new server)
$R = \{R_1, \ldots, R_r\}$: set of relations allocated to S_i, $i = 1, \ldots, s$
$R' = \{R'_1, \ldots, R'_r\}$: set of relations allocated to S_{s+1}
$R \cap R' = \emptyset$

$R_j \Rightarrow S_i$: permanent allocation of R_j to S_i, $j = 1, \ldots, r$; $i = 1, \ldots, s$
$R_j \rightarrow S_i$: temporary allocation of R_j to S_i, $j = 1, \ldots, r$; $i = 1, \ldots, s$
$O_o = \delta(\forall_j \in R, R_j \Rightarrow S_i)$, $i = 1, \ldots, s$
$O_{(s+1)j} = \delta(R_j \rightarrow S_{s+1})$ $j = 1, \ldots, r$

where $d(R_j \rightarrow S_i)$ and $d(R_j \Rightarrow S_i)$ is the objective cost function evaluated for $R_j \rightarrow S_i$ and $R_j \Rightarrow S_i$, respectively.

Each relation R_j must be allocated to a server and can be allocated to only one server at any given time. Therefore, For each $R_j \in (R \cup R')$, $\sum_{i=1}^{s+1} X_{ij} = 1$ where

$$X_{ij} = \begin{cases} 1 & R_j \to S_i \vee R_j \Rightarrow S_i \\ 0 & otherwise \end{cases}.$$

PARTIAL REALLOCATE

The steps of the Partial REALLOCATE algorithm are:

Step 1: Compute O_o.

Step 2: For each $R_j \in R$, Compute $O_{(s+1)j} = \delta(R_j \to S_{s+1})$, where for $R'' = R - R_j$, "R_k's $\in R''$, $R_k \neq> S_{s+1}$, $1 \leq k \leq (r-1)$.

Step 3: Compare $O_\Delta = \underset{j}{MIN}\, O_{s+1}$ to O_o. If $O_\Delta{}^{3}\, O_o$, O_o is the local optimum. If $O < O_o$, update O_o to O_Δ, $R' = R' + R_j$, $R = R - R_j$, $R_j \Rightarrow S_{s+1}$.

FUTURE TRENDS

Much research has been conducted in the data allocation problem. Partial REALLOCATE and Full REALLOCATE are particularly interesting because they solve the data allocation problem in an expanding system environment. Expansion of these algorithms for heterogeneous computer and wireless (Lin, Hu, & Lee, 2004) environments will offer many challenges and benefits for current and future business environments. Due to their linear complexity, the Partial and Full REALLOCATE algorithms can be used for large, complex problems while achieving good results as compared to the exhaustive search optimal.

CONCLUSION

Over 5,800 simulation experiments were run using only Full and Partial REALLOCATE (Chin, 2001; (Chin) Goyal 1994). The experiments were run to minimize system response time in the simple query environment. A "base case" simulation was run for each problem. The "base case" assumed there was only one server in the distributed database system, with all of the relations allocated to this one server. The Exhaustive Search algorithm was used for benchmark comparisons. However, due to the exponential growth in the search space of the Exhaustive Search algorithm, Exhaustive Search computations were restricted to a maximum of four servers and five relations (a combined search space of 1+32+243+1,024 = 1,300 possible allocations). The effect on the system response time of each of the following parameters (as independent variables) was studied: CPU processing rate, network transfer rate, number of queries, number of relations. The parametric studies demonstrated the consistency of the REALLOCATE algorithms across a broad range of parametric values.

Simulation experiments have shown that Partial REALLOCATE and Full REALLOCATE provide good solutions as compared to Exhaustive Search optimums. The Partial and Full REALLOCATE algorithms considerably reduced problem search space and, hence, the cost of testing relation-server combinations. Given the cost of each test to be one unit, implementing Partial REALLOCATE over Exhaustive Search resulted in a cost savings of S^R - R units; implementing Full REALLOCATE over exhaustive search resulted in a cost savings of $S^R - \sum_{r=1}^{R} r$ units.

Partial and Full REALLOCATE have different strengths. If the cost of testing relation-server combinations is high, Partial REALLOCATE is the algorithm of choice. This is because Partial REALLOCATE has a much smaller search space than Full REALLOCATE (R vs $\sum_{r=1}^{R} r$, where R is the number of relations in the distributed database system). If the cost of adding additional servers is high, Full REALLOCATE is the algorithm of choice. As demonstrated by the simulation experiments, Partial REALLOCATE is a server-hungry algorithm. It generally requires two to three times as many servers as Full REALLOCATE in order to find a comparable solution.

REFERENCES

Apers, P. M. G. (1988). Data allocation in distributed database systems. *ACM Transactions on Database Systems*, *13*(3), 263-304.

Babcock, B., Babu, S., Motwani, R., & Datar, M. (2003). Chain: Operator scheduling for memory minimization in data stream systems. *Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data*, 253-264.

Blankinship, R. (1991). *An iterative method for distributed database design*. Unpublished doctoral dissertation, University of Maryland at College Park.

Brunstrom, A., Leutenegger, S. T., & Simha, R. (1995). Experimental evaluation of dynamic data allocation strategies in a distributed database with changing workloads.

Proceedings of the 1995 International Conference on Information and Knowledge Management, 395.

Carey, M. J., & Lu, H. (1986, June). Load balancing in a locally distributed database system. *Proceedings of the ACM-SIGMOD International Conference on Management of Data*, 108-119.

Ceri, S., Navathe, S., & Wiederhold, G. (1983). Distribution design of logical database schemas. *IEEE Transactions on Computers*, *SE-9*(4), 487-504.

Chin, A. G. (1999). An Optimization Algorithm for Dynamic Data Migration in Distributed Database Systems. *Journal of Computer Information Systems*, *39*(4).

Chin, A. G. (2001). Incremental data allocation and reallocation in distributed database systems. *Journal of Database Management*, *12*(1).

(Chin) Goyal, A. (1994). *Incremental growth and reorganization in distributed database systems.* PhD Dissertation, University of Maryland at College Park.

Chu, W. W. (1969). Optimal file allocation in a multiple computer system. *IEEE Transactions on Computers*, *C-18*(10), 885-890.

Ciciani, B., Dias, D. M., & Yu, P. S. (1990). Analysis of replication in distributed database systems. *IEEE Transactions on Knowledge and Data Engineering*, *2*(2), 247-261.

Cornell, D. W. (1989). An Optimal Site Assignment for Relations in the Distributed Database environment. *IEEE Transaction on Software Engineering*, 15 (8), 1004-1009.

Date, C.J. (1991). *Introduction to database systems*. Addison-Wesley.

Daudpota, N. (1998). Five steps to construct a model of data allocation for distributed database systems. *Journal of Intelligent Information Systems*, *11*(2), 153-168.

Du, X., & Maryanski, F. J. (1988). Data allocation in a dynamically reconfigurable environment. *Proceedings of the Fourth International Conference on Data Engineering*, 74-81.

Hevner, A. R. & Yao, S.B. (1978). Optimization of Data Access in Distributed Systems. *Computer Science Department, Purdue University, Technical Report*, TR281.

Hevner, A. R. & Yao, S.B. (1979). Query Processing in Distributed Database Systems. *IEEE Transactions on Software Engineering*, *SE-5*(3).

Ladjel, B., Karlapalem, K., and Li, Q. (1998). An Iterative Approach for Rules and Data Allocation in Distributed Deductive Database Systems, *Proceedings of the seventh international conference on Information and knowledge management*, 356-363.

Laning, L. J., & Leonard, M. S. (1983). File allocation in a distributed computer communication network. *IEEE Transactions on Software Engineering*, *C-32*(3), 232-244.

Lee, H., & Liu Sheng, O. R. (1992). A multiple criteria model for the allocation of data files in a distributed information system. *Computers and Operations Research*, *19*(1), 21-33.

Levin, K. D. (1982). Adaptive structuring of distributed databases. *Proceedings of the National Computer Conference*, 691-696.

Levin, K. D., & Morgan, H. L. (1978). A dynamic optimization model for distributed databases. *Operations Research*, *26*(5), 824-835.

Lin, C.-W., Hu, H., & Lee, D.-L. (2004). Adaptive realtime bandwidth allocation for wireless data delivery. *Wireless Networks*, *10*(2), 103-120.

Porcar, H. (1982). *File migration in distributed computing systems*. Unpublished doctoral thesis, University of California at Berkeley.

Rivera-Vega, P. I., Varadarajan, R., & Navathe, S. B. (1990). Scheduling data redistribution in distributed databases. *Proceedings of the Symposium on Reliability in Distributed Software and Database Systems*, 166-173.

Segall, A. (1976). Dynamic file assignment in a computer network. *IEEE Transactions on Automatic Control*, *AC-21*, 161-173.

So, S.-K., Ahmad, I., & Karlapalem, K. (1998, October). Data allocation algorithm for distributed hypermedia documents. *The 1998 IEEE 17th Symposium on Reliable Distributed Systems*, 473-478.

Son, S.H. (1988, December). Replicated data management in distributed database systems. *SIGMOD Record*, *17*(4), 62-69.

Tamhankar, A., & Ram, S. (1998). Database fragmentation and allocation: An integrated methodology and case study. *IEEE Transactions on Systems, Man, and Cybernetics*, *28*(3), 288-305.

Theel, O. E., & Pagnia, H. (1996). Bounded dynamic data allocation in distributed systems. *The 1996 Third International Conference on High Performance Computing*, 126-131.

Wolfson, O., Jajodia, S., & Huang, Y. (1997). An adaptive data replication algorithm. *ACM Transactions on Database Systems*, *22*(2), 255-314.

KEY TERMS

Data Allocation: The process of determining what data to store at which servers in a distributed system.

Heuristic: An algorithm, particularly used to solve and achieve near-optimal solutions to intractable problems.

Incremental Growth Framework: A methodology for incrementally expanding a distributed database system. A new data allocation is computed with the introduction of each new server.

DBA (Database administrator): A person responsible for successfully maintaining a database system.

Partial REALLOCATE: A heuristic algorithm for computing a new data allocation, given some number of server and a system optimization parameter. This heuristic evaluates the effect of independently moving each database relation to the new server joining the distributed database system.

Full REALLOCATE: A heuristic algorithm for computing a new data allocation, given some number of server and a system optimization parameter. This heuristic is iterative, searching for stepwise solution refinement. This heuristic evaluates the effect of independently moving each database relation to the new server joining the distributed database system. It then holds this relation at the new server and reiterates with a reevaluation of moving an additional relation.

Inexperienced and Global Software Teams

Kim Man Lui
The Hong Kong Polytechnic University, Hong Kong

Keith C. C. Chan
The Hong Kong Polytechnic University, Hong Kong

INTRODUCTION

We dream of a way in which software can be quickly or even automatically produced. In the past, we achieved some success in some areas; however, we continue to face an excess of software demand over supply in general. The reason is obvious: Software is needed by not only home and commercial computers, but by almost every kind of electronic device, such as mobile phones, digital diaries, digital cameras, TVs, cars, and so on (Rischpater, 2001; Sapaty, 1999).

Given that the number of qualified programmers cannot be increased drastically and rapidly, software managers in most parts of the world will likely have to live with the human-resources shortage problem for some time (Information Technology of America, 2000). To deal with this shortage, we have to consider forming global software teams in which members are recruited from all over the world and software is developed in a "distributed" manner. Forming such a global software team can have many advantages. In addition to alleviating the problems caused by the scarcity of human resources, programmers on a global team would be free to work without the confines of physical location.

Although the idea of forming a global software team may increase the size of the pool of programmers that can be recruited, quality is a concern. Software managers want programmers who can deliver quality work. But software quality cannot be guaranteed and is hard to justify. Although the managers would prefer to establish a team consisting of experienced programmers, they are in reality faced with the problems of managing inexperienced programmers for a software project, especially in software-developing countries such as China, Poland, South Africa, and so forth.

It is not difficult to see that maintaining a team with a large proportion of inexperienced members significantly reduces running expenses (Figure 1) as there could be a tremendous salary gap between skilled and unskilled developers. Companies that operate with a tight cash flow will normally have an inexperienced software team as they try to minimize costs.

This paper discusses our experiences in running an inexperienced software team and a global software team. We believe that sharing our experience is useful to organizations thinking of exploiting relatively cheaper labor in developing countries such as China, Poland, and South Africa (Sanford, 2003).

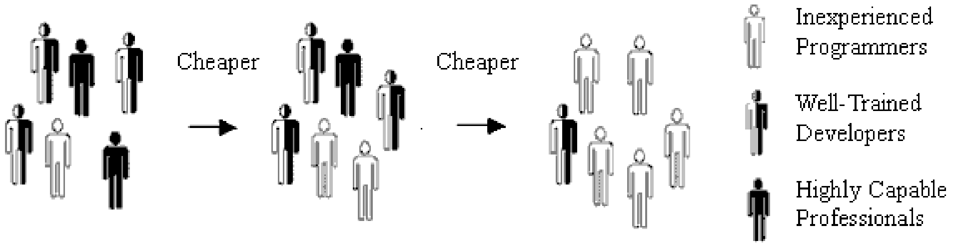
BACKGROUND

This section reviews real cases that have driven the formation of an inexperienced software team and a global software team. The mainstay of managerial decisions for building such teams is financial or environmental.

Discovering Developing Areas: Inexperienced Software Team

It has been a trend for active, rural industrialization, in which manufacturing plants move from more developed regions to less developed ones, to exploit the lower costs

Figure 1. From a professional team to an inexperienced team



of land, labor, and distribution channels (Otsuka, 2001). In order to manage these plants better, management information systems (MISs) need to be developed. Recruiting labor for manufacturing is easy in less developed regions, but recruiting IT professionals for developing an integrated, customized MIS in phases is difficult.

In developing countries, for example, China, the demand for IT professionals in larger cities is so high that it is almost impossible for any manufacturing plant in a rural area to recruit people. Workers in a poor rural area are usually inexperienced. Even though the alternative of employing expatriates might sound reasonable, it is not practical. Instead of in-house development, we might evaluate a third-party solution. The additional expenses incurred in purchasing vendor products, in consultancy services, maintenance, version upgrading, training, traveling, and so forth are expected to be much larger than what can be saved from exploiting cheaper labor. In less developed areas, many programmers do not receive proper training in computing. In addition, the turnover rate is typically high. As long as they have received some training, many workers will seek a job with better career prospects in a more developed city. This results in a vicious cycle in which the project manager always has to work with programmers who are inexperienced. The high turnover rate sometimes is aggravated by resignations without any advance notice. People tender their resignation and leave on the same day. They attempt to keep their current job while seeking any opportunity. Clearly, handing over work is very difficult if not impossible, and the team has to work understaffed constantly.

One may suggest that educating inexperienced people or allocating suitable jobs according to an individual's ability should fulfill the same purposes. However, when the knowledge and experience of staff members is not aligned with the tasks assigned, the learning curve can be steep and long (Amrine, Ritchey, Moodie, & Kmec, 1993). Nevertheless, when a staff member becomes well trained in some less developed regions in China or in a small company in Denmark, for example, his or her determination to look for better job prospects elsewhere will become stronger (Lui & Chan, 2000a; Lui & Chan, 2000b). Training, therefore, does not provide a promising solution in this case. Contradicting what we might expect in well-developed regions, certified professional programs psychologically encourage people to leave a company with little opportunity, or a less developed region, sooner. Some senior managers are disturbed by this phenomenon and say that they are always training another company's staff. The idea of allocating developers according to their skill set is not feasible when all team members are inexperienced. Human resource allocation can therefore be implemented only to a limited extent. Better knowledge management, rather than adopting conventional principles, is required.

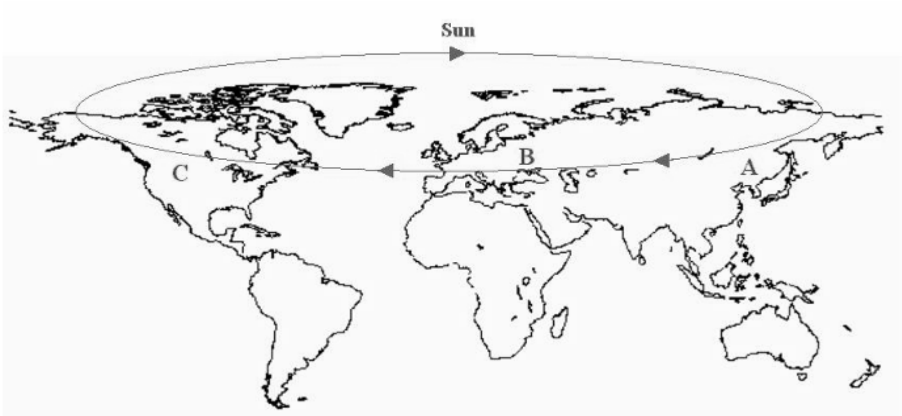
Around the Clock: Global Team

A small but ambitious company selling weight-loss and nutritional products, which was headquartered in New York, had a number of small offices of 40 employees in different parts of the world. For each such office, one to two staff members were hired to provide IT support. When the MIS system needed to be modified to meet requirements for local processing, requests for modifications would be sent to the head office. The result was that more resources were required at the head office to provide ongoing support to the branches. Although a larger software team was thus required at the head office, IT staff at the branches might have time to spare. The load-balancing problem got worse when the number of branches increased. The question, naturally, was to decide if it was possible to link people to establish a global software team. The ideal and more productive approach for the above could be a global software team. The team in each site then plays a role more or less as distributed agents following a communication scheme from a coordination agent.

A global software team can even be formed locally if the team is set up in different locations within the same country or in nearby countries or regions. This means that there may not be much difference in time zones and culture. In such situations, the term *multisite* software team can be used more generally to describe a software system developed by teams that are physically separated from each other in different cities of a country or in different countries. Compared with a global team, a multisite team in nearby time zones can be managed with less complexity and fewer challenges. Multisite software teams of this scale, when compared to global teams, have the constraint of relatively limited service hours. A client who sends a request over the Internet around the world normally demands a prompt reply. But if all teams are in the same time zone, meeting the demand immediately outside office hours is not easy. In any case, the management framework required for a global or multisite team should be very similar. To further explore around-the-clock development (see Figure 2) and global development, we realize the intrinsic difference is how synchronization of work in progress proceeds. We concluded that the challenges of managing around-the-clock tasking widely cover managerial and technical problems of non-around-the-clock global software development.

A global team with one site in Asia, one in Europe, and one in North America maximizes time use by working around the clock.

Figure 2. Around-the-clock development (also called around-the-sun development)



Around-the-clock development does release severe pressure to improve time-to-market by means of time zone differences. But, there has not been a model for this kind of global software development (Carmel, 1999; Karolak, 1998). It is easier to manage a global software team with less strict synchronization among different sites. For example, while waiting for the result from another site, this center could work on some other tasks for the same project. However, in around-the-clock development, the synchronization of work in progress and communications conforms to a rigid schedule based on natural rhythm, like the sun rising at 5:30 a.m. and setting at 6:00 p.m. The progress of a team relies on not only the team itself, but also on the progress of another team that posts the deliverables to the team by the end of the day.

A traditional framework does not provide a sound solution (McMahon, 2001). However, some light was shed when we recalled how quickly we copied our classmates' work at school to learn the part that would be included on an exam paper. For example, given a statistics question and its model answer, we are able to solve a similar problem by following each step of the standard solution, such as getting the sum first, then the mean, and the deviation, and so forth. If we are interrupted while solving that problem, any of our classmates can continue the calculation at the exact point where we stopped. When the problem has a large amount of data and needs a week to complete, the time of interruption will then be sunset. By forwarding our work to another colleague in a different time zone, the work can be continued. This process is repeated until it has been finished.

The above shows that the success of around-the-clock development greatly depends on what type of application we build and what methodology we use to manage a global team. Being too innovative implies that a similar system cannot be found for reference of development sites. In this case, our statistics example can be interpreted as the model answer corresponding to a generic system for a particular application such as a database or a Web system, and the similar assignment is a new IT project of the same type as that generic system. The bad news is that this approach is unable to cope with any new type of project completely. Fortunately, a high portion of information projects nowadays is related to commercial database applications and Web applications.

HOW THE TWO TEAMS ARE RELATED

At a glance, an inexperienced development team and a global software team conjure up two totally different pictures. However, some problems can be dealt with by the same common solution. We start with the examination of commonalities between them, shown in Table 1.

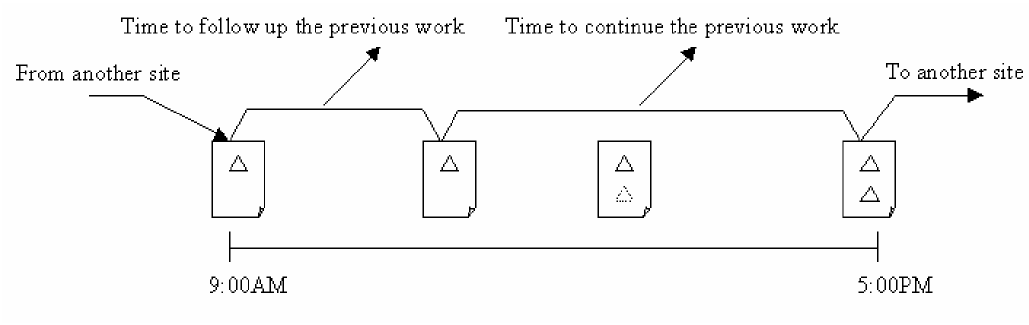
Turnover

Programmers are mostly inexperienced in less developed regions within the country. As long as they gain experience and receive some training, many of them will look for

Table 1. Characteristics of software development teams: An inexperienced team and a global team

	An Inexperienced Software Team in a Less Developed Region	A Global Software Team using Around-the-Clock Development
1	Personnel turnover	Task turnover
2	Weak IT knowledge and lack of IT project experience	Varied experience and knowledge at each site

Figure 3. Anatomy of the process of around-the-clock development



a new job in more developed cities in the same country or in other countries (Morris, 1995). As a result, the turnover rate is high. This situation is aggravated by the fact that many people who tender their resignation prefer to leave immediately in order to maximize their own benefits. Handing over work is very difficult if not impossible. In short, personnel turnover causes task turnover.

A phenomenon that shares the same characteristics of a rapid turnover rate is some global teams achieving around-the-clock software development. For the latter, a team that is to follow up on the work done by another team has to understand the code without being able to communicate with the other team, as the team members are asleep on the other side of the globe.

The basic challenge of around-the-clock work here is the time needed to follow up in order to continue the task delivered electronically from another site, and the remaining time for working and then relaying to another site at sunset, depicted in Figure 3. There are two unconventional problems. The subprocess in a site must be uninterrupted or not stopped; otherwise, the whole process will be suspended as is the progress. In addition, the sum of the working hours to follow up and to continue is just one day. Suppose each site works for 8 hours a day. The efficiency of around-the-clock development at one site will be as follows:

$$\text{Efficiency} = \frac{\text{8 hours} - \text{Time used to follow up previous works}}{\text{8 hours}} \times 100\%.$$

Around-the-clock software development has similar problems associated with a high personnel turnover rate. Neither guarantees that the expected outcome is achievable for sure, and both strongly require a very quick job hand-over without face-to-face, lengthy explanations from the previous developers. In the next section, we deliberate how plagiarism-based programming can be used to cope with a sudden hand-over.

Knowledge Management

To develop a database application, a software team may encounter many kinds of technical problems that require different skill sets, such as inserting records into a database, deleting records from a database, updating those records, controlling data integrity, controlling transactions, and the like (Taylor, 2003). In order to do programming, a software team should be equipped with minimum expertise that allows the team to complete part of, if not all, programming jobs. Developers below that level could do nothing by themselves. Figure 4 illustrates this idea.

Our goal is to lower the line of minimum expertise. But how? Let us look at an example. Suppose you have a group of people, say those with a learning disability, who are able to count numbers only but do not understand addition. If we want them to do the addition without a calculator, the best way seems to teach them the calculation. This would be the minimum expertise for this problem. Still, the learning curve may be long. (If they are your employees, your boss can not help but wonder why you would hire these people and make the office a learning center. Subsequently, you can not help but worry about the full support that your boss previously committed to you.) Another approach to getting the same work done is a mechanical method that asks the workers to follow a predefined mechanism for counting marbles. As for 4 + 3, the rule might be as follows.

Figure 4. Minimum expertise for programmers

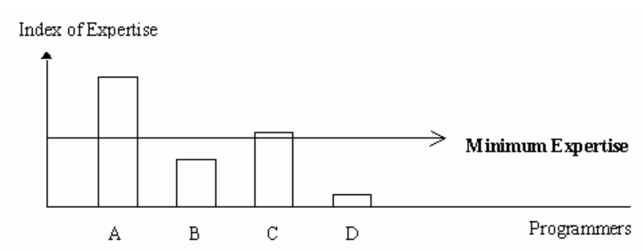
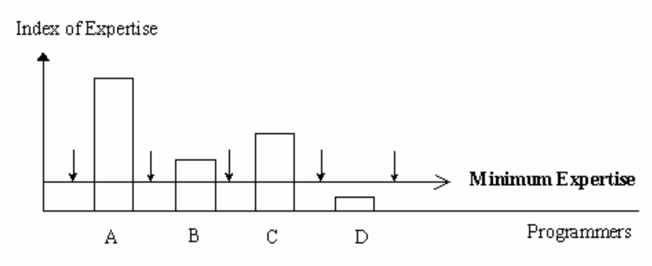


Figure 5. Lowering minimum expertise



Step 1: Count four marbles and put them aside.
Step 2: Count three marbles and put them aside as in Step 1.
Step 3: Mix them and count all the marbles.

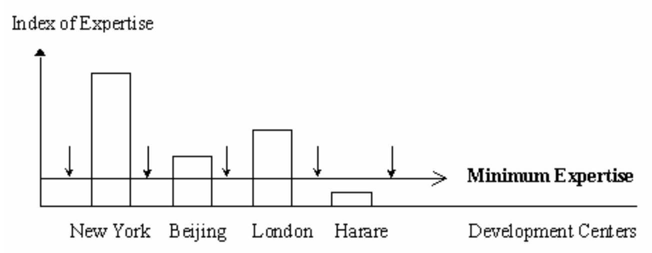
The minimum expertise is now counting marbles and putting counted marbles aside. Obviously, learning to add numbers requires the understanding of counting the numbers first. Putting counted marbles together means the workers follow the steps. In short, the minimum expertise is lowered to some degree, shown in Figure 5, although the approach might not appear to be intelligent. In fact, we found that management would be delighted with the approach as employees are able to start working and get the work done in a predictable way and time.

The problem of different technical skills among team members parallels the situation of different knowledge levels among many sites in the global software team, shown in Figure 6. For example, in Figure 6, programmers in Beijing may use some techniques in a program that developers in Harare would spend a whole day figuring out, whereas developers in London may follow up the work quickly and can continue to work.

FUTURE TRENDS

In the future, we expect the economic values gained from building those teams will become obvious when IT and our life are totally merged everywhere around the world. The demand for IT professionals can be predicted as an IT crisis if we now look at the rapid increment of Web sites. As a consequence, we can speculate that the number of job vacancies is huge. How many people are needed to develop and maintain them? There are two things we can do right away. One is resources reallocation, virtually (i.e., global software team) or physically. Another is to hire less qualified people who plagiarize the work of qualified professionals. One might argue this would be tricky. It is far from a long-term solution. Perhaps there is no long term in IT because it continues to advance almost every day. Remember, when you learn all the answers, IT has already changed all the questions.

Figure 6. Minimum expertise for development center



CONCLUSION

We addressed in depth some interesting problems in an inexperienced software team and in a global software team. Both require new managerial, technical, and social approaches at large. Knowing common relationships between two kinds of teams is the first step toward a framework for managing the team. Interested readers might consult the reference of our proposed methodology, called plagiarism-based programming (Lui & Chan, 2001, 2003).

REFERENCES

Amrine, H. T., Ritchey, J. A., Moodie, C. L., & Kmec, J. F. (1993). *Manufacturing organization and management.* Englewood Cliffs, NJ: Prentice Hall.

Carmel, E. (1999). *Global software teams: Collaborating across borders and time zones.* Upper Saddle River, NJ: Prentice Hall.

Information Technology of America. (2000). *Major new study finds enormous demand for IT workers: Research pinpoints hot jobs and skills needed, offers insights on employer preferred training approaches.* Retrieved from http://www.itaa.org/news/pr/PressRelease.cfm?ReleaseID=955379119

Karolak, D. W. (1998). *Global software development: Managing virtual teams and environments.* Los Alamitos, CA: IEEE Computer Society.

Lui, K. M., & Chan, K. C. C. (2000a). Managing design-coding for software development in China. *Proceedings of Software Engineering and Applications,* 2.9-2.19.

Lui, K. M., & Chan, K. C. C. (2000b). Managing inexperienced programmers by managing design-coding. *Pro-*

ceedings of European Software Process Improvement, 256-261.

Lui, K. M., & Chan, K. C. C. (2001). PbP: A programming paradigm for inexperienced software teams. *Proceedings of European Software Process Improvement,* 1.23-1.35.

Lui, K. M., & Chan, K. C. C. (2003). Inexperienced software team and global software team. In A. Gunasekaran, O. Khalil, & M. R. Syed (Eds.), *People, knowledge and information technology management: Human and social perspectives* (pp. 305-323). Hershey, PA: Idea Group Publishing.

McMahon, P. E. (2001). *Virtual project management: Software solutions for today and the future.* Boca Raton, FL: St. Lucie Press.

Morris, S. G. (1995). *Turnover among professionals: The role of person-culture fit and mentorin*g. Boulder, CO: University of Denver.

Otsuka, K. (2001). Book reviews: Growth and development from an evolutionary perspective. *Journal of Development Economics, 65*, 237-241.

Rischpater, R. (2001). *Palm enterprise applications: A Wiley tech brief.* New York:John Wiley.

Sanford, J. E. (2003). *Developing countries: Definitions, concepts and comparisons.* New York: Nova Science.

Sapaty, P. (1999). *Mobile processing in distributed and open environments.* New York: John Wiley & Sons.

Taylor, A. (2003). JD*BC: Database programming with J2EE.* Upper Saddle River, NJ: Prentice Hall.

KEY TERMS

Around-the-Clock Development: A software development style in which software teams that are geographically distributed make use of time zones to develop software.

Around-the-Sun Development: See "Around-the-Clock Development"

Efficiency of Around-the-Clock Development: An index used to indicate a ratio between the time required to follow up the previous work and time spent to continue the work.

Global Software Team: Software teams located in different countries collaborate as a single team for a clear objective project.

Inexperienced Software Team: Most members of a software team are graduates or inexperienced in disciplined software development.

Multisite Software Team: Software teams located in different cities and/or in different countries collaborate as a single team for a clear objective project.

Time-to-Market: A concept meaning to deliver a work product to the market the soonest or before the competitors do it to get more market share.

Infocratic Perspective on the Delivery of Personal Financial Services

Steve Gordon
Babson College, USA

Paul Mulligan
Babson College, USA

INTRODUCTION

In recent years, two factors – an increasingly relaxed regulatory environment and the growth of the Internet — have changed the competitive landscape of the personal financial services industry, particularly with regard to the potentiality of horizontal integration. Had financial regulation eased 10 years ago, most companies would have had little choice but to pursue a strategy of acquisition and merger to attain the capacity for providing integrated financial services. However, the Internet creates opportunities to build organizations, real and virtual, based on the control of information rather than the ownership of assets. Here we explore the potential for control of information to affect organizational strategy in the delivery of personal financial services.

BACKGROUND

The response of the financial service industry to deregulation is predicated, in part, on the will of companies to respond to consumer expectations and preferences. Pundits in the industry believe that consumers will demand a full integration of personal financial services. They believe, for example, that the consumer of the not-too-distant future would like, at a single Web site, to pay her bills, check her account balance, add money to her smart-cash card, check her credit card balance, check the current value of securities in her portfolio, buy and sell stocks and bonds, buy insurance, and issue instructions to reallocate the investment mix of payroll-deducted deposits into her 401K plan.

The quality of the integrated service experience will depend not only on the quality of the user interface, but also on the completeness of information and access to all the user's financial resources and on the features provided. Ideally, every financial instrument owned by a customer will be internally liquid and appear to be centrally located and controlled, restricted only by the legal and financial parameters of the instrument.

INTEGRATION OF PERSONAL FINANCIAL SERVICES – THREE MODELS

Three pure models and hybrids of these models can characterize the strategies of companies that have begun to offer integrated personal financial services (IPFS). The first model, which we call the Unified IPFS, describes companies that provide all or most of the services a consumer might want within a single corporate structure. Today, most Unified IPFS companies exist as holding companies, and, while their services are not yet well integrated, they are attempting to achieve a more seamless service delivery system through expanded investments in information technology. Unified IPFS providers currently differ in their degree of integration and the extent to which they can provide a full range of financial services.

The second model, which we call Allied IPFS, describes companies that provide diverse services through inter-organizational alliances. These companies focus on one primary area, such as banking, but provide a broad range of services through alliances with other companies. An example of such a company is Sovereign Bank, which provides, for example, investment products through their partner Lantern Investment Services and annuity products through IFS Agencies, Inc.

The third category, which we call Portal IPFS, describes companies that provide no direct, transactional services of their own but act as portals through which consumers can manage all of their financial services. An example of such a company might be Quicken, with its relationships with Ameritrade for brokerage services and Firstib.com for banking services. Although Quicken does not currently provide the seamless integration that we expect from the Portal IPFS firm of the future, it clearly hints at what this future might be.

Hybrid IPFS models are possible. For example, a company that appears to be a Unified IPFS provider might outsource some of its low volume products or services to companies that specialize in providing them. The result

Table 1. Customer perceptions

	Unified	Allied	Portal
Customer value proposition	One-stop shopping Lifetime relationships	Best in class Flexibility	One-stop viewing Unlimited providers
Perceived competitive vulnerability	Lost focus Exposure to niche players	Complex customer service Coordination/run costs Revenue sharing	Customer service Brand Ambiguous revenue model
Sources of customer resistance	Lack of diversity (performance) Perceived eggs in one basket	Customer servicing Limited choice in service providers	Security Privacy Permission granting

Table 2. Organizational components

	Unified	Allied	Portal
Anticipated competitive advantage	Cross-selling Brand development Integration options Full revenue capture	Internal focus Provider diversity Reduced start-up costs	Flexibility Lowest start-up costs Integration of non-financial data
Competitive Focus	Share of wallet	Share of customer	Share of information
Core Competencies	Intra-organizational integration	Inter-organizational integration	Cross-industry integration
IT Focus	Information management	Information transfer and integration	Screen scraping and presentation
Service focus	Service bundling	Service matching	Service aggregation
Examples	Schwab; Fleet; Citigroup	Sovereign Bank; Trustmark Bank	Quicken; InsurBank; Yahoo

would be a company that is dominantly Unified, but selectively Allied. A portal could also provide its own banking or brokerage services, creating a Portal/Allied or even a Portal/Unified hybrid.

Table 1 summarizes from the consumer's perspective the relative advantages, disadvantages, and risks of dealing with businesses pursuing each of the three IPFS strategies. Table 2 compares the strategies from the supplier's perspective as to competitive advantage, competitive focus, required core competencies, information technology focus, and service focus. It also provides some examples of organizations that are pursuing each strategy. The remainder of this section elaborates on each strategy.

Unified IPFS

A distinct advantage of the Unified approach is that a single corporate entity captures all revenues and maximizes its "share of wallet" with the customer. A unified company can increase the market share of each of the services it provides by cross selling. Such a company aims to "own" a customer from "cradle to grave," providing banking for the young, investment and retirement planning services for the middle-aged wealth accumulator, and trust management and reverse mortgage services for the elderly wealth distributor. The opportunity to establish brand recognition is also a benefit for a Unified IPFS provider.

The major disadvantage of the Unified IPFS strategy is the potential for a loss of corporate focus. The current diversity in investment products and continued advances in information technology make it difficult to be best in class for all products and services. The Unified providers compete directly with each other but also compete with more focused allied and portal players who, by incorporating product specialists into their networks, seek to provide premier service in a single product or market. From the consumer's perspective, the limitations in service and product options and choices and the thought of putting all one's financial "eggs" in a single basket can be problematic.

Allied IPFS

The primary benefits of the Allied approach are internal focus and external provision of choice to consumers. The enhanced focus gained by concentrating on a limited product-market offering eliminates resource deployment conflicts that may arise in more diverse Unified organizations. The firm can acquire, develop, and deploy human, capital, and technological resources in the development of a narrower set of competencies that are specific to their core business.

Proponents of the Allied model often boast of its apparent flexibility. Stronger partners can simply replace poor performing alliance members. However, this benefit is more illusory than real. Removing a partner for the alliance is extremely difficult unless all consumers opt to shift to the new alliance member. Adding a new member to provide alternatives to weak partners may violate contractual agreements or create confusion for customers.

The Allied model's major weakness for consumers, relative to the Unified strategy, is the challenge of providing a seamless integrative experience for the customer. Unified providers may find crossing internal *functional* boundaries difficult but the Allied firms face greater obstacles when crossing *organizational* boundaries. The coordination costs inherent in these relationships may impact financial performance by either shrinking margins, if prices are fixed, or compromising competitive position if the alliance raises prices to preserve margins. The goal for alliance members is to leverage focus such that the cost savings from internal operating efficiencies and excess returns earned by product specialists outweigh the added coordination costs.

Portal IPFS

The advantage of the Portal IPFS provider is that the choice of provider rests with the customer, as opposed to the Allied model in which the dominant provider preselects alliance partners. Furthermore, the portal model allows for multiple providers for the same product-service, which maximizes consumer freedom of choice. Portal IPFS providers also have the capability to incorporate non-financial services, such as frequent flyer accounts, news, and e-mail.

The major challenge for Portal providers is to placate consumer anxiety regarding privacy and security. The account consolidation benefits delivered by the Portal model also raise fears about identity theft. Customer servicing and accountability are also significant issues for this model. Today's Portal providers are not capable of providing customers service for such things as account registration changes, address changes and problem resolution. The technical standards imposed by the Internet enable this model but further establishment and maintenance of business process standards will be necessary to facilitate delivery on the Portal concept's full potential.

The Portal model is the newest of the three and currently services the smallest customer base. However, adoption rates continue to accelerate (Torris, Gross & Smit, 2001) as service levels improve and privacy concerns dissipate. The true power of the Portal will be realized when providers fully integrate account processing and augment transactional services with comprehensive add-on services.

THE COMING INFOCRACY

The viability of the organizational structures described previously depends as much on who controls the information central to the delivery of personal financial services as on who controls the other assets of the firms involved. Following Clawson, we use the term "infocracy" to describe a form of organization in which the basis of power is information (Clawson, 1999, 2000).

Clawson observes that the industrial revolution generated a shift in the dominant management paradigm from aristocracy to bureaucracy. He suggests that the information revolution is spawning a similar shift from bureaucracy to infocracy (see Table 3). In an aristocracy, with power derived from gender and lineage, decisions are made by those entitled to do so by birth. In a bureaucracy, decision-making authority is vested in those at the highest organizational levels; these office holders are assumed to know the most and to be best able to effect wise decisions. In an infocracy, decisions are made by those best equipped to interpret the relevant data, rather than those empowered to make decisions by dint of their positions.

The transition from bureaucracy to infocracy could have significant implications for how organizations oper-

Table 3. Clauson's management paradigms

	Aristocracy	Bureaucracy	Infocracy
Period	Pre Industrial Revolution (18th century and before)	Post Industrial Revolution (19^{th} and 20^{th} centuries)	Information Age (Now and future)
Power Source	Gender and lineage	Office and gender	Node Interpreter*
Assumption	"Father Knows Best"	"Boss Knows Best"	"Node Interpreter* Knows Best"

** Node interpreter: Person who has access to relevant and current data and the education and experience necessary to interpret and act upon them.*

ate. For example, Clawson hypothesizes that the most effective leadership styles will be those relying on information-based persuasion rather than command and control behavior. A large body of research has demonstrated that advances in information technology often affect the operation and structure of organizations (Bartezzaghi, Ciborra, DeMaio, Maggiolini & Romano, 1981; Carr, 2004; Dibrell & Miller, 2002; Drucker, 1988; Huber, 1990; Malone & Rockart, 1993; Wang, 1997; Whisler, 1970).

Clawson suggests that the transition from bureaucracy to infocracy will proceed faster than the move from aristocracy to bureaucracy. Bureaucracies, while designed to control information, are largely structured around the physical assets of the underlying organization. An infocracy has no comparable physical restrictions, which may explain the faster rate of transition. We believe that the information-intensive nature of financial services and lessening importance of associated physical assets positions this industry to be an early adopter of infocratic principles.

Yet there remains considerable research that affirms the value of bureaucracy in providing direction, structure, stability, and control (Adler, 1999), even in turbulent conditions (Perrow, 1970; Schellenberg & Miller, 1998). Despite predictions to the contrary, identifying any reduction in the bureaucratic structure of today's industry has proved to be devilishly difficult (Collins, 1997).

The internal adoption of and transition to infocratic organizational structures and processes does not ensure market acceptance. Many consumers accept, and perhaps derive assurance from, bureaucratic structure and policy. The movement to infocracy-supported channels and services will require firm investment in new skills for customer-facing employees and the acceptance of new rules of engagement for consumers. Service providers will require training, perhaps even licensing, in a broader set of products. Consumer education must address comfort and trust levels necessary to support increasingly disintermediated single-point-of-contact service encounters. Well-trained, efficient representatives will be needed to accelerate the development of consumer comfort and confidence.

FUTURE TRENDS

It is reasonable to expect a company's integration strategy to affect its ability to operate as an infocracy. The logical hypothesis is that a company's integration strategy should align with its strategy for control. A Unified IPFS, for example, might be expected to be bureaucratic, perhaps having a divisional structure, probably based on product rather than functional units, with responsibility and power delegated to those in charge of each division. Transactional and management information systems could be tightly integrated by the coordinated purchasing of enterprise systems. Although some matrix structures might be necessary to motivate and control cross-selling and joint product development, the overall control structure is likely to be highly bureaucratic.

An Allied IPFS would seem to require more dynamic control to respond to and nurture the inter-organizational fluidity inherent in its strategy. While each of the individual partners in the alliance might be somewhat bureaucratic, the governance of the alliance as a whole must have power both within and across organizational boundaries and must be sufficiently democratic to satisfy its respective partners. One might expect an Allied IPFS to exhibit a semi-bureaucratic or adhocratic structure, with distributed decision-making, some degree of democracy, and less control than one would find in a bureaucracy. Integration of transactional and management support systems would rely on adherence of all parties to emerging standards for XML and/or Web services, such as XBRL, IFX, and FpML.

A Portal IPFS, built solely on the flow of data with little to no control over its customers, would seem to require an infocracy if its internal structure, leadership, and decision making parallels the face it presents to the consumer. Transactional systems would require industry-wide adoption of standards pertaining to the exchange of data and the provision of secure e-commerce. Management systems would rely heavily on the widespread use of Web services to expose distributed data to the Portal entity.

An alternative hypothesis is that a company's integration strategy should oppose its strategy for information control. For example, one might hypothesize that a Portal IPFS would demand a bureaucratic structure to counterbalance the fluidity of its customer/supplier relationships by tightly controlling procedures and management oversight. Decision making would be based solely on the information that it collects as it executes Portal transactions, not on information supplied by its partners. Similarly, a Unified IPFS might need to operate as an infocracy to become nimble enough to compete with a Portal IPFS. This would imply, for example, that business units responsible for different types of financial transaction retain control over the information and the information systems that support their function. Rather than tightly integrated enterprise systems, these firms would rely on middleware and EAI products to provide an integrated view to the customer.

CONCLUSION

Regulatory change and the growth in the acceptance and use of the Internet have provided an environment in which financial service companies have the opportunity to pursue different strategic models for integrating the delivery of personal financial services. Three such models – Unified, Allied, and Portal – have been identified, and hybrids of these models have also been shown to exist. The financial services industry is identified as an ideal venue in which to study the coming infocracy – the movement from decision making based on official position and reporting structure to decision making based on information – because the different paradigms of service delivery provide different degrees of organizational control and have different needs for decision-making speed and flexibility.

REFERENCES

Adler, P.S. (1999). Building better bureaucracies. *Academy of Management Executive, 13*(4), 36-49.

Bartezzaghi, E., Ciborra, C., DeMaio, A., Maggiolini, P., & Romano, P. (1981). Computers, management and organization: Reflections on a pilot study. *Information & Management, 4*(5), 239-258.

Carr, N.G. (2004). In praise of walls. *MIT Sloan Management Review, 45*(3), 10-13.

Clawson, J.G. (1999). *Level three leadership: Getting below the surface*. Upper Saddle River, NJ: Prentice Hall.

Clawson, J.G. (2000, May/June). The new infocracies. *Ivey Business Journal,* 76-82.

Collins, D. (1997). Knowledge work or working knowledge? Ambiguity and confusion in the analysis of the "knowledge age". *Employee Relations, 19*(1), 38-50.

Dibrell, C.C., & Miller, T.R. (2002). Organization design: The continuing influence of information technology. *Management Decision, 40*(5/6), 620-627.

Drucker, P.F. (1988, January/February). The coming of the new organization. *Harvard Business Review*, 45-53.

Huber, G.P. (1990). A theory of the effect of advanced information technologies on organization design, intelligence, and decision making. *Academy of Management Review, 15*(1), 47-71.

Malone, T.W., & Rockart, J.F. (1993). How will information technology reshape organizations?: Computers as coordination Technology. In S.P. Bradley, J.A. Hausman & R.L. Nolan (Eds.), *Globalization, technology, and competition: The fusion of computers and telecommunications in the 1990s* (pp. 37-57). Boston: Harvard Business School Press.

Perrow, C. (1970). *Organizational analysis: A sociological view*. Belmont, CA: Wadsworth.

Schellenberg, K., & Miller, G.A. (1998). Turbulence and bureaucracies. *The Journal of Applied Behavioral Science, 34*(2), 202-221.

Torris, T., Gross, B., & Smit, M.K. (2001). *Exploiting finance portals*. Amsterdam: Forrester Research, Inc.

Wang, S. (1997). Impact of information technology on organizations. *Human Systems Management*, *16*(2), 83-90.

Whisler, T.L. (1970). *The impact of computers on organizations*. New York: Praeger.

KEY TERMS

Allied Model of IPFS: A model that describes the provision of integrated personal financial services through inter-organizational alliances.

Horizontal Integration: Horizontal integration in information management systems traditionally addresses information sharing across systems, which often means across functions or departments. In the context of organizational structure, horizontal integration typically refers to cross-business unit integration in pursuit of scope economies.

Hybrid Model of IPFS: The hybrid model is not, in the context of IPFS, simply a model that incorporates elements of two or more organizational forms. A hybrid IPFS model encompasses a more dynamic entity. Firms that adopt a hybrid IPFS structure typically maintain a general affinity towards a dominant form (e.g., Unified), but they temporarily create contrasting organizational relationships (e.g., Allied or Portal) that do not "fit" with the current organizational design. Much like evolution, the transition between dominant forms occurs over an extended period of time and as the result of numerous "hybridizations".

Infocracy: A form of organization in which information provides the underpinning of structure and the primary basis of individual power. In a more typical bureaucracy, hierarchical organizational structure dominates and is a primary driver of individual power. Information flows are then designed to support the structure and reinforce the power relationships. In an infocracy, information flows from point of origin to point of use, without regard to structure or pre-determined power relationships. As such, the organizational structure evolves to support required information flows.

Integrated Personal Financial Services (IPFS): The seamless integration of previously independent financial products and services. Prior independence of these financial products was a function of organizational, regulatory and/or technological constraints imposed upon the providers. For example, in a true IPFS relationship, funds would flow seamlessly between insurance, banking, investment and tax accounts.

Portal Model of IPFS: A model describing companies that provide no direct, transactional services of their own but act as portals through which consumers can manage all of their financial services.

Share of Wallet: The financial services industry defines "share of wallet" as the percentage of a customer's investable assets that are held by the provider. IPFS enables an organization to maximize their share of customer wallet by expanding the breadth of product and service offerings available.

Unified Model of IPFS: A model in which all or most of the financial services a consumer might want are provided within a single corporate structure.

Information and Communication Technology Tools for Competitive Intelligence

Dirk Vriens
Radboud University of Nijmegen, The Netherlands

INTRODUCTION

Competitive intelligence (CI) can be described as collecting and processing information about the environment for strategic purposes (cf. Kahaner, 1997). To formulate a strategy, an organization needs to collect and process information about its environment—about, for instance, competitors, customers, suppliers, governments, technological trends, or ecological developments. Collecting and processing environmental information has, of course, always been important. However, because of the increasing complexity and dynamics of the environment, the pressure to produce relevant, timely, "actionable" intelligence increases as well. At the same time, the amount of available data about the environment also increases.

To deal with this problem, it is necessary to structure intelligence activities, and many organizations use information and communications technology (ICT) to this end. They use, for instance, different types of Internet or intranet applications, so-called data warehouses, groupware applications, or applications specifically tailored to the organizations' intelligence needs. The number of possible ICT tools for CI is large, and an important question for organizations is which tools they should select and implement for their CI activities. Organizations face difficulties in selecting, implementing, and using these tools. Many ICT tools, for instance, lead to an information overload—to large collections of irrelevant data—or are expensive applications that only marginally contribute to the production of intelligence.

To select proper ICT tools for CI, an organization needs to understand the role of ICT for CI. This paper intends to address this understanding. In particular, it will present a short overview of the available tools and it will discuss a procedure for selecting appropriate ICT tools.

BACKGROUND

To understand the role of ICT for CI and to select proper ICT tools, the CI process needs to be clarified. To describe this process, several authors (cf. Fuld & Company, 2002; Gilad & Gilad, 1988; Herring, 1999; Kahaner, 1997) propose to use the intelligence cycle, consisting of four stages.

1. **Direction.** In this stage organizations determine their strategic information requirements; that is, they determine about what environmental aspects data should be collected. These aspects are also known as competitive intelligence needs (Fleisher, 2001) or key intelligence topics (Kahaner, 1997).
2. **Collection.** In this stage, the required data are collected. To this end, one needs to (a) determine what sources are available, and (b) access these sources and retrieve data from them.
3. **Analysis.** This stage focuses on interpreting data from a strategic point of view to determine their strategic relevance (if a piece of data is strategically relevant, the term intelligence is used; cf. Kahaner, 1997; Vriens, 2004). For this analysis, a model of what is relevant for the organization should be available. Often, tools such as Strengths, Weaknesses, Opportunities and Threats (SWOT) analysis, scenario analysis, war gaming, or competitor profiling are used to arrive at such a model (cf. Cook & Cook, 2000).
4. **Dissemination.** Here, the intelligence is made available for strategic decision making. That is, the intelligence should be presented clearly and distributed to relevant decision makers who use it to evaluate current strategic options and to generate, compare, select, and implement new ones.

Using this cycle as a background, many authors acknowledge the possible and actual use of ICT applications for CI activities (see, for instance, Chen, Chau, & Zeng, 2002; Cook & Cook, 2000; Fleisher & Blenkhorn, 2001; Fuld & Company, 2002). It is also a background for discussing the shortcomings of ICT applications for CI (e.g., Cook & Cook; Fuld & Company). In line with these authors, we will use this cycle of intelligence activities as a framework for reviewing, classifying, and selecting ICT applications, as will become apparent in the next sections.

A CLASSIFICATION OF ICT TOOLS FOR CI

ICT tools for CI can be classified according to two "dimensions": (a) their contribution to one or more stages of the intelligence cycle and (b) the specificity of the tool—that is, a tool can be either a general ICT tool used for intelligence activities (like a groupware application, used for direction activities) or a tool specifically tailored to one or more intelligence activities. Below, we first discuss the Internet as a "general" ICT tool for all CI activities. Next, we pay attention to other ICT tools, both general and specific. Finally, we discuss so-called business intelligence applications as a specific set of ICT applications.

The Internet as a Tool for CI

CI practitioners rely heavily on the use of the Internet for their intelligence activities. Some authors report the use of the Internet for supporting the direction, analysis, and dissemination stages, for example, Web Enabled Technology (WET) applications enhancing communication and collaboration regarding results of this stages (e.g., Teo & Choo, 2001). However, most attention has been paid to the use of the Internet for collection activities. Among the reported uses of Internet applications for collection activities are the following.

- **Search engines**
  The CI literature discusses many types of search engines and their application for CI (cf. Chen et al., 2002; Cook & Cook, 2000).
- **Tools for outsourcing collection activities**
  A part of the collection activities can be outsourced to some (automated) service or tool offered via the Web. Particularly popular is the use of commercial online databases (Kahaner, 1997; Chen et al., 2002). Another example is the use of Web robots or agents that "automatically traverse the hyperlink structure of the WWW [World Wide Web] to locate and retrieve information" (Tan & Kumar, 2002, p. 9).
- **Tools for text analysis**
  Tools supporting the collection of data in (large) Internet-based text files are used.
- **Tools for monitoring changes on the Web**
- **Tools for collecting data about the electronic behavior of Internet users**. These tools work, for instance, by identifying users' navigational patterns (cf. Tan & Kumar, 2002).
- **Internet tools for collaboration in collection activities.** Internet applications can be used to facilitate the collaboration in collection networks (networks of people performing collection activities).

General and Specific ICT Tools for CI

Besides using general Internet applications for CI purposes, other general ICT applications can be used as well. Examples are tools supporting the following.

- **Supporting the process of identifying strategic information needs** such as different types of groupware or software supporting group model building (e.g., Vennix, 1996).
- **Supporting specific methods used in analysis**, for instance, system-dynamics software enabling CI professionals to run simulations with certain data, and thus helps to establish their relevance.
- **Supporting (management of and collaboration in) the process of analysis**. Specific groupware applications may serve this purpose.
- **Supporting the dissemination of intelligence**—for example, applications supporting (a) the presentation of the intelligence in a suitable format and/or (b) the transmission of reports throughout the organization

There is also a class of ICT applications specifically designed for one or more CI activities. Fuld & Company (2002) analyzed and evaluated a number of such dedicated CI software packages. They concluded with the following.

(1) The "CI software cannot drive the CI process" (p. 2), but it can help in collecting data, in reporting and communicating intelligence, and in supporting the work flow and collaboration.
(2) No application can deal with all the intelligence stages adequately.
(3) No application can "truly conduct qualitative analysis" (p. 10), but some tools seem promising in assisting CI analysts to see novel linkages.

Business Intelligence Applications

For some time, the terms competitive intelligence and business intelligence (BI) were used as synonyms. Nowadays, however, the term business intelligence is often used to indicate a specific set of ICT tools. These BI tools refer to ICT tools enabling (top) management to produce overviews of and analyze relevant organizational data needed for their (strategic) decision making. BI tools typically include data warehouses and tools for extracting and presenting information from them (cf. Cook & Cook, 2000; Zanasi, 1998). In essence, a data warehouse is a large database in which data from many different databases (e.g., transactional, financial, or personnel databases) are copied and stored in such a way that they are readily accessible. To access, view, and analyze these data, specific tools are developed (e.g., data-mining tools for finding relations between classes of data). Using data warehouses and their associated tools to produce intelligence is not unproblematic. Examples of reported problems are the following (cf. Cook & Cook; Fuld & Company, 2002).

1. Most data warehouses only cover internal data (generated in the transactions of the organization; Fuld & Company, 2002). This is a major drawback when treating BI tools as CI tools. However, once data warehouses incorporate relevant *external* data, they may become valuable CI tools as well.
2. The focus of BI software is primarily on quantitative analysis while CI relies heavily on qualitative data.
3. Data warehouses cause cost and implementation issues: They require large budgets and much implementation time and effort.
4. Cook and Cook (2000) assert that one cannot expect BI software to produce intelligence. Data resulting from analyzing data in a warehouse should still "be analyzed and directly applied to a specific problem to become intelligence" (p.165).

HOW TO SELECT ICT FOR CI

To select appropriate ICT tools for CI, given specific organizational CI needs, organizations can use three classes of criteria: (a) criteria regarding the contribution of ICT applications to one or more stages of the intelligence cycle, (b) criteria regarding the CI infrastructure, and (c) criteria pertaining to costs. These three classes correspond to the criteria used in the selection of ICT applications in general (cf. Parker, Benson, & Trainor, 1988).

Criteria Regarding the Contribution to One or More Stages of the Intelligence Cycle

A (candidate) ICT application should contribute to one or more stages of the intelligence cycle. Criteria to evaluate the contribution of an ICT application for one or more stages may refer to its appropriateness to deliver the desired products for the stage, and to whether it fits the process leading to these products (cf. Fuld & Company, 2002). For example, collection applications should be evaluated regarding their appropriateness to collect the desired data (e.g., patent data). Applications should also match process aspects. For example, if particular methods are used to analyze data, such as a SWOT analysis, candidate applications can be evaluated regarding their appropriateness in supporting these particular methods.

Criteria Regarding the Relation with the CI Infrastructure

The CI infrastructure consists of three parts: (a) the technology supporting the CI activities (ICT hardware, software, and telecommunications technology), (b) the human resources needed to carry out the CI activities, and (c) the organizational structure of these activities, that is, how CI tasks and responsibilities are divided and allocated. Each possible ICT application should match all three parts—from these three parts, specific criteria can be derived.

The criteria regarding the technology part of the infrastructure focus on the question of whether the candidate application fits the current technology. This fit depends on the current set of ICT applications used to support the CI activities. Does the application fit into this set? Does it deliver more functionalities than this set? Is an easy link between the applications in this set (if desirable) possible? Other questions for judging the fit to the technology have to do with the implementation and maintenance of the application: Does the current hard- and software permit the implementation of the application, or does it require large changes? Is the application reliable? Yet another question for judging the fit to the technology has to do with the contribution of an application to the improvement of the infrastructure; for example, through a particular application, obsolete applications are replaced (or may be replaced more easily), or through a particular application, other state-of-the-art applications can be implemented and used more easily. A data warehouse, for instance, can be seen as a contribution to the current infrastructure because it enables all kinds of tools for visualizing and analyzing (internal) data.

The second set of criteria reflects the fit of the application to the human resources, that is, whether it fits the existing skills, knowledge, and attitudes of those who are carrying out CI activities. Important questions are, for instance, whether the required knowledge and skills are acquired easily, or whether the CI staff is motivated to integrate the ICT application into their daily routines.

The third set of infrastructural criteria reflects the fit of the application to the current definition and allocation of CI tasks and responsibilities (see Gilad & Gilad, 1988, for several ways of defining and allocating them). It does not make sense to install groupware for the direction stage if direction is not seen as a group process. The same holds for using ICT tools structuring the collection and analysis activities if the whole CI process has a highly informal nature. In some cases, the CI task structure may be designed poorly and ICT can act as a leverage to change it. An ICT tool can be valued because of its contribution to the improvement of the task structure. For instance, a groupware application may be valued for its contribution to structure the direction stage as a group process.

Criteria Pertaining to Costs

These criteria refer to the costs of the application itself, its implementation (e.g., project costs, training, etc.), and its maintenance. These costs may be calculated by different methods (cf. Parker et al., 1988).

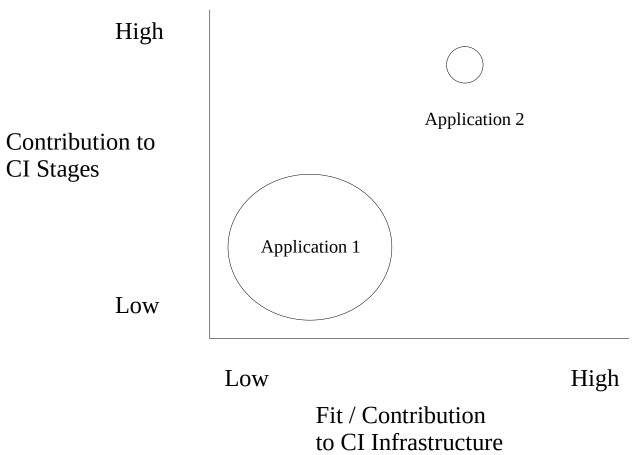
Selecting ICT for CI Using the Three Types of Criteria

To evaluate the appropriateness of an application for CI in a particular organization, the application should be "scored" regarding all three classes of criteria. To this end, the individual criterion in a class should each be identified, valued, and integrated into an overall score for the class. We will not treat all these substeps in this section; rather, we will indicate how these overall scores can be used to select of ICT for CI.

The overall scores of each class express (a) the contribution of a particular application to one or more of the intelligence stages, (b) the fit of the contribution to the CI infrastructure, and (c) the costs related to an application. These scores can be plotted in a graph (see Figure 1).

In this figure, the scores of several applications are depicted (the size of the circles reflects the costs related to the application). For instance, Application 1 may be a data warehouse (with an internal focus). This application

Figure 1. Classification of ICT applications for CI using three dimensions (see text)



may be very expensive, contributes only partly to the intelligence stages (it contributes to the collection stage, but due to its internal focus, its contribution to the CI stages is low). The data warehouse may contribute to the general infrastructure, but, in our view, contributes only partly to the CI infrastructure. In contrast, a data warehouse with explicit external linkages would still be very expensive, but may score higher on both other dimensions. Application 2 may be a groupware application. These applications are moderately expensive, may contribute to the direction stage, and fit the infrastructure in several ways.

An organization may treat several ICT tools to support their CI activities in this way and eventually select some.

FUTURE TRENDS

Many organizations recognize the importance of supporting CI activities by means of ICT. Although many ICT applications are already available, the possibilities of using ICT for CI will increase rapidly. Some of the trends that may be acknowledged are the following.

- A convergence of BI and CI applications (e.g., data warehouses and associated software also tied to external and qualitative data; cf. Li, 1999)
- Using ICT for qualitative data may increase (e.g., Chen et al. 2002)
- Using the Internet for more than just collection activities (e.g., for collaboration and dissemination purposes; cf. Teo & Choo, 2001)
- Improvement of Internet applications for collection (More efficient and effective collection applications will continue to emerge.)
- Implementing CI applications can be seen as a process by means of which the CI process and infrastructure can be reanalyzed.
- Improvement of analysis applications (cf. Fuld & Company, 2002)

CONCLUSION

To select and use proper ICT tools for supporting the CI process, organizations should know (a) what the CI process is, (b) what the role of ICT (tools) in this process can be, and (c) how to evaluate the role of ICT (tools) for their own CI process. In this chapter, we discussed the role of ICT tools in the CI process. We classified several types of ICT tools relevant for supporting CI activities: the Internet, general and specific applications for supporting CI activities, and business intelligence applications. In the last part of this chapter, we discussed three classes of criteria organizations can use in evaluating and selecting ICT tools for their CI processes.

The possibilities of ICT to support CI activities will probably only increase. To make full use of its potential and to avoid many of the pitfalls associated with using ICT for CI, organizations should be careful in selecting and implementing ICT applications for CI purposes. They should know the possibilities of ICT to deliver internal and external data and its capacity for supporting CI activities, and they should treat ICT as a part of the whole infrastructure.

REFERENCES

Chen, H., Chau, M., & Zeng, D. (2002). CI-spider: A tool for competitive intelligence on the Web. *Decision Support Systems*, *34*, 1-17.

Cook, M., & Cook, C. (2000). *Competitive intelligence*. London: Kogan Page.

Fleisher, C. G. (2001). An introduction to the management and practice of competitive intelligence (CI). In C. G. Fleisher & D. L. Blenkhorn (Eds.), *Managing frontiers in competitive intelligence* (pp. 3-18). Westport, CT: Quorum.

Fleisher, C. G., & Blenkhorn, D. L. (Eds.). (2001). *Managing frontiers in competitive intelligence*. Westport, CT: Quorum.

Fuld & Company. (2002). *Intelligence software report 2002*. Retrieved March 2003 from *http://www.fuld.com*

Gilad, B., & Gilad, T. (1988). *The business intelligence system*. New York: Amacon.

Herring, J. P. (1999). Key intelligence topics: A process to identify and define intelligence needs. *Competitive Intelligence Review*, *10*(2), 4-14.

Kahaner, L. (1997). *Competitive intelligence*. New York: Touchstone.

Li, C. (1999, Summer). ERP packages: What's next? *Information Systems Management*, *16* (3), 31-35.

Mahony, T. M. (1998). Data warehousing and CI: An evaluation. *Competitive Intelligence Review*, *9*(1), 38-43.

Parker, M. M., Benson, R. J., & Trainor, H. E. (1988). *Information economics: Linking business performance to information technology*. Englewood Cliffs, NJ: Prentice Hall.

Tan, P., & Kumar, V. (2002). Discovery of Web robot sessions based on their navigational patterns. *Data Mining and Knowledge Discovery*, *6*, 9-35.

Teo, T. S. H., & Choo, W. Y. (2001). Assessing the impact of using the Internet for competitive intelligence. *Information & Management*, *39*, 67-83.

Vennix, J. A. M. (1996). *Group model building*. Chichester, England: Wiley.

Vriens, D. (Ed.). (2004). *Information and communication technology for competitive intelligence*. Hershey, PA: IRM Press.

Vriens, D., & Philips, E. A. (1999). Business intelligence als informatievoorziening voor de strategievorming. In E. A. Philips & D. Vriens (Eds.), *Business intelligence*, (pp.11-44). Deventer, Netherlands: Kluwer.

Zanasi, A. (1998). Competitive intelligence through data mining public sources. *Competitive Intelligence Review*, *9*(2), 44-54.

KEY TERMS

Business Intelligence: The term business intelligence is used in two ways: (a) as a synonym for competitive intelligence and (b) to indicate a specific set of ICT tools to support managerial decision making. This set of tools often consists of a data warehouse and the tools to store, retrieve, and present the information it contains (e.g., data-mining software).

Competitive Intelligence: In the literature, two definitions are used: a product definition and a process definition. In the product definition, competitive intelligence is defined as information about the environment, relevant for strategic purposes. The process definition highlights producing and processing this environmental information. Process definitions often refer to the intelligence cycle.

ICT (Information and communication technology)**:** Can be used to indicate the organization's technological infrastructure (comprising of all hardware, software, and telecommunications technology) and to indicate one or more specific collections of hardware, software, and telecommunications technology (i.e., one or more ICT applications).

ICT Selection: The process of selecting proper parts of the technological infrastructure and/or specific applications. In the context of CI, specific criteria are used: criteria regarding costs, appropriateness in supporting intelligence activities, and their fit or contribution to the CI infrastructure.

Intelligence Cycle: This is a cycle of four stages (collections of intelligence activities). The stages are direction (determining the strategic [external] information requirements), collection (determining, locating, accessing, and retrieving sources to obtain data about the environment), analysis (determining the strategic significance of the environmental data), and dissemination (of the intelligence to strategic decision makers).

Intelligence Infrastructure (CI Infrastructure): The intelligence infrastructure comprises all technological, human-resources, and organizational means needed to manage, support, and perform the intelligence activities. It consists of three subinfrastructures: the technological, the human-resources infrastructure, and the organizational infrastructure.

Information Laws

Andrew S. Targowski
Western Michigan University, USA

INTRODUCTION

Humankind progresses in proportion to its wisdom which has roots in practice, acquired skills, available data and information, concepts and knowledge. To be wise, humankind needs to be informed and knowledgeable; otherwise it will not survive its own failures. Progress in knowledge was painfully slow as long as the racial memory was transmitted only by oral tradition. With the invention of writing and books the process of knowledge discovery and dissemination has been accelerated. Today, computers and their networks speed up that process far beyond our imagination. In the 2000s the Information Wave significantly controls the Agricultural and Industrial Waves through millions of computers. IT supports decision-making based on knowledge-oriented systems such as data mining that, for example, discovers knowledge about customers, organizational dynamics, and so forth to achieve competitive advantage.

Information and knowledge become the strategic resource as engineering science was in the Industrial Wave. However, the discovery of human cognition potential must be guided by knowledge science, which just emerges. One of the signs of any science is its set of scientific data, universal rules, laws, and systems of rules and laws. Hence, this article offers the first attempt to develop main laws of information that should increase our awareness about the Information Wave, which is a new stage of civilization's dynamics that is taking place at the beginning of the third millennium. The article also provides the framework for the analysis of the human capital from the information perspective. This set of considerations reflects a new emerging approach that I call macro-information ecology.

BACKGROUND

Macro-Information Ecology

Macro-information ecology is based on the premise that the growth rate in the new information (knowledge) discovery is the key determinant of macroeconomic activities in the service-industrial-global economy (so called the new economy). This new emerging school of macroeconomics can be called knowledgism.

Macro-information ecology is the study of information (cognition) as a whole and it is concerned with *aggregates* across nations and markets. Macro-information ecology studies the behaviour of society and economy (nationally and globally) — wide measures, such as:

- the value of human capital,
- the potential efficiency of human capital,
- knowledge output,
- economy output driven by knowledge in a given period, and so forth.
- It also studies measures derived from many individual nations:
  - markets such as the price of human capital or
  - the total structure of employed workers by such categories as production workers, in-person service workers, and information workers.

To control national output with the development of the global economy, knowledgists stress the need to control the growth of new knowledge discovery. Given the "long and variable lags" of knowledge and information policies and the difficulty in forecasting future economic events (such as recession), knowledgists question the ability of industrial or service-oriented macroeconomics to implement the "correct" economic policy.

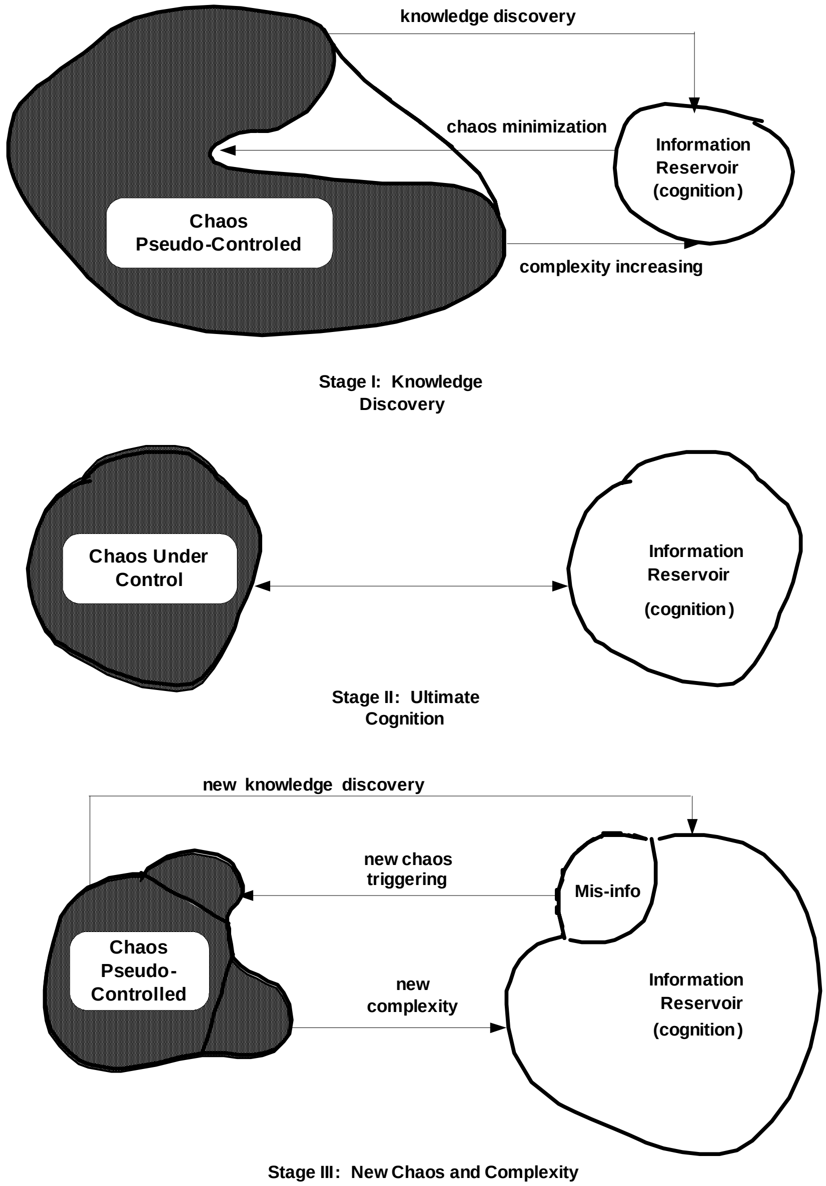
The knowledge approach suggests that direct government intervention within the economic system should be guided by the "predicted history of the futures". The knowledge policy is the key to this intervention; in this sense, the knowledge policy is closer to Keynesian interventionists than to "conservative" monetarists.

The supply and demand of information (knowledge) is the most basic model of information ecology (IE). However, prior to this model, we have to examine the stages of the information reservoir development. Figure 1 illustrates this process.

GENERAL INFORMATION LAWS

At the stage of knowledge discovery, the information reservoir (IR) minimizes or tries to "control" the chaos. Every increase in new information also increases a level of complexity of understanding. Based on the analysis of

Figure 1. The stages of information reservoir development



knowledge dynamics provided by Wojciechowski (1989), one can define the following laws of information:

- Law I: The complexity of the ecosystem (human, material, cognition, and nature) is growing accordingly with the level of the existing information reservoir.

The complexity is the state of a system whose components and relationships co-evolve through an enormous number of interconnections, creating dynamic structures either chaotic or orderly. The more information we have at our disposal, the more complex the ecosystem is perceived to be. The more we know, the less we understand. The founders of the Santa Fe Institute, which explores the new science of complexity, investigates such questions as why ancient ecosystems often remained stable for millions of years, only to vanish in a geological instant—and what such events have to do with the sudden collapse of Soviet Communism in the late 1980s.

- Law II: Information generates consequences, which it cannot foresee.

One of the forms of information is knowledge, for example, such as atomic physics. Atomic physics produced rules and techniques that allowed man to build the atomic bomb. The consequence was the tragedy that befell thousands of Japanese who lost their lives or, at the very least, their health in 1945. On the other hand, the Cold War, sustained by the balance of atomic weaponry, was practically bloodless. Should science stop research on atomic physics or gene engineering because consequences can get out of control? Or, being under control, can said research produce positive results, such as the Cold War, which eliminated another Bloody War?

- Law III: The precision and certainty of information is growing accordingly to the simplicity of the described object or inversely declining due to the complexity of the object.

Relatively simple material objects can be described by relatively simple information in natural science. On the other hand, complex social phenomenon requires complex description; this is sometimes contradictory if description is provided by more than one observer. For example, in the 1991 Persian Gulf War, there was a question among the Allied forces about whether to go to Baghdad and seek the surrender of the Iraqi military regime. Almost every observer of this war had his or her own answer (information) to this question.

- Law IV: The progress of the Information Wave generates relative ignorance and interdependence among people and globalizes humanity.

The advancement of mediated information requires information skills to access information infrastructure, systems, and services. People without this access are becoming more ignorant than those who can retrieve and apply required information anytime and anywhere in the synchronism of events. The information poor are becoming more dependent on the information rich; the latter are motivated to globally seek more useful information to become even richer. A college professor or a graduate who knows end-user computing has more chances to increase his or her material well-being than someone who does not know how to apply computers to gather and process important information and is ignorant about possible opportunities for him or her. Even a well-meaning businessperson who is ignorant about information technology may lose resources or at least not increase them if he/she does not know how to transform his/her business from *brick ’n mortar* to *brick ‘n click* format.

At Stage II — Ultimate cognition, the amount of information is equal to the amount of chaos. From humankind’s point of view, the equilibrium in macro-ecology never happens since the amount of time in which such equilibrium can be attained is infinite; in such disciplines as business management, perhaps, one can achieve short-term equilibrium. Therefore, the next law can be defined:

- Law V: The information reservoir has no saturation point.

Since the ecosystem is imperfect and still developing, the information about it has not become definite. What was right in the 19th century is revised in the 20th century, and what is right in the 20th century, perhaps, will be redefined in the 21st century, and so forth. Examples of Newtonian physics critique or post-modernism’s challenge of “scientific truth” provide data that prove this law in the 20th century. Each new discovery not only decreases the chaos; it also increases the confusion about the new directions and, *ipso facto*, requires more information to improve understanding.

Stage III — When the capacity of the information reservoir should exceed the capacity of chaos, new chaos and complexity are created by misinformation, which begins to penetrate the IR. It is only an assumption since, according to Law V, such a situation should not happen.

MACRO-INFORMATION ECOLOGY MODEL

The macro-ecology of the information equilibrium model (Figure 2) indicates that civilization, most of the time, operates in darkness. The mathematical model of the information reservoir is as follows:

Stage I: $I < E$

where: I = Information Reservoir capacity
E = Entropy, a measure of chaos
D = Darkness (or net entropy E-I)

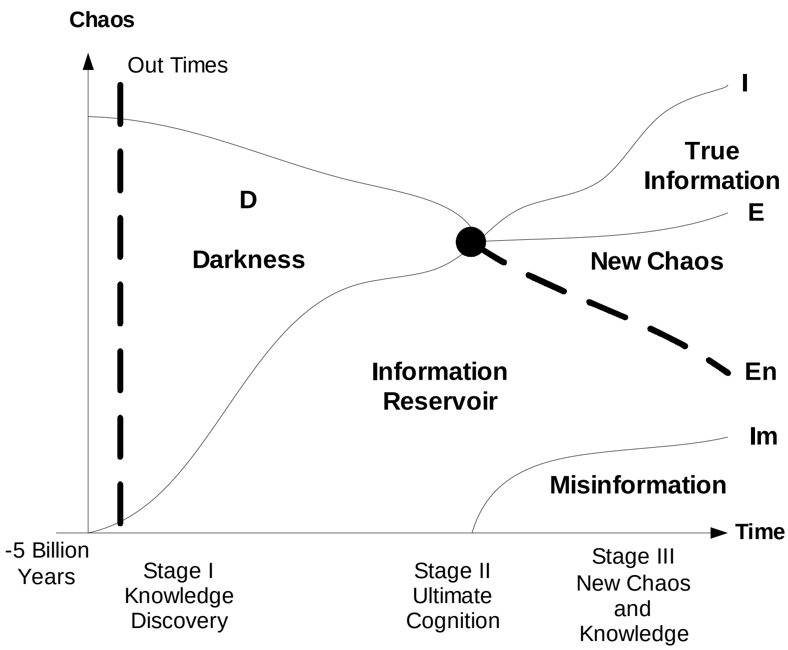
The macro-ecology goal is Dt ’! 0 and the task is to determine the elasticity of the increased entropy or information and how a user or organization responds to changes (+, -) caused either by the increase of information reservoir or by its “enemy” — entropy.

D (Et) = f (IMt, Ct) or D (It) = f (Et, Ct)

where: IM = Misinformation[1]
C = Complexity
t=time

The elasticity of information is a measure of the sensitivity or responsiveness of the information value de-

Figure 2. Information macro-ecology model



manded to create changes in price, revenue, unemployment, and in other factors of the Information Wave. Information elasticity will be one of the major indexes of the emerging Information Wave.

Macro-information ecology is also interested in the creation of human capital as a medium of knowledge generation and application. Particularly, a relation between human capital and economic development is a strategic inquiry of Information Ecology.

The mechanism of the material civilization in modern capitalistic theory was built on the rule that market growth triggers the specialization of human capital and the growth of its income as well as its level of living. In the Information Wave, the situation is different. The new motoric forces of economy can express the following law of the human capital.

INFORMATION LAW OF HUMAN CAPITAL

- Law VI: The human capital's growth in knowledge generates specialization and productivity and sustains the growth of income.

As Kevin Murphy [1] noticed that the old sequence in economic development in the Industrial Wave:

- (old) material sequence: market growth - specialization - more income transforms the Information Wave into a new sequence in economic development:
- (new) early information sequence: knowledge growth - specialization - more income

The new sequence is true as long as the specialization of human capital sustains the increase of productivity in the material sector or in the information sector. The necessary co-ordination of specialists, particularly those in the information sector, may consume the new "speed" of knowledge and not only contribute to the economic growth, but misguide it.

The most important question, however, is whether or not we should apply new knowledge to promote economic development by growth or whether we should just apply that knowledge's message, which says that zero growth is wiser and the only appropriate policy to achieve a sustainable society and economy.

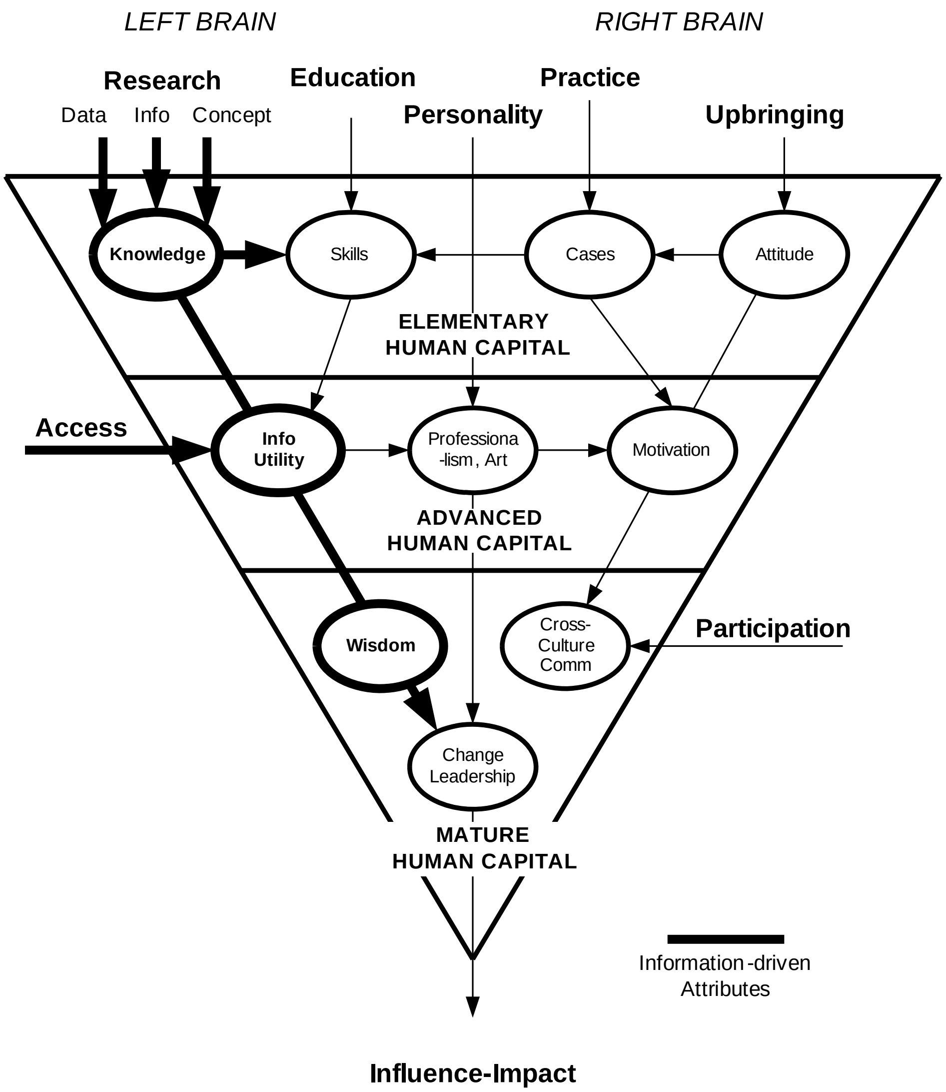
The uncontrolled development of the Information Wave may lead to another economic sequence:

- further information sequence: knowledge growth - specialization - collapse of economy (?)

This sequence produces an economy that is too specialized and productive, requiring a small work force without the means to create a demand for economic output. In current practice, robots and computers do not pay taxes.

Information ecology has to include a human dimension of the Information Wave in its inquiry. Better knowledge should provide a better level of living, not inspire self-destruction and limit progress to technology

Figure 3. The human capital in development stages



alone. Technology is not neutral; the new knowledge should define info-comunication tools of honourable and sustainable living. This is possible if we consider the Electronic Global Village as a tool of information and knowledge creation and distribution (bottom-up and top-down), and as a globally interconnected aware tribe.

The steered Information Wave should offer the following sequence of events:

- expected information sequence: knowledge growth - solutions - sustainable economy

If "human capital" becomes wise enough, this sequence should probably be implemented in the 21st century. Otherwise, population and ecological bombs (about 2050) will return us to the beginning stages of the history of humankind.

HUMAN CAPITAL DEVELOPMENT

Human capital in the 21st century will become the most important economic resource. This is a medium that generates and applies knowledge. Its architecture of "organs" is depicted in Figure 3.

Human capital develops in three stages:

- Stage I: Elementary Human Capital: Knowledge, Skills, Cases, Attitude

- Stage II: Advanced Human Capital: Info Utility Access, Professionalism and Artistry, Motivation
- Stage III: Mature Human Capital: Wisdom, Cross-Culture Communication, Change, Leadership

Only four attributes are information-driven: knowledge, skills, information utility access, and wisdom. This means that the development of human capital cannot be limited only to issues of information. The process of socialization plays a very important role in the estimation of human capital values; this process is culture-driven.

The measurement of human capital value can be done through the estimation of the value space of work force (macro-ecology) or through a given person (micro-ecology). Figure 4 illustrates the value space of human capital.

Each attribute (A) can be measured on the five-point scale. A sum total of all attribute points provides a value of given human capital (V). This is a comparative unit of human capital value. It can be applied in comparisons of economies, organizations, or persons. It can also be applied in the analysis of human capital efficiency potential.

FUTURE TRENDS

Along with the further development of the Information Wave, more laws of information should be defined and applied.

CONCLUSION

Macro-information ecology is just emerging, along with the development of Information Wave practice, and research should turn its attention into the application of the information laws and their further discovery and corrections in the analysis and design of values and tools of the Information Wave and civilization in general.

REFERENCES

Badaracco, J.L., Jr. (1991). *The knowledge link*. Boston: Harvard Business School Press.

Beach, L.R., Mitchell, T.R., Daeton, M.D., & Prothero. (1997). Information relevance, content, and source credibility in the revision of opinions. *Organizational Behavior and Human Performance*, *21*, 1-16.

Behm, D., & Peat, F.D. (1978). *Science, order and creativity*. New York: Bantam Books.

Bell, D. (1973). *The coming of the post-information society: A venture in social forecasting*. New York: Basic Books.

Blumenthal, A.L. (1977). *The process of cognition*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Drucker, P.F.M. (1988, January-February). The coming of the new organization. *Harvard Business Review, 88*(1), 45-53.

Drucker, P.F.M. (1993). *Post-capitalist society*. New York: HarperBusiness.

Ekecrantz, J. (1987). The sociological order of the new information society. In J.D. Slack & F. Fejes (Eds.), *The ideology of the information age*. Norwood, NJ: Ablex Publishing Corp.

Gore, A. (1991, January-February). Information superhighways: The next information revolution. *The Futurist*, 21-23.

Karraker, R. (1988, Spring). Highways of minds. *Whole Earth Review*, 4-15.

Laszlo, E. (1972). *Introduction to systems philosophy*. New York: Harper and Row.

McWhirter, B. (1993, March 29). Disposable workers of America. *Time*, 41-43.

Nowell, A., Perils, A., & Simon, H. (1987). What is computer science? *Science*, *157*, 1373-1374.

Parker, E. (1976, December). Social implications of computer/telecoms systems. *Telecommunications Policy*, *1*, 3-20.

Porat, M. (1977). *The information economy*. Washington, DC: US Office of Telecommunications.

Poster, M. (1990). *The mode of information*. Chicago: The University of Chicago Press.

Pricher, W. (1987). Tours through the back-country of imperfectly informed society. In J.D. Slack & F. Fejes (Eds.), *The ideology of the information age*. Norwood, NJ: Ablex Publishing Corp.

Sakaiya, T. (1991). *The knowledge-value revolution*. New York: Kodansha International.

Shannon, C.E. (1948). A mathematical theory of communication. *Bell Systems Tech Journal*, *3-4*.

Targowski, A. (1999). *Enterprise information infrastructure*. Hershey, PA: Idea Group Publishing.

Toffler, A. (1980). *The third wave*. New York: Bantam Books.

Wojciechowski, J. (1989, March). Progress of knowledge and right-left dichotomy: Are existing ideologies adequate? *Man & Development, XI*(1).

Van Doren, C. (1991). *A history of knowledge.* New York: Ballantine Books.

NOTE

[1] During his public lecture at Western Michigan University (13-12-1992).

KEY TERMS

Ecosystem: The entire ecological community composed of all living organisms interacting with the physical environment as one system.

Entropy: A measure of the degree of disorder or tendency toward the breakdown of any system. In physics this term is defined in the second law of thermodynamics, which states in part that "the entropy of the universe tends to a maximum".

Human Capital: Reflects a human potential in knowledge, skills, and attitude that can provide better solutions (meaning), efficiency (time) and effectiveness (money) of work.

Industrial Wave: A phase of civilization in which industries based on material processing and handling are dominant factors in the economy.

Information: A comparative unit of cognition that defines a change between the previous and present state of the natural, artificial, or semiotic systems.

Information Reservoir: A container of data, information, concepts, knowledge, and wisdom.

Information Wave: A phase of civilization in which industries based on information processing and handling are dominant factors in the economy.

Knowledge: A reasoning unit of cognition that creates awareness based on facts scientifically proven, rules, laws, coherent inferences, and well-defined methods.

Macro-Ecology of Information: The study of information (cognition) as a whole and concerned with *aggregates* across nations and markets.

Micro-Ecology of Information: The study of information (cognition) at the individual level of decision-making in professional, social, and personal undertakings.

Wisdom: A pragmatic unit of cognition that generates volition, a chosen way of acting and communicating. It is a process of choosing ordered routines that provide success and eliminate obstacles in performance.

ENDNOTE

[1] Misinformation means that an explanation of an object or process is incorrectly provided, either purposely or not purposely.

Information Modeling in UML and ORM

Terry Halpin
Northface University, USA

INTRODUCTION

The *Unified Modeling Language* (UML) was adopted by the Object Management Group (OMG) in 1997 as a language for object-oriented (OO) analysis and design. After several minor revisions, a major overhaul resulted in UML version 2.0 (OMG, 2003), and the language is still being refined. Although suitable for object-oriented code design, UML is less suitable for information analysis, since it provides only weak support for the kinds of business rules found in data-intensive applications. Moreover, UML's graphical language does not lend itself readily to verbalization and multiple instantiation for validating data models with domain experts.

These problems can be remedied by using a fact-oriented approach for information analysis, where communication takes place in simple sentences, each sentence type can easily be populated with multiple instances, and attributes are avoided in the base model. At design time, a fact-oriented model can be used to derive a UML class model or a logical database model. *Object Role Modeling* (ORM), the main exemplar of the fact-oriented approach, originated in Europe in the mid-1970s (Falkenberg, 1976), and has been extensively revised and extended since, along with commercial tool support (e.g., Halpin, Evans, Hallock & MacLean, 2003).

This article provides a concise comparison of the data modeling features within UML and ORM. The next section provides background on both approaches. The following section summarizes the main structural differences between the two approaches, and outlines some benefits of ORM's fact-oriented approach. The following section uses a simple example to highlight the need to supplement UML's class modeling notation with additional constraints, especially those underpinning natural identification schemes. Future trends are then briefly outlined, and the conclusion motivates the use of both approaches in concert to provide a richer data modeling experience, and provides references for further reading.

BACKGROUND

Detailed treatments of UML are provided in Booch, Rumbaugh, and Jacobson (1999); Jacobson, Booch, and Rumbaugh (1999); and Rumbaugh, Jacobson, and Booch (1999). The UML notation includes hundreds of symbols, from which various diagrams may be constructed to model different perspectives of an application (e.g., use case diagrams, class diagrams, object diagrams, statecharts, activity diagrams, sequence diagrams, collaboration diagrams, component diagrams, and deployment diagrams). This article focuses on data modeling, considering only the static structure (class and object) diagrams. UML diagrams may be supplemented by textual constraints expressed in the Object Constraint Language (OCL). For a detailed coverage of OCL 2.0, see Warmer and Kleppe (2003).

ORM pictures the world simply in terms of objects (entities or values) that play roles (parts in relationships). For example, you are now playing the role of reading, and this article is playing the role of being read. Overviews of ORM may be found in Halpin (1998a, 1998b) and a detailed treatment in Halpin (2001a). For advanced treatment of specific ORM topics, see Bloesch and Halpin (1997), De Troyer and Meersman (1995), Halpin (2000, 2001b, 2002a, 2002b, 2004), Halpin and Bloesch (1999), Halpin and Proper (1995), and ter Hofstede, Proper, and van der Weide (1993).

DATA STRUCTURES

Table 1 summarizes the main correspondences between high-level data constructs in ORM and UML. An uncommented "—" indicates no predefined support for the corresponding concept, and "†" indicates incomplete support. This comparison indicates that ORM's built-in symbols provide greater expressive power for capturing conceptual constraints in graphical data models.

A *class* in UML corresponds to an *object type* in ORM. ORM classifies objects into *entities* (UML objects) and *values* (UML data values—constants such as character strings or numbers). A *fact type* (relationship type) in ORM is called an *association* in UML (e.g., Employee works for Company). The main structural difference between ORM and UML is that ORM avoids *attributes* in its base models. Implicitly, attributes may be associated with roles in a relationship. For example, Employee.birthdate is modeled in ORM as the second role of the fact type: Employee was born on Date.

The main advantages of attribute-free models are that all facts and rules can be naturally verbalized as sentences, all data structures can be easily populated with multiple instances, models and queries are more stable

Table 1. Comparison of the main data constructs in ORM and UML

ORM	*UML*
Data structures:	**Data structures**:
object type: entity type;	object class
value type	data type
— { use fact type }	attribute
unary fact type	— { use Boolean attribute }
2^{+}-ary fact type	2^{+}-ary association
objectified association (nesting)	association class
co-reference	qualified association †
Predefined Constraints:	**Predefined Constraints**:
internal uniqueness	multiplicity of ..1 †
external uniqueness	— { use qualified association } †
simple mandatory role	multiplicity of 1^{+}.. †
disjunctive mandatory role	—
frequency: internal; external	multiplicity †; —
value	enumeration, and textual
subset and equality	subset †
exclusion	xor †
subtype link and definition	subclass, discriminator etc. †
ring constraints	—
join constraints	—
object cardinality	class multiplicity
— { use uniqueness and ring } †	aggregation/composition
—	initial value, changeability
User-defined textual constraints	**User-defined textual constraints**

† = incomplete coverage of corresponding concept

since they are immune to changes that reshape attributes as associations (e.g., if we later wish to record the historical origin of a family name, a family name attribute needs to be remodeled using a relationship), null values are avoided, connectedness via semantic domains is clarified, and the metamodel is simplified. The price paid is that attribute-free diagrams usually consume more space. This disadvantage can be offset by deriving an attribute-based view (e.g., a UML class or relation scheme) when desired (tools can automate this).

ORM allows relationships of any *arity* (number of roles). A relationship may have many readings starting at any role to naturally verbalize constraints and navigation paths in any direction. Fact type readings use *mixfix* notation to allow object terms at any position in the sentence, allowing natural verbalization in any language. Role names are also allowed. ORM includes procedures for creating and transforming models (e.g., verbalization of relevant information examples—these "*data use cases*" are in the spirit of UML use cases, except the focus is on the underlying data).

In an ORM diagram, roles appear as boxes, connected by a line to their object type. A predicate appears as a named, ordered set of role boxes. Since these boxes are set out in a line, fact types may be conveniently populated with tables holding multiple fact instances, one column for each role. This allows all fact types and constraints to be *validated by verbalization as well as sample populations*.

While supporting binary and longer associations, UML uses Boolean attributes instead of *unary* relationships. For example, the fact instance expressed in ORM as "Person 'Sam Spade' smokes" would typically be rendered awkwardly in UML as "SamSpade: Person.isSmoker = true." To be business friendly, UML should support unary fact types directly (e.g., Room has a window, Person smokes, etc.).

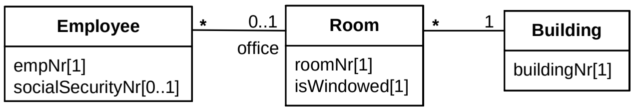
Each UML association has at most one name. Verbalization into sentences is practical only for infix binaries. Since roles for ternaries and higher arity associations are not on the same line, directional verbalization and multiple instantiation for population checks are ruled out. UML does provide object diagrams for instantiation, but these are convenient only for populating with one or two instances.

Both UML and ORM allow associations to be objectified as first class object types, called *association classes* in UML and *objectified* (or *nested) associations* in ORM. UML requires the same name to be used for the association and the association class, impeding natural verbalization, in contrast to ORM nesting based on linguistic nominalization (a verb phrase is objectified by a noun phrase).

CONSTRAINTS AND IDENTIFICATION SCHEMES

Business people communicate about things using value-based identification schemes, not memory addresses or

Figure 1. What semantics are missing?



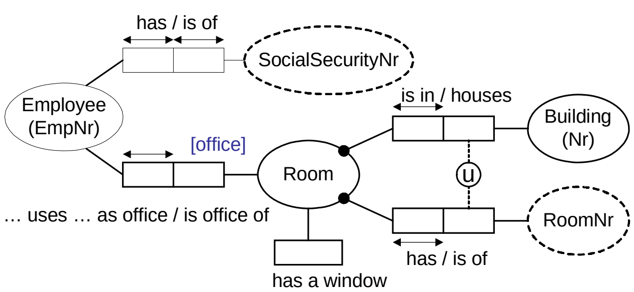
hidden object identifiers, and hence these need to be supported in any conceptual model. This entails rich support for what ORM calls internal and external uniqueness constraints, which may be used as a basis for identification of entities. UML 2.0 includes only a restricted form of internal uniqueness (maximum multiplicity of 1) and an even weaker form of external uniqueness (via qualified associations). Moreover, UML 2.0 does not require any value-based identification scheme for a class.

For example, consider the UML class model in Figure 1. Classes are depicted as rectangles, with the class name in the top compartment and attribute names in a lower compartment. Binary associations are depicted as lines connecting the classes involved. *Multiplicity constraints* may be applied to attributes and roles (association ends), using "1" for "exactly one," "0..1" for "at most one," and "*" for "zero or more.

Here, the attribute multiplicity constraints tell us that each employee has exactly one employee number and at most one social security number, each room has exactly one room number and exactly one value for isWindowed (a Boolean attribute to indicate whether a room has a window), and each building has exactly one building number. The multiplicity constraints on the associations tell us that each employee has at most one office, each room is the office of zero or more employees, each room is in exactly one building, and each building houses zero or more rooms. What semantics are missing?

To begin with, there is no declaration of how entities are identified by business people. In the business domain, employees are identified by their employee numbers, and rooms are identified by combining their local room number with the number of the building in which they are housed. The lack of the ability to declare these identification

Figure 2. ORM model captures the semantics lost in the UML model



schemes prevents an understanding of these business rules, and excludes any formal way to populate the data structures with fact instances. Moreover, the uniqueness constraint that each social security number applies to at most one employee is lost, because UML cannot declare the 1:1 nature of an association that has been modeled as an attribute.

One solution to this problem is to add additional constructs in UML to apply uniqueness constraints to arbitrary combinations of attributes or binary association roles that may be considered properties of the same class.

For comparison purposes, an ORM model for the same domain is shown in Figure 2. Entity types are depicted as named ellipses and value types as named, dotted ellipses. Fact types are depicted as role sets with at least one reading. Each role is depicted as a box connected to the object type that plays it.

Here we have one unary fact type: Room has a window. The other fact types are binary. For example, "Employee uses Room as office" and "Room is office of Employee" are forward and inverse readings of the same fact type. Role names may be added (e.g., office). Employees are identified by their employee number, and buildings by their building number. Simple identification schemes like these may be abbreviated in parentheses as shown. The arrow-tipped bars over roles depict internal uniqueness constraints (e.g., **each** Room is in **at most one** Building), and the solid dots are mandatory role constraints (e.g., **each** Room is in **at least one** Building). The external uniqueness constraint (circled "u') enables the composite identification scheme for Room. Even if we introduced a simple identifier for Room, this external uniqueness constraint still needs to be declared.

Given an n-ary association (n > 2), UML's multiplicity notation cannot express a mandatory role constraint on any association that has between 1 and n-2 mandatory roles. This is because multiplicity on one role is defined in terms of the other n-1 roles. This is fine for binary associations, but not for ternaries and beyond. For practical examples of such constraint patterns, see Halpin (2001b).

ORM includes many other graphical constraint primitives that go far beyond those found in UML class diagrams (e.g., inclusive-or, subset, exclusion, ring, and join constraints). Moreover, ORM's constraint primitives are far more orthogonal than those of UML. For discussions of these features, see Halpin (2000, 2001b, 2002a, 2004).

UML and ORM both permit users to add other constraints and derivation rules in a textual language of their choice. UML suggests OCL for this purpose (Warmer & Kleppe, 2003). Although OCL is an unambiguous language, its mathematical syntax renders it unsuitable for validating rules with non-technical domain experts. ORM's conceptual query language, ConQuer, is both formal and readily understandable to non-technical users, and its attribute-free nature makes it much more *semantically stable* than an attribute-based language such as OCL (Bloesch a& Halpin, 1997; Halpin & Bloesch, 1999).

FUTURE TRENDS

The OMG recommends its Model-Driven Architecture framework as a way to facilitate the generation of software artifacts from high-level models, starting with a Computation-Independent Model (CIM), and then moving down to a Platform-Independent Model (PIM), and finally the Platform-Specific Model (PSM) used in the actual implementation. For reasons given earlier, although useful for code design, UML currently has some shortcomings with regard to conceptual data analysis and the specification of business rules to be validated by business domain experts.

With a view to providing better support at the CIM level, in early 2003 the OMG received submissions to provide a truly semantic layer for business rules. Two of the three submissions incorporated features from ORM's fact-oriented approach, and at the time of writing it seems likely that a number of key aspects of fact-oriented modeling (e.g., support for unaries and high-level verbalization of business rules) will be included in the final adopted submission.

Many companies are now looking to model-driven development as a way to dramatically increase the productivity, reliability, and adaptability of software engineering approaches. It seems clear that both object-oriented and fact-oriented approaches will be increasingly utilized in the future to raise the percentage of quality code in software applications that can be generated from higher level models.

CONCLUSION

UML class diagrams are often more compact than ORM models, and they can be adorned with implementation detail for engineering to and from object-oriented programming code. Moreover, UML includes mechanisms for modeling behavior, and its adoption by the OMG is helping it gain wide support in industry, especially for the design of object-oriented software.

ORM is based on a small set of easily mastered, orthogonal concepts; its attribute-free nature facilitates model validation by verbalization and population and conveys semantic stability; and its graphical constraint language can formally capture many business rules. UML modelers willing to learn ORM can get the best of both approaches by using ORM as a front-end to their information analysis and then mapping their ORM models to UML, where ORM constraints with no UML counterpart can be captured in notes or formal textual constraints. This option will become more attractive once commercial tools provide automatic transformation between ORM and UML.

REFERENCES

Bloesch, A. & Halpin, T. (1997). Conceptual queries using ConQuer-II. In D. Embley & R. Goldstein (Eds.), *Proceedings of the 16th International Conference on Conceptual Modeling* (ER'97) (pp. 113-126). Berlin: Springer-Verlag.

Booch, G., Rumbaugh, J. & Jacobson, I. (1999). *The Unified Modeling Language user guide*. Reading, MA: Addison-Wesley.

De Troyer, O. & Meersman, R. (1995). A logic framework for a semantics of object-oriented data modeling. *OOER'95: Object-Oriented and Entity-Relationship Modeling* (pp. 238-249). Berlin: Springer-Verlag (LNCS 1021).

Falkenberg, E. (1976). Concepts for modeling information. In G. Nijssen (Ed.), *Modeling in data base management systems* (pp. 95-109). Amsterdam: North-Holland.

Halpin, T. (1998a). Object role modeling (ORM/NIAM). In P. Bernus, K. Mertins & G. Schmidt (Eds.), *Handbook on architectures of information systems* (pp. 81-101). Berlin: Springer-Verlag.

Halpin, T. (1998b). Object role modeling: an Overview. Retrieved from *www.orm.net/overview.html*.

Halpin, T. (2000). Integrating fact-oriented modeling with object-oriented modeling. In M. Rossi & K. Siau (Eds.), *Information modeling in the new millennium*. Hershey, PA: Idea Group Publishing.

Halpin, T. (2001a). *Information modeling and relational databases*. San Francisco: Morgan Kaufmann.

Halpin, T. (2001b). Supplementing UML with concepts from ORM. In K. Siau & T. Halpin (Eds.), *Unified Modeling Language: Systems analysis, design and development issues*. Hershey, PA: Idea Group Publishing.

Halpin, T. (2002a). Information analysis in UML and ORM: A comparison. In K. Siau (Ed.), *Advanced topics in database research, volume 1* (Chapter XVI, pp. 307-323). Hershey, PA: Idea Group Publishing.

Halpin, T. (2002b). Metaschemas for ER, ORM, and UML data models: A comparison. *Journal of Database Management, 13*(2), 20-30.

Halpin, T. (2004). Constraints on conceptual join paths. In J. Krogstie, T.A. Halpin & K. Siau (Eds.), *Information modeling methods and methodologies*. Hershey, PA: Idea Group Publishing.

Halpin, T. & Bloesch, A. (1999). Data modeling in UML and ORM: A comparison. *Journal of Database Management*, *10*(4), 4-13.

Halpin, T., Evans, K, Hallock, P. & MacLean, W. (2003). *Database modeling with Microsoft® Visio for enterprise architects.* San Francisco: Morgan Kaufmann.

Halpin, T. & Proper, H. (1995). Subtyping and polymorphism in object-role modeling. *Data & Knowledge Engineering, 15*(3), 251-281.

ter Hofstede, A., Proper, H. & van der Weide, T. (1993). Formal definition of a conceptual language for the description and manipulation of information models. *Information Systems, 18*(7), 489-523.

Jacobson, I., Booch, G. & Rumbaugh, J. (1999). *The unified software development process*. Reading, MA: Addison-Wesley.

OMG. (2003). *OMG Unified Modeling Language Specification*, *version 2.0*. Retrieved from *www.omg.org/uml.*

Rumbaugh, J., Jacobson, I. & Booch, G. (1999). *The Unified Modeling Language reference manual*. Reading, MA: Addison-Wesley.

Warmer, J. & Kleppe, A. (2003). *The Object Constraint Language: Getting your models ready for MDA* (2nd edition). Reading, MA: Addison-Wesley.

KEY TERMS

Arity: The number of roles in a fact type (unary = 1, binary = 2, ternary = 3, etc.). In ORM, fact types may be of arity 1 or more. In UML, fact types (associations) may be of arity 2 or more.

Business Rule: A constraint or derivation rule that applies to the business domain. A static constraint restricts the possible states of the business, and a dynamic constraint restricts the possible transitions between states. A derivation rule declares how a fact may be derived from existing facts, or how an object is defined in terms of existing objects.

Conceptual Schema: Specification of the structure of a business domain using language and terms easily understood by a non-technical domain expert. A conceptual schema typically declares the fact types and business rules that are relevant to the business domain.

Elementary Fact Type: In ORM, an elementary fact is an atomic proposition that applies a logical predicate to a sequence of one or more objects of a given type; it cannot be split into smaller facts without information loss. An elementary fact type is a kind of elementary fact. For example: Person smokes; Person was born in Country; Person introduced Person to Person. In UML, an elementary fact type is known as an elementary association or elementary relationship type.

Entity Type: An entity is a non-lexical object that in the real world is identified using a definite description that relates it to other things (e.g., the Country that has CountryCode 'US'). Typically, an entity may undergo changes over time. An entity type is a kind of entity, for example, Person, Country. In UML, an entity is called an object, and an entity type is called a class.

Object-Role Modeling (RM): A fact-oriented approach for modeling information at a conceptual level, using language that is easily understood by non-technical domain experts. ORM includes rich graphical and textual languages for modeling facts and business rules, and provides procedures for creating conceptual models and transforming them to lower level models for implementation.

Object Type: In ORM, an object is either an entity (non-lexical thing) or a value (lexical constant, such as a character string), and an object type is a kind of object (e.g., Person, CountryCode). In UML, the term "object" is restricted to entities (instances of classes), while the term "data value" is used for instances of data types.

Role: In ORM, a role is a part played in a fact type (relationship type). In UML, this is known as an association-end. For example, in the fact type *Person works for Company*, Person plays the role of employee, and Company plays the role of employer.

UML: The Unified Modeling Language adopted by the Object Management Group as a modeling language for object-oriented analysis and design of software systems. UML includes several sub-languages and diagram notations for modeling different aspects of software systems.

Information Resources Development Challenges in a Cross-Cultural Environment

Wai K. Law
University of Guam, Guam

INTRODUCTION

Western management theory considers information the lifeblood of organization. The sharing of information lubricates the interlocking divisions within the organization, promoting the effective achievement of organizational goals with external business partners. However, in many regions of the world, information represents power, and managers often try to accumulate as much of it as they can while denying access to others (Oz, 2002). For others, the disclosure of information is considered a threat to the span of management control (Rocheleau, 1999). In some cases, administrators could be more interested in the scale of the information system and its associated budget, than in the capability and functionality of the system (Kalpic & Boyd, 2000). These are examples of conflicting cultural values in a cross-cultural environment. The introduction of Western management approaches conflicts with regional administrative styles, diminishing the effectiveness of information systems (Shea & Lewis, 1996; Raman & Watson, 1997). Sensitivity to cultural differences has been recognized as an important factor in the successful global deployment of information systems. Minor information management issues potentially resolvable through improved communication in the West often manifest as major challenges in a cross-cultural environment.

BACKGROUND

The literature provided thorough coverage on designs, development, and implementation of computer-based information systems (CBISs). Numerous studies examined various systems solutions for organization needs (Applegate, 1995; McLeod, 1998; O'Brien, 2002). However, the projected value of information technology has been formulated based on a rough assessment of the possibilities without full appreciation of the limitations due to resistance to organizational and social changes (Osterman, 1991). Increasingly, management realized that massive deployment of information systems on a global basis, even with prudent management of the systems, has not been producing the desirable outcomes of value generation. Recent studies revealed the significant influence of cultures toward the success of transferring information technology beyond the Western world. National culture, organization culture, and MIS culture induced influence over the successful development and management of information resources (Hofstede, 1980; Raman & Watson, 1997). Shea and Lewis (1996) suggested the desirability of paying close attention to the user absorptive rate in the transfer of new technology into a different cultural environment. It became apparent that adaptation of information system designs to new cultural environments was insufficient to guarantee successful implementation. User selection of technological features, driven by cultural preferences, could be a key factor for designing information systems in a multicultural environment. Other studies reported the numerous obstacles of developing CBISs under various cultural settings, even with highly motivated leaders to support the deployment of information systems (Raman & Watson, 1997; Al-Abdul-Gader, 1999).

The information system function must enhance user effectiveness and efficiency in utilizing the information to improve value delivery for the organization (Parker, 1996). New challenges emerged as nontechnical issues and clouded the measurement of information system performance. A typical information system would be designed to provide information to users with common needs. Good data reports should contain all the required information with accurate representation of events. The reports needed to be generated in a timely fashion and in a format usable by the users (McLeod, 1998). However, individual users tended to value information systems for providing custom reports to meet individual needs in specific circumstances (Heeks, 1999). Inconsistent expectations in a cross-cultural environment crippled the effective management of information resources. Cultures carried different interpretations for timeliness, completeness, and relevancy of information.

Makeshift management decisions generated new dynamics in several ways. In the spirit of promoting free information exchange, the department that owned the information system became obligated to provide information to others (Oz, 2002). However, the new responsibility seldom came with additional resources. The information owners became reluctant to supply information, because doing so would take away that resource from other regular tasks (Davenport, 1997). Some managers shifted the data reporting responsibilities to other divisions, creating a bureaucratic nightmare for the users. Some ignored data requests, and others manipulated the data flows with respect to organizational politics (Oz, 2002; Rocheleau, 1999). Those working in the public sector faced the challenge of maintaining a delicate balance as they attempted to fulfill their responsibilities for both confidentiality and access to data (Osterman, 1991; Duncan, 1999). The problems would be more severe under a relationship-based culture, where favors could not be declined.

Cultural backgrounds shaped the preferential information system model. In some cultures, managers would be intuitive and feelings based and would have vague expectations for the performance of the information system. There would be more emphasis on group harmony and saving face than on actual problem solving (Bjerke, 1999). Others would be more interested in meeting obligations, ignoring the source and validity of the reports. The controlling manager seeked a complex and broad information system that provides qualitative data (Kluckhohn & Strodtbeck, 1961; Lane & DiStefano, 1992; Shea & Lewis, 1996). All these personality extremes coexist in a cross-cultural setting, making it more challenging to design systems than in a single-culture environment. The perceived value of information resources became less predictable in a cross-cultural environment.

CHALLENGES IN CROSS-CULTURAL INFORMATION SYSTEMS MANAGEMENT

The rapid expansion of Western influence on a global basis created an environment under the cross-currents of Western corporate culture and regional cultures. Nations in the Pacific Basin have established close relationships with the Western world. Heavy Western investments have transformed these nations into showcases for Western systems. However, underneath the formal display of the Western culture, local cultures retained strong influence on their societies. An influx of immigrants holding on to their traditions further diluted the penetration of Western influence in these regions. The predominating regional workforce challenged Western corporate culture through their deep-rooted traditions and work habits. For example, a massive absenteeism could be expected on festival days, even without approved leaves or holidays. Timely arrival at a meeting would be accepted as 15 minutes to several hours after the scheduled time. Mandated reports could be excused without penalty, and the uttermost concern, over efficiency, was to preserve group harmony. Sometimes, this meant ignoring facts to restore stability and group harmony. Periodic acquisition of technology would be celebrated even without the appropriate infrastructure support, preventing usage of the technology. Experience in the Pacific Basin provided a sampling of information resources management issues that became significant challenges in cross-cultural environments.

Challenge One: The design objectives of an information system must expand from efficiency orientation to adaptive accommodation of cultural habits. It becomes desirable to allow and to track dynamic modification of data-processing procedures according to shifting organizational and cultural influences.

While a primary design objective of an information system was to provide efficient transaction processing, often, the affected human system was slow to accept the implicit MIS culture embedded in the system design. Western culture emphasized timeliness and accuracy, which were less important to many cultures. For example, it often took months to update databases from paper documents. Some users relied on the information system for information, while others insisted on paper documents only. Hence, circulation of multiple versions of reports was common depending on the sources of the reports. Parallel operations to accommodate parallel cultures generated organizational conflict. Influential users and administrative interventions threatened the integrity of information systems. The full potential of the information system was suppressed to a preference for cultural norms, and only system features that would not threaten cultural practices would be allowed to remain. Some local cultures emphasized protecting family members more than performance appraisal. The value of information was not as much for improving decision making, but to endorse group position, to preserve relationship, and to avoid embarrassment.

Challenge Two: There is a need for clear definitions of data ownership and responsibilities for data acquisition, data quality control, and data distribution. This is especially challenging in cultural environments, where the political attributes of information interfere with the communicative value of information.

In many Eastern cultures, credible information was deferred to leaders and elders with power and status. Political relationships dictated the availability of information and the accessibility to organizational data. This was contrary to the basic assumptions of CBISs that promoted the free exchange of information (Oz, 2002; Osterman, 1991; Rocheleau, 1999). The bureaucratic procedures for the approval of data usage defeated the designed roles of the information system. A fully developed database supported very limited applications. The lack of explicit system objectives coupled with the practice of delegating data management responsibility to the lowest-level unskilled workers created data integrity problems. For example, withholding information to gain and maintain power was acceptable among many Asian cultures. Openness would be considered a sign of weakness. It would be critical to formally establish the credibility, relevancy, and accessibility of the data resource.

Challenge Three: Management must meticulously plan data acquisition, data preparation, data distribution, and data usage, and fully understand the required organizational incentive and associated costs for maintaining information flow within the organization. This is especially important in a cultural environment where data-driven decision making is a new practice.

An uncoordinated approach to information resource management created fragmented entities to process information for narrow applications. The fad of data-driven decision making created a mad race for data reports using every available political connection. The result would be a great assortment of data reports with massive details. Inconsistency occurred among data reports, depending on the data-processing methods and storage formats. For example, a report from an off-line, static database in a remote office could be given equal credibility as a report generated from a current database from the data center. In a cross-cultural environment, influential individuals would compete to justify the merit of their reports from their cultural perspectives. The heated debates, along with discrepancies among the reports, frustrated the end users and led to distrust of the information systems for the inability to produce usable information reports. Regretfully, the information systems were seldom designed to generate reports for decision support.

Challenge Four: Management must take leadership in establishing precise, formal data definitions, and communicate them to all potential data users and those assigned roles in data distribution. This is especially important where mastery of languages, cultural predisposition, level of information literacy, and social attitude could strongly influence the group dynamic of data usage.

Technology evolution increasingly placed information systems under the direct control of end users. However, end users often lack technical expertise, and few were committed to the development of information resources. Events and samples were confused with statistics. Relaxed practices in standards and data definitions created issues in data validity and data quality. Potential information was lost when processed data replaced raw data, while the time sensitivity of dynamic data was ignored. Time series data were deleted to preserve storage space. The information system would be blamed for the unfortunate chaos. In one incident, a user group maintained multiple versions of a database with the assistance of the data center. However, only selected workers in the user group and the data center were aware of the special arrangement. The different versions of the database were discovered only when two identical requests for information were returned with different outcomes. Top management, unwilling to escalate cultural tension, ignored the potential seriousness of the data integrity issue. In another incident, several users entered an unresolved dispute on their interpretations of a data definition according to their understanding of the language. The data definition used by the data center in maintaining the database was rejected.

Challenge Five: The increased complexity and frequency of usage of information reports is, in reality, a severe drain on budgetary resources. Management needs to develop a mechanism to track data usage and adjust resources appropriately. This could be more challenging under cultural environments that lack sophistication in information processing.

Modern management practices seek opportunities to replace physical resources with information. When management failed to adjust budgets to support the information services, those affected would try every means to discontinue information services. On the other hand, uncontrolled access encouraged abuse, wasting valuable resources. Ethics, disciplined usage, and an understanding of information value supported the information practices in Western society. The problems would be crippling in a culture with different appreciation for information under different ethical standards. A local culture of generosity would insist on the free distribution of fully colored documents. Another practice was to circulate printed copies of e-mail to avoid offending anyone. The practices quickly depleted the budget for supplies.

Challenge Six: Management must take an active role in controlling the flow of organizational data, both within the organization and to the external environment. Management should consider endorsement of an official organizational data to ensure consistency rather than leave the official data report to random actions. This is especially important in a cultural setting where it is impractical to correct public statements of social leaders, regardless of facts.

In cultures where subordinates would not question the positions of leaders, an information system must implicitly support the decisions and public statements of the leaders (Gannon, 2001). Officials of a local organization proposed an expensive marketing campaign, pointing to a decline in demand in the primary market. However, published data actually attributed the demand decline to the collapse of an emerging market. It would be an embarrassment to point out the omission, and the wrath of the society could be on those who allowed the facts to be publicized.

Challenge Seven: Management needs to play an active role in data planning and closely align the information report designs for decision support. This is especially challenging in a cultural environment that lacks appreciation for operational planning and control.

In cultures where gesture would be more important than details, systematic failure to collect information would be accepted and forgiven. Information systems applications were limited to payroll and accounting (Kalpic & Boyd, 2000). In some cases, the lack of adequate information was the key to assuring continued financial support. Organizations were unprepared to collect and store data to support meaningful decision support applications. Information systems were seldom utilized to their full potential under such cultural settings.

CHALLENGES IN PERCEPTION OF INFORMATION RESOURCE

Potential cultural myopia required great efforts when communicating the principles of information resource management.

Challenge One: Information Resources Solely as an Investment Issue

A common response to deficiencies in organizational data was to seek capital investment for new technology. This perception underestimated the requirements for system and facility maintenance, technical support, user training, data architecture development, data security, and distribution. Lack of organizational readiness stalled the deployment of information systems. Inflated expectations and uncoordinated usage of data services nullified the value of the information system. Erratic funding patterns destroyed development projects, making it extremely difficult to retain technical personnel. Poor maintenance damaged equipment and threatened data integrity. Cultural managers eager to modernize without fully understanding the implications of information resources management eventually abandoned their support for the Western ideas.

Challenge Two: Information Resources Development by Delegation

Management seeking an easy fix to the organizational data problem mandated data compilation activities by the functional divisions. Besides undermining information resources as a critical organizational asset, little technical assistance was provided. Unmotivated managers neglected data quality and resisted data distribution. The turnover of administrators caused discontinuity in data resources development.

Challenge Three: Limited Appreciation for Information Resources

Information resources could be compared to utility services, such as water, the value of which was suppressed until serious issues developed in quality and supply. Often neglected was the accountancy of the value contribution of the information system, beyond periodic technical improvements. Benchmark studies should identify cost performance as well as the critical roles of information resources within the organization.

FUTURE TRENDS

The historical development of information systems has followed the model of a rational manager (Kepner, 1965), with emphasis on openness, clear structure, innovative practices, and logical thinking. In regions where traditions and relationships resisted changes, information systems designers must consider the needs of emotional decision makers, with heavy emphasis on the concern to maintain social and cultural stability. Some cultures demand tight control of the information flow, while other cultures are casual about the absolute data quality. Some organizations integrate information systems to become organizational backbones, and others prefer to separate information in isolated pockets. Some prefer simple information systems, while others invest in sophisticated intelligence systems. Information systems for cross-cultural environments must deliver value to users with diversified backgrounds. Comparative study on information system features valued across cultural settings should improve the value delivery of the information system function.

CONCLUSION

Despite rapid technological development, information resource management is still a relatively new concept. Data reports preparation is often a laborious activity, and accepted practices and administrative preferences still drive decision making. Organizations that anticipate increasing exposure to multicultural environments should allow a longer time for organizational adjustment to technical development. Information systems originally developed as productivity tools, for data processing and report generation must undergo radical design evaluation to meet the diversified user expectations and information skills. An information resource manager must also carefully consider data ownership and data distribution issues. Cultural preferences and information values should be carefully considered in the justification of information services. The information system objective should be clearly distinguished from the information system capabilities, especially with different cultural interpretations of information value. Top management should play an active role in defining organizational data flow, with implementation of appropriate incentives. Special attention should be given to precise data definition, especially with a workforce with different training backgrounds under different cultural and language settings. Last, it is critical to emphasize strict standards for data quality, due to differences in expectations for information system performance.

REFERENCES

Al-Abdul-Gader, A. H. (1999). *Managing computer based information systems in developing countries: A cultural perspective*. Hershey, PA: Idea Group Publishing.

Applegate, L. (1995, September 26). *Designing and managing the information age IT architecture*. Cambridge, MA: Harvard Business School Press.

Bjerke, B. (1999). *Business leadership and culture: National management styles in the global economy*. Cheltenham, UK: Edward Elgar.

Davenport, T. (1997). *Information ecology: Mastering the information and knowledge environment*. New York: Oxford University Press.

Duncan, G. T. (1999). Managing information privacy and information access in the public sector. In G. D. Garson (Ed.), *Information technology and computer applications in public administration: Issues and trends* (pp. 99-117). Hershey, PA: Idea Group Publishing.

Gannon, M. J. (2001). *Understanding global cultures: Metaphorical journeys through 23 nations*. Thousand Oaks, CA: Sage.

Heeks, R. (1999). Management information systems in the public sector. In G. D. Garson (Ed.), *Information technology and computer applications in public administration: Issues and trends* (pp. 157-173). Hershey, PA: Idea Group Publishing.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Thousand Oaks, CA: Sage.

Kalpic, D., & Boyd, E. (2000). The politics of IRM: Lessons from Communism. In M. Khosrow-Pour (Ed.), *IT management in the 21st century* (pp. 72-73). Hershey, PA: Idea Group Publishing.

Kepner, C. H., & Tregoe, B. B. (1965). *The rational manager*. Princeton, NJ: Kepner-Tregoe, Inc.

McLeod, R. M. Jr. (1998). *Management information systems*. New York: Prentice Hall.

O'Brien, J. A. (2002). *Management information systems: Managing information technology in the e-business enterprise.* New York: McGraw-Hill.

Osterman, P. (1991). Impact of IT on jobs and skills. In M. S. Scott Morton (Ed.), *The corporation of the 1990s: Information technology and organizational transformation*. New York: Oxford University Press.

Oz, E. (2002). *Management information systems*. Boston, MA: Course Technology.

Parker, M. (1996). *Strategic transformation and information technology: Paradigms for performing while transforming*. New York: Prentice Hall.

Raman, K. S., & Watson, R. T. (1997). National culture, information systems, and organizational implications. In P. C. Deans & K. R. Karwan (Eds.), *Global information systems and technology: Focus on the organization and its functional areas* (pp. 493-513). Hershey, PA: Idea Group Publishing.

Rocheleau, B. (1999). The political dimensions of information systems in public administration. In G. D. Garson (Ed.), *Information technology and computer applications in public administration: Issues and trends* (pp. 23-40). Hershey, PA: Idea Group Publishing.

Shea, T., & Lewis, D. (1996). The influence of national culture on management practices and information use in developing countries. In E. Szewczak & M. Khosrowpour (Eds.), *The human side of information technology management* (pp. 254-273). Hershey, PA: Idea Group Publishing.

KEY TERMS

Cross-Cultural Environment: The coexistence of more than one cultural influence in different segments of a society, or the simultaneous adoption of different cultural practices at work, social events, and family life.

Cultural Habit: Accepted behaviors within a group of people, sharing some common background, such as language, family heritage, education, and living and socializing environment.

Data Definition: An elaborate statement of the representation of each piece of data, its source, storage method, and intended usage.

Data-Driven Decision Making: The practice of purposefully collecting, analyzing, and interpreting data according to accepted criteria, and using the outcomes to select and justify decisions.

Data Planning: The projection of expected future needs for data, with specifications on data sources, data collection and storage, data processing and presentation, data distribution, and data security.

Information Resources: Resources required to produce information, including hardware, software, technical support, users, facilities, data systems, and data.

Official Organizational Data: A selected version of dynamic data released for both internal and external usage to maintain consistency in communication.

Transaction-Processing System: This is a computer system designed to support the recording and completion of value exchange activities.

Information Resources Development in China

Maosheng Lai
Peking University, China

Xin Fu
University of North Carolina at Chapel Hill, USA

Liyang Zhang
Baidu.Com (Beijing) Co., Ltd., China

INTRODUCTION AND BACKGROUND

In its several thousand years of social progress, China has put continuous efforts on cultural development, which to a certain extent contributed to the exploitation and utilization of information resources.

Since the founding of the People's Republic of China in 1949, the government has been attaching great importance to information resources development. In 1956, the government set "Marching Towards Science" as the directing principle for the course of information resources management, and made a conscientious plan in information resources development with the emphasis on collecting, rearranging, analyzing, indexing, and reporting scientific and technical documents from home and abroad to serve the needs of professionals in various disciplines. By 1987, the scientific and technical information sector alone had already possessed 26,000 foreign periodicals, 6,000 domestic periodicals, 120 million patent manuals, and more than 32 million books. There were 236 abstracting and indexing journals published annually, covering more than 1.2 million documents and articles. Further, there were 2,038 public libraries at county and higher levels, collecting more than 200 million books. There were 745 academic libraries, collecting 250 million books. There were also more than 4,000 libraries in research institutes (Guan, 1988).

In late 1980s and early 1990s, however, information resources development was affected by the readjusting of China's economy. Non-profit libraries and information service institutions suffered from a severe shortage of money for collection development. As a result, information resources development was captured in severe logjam or even retrogress. Types of document collections in some libraries dropped by a half or even two-thirds (Fu, 1996). Many abstracting and indexing journals stopped publication. On the other hand, some new abstracting and indexing journals emerged, along with bibliographical databases, which catered to market demand.

Under the promotion of the international information technology revolution, China has been experiencing an upsurge in information development since the last decade of the 20th century. Information infrastructure construction keeps a rapid pace in development. The ownership of telephone, cellar phone, and computer has been increasing steadily. The overall scale of China's information infrastructure in terms of network capacity and number of users now ranks first in the world (China Telecommunications Yearbook, 2003; He, 2004). On the other side, however, information resources development is lagging far behind. The lack of information, especially Chinese information, in networks and information systems influences the benefit of investment in information technology, which has become a major obstacle not only to China's informationalization drive, but also to the competitiveness of Chinese economy.

Since the mid-1990s, under the promotion of the tide of information superhighway construction in many countries, information resources development in China entered a new phase. In 1997, the Chinese government constituted the "Draft on China's Informationalization," drawing the outline of China's information infrastructure (Zou, 1997), which includes six elements: information resource, national information network, information technology (IT) application, information industry, information professional, and information policy code and standard.

Information resource was set as the primary element among the six, which showed the state's emphasis on its development. This also indicated that people once again realized the importance of information resources development. Several years later, the proposal was accepted as a part of China's tenth "five-year plan," which marked that information resources development became the central task of China's informationalization drive.

The concept "information resources development" used in this item refers to collection, processing, organization, and dissemination of document resources, as well as their digitalization and networking. Factual and data resources ought to be included in the concept. However, China's progress in these aspects is relatively slow. In recent years, people started to realize the importance of factual and data resources development. The departments concerned have started to work out a plan for constructing the National Data Center.

MAJOR INITIATIVES IN CHINA'S INFORMATION RESOURCES DEVELOPMENT

Under the guidance of the policies introduced in the last section, the Chinese government initiated several major information resources development projects to change the current situation of inconsistency between information resources development and information network construction, as well as to lessen the discrepancy between information resources available and those required by the public.

CALIS (China Academic Library and Information System)

CALIS is an initiative under China's plan to build 100 key universities in the 21st century (named "211 Project" by the Ministry of Education). It aims at constructing a networked information resources sharing system based on the China Education and Research Network (CERNET) so as to parallel the development of a communication network and an information resources network, and provide university faculty and students as well as professionals in research institutions with easy access to a national information service system characterized by abundant information resources, advanced technologies, and a convenient service system.

The service system consists of a CALIS national management center, four CALIS national information centers (covering sciences and social sciences, engineering, agricultural science, and medical science, respectively), and seven CALIS regional information centers (in Beijing, Shanghai, Nanjing, Guangzhou, Xi'an, Chengdu, Wuhan, and Changchun, respectively). The system will also be linked to major information service systems outside China to form China's Academic Library and Information System. The construction of CALIS will greatly increase the amount of information available to academic libraries and also improve their capability in information services (data from www.calis.edu.cn).

Digitalization Projects

The China Digital Library Project was carried out under the coordination of the Ministry of Culture. In July 1997, the National Library of China (then Beijing Library), together with Shanghai Library and a few other institutions, started the Chinese Pilot Digital Library Project (CPDLP). Later in 1998, the Ministry of Culture formally put forward the proposal of constructing the China Digital Library. Various enterprises and organizations—such as China Telecom, the National Library of China, the Chinese Academy of Sciences, China Aerospace Industrial Corporation, Peking University, and Tsinghua University—participated in the project, called the China Digital Library Project.

As for the achievements of the project, it was expected that some 20 resource databases would be made available on the "China Cultural Information Network," which included the China Medical Science Resource Database, China Tourism Resource Database, China Economic Resource Information Database, among others. The network will become a significant channel of spreading the Chinese culture and strongly support China's project of "rejuvenating the nation through science and education" (Xu, 1999, 2000; Sun, 1999).

Besides the China Digital Library Project, various other digital library projects were also carried out. The construction of the Chinese National Science Digital Library (CSDL) was started in late 2001. The project, as part of the Knowledge Innovation Project of the Chinese Academy of Sciences, aims to build a digital information service system that meets the international developing trends of digital libraries, and caters to the development of the Chinese Academy of Sciences. It should be able to serve the needs of researchers and professionals in information accessing and knowledge innovation when it is finished in three to four years' time (Zhang, 2002).

In China's Taiwan Province, eight digital library initiatives are currently under way, including the construction of a Digital Library and Information Center and building of the Haoran Digital Library at Jiaotong University. Objectives of the initiatives are to promote information exchange among learning and research institutions in Taiwan, and coordinate their purchase of information resources such as databases from foreign countries. Another objective is to promote the research on Chinese culture, especially on Chinese history (Lv, 1999).

There are also digitalization projects other than construction of digital libraries. In January 1999, the Geology Department of the Chinese Academy of Sciences raised to the State Council a proposal on strategies of China's "Digital Globe" development, indicating the importance of building a national global information infrastructure and establishing a digital global spatial information sharing system (*Information Industry Newspaper*, November 22, 1999). In November 1999, the first "Digital Globe" International Conference was held in Beijing, showing that the Chinese government attached great importance to international cooperation in this area (*China Computer World*, December 6, 1999).

Construction of the China National Science and Technology Library (NSTL)

In June 2000, the China National Science and Technology Library was formally established through the cooperation of China's Ministry of Science and Technology, State Committee of Economics and Trade, Ministry of Agriculture, Ministry of Health, and the Chinese Academy of Sciences. As a virtual scientific and technical resource center, it consists of eight library and information institutions, including the Library of the Chinese Academy of Sciences, National Engineering Library of China, Library of the Chinese Academy of Agricultural Science, Library of the Chinese Academy of Medical Science, and Institute of Scientific and Technical Information of China. The center utilizes advanced technologies and methods to collect information from domestic and foreign sources. It also makes standards and criteria in information sharing. Moreover, the center serves as a bridge of cooperation between Chinese information resources management professionals and their foreign counterparts (Yuan, 2001).

Special-Topic Information Resources Development

Government Information Resources Development—China Online Government Project

On January 22, 1999, the China Online Government Project Start-Up Conference was held in Beijing, sponsored by China Telecom and the State Economic and Trade Commission, together with the information sectors of more than 40 ministries. During the conference, the China Online Government Project was started. Subsequently, the main Web site of the China Government Portal, *www.gov.cn*, was established in 1999.

According to the "White Paper on China Online Government Project," accessed through the portal, the project refers to the practice by the government at all levels to establish their formal Web sites to promote office automation in government work, offer public services via the Internet, and fulfill the roles of management and service in the fields of society, economy, and social life.

Patent Information Resources Development

Patent information is an essential part of a country's technical information resources. To meet the users' requirement of searching and utilizing patent information, China Patent InfoNet was established by the Retrieving and Consulting Center under the State Intellectual Property Office in May 1998. In January 2002, its new version (*www.patent.com.cn*) was published online and started to offer to patent users and researchers comprehensive services, such as patent information retrieval, introduction of patent laws and regulations, guidelines for patent application.

Construction of the National Institute for Information Resources Management

Three national research centers for information resources management have recently been set up in Beijing, Nanjing, and Wuhan to promote research on theories, applications, policies, and technologies in IRM, affiliated to the Department of Information Management of Peking University, Department of Information Management of Nanjing University, and School of Information Management of Wuhan University.

INITIAL IMPACT OF SOME INITIATIVES

CALIS

Started in November 1998, CALIS completed its first phase of construction by the end of 2001. Currently, the system can provide an online public access catalog, interlibrary loan (ILL), Internet navigation, online cataloging, cooperative literature purchasing, and various other services through digitalization of information resources, networking of information services, and cooperation among participating academic libraries. As a result, universities and colleges in China now possess information resources greatly, more than ever before: the variety of foreign periodicals increased by one-third, 95% of the Chinese literature and 80% of the foreign literature are now available, and more than 100 academic libraries offer 24-hour online information services. In addition, 25 distinctive databases and 194 disciplinary navigation databases are built.

In its second phase of construction starting from 2002, CALIS aims to further strengthen the document-supporting ability of academic libraries. It plans to automate and network about 1,000 academic libraries, among which 100 will be completely automated and networked, becoming the backbones in information resource sharing. Some 20 academic libraries will be developed into digital library bases, acting as the kernels of information service systems and distributing centers of information resources. Besides, digitalized information resources imported from foreign countries are expected to cover all subject areas, while domestic information resources will be as much as several Tera Bytes (Zhu, 2001).

China Digital Library Project and NSTL Construction

Construction of the China Digital Library and the National Science and Technology Library have been advancing smoothly. In April 1999, the China Cultural Information Net started operation as the top level of the China Digital Library. In November 1999 and February 2000 respectively, the Capital Library and China Radio International (CRI) became experimental units of the China Digital Library Project. It should include information resources not only from libraries, but also from the government, even from international channels. The ultimate goal of the project is to build a "Digital China."

The National Culture Information Resources Sharing Project has been carried out. The culture information resources base composed of 40 multimedia resource bases has included 300 films, 132 local operas, 300,000 pictures, and nearly 100 lectures. All resources have been published through Web pages to allow access for people in rural and remote areas.

The initiative of building the National Science and Technology Library is near conclusion. Through two years of construction, participating libraries now collect more than 16,000 types of foreign scientific and technical literature (including periodicals, conference proceedings, technical reports, etc.), as compared to no more than 4,000 types in 1996. Some 6.5 million bibliographical records had been put online by the end of March 2002, and this number is expected to increase at a rate of two million per year. The network service system provides 24-hour, free secondary literature retrieval service to Internet users. In March 2002 alone, 1.37 million users visited the system, as compared to 150,000 when the system was started in January 2001. More than 60,000 users have received full-text document service.

Development of Commercial Information Products

Information resource development in a market-oriented approach achieves great effect. Many database and information service providers (such as ICP, ISP) come into operation, among which the China Academic Journals CD-ROM database, ChinaInfo Group, Chongqing Weipu Information Consulting Corporation Ltd. (*www.vipinfo.com.cn*), Beijing Scholar Sci-Tech Co., Ltd., and China Infobank enjoy nationwide recognition. The Chinese Journal Full-Text Database includes more than 6.1 million articles from 6,600 major periodicals published in Mainland China since 1994, as well as more than 15 million bibliographical records. The database is available both online and in CD-ROM form.

In a broader context, Internet-based information resources have also undergone rapid development. According to statistics from the "Survey on Information Resources in China," which was released by the China Internet Network Information Center (CNNIC) in September 2001 (www.cnnic.net.cn/tj/rep2001.shtml), there were 692,490 registered domain names, 238,249 Web sites, 159,460,056 Web pages, and 45,598 online databases within China. A recently released report shows that the number of Web sites has increased to 293,213 by June 2002 (www.cnnic.net.cn/develst/2002-7e/index.shtml).

FUTURE TRENDS AND CONCLUSION

The projects that we have introduced above lay a solid foundation for the further development of information resources in China. Recently, the Ministry of Science and Technology has started up the Science and Technology Documents Resource and Service Platform. Founding of the National Informationization Directing Committee also boosted the development and utilization of information resources.

Looking into the future, we can feel the long way ahead for China to improve its information resources. Efforts need to be made in a number of areas: first, the digital library projects need to be further expanded; second, the government should continue its support for information resources development projects; and third, the issue of nationwide cooperation across industries needs to be addressed properly as one of the most important problems in information resources development.

REFERENCES

China Telecommunications Yearbook. (2003). Beijing: Posts and Telecom Press.

Fu, L. (1996). Some thinking on the strategies of China's sci-tech information resources construction. *Journal of* *the China Society for Scientific and Technical Information*, 15(5), 374-377.

Guan, J. (1988). *Information work and information science development strategy*. Beijing: Sci-Tech Document Publishing House.

He, W. (2003). *Rapid development of the communication industry in China*. Retrieved from *www.cnii.com.cn/20030915/ca226082.htm*.

Lv, Y. (1999). Digital library in Taiwan. *China Computer World*, (May 31).

Sun, C. (1999). Towards a digital library. In Wang, R. & Lu, Z. (Eds.), *Collected thesis on digital library*. Shenyang: NEU Press.

Xu, W. (1999). Constructing the China Digital Library Project. *Guangming Daily,* (May 7).

Xu, W. (2000). Great significance in building the China Digital Library. *Guangming Daily,* (March 8).

Yuan, H. & Meng, L. (2001). A practice on information resource sharing in a Web-based environment—construction and development of national research centers for information resources management. *Proceedings of the Academic Seminar Commemorating the 45th Anniversary of China's Scientific and Technical Information Cause*, Beijing.

Zhang, X. (2002). The China Scientific Digital Library: A user-oriented digital information service system. *Sci-Tech International*, (4), 21-23.

Zhu, Q. (2001). A rewarding practice on information resource sharing oriented towards the 21st century—advances in the construction of the China Academic Library and Information System. *Proceedings of the 2001 Annual Meeting of the China Society for Library Science,* Chengdu.

Zou, J. (1997). Promoting national informationalization. *China Electronics Daily,* (September 16).

KEY TERMS

Digital Library: A cultural infrastructure that collects and stores information in electronic format and supports its users in accessing a large collection of information effectively through digital means.

Information Resource: A collection of valuable information generated by human activities. In a broader sense, it also includes related equipment, personnel, and capital.

Information Resources Development: The process of collecting, processing, storing, disseminating, and utilizing information resources according to the need of the society.

Information Resources Management: The planning, organization, allocation, and utilization of information and related resources through legal, technological, and other methods to support institutional goals and missions.

Information Service: The activity of providing information products and related services according to users' needs. In a broader sense, it refers to providing users with information through any forms of product or service.

Informationalization: The process of social advances in which human society transforms from industrial society to information society.

Information Systems and Small Business

M. Gordon Hunter
University of Lethbridge, Canada

INTRODUCTION

The subject area of the application of information systems to small business is a thoroughly interesting, yet relatively under-researched topic. Small business is an important part of any economy. In the United Kingdom, 25% of the gross domestic product is produced by small business, which employs 65% of the nation's workers (Ballantine et al., 1998). In Canada, 43% of economic output is accounted for by small business, employing 50% of private sector employees (Industry Canada, 1997). Further, governments view the small business sector as that component of the economy that can best contribute to economic growth (Balderson, 2000). Given the importance of this sector of the economy, it is incumbent upon researchers and managers of small business to develop a better understanding of how information systems may contribute to the operation and growth of individual businesses as well as the overall sector.

The objective of this article is to provide an overview of information systems used by small business. Research projects are presented that describe the current situation. Recommendations are then proffered for various stakeholders who should contribute to a more effective use of information systems by small business.

BACKGROUND

There does not seem to be a commonly accepted definition of a small business. Thus, individual researchers have adopted a definition for their specific projects. Some definitions include annual revenue, amount of investment, or number of employees. The definition mostly used is number of employees (Longnecker et al., 1997). The European Parliament (2002) has also adopted number of employees as a definition and has further refined the category. Thus, 0 to 10 employees represent micro businesses, small businesses include 10 to 50 employees, and medium businesses have 50 to 250 employees.

Beyond the size aspect of small business, there are others that differentiate them from large businesses. Stevenson (1999) has determined that from a strategic perspective, managers of small businesses tend to respond to opportunities presented by their environment in a multi-staged approach by committing a minimum of resources. Another differentiating factor is "resource poverty" (Thong et al., 1994). This term refers to the lack of time, finances, and human resources.

Laudon and Laudon (2001) suggest that an information system is "interrelated components working together to collect, process, store, and disseminate information to support decision making, coordination, control, analysis, and visualization in an organization" (Laudon & Laudon, 2001, p. 7). As indicated previously, managers of small businesses emphasize short-term decisions in their allocation of scarce resources. However, most information systems require a long-term plan with a significant one-time initial financial commitment. This conflict may result in inefficient investment in information systems, which in turn may negatively impact the financial situation of the small business.

Recent research has supported the contention that the use of information systems by small business represents a unique approach. For instance, Belich and Dubinsky (1999) and Pollard and Hayne (1998) determined that the issues being faced by small business managers (lack of time, skills, and financial resources) are different than those faced by large business managers. Further, Taylor (1999) investigated the implementation of enterprise software in small businesses and found that neither the businesses themselves, nor the software vendors were fully cognizant of the unique problems (matching system capability to functional requirements) encountered by small business managers. Finally, Hunter et al. (2002) identified two major themes regarding small business use of information systems. These themes are "dependency" and "efficiency". The authors suggest that the adoption of information systems increased the small business' dependency on an internal champion, and a series of external stakeholders, including consultants and suppliers. Hunter et al. (2002) suggest this increased dependency results from the approaches to business (Stevenson, 1999) taken by the manager and the concept of resource poverty (Thong et al., 1994). The efficiency theme suggests that small business managers primarily use information systems as an operational tool to help complete daily activities.

Earlier research (Nickell & Seado, 1986) determined that small business was mainly using information systems for accounting and administrative purposes. Research conducted in the 1990s (Berman, 1997; Canadian Federation of Independent Business, 1999; Fuller, 1996; Lin et al, 1993; Timmons, 1999) noted a growing interest by small business in employing information systems for daily operations. While small business has been more than prepared to exploit the use of information systems to support daily operations (El Louadi, 1998), there exists little evidence that they are prepared to employ the technology in a strategic manner (Berman, 1997). Bridge and Peel (1999) determined that small businesses employed computers mainly to support daily operations and tended not to use them to support decision-making or long-term planning. Current research suggests this situation has not changed. For instance, Dandridge and Levenburg (2000) found that information systems were being employed for daily operations and there was little use of computerization for competitiveness aspects such as accessing the Internet.

A number of research projects have identified that small businesses have not adopted Internet use because of lack of knowledge and experience (Damsgaard & Lyytinen, 1998; Iacovou et al., 1995; Kuan & Chau, 2001). Another set of contributing factors relates to the lack of personnel and time (Bennett et al., 1999). Even when time and personnel are available, there seems to be reluctance by small businesses to investigate the use of the Internet (Chapman et al., 2000).

Burgess and Trethowan (2002) examined the use of Web sites by small businesses, represented by general practitioners, in Australia. They found that while there was reasonably high use of computers to improve efficiency and lower costs, there was not much use of computers for Web sites. Those who had Web sites mainly employed them to provide basic information and contact details.

FUTURE TRENDS

This section presents a number of recommendations for various stakeholders intent on increasing the use of information systems within the small business sector. These recommendations represent suggestions for the future and reflect a synthesis of available literature, presented in the previous section, in the area in conjunction with, and in the specific context of, previous research (Hunter, 2002; Hunter & Long, 2003;Hunter et al., 2002).

Small Business Manager

To overcome the limitations of being dependent upon others' expertise, managers need to gain an understanding of the capabilities of information systems. While managers do not need to know how to design or develop information systems, they do need to understand how technology might be used as a key resource in adding value to the firm's core business products or services.

Further, the small business manager should establish a relationship with a specific individual regarding a source for advice. The recommendation is for the manager to establish a relationship with someone who is independent of a specific solution and who will be prepared to play a strategic role, taking a long-term perspective. It is incumbent upon the manager to review the relationship to ensure that the recommendations being proffered are appropriately contributing to the long-term success of the firm.

Also, managers should take a proactive approach toward the adoption of information systems. This would involve actively seeking out ways to leverage information systems to create or improve products or services offered to customers. However, the manager should avoid being an early adopter of new hardware and/or software applications.

Consultant

It is important for consultants to recognize that in regard to the nature, timing and acquisition of resources, the small business manager generally aims to minimize the amount of resources used at each stage of the firm's growth. Generally, consultants need to be able to provide opportunities for small businesses to "phase in" information systems in stages. Doing so will accommodate small business practice and form the foundation for a mutually beneficial ongoing relationship.

Vendor

Vendors should make a visible commitment to small business through the establishment of an entity specifically directed at small business. The small business sector is a large and important one; thus target marketing this sector makes good business sense. Also, software vendors must ensure that an application performs the necessary functions for small business. It is incumbent upon the vendor to ensure the hardware or software addresses the appropriate functionality of the small business.

Government

Government can help overcome resource poverty by providing advice and financial incentives to small business. By initiating relationships with small business managers, individuals representing government services can more effectively support these managers by tapping into their informal networks to exchange required information.

The role of tax and other financial incentives may be employed to encourage the expanded use of information systems.

CONCLUSION

Information systems have increased the efficiency of daily operations for small business.

Generally, evidence suggests that information systems expenditures are being made in reaction to needs or problems, rather than as a result of a long-term coordinated plan. The suggestions included here should contribute to more effective use of information systems by small business.

REFERENCES

Balderson, D.W. (2000). *Canadian entrepreneurship and small business management*. Toronto: McGraw-Hill Ryerson.

Ballantine, J., Levy, M., & Powell, P. (1998). Evaluating information systems in small and medium-sized enterprises: Issues and evidence. *European Journal of Information Systems, 7*, 241-251.

Belich, T.J., & Dubinsky, A.J. (1999, Fall). Information processing among exporters: An empirical examination of small firms. *Journal of Marketing Theory and Practice, 7*(4), 45-58.

Bennett, J., Polkinghorne, M., Pearce, J., & Hudson, M. (1999, April). Technology transfer for SMEs. *Engineering Management Journal,* 75-80.

Berman, P. (1997). *Small business and entrepreneurship*. Scarborough, Ontario: Prentice Hall.

Bridge, J., & Peel, M.J. (1999, July-September). A study of computer usage and strategic planning in the SME sector. *International Small Business Journal, 17*(4), 82-87.

Burgess, S., & Trethowan, P. (2002). GP's and their Web sites in Australia: Doctors as small businesses. *Proceedings of ISOneWorld Conference,* Las Vegas, NV.

Canadian Federation of Independent Business.(1999). Results of members' opinion surveys #37-42. Retrieved August 29, 2000, from *http://www.cfib.ca/research/98internet.asp*

Chapman, P., James-Moore, M., Szczygiel, M., & Thompson, D. (2000). Building Internet capabilities in SMEs. *Logistics Information Management, 13*(6).

Damsgaard, J., & Lyytinen, K. (1998). Contours of diffusion of electronic data interchange in Finland: Overcoming technological barriers and collaborating to make it happen. *Journal of Strategic Information Systems, 7,* 275-297.

Dandrige, T., & Levenburg, N.M. (2000, January-March). High-tech potential? An exploratory study of very small firms' usage of the Internet. *International Small Business Journal, 18*(2), 81-91.

El Louadi, M. (1998). The relationship among organizational structure, information technology and information processing in small Canadian firms. *Canadian Journal of Administrative Sciences, 15*(2), 180-199.

European Parliament. Retrieved July 6, 2002, from *www.europarl.eu.int/dg4/factsheets/en/4_14_0.htm*

Fuller, T. (1996). Fulfilling IT needs in small businesses: A recursive learning model. *International Journal of Small Business, 14*(4), 25-44.

Hunter, M.G. (2002). Information systems development outcomes: The case of song book music. In S. Burgess (Ed.), *Managing information technology in small business: Challenges and solutions* (ch. 3). Hershey, PA: Idea Group Publishing.

Hunter, M.G., Diochon, M., Pugsley, D., & Wright, B. (2002). Unique challenges for small business adoption of information technology: The case of the Nova Scotia Ten. In S. Burgess (Ed.), *Managing information technology in small business: Challenges and solutions* (ch. 6). Hershey, PA: Idea Group Publishing.

Hunter, M.G., & Long, W.A. (2003). Adopting the entrepreneurial process in the study of information systems and small business. In G. Gingrich (Ed.), *Managing information technology in government, business, and communities* (ch. 1). Hershey, PA: IRM Press.

Iacovou, C., Benbasat, I., & Dexter, A. (1995). Electronic data interchange and small organizations: Adoption and impact of technology. *MIS Quarterly, 19*(4), 465-485.

Industry Canada. (1997). *Your guide to government of Canada services and support for small business: Trends and statistics* (Catalogue No. C1-10/1997E). Ottawa: Canadian Government Publishing Centre.

Kuan, K., & Chau, P. (2001). A perception-based model of EDI adoption in small businesses using a technology-organized environment framework. *Information and Management, 38,* 507-521.

Laudon, K.C., & Laudon, J.P. (2001). *Essentials of management information systems – organization and tech-*

nology in the networked enterprise (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Lin, B., Vassar, J., & Clack, L. (1993). Information technology strategies for small business. *Journal of Applied Business Research*, *9*(2), 25-29.

Longnecker, J., Moore, C., & Petty, J. (1997). *Small business management*. Cincinnati: South-Western College Printing.

Nickell, G., & Seado, P. (1986). The impact of attitudes and experiences on small business computer use. *American Journal of Small Business*, *101*, 37-48.

Pollard, C., & Hayne, S. (1998). The changing faces of information systems issues in small firms. *International Small Business Journal*,*16*(3), 70-87.

Stevenson, H.H. (1999). A perspective of entrepreneurship. In H.H. Stevenson, H.I. Grousebeck, M.J. Roberts & A. Bhide (Eds.), *New business ventures and the entrepreneur* (pp. 3-17). Boston: Irwin McGraw-Hill.

Taylor, J. (1999). Fitting enterprise software in smaller companies. *Management Accounting*, *80*(8), 36-39.

Thong, J., Yap, C., & Raman, K. (1994). Engagement of external expertise in information systems implementation. *Journal of Management Information Systems*, *11*(2), 209-223.

Timmons, J.A. (1999). *New venture creation* (5th ed.). Boston: Irwin McGraw-Hill.

KEY TERMS

Dependency: The requirement to rely upon another individual or organization.

Effectiveness: The ability to accomplish a task with fewer errors.

Efficiency: The ability to accomplish a task with few resources.

Information Systems: Interrelated components working together to collect, process, store, and disseminate information to support decision-making, coordination, control, analysis, and visualization in an organization.

Internal Champion: Highly respected individual within the organization who possesses expertise in a specific area, specifically information systems.

Resource Poverty: The lack of time, finances, and human resources.

Sales/Revenue: Receipt of income for the exchange of goods or services.

Small Business: Various categories may be employed to define this term. Some examples are as follows:

Employees:

- Micro business: 0 – 10 employees
- Small business: 10 – 50 employees
- Medium business: 50 – 250 employees

Stakeholder: An independent party who may have the ability to impact another party.

Strategy: A business plan of action.

Information Systems and Systems Theory

Ovsei Gelman
National Autonomous University of Mexico (UNAM), Mexico

Manuel Mora
Autonomous University of Aguascalientes (UAA), Mexico

Guisseppi Forgionne
University of Maryland Baltimore County, USA

Francisco Cervantes
National Autonomous University of Mexico (UNAM), Mexico

INTRODUCTION

The information systems (IS) field has been recognized as a scientific discipline since the 80's, as indicated by: (i) the existence of an intellectual community related with doctoral programs and research centers around the world that generates scientific knowledge and solves practical problems using standard scientific procedures accepted and regulated by this community, and (ii) the diffusion of scientific knowledge related with IS through research outlets and research conferences under a rigorous peer-based review process.

Nonetheless, the discipline of information systems has been critiqued by: (i) the lack of formal theories (Farhoomand, 1987, p.55); (ii) the scarce utilization of deductive and formal (e.g., logical-mathematical) research models and methods (idem, p.55); and (iii) the lack of a formal and standard set of fundamental core well-defined concepts associated with the central object of study in this discipline (Alter, 2001, p.3; Banville & Landry, 1989, p.56; Wand & Weber; 1990, p.1282). Consequently, a common-sense language based on informal, conflicting and ambiguous concepts is used as the communicational system in this discipline (Banville & Landry, 1989), and this approach hinders the development of a cumulative research tradition and delays the maturation of the field (Farhoomand, 1987; Wand & Weber, 1990). Furthermore, a deep examination (Mora, Gelman, Cervantes, Mejia, & Weitzenfeld, 2002) of definitions of the term information system, reveals that fundamental concepts are based on few and misused core concepts from the theory of systems (Ackoff, 1960, 1971) and the few formalization proposals (Alter, 2001; Mentzas, 1994; Wand & Weber, 1990) are incomplete.Therefore, the reduction of the lack of formalization of the core concepts used in the IS discipline becomes a relevant and mandatory research purpose. This article contributes to the IS literature with the adaptation and extension of previous formal definitions reported of the terms *system* (Ackoff, 1971; Gelman & Garcia, 1989) and *organization* (Mora, Gelman, Cervantes, Mejia, & Weitzenfeld, 2002) based on the core principles from the Theory of Systems and with the proposal of a formal definition of the term *information systems*. The article also examines the implications for IS research and practice.

BACKGROUND

The term *information system (IS)* has been widely defined in textbooks. Table 1 shows a sample of the main definitions posed in the literature. An examination of these definitions suggests that the IS notion: (i) lacks fundamental standardized and formal concepts (Alter, 2001); (ii) lacks competitive formal macro-structures to cumulate

Table 1. A sample of informal definitions of "what is an information system"

Definition	Reference
"An IS is a system composed of subsystems of hardware, programs, files and procedures to get a shared goal."	(Senn, 1989, p.23)
"An IS is a system composed of application software, support software, hardware, documents and training materials, controls, job roles and people that uses the software application".	(Hoffer, George & Valacich, 1996, p.8)
"An IS is a system composed of inputs, models, outputs, technology, data bases and controls."	(Burch & Grudnitski, 1989, p.58)

theories (Farhoomand & Drury, 2001, p. 14); and (iii) has an excessive variety of micro-theories (Barkhi & Sheetz, 2001).

There have been few, if any, efforts to formalize the discipline. Despite attempts to reduce ambiguity, the proposals (Alter, 2001; Wand & Weber, 1990) have been underpinned on partial views – e.g., syntactical and structural perspectives that hide core semantic information – of what is a *system* (Mora, Gelman, Cervantes, Mejia, & Weitzenfeld, 2002; Sachs, 1976). Others (Mentzas, 1994) offer a more articulated definition than exhibited on Table 1 – by the identification of five subsystems and their functional properties – still lack formalization due to they were developed using a common-sense language critiqued in the IS literature (Banville & Landry, 1989). Therefore, the concept *information system* has still multiple meanings. A systems-based research stream (Alter, 2001; Mora, Gelman, Cervantes, Mejia, & Weitzenfeld, 2002; Paton, 1997) combined with an ontological perspective (Wand & Weber, 1990) suggest that formal foundations from the Theory of Systems (Xu, 2000, pp.113) can reduce this ambiguity and strengthen the rigor that a scientific discipline requires to mature and simultaneously to be relevant and useful for practitioners.

MAIN THRUST OF THE ARTICLE

Formalization reported in this article is adapted and extended from previous work by the authors on the formal concepts of *system* (Gelman & Garcia, 1989) and *organization* and *business process* (Mora, Gelman, Cervantes, Mejia, & Weitzenfeld, 2002). This conceptual development follows a ontological path to define primitive concepts and postulates to derive updated definitions of the constructs *system-I, system-II, general-system, organization, business process* and finally *information system*. A similar approach was used by Wand and Weber (1990) and Wand and Woo (1991) to define what is an *information system* and what is an *organization*.

Formal Definition of System-I. An object of study **X,** formalized as ***system-I*** and denoted as $\mathbf{S_I(X)=<B(X),RB(X),E(X)>}$, is a whole **X** that fulfills the following conditions: (I.1) it has a *conceptual structure* **§(X)** that defines its set of *attributes* **B(X)**, its set of *events* **E(X)** and its set of *range of attributes* **RB(X)**; (I.2) for any subset **B'(X)** of *attributes* of **B(X)**, the set of *events* **E(X)** associated with **B(X)** differs in at least one element from the set of *events* **E'(X)** associated with **B'(X).**

Therefore, to define a situation of study as a *system-I* implies to specify $\mathbf{S_I(X) = <§(X)> =<B(X), E(X), RB(X)>}$ and to fulfill the condition (I.2).

Formal Definition of System-II. An object of study **X,** formalized as ***system-II*** and denoted as $\mathbf{S_{II}(X)=<C_X, \Re_S(C_X')>}$ is a whole **X** that fulfills the following conditions: (II.1) the whole **X** is a set $\mathbf{C_X}$ of elements $\mathbf{X_1, X_2, ..., X_k}$, called *subsystems*, where each $\mathbf{X_i}$ for i=1,2,..., k can be formalized as $\mathbf{S_I(X_i)}$ or $\mathbf{S_{II}(X_i)}$; (II.2) there is a collection finite $\mathbf{\Re_S(C_X')}$ of *set-relations* where $\mathbf{\Re_S(C_X')=\{\Re_1(C_X'), \Re_2(C_X'),\ldots\}}$ on the set $\mathbf{C_X'=\{C, S_I(X)\}}$ and where each *set-relation* $\mathbf{\Re_p(C_X')=\{ {}_p\Re_1, {}_p\Re_2, \ldots \mid {}_p\Re_n = < X_i, {}_ia_j, X_j > or\ {}_p\Re_n = < X_i, {}_ia_X, S_I(X)> or\ {}_p\Re_n = < S_I(X), {}_Xa_j, S_j>}$ and ${}_i\mathbf{a}_j$ stands by the output-input parameters or acts between the two elements}; and (II.3) exists at least a *non-directed-path* among two any items X_i and X_j in the *set-relation* $\mathbf{\Re_S(C_X')}$.

It must be noted that: (i) condition II.3 assures that for any two elements X_i and X_j in the multi-digraph **X**, X_i is reachable from X_j and vice versa; (ii) it is a recursive definition to let a *subsystem* has *subsystems;* and (iii) this definition updates previously reported by authors to consider the output/input relationships between any *subsystem* and the whole *system*. Therefore, to define a situation of study as a *system-II* implies to specify: $\mathbf{S_{II}(X)=<C_X, \Re_S(C_X')>}$ where $\mathbf{C_X=\{ S_I(X_i)\ or\ S_{II}(X_i) \}}$ for i = 1,2, .., k; $\mathbf{\Re_S(C_X')=\{\Re_1(C_X'), \Re_2(C_X'), \ldots\}}$ and the fulfillment of the condition II.3.

Formal Definition of System as General-System. An object of study **X,** formalized as ***general-system*** and denoted as $\mathbf{S_G(X)}$, is a whole **X** that can be defined simultaneously as a *system-I* $\mathbf{S_I(X)}$ and as a *system-II* $\mathbf{S_{II}(X)}$.

Postulate 1. Any *general-system* $\mathbf{S_G(X)}$ defined as *system-I* $\mathbf{S_I(X)}$ can be mapped to a *system-II* $\mathbf{S_{II}(X)}$ and vice versa.

Auxiliary Definition 1. Suprasystem. A whole **SX** is called the ***suprasystem of a system X*** and it is denoted as **SS(X)** if (IV.1) the whole **X** is a *subsystem* of **SX**; and (IV.2) **SX** can be formalized as $\mathbf{S_I(SX)}$ or $\mathbf{S_{II}(SX)}$.

Auxiliary Definition 2. Envelope. A whole **EX** is called the ***envelope of a system X*** and it is denoted as **EE(X)**, if (V.1) the whole **EX** is the *suprasystem* of the *suprasystem* of **X**; and (V.2) **EX** can be formalized as $\mathbf{S_I(SX)}$ or $\mathbf{S_{II}(SX)}$.

Auxiliary Definition 3. Environment. A whole **WX** is called the ***environment of a system X*** and it is denoted as **W(X)**, if (VI.1) **WX** can be formalized as $\mathbf{S_I(WX)}$ or $\mathbf{S_{II}(WX)}$ and (VI.2) **W(X)={SS(X), EE(X)}.**

Postulate 2. Any *general-system* $\mathbf{S_G(X)}$ has a *suprasystem* **SS(X)** and an *envelope* **EE(X)**.

The first formal definition of the concept *system* – for example, *system-I* – accounts for the conception of an external view that sees the *system* as a single-unit with special characteristics – called, *attributes* – and potential acts to execute – called, *events*. In turn, the second formal definition – for example, *system-II* – represents the more usual view – for example, the internal view – that sees the system as a digraph. Furthermore, the definitions of the *set-relations* $\mathbf{\Re_1(C_X'), \Re_2(C_X'), ..., \Re_m(C_X')}$ consider the *system* as a multi-digraph instead of digraph and therefore eliminates some limitations of classic digraph-alike definitions critiqued in the systems theory's literature (Sachs, 1976). Auxiliary definitions and the second postulate help to support the expansionist systemic perspective that indicates that every system always belongs to other larger system (Ackoff, 1971).

Formal Definition of Business Process as System. An object of study **X** is called a ***business process*** and denoted as **BP(X)** if satisfies the following conditions: (VII.1) **X** can be defined as a *system-II* $\mathbf{S_{II}(X)=<C_X, \mathfrak{R}_S(C_X')>}$ where $\mathbf{C_X =\{S_G(CSS), S_G(OSS), S_G(ISS)\}}$ are respectively called the control, operational and information subsystems; and $\mathbf{\mathfrak{R}_S(C_X')= \mathfrak{R}_S(X, S_I(X)) =\{\mathfrak{R}_1(C_X'), \mathfrak{R}_2(C_X'), \ldots, \mathfrak{R}_{12}(C_X') \mid \mathfrak{R}_1(C_X')=\{ {}_1\mathfrak{R}_1 = < X_{CSS}, {}_{CSS}a_{OSS}, X_{OSS}>\}, \ldots, \mathfrak{R}_{12}(C_X')=\{ {}_{12}\mathfrak{R}_1 = <X_{CSS}, {}_{CSS}a_X, S_I(X)> \} \}}$; and (VII.2) $\mathbf{\mathfrak{R}_S(X, S_I(X))}$ satisfies condition (II.3).

It must be noted that the *item-relations* $\mathfrak{R}_{S1}$ to $\mathfrak{R}_{S12}$ take into account at least all possible interrelationships between any two subsystems and of them with the whole system **BP(X)**.

Formal Definition of Organization as a System. An object of study **X** is called an ***organization*** and denoted as **O(X)** if satisfies the following conditions: (VIII.1) **X** can be defined as a *system-II* $\mathbf{S_{II}(X)=<C_X, \mathfrak{R}_S(C_X')>}$ where $\mathbf{C_X =\{S_G(X.1), S_G(X.2), \ldots S_G(X.k)\}}$ and $\mathbf{\mathfrak{R}_S(C_X')= \mathfrak{R}_S(X, S_I(X))}=\{\mathfrak{R}_1(\mathbf{C_X}'), \mathfrak{R}_2(\mathbf{C_X}'), \ldots, \mathfrak{R}_P(\mathbf{C_X}') \mid$ for N=1,2, … , p, each set-relation $\mathfrak{R}_N(\mathbf{C_X}')$ has item-relations ${}_N\mathfrak{R}_1$, ${}_N\mathfrak{R}_2$, ${}_N\mathfrak{R}_3$… of the format $< X.i, {}_{.i}a_j, X.j>$ or $< X.i, {}_{.i}a_X, S_I(X)>$ or $< S_I(X), {}_{.X}a_j, X_j>$ and ${}_\mathbf{i}\mathbf{a}_\mathbf{j}$ stands by the output-input parameters or acts between the two elements X.i and X.j }; (VIII.2) $\mathbf{\mathfrak{R}_S(X, S_I(X))}$ satisfies condition (II.3); and (VIII.3) for j=1,2, … , k either $\mathbf{S_G(X.j)= BP(X.j)}$ or $\mathbf{S_G(X.j)= S_{II}(X.j)=<C_{X.j}, \mathfrak{R}_S(C_{X.j}')>}$ where $\mathbf{C_{X.j} =\{BP(X.j.1), BP(X.j.2), \ldots BP(X.j.n)\}}$ and $\mathbf{\mathfrak{R}_S(C_{X.j}')}$ exists.

Therefore, a *business process* can be considered a *system* with three *subsystems*: control, operational and informational. In turn, an *organization* is a *system* composed of at least two *general-systems* that in turn can be either a *business process* or a *system-II* composed by *business process*. This general definition offers modeling freedom according to the scope of a specific research. In this article, we will use a cybernetic management perspective of organizations that suggests that any organization is composed by two *subsystems* called the *driving* or *management* and the *driven* or *productive subsystems*, respectively (Gelman & Negroe, 1981, 1982). *Management subsystem* can be composed of the planning, decision-making, financial and controlling subsystems, and the *productive subsystems* in turn can be composed by manufacturing, logistic, marketing, and sales/post-sales *subsystems*. However, it must be noted that general definitions of what is an *organization* can adopt other management perspectives such as Porter's added-value chain. In that case, it would be necessary to consider three *subsystems*: of the primary activities, of the support activities and of the top management activities. With these previous antecedents, the formalization of what is an *information system* is straightforward.

Formal Definition of *Information System* as System. An object of study **X** is called an ***Information Systems*** and is denoted as **IS(X)** if fulfills the following conditions: (IX.1) **X** is a *system-II* $\mathbf{S_{II}(X)=<C_X, \mathfrak{R}_S(C_X')>}$ where **$\mathbf{C_X}$={$\mathbf{S_G}$(IT="Information Technology"), $\mathbf{S_G}$(IR=" Information Resources"), $\mathbf{S_G}$(IW="Information Workers"), $\mathbf{S_G}$(IP="Information Procedures") }** and $\mathbf{\mathfrak{R}_S(C_X')= \mathfrak{R}_S(X, S_I(X))=\{\mathfrak{R}_1(C_X'), \mathfrak{R}_2(C_X'), \ldots, \mathfrak{R}_{32}(C_X') \mid \mathfrak{R}_1(C_X')=\{ {}_1\mathfrak{R}_1 = < IT, {}_{IT}a_{IR}, IR> \}, \ldots, {}_{32}\mathfrak{R}_1(C_X')= <IR, {}_{IR}a_X, S_I(X)> \} \}}$; (IX.2) $\mathbf{\mathfrak{R}_S(X,S_I(X))}$ satisfies the condition (II.3); and (IX.3) the *suprasystem* **SS(IS(X))** is a *business process* **BP(Y)** and the **IS(X)** corresponds to the information *subsystem* $\mathbf{S_G}$**(ISS)** of the *business process* **BP(Y).**

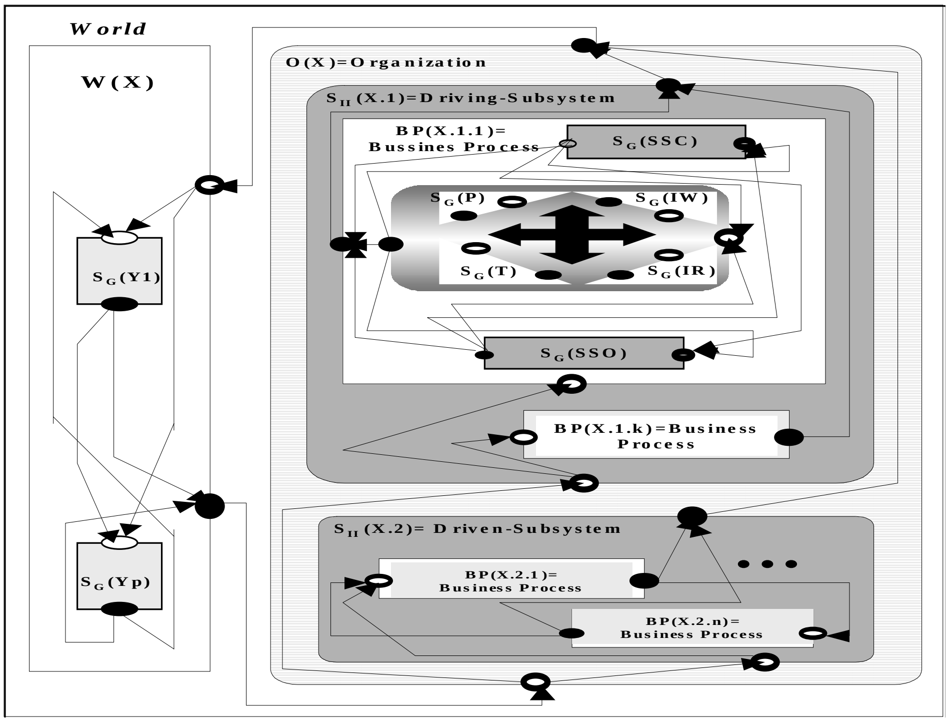
Figure 1 presents a visualization of all formal concepts reported. Due to space limitations, all notation of the *item-relations* of all *systems-II* are hidden. Also, all *item-relations* into the *subsystems* of the *information system* are represented by the four arrows grouped in a single shape. The formal definition posed indicates that an *information system*, as a *system*, is a part of a larger *system*, that is, a *business process system*, that in turn, is the *suprasystem* of the *IS* denoted as SS(IS). Full dots represent the concentration of the subsystems' outputs towards the environment of the system. In turn, clear dots represent the concentration of the environment's inputs into the system. Therefore full and clear dots are useful to take into account the dual specification of any system either as *system*-I – for example, a monolithic view – or as *system-II* – for example, a multi-digraph view.

FUTURE TRENDS

We claim that the proffered formal definition offers a flexible framework to model and specify an *information system* in the level of detailed demanded by the researchers or practitioners. Four possible applications are briefly presented under a theoretical and practice perspective: (i) formal integration of the common-sense concepts used into the informal definitions of what is an *information system*; (ii) elimination of the ambiguity caused by the "Siamese Twin Problem" (Alter, 2001, pp. 30-31); (iii) analysis of failures on information systems through the systemic classification of problems; and (iv) generation of conceptual and dynamic simulation models of information systems.

(i) The formal approach offers a flexible framework consistent with previous definitions of an *IS*. For example, Table 2 exhibits a straightforward way to

Figure 1. Diagram of the overall and systemic perspective of "what is an information system"



formalize the informal definition from Hoffer et al. (1996). Consider for instance the specification of the <job roles> potentially performed by <people> in an *IS*. In this case, the S_G(IW) can be modeled as a *system-I* and the <job roles> can be specified as a system's attribute B(IW)={ $_{IW}b_1$=<job roles>} and specify all possible <job roles> by the definition of the range of attributes RB(IW)={ $_{IW}Rb_1$={operational-user; staff-user; executive-user; etc } }. Another modeling alternative is the representation through system's events E(IW)={ $_{IW}e_1$, $_{IW}e2$, ... }, depending on the interest of the researcher or practitioner. Furthermore, some *item-relations* from the formal definition can be useful to model *subsystem's* acts between the subsystems, the whole *IS*, and its environment. Hoffer et al.'s definition of *IS*, does not account for the information component per se, even though such a component could be considered as the main output of the Hoffer et al.'s IS. Using the formal construct, this component can be modeled as a *subsystem* S_G(IR). In turn, S_G(IR) can be modeled as a *system-I* with system*'s attributes* B(IR)={ $_{IR}b_1$, $_{IR}b_2$, ...}, such as quality, clarity, accuracy, accessibility, appropriateness, among others (Burch, Strater, & Grudnitski, 1979).

(ii) Systemic properties of *organization* and *hierarchy* mapped into the concept of *IS* are useful to resolve the "Siamese Twin Problem". This problem involves the fallacy of studying *IS* without studying the *work system*. The *hierarchy* property of *systems* and the *expansionism approach* suggest that *IS* should not be studied in isolation. Instead, an *IS* should be viewed as part of a larger *system*, that is, the *business process system* or Alter's *work system* concept. Moreover, our formal definition is in agreement with the synthetic thinking view that the behavior of every system can be best understood when studied in relationship to its *suprasystem*, as established by the condition (IX.3).

(iii) Potential deep explanations of failures of *IS* implementations or studies of effectiveness of IS can be theoretically supported through the classification of problems according to the theory of systems. Ackoff (1973, p.666) suggests that failures on *systems* are caused by the conflicts between the system's purpose and the environment's purpose, the system's purpose and the subsystems' purposes and the systems with itself, called respectively the environmentalization, humanization and self-control problems. Formal constructs of *systems* and *IS*, and with conceptual simple tools such Figure 1 and Table 2, can help researchers and practitioners to acquire a holistic perspective of the situation under study and to recognize the influence of the interaction between the systems' parts and the whole

Table 2. Example of formalization of Hoffer et al.'s (1996) definition of information systems

<table>
<tr><td rowspan="14">W(X)</td><td rowspan="9">O(X) =
Organization</td><td rowspan="8">S_{II}(X.1)=
Driving-Subsystem</td><td rowspan="6">BP(X.1.1)</td><td>S_G(CSS)</td><td></td></tr>
<tr><td>S_G(OSS)</td><td></td></tr>
<tr><td rowspan="4">S_G(ISS)=
Information Systems</td><td>S_I(IT)=< B(IT)={S&H support}, E(IT), RB(IT ></td></tr>
<tr><td>S_I(IR)= < B(IR), E(IR), RB(IR) ></td></tr>
<tr><td>S_I(IW)=< B(IW)={people; job roles }, E(IW), RB(IW) ></td></tr>
<tr><td>S_I(IP)= < B(IP)={documents; training materials; controls}, E(IP), RB(IP) ></td></tr>
<tr><td>…</td><td colspan="2"></td></tr>
<tr><td>BP(X.1.k)</td><td colspan="2"></td></tr>
<tr><td rowspan="3">S_{II}(X.2)=
Driven-Subsystem</td><td>BP(X.2.1)</td><td colspan="2"></td></tr>
<tr><td></td><td>…</td><td colspan="2"></td></tr>
<tr><td></td><td>BP(X.2.n)</td><td colspan="2"></td></tr>
<tr><td>S_G(Y1)</td><td colspan="4"></td></tr>
<tr><td>…</td><td colspan="4"></td></tr>
<tr><td>S_G(Yp)</td><td colspan="4"></td></tr>
</table>

within its environment. Furthermore, since the formalisms posed can be extended through the addition of parameters in the set of *a-actions* in the *item-relations*, it is possible to model social, technical and political features through the specification of *attributes* and *events* assigned to *subsystems* in the *system's environment* W(BP(X)), where BP(X) represents the *suprasystem* of the *IS* under study. For example, soft features as <top management support>, <environmental hostility>, <environmental dynamism> and <organizational climate>, once specified and measured, can be easily assigned as subsystems' *attributes* of W(BP(X)).

(iv) Complex conceptual and dynamic-based simulations models of *IS* can be specified. A case in point is the recent development of a conceptual and system dynamic-based simulation model of the implementation process of Decision-Making Support Systems (Mora, Cervantes, Gelman, & Forgionne, 2003, 2004) that generated satisfactory predictive results from several well-known reports in the literature.

In short, this article suggests that the formal definition of the term *information system*, developed from the formal principles of *theory of systems*, enables researchers and practitioners to: (i) avoid ambiguity from informal definitions; (ii) account for practically all informal definitions; (iii) specify and customize a structure of the concept *IS* with the level of detail demanded by the modelers; and (iv) help to build complex systemic models of organizations that use *IS*. However, since formalisms in the information systems field have been scarcely reported, it is strongly recommended that further research be conducted to deploy their benefits and cope their limitations.

CONCLUSION

This article supports the thesis that formalisms – from the theory of systems – are required in the field of information systems to reach the rigor and maturation needed for a respectable scientific discipline. However, gains will not be free since researchers and practitioners could be required to add to his/her conceptual tools the utilization of logical-mathematical models. Perhaps, then, the information systems discipline will no longer be a "fragmented adhocracy" (Banville & Landry, 1989, p.56). Formalisms posed offer a non-trivial way to understand and model an *information systems* and therefore contribute toward the maturation of our field.

REFERENCES

Ackoff, R. (1960). Systems, organizations and interdisciplinary research. *General System Yearbook, 5*, 1-8.

Ackoff, R. (1971).Towards a system of systems concepts. *Management Science, 17*(11), 661-671.

Ackoff, R. (1973). Science in the systems age: beyond IE, OR and MS. *Operations Research, 21*(3), 661-671.

Alter, S. (2001). Are the fundamental concepts of information systems mostly about work systems? *Communication of AIS, 5*(11), 1-67.

Banville, C., & Landry, M. (1989). Can the field of MIS be disciplined. *Communications of the ACM*, *32*(1), 48-60.

Barkhi, R. & Sheetz, S. (2001). The state of theoretical diversity of information systems. *Communication of AIS*, *7*(6), 1-19.

Burch, J.G. & Grudnitski, G. (1989). *Design of information systems*. New York: Wiley & Sons.

Farhoomand, A. (1987). Scientific progress of management information systems. *Database*, Summer, 48-57.

Farhoomand, A. & Drury, D. (2001). Diversity and scientific progress in the information systems discipline. *Communication of AIS*, *5*(12), 1-22.

Gelman, O., & Garcia, J. (1989). Formulation and axiomatization of the concept of general system. *Outlet IMPOS (Mexican Institute of Planning and Systems Operation)*, *19*(92), 1-81.

Gelman, O., & Negroe, G. (1981). Role of the planning function in the organizational conduction process. *Outlet IMPOS (Mexican Institute of Planning and Systems Operation)*, *11*(61), 1-17.

Gelman, O., & Negroe, G. (1982). Planning as organizational conduction process. *Journal of the Mexican National Academy of Engineering*, *1*(4), 235-270.

Hoffer, J., George, J., & Valachi, J. (1996). *Modern systems analysis and design*. Menlo Park, CA: Benjamin/ Cummings.

Mentzas, G. (1994). Towards intelligent organizational information systems. *International Journal of Information Management*, *14*(6), 397-410.

Mora, M., Cervantes, F., Gelman, O., & Forgionne, G. (2003). Theoretical foundations of the systems approach and its application to the study of the dynamic of the implementation process of decision making support systems. Doctoral dissertation. School of Engineering of the National Autonomous University of Mexico, Mexico: UNAM Press.

Mora, M., Cervantes, F., Gelman, O., & Forgionne, G. (2004). Understanding the strategic process of implementing decision making support systems (DMSS): A systems approach. Paper accepted in the *2004 IFIP International Conference on Decision Support Systems (DSS2004)*. Prato, Italy, July 1-3.

Mora, M., Gelman, O., Cervantes, F., Mejia, M., & Weitzenfeld, A. (2002). A systemic approach for the formalization of the information system concept: why information systems are systems? In J. Cano (Ed.), *Critical reflections of information systems: A systemic approach*, (pp. 1-29). Hershey, PA: Idea Group Publishing.

Paton, G. (1997). Information system as intellectual construct –its only valid form. *Systems Research and Behavioral Science*, *14*(1), 67-72.

Sachs, W. (1976). Toward formal foundations of teleological systems science. *General Systems*, *XXI*, 145-154.

Senn, J.A. (1989). *Analysis & design of information systems*. New York: McGraw-Hill.

Xu, L. (2000).The contributions to systems science to information systems research. *Systems Research and Behavioral Science*, *17*, 105-116.

Wand, Y., & Weber, R. (1990). An ontological model of an information system. *IEEE Transactions on Software Engineering*, *16*(11), 1282-1292.

Wand, Y., & Woo, C. (1991). An approach to formalizing organizational open systems concepts. ACM Library Digital. Accessed in March 5, 2002.

KEY TERMS

Attribute (Informal Definition): A substantial feature of a whole that is perceived by an observer with the potential to produce or cause a product or effect.

Event (Informal Definition): An act performed by a whole or to the whole that is perceived by an observer directly or through its consequences on other(s) whole(s).

Subsystem: Any immediate inner system that is subsumed to the system.

Suprasystem: The immediate outer system that subsumes any system.

System (Informal Definition): A whole composed of subsystems, and at the same time included in a suprasystem in such way that some particular properties of the whole and of the subsystems are getting lost when they are considered analytically, for example, by separation of the parts of the whole.

Theory of Systems: A research inquiry paradigm based on an expansionist world view, a synthetic and holistic thought and teleological principles, which makes it suitable and effective for studying complex phenomena.

Information Systems and Technology in South Africa

Anesh Maniraj Singh
University of Kwazulu Natal, South Africa

INTRODUCTION

South Africa, like most of its African neighbors, has a dual economy that sees formal and informal trading taking place side by side. Walking down many of the main streets of South African cities, one can immediately see the informal traders conducting their trade on the doorsteps of established retailers. Many of the formal traders complain about the informal activity and its impact on their businesses. However, there seems to be a symbiotic relationship that ensures their peaceful coexistence. For many foreigners, there is the perception that South Africa is a jungle with lions and tigers running around the streets and wind-up telephones as a means of communication. This is, however, far from the reality. Technology in the formal economy is almost as good as and sometimes better than it is in the United States and other first-world countries. Cellular telephone technology is one generation ahead of the United States. However, like the economy, a technology divide exists, where some parts of the population have access to technology, while others do not. This article will examine the trends in technology, outlining the use of technology in South African industry. The digital divide will also be discussed, looking at the problems and how the State in partnership with the private sector can bridge the gap.

BACKGROUND

Technology has always been existent in South Africa. However, it has never kept pace with the developments in first-world countries. During the apartheid era, trade sanctions resulted in extremely high costs in acquiring technology, and as such, South African organisations were unable to keep pace with the latest developments in technology in developed countries. However, at present, organisations are capable of implementing the latest systems and technologies as they roll off the production line, as is evidenced by the following scenarios.

The Private Sector

Manufacturing

According to Hodge and Miller (1997), the manufacturing sector spent about R900 million on hardware and a billion Rand on software. In order to improve productivity and manage costs more effectively, manufacturers rely heavily on MRP (materials requirements planning) software (Faull, 1995). Toyota South Africa has one of the most technologically advanced production lines in the country. SAPREF (South African Petroleum Refiners) monitors and controls the entire petroleum refining process from a centralized control room.

Retail

The retail industry has seen an evolution of systems from the basic adding machine to the latest in scanning and optical character recognition systems. Companies such as Pick ’n Pay use their point-of-sale systems not only for calculating and receiving payments but also for EFTPOS (electronic funds transfer at point of sale). Pick ’n Pay is linked electronically to almost all South African banks, allowing customers to draw cash and make account payments to third parties at the supermarket cashier.

Banking

According to the Reserve Bank, the banks and financial services sector spends almost as much on technology as the manufacturing sector. SASWITCH was introduced in the 1980s and linked the various institutions’ ATMs (automated teller machines) on a single network, allowing customers to use any bank’s ATM to access their accounts at another bank. Standard banks were among the pioneers to offer electronic account payments from an ATM. They also formed a relationship with MTN (a cellphone network operator) to allow customers to top up their cell phone airtime from an ATM. All the banks offer Internet banking facilities. Investment houses such as Old Mutual, Sanlam, and Liberty have Web sites where the customer can view his or her investment portfolios, check exchange rates, and conduct secure transactions.

Banks have been moving steadily in developing their offerings electronically, with a long-term vision of reducing branches and services by developing self-service centres. Reducing branches and staff would lower overheads, allowing the banks to become lower-cost providers and enabling them to become more profitable and competitive.

Telephony

Telkom is South Africa's sole provider of fixed-line telecommunications. Most of South Africa was connected via copper wire. However, in 1981, fibre optic was introduced in the then Transvaal. Beltel, a text-based service, was introduced in 1986, allowing subscribers to gain access to banks, entertainment, adult entertainment, and static text-based advertisements. ISDN was introduced in 1994, and the first cellular phone networks were introduced (Highlights of the…, 2004). MTN, Vodacom, and Cell-C are the only cellular service providers. Apart from providing voice communications, they also offer WAP (wireless application protocol), GPRS (general packet radio service), SMS (short message service), and MMS (multimedia messaging service). The cellular phone industry is extremely advanced and also offers value-added services, such as mobile banking. The advances in telephony have come about due to the lifting of trade sanctions imposed during the apartheid era. Accessing the latest technologies from foreign developers as they unfold is much easier than in the past. South African companies have also chosen, in some spheres, to leapfrog certain intermediate developments in technology rather than gradually evolve from one technology to the next. This is why, in certain areas of technology, South Africa is ahead of some Western countries.

However, according to Weidemann (2004a), South Africa is losing its role as Africa's technology leader due to its restrictive information and communications technology (ICT) legislation. He goes on further to state that legislation is affecting new technologies, such as voiceover Internet protocol (VOIP) and wireless fidelity (WIFI). South Africa has been active in attracting international call centres to its shores. However, according to Jones (2004), because of the prohibitively high costs of fibre optic based communication, providers want to use VOIP, which would drive tariffs down and allow South Africa to compete with call centre hubs in India.

Videoconferencing came into being in 1991, and today, 80% of the top 500 companies in the country use it (Weidemann, 2004b). Videoconferencing is being used extensively for meetings across provinces, e-learning and training, and, more recently, for conducting interviews, saving on travel and accommodation costs. The University of Kwazulu Natal uses videoconferencing to conduct meetings across its five campuses.

The Public Sector

Unlike the private sector, the public sector has trailed significantly behind with regard to technology implementation and technology usage. According to Hodge and Miller (1997), it was estimated that seven in 100 government employees had access to computers; this figure has hardly changed, with approximately 16 in 100 employees having access to a PC (Source Withheld). The Home Affairs department that handles birth, identification, travel, and death documents does not have PCs at the counters. Information is recorded manually in books, causing long delays in the event of a query. Furthermore, there is no document management system to track documents sent from regional offices to national offices and vice versa.

In order to improve IT skills, usage, and backup service, SITA, the State Information Technology Agency, was set up as an "outsource" organisation to ensure that proper procedures were developed and implemented in procuring IT, providing end user support, and training, developing, and maintaining information systems. However, there are departments that believe that SITA has not delivered on its mandate due to its being an organ of the State. It is believed that SITA has only benefited central government officers and offices, whereas provincial departments do not enjoy the same service delivery, and local and municipal governments are totally excluded.

Although there are departments that have inadequate information technology (IT) and information systems (IS), there are others that have very sophisticated systems. NATIS (national traffic information system) is a database of all driver details and records of road offences. Traffic officers on the street, armed with laptops can access the database and determine if someone who has been stopped has previous unpaid fines or active warrants of arrest. The Prince Alfred Luthuli Hospital in Durban is the first paperless hospital in the country. All records are stored electronically and can be easily accessed, ensuring proper patient care and administration of medication. The Prince Alfred Luthuli project will be used as a case study and benchmark for the conversion of other state hospitals to paperless institutions.

The use of technology in education displays a digital divide of its own. Public primary and secondary schools use outdated technologies for administration and teaching, such as pentium® 1 computers and software such as

Windows® 95. Pupils taking computer education classes in secondary school are taught Turbo Pascal® which is totally outdated. Private schools, on the other hand, have computerised administrative systems and use the latest technology and programs for teaching purposes. Tertiary institutions are well ahead of schools and use specialized software, such as Integrated Tertiary Software, that tracks a student's progress, fees, and related academic activities from the date of first registration.

University and Technikon academics use IT extensively for research. The Internet is used daily to search for information, university libraries are linked to some of the leading international electronic journals and databases. Database packages are used for capturing research data, and analysis is conducted using some of the latest statistical packages such as SPSS.

Many of the tertiary institutions are partnering with big brands, such as Microsoft, IBM, SAP, Oracle, 3Com, Novell and Cisco among others, creating specialist training academies to provide stand-alone courses and to incorporate some modules into the academic curriculum.

The Internet in South Africa

The Internet is growing rapidly in South Africa. Internet usage in 2001 was a mere 2.4% but has grown to 6.8% in 2003 (Internet usage statistics…, 2004). Initially, the Internet was used mainly by academics. However, this has been rapidly overtaken by e-commerce, both business-to-business (B2B) and business-to-consumer (B2C). B2B commerce was predicted to reach R28.3 billion and B2C to account for R2.5 billion (Goldstuck, 2001). All the retail banks in South Africa have a Web presence, while a number of large retailers have extended their services online. Pick 'n Pay and Woolworths have shopping Web sites and offer deliveries to customers. Pick 'n Pay has not shown a profit from its online store. However, in line with its 5 year strategy, the online losses have been decreasing each year (Mercury, 2004).

Kulula.com, South Africa's first online airline, allows passengers to purchase tickets online in three easy steps. The airline has shown a profit from day one (Scholtz, 2003). The Government has also embraced the Internet by developing an online presence where information in the form of articles, green papers, and white papers are made available for public scrutiny. The South African Revenue Services (SARS) has an interactive site that allows citizens to file taxes online and to effect online payments.

It is evident from the discussion that South Africa is on par with first-world countries in terms of private-sector usage of technology. However, it has also surfaced that a divide exists between public- and private-sector usage of technology.

The Digital Divide in South Africa

The digital divide separates those who have access to technology from those who do not. The term "digital divide" stems from the global information society (GIS), where it is intimated that those people, countries, or organizations that do not have access to the GIS will become marginalised and will not share in the benefits that the Internet brings with it. More than 80% of people around the world have never heard a dial tone, let alone surfed the Web. The gap between the information haves and have-nots is widening (Bridging the Digital Divide, 1999). Developing countries only account for 4.25% of Internet users worldwide. Poverty, illiteracy, lack of infrastructure, and inadequate government interest or intervention has led to the information have-not situation. According to Stillkind (1996), those countries that are lagging behind in access to the Internet are bound to fall further behind to the point of exclusion. Research pertaining to the digital divide attributes its existence to developed and developing countries, but is this truly the case?

Factors that Contribute to the Digital Divide

The digital divide is not only the product of the difference between first and third world standards. The digital divide also exists, among others, gender, physical disability, racial segregation and age.

- *Gender*: In some countries and organisations, females have less access to the Internet than males, but in some instances, males tend to be marginalised as well. In a survey by Lake (1997), it was found that fewer males (38%) had limited access to the Internet than females (41%). This indicates that gender creates a divide between the information haves and the have-nots; however, the context may differ from place to place.
- *Racial segregation*: According to a study by Novak and Hoffman (1998), African Americans differ significantly from whites in their access to computers and web usage. The study revealed that 44% of White Americans had home PCs, whereas only 29% of Blacks had home PCs. This limited the access that Blacks had to the Internet. Similarly, in South Africa, Blacks up to the age of 35, have the least access to computers and the Internet (Saner, 2003).
- *Age*: According to a study carried out by Singh (2001), persons aged 15–24 (45%) used the Internet

daily. Older respondents, especially those in the 45–54-year-old category (27%), used the Internet once a month. It is clear from this study that a digital divide exists between age groups, because youth are more exposed to technology and are willing to use it, whereas older people are resistant to change and avoid the use of technology.

- *Literacy*: A major problem in South Africa is illiteracy (Aitchison, 1998). Due to the low levels of literacy and poor schooling infrastructure, many South Africans are excluded from the world of technology and the Internet.

The digital divide in South Africa is extremely pronounced in that only 4.5% of the total population comprises information haves and 95.5% comprises information have-nots (South African Web Usage Behavior, 2000). The main reason for the digital divide in South Africa is the apartheid legacy that promoted separate development, which provided inferior education and poor or no access to learning opportunities for nonwhites. The existence of the digital divide is also attributed to high levels of poverty, a lack of telecommunications infrastructure, and high costs of connectivity. The foregoing illustrates that a digital divide exists in any situation where there are information haves and have-nots, and it is not limited exclusively to a divide between first- and third-world nations.

Solutions to the Digital Divide

The solutions to the digital divide range from micro interventions to overarching policy developments and investment by the government.

- *Computer recycling*: Many large corporations and institutions upgrade their hardware at least once in 3 years. The old machines should be refurbished and donated to schools for use in the classroom and by administrators.
- *Training and development*: There are many sector education and training authorities (SETAs) that have funds available for training. These funds could be utilized to train the masses of unemployed into computer end users.
- *IT incubators*: The Cape IT Initiative (CITI) is aimed at promoting Cape Town as the hub into Africa (About Citi, 2004). In creating jobs within the IT sector, CITI has developed a number of linkages, both locally and internationally. According to Annecke (2004), Microsoft is helping to develop "digital villages," where locals teach themselves computing and then transmit these skills among the community. Many entrepreneurs have emerged from this initiative. One man opened up a CV writing business based on the skills learned from the village.
- *Public–private partnerships*: The Government does not have the funds or the skills to bridge the digital divide. Therefore, they need to partner with the private sector in order to obtain funding and skills to bridge the gap between the technology haves and the have nots. With the introduction of The Digital Partnership, South Africa has seen the development of over 20 partnerships to deliver e-learning centres to disadvantaged schools (The Government of…, 2002). However, public–private partnerships need not only be for the development of citizens but should also seek to develop the IT skills within Government. Specialists from industry could be engaged to help the Government make strategic IT plans and help them implement these plans.

The Legal Environment

IT law is becoming very competitive and is growing into a discipline of its own, due to the increase in white collar crime, especially those perpetrated using technology (Vecchiatto, 2004). The legal fraternity is becoming more involved in issues such as intellectual property law, privacy, spam, and online contracts.

In order to balance the rights of consumers, employers, and employees using the Internet, the Electronic Communications and Transactions Act (ECTA) was promulgated in 2002. ECTA was developed to establish a formal structure to define, develop, regulate, and govern e-commerce in South Africa (Michaelson, 2003). The Act also covers issues unique to electronic environments, such as domain name registration, cryptography standards, and authentication service providers. Other supporting legislation includes the following: The Access to Information Act that determines what information can and must be kept private, and what can and must be made easily available to people requesting it; The Interception and Monitoring Prohibition Act that determines who is allowed to intercept communications; and The Interception and Monitoring Prohibition Act that makes it illegal for employers to monitor employee usage of Internet facilities at work. However, due to the increase in employee abuse of Internet facilities and electronic industrial espionage, representation has been made to have the Act amended to allow employers to monitor and intercept employee's electronic communications.

It is evident that ICTs are playing a major role in South Africa, hence the interest and commitment to developing standards to regulate electronic activity.

FUTURE TRENDS

As previously mentioned, information systems and technology play a mixed role in South Africa, with the private sector using cutting-edge technology and the public sector lagging behind in the technology and the systems used. Public schools have a major role to play in spreading IT skills and knowledge to the masses, which will require investment and support from Government and the private sector. Apart from the contribution made by Mark Shuttleworth and a few others, South Africans tend to be adopters rather than innovators of information systems and technology; hence, there are no new "killer applications" or products that have emanated from the country. However, most organizations are quick to adopt the best systems and technology from the East and West and implement them very effectively.

Cellular telephony seems poised for great success in the event of a breakthrough business application being developed. SMS is a very profitable application for the cellular networks; however, this platform needs further development for transaction purposes. Wireless technology is finding its way into large corporations, which will make wired LANs obsolete. Hot spots are finding their way in public arenas, such as in the country's international airports and Mugg & Bean Coffee Shops, making Internet connections available to traveling businesspeople among others. Community centres, Internet café's public information terminals (PITS), and digital villages will make technology available even in the most remote rural areas. Bank branches will become smaller, offering limited personal service as customers embrace Internet banking and the use of ATMs.

CONCLUSION

It is evident that like most developing countries, South Africa has a first- and third-world approach to information systems and technology. This has created a digital divide that can only be bridged due to Government intervention in reducing poverty, improving literacy, and providing access to ICT. Furthermore, Government has to develop policies and provide guidelines for the ethical use of technology in society and also provide protection for businesses and individuals engaging in e-commerce. Furthermore, private-sector investment is needed to boost telecommunications and to bridge the digital divide. It is evident that although South Africa is not a technology innovator, the effective implementation of technology strategies has enabled certain sectors, such as banking and telephony, to dominate in the effective use of technology. Mass access and usage of technology, especially computer technology, will only take place when technology becomes more affordable to the masses. Democratic South Africa is undergoing rapid change, and perhaps all that is needed to get information systems and technology up to first-world standards, is time.

REFERENCES

About Citi. (2004). Retrieved March, 2004, from *www.citi.org.za*

Aitchison, J. J. W. (1998). *A review of adult basic education and training in South Africa*. Retrieved July, 2000, from www.fsu.edu/~/vadca/english/adeas.html

Annecke, T. (2004). Entrepreneurs bred in digital villages. *Business Times*. Retrieved March, 2004, from *http://www.btimes.co.za/99/0926/survey/survey06.htm*

Bridging the Digital Divide. (1999). Retrieved July, 2004, from *http://news.bbc.co.uk/1/hi/specialreport/1999/10/99/information_rich_information_poor/46651.stm*

E-Tailers Running at a Loss. (2004). *Natal Mercury*. Independent Newspapers: Durban.

Faull, N. (1995, April). Examining the role of IT in world class manufacturing. IEC Conference.

Goldstuck, A. (2001). E-Commerce potential in South Africa. Business Day. Retrieved March, 2004, from *http://www.sapoaonline.co.za/news/article.aspx?idArticle=978*

Highlights of the Telecommunications History of South Africa. (2004). Retrieved March, 2004, from *http://www.telkom.co.za/news/telecomshistory.jsp*

Hodge, J., & Miller, J. (1997). Information technology in South Africa. The state of the art and implications for national policy. University of Cape Town. Retrieved March, 2002, from *http://news.bbc.co.uk/hi/english/special...nformation_poor/newsid_466000/466651.stm*

Internet Usage Statistics for Africa. (2004). Retrieved January, 2004, from *www.internetworldstats.com/stats1.htm*

James, A. (2004, May 17). Developing new initiatives in the African markets. *Daily News*. Independent Newspapers, Durban.

Lake, A. (1997). Out of the abyss—Surviving the information age. Retrieved July, 2002, from *www.flexibility.co.uk/issues/info-overload.htm*

Michaelson, L. (2003, May). You and the law—The ECT Act uncovered. Windows Southern Africa. Gauteng: Information and Technology Publishing Co. Ltd.

Novak, T. P., & Hoffman, D. L. (1998). Bridging the digital divide: The impact of race on computer access and Internet use. Retrieved July, 2002, from *http://elab.vanderbilt.edu/research/papers/html/manuscripts/race/science.html*

Saner, R. (2003). The effect of the Internet on South African youth in the next 5 years: 2003–2008. *The Edge: The Future*. Retrieved July, 2004, from *http://www.youth.co.za/theedge/the_future/the-future20.asp*

Scholtz, C. (2003). The trials and tribulations of an online airline. Presented at the Fifth World Wide Web Applications Conference, Durban.

Singh, A. M. (2001). The effectiveness of Internet marketing—A management and consumer perspective. Unpublished doctoral thesis, University of Durban Westville, Durban.

South African Web Usage Behavior. (2000). Retrieved November, 2000, from *.co.za/library_marketing.html*

Stilkind, J. (1996). The electronic revolution and developing countries. Retrieved March, 2002.

Weidemann, R. (2004a, May 27). South Africa losing role as Africa's technology leader. *ITWEB News*. Retrieved March, 2004, from *www.afrol.com/articles/12786*

Weidemann, R. (2004b, June 11). South Africa losing role as Africa's technology leader. *ITWEB News*. Retrieved March, 2004, from *www.allafrica.com/stories/printable/200406110621.html*

KEY TERMS

Digital Divide: Refers to the gap between those who have access to the Internet and those who do not.

Electronic Commerce: Refers to the use of electronic technologies, especially the Internet, to conduct buying and selling.

Global Information Society (GIS): Refers to the linking of people using the Internet to share information for the benefit of all societies around the globe.

Information Systems: Refers to a set of people, procedures, and resources used to collect, process/transform, and disseminate information in an organisation.

Information Technology: Refers to the physical components that are needed to create information products; normally refers to the hardware, such as computers, monitors, printers, and other computer components.

Private Sector: The private sector refers to businesses that are engaged in commerce for the sole purpose of making a profit and increasing shareholder wealth.

Public Sector: This refers to the government and its various departments that serve the citizens of a country. These departments are nonprofit-making organs of the State.

Information Technology Strategic Alignment

Fernando José Barbin Laurindo
University of São Paulo, Brazil

Marly Monteiro de Carvalho
University of São Paulo, Brazil

Tamio Shimizu
University of São Paulo, Brazil

INTRODUCTION

Information technology (IT) has assumed an important position in the strategic functioning of leading companies in competitive markets (Porter, 2001). Particularly, e-commerce and e-business have been highlighted among IT applications (Porter, 2001). Two basic points of view can be used for understanding the role of IT: the acquisition of a competitive advantage at the value chain and the creation and enhancement of core competencies (Porter & Millar, 1985; Duhan et al., 2001).

Several problems have been discussed concerning IT project results and effectiveness of their management. Effectiveness, in the context of this article, is the measurement of the capacity of the outputs of an information system (IS) or of an IT application to fulfill the requirements of the company and to achieve its goals, making this company more competitive (Laurindo & Shimizu, 2000; Walrad & Moss, 1993; Maggiolini, 1981; Drucker, 1963).

There is a general consensus about the difficulty in finding evidence of returns over the investments in IT (the "productivity paradox"), even though this problem can be satisfactorily explained (Brynjolfsson, 1998). In order to better use these investments, organizations should evaluate IT effectiveness that allows for the strategic alignment of objectives of implemented IT applications and their results with the company business vision (Laurindo et al., 2003; Laurindo, 2002; Smithson & Hirscheim, 1998).

The comparison and evaluation of business and IT strategies and between business and IT structures must be a continuous process, because the company situation is constantly changing to meet market realities and dynamics.

FINDING STRATEGIC IT APPLICATIONS

Critical success factors (CSF) is a widespread method used for linking IT applications to business goals, and for planning and prioritizing IS projects. This method was proposed by Rockart (1979), although King and Cleland (1977) had suggested a similar idea (critical decision areas) before.

According to this method, the information systems, especially executive and management information systems, are based on the current needs of the top executives. These information needs should focus on the CSF. Rockart defined CSF as the areas where satisfactory results "ensure successful competitive performance for the organization." This author states that CSF prime sources are the structure of the industry, competitive strategy, industry position, geographic location, and environment and temporal factors.

Basically, the CSF method includes the analysis of the structure of the particular industry and the strategy and the goals of the organization and its competitors. This analysis is followed by two or three sessions of interviews with the executives, in order to identify the CSF related to business goals, define respective measures (quantitative or qualitative) for the CSF, and define information systems for controlling CSF and their measures (Carvalho & Laurindo, 2003).

For Rockart, this process can be useful at each level of the company and should be repeated periodically, because CSF can change through the time and also can differ from one individual executive from another.

The CSF method had an important impact on managerial and strategic planning practices, even though it was primarily conceived for information systems design, especially management and executive information systems. Besides the utilization in information systems planning and information systems project management, it has been used in strategic planning and strategy implementation, for management of change, and as a competitive analysis technique (Pollalis & Frieze, 1993). This method leads to a policy-oriented approach by focusing on the essential issues of companies. Furthermore, the continuous measurement of the CSF allows companies to identify strengths and weaknesses in their core areas, processes, and functions (Rockart, 1979; Sullivan, 1985).

More details of the process of implementation of the CSF method can be found in Rockart and Crescenzi (1984) and Martin (1982).

McFarlan (1984) proposed the strategic grid that allows the visualization of the relationship between IT strategy and business strategy and operations. This model analyzes the impacts of IT existent applications (present) and of applications portfolio (future), defining four boxes, with each representing one possible role for IT in the enterprise: "Support," "Factory," "Turnaround," and "Strategic" (Figure 1):

- *Support*: IT has little influence in present and future company strategies.
- *Factory*: Existent IT applications are important for a company's operations success, but there is no new strategic IT application planned for the future.
- *Turnaround*: IT is changing from one situation of little importance ("support" box) to a more important situation in business strategy.
- *Strategic*: IT is very important in business strategy in the present, and new planned applications will maintain this strategic importance of IT in the future.

In order to assess the strategic impact of IT, McFarlan proposed the analysis of five basic questions about IT applications, related to the competitive forces (Porter, 1979):

- Can IT applications build barriers to the entry of new competitors in the industry?
- Can IT applications build switching costs for suppliers?
- Can IT applications change the basis of competition?
- Can IT applications change the balance of power in supplier relationships?
- Can IT applications create new products?

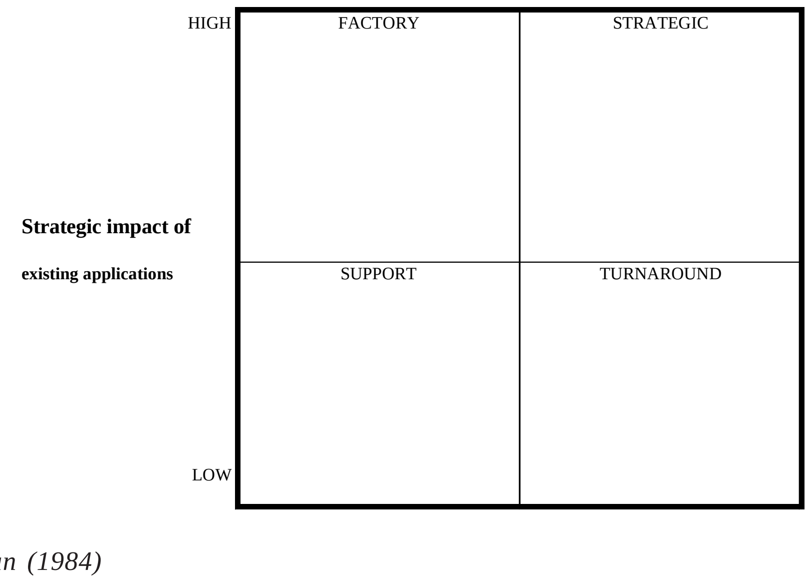
These questions should be answered considering both present and planned future situations.

Thus, IT may present a smaller or greater importance, according to the kind of company and industry operations. In a traditional manufacturing company, IT supports the operations, because the enterprise would keep on operating even when it could not count on its IS. However, in a bank, IT is strategic for business operations, because it is a source of competitive advantage, and a bank cannot operate without their computerized IS.

Henderson and Venkatraman (1993) proposed the Strategic Alignment Model that analyzes and emphasizes the strategic importance of IT in the enterprises. This model is based on both internal (company) and external (market) factors.

Two fundamental concepts in this model are strategic fit (interrelationships between strategy and infrastructure) and functional integration (integration between

Figure 1. Strategic grid of impacts of IT applications



Source: McFarlan (1984)

business and IT, in the strategic and infrastructure aspects). The authors emphasize that strategy should consider both internal and external domains of the company. The internal domain is concerned with the administrative structure of the company. The external domain is concerned with the market and the respective decisions of the company. Thus, according to this model, four factors (that the authors called domains) should be considered for planning IT:

1. Business strategy
2. IT strategy
3. Organizational infrastructure and processes
4. IS infrastructure and processes

The Strategic Alignment Model brings the premise that the effective management of IT demands a balance among the decisions about those four domains above.

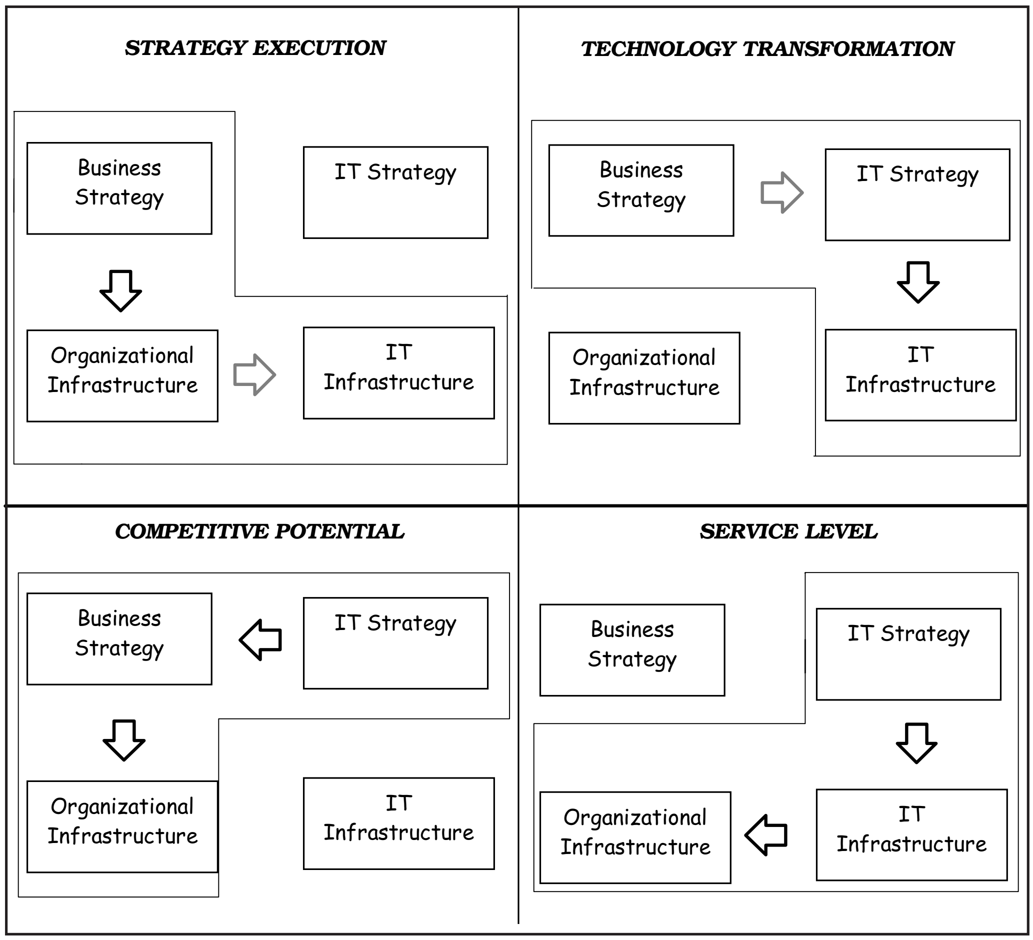
According to Henderson and Venkatraman, there are four main perspectives of strategic alignment, through the combination of the four factors, starting from business strategy or from IT strategy, as shown in Figure 2.

One important innovation of this model is that IT strategy could come first and change business strategy, instead of the usually general belief that business strategy come before IT planning.

This planning should be a continuous process, because external factors are in permanent changing situations. If the company does not follow these changes, it will be at a serious disadvantage in the fiercely competitive market. This is particularly true when a new technology is adopted by almost all companies in an industry, passing from a competitive advantage for those that have it to a disadvantage to those that do not use it.

Thus, in this sense, the strategic alignment differs from the classic vision of strategic plan, which does not present the same dynamic approach.

Figure 2. Perspectives of strategic alignment (Henderson &Venkatraman, 1993)



After the proposal of these four perspectives above, Luftman (1996) described four new perspectives that start in the infrastructure domains, instead of the strategies domains:

- Organizational IT infrastructure perspective: Organizational infrastructure → IT infrastructure → IT strategy
- IT infrastructure perspective: IT infrastructure → IT strategy → Business strategy
- IT organizational infrastructure perspective: IT infrastructure → Organizational infrastructure → Business strategy
- Organizational infrastructure perspective: Organizational infrastructure → Business strategy → IT strategy

Luftman (1996) also proposed that in some situations, a fusion of two perspectives might occur. In these cases, two perspectives can be simultaneously assessed and can impact the same domain: IT infrastructure fusion, organizational infrastructure fusion, business strategy fusion, and IT strategy fusion.

Research was developed in order to find the enablers of strategic alignment. Luftman (2001) listed five of them: senior executive support for IT, IT involved in strategy development, IT understands the business, business–IT partnership, well-prioritized IT projects, and IT demonstrates leadership. The absence or poor performance of these same factors is considered an inhibitor of strategic alignment.

Some authors, like Ciborra (1998), state that the strategic success of IT applications might be achieved through a tentative approach, rather than through structured methods of strategic IT planning. These authors argue that frequently the drivers of strategic IT applications are efficiency issues, instead of a result of a strategic IT plan. Some important and well-known successful information systems, with clear strategic impacts, do not present evidence of being previously planned, which seems to be in agreement with this kind of thinking (Eardley et al., 1996).

CONCLUSION

The concepts described above show the importance of a broad view for analyzing IT strategic alignment. Each of the described models (CSF, strategic grid, and strategic alignment) focuses on specific aspects of this issue.

These three widespread known models, in fact, have complementary characteristics, and concomitant use of them allows for better comprehension of the IT role in an organization.

On the other hand, even the use of the three models does not solve the complexity of IT alignment in organizations. As highlighted by several authors, sometimes a tentative and evolutionary approach can be successfully adopted in circumstances in which structured methods do not work properly.

Further studies would be necessary for a better and deeper understanding of the importance of IT effectiveness for the success of competitive companies. However, the intent of this article was to help find a way to reach this understanding.

REFERENCES

Brynjolfsson, E., & Hitt, L. M. (1998). Beyond the productivity paradox. *Communications of the ACM*, August.

Carvalho, M. M., & Laurindo, F. J. B. (2003). Linking strategy with a network of performance indicators: A Brazilian Research Centre case study. *International Journal of Business Performance Management, 5*(4), 285-301.

Carvalho, M. M., Laurindo, F. J. B., & Pessôa, M. S. P. (2003). Information technology project management to achieve efficiency in Brazilian companies. In S. Kamel (Ed.), *Managing globally with information technology* (pp. 260-271). Hershey, PA: Idea Group Publishing.

Ciborra, C. U. (1998). Crisis and foundations: An inquiry into the nature and limits of models and methods in the information systems discipline. *Journal of Strategic Information Systems, 7*, 5-16.

Drucker, P. F. (1963). Managing for business effectiveness. *Harvard Business Review, May/June*, 53-60.

Duhan, S., Levy, M., & Powell, P. (2001). Information systems strategies in knowledge-based SMEs: The role of core competencies. *European Journal of Information Systems, 10*(1), 25-40.

Eardley, A., Lewis, T., Avison, D., & Powell, P. (1996). The linkage between IT and business competitive systems: A reappraisal of some “classic” cases using a competitive analysis framework. *International Journal of Technology Management, 11*(3/4), 395-411.

Henderson, J. C., & Venkatraman, N. (1993). Strategic alignment: Leveraging information technology for transforming organizations. *IBM Systems Journal, 32*(1), 4-16.

Hirscheim, R., & Smithson S. (1998). Analysing information systems evaluation: Another look at an old problem. *European Journal of Information Systems, 7*(3), 158-174.

King, W. R., & Cleland, D. (1977). Information for more effective strategic planning. *Long-Range Planning, 10*, 2.

Laurindo, F. J. B. (2002). *Tecnologia da Informação: Eficácia nas organizações - São Paulo, Editora Futura*, 248 pp.

Laurindo, F. J. B., Carvalho, M. M., & Shimizu, T. (2003). Information technology strategy alignment: Brazilian cases. In K. Kangas (Ed.), *Business strategies for information technology management* (pp. 186-199). Hershey, PA: Idea Group Publishing.

Laurindo, F. J. B., & Shimizu, T. (2000). Evaluating strategies in information technology. In A. Neely (Ed.), *Performance measurement 2000 conference—Past, present and future, Proceedings* (pp. 323-220). Cambridge: Inglaterra.

Luftman, J. N. (1996). Applying the strategic alignment model. In J. N. Luftman (Ed.), *Competing in the information age—Strategic alignment in practice* (pp. 43-69). New York; London: Oxford University Press.

Luftman, J. N. (2001). Business–IT alignment maturity. In R. Papp (Ed.), Strategic information technology: Opportunities for competitive advantage (pp. 105-134). Hershey, PA: Idea Group Publishing.

Maggiolini, P. (1981). Costi E Benefici Di Un Sistema Informativo. Italy: Etas Libri.

Martin, E. W. (1982). Critical success factors of chief MIS/DP executives. *MIS Quarterly*, June.

McFarlan, W. E. (1984). Information technology changes the way you compete. *Harvard Business Review, 62*(3, May/June), 98-103.

Pollalis, Y. A., & Frieze, I. H. (1993). A new look at critical success factors in IT. *Information Strategy: The Executive's Journal*, *10*(1, Fall), 24-34.

Porter, M. E. (1979). How competitive forces shape strategy. *Harvard Business Review, 57*(2), 137-145.

Porter, M. E. (2001). Strategy and the Internet. *Harvard Business Review*, March, 63-78.

Porter, M. E., & Millar, V. (1985). How information gives you competitive advantage. *Harvard Business Review, 63*(4, July/August), 149-160.

Rockart, J. F. (1979). Chief executives define their own data needs. *Harvard Business Review, 57*(2, March/April), 81-92.

Rockart, J., & Crescenzi, A. D. (1984). Engaging top management in information technology. *Sloan Management Review, 25*(4, Summer), 3-16.

Sullivan, C. H. (1985). Systems planning in information age. *Sloan Management Review, 26*(2, Winter), 3-12.

Venkatraman, N., & Henderson, J. C. (1998). Real strategies for virtual organizing. *Sloan Management Review*, Fall, 33-48.

Walrad, C., & Moss, E. (1993) Measurement: The key to application development quality. *IBM Systems Journal*, *32*(3), 445-460.

KEY TERMS

Competitive Forces: According to Porter and Millar (1985), the state of the competition in a particular industry depends on five basic forces: new competitors, bargaining power of suppliers, bargaining power of customers, rivalry among current competitors, and substitute products or services.

Critical Success Factors (CSF): The areas where satisfactory results, according to Rockart (1979), "ensure successful competitive performance for the organization."

Effectiveness: In the context of IT, this is the measurement of the capacity of the outputs of an information system or of an IT application to fulfill the requirements of the company and to achieve its goals, making this company more competitive. In a few words, effectiveness can be understood as the ability to "do the right thing."

Productivity Paradox: The discussion about the lack of evidence about the return of investments on IT in the economy productivity indicators.

Strategic Alignment: The IT Strategic Alignment Model was proposed by Henderson and Venkatraman (1993) and consists of a framework for studying IT impacts on business and understanding how these impacts influence IT organization and strategy, as well as it enables us to analyze the market availabilities of new information technologies.

Strategic Grid: McFarlan (1984) proposed the Strategic Grid that allows the visualization of the relationship between IT strategy and business strategy and operations. This model analyzes the impacts of IT existent applications (present) and of the applications portfolio (future), defining four boxes, with each representing one possible role for IT in the enterprise: *Support*, *Factory*, *Turnaround*, and *Strategic*.

Value Chain: According to Porter and Millar (1985), this is the set of technologically and economically distinct activities a company performs in order to do business.

Information Technology Usage in Nigeria

Isola Ajiferuke
University of Western Ontario, Canada

Wole Olatokun
University of Ibadan, Nigeria

INTRODUCTION

The term information technology (IT) came into common use in the late 1980s, supplanting earlier terms such as electronic data processing systems (EDP), management information systems (MIS), and office systems (IS). Oliver, E. C., Chapman, R. J. and French, C. S. (1990) defined IT as the technology which supports activities involving the creation, storage, manipulation and communication of information (principally computing, electronics and electronic communications) together with their related methods, management and applications. Such an all-embracing term is clearly open to a number of interpretations depending on in which context it is used. In some contexts, the term "information technology" means computers, ancillary equipment, software and firmware (hardware) and similar procedures, services (including support services) and related resources. Also, it includes any equipment or interconnected system or subsystem of equipment that is used in the automatic acquisition, storage, manipulation, management, movement, control, display, switching, interchange, transmission or reception of data or information.

The emergence, development and diffusion of information technology has changed the society dramatically into what is now sometimes called an information society. IT makes it possible to collect, process and transmit information much faster and much cheaper than before. The use of IT has had profound effects on the economy, production, services and the society as a whole, and its areas of application include education, health care, commerce, publishing, manufacturing, finance, and banking (Hanna, N., Guy, K. & Arnold, E,1995).

BACKGROUND

Even though the use of IT is spreading fast, most of the market remains geographically concentrated in the advanced industrialised countries, notably the United States, Western Europe and Japan. There are great differences between industrial and developing countries: Hanna, Guy and Arnold (1995) concluded that developing countries are poor in the infrastructure that is the key to IT diffusion. Lack of trained manpower, the high cost of telecommunications and document delivery, and the cost of information itself are other barriers developing countries are struggling with (Conceicão Calomon Arruda, 1997).

Apart from the problems identified above, the other major impediment to the successful adoption of IT in developing countries is the non-availability of a well-articulated information technology policy. According to Shila (1994), such policies are needed to:

(i) solve coordination problems in IT such as haphazard development or importation of systems without regard for nationwide concerns and priorities;
(ii) create a critical mass of expertise in the public sector;
(iii) raise general awareness about the social and economic consequences of IT;
(iv) increase the efficiency of government computerization; and
(v) launch specific projects that cut across agency lines in the fields of infrastructure development, standardization, human capital formation, and technology support for the private sector.

Shila's view is corroborated by Mulira (1995), who states that the purpose of having an appropriate national IT policy is to create an environment where economic and social benefits may be achieved; where utilization of resources may be optimized; where domestic technological capabilities may be encouraged; and where procurement decisions can be taken rationally. Hence, the objectives of this chapter are to review the IT usage in a typical developing country, Nigeria, and to evaluate the appropriateness of its recently formulated IT policy.

MAIN THRUST OF THE ARTICLE

The Federal Republic of Nigeria is situated on the West African coast, and shares common borders with the Benin Republic in the West, Niger to the North, Chad Republic to the North-East, and Cameroon Republic to the East. Nigeria is the most populous country in Africa, with an estimated population of about 123 million people (Encyclopedia4u, 2000). The climate is wholly tropical, and even though more than 75% of the people live in rural areas, agriculture is not the mainstay of the Nigerian economy. The expansion in the oil sector in the 1970s and early 1980s has led to a considerable decline in agriculture.

IT usage in Nigeria cuts across the major sectors of the economy but with varying degrees of applications. The use of computers in manufacturing - with the exception of certain high-technology process industries in which the process computer is essentially a part of the process machinery (e.g. petroleum refining and the steel mill) - is only a recent phenomenon. It is in the area of coordination, or management, which the greatest advance has been made. In fact, computerization has been most widespread in the area of financial management, including payroll, accounts, general ledger, sales, and invoicing. More than 80 per cent of computer installations are used in this way. There are also many instances of companies that have not installed computers but have their accounts and payroll batch-processed on a bureau computer owned by a vendor or an agency.

In the banking and financial industry, computerization is still limited to ledgers, communication, and current account management (Ugwu, L.O., Oyebisi, T.O., Ilori, M.O., & Adagunodo, E.R., 2000; Idowu, P.A., Alu, A.O., & Adagunodo, E.R., 2002; Uche, 2003). There are no automated telling machines, nor are multi-branching facilities available. However, the service to the customer is improving in many respects, including "quick service" cash counters, and prompt and regular monthly statements of account. In general, the thrust of computerization in banks is in the direction of more automation and networking, but the rate of progress is limited by the ineffective telecommunications infrastructure in the country. Other service industries, such as advertising, also use IT mainly for word processing and accounting functions (Ehikhamenor, 2002; Odesanya and Ajiferuke, 2000; Adetayo, J.O., Sanni, S.A., & Ilori, M.O., 1999).

In the communications field, the Nigerian Telecommunications Limited (NITEL), the national carrier, is heavily computerized, with a huge installed capacity including mainframes, minis, and micros. These computers are located in four regional headquarters as well as in the national headquarters in Lagos. The machines are used for administrative purposes and for the management of the telephone network. At present, they function on a stand-alone basis, but it is known that NITEL is interested in interconnecting these machines in both local and wide area networks for greater efficiency and increased flexibility. The National Television Authority (NTA) is computerized, and so are the leading telecommunications outfits represented in Nigeria. However, the postal agencies are yet to adopt the use of IT in sorting mails.

Government departments are rapidly computerizing (Tiamiyu, 2000). As many as 13 federal ministries have computers; the Ministry of Defence alone has 9 installations. The Ministry of Works operates a computerized maintenance system. The Federal Office of Statistics, the national body responsible for the gathering and compilation of statistical data ranging from trade statistics to commodity prices and population data, is fully computerized. Among the state governments, it is usual to install a central computer facility in the Ministry of Finance for financial administration. In almost all the 25 federal universities there are well-staffed computing centres equipped with time-shared multi-user mainframe computers used for teaching and research (Ehikhamenor, 2003). In addition, several departments and faculties have their own computer facilities, consisting mainly of micros. Many university computing centres also provide computer services for the administrative departments, such as the bursary, the registry, and the library.

For the general public, only a few homes have computers but IT training centres and cybercafés are spread across Nigeria, especially in the major cities (Adomi, E.E., Okiy, R.B., & Ruteyan, J.O., 2003). These centres provide IT training as well as allow people to do word processing, send e-mails, and browse the Internet.

Overall, IT usage in Nigeria is widespread but the depth is very shallow. The reasons for these, which are similar to those in other developing countries, include: very low literacy level; extremely low density of telecommunications facilities and services; underdeveloped computing infrastructures and culture; government regulations; corruption in both public and private organizations; and lack of a coordinated national IT policy (Eze, 2002, Oyebisi & Agboola, 2003).

FUTURE TRENDS

To increase the depth and effective usage of IT in Nigeria in the future, solutions would have to be found to the problems identified above. Telecommunications is probably the first critical requirement because it provides the network through which various people, organizations and regions of Nigeria can be linked together, and to the rest of the world. Telecommunications facilitates not only Internet connectivity, but also interpersonal voice, radio and TV communications. However, investors in the sector must be ready to focus on investing in the short and medium terms, and recovering their investments in the long run. They must also be willing to popularize the telephone culture through appropriate pricing.

Computing infrastructure must also become more widespread in order for people to use computers to exploit Internet-based information service opportunities. But computers are still relatively expensive to acquire for most people and organizations. Nevertheless, an area requiring urgent attention is the educational sector - primary, secondary and tertiary institutions, because graduates from such institutions would eventually constitute the future work force. Hence, we should begin by introducing and expanding computing laboratories and networks in educational institutions. Indeed, the development of these laboratories and centres in schools, colleges, organizations, and shopping plazas will quickly popularize computing culture among students, employees, and the masses in general.

The presently low literacy level in Nigeria will take time to improve. Related to the low literacy level are the two other underdeveloped aspects of the Nigerian socio-economy: restriction of universal access to public information, and inadequate appreciation and application of science and technology. A high literacy level will empower the masses to request for intellectually nourishing information as well as demand for their citizenship right of access to public information. Also, a high literacy level will enable the masses to appreciate science and technology better.

In response to the lack of an enabling national IT policy, the Nigerian Government, in February 2001, put together a national IT Policy for Nigeria (Federal Government of Nigeria, 2001). This policy is in the early stages of implementation, and we shall evaluate its adequacy below under the following broad headings: scope and coverage, relevance of the policy to the IT needs of Nigeria, acceptability among stakeholders, the realism of the objectives and strategies set out in the policy, and the implementation strategy.

(i) **Scope and coverage**: Based on the content analysis of the Nigerian IT policy, the policy was found to be very comprehensive in scope. The policy gave recognition to fifteen sectoral areas; these are human resource development, infrastructure, governance, research and development, health, agriculture, trade and commerce, fiscal measures, government and private sector partnerships, arts, culture and tourism, national security and law enforcement, legislation, global considerations, and IT popularization and awareness. However, as comprehensive and all-embracing as the policy is in terms of scope, there are still some gaps. For instance, a strategic sector like education was glaringly omitted. The role of education, as producer and user of IT products and systems, was mentioned in connection with the human development sector. But one finds it difficult to believe that education can be so relegated to the background in a national IT policy for Nigeria where illiteracy rate is still pretty high.

(ii) **Acceptability among stakeholders**: The policy is far from being acceptable to some of the stakeholders. For instance, prior to the formulation of the policy, both the Information Technology Association of Nigeria (ITAN) and the Computer Association of Nigeria (COAN) submitted proposals on what the content of an IT policy for Nigeria should look like. In the final document, some of the suggestions made by these two associations were ignored. Thus, the present IT policy is far from being acceptable to all the stakeholders. This is why National Information Technology Development Agency (NITDA) has been finding means of reviewing the policy both for meeting new challenges, and also to make it more acceptable to the different stakeholders.

(iii) **The realism of the objectives and strategies set out in the policy:** The vision statement of the policy was "*to make Nigeria an IT capable country in Africa and a key player in the information society by the year 2006, using IT as the engine for sustainable development and global competitiveness.*" The question is how realizable is this vision statement in a country where computer illiteracy rate is still pretty high, and the information and telecommunications infrastructures are very poor. Also, the mission statement identified four key areas for the use of IT: creation of wealth, poverty eradication, job creation, and global competitiveness. It neglects education as a key area in which IT could be applied. For instance, how can IT be used in creating jobs and eradicate poverty when the majority of the population is still not knowledgeable about IT?

On governance, one of the objectives is "*to replace traditional governance with electronic governance*". The state of the telecommunications infrastructure and the level of computer literacy among Nigerians create serious doubts about the capability to achieve this objective. E-governance, at best, will distance government from the people and neglect the generality of Nigerians owing to the low levels of literacy and computer literacy among Nigerians.

(iv) **Policy Implementation strategy:** One of the important strategies set out for the implementation of the IT policy is contained under policy implementation sub-section 16.3 (i) as follows:

"Establishing a coordinated program for the development of a National, State and Local Informa-

tion Infrastructure (NII, SII, LII) backbone by providing emerging technologies, such as satellite (including VSAT), fibre optic networks, high-speed gateways and broad bandwidth/multimedia within the next eighteen months but not later than the fourth quarter of 2002"

The above strategy is well thought out and of great import. But one would have thought that the policy would have been patterned after that of India by promoting state-owned enterprises in such a way that each state in the country will be made to have its own IT policy and thus develop the aspect of the State Information Infrastructure (SII) while the local governments will be made to develop the aspect of the Local Information Infrastructure (LII). This pattern, if adopted will be with a view to integrating SII and LII with the National Information Infrastructure (NII). It is believed that with this approach, there will be a rapid progress in the development of the National Information Infrastructure (NII). This perception applies to other aspects of the policy, such as IT popularization and awareness, governance, IT human resource development, national security and law enforcement as well as other sectoral applications.

CONCLUSION

In this chapter, we have provided an overview of IT usage in Nigeria as well as discuss the national IT policy. It is our view that Nigeria needs to modify its national IT policy in such a way that technology is put at the centre of the developmental planning. Along this line, we submit that policy implementation requires an institutional framework without which any planning effort may be futile, no matter how elaborate or well intentioned. Thus, we conclude by saying that the future of IT and related technologies in Nigeria cannot but be bright provided public and private sector initiatives and policies are implemented to remove the obstacles identified above.

REFERENCES

Adetayo, J.O., Sanni, S.A., & Ilori, M.O. (1999). The impact of information technology on product marketing: A case study of a multinational company in Nigeria. *Technovation, 19*(11), 691-699.

Adomi, E.E., Okiy, R.B., & Ruteyan, J.O. (2003). A survey of cybercafés in Delta State, Nigeria. *Electronic Library, 21*(5), 487-495.

Conceicão Calomon Arruda, M. da (1997). Libraries, new technologies and human resources: The challenge to the 21st Century. *63rd IFLA General Conference - Conference Programme and Proceedings***,** 1-14.

Ehikhamenor, F.A. (2003). Internet facilities: Use and non-use by Nigerian university scientists. *Journal of Information Science*, *29*(1), 35-48.

Ehikhamenor, F.A. (2002). Socio-economic factors in the application of information and communication technologies in Nigerian print media. *Journal of the American Society for Information Science and Technology*, *53*(7), 602-611.

Encyclopedia4u. (2000). *Demographics of Nigeria*. Retrieved February 13, 2004, from *http://www.encyclopedia4u.com/d/demographics-of-nigeria.html*

Eze, E. (2002). A critical examination of strategic planning and implementation of information technology in developing countries. *International Journal of Services Technology & Management*, *3*(4), 429-440.

Federal Government of Nigeria (2001). *Nigerian national policy for information technology*. Abuja: National Information Technology Development Agency.

Hanna, N., Guy, K. & Arnold, E. (1995). *The diffusion of information technology: Experience of industrial countries and lessons for developing countries*, World Bank Discussion Papers 281, Washington, p. 7.

Idowu, P.A., Alu, A.O., & Adagunodo, E.R. (2002). The effect of information technology on the growth of the banking industry in Nigeria. *The Electronic Journal on Information Systems in Developing Countries*, *10*(2), 1-8.

Mulira, N. K. (1995). Managing information technology in Uganda: Strategies and policy formulation. *Information Technology for Development*, *6*, 95-105.

Odesanya, O.A., & Ajiferuke, I. (2000). Information technology usage by advertising agencies in Lagos, Nigeria. *African Journal of Libraries, Archives & Information Science*, *10*(2), 113-123.

Oliver, E. C., Chapman, R. J. and French, C. S. (1990). *Data processing and information technology – An instructional manual for business and accountancy students* (8th ed.). London: DP Publications.

Oyebisi, T.O. & Agboola, A.A. (2003). The impact of the environment on the growth of the Nigerian IT industry. *International Journal of Information Management*, *23*(4), 313-321.

Shila, H. (1994, April). Case study effectiveness of informatics policy instruments in Africa: Tanzania. A paper presented at the meeting of the *Committee of Development Information* (CODI), Addis Ababa, Ethiopia, Addis Ababa: ECA.

Tiamiyu, M.A. (2000). Availability, accessibility and use of information technologies in Nigerian federal agencies: Preliminary survey. *Information Technology for Development*, *9*(2), 91-104.

Uche, C.U. (2003). Information technology and commercial banking in Nigeria. *Journal of International Banking Law and Regulation*, *18*(3), 102-109.

Ugwu, L.O., Oyebisi, T.O., Ilori, M.O., & Adagunodo, E.R. (2000). *Technovation*, *20*(12), 711-721.

KEY TERMS

Cybercafé: A café that makes available a number of personal computers that are connected to the Internet for the use of the general public.

Developing Country: A country with a relatively low per capital gross national product (GNP).

Information Economy: An economy which is dominated by information-related activities.

Information Society: A society in which information is used heavily in everyday life of most citizens.

Information Technology: Encompasses all forms of technology used in processing and disseminating information.

Policy: A course of action developed for the achievement of a set of goals.

Strategy: A plan or method for obtaining a specific result.

I

Informationbase – A New Information System Layer

Kovach Dragan
SUPER-King, Inc., Croatia

INTRODUCTION

Information is a set of data purposefully organized into a report. In the database of an information system, only the raw material for the information is stored, that is, the data. To get the information one needs:

- To find an adequate option in the information system menu.
- To understand the possibilities of the option and to enter appropriate values for the parameters required.
- Execute the option and …
- … wait to get the report.

The concept of the informationbase eliminates these steps and enables the user to get the information directly and instantly.

Informationbase is a top layer of the information system that uses information systems database and information system reporting options to automatically maintain set of reports and graphs, permanently available to the user. Hence, it acts as an information manager that:

- enables free design of the informationbase contents and structure,
- automatically activates predefined information system options ...
- ... and delivers parameters for those options to produce up to date reports,
- reports produced automatically and/or manually stored in an informationbase, where they could be easily accessed by authorized users, and
- manages authorization of informationbase usage.

The rationale for the informationbase concept is the following:

Instead of being repeatedly and randomly produced by individual users, information could be once generated by the editor and/or automatically generated and stored in the informationbase, that is, be permanently available to all potential users.

Thus, the final product of the information system - information - could be available as up to date reports, rather than as a vague potential possibility, hidden in the complexity of the information system.

One of advantages of the informationbase is that it solves the problem of computer capacity overload, so commonly present and caused by simultaneous information retrieval of many users.

As Zygmont (1999) states: “The usual lifecycle goes like this: The user first accepts the reporting tools and templates provided by the ERP vendor. The honeymoon ends shortly after implementation, when workers become comfortable enough with a new system to begin making demands”.

“Structure of data within ERP systems generaly prevents efficient reporting. Data tables are very normalized, containing minimal cross references, so that transactions move rapidly. But that means to get reports, you have to hit up against many different tables. That eats a lot of CPU cycles.”

But even more important is that it makes easy to grasp what are the most important deliveries of the information system to the user organization. Namely, the scope of information system coverage is carefully structured and clearely presented in the informationbase. Thus, the user gets the understanding of the system far faster, more completely and with less effort than studying system manuals. The cognitive and organizational learning dimensions of the information system are recognized and considered as important issues in the literature (van Stijn, 2001, p. 502).

BACKGROUND

Informationbase concept is developed by the author of this paper, working for SUPER-KING Inc., Zagreb, Croatia, and implemented in TOMAS (**TO**tal **MA**nagement **S**ystem) ERP (enterprise resource planning) software. In the Croatian version, the name of software is SUPER (**S**ustav **U**pravljanja **P**oslovanjem **E**lektronickim **R**acunalom). The first version of informationbase was implemented in 1994 (Kovach, 1996). SUPER has been implemented in an organization with a multimilion transactions database. This caused problems with computer overload during the day, for many users had been making reports that required handling large database. First, we tried to solve that so that computer operators were producing the most frequently needed reports after normal working hours and during the night, so that the next day printed reports were distributed to users. The first version of informationbase was a report management system that enabled storage of generated reports and retrieval at will.

In this article the second version, developed in 1996, is presented, to describe the concept first presented in Kovach (2001, 2004). In this version, capabilities of informationbase have been extended to the limits of our imagination and remain to the present time unchanged.

PROPERTIES OF THE INFORMATIONBASE

Informationbase is a structured set of reports and/or graphs ready to use. TOMAS' informationbase offers the following capabilities:

- Informationbase is built for the user organization and a calendar year, but any number of informationbases could be kept simultaneously.
- Each informationbase could have:
    - up to 99 chapters,
    - up to 34 reports/chapter, and
    - unlimited number of pages/report.
- Structure of chapters and reports could be defined by the user. The default structure of the informationbase is available, but the user could modify or change it completely.
- Access to the informationbase is controlled by an authorization system. User password determines what chapters and/or reports are available to the user and whether the user is allowed to change informationbase structure.
- Reports in the informationbase could be updated automatically and/or manually. For each informationbase entry to be updated automatically, there should be stored parameters that are described next.

  Based on these parameters, informationbase is automatically updated, using standard information system options for report generation, which are used apart from the informationbase as well.

Hence, informationbase is a new layer of an information system for maintaining a permanently available and up to date set of reports and graphs predefined by the user.

CREATING THE STRUCTURE OF THE INFORMATIONBASE

Informationbase structure is created in the following way:

- Chapter is defined by the chapter number and the chapter title.
- Chapter entry, that is, report or graph, is defined by entry number and entry title.
- For each chapter entry (report or graph) to be generated automatically, the following parameters are to be stored:
    - Number of the option from the information system menu that makes the report. These options are used to produce reports normally, apart from informationbase, but also could be used for automatic report generation for the the informationbase. This requires additional considerations in information system development.
    - Option template number, with specified values of parameters. Options for report generation in TOMAS have usually 15 - 20 parameters and the capability of producing billions of report variations, depending on user chosen values of parameters. Combinations of parameters values could be stored as option templates.
    - Type of date interval – reports are always made for some period of time. This period is defined by several parameters. Type of date interval is one of parameters and refers to one of the following:
        - date of data entry
        - document date
        - payment due date
        - financial month/year
        - planned due date
        - order delivery date, and so forth
    - Relative report time period – defines the period for which the report is to be made in relative terms, which could be:
        - yesterday
        - today
        - previous month
        - current month
        - current year up to today
        - current year up to the end of previous month, and so forth
        - to previous could be added +/- N days, where N is defined by the user

This makes it possible that time cycle for each report in the informationbase could be defined. Also, this parameter is needed because in the program unit for making report, one must specify the date interval for which report is to be made. Hence, when the option is used for the informationbase, the informationbase driver has to determine actual time limits for "yesterday," "previous month" and so forth and insert them into option parameters so that the report could be produced.

- Time activator, that is, trigger for automatic update of the report in the information base, could be:
  - calendar day
  - calendar month
  - calendar year
  - financial (booking) month and so forth

So, if the trigger is calendar day, the report is produced automatically each day. If the trigger is booking month, the report is produced whenever the booking month is changed in the database.

UPDATING THE INFORMATIONBASE

There are three ways to update the informationbase:

- Manually, that is, any report could be stored in any of the chapters of the informationbase,
- Semiautomatic – the user starts the option for automatic updating of the whole informationbase or only part of it, for example some chapters and/or reports.
- Fully automatic – the option for updating the informationbase is set to start every day at a specified time.

When the informationbase update is started, the system scans time activators and checks what reports are to be updated, starts appropriate options, that is, application programs for report generation with predefined option templates, and updates the informationbase.

ADVANTAGES OF THE INFORMATIONBASE

The informationbase concept brings the following advantages:

- Information is comprehensive and well structured: in 10 – 15 chapters and 150 – 200 reports carefully defined, all aspects of business are covered:
  - In typical time segments (yesterday, today, current month, current year, etc.)
  - Summarized by relevant entities (materials, vendors, etc.)
  - Presented as ABC analysis, that is, sorted by relevant value, with cumulative percentages
  - Due dates approaching in next period, for example orders to be delivered in the next 5 days
  - Due dates that are already late, for example undelivered orders or unpaid invoices
  - Exceptions to the rules, for example inventories below minimum and so forth

The user could acquire information by normal use of an information system, as well, using system options in a standard way, but the major part of information needs should be covered by the informationbase.

What is going to be in the informationbase depends upon the potential of the information system and user judgment as to what is relevant. Still, whatever the quality of the information system is, the informationbase concept makes possible that full potential of information system output repertoire is carefully structured, completely transparent and permanently available to the user.

- Information system is much easier to grasp: Full potential of information system output repertoire that is relevant to the organization is transparently presented by the informationbase structure and contents. Hence, it is easy to grasp what most important information system capabilities are.
- Information is always up to date - updating is done automatically.
- Full availability and direct use of information – the user immediately gets the information, without intermediates, at any time, from any place (remote access and/or via INTERNET), and no computer literacy is required.
  Therefore, users can forget about all intricacies of the information system – all one needs to do is to enter the informationbase menu and choose the information.
- Directs attention – titles of chapters and reports indicate aspects to be scrutinized, even if the user has no idea what to look for.
- Efforts and costs minimized – information is made once, automatically and used by many, so human and computer resources required are minimized.
- Resultant effect of informationbase is that all users, especially management, are better informed, with less total costs and effort.

STANDARD (DEFAULT) INFORMATIONBASE

To help the user, information system offers standard, that is, default informationbase, with all parameters ready made. As has been mentioned already, the user could change it or define a completely different structure, without programming. The structure of chapters and content of some chapters for standard TOMAS informationbase is shown as follows.

FUTURE TRENDS

As far as TOMAS, that is, SUPER is concerned, there are no plans for further development of the informationbase software engine. Rather, we work on informationbase content, that is, structure of the informationbase chapters for various business functions (e.g., sales, production control, etc.) and types of organizations (e.g., retail, manufacturing, etc.); that is, we study the morphology of logical informationbase. Still, if the concept of informationbase would be accepted, it is conceivable that informationbase software could be developed for existing ERP packages and/or information systems of large organizations.

CONCLUSION

The informationbase is a new layer of the information system, which uses information systems database and application programs to automatically maintain the final product of the information system, that is, information. Information produced in this manner is comprehensive, carefully structured, permanently up to date and available to the user, without the need for users to run any programs. Hence, informationbase is a sort of information system management that permanently keeps available full potential of the information system, ensuring maximum efficiency in getting information and minimum of human and computer resources required.

TOMAS - Informationbase: C H A P T E R S

Organization: CROATIA – INVEST Inc. Year: 2003

	C O M M E R C E		**F I N A N C E**
1	SALES		
2	PURCHASING		**Accounts:**
3	INVENTORIES	6	RECEIVABLES
		7	PAYABLES
	M A N U F A C T U R I N G		
4	PRODUCTION PLAN	8	COST CONTROL
5	PRODUCTION CONTROL	9	BALANCES
		10	FINANCIAL ANALYSIS

TOMAS - Informationbase: F I N A N C I A L A N A L Y S I S

Organization: 30 CROATIA - INVEST Inc. Year: 2003

	Planned vs. actual results:	5	Why profit has been changed
1	Company as a whole		
2	Organizational units	6	What is required to increase the profit
		7	What if ...
	Key business indicators:		
3	Company as a whole		
4	Organizational units		

TOMAS - Informationbase: S A L E S			
Organization: 30 CROATIA – INVEST Inc.		Year: 2003	
	SALES ORDERS		**ACTUAL SALES**
1	In previous month	13	Yesterday
2	In current month	14	Today
3	With due date in up to 10 days	15	Current month
4	Late for more than 3 days	16	Previous month
		17	Year to previous month
	Sales orders summary by:		
5	Buyer		**ABC analysis by:**
6	Class of product	18	Warehouse
7	Product	19	Buyer
		20	Class of product
		21	Product
	Bar charts:		
8	Value of orders		**Periods comparison by:**
9	Ordered quantities	22	Buyer
		23	Product
	PRICE CONTROL		
10	Sales price movements		**Bar charts:**
11	List of invoice prices	24	Values for monthly sales
12	Actual/planned prices	25	Quantities of monthly sales

REFERENCES

Kovach, D. (1996, June). Back to the future – New possibilities for the design of information systems. *Proceedings of the CASE 8 Conference*, Opatija, Croatia.

Kovach, D. (2004, June). Intelligent financial system. *Proceedings of the CASE 16 Conference,* Opatija, Croatia.

Kovach, D., & Fertalj, K. (2001, May). Informationbase – The new layer of the information system. *IRMA 2001 Conference*, Information Resource Management Association, Toronto.

van Stijn, E. (2001). Beyond ERP systems as a hype. *2001 Information Resources Management Association International Conference,* Toronto.

Zygmont, J. (1999). ERP system performance - rethink reporting. *Datamation IT Management Update, 19.5*.

KEY TERMS

Authorization for Informationbase: Regulates access to informationbase, that is, what chapters and entries are available to a particular person, authority to initiate informationbase updating and to change informationbase structure.

Informationbase: Set of reports and graphs, structured by means of chapters, updated automatically and/ or manually.

Informationbase Chapter: Part of informationbase, group of reports or graphs.

Informationbase Entry: Part of an informationbase chapter, single report or graph,

Relative Time Period: Informationbase parameter that defines relative time period for which the report is to be produced, for example yesterday, today, previous month and so forth. Informationbase driver translates this into actual dates for data processing.

Time Activator: Trigger for initiating automatic update of the report in the informationbase. Could be calendar day, month or year, financial (booking) month and so forth. So, if the trigger is calendar day, the report is produced automatically each day. If the trigger is booking month, the report is produced whenever booking month is changed in the database.

Type of Date Interval: Informationbase parameter that defines the relative time period. This could be document date, delivery due date, payment due date and so forth.

InfoSec Policy – The Basis for Effective Security Programs

Herbert J. Mattord
Kennesaw State University, USA

Michael E. Whitman
Kennesaw State University, USA

INTRODUCTION

The success of any information security program lies in policy development. The lack of success in any particular program can often be attributed to this unmet need to build the foundation for success. In 1989, the National Institute of Standards and Technology addressed this point in *Special Publication SP 500-169: Executive Guide to the Protection of Information Resources* (1989):

The success of an information resources protection program depends on the policy generated, and on the attitude of management toward securing information on automated systems. You, the policy maker, set the tone and the emphasis on how important a role information security will have within your agency. Your primary responsibility is to set the information resource security policy for the organization with the objectives of reduced risk, compliance with laws and regulations and assurance of operational continuity, information integrity, and confidentiality (p.1).

Policy is the essential foundation of an effective information security program. As stated here by Charles Cresson Wood, in his widely referenced book *Information Security Policies Made Easy* (2003),

The centrality of information security policies to virtually everything that happens in the information security field is increasingly evident. These policies will stipulate the type of services that should be permitted, how to authenticate the identities of users, and how to log security-relevant events. An effective information security training and awareness effort cannot be initiated without writing information security policies because policies provide the essential content that can be utilized in training and awareness material (p.1).

Policy is essential because it is the primary mechanism an organization possesses to inform and enforce expected behaviors in employees. Policy has the effect of law within the confines of the institutions. However, while *ignorantia legis neminem excusat* (ignorance of the law is no excuse) is prevalent in the public domain, ignorance of policy is legally defensible.

Although information security policies are among the least expensive information security controls to create, they are often the most difficult to implement. Policy-based controls typically cost only the time and effort the management teams spends to create, approve, and communicate them, and the time and effort employees spend integrating the policies into their daily activities. Even when the management team hires an outside consultant to assist in the development of policy, the costs are minimal compared to the other forms of control, especially technical controls (Whitman & Mattord, 2004).

BACKGROUND

Policy is "a plan or course of action, as of a government, political party, or business, intended to influence and determine decisions, actions, and other matters" (Merriam-Webster, 2002). In other words, policies are a set of rules that dictate acceptable and unacceptable behavior within an organization. Policies must also specify the penalties for unacceptable behavior, and define an appeal process. An example of a policy would be an organization's prohibiting the viewing of pornographic Web sites at the workplace.

To execute this policy, the organization must implement a set of standards. A standard is a more detailed statement of what must be done to comply with policy. In the implementation of the anti-pornography policy, the organization may create a standard that the network will block access to pornographic Web sites. Practices (i.e., procedures and guidelines) explain how employees will comply with policy.

For policies to be effective they must be properly disseminated, via personnel manuals, organizational intranets, periodic supplements, staff meetings and/or training (to name a few). All members of the organization must read, understand, and agree to abide by the organization's policies. Failure to ensure each of these requirements can negate the regulatory effect of policy. Policies require constant modification and maintenance. As the needs of the organization evolve, so must its policies.

Some basic rules must be followed when shaping any policy, including information security policy:

- Policy should never conflict with law.
- Policy must be able to stand up in court, if challenged.
- Policy must be properly supported and administered.

Since policy is often difficult to implement, Bergeron and Bérubé (1990) have proposed guidelines for the formulation of computer policy, which are also directly applicable to information security policy:

1. "All policies must contribute to the success of the organization.
2. Management must ensure the adequate sharing of responsibility for proper use of information systems.
3. End users of information systems should be involved in the steps of policy formulation" (p. 16).

Bergeron and Bérubé further note that while it is an admirable goal for policies to be complete and comprehensive, too many policies or policies that are too complex can lower end-user satisfaction.

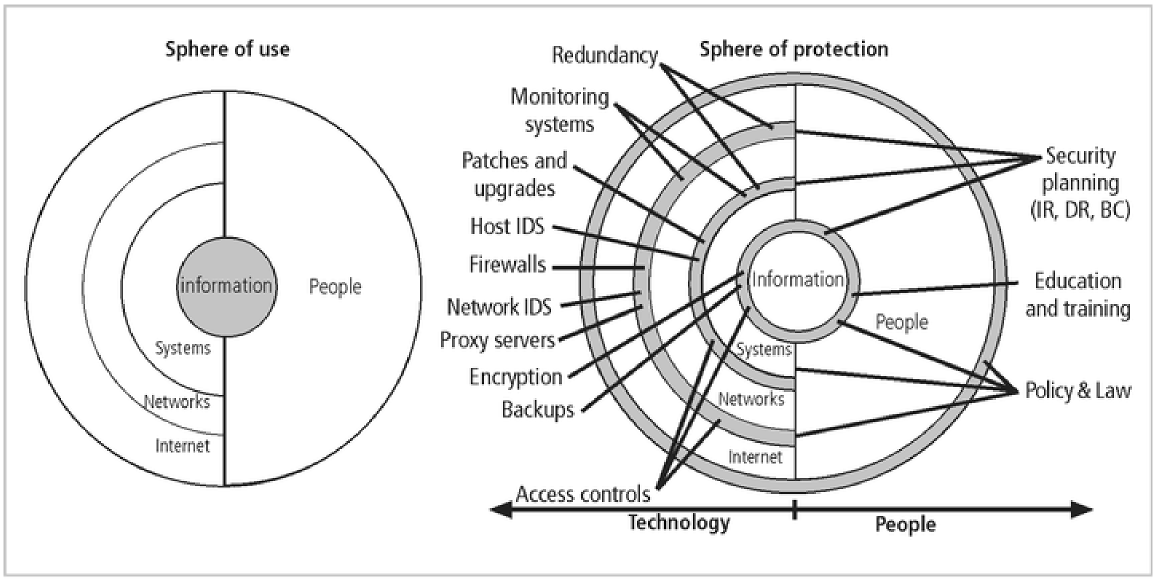
As is evidenced in Figure 1, in order to secure information an organization must place protection mechanisms at multiple points. This is easily done in the electronic arena, where most threats come through the Internet, to the internal network, to the systems that house information, and finally to the information itself. However, inside an organization you may only have a few opportunities to protect information from those that use it. These opportunities include security education, training and awareness programs (SETA) and policy.

The use of multiple layers of protection is a concept called defense-in-depth, whereby security components at multiple layers serve to back each other up in the event that one layer's controls fail. Until sound and useable IT and information security policy is developed, communicated, and enforced, no additional resources should be spent on controls other than policy.

EFFECTIVE INFORMATION SECURITY POLICIES

To produce complete information security policy in the organization, management must use three types of information security policies. These three types are based on National Institute of Standards and Technology Special Publication 800-14 (1996), which outlines the requirements of writing policy for senior managers. This document is recommended for professionals involved in creat-

Figure 1. Spheres of use and protection of information (Whitman & Mattord, 2003)



ing policy, and can be found at http://csrc.nist.gov/publications/nistpubs/800-14/800-14.pdf. The three types of policy are:

- Enterprise information security program policy (EISP)
- Issue-specific information security policies (ISSP)
- Systems-specific information security policies (SysSP)

Enterprise Information Security Policy

An enterprise information security policy (EISP)—also known as a security program policy, general security policy, IT security policy, high-level information security policy or information security policy—sets the strategic direction, scope, and tone for all of an organization's security efforts. The EISP assigns responsibilities for the various areas of information security, including maintenance of information security policies, and the practices and responsibilities of end users. In particular, the EISP guides the development, implementation, and management requirements of the information security program, which must be met by information security management, IT development, IT operations and other specific security functions.

The EISP is an executive-level document, drafted by the Chief Information Security Officer (CISO) in consultation with the Chief Information Officer (CIO), and shapes the security philosophy in the IT environment. The EISP usually does not require repeated or routine modification, unless there is a change in the strategic direction of the organization.

The EISP plays a number of vital roles, not the least of which is to state the importance of information security in support of the organization's mission and objectives. Information security strategic planning derives from the IT strategic policy (if the Information Security department is placed under the control of the CIO), which is derived from the organization's strategic planning. Unless the EISP directly reflects this association, the policy will likely become confusing and counter-productive.

Though specifics of EISPs vary from organization to organization, most EISP documents should provide the following:

- An overview of the corporate philosophy on security
- Information on the structure of the information security organization and individuals that fulfill the information security role
- Fully articulated responsibilities for security that are *shared by all members* of the organization (employees, contractors, consultants, partners and visitors)
- Fully articulated responsibilities for security that are *unique to each role* within the organization

The formulation of program policy in the EISP establishes the overall information security environment. As noted earlier, there are any number of specific issues that require policy guidance beyond what can be offered in the EISP. The next level of policy document, the issue-specific policy, delivers the specificity. The components of a good EISP are as follows:

1. Statement of Purpose - Answers the question "What is this policy for?" Provides a framework for the helps the reader to understand the intent of the document, that is, "This document will:

- Identify the elements of a good security policy
- Explain the need for information security
- Specify the various categories of information security
- Identify the information security responsibilities and roles
- Identify appropriate levels of security through standards and guidelines" (WUSTL, 2002).

2. Information Security Elements - Defines information security. It can also lay out security definitions or philosophies in order to clarify the policy. For example - "Protecting the confidentiality integrity and availability of information while in processing, transmission and storage, through the use of policy, education & training, and technology…" (WUSTL, 2002).

3. Need for Information Security - Provides information on the importance of information security in the organization and the obligation (legal and ethical) to protect critical information whether regarding customers, employees, or markets.

4. Responsibilities and Roles - Defines the organizational structure designed to support information security within the organization. Includes identification of categories of individuals with responsibility for information security (IT dept, management, users) and their information security responsibilities, including maintenance of this document.

5. Reference to Other Standards and Guidelines - Outlines lists of other standards that influence and are influenced by this policy document. These could include relevant laws, federal and state, as well as other polices in place in the organization. (Note: this outline was derived from a number of sources, the most notable of which is WUSTL, 2002.)

Issue-Specific Security Policy (ISSP)

A sound issue-specific security policy provides detailed, targeted guidance to instruct all members of the organization in the use of technology-based systems. The ISSP should begin with an introduction of the fundamental technological philosophy of the organization. It should assure the members of the organization that the purpose of the policy is not to provide a legal foundation for persecution or prosecution, but to provide a common understanding of the purposes for which an employee can and cannot use the technology. Once this understanding is established, employees are free to use the technology without seeking approval for each type of use. This serves to protect both the employee and the organization from inefficiency and ambiguity. According to Whitman et al., (1999) an effective ISSP:

- Articulates the organization's expectations about how the technology-based system in question should be used
- Documents how the technology-based system is controlled and identifies the processes and authorities that provide this control
- Serves to indemnify the organization against liability for an employee's inappropriate or illegal system use

An effective ISSP is a binding agreement between parties (the organization and its members) and shows that the organization has made a good faith effort to ensure that its technology is not used in an inappropriate manner. An ISSP may be drafted to cover many topics, including e-mail, use of the Internet and World Wide Web, office computing equipment, and a host of other fair and responsible use areas. The specific situation of any particular organization dictates the exact wording of the security procedures as well as issues not covered within these general guidelines. There are seven major sections of a good ISSP (Whitman, 2003). These are described here in detail.

1. Statement of Purpose - a clear statement of purpose that outlines the scope and applicability of the policy, addressing the purpose of this policy, who is responsible and accountable for policy implementation and what technologies and issues the policy document addresses.
2. Authorize Access and Usage of Equipment - who can use the technology governed by the policy, and for what purposes. This section defines "fair and responsible use" of equipment and other organizational assets, as well as addressing key legal issues, such as protection of personal information and privacy.
3. Prohibited Usage of Equipment - what the issue or technology cannot be used for, that is, personal use, disruptive use or misuse, criminal use, offensive or harassing materials, and infringement of copyrighted, licensed, or other intellectual property. Unless a particular use is clearly prohibited, the organization cannot penalize employees for such usage.
4. Systems Management - the users' relationships to systems management, including systems maintenance and storage authorization and restriction. The Systems Management section should specify users' and systems administrators' responsibilities.
5. Violations of Policy - the penalties and repercussions of violating the usage and systems management policies, as well as instructions on how to report observed or suspected violations, either openly or anonymously.
6. Policy Review and Modification - procedures and a timetable for periodic review. This section should contain a specific methodology for the review and modification of the ISSP, to ensure that users always have guidelines that reflect the organization's current technologies and needs.
7. Limitations of Liability - a general statement of liability or set of disclaimers. If an individual employee is caught conducting illegal activities with organizational equipment or assets, management does not want the organization held liable. Therefore, if employees violate a company policy or any law using company technologies, the company will not protect them, and is not liable for their actions, assuming that the violation is not known or sanctioned by management.

Systems-Specific Policy (SysSP)

While issue-specific policies are formalized as written documents, distributed to users, and agreed to in writing, systems-specific policies (SysSPs) are frequently codified as standards and procedures used when configuring or maintaining systems. One example of a SysSP is a document describing the configuration and operation of a network firewall. This document could include a statement of managerial intent, guidance to network engineers on selecting, configuring, and operating firewalls, and an access control list that defines levels of access for each authorized user. Systems-specific policies can be organized into two general groups, management guidance and technical specifications.

Management Guidance SysSPs

A management guidance SysSP is created by management to guide the implementation and configuration of technology intended to support the security of information. For example, while the specific configuration of a firewall belongs in the technical specifications SysSP, the general construction and implementation of the firewall must follow guidelines established by management. For example, an organization may not want its employees to have access to the Internet via the organization's network; the firewall would have to be implemented according to this rule.

Systems-specific policies can be developed at the same time as ISSPs, or they can be prepared in advance of their related ISSPs. Before management can craft a policy informing users what they can do with the technology and how they may do it, it might be necessary for system administrators to configure and operate the system. Some organizations may prefer to develop ISSPs and SysSPs in tandem, so that operational procedures and user outcomes are developed at the same time.

Technical Specifications SysSPs

While a manager may work with a systems administrator to create managerial policy as specified previously, the system administrator may need to create a different type of policy to implement the managerial policy. Each type of equipment has its own type of policies, which are used to translate the management intent for the technical control into an enforceable technical approach. For example, an ISSP may require that user passwords be changed quarterly; a systems administrator can implement a technical control within a specific application to enforce this policy. There are two general methods of implementing such technical controls, access control lists – which include the user access lists, matrices, and capability tables that govern the rights and privileges of users, and configuration rules - the specific configuration codes entered into security systems to guide the execution of the system when information is passing through it.

FUTURE TRENDS

In order to deal with the complexities of developing and implementing policies, organizations are increasingly turning to alternate solutions. These alternate solutions provide options ranging from templates based on established experts in the field (i.e., Charles Cresson Wood), to automated policy approval and distribution systems like Security Policy Management from NetIQ (Security Policy Management, 2004). These systems simplify the onerous task of drafting policy, obtaining management approval, distributing to end users, and documenting compliance with policy but creating a structure in which the draft policy is placed. Control and approval is passed from author to reviewer, and eventually published to end users. Once users have read the policy, the system documents their activities and eventually can provide quizzes on policy content. Use of systems like these greatly improves the organization's ability to issue and manage policy as an effective tool in supporting ongoing operations.

CONCLUSION

The early years of the 21st century have seen the emergence of information security as both a practical area of specialization in information technology and as an academic discipline in post-secondary education. As many new members join the information security community, it is important that the primary role of policy as the mechanism whereby an organization defines what is to be secured is clearly understood. Without sound policy as a foundation, policy constructed with the same care and attention to detail required by all parts of the information security mission, an organization is less likely to be successful in its mission to protect information assets.

REFERENCES

Bergeron, F., & Bérubé, C. (1990, December). End users talk computer policy. *Journal of Systems Management, 41*(12), 14-17.

Merriam-Webster Online. Retrieved June 24, 2002, from *http://www.m-w.com/cgi-bin/dictionary*

NIST. (1989). Special publication 500-169: Executive guide to the protection of information resources. National Institute of Standards and Technology. Retrieved September 5, 2002, from *http://csrc.nist.gov/publications/nistpubs/500-169/sp500-169.txt*

NIST. (1995). Special publication 800-12: An introduction to computer security: The NIST handbook. National Institute of Standards and Technology. Retrieved August 15, 2002, from *http://csrc.nist.gov/publications/nistpubs/800-12/*

NIST. (1996). Special publication 800-14: Generally accepted principles and practices for securing information technology systems. National Institute of Standards and Technology. Retrieved September 10, 2002, from *http://csrc.nist.gov/publications/nistpubs/800-14/800-14.pdf*

Security Policy Management. (2004). Create, communicate and enforce security policy across your organization. NetIQ Corporation. Retrieved June 1, 2004, from *http://www.netiq.com/solutions/security/policy.asp*

Whitman, M.E. (2003, August). Enemy at the gates: Threats to information security. *Communications of the ACM, 46*(8), 91-96.

Whitman, M.E., & Mattord, H.J. (2003). *Principles of information security*. Boston: Course Technology.

Whitman, M.E., & Mattord, H.J. (2004). *Management of information security*. Boston: Course Technology.

Whitman, M.E., Townsend, A.M., & Aalberts, R.J. (1999, June). Considerations for an effective telecommunications-use policy. *Communications of the ACM*, *42*(6), 101-108.

Wood, C.C. (2003). *Information security policies made easy* (9th ed.). NetIQ Corporation.

WUSTL. (2002). Policies and procedures: Information security policy. Washington University, St. Louis. Retrieved April 15, 2002, from *http://www.wustl.edu/policies/infosecurity.html*

KEY TERMS

Access Control List (ACL): A list of people or other entities permitted to access a computer resource.

Capability Table: Synonymous with capabilities table. A list that specifies data items or physical devices (for example printers) that users are authorized to access.

Defense in Depth: The multiple levels of security controls and safeguards that an intruder faces.

Enterprise Information Security Policy (EISP): A policy that sets the strategic direction, scope, and tone for all of an organization's security efforts.

Issue-Specific Security Policy (ISSP): Policies that provides detailed, targeted guidance to instruct all members of the organization in the use of technology-based systems.

Policy: A body of expectations that describe acceptable and unacceptable behaviors of employees in the workplace.

Standards: Detailed statements of actions that comply with policy.

Systems-Specific Security Policy: Policies codified as standards and procedures used when configuring or maintaining systems.

Innovation Link Between Organization Knowledge and Customer Knowledge

Helen Mitchell
Unitec, Auckland, New Zealand

INTRODUCTION

In 1959 Penrose referred to the importance of knowledge for using resources more innovatively and profitably, and in the same year Drucker indicated greater value should be placed on knowledge workers. An article by Nonaka (1991) suggested that the source of lasting competitive advantage is knowledge. Since then there has been a growing interest in knowledge, and an emerging view that the sharing of knowledge and creation of new knowledge leading to innovations is the key to providing future economic growth. In 1998, Amidon referred to the survey in 1996 produced by *The Economist* describing how rich economies will increasingly depend upon "their ability to innovate" (p. 24).

THE ORGANIZATION

At the corporate level it is important for organizations to recognize the value of knowledge. An environment accepting that knowledge sharing is the norm provides a conduit for the generation of innovations. Continuous innovation is necessary for survival in what is now a highly competitive global environment. Innovation springs from the knowledge within the organization. Corporate knowledge is accumulated through its systems and procedures, its databases and its technology, and from its employees sharing and applying their knowledge to the operational requirements of the business. The interaction of employees with external parties such as customers, suppliers, industry contacts and those residing in the wider community also provides a considerable source of knowledge for the organization.

Employees bring to an organization their individual knowledge bank gathered as a result of their education and experience of life. The longer they work in an organization, the more knowledge they accumulate about its operations, products and customers. Nonaka (1991) refers to the spiral of knowledge in the organization and says that knowledge begins with the individual. Taking this further, Inkpen (1996), when referring to the spiral of organizational knowledge creation, indicates the importance of moving knowledge from the level of the individual to group level and on upwards to organizational level.

It is perhaps worth taking cognizance of Davenport and Prusak (1998, p.108), who say, "Despite the corporate mantra that employee knowledge is a valuable resource, most firms do not make concerted efforts to cultivate the knowledge oriented activities of their personnel". Although knowledge comes from employees, Teece (1998, p. 62) indicates it is the organization that provides "the physical, social, and resource allocation structure so that knowledge can be shaped into competences". It is worth remembering that Drucker (1994, p. 71) said it is the knowledge workers who own the tools of production. Previously, "The industrial workers needed the capitalist infinitely more than the capitalist needed the industrial worker," but in the knowledge society organizations "…need knowledge workers far more than knowledge workers need them". It should be kept in mind that employees take their knowledge home with them each evening, and in some situations they may not return, thus depriving the organization of valuable knowledge. However, if it is recognized that employees have knowledge of value, then organizations need to maximize the benefit of that knowledge, and determine how to harness it to the advantage of the organization and the individual (Mitchell, 2002a).

When assessing the potential of knowledge, Carneiro (2000, p. 87) suggests organizations "…need to look for the knowledge that is able to add value. Value adding knowledge is very different to an information-mix." Lester (2001) also advises, "Managing knowledge is not like managing information, it involves focusing on, nurturing and winning the trust of the professional knowledge workers and the confidence of customers and suppliers" (p. 172). This means organizations need to look closely at encouraging a culture involving the development of good relationships based around trust, and in which knowledge is readily shared. Sharing knowledge increases the knowledge among those who share and adds value to the organization while at the same time making it difficult for competitors to copy.

It is, therefore, important for organizations to try to move the tacit knowledge of its employees into explicit knowledge to be embedded in the organization's knowledge repository for all to share. Codifying tacit knowledge is not easy, and all the knowledge the employee holds will never be captured. However, knowledge that can be codified becomes available for sharing with others in the organization. Knowledge sharing and creation of new knowledge emerges as a result of working in teams where problems that arise need to be resolved. Opportunities for social interaction should be encouraged – it is not unknown for useful knowledge to be exchanged in such an environment. While sharing knowledge for the benefit of the organization involves everyone, knowledge is also shared, wittingly, or unwittingly, with those external to it. This can present a two-edged sword. While knowledge can be lost, it can also be gained. However, from the knowledge of all those involved in the knowledge sharing process, along with the knowledge held within the structure of the organization, there emerges collective knowledge. Whether it is focused within the organization, or encompassing knowledge external to it, collective knowledge provides a catalyst for the development of innovations (Mitchell, 2002b).

THE ORGANIZATION AND INNOVATION

According to Kanter (1996, p. 94), "Structures and practices that may work well for the perpetuation of the known tend to be at odds with innovation." Creativity is a product of knowledge, and shared knowledge provides opportunities for organizations to be innovative – an important element for moving ahead (Mitchell, 2002b). Innovative organizations are growing organizations, with 3M being a classic example; therefore the organizational environment is important if innovation is to occur.

Amidon (1997) refers to Japanese firms placing customers at the heart of the innovation process. She goes on to say that innovation should be a strategic focus and to make a commitment to work with customers. Amidon provides a framework to assess the ability for organizations to develop ideas and take them to market. The innovation assessment encompasses 10 modules, identified as: collaborative process; performance measures; education/development; learning network; market positioning; products/services; market penetration; market image, leadership/leverage; technology/internet. For each module there are questions to be asked, and answered, and the model used views the organization from both an internal and external perspective. According to Amidon, executives who have followed the process have found creative ways of improving efficiency and effectiveness. Innovation is not solely dependent on new knowledge. Ideas emerge from unexpected sources and applying knowledge, recognizing potential and seeing opportunities is what brings innovative ideas to fruition (Mitchell, 2002d). As Drucker (1998) indicated, innovation needs knowledge, ingenuity and focus.

The view put forward by Johannessen, Olaisen, and Olsen (1999, p. 116) is that "Managers need to focus their attention on innovation. …create commitment…initiate change. To enhance innovation managers also need a high level of integrity. This means the ability to create trust …". Johannessen et al. identified the characteristics of people who manage the innovation process as being proactive in creating opportunities; willing to take risks; and goal setting. It is also necessary for managers to recognize the importance of innovation and its relationship with the knowledge of employees, its customers, suppliers and other external parties, and the need to promote an innovative culture.

However, management must thoroughly understand the level of knowledge, expertise, skills and competencies resident within the organization, and be able to identify what additional capabilities are needed if it is to successfully work for the development of innovations in a collaborative environment. Organizations that are aiming for success will maximize the benefit, for example, of cross-functional teams from which a montage of knowledge provides the opportunity for the creation of new knowledge from which new innovations can emerge. It may, therefore, be necessary for organizations to be prepared for changes arising from introducing an innovative environment that encourages closer working relationships with customers. No organization stands in isolation. Every organization is dependent upon and has responsibilities for its customers, suppliers, stakeholders and the wider community. There are real opportunities for organizations to effectively work together to become more innovative

The Organization, Innovation and Customer Relationships

"Customer loyalty is an asset" (Brooking, 1996, p. 26). Getting to know about customers – their needs, wants, expectations – is critical to the ongoing success of the company. Brooking identifies what she terms "Customer Audit Questions" that include: who are we selling to, why do those customers buy from us, who are repeat buyers, how do we monitor such events, and how often are customers contacted. Armed with such a knowledge base, the organization is in the position of being not only informed about its customers but also in a positive situation when interacting with them.

Research carried out in 1999 by Mitchell (2002c) found that organizations routinely increase their knowledge of customers in many ways, for example through the salesperson, credit cards, sales analysis, warranty cards, customer databases, surveys, and customer focus groups. However, many organizations do not give sufficient recognition to the importance of mining those sources for hidden gems of knowledge that could provide the means through which to enhance the potential of the business and at the same time more effectively meeting the needs of its customers. While knowledge about customers can be obtained through the organization's resources, more importantly it will come from associating closely with customers (Amidon, 1997; Byrne, 1993; Evans & Wurster, 1997; Fahey & Prusak, 1998; Kanter, 1996; Jordan & Jones, 1997; Teece, 1998).

Working collaboratively with a customer to jointly develop a product, or for the introduction of a new technology, brings benefits to both parties and there is a likely spin-off for other customers (Tidd, Bessant & Pavitt, 2002). Tidd et al. make reference to the approach taken by international energy and construction firms of having "policies of collaboration with customers overseas and source technologies from countries in which they wish to sell products" (p.178). Actively promoting a collaborative approach opens the way to being able to respond positively to opportunities as they arise.

Organizations working alongside customers on new products, or process innovation, or designing innovative solutions to meet specific needs, adds value to both the organization and its customers (Grant, 1997). Rowley (2002, p. 500) also supports the employee/customer connection: "Another rich seam of customer knowledge resides with the people who interact with the customer, whether they be the promoters of small business in B2B interacts or the service agent or retail assistant in B2C interactions." Suppliers to organizations must not be forgotten; after all, the organization is the suppliers' customer and there is much to gain from working collaboratively.

Research being undertaken by Mitchell in 2002/3 indicates that businesses are reaping benefits through working closely with customers. Some of those businesses have produced a number of innovative products, increased their intellectual property portfolio, and as a result of commercialization of those products have created increased wealth. Their customers are benefiting by gaining new and enhanced products. The following are some examples found by Mitchell of organizations successfully working collaboratively with customers:

A food manufacturing company that generates a number of ideas works closely with its customers and suppliers for all new products. A customer with an idea will put it to the company to assess its feasibility. One product the company is involved with was the result of such an approach and it has triggered a revolutionary method of production by the food company, resulting in a new product. The technological development was hailed as a breakthrough at an international trade fair, and has been welcomed by other companies involved in making similar products. An additional service provided by the company was the development of a document of agreement for its clients to protect their rights.

A packaging company worked alongside a customer to design a sophisticated system of packaging to prevent damage to a very delicate, perishable product. The success of the packaging has resulted in reducing damage claims to virtually zero. The customer is very satisfied, as are the customer's customers, and the packaging company developed technology that can be transferred to other products in its range.

Sales representatives working for an electronic products distributor keep a weekly log on feedback and suggestions and all suggestions are reviewed and followed up. A Web site has been developed for customer feedback. It was found that children, because of their interest in electronic games, give very good feedback and provide very good ideas. On the commercial side, customers will suggest specific types of software be developed to complement a product, or ask for the package to be designed as a mobile system.

Other benefits of working with customers are that they have considerable knowledge not only about an organization's products and how they are used, but also about the products of other suppliers. Observing how products are used provides valuable feedback that may instigate modifications, new customer solutions, or the customizing of products. It is also important to know what customers are thinking, and their rationale for changing to another supplier (Fahey & Prusak, 1998), and this is another reason to keep in close contact. Kanter (1996) suggests focusing on demanding customers. Difficult as these customers may be, it is a sensible approach as it is those customers who are likely to produce the greatest and most valuable knowledge, which may well lead to new innovations.

CONCLUSION

Without customers, organizations will not survive. As competition in the global market increases, the value of knowledge will become more important and working collaboratively with customers more critical. The future of organizations is dependent upon knowledge and the directing of it towards innovation. Organizations need to satisfy their customers, and their customers' customers. Linking the knowledge of the organization with the knowledge of the customer has the potential to lead to innovative activity opening up an avenue through which to combat competition while at the same time providing a sense of achievement and satisfaction to those involved.

REFERENCES

Amidon, D.M. (1997). *The ken awakening*. Butterworth-Heinemann, Newton.

Amidon, D.M. (2000). Blueprint for 21st century innovation management. *Journal of Knowledge Management, 2*(1), 23-31.

Byrne, J. (1993, February 8). The virtual corporation. *Business Week,* 98-102.

Carneiro, A. (2000). How does knowledge management influence innovation and competitiveness? *Journal of Knowledge Management, 4*(2), 87-98

Davenport, T.H., & Prusak, L. (1998). *Working knowledge*. Harvard Business School Press.

Drucker, P.F. (1994, November). The age of social transformation. *The Atlantic Monthly*, 53-80.

Drucker, P.F. (1998, November/December). The discipline of innovation. *Harvard Business Review,* 149-157.

Evans, P.B., & Wurster, T.S. (1997, September/October). Strategy and the new economics of information. *Harvard Business Review,* 71-82

Fahey, L., & Prusak, L. (1998, Spring). The eleven deadliest sins of knowledge management. *California Management Review, 40*(3), 265-276.

Grant, R.M. (1997). The knowledge-based view of the firm: Implications for management practice. *Long Range Planning, 30*(3), 450-454.

Inkpen, A.C. (1996, Fall). Creating knowledge through collaboration. *California Management Review, 39*(1), 123-140.

Johannessen, J., Olaisen, J., & Olsen, B. (1999). Managing and organizing innovation in the knowledge economy. *European Journal of Innovation Management, 2*(3), 116-128.

Jordan, J., & Jones, P. (1997). Assessing your company's knowledge management style. *Long Range Planning, 30*(3), 392-398.

Kanter, R.M. (1996). When a thousand flowers bloom: Structural, collective, and social conditions for innovation in organizations. In P.E. Myers (Ed.), *Knowledge management and organizational design.* Butterworth-Heinemann, Newton.

Lester, M. (2001, September). Innovation and knowledge management: The long view. *Creativity and Innovation Management, 10*(3), 165-176.

Mitchell, H.J. (2002a, February). Recognising the value of knowledge to create innovations. *Sixth International Research Conference on Quality, Innovation and Knowledge Management*, Kuala Lumpur, Malaysia.

Mitchell, H.J. (2002b). *Knowledge sharing for innovation and wealth.* Unpublished paper.

Mitchell, H.J. (2002c). The innovation link between organisation knowledge and customer knowledge. In D.E. White (Ed.), *Knowledge mapping and management*. Hershey, PA: IRM Press.

Mitchell, H.J. (2002d, November). The knowledge pathway to innovation. *Second Doing Business Across Boarders Conference,* Newcastle, Australia.

Nonaka, I. (1991, November/December). The knowledge-creating company. *Harvard Business Review,* 97-104

Penrose, E.T. (1963). *The theory of the growth of the firm*. UK: Oxford Basil Blackwell.

Rowley, J. (2002). Eight questions for customer knowledge management in e-business. *Journal of Knowledge Management, 6*(5), 500-511.

Teece, D.J. (1998, Spring). Capturing value from knowledge assets: The new economy, markets for know-how, and intangible assets. *California Management Review, 40*(3), 55-79.

Tidd, J., Bessant, J., & Pavitt, K. (2002). Learning through alliances. In J. Henry & D. Mayle (Eds.), *Managing innovation and change.* London: Sage Publishers.

KEY TERMS

Collective Knowledge: The knowledge of the organization, including the knowledge of its employees, customer/supplier and industry knowledge that can be actioned to bring about innovation.

Corporate Knowledge: The knowledge owned by an organization – its databases, technology, beliefs, culture, structure, processes and procedures. The organization has access to, but does not own, the knowledge of employees, and through research acquires knowledge of the external environment.

Explicit/Codified Knowledge: Knowledge that can be written down, for example a formula, policies and procedures, or a recipe, and readily available for sharing with others.

Innovation: Something new. It can be a product or a process and must provide benefit to the organization and to the advancement of society. An invention only becomes an innovation when it has become commercially successful.

Knowledge Sharing: Involves giving people some of the knowledge you have. It may be about how to do a particular job, where to find information, knowledge of a customer's products, and of industry developments.

Tacit Knowledge: Knowledge owned by the individual. It is personal and is developed through life experiences and education and is not readily shared or made explicit.

Innovations for Online Collaborative Learning in Mathematics

Rodney Nason
Queensland University of Technology, Australia

Earl Woodruff
OISE - University of Toronto, Canada

INTRODUCTION

The field of computer-supported collaborative learning (CSCL) has been growing in a number of areas and across a number of subjects (Koschmann, 1996; Koschmann, Hall, & Miyake, 2002; Wasson, Baggetun, Hoppe, & Ludvigsen, 2003). One of the most promising pedagogical advances, however, for online collaborative learning that has emerged in recent years is Scardamalia and Bereiter's (1996) notion of knowledge-building communities. Unfortunately, establishing and maintaining knowledge-building communities in CSCL environments such as Knowledge Forum® in the domain of mathematics has been found to be a rather intractable problem (Bereiter, 2002b; Nason, Brett, & Woodruff, 1996). In this chapter, we begin by identifying two major reasons why computer-supported knowledge-building communities in mathematics have been difficult to establish and maintain.

1. The inability of most "textbook" math problems to elicit ongoing discourse and other knowledge-building activity
2. Limitations inherent in most CSCL environments' math representational tools

Therefore, in this chapter, we argue that if mathematics education is to exploit the potentially powerful new ways of learning mathematics being provided by online knowledge-building communities, then the following innovations need to be designed and integrated into CSCL environments:

1. authentic mathematical problems that involve students in the production of mathematical models that can be discussed, critiqued, and improved, and
2. comprehension-modeling tools that (a) enable students to adequately represent mathematical problems and to translate within and across representation modes during problem solving, and (b) facilitate online student-student and teacher-student hypermedia-mediated discourse.

Both of the above innovations are directed at promoting and sustaining mathematical discourse. The requirement that the mathematical problems need to be authentic ensures that the students will have the contextual understanding necessary to promote a discussion about the mathematical models. Comprehension-modeling (Woodruff & Nason, 2003) further promotes the discourse by making student understanding yet an additional object for discussion.

Most textbook math problems do not require multiple cycles of designing, testing, and refining (Lesh & Doerr, in press), and therefore do not elicit the collaboration between people with special abilities that most authentic math problems elicit (Nason & Woodruff, 2004). Another factor that limits the potential of most textbook math problems for eliciting knowledge-building discourse is that the answers generated from textbook math problems do not provide students with much worth discussing (Bereiter, 2002b).

Another factor that has prevented most students from engaging in ongoing discourse and other mathematical knowledge-building activity within CSCL environments is the limitations inherent in their mathematical representational tools (Nason et al., 1996). Most of these tools are unable to carry out the crucial knowledge-building functions of (a) generating multiple representations of mathematical concepts, (b) linking the different representations, and (c) transmitting meaning, sense, and understanding.

Two clear implications can be derived from this review of the previous research. First is that different types of mathematical problems that have more in common with the authentic types of mathematical problems investigated by mathematics practitioners than most existing types of textbook math problems need to be designed and integrated into CSCL environments. Second, a new generation of iconic mathematical representation tools also needs to be designed and integrated into CSCL environments. In order to differentiate these tools from previous iconic math representation tools, we have labeled our new generation of tools as comprehension-modeling tools.

Each of these two issues will be discussed in the next two sections of this chapter.

AUTHENTIC MATH PROBLEMS

Credence for the viewpoint that the integration of more authentic types of mathematical problems into CSCL environments may lead to conditions necessary for the establishment and maintenance of knowledge-building activity is provided by the findings from two recent research studies conducted by the coauthors. Although both of these studies were situated within elementary schools, it should be noted that the same math problems used in these research studies could also be used within online CSCL environments to facilitate the development of mathematical subject-matter knowledge in high school students and preservice teacher-education students. Therefore, we believe that the findings from these two studies have much relevance for the establishment and maintenance of math knowledge-building communities not only in elementary schools, but also in secondary school and higher education institutions, too.

In a series of research studies, Nason, Woodruff, and Lesh have been investigating whether having students engage in model-eliciting mathematical problems with collective discourse mediated by Knowledge Forum would achieve authentic, sustained, and progressive online knowledge-building activity. In this section, we focus on two of these research studies.

In the first of the research studies (Nason & Woodruff, 2004), 21 students in a Grade-6 class at a private urban Canadian school for girls were asked to devise an alternative model that could be used for ranking nations' performance at the Olympic games that de-emphasized the mind-set of "gold or nothing." In the second research study (Nason, Woodruff, & Lesh, 2002), 22 students in another Grade-6 class at the same school were asked to build a model that could help rank Canadian cities in terms of quality of life.

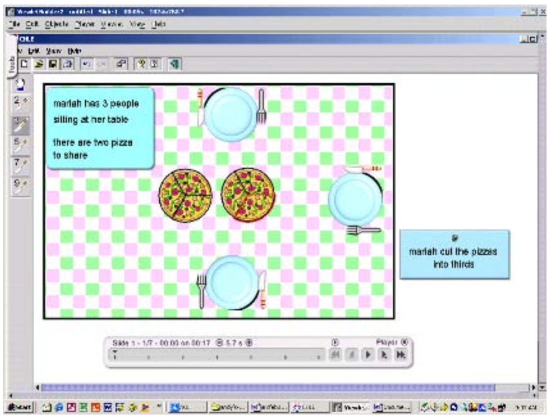
In both studies, the students were initially presented with an article setting the scene for the model-eliciting activity and a set of focus questions based on the article. After this 45-minute warm-up activity, the students went through the phases of (a) initial model building (Phase 1, one session of 45 minutes), (b) sharing of initial models (Phase 2, one session of 45 minutes), and (c) iterative online critiquing and revision of models within Knowledge Forum (Phase 3, four sessions of 45 minutes). The sharing of the initial models in Phase 2 was done face to face within the classroom. After the face-to-face sharing of the initial models had been completed, each group attached their math model to a Knowledge Forum note where it could be viewed and evaluated by other participants within the online CSCL community. During the online critiquing and revision of models in Phase 3, Knowledge Forum provided the contexts and scaffolds for intergroup online discourse.

Five important elements of activity consistent with Scardamalia's (2002) principles of knowledge building were observed during the course of these two studies.

1. Redefinition of the problems, which highlights Scardamalia's principles of improvable ideas and rising above
2. Inventive use of mathematical tools, which highlights Scardamalia's principle of improvable ideas
3. Posing and exploration of conjectures, which highlights Scardamalia's principles of idea diversity and knowledge-building discourse
4. Collective pursuit of the understanding of key mathematical concepts, highlighting Scardamalia's principles of community knowledge and collective responsibility
5. Incremental improvement of mathematical models, which highlights Scardamalia's principle of improvable ideas

Much of the success in establishing and maintaining the online mathematics knowledge-building communities in these two studies can be attributed to the rich context for mathematical knowledge-building discourse provided by the model-eliciting problems. In both problems, students were required to produce a mathematical model for issues that the students found meaningful and relevant. Therefore, they were willing to proceed through multiple cycles of developing, evaluating, and revising their models. This process of proceeding through multiple cycles encouraged much online discourse between the groups in each classroom. The model-eliciting problems also had many different possible solutions. Because of this, there was much heterogeneity in the initial models produced by the groups of students. In order to understand other groups' models and also to explain their own model to other groups, each group had to engage in much iterative online discourse with other groups. During this discourse, they had to ask good questions, propose how other groups' models could be improved, and elaborate on and/or modify their explanations. Finally, the models themselves provided students with artifacts that could be discussed, evaluated, compared, and improved (just like the artifacts built by mathematics practitioners). Unlike the answers produced in most textbook problems that tend to only enable discourse about correctness (or incorrectness), the models produced from the model-eliciting problems were artifacts that could be evaluated

Figure 1. Screen shot of the problem students are attempting to solve



and discussed in terms of not only correct usage of mathematical concepts and processes, but also in terms of subjective, nonmathematical factors.

COMPREHENSION MODELING

Evidence to support the notion that the inclusion within a CSCL learning environment of comprehension-modeling tools can do much to facilitate knowledge-building discourse has been provided by research during the development of CHiLE (constructivist hypermedia interactive learning environment; Charles & Nason, 2000). CHiLE situates the learning of fractions in the context of a restaurant in which the children play the role of a waiter and are asked to partition and share out equal objects such as pizza and apple pies to customers sitting at the restaurant table (see Figure 1). The number of customers sitting at the table and the number of objects to be partitioned and shared can be varied. CHiLE provides the children with five different slicers (a knife-like tool) that enable objects to be cut in halves, thirds, fifths, sevenths, and ninths. With these slicers, the children can also create other fractions such as quarters (by halving the halves) and sixths (by halving the thirds).

CHiLE enables children to generate multiple representations of fraction problems and provides the iconic tools for facilitating synchronous hypermedia-mediated, knowledge-building, child-child and teacher-child discourse. CHiLE, however, has an added facility that enables teachers and children to also engage in online asynchronous knowledge-building discourse. With CHiLE, children are able to make an animated sequence of slides with accompanying text that not only enables them to communicate the solution to a fraction problem, but also the process (or model) that was used to generate the solution. CHiLE thus uses hypermedia as a way to animate and promote mathematical discourse: The strategy (or model) is reified on the screen via the iconic representation, the animation shows the “story,” and everything is recorded, thus promoting reflection and revisitation. This is illustrated below in a series of figures (see Figures 1 to 4) generated by two 8-year-old children who had been asked to share one pizza fairly between three people.

Nason and Woodruff (2004) have found that online knowledge-building discourse facilitated by CHiLE operates on two different levels. First, the discourse can occur at the global level and focus on the overall strategy (or model). For example, another group of students, when given the same problem as in Figure 1, decided to slice the pizza into sixths and give two sixths to each person. After looking at one another’s models, the two groups of students engaged in robust online debate about which

Figure 2. Early screen shot of students’ initial steps toward a solution

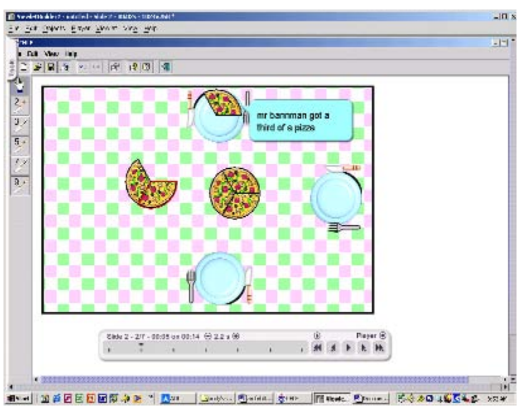


Figure 3. Screen shot midway toward a solution

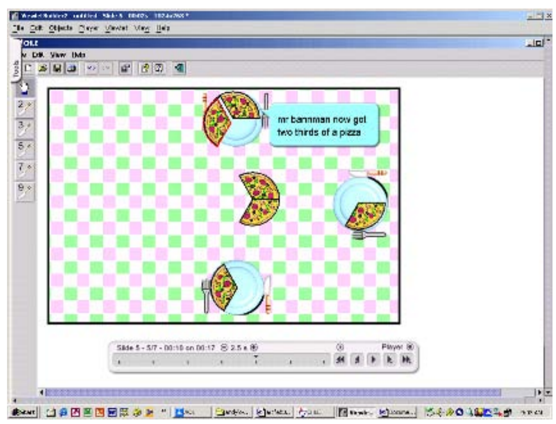
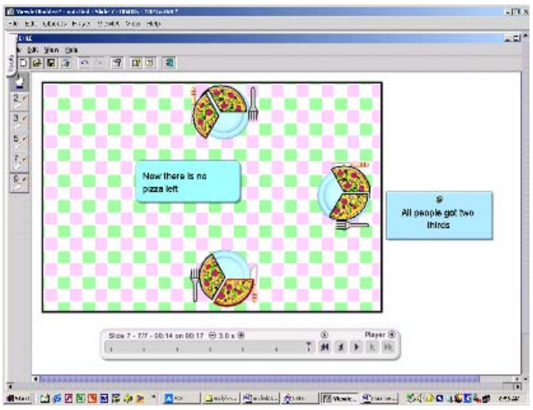


Figure 4. Screen shot of students' proposed solution



was the "better" strategy (and solution). During this debate, they were able to identify similarities and differences between the strategies, but more importantly, build conceptual links between thirds and sixths. Second, the discourse can occur at the language level. For example, there can be discourse about the best language to insert in a sequence of slides. This discourse often provides the contexts for the introduction of formal mathematical language as a more precise way of communicating meaning within mathematical contexts than natural language.

The hypermedia facilities provided by CHiLE enabled children to engage in online knowledge-building discourse synchronously and asynchronously via iconic, natural language, and/or mathematical language representations. CHiLE thus provided one of the most important dynamics that Scardamalia (2002) identified as being a technical determinant of knowledge building and knowledge advancement within online CSCL environments.

CHiLE also provided two other dynamics that Scardamalia indicated were technological determinants of knowledge building and knowledge advancement. First, there is her notion that computer technology should include facilities for bringing together different ideas in such a way that productive use can be made of diversity. The iconic tools provided by CHiLE met this criterion by enabling children to readily

1. generate diverse solutions and solution processes to the same mathematics problem, and
2. communicate both synchronously and asynchronously via the iconic models, natural and mathematical language, and mathematical symbols their diverse solutions and solution processes to others within the online learning community.

Scardamalia also indicated that the computer technology also should provide children with the opportunity and the means to make revisions. Without this, she claimed that children will not be able to work continuously to improve the quality, coherence, and the utility of their ideas. One of the major qualities of CHiLE is the ease with which children can revisit and revise the sequences of slides and their accompanying text. The comprehension-modeling tools provided by CHiLE thus promoted idea diversity, improvable ideas, and knowledge-building discourse, three of the sociological and technological determinants of knowledge building identified in Scardamalia (2002).

FUTURE TRENDS

The research in progress reported in the previous two sections indicates that the inclusion of model-eliciting problems and of comprehension-modeling tools (such as CHiLE) into online collaborative learning environments both have the potential to facilitate the establishment and maintenance of online collaborative mathematics knowledge-building communities in schools and higher education institutes.

However, two important issues still need to be addressed before this potential can be realized. First, the set of principles for informing the design of model-eliciting problems developed by Lesh and Doerr (2003) need to be modified to take cognizance of the differences between online collaborative and traditional classroom environments. Second, the theoretical framework informing the design of comprehension-modeling tools needs to be modified to include not just ideas from research into external mathematical representations that were used to inform the design of CHiLE (e.g., Kaput, 1992; Olive, 2000), but also ideas from research conducted in other areas such as online collaboration (e.g., Klopfer & Woodruff, 2002), cognitive science, and multimedia learning (e.g., Mayer, 2001; Sweller, 1999). Both these issues are the foci of a series of design experiments (Bereiter, 2002a) currently being conducted by Nason and Woodruff.

CONCLUSION

In this chapter, we identified two major reasons why mathematics educators have had limited success in establishing and maintaining online knowledge-building communities.

1. The inability of most textbook math problems to elicit ongoing discourse and other knowledge-building activity

2. Limitations inherent in most CSCL environments' math representational tools

We then proposed how these two problems could be overcome, namely, by the inclusion of mathematical problems that children can analyze and describe through a mathematical model (such as the steps necessary to divide two pizzas among three people) and comprehension-modeling tools (that allow observers to later see how the students have solved the problem) within CSCL environments.

We have targeted our discussion within one CSCL groupware product called Knowledge Forum, but we believe the same principles will apply to any online computer-supported collaborative learning system. To that end, we argued that the development of model-eliciting problems suitable for use in online CSCL environments and of comprehension-modeling tools is being restricted by the lack of adequate theoretical frameworks to inform the research and development of these two types of artifacts. Therefore, we have proposed that the development of adequate theoretical frameworks to inform the design of these two types of artifacts should be a major research priority in this field.

REFERENCES

Bereiter, C. (2002a). Design research for sustained innovation. *Cognitive Studies, Bulletin of the Japanese Cognitive Science Society*, *9*(3), 321-327.

Bereiter, C. (2002b). *Education and mind in the knowledge age*. Mahwah, NJ: Erlbaum.

Charles, K., & Nason, R. A. (2000). Towards the specification of a multimedia environment to facilitate the learning of fractions. *Themes in Education*, *1*(3), 263-288.

Kaput, J. J. (1992). Technology and mathematics education. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 515-556). New York: Macmillan.

Klopfer, E., & Woodruff, E. (2002). *The impact of distributed and ubiquitous computational devices on the collaborative learning environment.* Proceedings from the Annual CSCL Conference, Boulder, CO.

Koschmann, T. (Ed.). (1996). *CSCL, theory and practice of an emerging paradigm*. Mahwah, NJ: L. Erlbaum.

Koschmann, T., Hall, R., & Miyake, N. (Eds.). (2002). *CSCL2: Carrying forward the conversation.* Mahwah, NJ: L. Erlbaum.

Lesh, R., & Doerr, H. (2003). Foundations of a models and modelling perspective on mathematics teaching, learning and problem solving. In H. Doerr & R. Lesh (Eds.), *Beyond constructivism: A models and modelling perspective on mathematics learning, problem solving and teaching*. Mahwah, NJ: Erlbaum.

Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.

Nason, R. A., Brett, C., & Woodruff, E. (1996). Creating and maintaining knowledge-building communities of practice during mathematical investigations. In P. Clarkson (Ed.), *Technology in mathematics education* (pp. 20-29). Melbourne: Mathematics Education Research Group of Australasia.

Nason, R. A., & Woodruff, E. (2004). Online collaborative learning in mathematics: Some necessary innovations. In T. Roberts (Ed.), *Online learning: Practical and theoretical considerations* (pp. 103-131). Hershey, PA: Idea Group Inc.

Nason, R. A., Woodruff, E., & Lesh, R. (2002). Fostering authentic, sustained and progressive mathematical knowledge-building activity in CSCL communities. In B. Barton, C. Irwin, M. Pfannkuch, & M. O. J. Thomas (Eds.), *Mathematics education in the South Pacific* (Proceedings of the Annual Conference of the Mathematics Education Research Group of Australasia, Auckland, pp. 504-511). Sydney, Australia: MERGA.

Olive, J. (2000). Computer tools for interactive mathematical activity in the elementary school. *International Journal of Computers for Mathematical Learning*, *5*(3), 241-62.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67-98). Chicago: Open Court.

Scardamalia, M., & Bereiter, C. (1996). Adaptation and understanding: A case for new cultures of schooling. In S. Vosniadou, E. De Corte, R. Glaser, & H. Mandel (Eds.), *International perspectives on the psychological foundations of technology-based learning environments* (pp. 149-165). Mahwah, NJ: Lawrence Erlbaum Associates.

Sweller, J. (1999). *Instructional design in technical areas*. Melbourne: ACER.

Wasson, B., Baggetun, R., Hoppe, U., & Ludvigsen, S. (Eds.). (2003). *CSCL2003: Community events, communication and interaction*. Bergen, Norway: University of Bergen.

Woodruff, E., & Nason, R. (2003). Math tools for knowledge-building and comprehension modeling in CSCL. In B. Wasson, R. Baggetun, U. Hoppe, & S. Ludvigsen (Eds.), *International Conference on Computer Support for Collaborative Learning, CSCL 2003: Community Events, Communication and Interaction* (pp. 31-34). Bergen, Norway: University of Bergen.

KEY TERMS

Comprehension-Modeling Tools: Math representation tools that enable users to (a) generate multiple representations of mathematical concepts and processes, (b) dynamically link the different representations, (c) communicate the mathematical ideas they have constructed, and (d) make movie-like sequences of animation slides that enable others to replay the process used to generate the solution.

Computer-Supported Collaborative Learning: collaborative learning mediated by computers.

CSCL: Acronym for computer-supported collaborative learning.

Knowledge Building: Production and improvement of knowledge objects that can be discussed, tested, compared, hypothetically modified, and so forth, and not simply the completion of school tasks.

Knowledge Forum ®: A single, communal multimedia database designed to facilitate computer-supported collaborative learning.

Mathematical Representations: concrete, pictorial, and symbolic models used to represent mathematical ideas.

Model-Eliciting Problems: Mathematical problems that involve producing models for constructing, describing, explaining, manipulating, predicting, and controlling complex systems (Lesh & Doerr, 2003).

Problem Solving: Situation involving an initial state, a goal (or solution) state, and a blockage between the initial and goal states that requires the construction of new knowledge to proceed from the initial to the goal state.

Innovative Thinking in Software Development

Aybüke Aurum
University of New South Wales, Australia

INTRODUCTION

As we enter the third millennium, organizations have to cope with accelerating rates of change in technology and increased levels of competition on a global scale more than ever before. In order to stay competitive within this changing business environment, organizations are forced to constantly pursue new strategies to differentiate themselves from their competition, such as offering a stream of new products and services (Satzinger et al., 1999). Furthermore, there is growing recognition that an organization's capability to deal with change, improve services and quality, cut costs, develop new products, and compete in a global market will depend upon the level of creative and innovative thinking of its workforce (Covey, 1989). In short, in order to remain competitive in an era of increasing uncertainty and market globalization, organizations must constantly be creative and innovative with their products and services.

Software has been widely considered as central to all sophisticated innovations. In the age of the Internet the challenge is to identify and evaluate new ideas, processes and applications. In many of the fastest growing industries, including computer, entertainment, communications, advertising, logistics and finance, software has been the end product itself, or the highest value component in the end product. In other cases, software has been used to support value creation and innovation processes. The growing importance of software-based innovations suggests the need to improve the creative skills of IT professionals. This need, in turn, requires an appropriate response from the IT education and training sector. Moreover, IT education and training should better nurture students' creativity, so that they can be successful in their future roles as innovative professionals and business people. It is particularly important that IT students be given an opportunity to develop and apply creative and innovative skills to software processes and products.

Given the crucial importance of creativity and innovativeness for success in a knowledge economy, the main purpose of this article is to explore concepts about creativity and how they relate to software development by providing empirical research examples in IT education.

CONCEPT OF CREATIVITY

The literature offers diverse conceptual definitions of creativity. Glass (2001) argues that creativity is hard to define, hard to judge and hard to quantize. Kappel and Rubenstein (1999) reason that this is due to fact that creativity is used to describe a variety of things; that is, supporting the creativity process, the creative person or the creative product present different requirements for the definition of the creativity. Tomas (1999), for example, defines creativity in terms of an original idea. Shalley and Perry-Smith (2001) point out that it is not enough to only be original; also, appropriateness is vital in order to distinguish creative ideas from surreal ideas that may be unique, but have unlawful or highly unrealistic implications.

Central to creativity is the ability to generate ideas. Some psychologists and philosophers have argued that idea formation can be explained by way of association (Mednick & Mednick, 1964). This theory suggests that association occurs when two stimuli take place together (contiguity), when two stimuli are similar to each other (similarity), or when two stimuli are different from each other (contrast). Associations may be stimulated by environmental factors, by previous associations, or may be mediated by ideas related to other associates. Therefore, it is possible to have many combinations and permutations. Associations can vary in strength, depending on how often associated ideas occur together or separately.

Lateral thinking is an aid to creativity when one needs to have diverse ideas. It is a function of knowledge and imagination that may bring out discovery, innovation, imagination, and exploration. Lateral thinking consists of seeking as many alternative options as possible to the extent of one's adventurousness. In other words, it is a mental activity involving making connections between knowledge and ideas that were previously unrelated. The basis of lateral thinking is that since many problems require a different perspective to be solved successfully, individuals should suspend their judgment about what is relevant to a course of action.

CREATIVITY TECHNIQUES

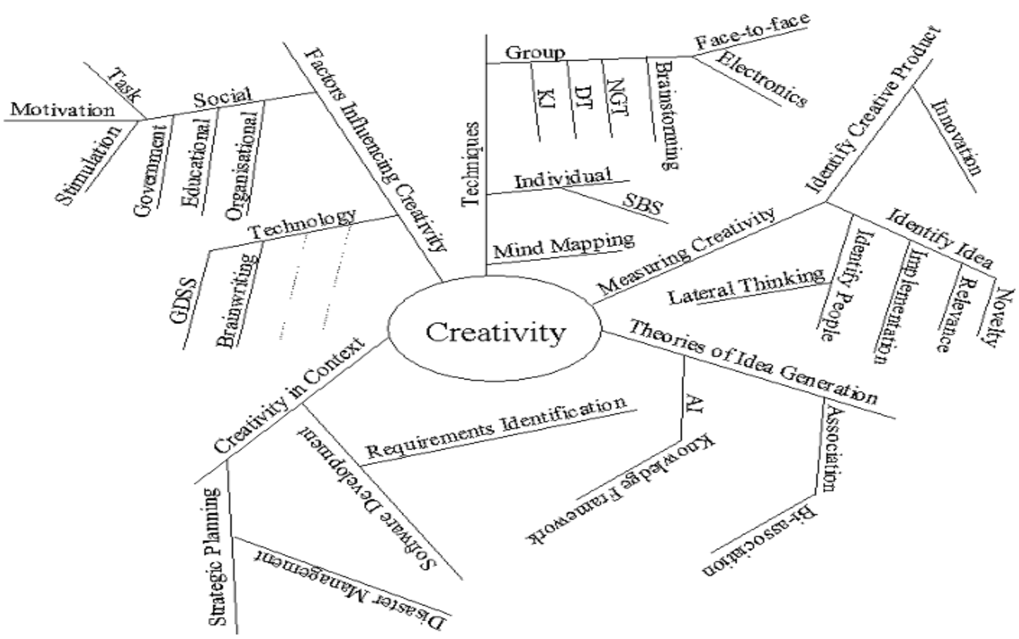
Consistent with the view that creative thinking can be learnt by appropriate stimulation and instruction, a variety of formal techniques have been developed to assist the production of novel ideas including brainstorming, mind mapping or solo brainstorming. Brainstorming and similar idea generation techniques aim to increase the production of novel ideas. The objective is to promote creativity by appropriately managing interaction within group as well as enhancing the creative environment. The procedures involved in the following examples are not difficult and may involve "lateral thinking," where ideas are stimulated by members of the group.

Brainstorming is an idea generation technique that was conceptualized by Walt Disney in the late 1920s and then expanded by Alex Osborn (1957). The objective of brainstorming is to encourage associations. The basic assumption is that it is possible for an individual(s) to generate many ideas, provided that he or she is exposed to stimuli and has experience, knowledge, and the personal flexibility to develop various permutations and combinations, and the capacity to make correct selections. This method initially emphasizes the quantity of ideas generated, leaving the assessment of quality to a later stage. Brainstorming sessions can be conducted electronically or verbally. In electronic brainstorming systems (EBS), group members share their knowledge and ideas by sending their ideas to each other, and by viewing the ideas of other members. Ideas generated from a brainstorming session can be recorded and stored in electronic files, making them easily accessible for printing or later reference (Nunamaker et al., 1991).

Another free association technique is mind mapping. This method begins with writing down a main idea in the centre of the page, and then working outward in all directions, producing a growing and organized structure composed of key words and key images, as illustrated in Figure 1. Mind mapping therefore relies on association (and clustering) of concepts/issues. The association process underling construction of the mind map actually facilitates making connections between concepts, and hence tends to generate new ideas and associations that have not been thought of before.

An example of an individual creativity technique is solo brainstorming (SBS), originally proposed by Aurum (1997). This technique is especially suited to environments where sentential analysis is appropriate, or information sources are document-based (e.g., reports, abstracts). SBS requires the individuals to adhere to a formal protocol, where a series of documents are examined and then edited. The ultimate aim in an SBS session is to determine a sufficient set of issues. As applications of the SBS protocol have been computer-based, all issues are automatically available in electronic form for further analysis. The SBS protocol touches upon an important research issue in the area of knowledge management: whether an increase in an individual's level of domain knowledge will necessarily increase their capacity to be creative within that domain. Central to the SBS protocol is the encourage-

Figure 1. Mind mapping (Aurum & Gardiner, 2003)



ment of participants to use their cognitive abilities by asking them to make "lateral comments".

CREATIVITY IN SOFTWARE DEVELOPMENT

Software engineering is another domain in which creativity plays an important role. The value of creativity is also well recognised in the field of system requirements determination. Robertson (2001) addresses requirements determination as "requirements discovery," which suggests that many users may not even be aware of their true requirements (e.g., unconscious requirements) without application of techniques for reflection and creativity.

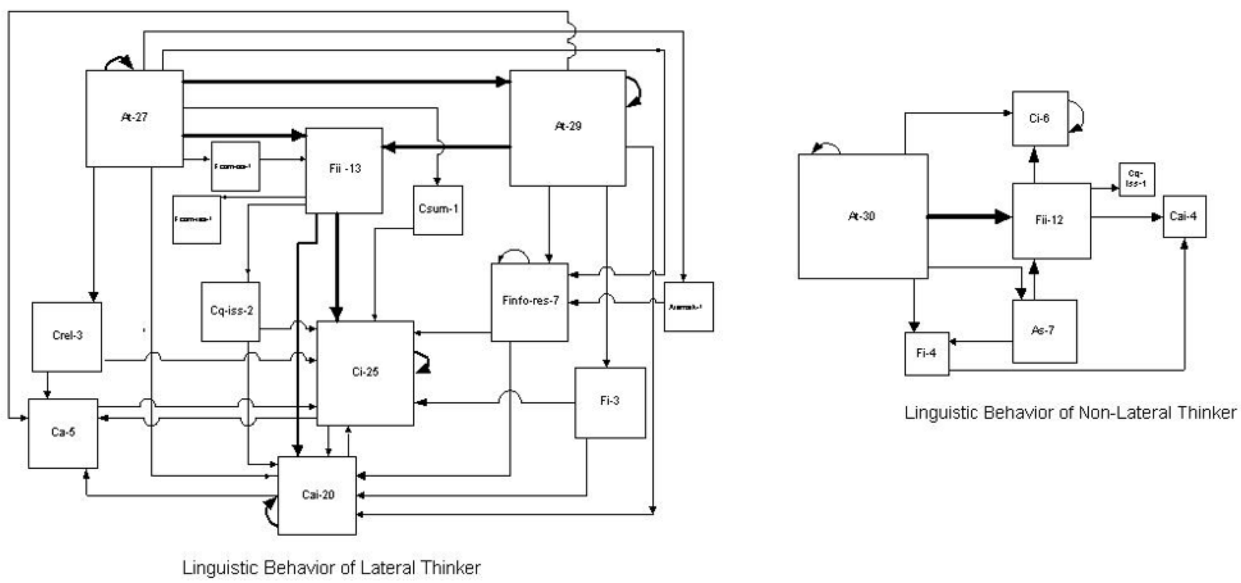
In an experiment focusing upon requirements elicitation, Aurum and Martin (1999) applied the SBS protocol to determine whether application of the protocol would deliver a richer set of requirement statements and insights. An experiment was conducted in which participants were told to adopt the role of a systems analyst retained by a fictitious organization, The Cultural Heritage Authority (CHA), to write requirements specification for their main information systems. The types of documents used as input (external information) to this study included fictitious interviews with users and abstracts from published articles addressing either heritage or marketing issues. Participants' task was to generate ideas and identify issues to be included in CHA's software requirements specification (SRS).

The objective of the research was to investigate whether the application of the SBS protocol had indeed led to a richer level of requirement specifications. The following two research questions are investigated (a) whether application of the protocol would result in identification of more relevant, workable, and original requirements issues, (b) to measure the lateral thinking by examining the linguistic behavior of participants.

The main findings of this study indicated that an SBS-based learning tool had a positive effect on participants' creative performances in development of SRS. This outcome was evident in the originality aspect of task performance, but not relevance and workability. Users were found to generate significantly more original ideas as the result of their interaction with the tool, while maintaining similar levels of relevance and workability (Aurum et al., 2003).

Aurum (1997) suggests that the level of "laterality" for any thought for a given problem can only be assessed with respect to the thoughts generated by others for the same problem. Aurum also found that documents generated from SBS session exhibited some unique characteristics. From linguistic analysis, it was possible to identify those users who were able to think laterally in the SBS session. Furthermore, lateral thinkers displayed a more

Figure 2. Linguistic behavior of SBS participants



complex linguistic pattern than non-lateral thinkers, as illustrated in Figure 2. Participants who generated many ideas and identified many issues were also found to be the lateral thinkers. The findings showed that lateral thinkers wrote many "issue loaded" or "idea loaded" sentences, whereas non-lateral thinkers produced fewer ideas. However, the distinction between these two groups was not clear-cut, but rather a continuum.

The results of this study indicate that the SBS is a promising method for stimulating creative thinking and idea generation in a software development task. Essentially, the brainstorming session helped students uncover ideas without being constrained, stimulate their own thinking by external influences, and capture their thoughts.

These findings also have some important implications for software development and IT education. They demonstrate that creativity can be improved, leading to higher quality software designs. The findings also suggest that the type of tool tested here may be a useful teaching tool in a variety of IT courses involving creative thinking and problem solving. Furthermore, the tool is likely to be most valuable in situations where the problem is unstructured, goals indistinct, and where the outcome of an action cannot always be clearly identified. The tool is a relatively generic one, since it uses a technique that can be applied to a variety of scenarios and can help people process relevant documents whilst identifying issues. These documents act like a 'trigger to stimulate domain-specific ideas from users.

CONCLUSION

Many organizations have come to realize that the creativity of their management and employees is an important source for competitive advantage. However, arguably more can be done within these organizations to promote a creative culture – for example, more organizations should seek to reward management and employees for creative (or divergent) displays, and make creativity supporting technologies more readily available to them.

A number of techniques have been developed to facilitate creativity, with many techniques based upon some form of brainstorming. One theme common to some of the more recent studies on creativity is the importance of a rich source of stimuli to support the creative process, whether the stimuli are: documents, as in Aurum and Martin (1999); group memory, as in Satzinger et al. (1999); or models, as in Shalley and Perry-Smith (2001). Other forms of stimuli include: text, audio, graphics, simulations, video, and so forth (Kletke et al., 2001). Indeed, the effectiveness of the brainstorming technique relies on participants being stimulated by the ideas contributed by others. The potential to cascade ideas is referred to as synergy (Dennis & Valacich, 1993) – that is, the ability of an idea from one participant to trigger in another participant a new idea that would otherwise not have been produced. Another technique is formalizing the creative process through some protocol that can be an effective strategy in terms of supporting the level of intrinsic motivation and mental effort required by participants undertaking a creativity task (Aurum 1999; Paulus & Yang, 2000). Application of a formal protocol is usually at the heart of a creativity technique, and ensures a more systematic and thorough approach to information analysis, which is essential for many creativity tasks.

REFERENCES

Aurum, A. (1997). *Solo brainstorming: Behavioral analysis of decision-makers.* PhD thesis. University of New South Wales, Australia.

Aurum, A., & Gardiner, A. (2003). Creative idea generation. In H. Hasan & M. Handzic (Eds.), *Studies on knowledge management* (pp. 57-91). University of Wollongong Press.

Aurum, A., Handzic, M., & Gardiner, A. (2003). Preparing IT professionals for creative development. In T. McGill (Ed.), *Supporting creativity in requirements engineering: An application in IT education.* Hershey, PA: Idea Group Publishing.

Aurum, A., & Martin, E. (1999). Managing both individual and collective participation in software requirements elicitation process. *14th International Symposium on Computer & Information Sciences,* Kusadasi, Turkey (pp. 124-131).

Covey, S.R. (1989). *The 7 habits of highly effective people.* New York: Rockfeller Center.

Dennis, A.R., & Valacich, J.S. (1993). Computer brainstorms: More heads are better than one. *Journal of Applied Psychology*, *78*(4), 531-537.

Glass, R.L. (2001, September/October). A story about the creativity involved in software work. *IEEE Software*, 96-97.

Kappel, T.A., & Rubenstein, A.H. (1999). Creativity in design: The contribution of information technology. *IEEE Transaction on Engineering Management, 46*(2), 132-143.

Kletke, M.G., Mackay, J.M., Barr, S.H., & Jones, B. (2001). Creativity in the organization: The role of individual creative problem solving and computer support. *International Journal of Human-Computer Studies, 55,* 217-237.

Mednick, S.A., & Mednick, M.T. (1964). An associative interpretation of the creative process. In C.W. Taylor (Ed.), *Widening horizons in creativity*. New York: John Wiley & Sons.

Nunamaker, J.F., Dennis, A.R., Valacich, J.S., Vogel, D.R., & George, J.F. (1991). Electronic meeting systems to support group work. *Communications of the ACM, 34*(7), 40-61.

Osborne, A. (1957). *Applied imagination: Principles and procedures of creative thinking*. New York: Charles Scribner's Sons.

Paulus, P.B., & Yang, H. (2000). Idea generation in groups: A basis for creativity in organizations. *Organizational Behavior and Human Decision Processes, 82*(1), 76-87.

Robertson, S. (2001). Requirements trawling: Techniques for discovering requirements. *International Journal of Human-Computer Studies, 55*, 405-421.

Satzinger, J.W., Garfield, J.M., & Nagasundaram, M. (1999). The creative process: The effects of group memory on individual idea generation. *Journal of Management Information Systems, 14*(4), 143-160.

Shalley, C.E., & Perry-Smith, J.E. (2001). Effects of social-psychological factors on creative performance: The role of informational and controlling expected evaluation and modelling experience. *Organizational Behavior and Human Decision Processes, 84*(1), 1-22.

Tomas, S. (1999). Creative problem-solving: An approach to generating ideas. *Hospital Material Management Quarterly, 20*(4), 33-45.

KEY TERMS

Creativity: There are many views about the definition of creativity. In the context of discovery, creativity is the ability to generate or recognize ideas, alternatives that might be useful solving problems. There are several aspects of creativity, including creative product or value, creative person/people, creative environment, creative symbols and creative process.

Electronic Brainstorming Systems (EBS): A computer-based system that facilitates brainstorming between group members.

Information System (IS): A system that uses IT to capture, transmit, store, retrieve, manipulate or display data for business processes in an organization.

Information Technology (IT): Computer hardware and software, as well as the peripheral devices closely associated with computer-based systems that facilitate data processing tasks, such as capturing, transmitting, storing, retrieving, manipulating or displaying data. IT includes matters concerned with design, development, and implementation of information systems and applications.

Internet: A worldwide network of computer networks that use the TCP/IP network protocols to facilitate data transmission. It provides access to a vast amount of information resources including multimedia (movies, sound, and images), software, text documents, news articles, electronic journal, travel information and so forth. It also provides an environment for buying and selling products and services over a network.

Knowledge Economy: Economic growth is driven by the accumulation of knowledge, which is the basic form of capital. A knowledge driven economy is one in which the generation and exploitation of knowledge plays the predominant part in the creation of wealth.

Knowledge Management (KM): The collection of processes that manage the creation, dissemination, and utilization of knowledge for learning, problem solving, and decision-making. KM often encompasses identifying intellectual assets within organizations. The management of knowledge is regarded as a main source of competitive advantage for organizations. KM brings together three organizational resources: people, process and technologies, and enables the organization to use and share information more effectively

Software Requirements Specification: A document that contains all requirements, for example functional and non-functional requirements and project issues, of the system as agreed upon by customers and software developers.

Instant Messaging Moves from the Home to the Office

Ha Sung Hwang
Temple University, USA

Concetta M. Stewart
Temple University, USA

INTRODUCTION

Instant messaging (IM) quickly established itself as one of the most popular modes of communication, with millions of people logging in at home, at the workplace, and at school. IM is an Internet protocol (IP)-based application that provides convenient communication between people using a variety of different device types. IM enables two people to exchange messages and hold simultaneous conversations without incurring long distance fees, as long as they use the same IM application. While corporate users employ proprietary systems, end users have several commercial services available to them, such as AOL Instant Messenger and Yahoo! Instant Messenger. With IM, users can exchange short text messages simultaneously as well as learn the online status of other users. This is IM's key feature.

While IM is the newest and most popular "near-synchronous" text-based chat technology, other chat technologies, such as multiuser domains (MUDs) and Internet Relay Chat (IRC) have been used for nearly two decades (Grinter & Palen, 2002). However, IM technology distinguishes itself from these previous chat technologies in several ways. While MUDs and IRC technology provide a chat room that allows a group of people to type in messages that are seen by everyone in the room, IM technology allows just two people to exchange messages. IM users adopt a screen name to represent themselves in the messaging service. So, before one can contact someone through IM, he or she must know his or her screen name. Most IM programs also allow users to search for a screen name by e-mail address, name, or area of interest.

HISTORY OF INSTANT MESSAGING

The first IM system, ICQ (I Seek You), was developed in 1996 by Mirabilis. Realizing that millions of people were connected to one huge worldwide network—but were not interconnected with one another—they created a technology that would enable users to locate each other online. They released ICQ in November of 1996, and within 6 months, there were 850,000 ICQ users. AOL acquired ICQ in 1998, but ICQ and AIM are still separate systems. According to a recent report by *Business Wire* (2004), with over 175 million registered users worldwide, ICQ is one of the leading communications communities in the world.

Seeing ICQ's mass appeal and usage, AOL released a free IM software application, named AOL Instant Messenger (AIM) in May 1997. Since 1998, the IM market has exploded: Yahoo launched its own version of IM, Yahoo Messenger, and in 1999, Microsoft's MSN introduced MSN Messenger.

FEATURES OF IM

Upon registering for the IM service, a user adopts a screen name that serves as his or her address for sending and receiving instant messages. Users can use e-mail and other applications when using IM. IM users can also create personal profiles that include the types of people with whom they want to chat. While surfing the Web, IM alerts the user through a window that pops up in one corner of the computer screen when a friend logs in. When not in use, that window can remain hidden until the user wants to send a message to notify that an instant message from a friend or coworker is incoming.

Most IM systems provide awareness technology that allows users to monitor the online state of others, using what is commonly known as a "buddy list." This list displays information about contacts and can be sorted into user-defined categories, such as "friend," "coworker," "classmate," or "family."

The main function of the buddy list is to monitor when friends or coworkers are online. To start a session with a buddy, a user double clicks on the "buddy's" screen name. The service gives or denies authorization to send a pop-up window to the receiver's computer. Such "presence awareness" technology is still growing and developing with more options among IM programs. For instance, MSN Messenger provides several status options to let other users know someone's online presence. Options include *Online*, *Offline*, *Busy*, *Be Right Back*, Away, *Out to Lunch*, and *On The Phone*.

The "blocking" option is another common feature of IM systems. With this feature, a user can control who appears on his or her list. For instance, while chatting with someone, you receive unwanted messages. You can click the *Block* button to refuse all messages from that person. Once that person is blocked, he or she will not be able to send messages to you until you decide to remove that person from your block list. You can add or delete people from your block list. Other common features include variation in fonts and text color, the ability to send files and photos, and Web links to favorite Web sites.

IM ADOPTION AND USE

IM has been widely adopted among the general population, and the number of users worldwide is growing exponentially. While the figures vary among research firms, most studies suggest that the number of worldwide IM users is large and will continue to grow. For instance, Conley (2002) reported that the number of IM users worldwide has already reached over 445 million, with more than 225,000 new users per day. International Data Corporation (IDC) estimated that users sent about 900 million instant messages on a typical day in 2000 and will send about 7 billion a day by 2004. IDC also estimates that the number of global IM users will reach over 500 million by 2005 (as cited in Disababatino, 2001).

The United States has emerged as the leading user of IM, according to Jupiter Media Metrix. Their research revealed that the number of IM users in U.S. homes increased 28%—from 42 million in September 2000 to 53.8 million in September 2001. Jupiter Media Metrix also reported that U.S. IM users at work increased by 34%—from 10 million in September 2000 to 13.4 million in September 2001 (as cited in Woods, 2002).

According to the Pew project (2001), the heaviest users of IM are teenagers and college students for whom IM is part of their daily routine The study showed that almost 13 million teenagers (74% of online teens) use IM, and 69% of teen IM users use IM several times a week. Nielsen/NetRatings (2002) reported that in July, 2002, IM drew 11.5 million kids and teens, with this number representing nearly 24% of the total at-home IM population. According to the Pew project (2002), college Internet users are twice as likely to use IM on any given day compared to the average Internet user. On a typical day, 26% of college students use IM as compared to 12% of other Internet users.

AOL is the predominant IM system in the United States, according to Nielsen/NetRatings (2002). In May of 2002, AOL Instant Messenger became the number one IM application in the United States, attracting more than 22 million users or almost 21% of the total Internet population. Combined, AOL Instant Messenger, MSN Messenger, Yahoo Messenger, and ICQ are the most popular IM services in the United States. Over 41 million American consumers (nearly 40% of the active U.S. online users) used one of these four IM services in the month of May, 2002.

IM IN THE WORKPLACE

After widespread adoption of IM for personal use, IM began to move beyond its consumer-based origins and establish itself as a tool for business communication. In 2002, Jupiter Media Metrix reported that the total number of minutes spent using IM at work increased 110%, from 2.3 billion minutes in September, 2000, to 4.9 billion in September, 2001. By September, 2003, one or more workers in 90% of large U.S. organizations used IM. IM has revolutionized business communication, because it is well suited for communication among dispersed coworkers who face challenges in maintaining ongoing projects. Unlike e-mail, though, IM allows spontaneous interaction among its users, and thereby, facilitates more opportunities for conversation. In this way, IM increases the efficiency of communication among geographically dispersed coworkers.

Studies on IM in the workplace demonstrate IM's value as a tool for supporting brief and informal communication among business partners. However, some studies have found that IM is also used for complex work-related discussions. Nardi et al. (2000) found that IM's primary function is in terms of interaction: quick questioning and clarifications, coordinating work-related activities, scheduling impromptu social meetings, and keeping in touch with friends and family. In particular, they emphasized the utility of IM for facilitating "outeraction"—that is, communication that facilitates further interaction, such as using IM for negotiating availability and scheduling. Muller et al. (2003) found that IM was a means for avoiding certain kinds of communication in the workplace, i.e., to obtain a speedy response to a question, to avoid long telephone conversations, and to seek quick clarification of a question.

Vos et al. (2004) found that most workers used their business IM account mainly on workdays and during work time, and that this IM use was more connected to work-related activities than to personal purposes. Issacs et al. (2002) pointed out the three main characteristics of workplace IM conversations: (1) IM conversations are brief, (2) media switching is prevalent, and (3) multitasking is common. Media switching reveals IM's limitations. When the conversation becomes too complex for the medium of IM, participants agree to switch to phone or face-to-face conversation. Multitasking, however, is considered to be a major benefit of IM. While carrying out other activities, such as e-mailing and talking on the phone, workers can quickly respond to questions from coworkers through IM. They note that frequent users are more likely than infrequent users to utilize IM for a greater variety of tasks.

According to Hansen and Damm (2002), IM is particularly useful for such tasks as problem-solving and creating new software designs, suggesting that IM is an effective "distributed collaboration tool." Other studies suggest how the designs of IM systems might need to evolve to meet the needs of workplace users. Huang et al. (2004), for instance, suggested five factors that affect the adoption of IM use in the workplace: *task specificity and integration*, *tool flexibility and generality*, *visibility and exposure to others' interactions*, *low barriers to use*, and a *dedicated core group of users*.

EMERGING ISSUES FOR IM

A common finding among all of these studies, however, is that the perception of IM as a communication tool among users is mixed. For instance, Herbsleb et al. (2002) indicated that the perception of IM utility is a difficult issue to determine, depending on how users view the importance of informal conversation. According to their study, new users may not see any advantages in IM. Several of their respondents reported that "it's just as easy to pick up a phone and call someone," and others questioned the value of the presence awareness function, "it is not that valuable to know that someone is sitting at their desk" (p. 176).

However, other studies report users who found this presence awareness function useful. Segerstad and Ljungstrand (2002) investigated how awareness of presence affects the content of IM among college students who use the IM service called "Webwho." Key findings included that task-related messages were more often passed among distributed college students in different places, while greeting messages were more often passed among students in the same place, and messages with sexual content were more often sent among students in the same computer lab or at home than among students in different computer labs. The study also suggests that IM is used to support collaborative work and coordinate social activities, as well as to simply keep in touch or have fun.

Teenagers' use of IM has recently emerged as a focus of media research, and these findings suggest that IM is used mainly for the fundamental need to interact with others. Noteworthy here is Grinter and Palen's (2002) finding that exchanging IM messages among teenagers is restricted to those who first met face-to face in a physical setting, such as at school or at summer camp. The study also suggests that for teenagers, peer pressure is an important facilitator in IM adoption.

Though IM has overwhelmingly been the province of teenagers, the Pew project (2003) found that nearly half of all American adults who are active online have sent an instant message at one time or another. The Pew project (2003) showed, however, that when it comes to adults, IM proved to be an activity where experience matters. Compared to those with only a year or less of online experience, those with four or more years of online experience were almost three times as likely to have used IM on an average day.

TRENDS

IM use is expanding dramatically, along with other online activity. The Pew project (2003) reported that between March, 2000, and August, 2003, the American online population expanded from 86 million to 126 million, or by 47%. However, the desire for a sense of control is increasingly influencing the choice of online activity. IM is growing, while other forms of real-time online communications are losing popularity. Although initial adoption of the Internet is in large part due to the popularity of chat rooms, the growth of IM may now be contributing to a decrease in the popularity of chat rooms. With IM, users can exert more control over their social interaction than they can with chat rooms, which may be a factor for populations who feel more vulnerable in open chat room situations, such as with young people and women.

According to the Pew project report (2003), the number of Internet users who have participated in chat rooms grew only 21% (from 24–29 million) between March, 2000, and July, 2002. IM, on the other hand, grew about 33% (from 39–52 million) in the same period. Moreover, American Internet users were almost twice as likely to use IM as they were to participate in a chat room. Compared to chatting, they were also three times as likely to use IM on a typical day.

Most recently, IM has been incorporated into emerging technologies, such as mobile phones and Internet telephony. With this development, IM is poised to expand its reach and convenience even further. For instance, the mobile phone industry has eagerly adopted IM technology by creating a service called "short messaging service." Through this service, millions of short messages are being sent daily to and from mobile phones throughout the world. MSN is preparing to launch Mobile Messenger in Asia (as cited in Peng, 2002). And, AIM now provides the service called "AIM Talk" for online conversations.

AOL seems to be extending its dominance in IM to the wireless world. The owner of ICQ and AOL Instant Messenger (AIM), AOL announced in February, 2003, that its ICQ service would provide messaging services for wireless companies in such countries as Germany, Israel, Taiwan, and Canada (Hu, 2003). ICQ specifically would offer a number of wireless messaging applications on a variety of software platforms, including short message service (SMS), two-way SMS, Wireless Application Protocol (WAP), and Java for cell phones. AOL and Yahoo have also recently established agreements with wireless companies to allow their IM users to send text messages to cell phones (Hu, 2003). Interestingly, however, though mobile IM via SMS has become popular in Europe, it has yet to do the same in the United States.

CONCLUSION

Overall, recent studies on IM use suggest that while young people are still the heaviest users of IM, its popularity has spread across age groups and social settings. Individuals' perceptions of IM utility differ greatly, though. While some IM adopters perceive IM as useful and find it rewarding, some new users do not see any advantages to IM and use other electronic communication media (e.g., telephone and e-mail). For teenagers, though, IM serves as a fundamental tool for socializing with their peers. IM's use, however, goes beyond the social realm and is an increasingly useful tool for workplace communication. In addition to keeping up with friends and family, IM helps with a variety of work-related communication tasks, such as short query, scheduling, and coordinating meetings among distributed business people. IM use has also spread worldwide, moving beyond the borders of the United States, with other countries taking the lead in adopting some of IM's newer features.

REFERENCES

Business Wire, Inc. (2004, April 20). *ICQ 4.0 Delivers dynamic personal communications platform to its growing global community*. Retrieved June 20, 2004, from *http://www.businesswire.com*

Conley, C. (2002, August 21). Instant messaging™craze captures America's teens. *Palo Alto Weekly Online Edition*. Retrieved September 17, 2003, from *http://www.paweekly.com/*

Disababatino, J. (2001, June). *Microsoft to embed instant messaging in Windows XP*. Retrieved August 26, 2003, from *http://www.computerworld.com/*

Grinter, E. R., & Palen, L. (2002). Instant messaging in teen life. In *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work* (pp. 21–30). New York: ACM Press.

Hansen, K., & Damm, C. (2002). *Instant collaboration: Using context-aware instant messaging for session management in distributed collaboration tools.* Paper presented at NordiCHI, October 19-23, 2002, Denmark.

Herbsleb, J. D., Atkins, D. J., Boyer, D. G., Handel, M., & Finholt, T. A. (2002). Introducing instant messaging and chat in the workplace. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing Our World, Changing Ourselves* (pp. 171-178). New York: ACM Press.

Hu, J. (2003, February 18). *ICQ goes wireless around the world*. Retrieved March 16, 2004, from *http://news.com.com/2100-1033-984945.html*

Huang, M. E., Russell, M. D., & Sue, E. A. (2004). IM here: Public instant messaging on large, shared displays for workgroup interactions. In *Proceedings of the 2004 Conference on Human Factors in Computing Systems* (pp. 279–286). New York: ACM Press.

Isaacs, E., Walendowski, A., Whittaker, S., Schiano, J. D., & Kamm, C. (2002). The character, functions, and styles of instant messaging in the workplace. In *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work* (pp. 11-20). New York: ACM Press.

Muller, J. M., Raven, E. M., Kogan, S., Millen, R. D., & Carey, K. (2003). Introducing chat into business organizations: Toward an instant messaging maturity model. In *Proceedings of the 2003 International ACM Conference on Supporting Group Work* (pp. 50-57). New York: ACM Press.

Nardi, A. B., Whittaker, S., & Bardner, E. (2000). Interaction and outeraction: Instant messaging in action. In *Proceedings of CSCW 2000* (pp. 79-88). New York: ACM Press.

Nielsen/NetRatings. (2002, June 17). Instant messaging used by more than 41 million home Internet surfers. Retrieved March 15, 2003, from *http://www.nielsennetratings.com/*

Nielsen//NetRatings. (2002, August 13). Nearly 20 percent of the active online population are kids and teens, creating opportunities for marketers. Retrieved March 15, 2003, from *http://www.nielsennetratings.com/*

Peng, S. L. (2002, August). Instant reply needed? Try instant messaging. Retrieved April 20, 2003, from *http://www.internetnews.com/*

Pew Internet & American Life. (2001). The rise of the instant messaging generation. Retrieved April 4, 2003, from *http://www.pewinternet.org*

Pew Internet & American Life. (2002). The Internet goes to college: How students are living in the future with today's technology. Retrieved January 13, 2003, from *http://www.pewinternet.org*

Pew Internet & American Life. (2003). America's online pursuits: The changing picture of who's online and what they do. Retrieved March 2, 2004, from *http://www.pewinternet.org*

Segerstad, H. Y., & Ljungstrand, P. (2002). Instant messaging with WebWho. *Human–Computer Studies*, *56*, 147-171.

Vos, H., Hofte, H., & Poot, H. (2004). IM [@Work] adoption of instant messaging in a knowledge worker organization. In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences* (p. 10019). Washington, D.C.: IEEE Computer Society.

Woods, B. (June 17, 2002). U.S. in-home IM usage hits 41M. Retrieved March 18, 2003, from *http://www.internetnews.com/stats/*

KEY TERMS

Blocking: Blocking allows an IM user to control the ability of his or her online status. When you block someone, that person will not be able to send messages to you until you remove that person from your block list.

Buddy List: Presence awareness technology that allows users to monitor the online status of others. A buddy list window shows whether buddies are online or offline. Users double click on a screen name of an active friend, and a message is automatically initiated.

Chat/Chat Room: A real-time online interactive discussion group.

Instant Messaging (IM): Near-synchronous Internet-based one-to-one communication technology. IM allows two users to exchange short text message in real time.

Internet Relay Chat (IRC): Internet-based chat technology that allows a group of people to exchange text messages in real time. To join group chatting, one needs an IRC program to connect to an IRC server. Examples include mIRC, Pirch, and Virc for Windows and Homer or Ircle for Macintosh.

MUD (Multiuser Domains): Network-based text application that was initially developed for role-playing games in which players interact with each other in a computer-created fantasy environment.

Pop-Up Window: A window that suddenly appears (pops up) when a user selects an option. In IM, a pop-up window signals when a users' buddies come "online" and go "offline," while the user is engaged in other computer applications.

Screen Name: This is an IM user name that serves as the address for sending and receiving messages. One can change his or her screen name at any time, and most IM programs also allow users to search for a screen name by e-mail address, name, or area of interest.

Surfing: To navigate the Internet searching for topics of interest, usually considered an undirected form of browsing.

Institutional Dimensions of Information Systems Evaluation[1]

Vassilis Serafeimidis
Atos KPMG Consulting, UK

INTRODUCTION

Information systems (IS) evaluation is a complex organizational and social decision making process. IS evaluation has attracted a lot of interest from the academic community and practitioners during the last few decades. This interest in some way has been generated by the tremendous increase of IT investments and the fact that IT/IS have become an organizational "necessity" in order to, for example, support routine data processing operations, initiatives for competitive advantage, and business transformation exercises.

IS evaluation is highly subjective and context dependent, as well as covering a wide area of situations and activities. A number of good definitions are listed next:

"Evaluation is a series of activities incorporating understanding, measurement, and assessment. It is either a conscious or tacit process which aims to establish the value of or the contribution made by a particular situation. It can also relate to the determination of the worth of an object." (Remenyi & Sherwood-Smith, 1997, p. 46).

"IS evaluation is a process for searching and for making explicit, quantitatively or qualitatively, all the impacts of an IT project and the programme and strategy of which it is a part." (Farbey, Land & Targett, 1999, p. 190)

BACKGROUND

Information Systems Evaluation – The "journey"

In the 1980s organizations started to realize that successful IT outcomes do not occur by default; they are highly uncertain and in order to achieve organizational success, IS have to be managed effectively and be considered broadly within their context. The difficulties in identifying and measuring potential benefits and costs, deriving from current organizational practices, forced many organizations to establish management control mechanisms. Among these mechanisms are the thorough "appraisal" of potential IT investments and the "evaluation" of their expected deliverables.

Evaluation happens in many ways (e.g., formally, informally), uses diverse criteria (e.g., financial, technical, social), follows rigorous methodologies or "gut feelings," and often becomes a political instrument that influences the balance of organizational power and stimulates organizational changes.

The role that evaluation plays as an organizational process varies. It is strongly related to other management and decision making processes. The management expectation from IS evaluation is about establishing by quantitative and/or qualitative means the worth of IT to the organization (Farbey et al., 1993) and IT's contribution to the organizational growth (Bakos & Kemerer, 1992; Hitt & Brynjolfsson, 1996). This can be achieved by effective IS evaluation which ranks alternatives (Clemons, 1991) and forms a central part of a complex and incremental planning, decision-making and control (diagnosis) process (Hawgood & Land, 1988). Evaluation is then a crucial feedback function (Baker, 1995), which helps the organization learn (Walsham, 1993) and thereby reduces the uncertainty of decisions. This feedback helps trace and understand the underlying factors leading to the success or otherwise of an IT investment. In many cases (Farbey et al., 1995) evaluation is a mechanism for gaining commitment and, in highly politically influenced environments, for legitimization, and in some other occasions is a mechanism for exploration and discovery.

In order to understand IS evaluation we have to focus on the why, what, which aspects, when, who and how dimensions. Three different streams of IS evaluation theoretical and practical developments have been dominating this research field: the technical/functional; the economic/financial; and the "interpretive".

The traditional (formal-rational or functionalist) conception sees evaluation as an external judgement of an information system that is treated as if it existed in isolation from its human and organizational components and effects. It also places excessive emphasis on the technological and accounting aspects at the expense of the organizational and social aspects. Therefore, it neglects the organizational context and process of IS development and its content, elements that are critical to the successful application of IT in support of the business. In general, more attention has been focused over the years on pre-scribing how to carry out evaluations (with technically-driven and cost-focused frameworks) rather than analyzing and understanding their role, interactions, effects and organizational impacts (Smithson & Hirschheim, 1998).

Both the *functional and economic* streams promote a logical rationalistic philosophy that searches for the efficiency and effectiveness of an information system in technical and business terms. Developments under these areas have attracted attention for a long period of time as they have addressed necessary questions regarding the performance and the financial aspects of the technical components and their investment returns. These two modes, although necessary and complementary, suffer a number of deficiencies. Their limitations include the:

- limited consideration of the organizational context,
- narrow purposes deriving from the formal/rational paradigm,
- lack of consideration of the new content elements and relevant measures,
- confined and fragmented time horizon,
- neglect of human aspects of evaluation, and
- narrow methodological focus.

Table 1. Summary of the characteristics of the IS evaluation modes (based on Serafeimidis, 2001)

	Technical stream	**Economic stream**	**Interpretive alternatives**
Why - Purpose/ Reasons	• Technical performance (e.g., quality) • Control of resources (e.g., costs)	• Quality and utilization of IS outputs (e.g., accuracy of information)	• Context-sensitive (i.e., contingent, emergent) • Understanding of social actions • Organizational learning
What - The subject of evaluation **Criteria and measurement**	• IT system • Automate - Cost reduction	• IS outputs • Informate - Productivity - Business value - User satisfaction • Uncertainty/Risks	• Broad portfolios of processes and systems • Intermediate relevant measures (e.g., more reliable systems)
When - Time frame	• *Ex ante* and *ex post* in relation to the systems development life cycle	• *Ex ante* and *ex post* in relation to the systems development life cycle	• Continuous benefits management
Who - People	• IT experts	• IT experts • Finance experts • Business managers	• "Evaluation party" including internal and external stakeholders
How – Methodologies /Tools	• Quality-related (e.g., TQM, software metrics) • Cost-related (e.g., COCOMO, function point analysis)	• Economic oriented (e.g., agency theory) • Finance oriented (e.g., CBA, SESAME, DCF, IRR) • Behavioral science driven (e.g., ROM, value analysis)	• Meta-methodologies • Contemporary methods (experimental and exploratory)
Strengths	• Addresses the technical system performance • Assesses the efficiency of IT/IS	• Focuses on financial and economic impact of the investment • Assesses the effectiveness of IT/IS	• Increases the understanding of the role and impact of the investment • Assesses the fit and the contributions of IS/IT within its organizational context
Weaknesses	• Does not address uncertainty of business requirements and the organizational fit of IT • Prescriptive	• Approaches require specialist expertise • Dominated by economic results	• Applicability and use are challenging

As a response to the limitations identified previously, alternative evaluation approaches based on the *interpretive* paradigm have been proposed by authors such as Irani and Love (2001), Symons (1993), and Walsham (1995).

"Interpretive methodologies of evaluation actively analyse the experience of organizational reality, focusing on stakeholder interests and perspectives. ... They increase the effectiveness of organizational activity by ensuring a match between elements of organizational change and the dynamics of organizational culture and politics." (Symons, 1993, p. 74)

The more interpretive approaches come closer to the social and "real" nature of IS and evaluation, but they have not managed to provide a completely "rich picture" of the phenomenon. Moreover, these attempts lack empirical evidence for their wide acceptance by practitioners. A summary of the major issues addressed in each stream is presented in Table 1.

TRENDS IN IS EVALUATION

The latest trends in IS evaluation (Irani & Love, 2001; Smithson & Hirschheim, 1998) recognize the need for multi-dimensional perspectives which consider the dimensions discussed previously but also focus on the social and organizational aspects of an IT/IS investment.

The organizational context (both internal and external) strongly influences the role that evaluation is expected to play and consequently the performance of the evaluation roles. The context includes the organizational characteristics (e.g., culture, norms), structural limitations, the organizational expectations from evaluation and external requirements. All these factors restrict and enable its impacts and contributions. Considering this relationship, four different orientations of IS evaluation can be identified (Serafeimidis & Smithson, 2003): a control; a social learning; a sense making and an exploratory one. The study of IT/IS evaluation through the institutional dimensions offers an alternative approach to the traditional ways. The main principle remains the fact that different evaluation orientations require different treatments in terms of their integration with existing organizational properties.

CONCLUSION

IS evaluation is an important element of the decision making process related to the investments in IT. IS evaluation is a complex, human-centered activity that cannot be divorced from its organizational context and the stakeholders involved. Multiple theories and methodologies have been developed focusing on a variety of areas and dimensions of evaluation, mainly technical, financial and organizational. It is evident that no single method or approach is adequate to provide a complete picture to decision makers. Empirical evidence (Coleman & Jamieson, 1994; Farbey, Land & Targett, 1993; Serafeimidis & Smithson, 2003; Willcocks, 1996) suggests that organizations use different methods and techniques for different types of IT/IS projects depending on the objectives of the evaluation, the time of it and the stakeholder interests. Therefore, a multidimensional consideration which will need to understand the organizational (institutional) dimensions of IS evaluation as well as the needs of the decisions makers at that point in the decision process should be adopted.

REFERENCES

Baker, B. (1995). The role of feedback in assessing information systems planning effectiveness. *Journal of Strategic Information Systems*, *4*(1), 61-80.

Bakos, J.Y., & Kemerer, C.F. (1992). Recent applications of economic theory in information technology research. *Decision Support Systems, 8*, 365-386.

Clemons, E.K. (1991). Evaluation of strategic investments in information technology. *Communications of the ACM, 34*(1), 23-36.

Coleman, T., & Jamieson, M. (1994). Beyond return on investment. In L. Willcocks (Ed.), *Information management. The evaluation of information systems investment* (pp. 189-205). London: Chapman & Hall.

Farbey, B., Land, F., & Targett, D. (1993). *How to assess your IT investment. A study of methods and practice.* Oxford: Butterworth Heinemann.

Farbey, B., Land, F., & Targett, D. (Eds.). (1995). *Hard money - soft outcomes. Evaluating and managing the IT investment.* Oxon: Alfred Waller Ltd.

Farbey, B., Land, F., & Targett, D. (1999). Moving IS evaluation forward: Learning themes and research issues. *Journal of Strategic Information Systems*, *8*(2), 189-207.

Hawgood, J., & Land, F. (1988). A multivalent approach to information systems assessment. In N. Bjorn-Andersen & G.B. Davis (Eds.), *Information systems assessment: Issues and challenges* (pp. 103-124). North Holland, Amsterdam.

Hitt, L.M., & Brynjolfsson, E. (1996, June). Productivity, business profitability, and consumer surplus: Three *different* measures of information technology value. *MIS Quarterly*, 121-142.

Irani, Z., & Love, P.E.D. (2001). The propagation of technology management taxonomies for evaluating investments in information systems. *Journal of Management Information Systems, 17*(3), 161-177.

Remenyi, D., & Sherwood-Smith, M. (1999). Maximising information systems value by continuous participative evaluation. *Journal of Logistics and Information Management, 12*(1-2), 14-31.

Serafeimidis, V. (2001). A review of research issues in evaluation of information systems. In W. Van Grembergen (Ed.), *Information technology evaluation methods and management* (pp. 58-77). Hershey, PA: Idea Group Publishing.

Serafeimidis, V., & Smithson, S. (2003). Information systems evaluation as an organizational institution - Experience from a case Study. *Information Systems Journal, 13*(3), 251-274.

Smithson, S., & Hirschheim, R. (1998). Analysing information systems evaluation: Another look at an old problem. *European Journal of Information Systems, 7*(3), 158-174.

Symons, V.J. (1993). Evaluation and the failure of control: Information systems development in the processing company. *Accounting, Management and Information Technology, 3*(1), 51-76.

Walsham, G. (1993). *Interpreting information systems in organizations.* Chichester, England: John Wiley & Sons.

Walsham, G. (1995). The emergence of interpretivism in IS research. *Information Systems Research, 6*(4), 376-394.

Willcocks, L. (Ed.). (1996). *Investing in information systems. Evaluation and management.* London: Chapman & Hall.

KEY TERMS

Economic Evaluation: This approach to evaluation views IT/IS as an investment or a business facilitation project. Therefore, the focus of evaluation shifts from the performance of the IS *per se* to the quality of its outputs (e.g., information) and their utilization (e.g., customer satisfaction, creation of business value). Here the evaluation of IT investments is based on an organizational analysis that emphasizes the achievement of predetermined outcomes as a measure of effectiveness (e.g., critical success factors, business objectives/strategy). In other words the "worth" of an IS is sought in the system's performance and financial profitability.

Institutional Dimensions: Considering the need for integrated evaluation of business ideas/solutions enabled by IT, it is emergent that there are organizational dimensions (or institutions). They include the context where evaluation is integrated (e.g., the system's development life cycle, the IS management practices and processes, the people's roles, the organizational culture) and furthermore, the understanding of stakeholders' behavior within this context.

Interpretive Evaluation: The formal-rational approaches (technical and economic) emphasize the technology features at the expense of the organizational and social aspects. In this context interpretive evaluation is a move towards the "interactionist" role of the technology with the organizational structures, culture and stakeholders. IS evaluation focuses on the analysis and understanding of the social and subjective nature of the phenomenon.

IS/IT Evaluation: There are many definitions in the literature and a few are quoted previously. They are considered as the most representative for the purposes of this article.

Organizational Context: Context is concerned with the multi-level identification of the various systems (e.g., social, political, economic, technical) and structures (e.g., organizational processes, policies, individuals' roles) as well as informal networks and personal relationships within which the organization is "located".

Technical/Functional Evaluation: This type of evaluation focuses on the technical (IT) components. The evaluation is concerned with the relationship between inputs and outputs. Its focal point is on efficiency in terms of the technical performance and the control of resources. Further interests lie around the increased and improved capacity, fewer errors, greater reliability, humanpower savings, and so forth and also in terms of software performance developments and quality (e.g., TQM). The driving force for the timing of evaluation is the traditional system's development life cycle.

ENDNOTE

[1] The statements and opinions in this article are in all respects those of the author and do not represent the views of Atos KPMG Consulting.

Integrated–Services Architecture for Internet Multimedia Applications

Zhonghua Yang
Nanyang Technological University, Singapore

Yanyan Yang
University of California, USA

Yaolin Gu
Southern Yangtze University, China

Robert Gay
Nanyang Technological University, Singapore

A HISTORICAL PERSPECTIVE

The Internet has gone from near invisibility to near ubiquity and penetrated into every aspect of society in the past few years (Department of Commerce, 1998). The application scenarios have also changed dramatically and now demand a more sophisticated service model from the network. In the early 1990s, there was a large-scale experiment in sending digitized voice and video across the Internet through a packet-switched infrastructure (Braden, Clark, & Shenker, 1994). These highly visible experiments have depended upon three enabling technologies: (a) Many modern workstations now come equipped with built-in multimedia hardware, (b) IP multicasting, which was not yet generally available in commercial routers, is available, and (c) highly sophisticated digital audio and video applications have been developed. It became clear from these experiments that an important technical element of the Internet is still missing: Real-time applications often do not work well across the Internet. The Internet, as originally conceived, offers only a very simple quality-of-service (QoS), point-to-point, best-effort data delivery. However, for a real-time application, there are two aspects of the problem with using this service model. If the sender and/or receiver are humans, they simply cannot tolerate arbitrary delays; on the other hand, if the rate at which video and audio arrive is too low, the signal becomes incomprehensible. To support real-time Internet applications, the service model must address those services that relate most directly to the time of delivery of data. Real-time applications like video and audio conferencing typically require stricter guarantees on throughput and delay. The essence of real-time service is the requirement for some service guarantees in terms of timing. In response to these demands of real-time multimedia applications, the Internet Engineering Task Force (IETF) has significantly augmented the Internet protocol stack based on the Internet integrated-services model, which is the focus of this article.

THE INTERNET INTEGRATED-SERVICES MODEL

An Internet *service model* consists of a set of service commitments; that is, in response to a service request, the network commits to deliver some service. The Internet is conventionally designed to offer a very simple service model, best effort, providing no guarantee on the correct and timely delivery of data packets. Each request to send is honored by the network *as best as it can*. This is the worst possible service: Packets are forwarded by routers solely on the basis that there is a known route, irrespective of traffic conditions along that route. This simplicity has probably been one of the main reasons for the success of IP technology. The best-effort service model, combined with an efficient transport-layer protocol (TCP [transmission-control protocol]), is perfectly suited for a large class of applications, which tolerate variable delivery rates and delays. This class of applications is called elastic applications.

However, demanding real-time applications require more sophisticated service models beyond the best effort. There has been a great deal of effort since 1990 by IETF to add a broad range of services to the Internet service model, resulting in the Internet integrated service model (Braden et al., 1994; Crowcroft, Handley, & Wakeman, 1999). The Internet integrated services model defines five classes of services that should satisfy the requirements of the vast majority of future applications.

1. *Best effort*: As described above, this is the traditional service model of the Internet.
2. *Fair*: This is an enhancement of the traditional model where there are no extra requests from the users, but the routers attempt to partition network resources in some fair manner. This is typically implemented by adopting a random-drop policy when encountering overload, possibly combined with some simple round-robin serving of different sources.
3. *Controlled load*: This is an attempt to provide a degree of service guarantee so that a network appears to the user as if there is little other traffic, and it makes no other guarantees. The admission control is usually imposed so that the performance perceived is as if the network were overengineered for those that are admitted.
4. *Predictive service*: This service gives a delay bound that is as low as possible, and at the same time, is stable enough that the receiver can estimate it.
5. *Guaranteed service*: This is where the delay perceived by a particular source or to a group is bounded within some absolute limit. This service model implies that resource reservation and admission control are key building blocks of the service.

The level of QoS provided by these enhanced QoS classes is programmable on a per-flow basis, and end-to-end QoS commitment for the data flow is built using a unified multimedia protocol stack and through resource reservation.

THE INTERNET MULTIMEDIA PROTOCOL ARCHITECTURE

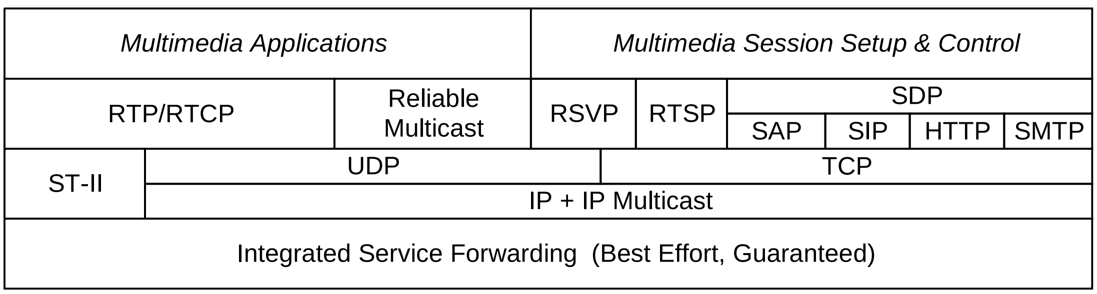
The integrated-services Internet offers a class of service models beyond the TCP/IP's (Internet protocol) best-effort service, and thus it imposes strict new requirements for a new generation of Internet protocols. The set of Internet real-time protocols, which constitute the Internet multimedia protocol architecture, represents a new style of protocols. The new style of protocols follows the proposed principles of *application-level framing* (ALF) and *integrated layer processing* (Clark & Tennenhouse, 1990). In this approach to protocol architecture, the different functions are next to each other, not on top of one another. The Internet multimedia protocol architecture is shown in Figure 1.

As shown in Figure 1, the overall multimedia data and control architecture currently incorporates a set of real-time protocols, which include the real-time transport protocol (RTP) for transporting real-time data and providing QoS feedback, the real-time streaming protocol (RTSP) for controlling delivery of streaming media, the session-announcement protocol (SAP) for advertising multimedia sessions via multicast, and the session-description protocol (SDP) for describing multimedia sessions. In addition, it includes the session-initiation protocol (SIP), which is used to invite the interested parties to join the session. But the functionality and operation of SIP does not depend on any of these protocols. Furthermore, the resource-reservation protocol (RSVP) is designed for reserving network resources. These protocols, together with reliable multicast (Handley, Floyd, Whetten, Kermode, Vicisano, & Luby, 2000), are the underlying support for Internet multimedia applications. While all the protocols above work on top of the IP protocol, the Internet stream protocol, version 2 (ST-II), is an IP-layer protocol that provides end-to-end guaranteed service across the Internet.

The Real Time Transport Protocols: RTP and RTCP

The real-time transport protocol, named as a transport protocol to emphasize that RTP is an end-to-end protocol, is designed to provide end-to-end delivery services for data with real-time characteristics, such as interactive audio and video (Schulzrinne, Casner, Frederick, & Jacobson, 2003). Those services include payload-type identification, sequence numbering, time-stamping, and

Figure 1. Internet protocol architecture for real time applications



nb: For the acronyms, refer to "Key Terms" delivery monitoring. Applications typically run RTP on top of UDP (user datagram protocol) to make use of its multiplexing and checksum services; both protocols contribute parts of the transport protocol functionality. However, RTP may be used with other suitable underlying network or transport protocols. RTP supports data transfer to multiple destinations using multicast distribution if provided by the underlying network.

RTP consists of two closely-linked parts:

- the real-time transport protocol to carry data that has real-time properties, and
- the RTP control protocol (RTCP) to monitor the quality of service and to convey information about the participants in an ongoing session. This functionality may be fully or partially subsumed by a separate session control protocol, which is beyond the scope of this document.

The RTP defines a fixed data-packet header for the set of functions required in common across all the application classes that RTP might support. However, in keeping with the ALF design principle, an extension mechanism is provided to allow the header to be tailored through modifications or additions defined in a profile specification.

RTCP control packets supplement each RTP flow and are periodically transmitted by each participant in an RTP session to all other participants. RTCP performs the following four functions.

1. Provides feedback information to application
2. Identifies RTP source
3. Controls RTCP transmission interval
4. Conveys minimal session-control information

Internet Stream Protocol: ST-II

ST-II has been developed to support efficient delivery of streams of packets to either single or multiple destinations in applications requiring guaranteed data rates and controlled delay characteristics (Schulzrinne, Rao, & Lanphier, 1998). ST-II is an Internet protocol at the same layer as IP (Figure 1). ST-II differs from IP in that every intervening ST-II entity maintains state information for each stream that passes through it. The stream state includes forwarding information, including multicast support for efficiency, and resource information, which allows network or link bandwidth and queues to be assigned to a specific stream. This preallocation of resources allows data packets to be forwarded with low delay, low overhead, and a low probability of loss due to congestion, and thus to support efficient delivery of streams of packets to either single or multiple destinations in applications requiring guaranteed data rates and controlled delay characteristics.

Protocols for Multimedia Session Setup and Control

There are two basic forms of multimedia session-setup mechanisms. These are session advertisement and session invitation. Session advertisements are provided using a session directory, and session invitation (inviting a user to join a session) is provided using a session-invitation protocol such as SIP or packet-based multimedia communication systems standard H.323 (ITU, 1998).

Before a session can be advertised, it must be described using the session-description protocol. SDP describes the content and format of a multimedia session, and the session-announcement protocol is used to distribute it to all potential session recipients.

The Session-Description Protocol

The session-description protocol is used for general real-time multimedia session-description purposes and is purely a format for session description (Handley & Jacobson, 1998). SDP is intended for using different transport protocols as appropriate.

SDP serves two primary purposes. It is a means to communicate the existence of a session, and is a means to convey sufficient information to enable joining and participating in the session. A session description contains the following information.

- Session name and purpose
- The media comprising the session, such as the type of media (video, audio), the transport protocol (RTP/UDP/IP, H.320), and the format of the media (H.261 video, MPEG video)
- Time(s) the session is active.
- Information for receiving those media (addresses, ports, formats, and so on)

The SDP description is announced using the session-announcement protocol.

Session-Announcement Protocol

SAP defines an announcement protocol to be used to assist the advertisement of multicast multimedia conferences and other multicast sessions, and to communicate the relevant session-setup information to prospective participants (Handley, Perkins, & Whelan, 2000). Sessions are described using the session-description protocol, and the session description is the payload of the SAP packet.

SAP supports session announcement and deletion. However, SAP defines no rendezvous mechanism, the SAP announcer is not aware of the presence or absence of any SAP listeners, and no additional reliability is provided over the standard best-effort UDP/IP semantics. A SAP announcer periodically sends an announcement packet to a well-known multicast address and port. A preannounced session can be modified by simply announcing the modified session description. A previously announced session may be deleted.

The announcement contains an authentication header for verifying that changes to a session description or deletion of a session are permitted. It can also be used to authenticate the identity of the session creator.

Session-Initiation Protocol

Not all sessions are advertised, and even those that are advertised may require a mechanism to explicitly invite a user to join a session. The session-initiation protocol is an application-layer control (signaling) protocol that can establish, modify, and terminate multimedia sessions or calls (Rosenberg et al., 2002). SIP can also invite participants to already existing sessions, such as multicast conferences. Media can be added to (and removed from) an existing session. SIP invitations used to create sessions carry session descriptions that allow participants to agree on a set of compatible media types. SIP runs on top of several different transport protocols.

SIP supports five aspects of establishing and terminating multimedia sessions.

- User location: determination of the end system to be used for communication
- User availability: determination of the willingness of the called party to engage in communications
- User capabilities: determination of the media and media parameters to be used
- Session setup: “ringing”; establishment of session parameters at both the called and calling party
- Session management: including transferring and terminating sessions, modifying session parameters, and invoking services

Note that SIP is not a vertically integrated communications system. SIP is, rather, a component that can be used with other Internet protocols to build a complete multimedia architecture (e.g., RTP, RTSP).

Controlling Multimedia Servers: RTSP

A standard way to remotely control multimedia streams delivered, for example, via RTP, is the real-time stream-control protocol. Control includes absolute positioning within the media stream, recording, and possibly device control. RSTP is primarily aimed at Web-based media-on-demand services, but it is also well suited to provide VCR-like controls for audio and video streams, and to provide playback and record functionality of RTP data streams. A client can specify that an RTSP server plays a recorded multimedia session into an existing multicast-based conference, or can specify that the server should join the conference and record it. RTSP acts as a “network remote control” for multimedia servers.

The protocol supports retrieval of media from media servers, the invitation of a media server to a conference, and the addition of media to an existing presentation.

Resource Reservation Protocol

RSVP is an IP resource-reservation setup protocol designed for an integrated-services Internet, and it provides receiver-initiated setup of resource reservations for multicast or unicast data flows that take a significant fraction of the network resources (Braden, Zhang, Berson, Herzog, & Jamin, 1997). Using RSVP, the resources necessary for a multimedia session are reserved, and if no sufficient resource is available, the admission is rejected. The reservations and route setups apply only to packets of a particular session, and RSVP identifies a particular session by the combination of destination address, transport-layer protocol type, and destination port number.

RSVP makes receivers responsible for requesting a specific QoS. A QoS request from a receiver host application is passed to the local RSVP process. The RSVP then carries the request to all the nodes (routers and hosts) along the reverse data path(s) to the data source(s), but only as far as the router where the receiver’s data path joins the multicast distribution tree.

Quality of service for a particular data flow is ensured by mechanisms collectively called *traffic control*, and they are a packet classifier, admission control, and “packet scheduler,” or some other link-layer-dependent mechanism to determine when particular packets are forwarded.

During reservation setup, an RSVP QoS request undergoes *admission control* and *policy control*. Admission control determines whether the node has sufficient available resources to supply the requested QoS. Policy control determines whether the user has administrative permission to make the reservation.

RSVP does not transport application data, but is rather an Internet control protocol. It uses underlying routing protocols to determine where it should carry reservation requests. As routing paths change, RSVP adapts its reservation to new paths if reservations are in place.

THE FUTURE TRENDS

Technically, there are several types of communication networks that are used to provide multimedia communication services, which include data networks, broadband television networks, integrated-services digital networks, and broadband multiservice networks. The Internet is the most widely deployed data network, and significant advancement has been made to provide communication services beyond its original design. This advancement is centered on QoS provisioning on the Internet so that QoS-sensible applications can be supported across the Internet.

Generally, work on QoS-enabled Internet has led to two distinct approaches: (a) the integrated services architecture (often called *Intserv*) and its accompanying signaling protocol, most importantly, RSVP, and (b) the differentiated services architecture (often called *Diffserv*).

As described in this article, the Internet integrated-services architecture is an extension of the Internet architecture and protocols to provide integrated services, that is, to support real-time as well as the current non-real-time service of IP. This extension is necessary to meet the growing need for real-time service for a variety of new applications, including teleconferencing, remote seminars, telescience, and distributed simulation. It allows sources and receivers to exchange signaling messages that establish packet classification and forwarding state on each node along the path between them. In the absence of state aggregation, the amount of state on each node scales in proportion to the number of concurrent reservations, which can be potentially large on high-speed links. Integrated services are considered not scalable and best suited in intranet and LAN (local area network) environments.

In the last few years, we have witnessed the development of other models of service differentiation in order to facilitate the deployment of real-time applications, including the differentiated-services architecture, relative priority marking, service marking, label switching, and static per-hop classification. The differentiated-services architecture has received the most attention in the Internet community.

In contrast to integrated services, which use the more stringent and complex quality-of-service approach, the differentiated-services architecture has emerged as an alternative for implementing scalable service differentiation on the Internet. This architecture achieves scalability by aggregating traffic classification states, which are conveyed by means of IP-layer packet marking. Packets are classified and marked to receive a particular per-hop forwarding behavior on nodes along their path. Sophisticated classification, marking, policing, and shaping operations need only be implemented at network boundaries or hosts. Network resources are allocated to traffic streams by service-provisioning policies that govern how traffic is marked and conditioned upon entry to a differentiated-services-capable network, and how that traffic is forwarded within that network. A wide variety of services can be implemented on top of these building blocks (Blake, Black, Carlson, Davies, Wang, & Weiss, 1998). We can foresee that in the future, support for multimedia applications on the Internet, both integrated services and differentiated services, will be developed and deployed, and will complement each other.

CONCLUSION

The Internet has evolved from a provider of simple TCP/IP best-effort service to an emerging integrated-service Internet. This development provides tremendous opportunities for building real-time multimedia applications over the Internet. The protocol stack necessary for classes of quality of service, including real-time applications, is presented as the Internet integrated-service architecture that supports the various service models. The constituent real-time protocols of this architecture are the foundations and the critical support elements for the building of Internet real-time multimedia applications.

REFERENCES

Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., & Weiss, W. (1998). *An architecture for differentiated services* [RFC 2475]. Internet Engineering Task Force.

Braden, R., Clark, D., & Shenker, S. (1994). *Integrated services in the Internet architecture: An overview* [RFC 1633]. Internet Engineering Task Force.

Braden, R., Zhang, L., Berson, S., Herzog, S., & Jamin, S. (1997). *Resource ReSerVation protocol (RSVP): Version 1, functional specification* [RFC 2205]. Internet Engineering Task Force.

Clark, D. D., & Tennenhouse, D. L. (1990). Architectural considerations for a new generation of protocols. *Computer Communications Review, 20*(4), 200-208.

Crowcroft, J., Handley, M., & Wakeman, I. (1999). *Internetworking multimedia*. Morgan Kaufmann Publishers.

Department of Commerce. (1998). *The emerging digital economy*.

Handley, M., Floyd, S., Whetten, B., Kermode, R., Vicisano, L., & Luby, M. (2000). *The reliable multicast design space* *for bulk data transfer* [RFC 2887]. Internet Engineering Task Force.

Handley, M., & Jacobson, V. (1998). *SDP: Session description protocol* [RFC 2327]. Internet Engineering Task Force.

Handley, M., Perkins, C., & Whelan, E. (2000). *Session announcement protocol* [RFC 2974]. Internet Engineering Task Force.

ITU. (1998). *Packet-based multimedia communication systems recommendation H.323*. Geneva, Switzerland: Telecommunication Standardization Sector of ITU.

Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., et al. (2002). *SIP: Session initiation protocol* [RFC 3261]. Internet Engineering Task Force.

Schulzrinne, H., Casner, S., Frederick, R., & Jacobson, J. (2003). *RTP: A transport protocol for real-time applications* [RFC 3550]. Internet Engineering Task Force.

Schulzrinne, H., Rao, A., & Lanphier, R. (1998). *Real time streaming protocol (RTSP)* [RFC 2326]. Internet Engineering Task Force.

Topolcic, C. (Ed.). (1990). *Experimental Internet stream protocol, version 2 (ST-II)* [RFC 1190]. Internet Engineering Task Force.

KEY TERMS

Elastic applications: A kind of network applications that will always wait for data to arrive rather than proceed without it.

HTTP: Short for *hypertext transfer protocol*, the underlying protocol used by the World Wide Web.

IP: Abbreviation of *Internet protocol*. IP specifies the format of packets, also called datagrams, and the addressing scheme. Most networks combine IP with a higher level protocol called transmission-control protocol (TCP), forming TCP/IP networks.

Multicast: Means transmitting a single message to a select group of recipients. A simple example of multicasting is sending an e-mail message to a mailing list.

Protocol: A set of rules that govern the operation of functional units to achieve communication.

Protocol architecture: An organization structure of the communication system, which comprises constituent protocols and the relationships among them.

Real-time applications: One class of applications needs the data in each packet by a certain time, and if the data has not arrived by then, the data is essentially worthless.

RSVP: The Internet standard protocol *resource-reservation protocol* (RSVP).

RTP/RTCP: The Internet standard *real-time transport protocol* (RTP) and *RTP control protocol* (RTCP).

RTSP: The Internet standard *real-time streaming protocol* (RTSP) for controlling delivery of streaming of multimedia data.

SAP: The Internet standard *session-announcement protocol* (SAP) for advertising multimedia sessions via multicast.

SDP: The Internet standard *session-description protocol* (SDP) is used for general real-time multimedia session-description purposes and is purely a format for session description.

Service model: Consists of a set of service commitments. In response to a service request, the network commits to deliver some service.

SIP: The Internet standard *session-initiation protocol* (SIP), a signaling protocol for Internet conferencing, telephony, presence, events notification, and instant messaging.

SMTP: Short for *simple mail-transfer protocol*, a protocol for sending e-mail messages between servers. Most e-mail systems that send mail over the Internet use SMTP to send messages from one server to another. In addition, SMTP is generally used to send messages from a mail client to a mail server.

ST-II: The Internet *stream protocol, version 2* (ST-II), an IP-layer protocol that provides end-to-end guaranteed service across the Internet.

TCP: Abbreviation of *transmission-control protocol*, one of the main protocols in TCP/IP networks. Whereas the IP protocol deals only with packets, TCP enables two hosts to establish a connection and exchange streams of data.

UDP: Abbreviation for *user datagram protocol*, a connectionless protocol that, like TCP, runs on top of IP networks. Unlike TCP/IP, UDP/IP provides very few error-recovery services so is less reliable than TCP, offering instead a direct way to send and receive datagrams over an IP network.

I

Integrating Requirements Engineering Techniques and Formal Methods

María Virginia Mauco
Universidad Nacional del Centro de la Pcia. de Buenos Aires, Argentina

Daniel Riesco
Universidad Nacional de San Luis, Argentina

INTRODUCTION

Formal methods help to develop more reliable and secure software systems, and they are increasingly being accepted by industry. The RAISE[1] Method (George et al., 1995), for example, is intended for use on real developments, not just toy examples. This method includes a large number of techniques and strategies for formal development and proofs, as well as a formal specification language, the RAISE Specification Language (RSL) (George et al., 1992), and a set of tools (George et al., 2001).

Formal specifications may be used throughout the software lifecycle, and they may be manipulated by automated tools for a wide variety of purposes such as model checking, deductive verification, formal reuse of components, and refinement from specification to implementation (van Lamsweerde, 2000a). However, they are not easily accessible to people who are not familiar or comfortable with formal notations. This is particularly inconvenient during the first stages of system development, when interaction with the stakeholders is very important. In common practice, the analysis of a problem often starts from interviews with the stakeholders, and this source of information is heavily based on natural language.

System requirements must be described well enough so that an agreement can be reached between the stakeholders and the system developers on what the system should and should not do. A major challenge with this is that the stakeholders must be able to read and understand the results of requirements capture. To meet this challenge we must use the language of the stakeholders to describe these results (Jacobson, Booch & Rumbaugh, 1999). Stakeholder-oriented requirements engineering techniques help to improve communication among stakeholders and software engineers as they ease the development of a first specification of a system which could be validated with stakeholders and which could be the basis for a formal development. Thus, we could take advantage of both techniques to improve the final product.

Among the techniques proposed to formalize requirements elicitation and modeling, Leite's Requirements Baseline can be mentioned (Leite et al., 1997). Two of its models are the Language Extended Lexicon (LEL) and the Scenario Model. LEL and scenarios provide a detailed description of an application domain, and as they are written in natural language, they are closer to stakeholder's world. However, an important point is how to fruitfully use all this information during the software development process.

To address the problems stated above, we have been working in the integration of stakeholder-oriented requirements engineering techniques with formal methods in order to take advantage of the benefits of both of them. In particular, our work is focused on the Requirements Baseline and the RAISE Method. We have proposed a three-step process to help in the definition of an initial formal specification in RSL of a domain starting from natural language models such as LEL and scenarios. These steps are the derivation of types, the derivation of functions, and the definition of modules. We have developed a preliminary set of heuristics that show how to derive types and functions, and how to structure them in modules by using LEL and scenarios information. We have also proposed to represent the hierarchy of RSL modules obtained using a layered architecture. This layered architecture is then the basis to start applying the steps of the RAISE Method.

BACKGROUND

In spite of the wide variety of formal specification languages and modeling languages, such as the Unified Modeling Language (UML) (Jacobson et al., 1999), natural language is still the method chosen for describing software system requirements (Jacobson et al., 1999; Sommerville & Sawyer, 1998; van Lamsweerde, 2000a). However, the syntax and semantics of natural language, even with its flexibility and expressiveness power, is not formal enough to be used directly for prototyping, implementation, or verification of a system. Thus, the requirements document written in natural language has to be reinterpreted by software engineers into a more formal design on the way to a complete implementation. Some recent works (Lee, 2001; Lee & Bryant, 2002; Moreno Capuchino, Juristo & Van de Riet, 2000; Nuseibeh & Easterbrook, 2000, van Lamsweerde, 2000b) present different strategies for mapping requirements to, for example, object-oriented models or formal specifications.

When using the RAISE Method, writing the initial RSL specification is the most critical task because this specification must capture the requirements in a formal, precise way (George, 2002). RSL specifications of many domains have been developed by starting from informal descriptions containing synopsis (introductory text that informs what the domain is about), narrative (systematic description of all the phenomena of the domain), and terminology (list of concise and informal definitions, alphabetically ordered). Others also include a list of events. They can be found in UNU/IIST´s Web site *(*www.iist.unu.edu*)*. The gap between these kind of descriptions and the corresponding RSL formal specification is large, and thus, for example, it is difficult and not always possible to check whether the formal specification models what the informal description does and vice versa.

As we had some experience in using the Requirements Baseline, and we knew it had been used as the basis to an object conceptual model (Leonardi, 2001), we consider the possibility of using it as the first description of a domain from which a formal specification in RSL could be later derived.

THREE-STEP PROCESS TO DERIVE A FORMAL SPECIFICATION

As an attempt to reduce the gap between stakeholders and the formal methods world, we propose a technique to derive an initial formal specification in RSL from requirements models, such as LEL and scenarios that are closer to stakeholders' language. The derivation of the specification is structured in three steps: Derivation of Types, Derivation of Functions, and Definition of Modules. They are not strictly sequential; they can overlap or be carried out in cycles. For example, function definitions can indicate which type structures are preferable.

Derivation of Types

This step produces a set of abstract as well as concrete types that model the relevant terms in the domain. We perform the derivation of the types in two steps. First we identify the types, and then we decide how to model them. This way of defining types follows one of the key notions of the RAISE Method (George et al., 1995): the step-wise development.

The main goal of the identification step is to determine an initial set of types that are necessary to model the different entities present in the analyzed domain. This initial set will be completed, or even modified, during the remaining steps of the specification derivation. For example, during the Definition of Modules Step, it may be necessary to define a type to reflect the domain state. Also, when defining functions, it may be useful to define some new types to be used as result types of functions. The LEL is the source of information during this step, as LEL subjects and some objects represent the main components or entities of the analyzed domain. In general, LEL subjects and objects will correspond to types in the RSL specification. In some cases, LEL verbs may also give rise to the definition of more types, as when they represent an activity that has its own data to save. However, in order to define just the relevant types, we have suggested some heuristics that can be found in Mauco, Riesco, and George (2001a).

Once a preliminary set of types is defined and in order to remove under-specification, we propose to return to the information contained in the LEL and the Scenario Model. In particular, the analysis of the notion, and sometimes the behavioral response of each symbol that motivated the definition of an abstract type, can help to decide if the type could be developed into a more concrete type. All the developments we suggest satisfy the implementation relation. In Mauco et al. (2001a), some heuristics to assist in this task can be found. During this step, it might be necessary to introduce some type definitions that do not correspond to any entry in the LEL. They appear, in general, when modeling components of some other type. Symbols without an entry in the LEL may represent an omission or a symbol considered outside the application domain language. When an omission is detected, it is necessary to return to the LEL to add the new definition, and update the Scenario Model to maintain the consistency between its vocabulary and the LEL itself.

Definition of Modules

This step helps to organize in modules all the types produced by the Derivation of Types Step in order to obtain a more legible and maintainable specification. These modules would be later completed with the definition of functions in the next step, and probably they will be completed with more type definitions.

A summary of the heuristics we propose to define for the modules can be found in Mauco et al., (2001b). In defining these heuristics, we closely followed the features RSL modules should have according to the RAISE Method (George et al., 1995; George, 2002). For example, each module should have only one type of interest, defining the appropriate functions to create, modify, and observe values of the type, and the collection of modules should be, as far as possible, hierarchically structured.

Then, we identify class expressions to define schemes, and we assemble these schemes defining objects to express dependencies between them.

The modules obtained by applying the heuristics we propose can be hierarchically organized to show the system module structure. In addition, this hierarchy of modules can be represented using a layered architecture composed of three layers (specific layer, general layer, and middleware layer), where each layer is a set of RSL modules that share the same degree of generality (Mauco et al., 2002).

Derivation of Functions

This step results in a set of functions that model the functionality in the application domain. Scenarios are the main source of information when defining functions, as they are natural language descriptions of the functionality in the domain. Each scenario can be classified as modifying or observing depending on whether it produces a change in the domain or not, and so corresponds to a generator or observer function. A scenario that just reports information will be observing.

Functions are usually identified at the top level as scenarios help to generate them there. Functions at one level in the hierarchy of modules frequently have counterparts at lower levels, but with different parameters. For each function in the top level module, we model the necessary functions in lower level modules, in order to simplify the legibility and maintainability of the specification. In Mauco et al. (2001c), we describe some heuristics to derive the functions. These heuristics help to identify and to model the functions by showing how to derive arguments and result types of functions, how to classify functions as partial or total, and how to define function bodies by analyzing the components of scenarios.

FUTURE TRENDS

Formal methods offer a wide spectrum of possible paths towards developing high-quality software. Their successful use for real systems is steadily growing from year to year.

Requirements engineering plays an important role in the development of a software system. Success with requirements capture and modeling is considered crucial as the remainder of the lifecycle activities of the software development process is highly dependent on this early foundation. As we have mentioned before, in the requirements phase, requirements are often specified informally with a language in which stakeholders are familiar.

Then, among the many challenges for the future, bridging the gap between requirements engineering and formal specifications is an important one to work in. The former offers much richer modeling abstractions, while the latter offers many advantages in constructing the software system. So, to take profit of both of them, one should look at ways for mapping the conceptually richer world of requirements engineering to the formal methods world.

CONCLUSION

The advantage of formal methods such as RAISE is they help to avoid requirements ambiguities and misinterpretations, and they provide a correct software development process based on mathematical proofs. However, formal specifications are not easy to understand by stakeholders. LEL and scenarios are valuable for supporting interaction with stakeholders in the initial phases of software development. Then, to contribute to bridging the gap between stakeholders and formal methods specialists, we have defined a three-step process to derive an initial specification in RSL from natural language models. In addition, this process contributes to fruitfully using the large amount of information usually available after problem analysis. We have applied this three-step process to a complete case study (Mauco & George, 2000), the Milk Production System.

The heuristics we have proposed are guidelines on how to start with the definition of an initial specification, taking into account the structured description of a domain provided by LEL and scenarios. The LEL provides structural features of the relevant terms in the domain (Leite, Hadad, Doorn & Kaplan, 2000), thus limiting the definition of types to those that correspond to significant terms in the domain. Using the behavioral description represented in the scenarios, it is possible to identify the main functionality to model in the specification. In addition, the structure proposed in Leite et al. (2000) to describe each scenario makes simpler the derivation of function signatures. However, even though LEL and scenarios have a precise structure and it is established what to write in their components, the same semantics may be usually expressed with many different natural language sentences. But, we think some of the problems found in the derivation, and associated with natural language expressiveness and flexibility, could be overcome if stronger standards or guidelines were imposed to the way of describing LEL symbols and scenarios.

Once the initial specification is derived, the process continues with the steps proposed in the RAISE Method. For example, the initial applicative and partially abstract specification derived could be developed into a concrete one to make use of the SML translator (George, 2001), and thus obtain a quick prototype to validate the specification and get a feeling of what it really does.

We plan to improve the three-step process we proposed by refining and completing the heuristics presented in this work, though obviously a complete automatic derivation is by no means possible, as LEL and scenarios contain all the necessary and unavoidable ambiguity of the real world, while the specification contains decisions about how to model this real world. Besides, a Web-based tool to assist in the derivation process is under development.

REFERENCES

George, C. (2001). *RAISE tools user guide*. UNU/IIST, Macau, Research Report 227. Online: *http://www.iist.unu.edu*

George, C. (2002). *Introduction to RAISE*. UNU/IIST, Macau, Technical Report 249. Online *http://www.iist.unu.edu*

George, C., Haff, P., Havelund, K., Haxthausen, A., Milne, R., Nielsen, C.B., Prehn, S. & Wagner, K.R. (1992). *The RAISE Specification Language*. BCS Practitioner Series. UK: Prentice-Hall.

George, C., Haxthausen, A., Hughes, S., Milne, R., Prehn, S. & Pedersen, J.S. (1995). *The RAISE Development Method*. BCS Practitioner Series. UK: Prentice-Hall.

Jacobson, I., Booch, G. & Rumbaugh, J. (1999). *The unified software development process*. Reading, MA: Addison-Wesley.

Lee, B. (2001). Automated conversion from a requirements documentation to an object-oriented formal specification language. *Proceedings of the 16th IEEE International Conference on Automated Software Engineering*. Los Alamitos, CA: IEEE Computer Society Press.

Lee, B. & Bryant, B. (2002). Prototyping of requirements documents written in natural language. *Proceedings of SESEC 2002, the 2002 Southeastern Software Engineering Conference*. USA.

Leite, J.C.S, Rossi, G., Balaguer, F., Maiorana, V., Kaplan, G., Hadad, G. & Oliveros, A. (1997). Enhancing a requirements baseline with scenarios. *Proceedings of the IEEE 3rd International Requirements Engineering Symposium* (pp. 44-53). Los Alamitos, CA: IEEE Computer Society Press.

Leite, J.C.S.P, Hadad, G., Doorn, J. & Kaplan, G. (2000). A scenario construction process. *Requirements Engineering Journal*, *5*(1), 38-61.

Leonardi, C. (2001). *Una estrategia de modelado conceptual de objetos basada en modelos de requisitos en lenguaje natural*. Master's thesis. Universidad Nacional de La Plata, Argentina.

Mauco, M.V. & George, C. (2000). *Using requirements engineering to derive a formal specification*. UNU/IIST, Macau, Research Report 223. Online *http://www.iist.unu.edu*

Mauco, M.V., Riesco, D. & George, C. (2001a). Deriving the types of a formal specification from a client-oriented technique. *Proceedings of the 2nd International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing* (pp. 1-8), Japan.

Mauco, M.V., Riesco, D. & George, C. (2001b). Heuristics to structure a formal specification in RSL from a client-oriented technique. *Proceedings of the 1st Annual International Conference on Computer and Information Science* (ICIS'01) (pp. 323-330), USA.

Mauco, M.V., Riesco, D. & George, C. (2001c). Using a scenario model to derive the functions of a formal specification. *Proceedings of the 8th Asia-Pacific Software Engineering Conference* (APSEC'01) (pp. 329-332). Los Alamitos, CA: IEEE Computer Society Press.

Mauco, M.V., Riesco, D. & George, C. (2002). A layered architecture for a formal specification in RSL. *Proceedings of the International Conference on Computer Science, Software Engineering, Information Technology, E-Business and Applications* (CSITeA'02) (pp. 258-263), Brazil.

Moreno Capuchino, A., Juristo, N. & Van de Riet, R.P. (2000) Formal justification in object-oriented modeling: A linguistic approach. *Journal of Data and Knowledge Engineering*, *33*(1), 25-47.

Nuseibeh, B. & Easterbrook, S. (2000). Requirements engineering: A roadmap. *Proceedings of the Conference on the Future of Software Engineering* (pp. 35-46). ACM Press.

Sommerville, I. & Sawyer, P. (1998). *Requirements engineering: A good practice guide*. New York: John Wiley & Sons.

van Lamsweerde, A. (2000a). Formal specification: A roadmap. *Proceedings of the Conference on the Future of Software Engineering* (pp. 147-159). ACM Press.

van Lamsweerde, A.(2000b). Requirements engineering in the year 00: A research perspective. *Proceedings of the 22nd International Conference on Software Engineering* (pp. 5-19). ACM Press.

KEY TERMS

Formal Methods: The variety of mathematical modeling techniques that are applicable to computer system (software and hardware) design. Formal methods may be used to specify and model the behavior of a system and to mathematically verify that the system design and implementation satisfy system functional and safety properties. These specifications, models, and verifications may be done using a variety of techniques and with various degrees of rigor.

Formal Specification: The expression, in some formal language and at a some level of abstraction, of a collection of properties some system should satisfy. A specification is formal if it is expressed in a language made of three components: the syntax (rules for determining the grammatical well-formedness of sentences), the semantics (rules for interpreting sentences in a precise, meaningful way in the domain considered), and the proof theory (rules for inferring useful information from the specification) (van Lamsweerde, 2000a).

RAISE Method: Encompasses formulating abstract specifications, developing them to successively more concrete specifications, justifying the correctness of the development, and translating the final specification into a programming language. The method is based on a number of principles such as separate development, stepwise development, invent and verify, and rigor. RAISE is an acronym for "Rigorous Approach to Industrial Software Engineering"; it gives its name to a formal specification language, the RAISE Specification Language, the associated method, and a set of tools. (George et al., 1995; George, 2002).

RAISE Specification Language (RSL): A formal specification language intended to support the precise definition of software requirements and reliable development from such definitions to executable implementations. It supports specification and design of large systems in a modular way, and thus it permits separate subsystems to be separately developed. It also provides a range of specification styles (axiomatic and model-based; applicative and imperative; sequential and concurrent), as well as supports specifications ranging from abstract (close to requirements) to concrete (close to implementations) (George et al., 1992).

Requirements: Descriptions of how the system should behave, or of a system property or attribute. They are defined during the early stages of a system development as a specification of what should be implemented. They should be statements of what a system should do rather than a statement of how it should do it (Soomerville & Sawyer, 1998).

Requirements Baseline: A mechanism proposed to formalize requirements elicitation and modeling, or a structure that incorporates descriptions about a desired system in a given application domain. Although it is developed during the requirements engineering process, it continues to evolve during the software development process. It is composed of five complementary views: the Lexicon Model View, the Scenario View, the Basic Model View, the Hypertext View, and the Configuration View (Leite et al., 1997).

Requirements Engineering: Comprehends all the activities involved in eliciting, modeling, documenting, and maintaining a set of requirements for a computer-based system. The term "engineering" implies that systematic and repeatable techniques should be used to ensure that system requirements are consistent, complete, relevant, and so forth (Soomerville & Sawyer, 1998).

ENDNOTE

[1] RAISE is an acronym for Rigorous Approach to Industrial Software Engineering.

Integrating Security in the Development Process with UML

Folker den Braber
SINTEF ICT, Norway

Mass Soldal Lund
SINTEF ICT, Norway

Ketil Stølen
SINTEF ICT, Norway

Fredrik Vraalsen
SINTEF ICT, Norway

INTRODUCTION

Today, most business processes and communications as well as a lot of everyday life situations involve IT technology. Apart from requirements on functionality, this development of IT systems has increased the need for security. Security issues are reaching the main headlines in the media on a regular basis. Virus, worms, *mis*configuration and program bugs are common problems in a world where new releases and updates are almost as frequently announced as spam e-mail pop-ups in our inboxes.

Having a closer look at the world of IT technology through security-colored glasses, we observe that security is often one step behind functionality. Security issues are mainly addressed as a reaction to existing problems. Like firemen and emergency units, security measures do not come into action before it is too late.

The concept of IT security contains a lot of different aspects. One of the IT security aspects is security risk analysis, in the sequel referred to as security analysis. Security analysis is an inevitable and crucial activity for every system developer, system user or system owner in order to get control over and knowledge about the security level of the actual system.

Security analyses are costly and time consuming and cannot be carried out from scratch every time a system is updated or modified. This motivates the need for specific methodology addressing the integration of security analysis and system development, providing access to, storage of, and maintenance of analysis results.

CORAS (2004) provides such a methodology in the form of so-called “model based security analysis”. The CORAS methodology combines traditional risk analysis techniques like HazOp (Redmill, Chudleigh, & Catmur, 1999), FTA (IEC 1025, 1990) and FMEA (Bouti & Ait Kadi, 1994) with system development techniques like UML (OMG, 2003b) and UP (Jacobson, Rumbaugh, & Booch, 1999). It builds on international standards for risk management: the Australian/New Zealand AS/NZS 4360 (1999), “Risk Management”; the ISO/IEC 17799 (2000), “Code of Practice for Information Security Management”; the ISO/IEC 13335 (2001), “Guidelines for the management of IT Security”; and system documentation in the form of the Reference Model for Open Distributed Processing (RM-ODP) (ISO/IEC 10746, 1995).

BACKGROUND

The CORAS methodology incorporates a documentation framework, a number of closely integrated security analysis techniques, and a risk management process based upon widely accepted standards. It gives detailed recommendations for the use of modeling with UML and similar languages in conjunction with security analysis in the form of guidelines and specified diagrams. Security analysis requires a firm, but nevertheless easily understandable, basis for communication between different groups of stakeholders. Graphical, object-oriented modeling techniques have proven well suited in this respect for requirements capture and analysis. We claim they are as equally suited as part of a language for communication in the case of security analysis. Class diagrams, use case diagrams, sequence diagrams, activity diagrams, dataflow diagrams, and state diagrams represent mature paradigms used daily in the IT industry throughout the world. They are supported by a wide set of sophisticated case tools, are to a large extent complementary, and together support all stages of system development.

The CORAS methodology may be separated into three different components: tools, processes and languages. This is shown in Figure 1. The language part defines common languages to support the methodology. The process part includes instructions for how to "execute" the methodology, that is, descriptions of what should be done, and how and when it should be done. The CORAS methodology for model based security analysis (MBSA) integrates aspects from partly complementary risk analysis methods and state-of-the-art modeling methodology. During execution of the methodology, one may require the use of different tools. The CORAS methodology includes a computerized platform that can be integrated with third party modeling tools and risk analysis tools. It also includes two languages: a UML-based specification language, the UML profile for security analysis (OMG, 2003c), targeting security risk analysis, and an XML markup for exchange of risk analysis data (World Wide Web Consortium, 2000).

In the section called *Model-Based Security Analysis*, the core of model-based security analysis, based on integration of risk management and system development, is explained.

Related Approaches

The CORAS methodology, with its unique approach to security analysis from a modeling point of view, addresses a problem area in which also other methods and technologies exist. Some important ones are mentioned here.

The Common Criteria (CC)

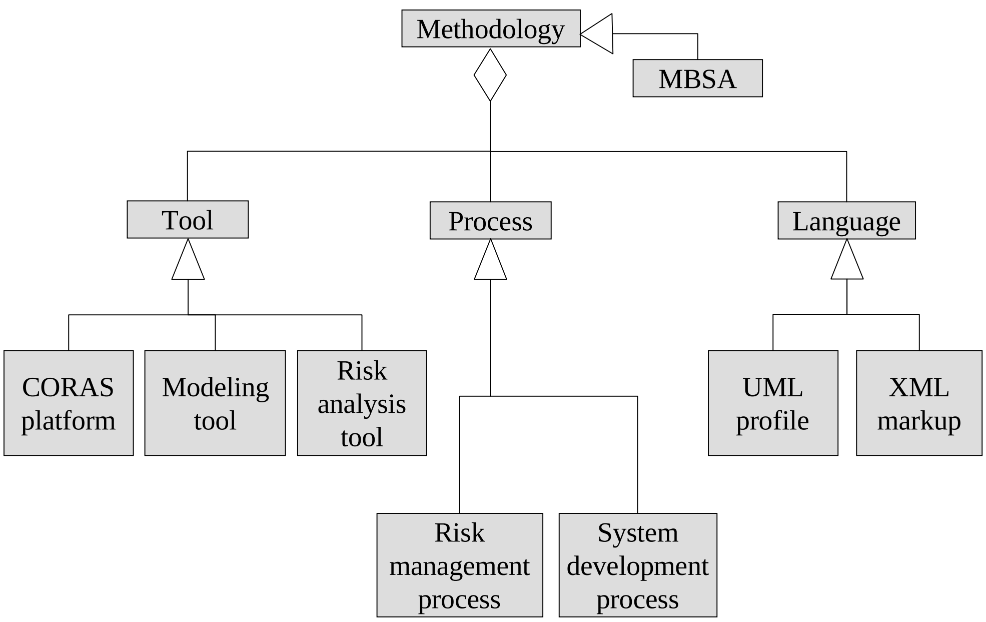
Since 1990, work has been going on to align and develop existing national and international schemes into one mutually accepted framework for testing IT security functionality. The Common Criteria (CC) (CCO, 2002) represents the outcome of this work. The Common Criteria project harmonizes the European "Information Technology Security Evaluation Criteria (ITSEC)" (Communications-Electronics Security Group, 2002), the "Canadian Trusted Computer Product Evaluation Criteria (CTCPEC)", and the American "Trusted Computer System Evaluation Criteria (TCSEC)" and "Federal Criteria (FC)". Increasingly, it is replacing national and regional criteria with a worldwide set accepted by the International Standards Organization (ISO15408) (ISO/IEC, 1999).

The CC and CORAS are orthogonal approaches. The CC provides a common set of requirements for the security functions of IT products and systems as well as a common set of requirements for assurance measures that are applied to the functions of IT products and systems during a security evaluation. CORAS provides a concrete methodology for model based security analysis.

Surety Analysis (SA)

Surety Analysis (SA), developed in Sandia National Laboratories (2003), is a methodology based on the creation of

Figure 1. Structure of the CORAS methodology



an explicit model that covers several aspects of the system's behavior. The modeling framework in SA is proprietary, whereas in CORAS the standardized RM-ODP is used as a common basis. SA supports modeling by means of basic techniques such as interaction, state and dataflow diagrams. CORAS uses the full descriptive power of UML/OCL (Object Constraint Language) (OMG, 2003a) enhanced with aspects of other modeling paradigms specific to security modeling.

COBIT

The control objectives for information and related technology (COBIT) (2003) addresses the management of IT. COBIT and CORAS are orthogonal approaches. COBIT focuses on control objectives defined in a process-oriented manner following the principles of business reengineering.

CRAMM

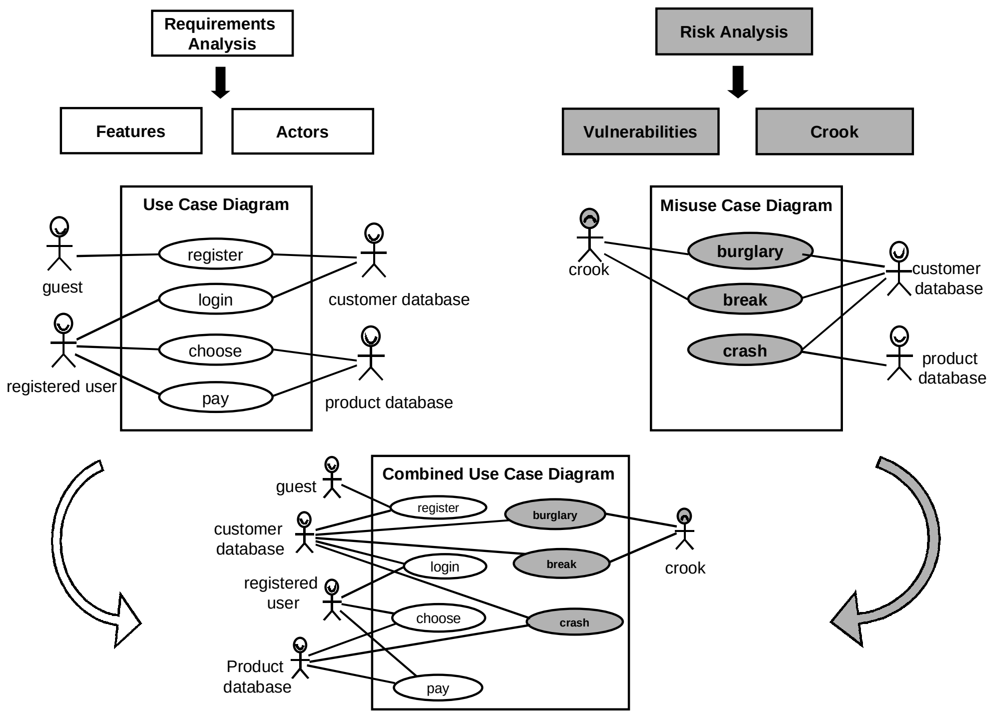
CCTA Risk Analysis and Management Methodology (CRAMM) (Barber & Davey, 1992) was developed by the British Government's Central Computer and Telecommunications Agency (CCTA) with the aim of providing a structured and consistent approach to computer security management for all systems. The UK National Health Service considers CRAMM to be the standard for security analysis of information systems within health care establishments. CRAMM has been an important source of inspiration for CORAS, and aspects of CRAMM have been incorporated in CORAS. Contrary to CRAMM, CORAS provides a security analysis process in which state-of-the-art modeling methodology in the form of UML is tightly integrated.

MODEL-BASED SECURITY ANALYSIS

As mentioned before, "model-based security analysis" is a core concept in the CORAS methodology. This section explains what "model-based security analysis" is, and gives a glimpse on how it is to be applied.

The basis of model-based security analysis is about identifying negative functionality and making this *mis*functionality (Sindre & Opdahl, 2000) visible and known in order to protect the system. This misfunctionality can be everything from vulnerabilities that open for hacker attacks to bad user interface design increasing the chance for crucial mistakes. The idea is that the well-accepted techniques for specification and design of functionality are also suited for model misfunctionality. This is illustrated by Figure 2.

Figure 2. Model-based security analysis



The requirements analysis forms the basis for the traditional kind of modeling (Cockburn, 1997) with focus on desired behavior. Security analysis has a similar role but focuses on unwanted behavior. Every designed system carries its own risks, and it is important that they are known. It is therefore not enough to model only the desired behavior but also the unwanted behavior. In addition, the malicious actors, or crooks, need to be identified just like the normal actors. Model-based security analysis is about documenting the results of traditional risk analysis techniques in the same way as we are used to doing for system requirements. The design process needs to take into account both wanted and unwanted behavior and designed actors and malicious actors. As shown here, UML like graphical techniques can be used for both aspects providing the complete documentation of the system design from both good and bad angles.

UML Profile

The CORAS methodology defines its own UML stereotypes and methods for describing UML models related to security analysis. These specific security related UML aspects are caught in the CORAS UML profile (OMG, 2004). This UML profile for security analysis introduces a meta model that defines an abstract language for supporting model-based security analysis. Furthermore, the profile provides a mapping of classes in the meta model to UML modeling elements by defining so-called stereotypes, and introduces special symbols (icons) for representing the stereotypes in UML diagrams.

The motivation for the profile is the practical use of UML to support security management in general, and security analysis, in particular. In the model-based security analysis methodology of CORAS, UML models are used for three different purposes:

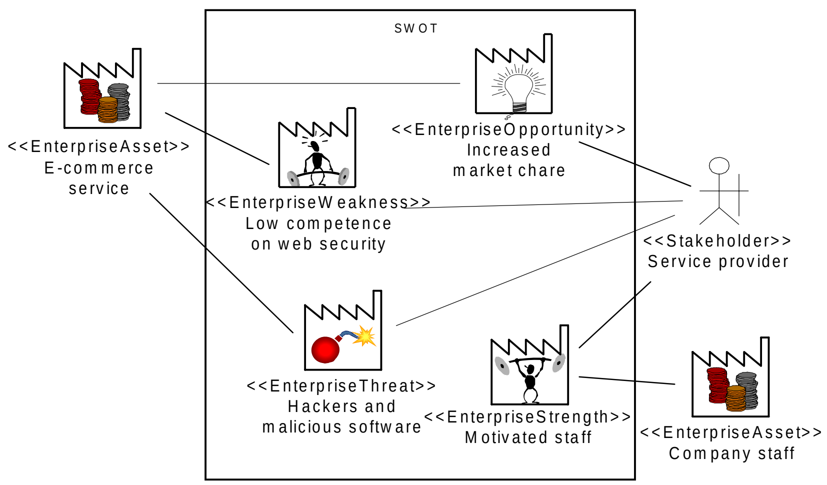
- Target of evaluation: Describing the target of evaluation at the right level of abstraction.
- Communication: Facilitating communication and interaction between different groups of stakeholders involved in a security assessment.
- Documentation: Documenting security assessment results and the assumptions on which these results depend to support reuse and maintenance.

The target of evaluation in a security analysis is a system or part of a system to be assessed, but also the context and assumptions under which the system is assessed. The CORAS profile supports describing the target of evaluation by allowing explicit modeling of the context of assessments, like stakeholders and assets. An example of this is shown in the SWOT diagram in Figure 3, where the assets and stakeholders are directly related to main characteristics of the system to be assessed.

The effectiveness and success of security analysis depends on the extent to which the involved stakeholders and analysts understand and are understood by each other. Users, system developers, decision makers and system managers are different examples of such stakeholders. These will have different backgrounds and competencies, and misunderstandings are likely to occur. The CORAS profile supports communication by introducing easy to understand icons representing the various concepts of the ontology, as shown in the SWOT diagram of Figure 3.

Imagine the following security analysis results generated by a security analysis session as defined in the CORAS methodology.

Figure 3. SWOT (Strength-Weakness-Opportunity-Threat) diagram



"Regarding the availability of service as an important asset of the system, it was identified that some persons with the right knowledge and the wrong intentions could threaten this availability of service by flooding the system. The indicated flooding threat might generate a Denial-of-Service attack which directly affects the availability of service. One way of securing the system against this risk is to apply authentication on requests."

The same results represented in UML using the CORAS profile are shown in Figure 4. Clearly, this representation is easier to communicate and leaves less room for misunderstanding. It also shows how the CORAS methodology allows for specifying discovered risks and treatments related to threats. One might argue that the result in text is just as clear, but this will typically only be the case for relatively simple cases. The CORAS methodology advices to use the given diagrams and modeling techniques, but does not restrict one from using normal text in addition.

In CORAS, security analysis results are documented by the means of elements (diagrams, tables, etc.). Each of these elements belongs to an activity of the risk management process. The CORAS profile supports documentation by introducing types of diagrams supporting various activities of the risk management process being founded in a meta model conforming to the underlying data structure.

The diagrams mentioned here become part of the system documentation and are also stored like that. This documentation process is supported by the CORAS platform, a tool containing both guidance and storage functionality. Recent security analysis results are stored while the process is supported by results from past security analysis experiences. This opens for running the CORAS methodology parallel to the traditional system development process in addition to being able to follow the maintenance and upgrading phases by similarly maintaining and upgrading the security analysis results.

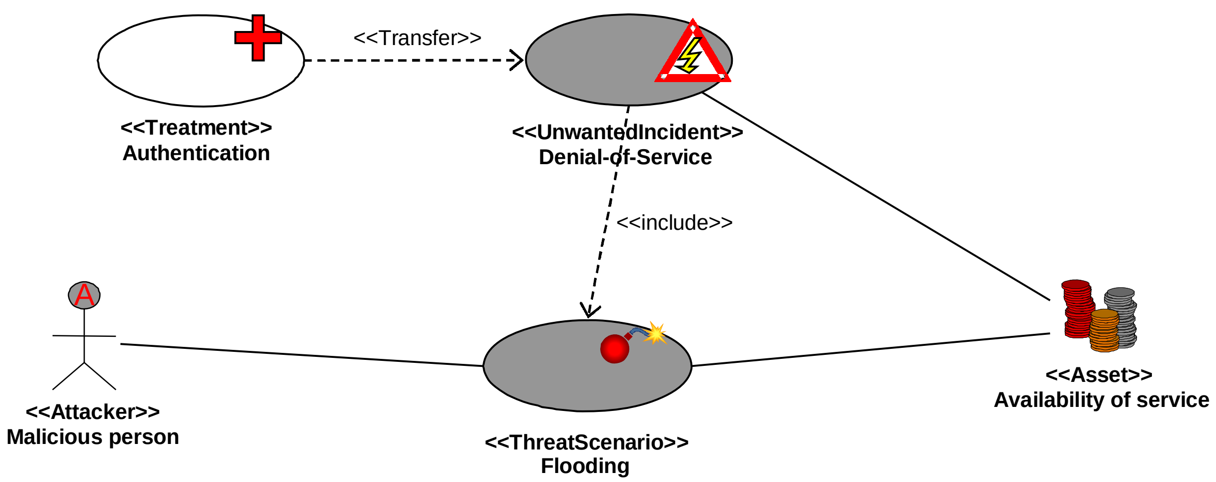
FUTURE TRENDS

A main future trend regarding security analysis is the integration of security in the development process of IT systems. Developers will have to start to think security at the very start of the development process. Today, security and dealing with risks do not always get the attention they deserve: It is often forgotten or simply ignored. Apart from door locks and bank safes, most people let security "live its own life" and often just hope for the best. This attitude is probably necessary in our everyday life: There are so many things that can go wrong, and we cannot worry about all of them. When it comes to IT systems, and especially security critical IT systems, such an attitude could be fatal. We, therefore, need to break the habit of living on hope and turn it into informed, calculated, and constructive actions. To be able to do this, methods and tools are needed to guide us in our security analyses.

CONCLUSION

The CORAS methodology and the corresponding developed CORAS platform help developers in keeping focus on security and provide guidelines on how security analysis can become a natural part of the system development process. The combination of security analysis with system development methodologies like the Unified Process makes security related adjustments possible during all phases of the development process.

Figure 4. Threat diagram



A security analysis identifies and documents the undesirable system behavior. The CORAS UML profile employs the same kind of modeling techniques to document the undesirable behavior that leading system development methodologies employ to capture requirements to the desirable system behavior. For example, as pointed out previously, in the same way as use cases are employed to capture desired functionality, misuse cases may be employed to capture threat scenarios. The fact that the CORAS methodology employs techniques well known to system engineers simplifies security analysis during system development.

The experiences from applying the CORAS methodology in major field trials within e-commerce and telemedicine are promising. The CORAS project involved seven major field trials, each of which made use of the earlier versions of the CORAS methodology and tool to analyze the security of an already existing system or application. The field trials were all of industrial size. Two field trials within telemedicine are documented in Stathiakis et al. (2003) and Stamatiou et al. (2003), while Dimitrakos et al. (2002) and Raptis, Dimitrakos, Gran and Stølen (2002) document field trials within e-commerce. More documentation on the CORAS methodology can be found on the CORAS Web site (CORAS, 2004).

The research on which this article reports has partly been funded by the Research Council of Norway projects COBRA (152209/431) and SECURIS (152839/220) and partly by the 5th Framework EU project CORAS (IST-2000-25031). The CORAS consortium consisted of 11 partners from four countries: CTI (Greece), FORTH (Greece), IFE (Norway), Intracom (Greece), NCT (Norway), NR (Norway), QMUL (UK), RAL (UK), SINTEF (Norway), Solinet (Germany) and Telenor (Norway). The results reported in this article have benefited from the joint efforts of the CORAS consortium.

REFERENCES

Australian/New Zealand Standard AS/NZS 4360 (1999). Risk management.

Barber, B., & Davey, J. (1992). The use of the CCTA risk analysis and management methodology CRAMM. *Proc. MEDINFO92*, North Holland (pp. 1589-1593).

Bouti, A., & Ait Kadi, D. (1994). A state-of-the-art review of FMEA/FMECA. *International Journal of Reliability, Quality and Safety Engineering, 1,* 515-543.

Cockburn, A. (1997). Structuring use cases with goals. *Journal of Object-oriented Programming*, Sep/Oct, 35-40, Nov/Dec, 56-62.

Common Criteria Organization. (2002). Common criteria for information technology security evaluation. Retrieved in 2003 from the World Wide Web at *http://www.commoncriteria.org*

Communications-Electronics Security Group. (2002). Information security evaluation criteria. Retrieved in 2003 from the World Wide Web at *http://www.iwar.org.uk/cip/resources/uk/*

Control objectives for information and related technology (COBIT). Retrieved in 2003 from the World Wide Web at *http://www.isaca.org/*

CORAS (2003). A platform for risk analysis of security critical systems. IST-2000-25031. Retrieved in 2004 from the World Wide Web at *http://coras.sourceforge.net/*

Dimitrakos, T., Ritchie, B., Raptis, D., Aagedal, J.Ø., den Braber, F., Stølen, K., & Houmb, S.H. (2002). Integrating model-based security risk management into e-business systems development " The CORAS approach. In *Proc. 2nd IFIP Conference on E-Commerce, E-Business, E-Government (I3E 2003)*, (pp.159-175). Kluwer.

IEC 1025. (1990). Fault tree analysis (FTA).

ISO/IEC (1999). Information technology – Security techniques – Evaluation Criteria for IT Security ISO/IEC, 15408-1.

ISO/IEC 10746. (1995). Basic reference model of open distributed processing.

ISO/IEC 17799. (2000). Information technology – Code of practice for information security management.

ISO/IEC TR 13335. (2001). Information technology – Guidelines for the management of IT security.

Jacobson, I., Rumbaugh, J., & Booch, G. (1999). *The unified software development process*. Object Technology Series. Addison-Wesley.

OMG. (2003a). Object constraint language specification. Part of the UML specification. OMG document number: ad/03-01-07.

OMG. (2003b). Unified modeling language specification. Version 2.0. OMG document number: ptc/03-09-15.

OMG. (2004). UML for QoS & fault tolerance. OMG document number: ptc/04-06-01.

Raptis, D., Dimitrakos, T., Gran, B.A., & Stølen, K. (2002). The CORAS approach for model-based risk analysis applied to the e-commerce domain. In *Proc. Communication and Multimedia Security (CMS 2002)*, (pp.169-181). Kluwer.

Redmill, F., Chudleigh, M., & Catmur, J. (1999). *Hazop and software Hazop*. Wiley.

Sandia National Laboratories, Surety Analysis. Retrieved in 2003 from the World Wide Web at *http://www.sandia.gov*

Sindre, G., & Opdahl, A.L. (2000). Eliciting security requirements by misuse cases. In *Proc. TOOLS_PACIFIC 2000*, (pp.120-131). IEEE Computer Society Press.

Stamatiou, Y., Skipenes, E., Henriksen, E., Stathiakis, N., Sikianakis, A., Charalambous, E., Antonakis, N., Stølen, K., den Braber, F., Lund, M. S., Papadaki, K., & Valvis, G. (2003). The CORAS approach for model-based risk management applied to a telemedicine service. In *Proc. Medical Informatics Europe (MIE 2003)*, (pp.206-211). IOS Press.

Stathiakis, N., Chronaki, C., Skipenes, E., Henriksen, E., Charalambous, E., Sykianakis, A., Vrouchos, G., Antonakis, N., Tsiknakis, M., & Orphanoudakis, S. (2003). Risk assessment of a cardiology eHealth service in HYGEIAnet. In *Proc. Computers in Cardiology (CIC 2003)*.

World Wide Web Consortium. (2000, October 6). Extensible Markup Language (XML) v1.0, W3C recommendation (2nd ed.).

KEY TERMS

Risk: The chance of something happening that will have an impact upon objectives. It is measured in terms of consequence and likelihood.

Risk Analysis: A systematic use of available information to determine how often specified events may occur and the magnitude of their consequences.

Risk Assessment: The overall process of risk analysis and risk evaluation.

Risk Evaluation: The process used to determine risk management priorities by comparing the level of risk against predetermined standards, target risk levels or other criteria.

Risk Treatment: Selection and implementation of appropriate options for dealing with risk.

Security Analysis: Thorough analysis of a system in order to get a complete picture of its security level. Vulnerabilities, risks, treatments and their costs are identified.

Stakeholders: Those people and organizations who may affect, be affected by, or perceive themselves to be affected by, a decision or activity.

SWOT Analysis: Strength, Weakness, Opportunity and Effect Analysis.

Target of Evaluation (TOE): An IT product or system and its associated administrator and user guidance documentation that is the subject of an evaluation.

Virus: Program that can infect other programs by modifying them; the modification includes a copy of the virus program, which can then go on to infect other programs.

Integration Framework for Complex Systems

Edward Szczerbicki
The University of Newcastle, Australia

INTRODUCTION

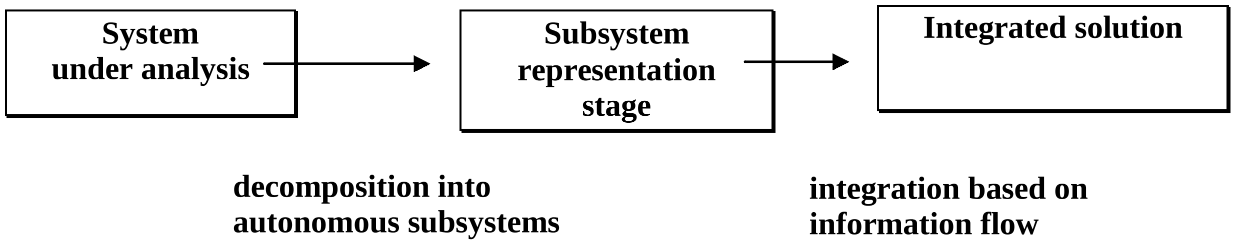
Information is seen as one of the main resources that systems analysts try to use in an optimal way. In this chapter we show how this resource can be used in integration issues. We introduce the problem of information-based integration, propose a solution, and briefly discuss future trends in this area. Systems become increasingly complex. Their decomposition into smaller units is the usual way to overcome the problem of complexity. This has historically led to the development of atomized structures consisting of a limited number of autonomous subsystems that decide about their own information input and output requirements, that is, can be characterized by what is called an information closure. In a real-world context, autonomous subsystems consist of groups of people and/or machines tied by the flow of information both within a given subsystem and between this subsystem and its external environment (Esteve, 2002; Szczerbicki, 2003; Tharumarajah, 1998). Autonomous subsystems can still be interrelated and embedded in larger systems, as autonomy and independence are not equivalent concepts. These ideas are recently gaining very strong interest in both academia and industry, and the atomized approach to complex systems analysis is an idea whose time has certainly come (Liu & Ling, 2003; Orlowski, 2002).

Complex systems (for example manufacturing) are often viewed as sets of components (agents, subsystems) supporting separate functions. Many organizations operate in this highly compartmentalized manner. It appears that the general direction of systems in the future, however, is toward linking together function-specific agents into fully integrated entity. An integrated system is a system that consists of agents/subsystems that efficiently contribute to the task, functional behaviour, and performance of a system as a whole. It is believed that such an integration can be achieved through the flow of information. “Integration,” as used in this article, should not be confused with integration at a physical level by means of computer networks or computer buses. Rather, the semantics of integration is addressed - the information that subsystems should share. While structuring the approach presented, the first consideration was to design some tools that could be easily implemented as components of an intelligent system supporting development of system configurations integrated by the flow of information. Due to the complexity and creativity associated with the early stages of such a development it is quite clear that the way a practicing analyst solves a system configuration problem cannot be easily implemented. This explains the need for new tools and approaches that solve the problem but at the same time can be supported by a computer. Elements of such an approach are presented in this article.

BACKGROUND

We propose a three-stage approach for the development and analysis of complex systems. The involved stages are systems decomposition, subsystem modelling represen-

Figure 1. Overview of a three-stage approach to complex systems integration



tation, and integration at the level of information flow. Figure 1 depicts the general underlying idea behind this approach.

The essence of information-based integration problem can be formulated as follows and illustrated as in Figure 2.

Theoretical framework to provide support for systems integration as outlined includes the following fundamentals:

- the syntax for connections of autonomous subsystems,
- the mechanism for guiding the generation of such connections,
- the integration algorithm.

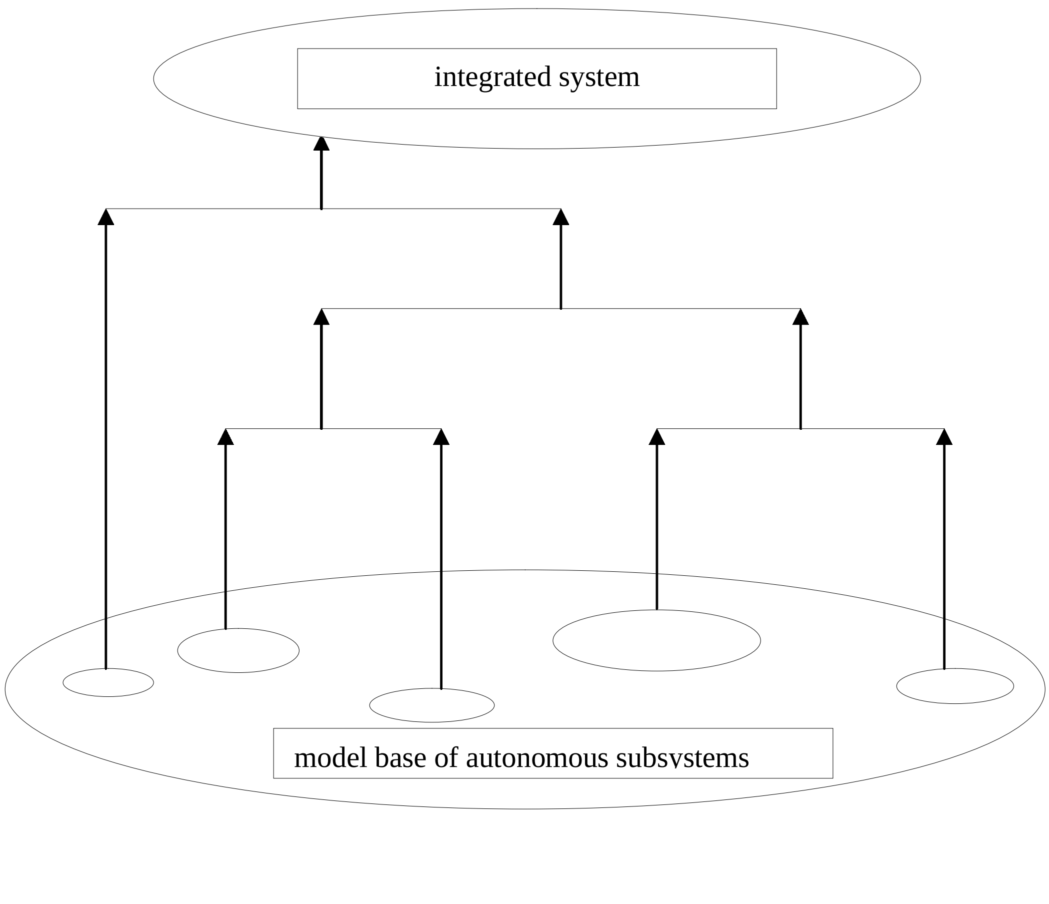
The aim of this article is to briefly outline these fundamentals.

The author, together with his collaborators, has been researching information-based integration issues since the early nineties, starting at The University of Iowa (Kusiak, Szczerbicki & Vujosevic, 1990, 1991), Iowa City, USA, continuing at The GMD FIRST, Berlin, Germany (Szczerbicki, 1994) and currently working on these issues at The University of Newcastle, Newcastle, Australia (Szczerbicki, 2003). The aim of this entry, which is based on previous research publications by the author, is to overview the integration problem from the perspective of the author's experience and to place it among the work of others.

Integration problem is recently gaining very strong interest in both academia and industry. This is particularly apparent in the area of design and modelling of manufacturing systems (O'Grady, 1999). Model development and synthesis (that resembles autonomous systems develop-

Figure 2. Illustrative example of a three-level hierarchical tree of the bottom-up integration process

Given the informational inputs and outputs of autonomous subsystems, find the overall system being designed that meets the desired functions and is integrated through the flow of information.



ment and integration) is frequently based on the general systems theory and it uses hierarchical structures and a number of model base concepts (Esteve, 2002; Rolstadas & Andersen, 2000; Wyzalek 1999). In Raczkowsky and Reithofer (1998), the future development of a hierarchical communication model for coordination of a set of agents performing several functions is addressed. A conceptual modeling approach to represent the complexities in CIM (Computer Integrated Manufacturing) systems, including such issues as information acquisition, storage, dissemination and the time and costs associated with such informational activities is proposed in Tharumarajah (1998). The problem of coordination of multiagent manufacturing systems developed to fulfill their functional requirements advocating a decentralized approach in which each agent has relative autonomy over its own actions is discussed in Pacholski (1998). In O'Grady (1999), the application of modular paradigm to integration of systems in which planning, grouping, and scheduling are the central functional areas is described. The role of the flow of information in the process of integration is discussed in Prakken (2000) and Orlowski (2002). In Kamrani and Sferro (1999), the integration of manufacturing agents (information islands and automation islands) using knowledge-based technology is proposed for the factory of the future.

Traditionally, information system analysts have been solving the integration problem in an ad hoc manner. What we propose through our research is a formal integration approach suitable for computer implementation.

Information-based integration problem should be seen as one of the challenges within the broader, more general aim to enhance systems performance through the flow of information. Engineering, operations research, information science and management science use scientific and engineering processes to design, plan, and schedule increasingly more complex systems in order to enhance their performance. One can argue that systems have grown in complexity over the years, mainly due to increased strive for resource optimization combined with a greater degree of vagueness in the system's environment. Information is seen as one of the main resources that analysts try to use in an optimal way in complex systems. Proper design of information flow, its management, its use, its maintenance, that is, information engineering, is critical in systems' abilities to act appropriately in an uncertain environment – to act intelligently (Baba et al., 2001; Bogdan, 2000; Gunasekaran & Sarhadi, 1997; Morabito, 1997; Prakken, 2000; Tharumarajah, 1998).

INFORMATION BASED INTEGRATION

Autonomous subsystems are matched using informational input and output defined at representation stage (Figure 1). For example, in the domain of environmental engineering, information may represent geographical positioning, urban planning, sources of pollution, traffic data, and the like. In the domain of manufacturing systems, informational inputs and outputs may represent material availability, tool availability, machine availability, number of parts produced, and number of products assembled. After the matching has been accomplished, the informational input variable of a given subsystem represents the value of the informational output variable of the subsystem to which it has been connected. For example, if the input variable X of autonomous subsystem AS2 is matched with the output variable Y of AS1, then the syntax of this connection will be given as:

$$AS1.Y \longrightarrow AS2.X \tag{1}$$

Similar simple syntax can be used for all structures that can be produced during the integration process. These structures are enclosed into higher-level subsystems using ports. The informational input and output ports provide an interface to the subsystem environment. This interface is used to develop hierarchical structures. With the syntax of autonomous subsystems connections in place, the mechanism for guiding the generation of such connections is required. The mechanism needs to represent qualitative system theoretic knowledge, and so it is based on IF... THEN production rules.

Generation of connections between elements in model base of autonomous subsystems is guided by the following production rules (Kusiak, Szczerbicki & Vujosevic, 1990, 1991; Szczerbicki, 2003):

- Rule 1
  IF there is only one element left
  THEN do not generate connections
- Rule 2
  IF a single element that is left includes boundary inputs and outputs only
  THEN it is an overall system
- Rule 3
  IF there are more than one element
  THEN select a connection for an input boundary element
- Rule 4
  IF there are elements other than the boundary elements
  THEN do not specify any connections that involve boundary elements only
- Rule 5
  IF an element is an input boundary element
  THEN it cannot accept an input from any other element

- Rule 6
  IF an element is an output boundary element
  THEN it cannot provide an input to any other element
- Rule 7
  IF two elements have identical output and input variables
  AND there are no production rules that prevent from connecting them
  THEN specify the connections for these elements
- Rule 8
  IF there are no elements with identical input and output variables
  AND there are elements with partially identical input and output variables
  AND there are no production rules that prevent from connecting them
  THEN specify the connection for these elements beginning with the closest match
- Rule 9
  IF a connection for an input boundary element has been specified
  THEN continue with selecting connections for elements that have not been listed in the specifications
- Rule 10
  IF there are boundary elements only
  THEN specify connections between them

These production rules are domain independent and were structured using the underlying general systems theory. The analyst may, however, add domain-based production rules. They may follow, for example, the safety requirements, emission data, traffic data, or other constraints imposed by the analyst.

The last tool needed for simulation of the information-based integration process is an integration algorithm. The algorithm guides the simulation process across various levels of integration illustrated in Figure 2. It was developed with the assumption that in order to enter the next level of integration it is enough to generate just one connection in a given step. Elements taking part in integration that are not matched at integration level i are considered for matching at level $i+1$. The algorithm terminates at the level at which it will no longer be possible to match subsystems into pairs (no connections will be generated). At each integration level, production rules presented previously are fired during simulation process to generate connections between remaining integration elements. The integration algorithm is the last element of fundamentals of a framework to support systems integration process. The algorithm includes the following simple steps (Kusiak, Szczerbicki & Vujosevic, 1990, 1991; Szczerbicki, 2003):

- Define database of autonomous sub systems (bottom level of illustrative example of integration process in Figure 2)
  Set level = 1.
- Generate connections between elements at current•
- If no connections are generated, stop;
  Otherwise, match elements in database into pairs using the existing connections.
- Define informational input and output variables for subsystems generated by the matching process.
- Remove from the database all elements that have taken part in the matching process.
- Add to the database all subsystems generated by the matching process.
- . Set level = level + 1 and go to Step 2.

The presented framework for information-based integration has been applied in numerous real-life based cases in which simulation was used to arrive at integrated problem solution. Some particularly successful applications included system modelling and integration for a coal mine (Szczerbicki & Charlton, 2001), integrated agile manufacturing strategy (Szczerbicki & Williams, 2001), modelling for steel processing (Szczerbicki & Murakami, 2000) and integration of maintenance services (Szczerbicki & White, 2003). In all these cases all three stages of systems decomposition, representation and integration were present (see Figure 1), integration representing the last step in the process of arriving at a systems structure integrated by the flow of information.

FUTURE TRENDS

Information-based system integration is gradually becoming one of the main challenges of our new millennium of information age. The most prominent future trend in this area is focused on using Web-based technologies for tackling problems of integration. Within this trend we have a rapid increase of research efforts towards developing metamodelling architectures and interoperability of Web-enabled information flows (Terrase et al., 2003), semistructured data integration (Liu & Ling, 2003), schema integration (Castana et al., 2003), and Web-based aggregation architectures (Bussler, 2003). The explosive popularity of the Web makes it an ideal integration tool as it opens the possibility to integrate geographically distributed systems. Also, in the future the Web may become a universal platform to synthesize a number of integration approaches (for example like the one presented in this article) into one universal Web-based integration interface.

CONCLUSION

Information flow integration is one of the major activities of the design process of an integrated system. The outcome of the integration process is the overall system integrated through the flow of information.

In this entry, integration problem is formulated as follows: Given the informational inputs and outputs of autonomous subsystems, find the overall system being designed that meets the desired functions and is integrated through the flow of information.

Autonomous subsystems are integrated using an algorithm into an overall system that has a hierarchical structure. General production rules supporting generation of connections for subsystems relate to the underlying systems theory. They are structured independently of the system's domain and cannot be modified by a system analyst. Production rules ensure that only feasible variants of the designed system are explored.

REFERENCES

Baba, N., Jain, L.C., & Howlett, L.J. (2001). *Knowledge-based intelligent information engineering systems and allied technologies*. Osaka: IOS Press.

Bogdan, R.J. (2000). *Minding minds*. MIT Press.

Bussler, C. (2003). Application service provider aggregation architecture. In A. Dahanayake & W. Gerhardt (Eds.), *Web-enabled systems integration: Practices and challenges*. Hershey, PA: Idea Group Publishing.

Castana, S., Antonellis V.D., Capitani S., & Melchiori M. (2003). Data schema integration in Web-enabled systems. In A. Dahanayake & W. Gerhardt (Eds.), *Web-enabled systems integration: Practices and challenges*, IDEA: London.

Esteve, Y.V. (2002). Notional logic of systems. *Cybernetics and Systems, 33,* 189-202.

Gunasekaran, A., & Sarhadi M. (1997). Planning and management issues in enterprise integration. *Concurrent Engineering: Research and Application, 5,* 98-102.

Kamrani, A.L., & Sferro, P.R. (1999). *Direct engineering: Toward intelligent manufacturing*. Berlin: Kluwer.

Kusiak, A., & Szczerbicki, E. (1990, July 18-20). Rule-based synthesis in conceptual design. In M. Jamshidi & M. Saif (Eds.), *Proceedings of The 3rd International Symposium on Robotics and Manufacturing: Research, Education, and Applications (ISRAM'90)*, Vancouver, B.C., Canada (vol. 3, pp. 757-762).

Kusiak, A., Szczerbicki, E., & Vujosevic, R. (1991). Intelligent design synthesis: An object oriented approach. *International Journal of Production Research, 29*(7), 1291-1308.

Liu, M., & Ling, T.W. (2003). Towards semistructured data integration. In A. Dahanayake & W. Gerhardt (Eds.), *Web-enabled systems integration: Practices and challenges*, Hershey, PA: Idea Group Publishing.

Morabito, F.C. (1997). *Advances in intelligent systems.* Amsterdam: IOS Press.

O'Grady, P. (1999). *The age of modularity*. IA: Adams and Steele Publishers.

Orlowski, C. (2002). Soft modelling support for managing knowledge-based information technology projects. *Cybernetics and Systems, 33,* 401-412.

Pacholski, L. (1998). Fuzzy logic application in ergonomic renewal of multiagent manufacturing systems. *Cybernetics and Systems, 29,* 715-728.

Prakken, B. (2000). *Information, organization and information systems design*. Berlin: Kluwer.

Raczkowsky, J., & Reithofer, W. (1998). Design of Consistent Enterprise Models, *Cybernetics and Systems.* Vol. 29, pp. 525-552.

Rolstafas, A., & Andersen, B. (2000). *Enterprise modelling*, Kluwer: Berlin.

Szczerbicki, E., (1994). Flow of information - the semantics of integration. *Systems Analysis, Modelling, Simulation*, Vol. 16, pp. 229-243.

Szczerbicki, E. (2003). Knowledge based integration for multi-agent systems in information rich environments, *International Conference on Information Technologies in Environmental Engineering*, Gdansk, Poland, 2003 (proceedings on CD ROM).

Szczerbicki, E., & Charlton, G. (2001). System modelling and simulation for a coal mine, *Systems Analysis, Modelling, Simulation* , 40, 111-123.

Szczerbicki, E., & Williams, J. (2001). Developing agile manufacturing strategy with AweSim simulation platform, *Cybernetics and System*, 32, 719-736.

Szczerbicki, E., & Murakami, R. (2000). Simulation-based decision support platform for steel processing, *Systems Analysis Modelling. Simulation* , 39, 461-478.

Szczerbicki, E., & White, W. (2003). Simulation modelling for managing complex systems: case study for maintenance services, International Conference on Information

Systems Application and Technology, ISAT'2003, Wroclaw, Poland, September 2003, 15-22.

Terrasse, M-N., Becker, G., & Savonnet, M. (2003). Metamodeling Architectures and Interoperability of Web-Enabled Information Systems, In *Web-enabled systems integration: Practices and challenges*, A. Dahanayake & W. Gerhardt (Eds), Hershey, PA: Idea Group Publishing.

Tharumarajah, A. (1998). A self-organising model for scheduling distributed autonomous manufacturing agents, *Cybernetics and Systems*. 29, 461-480.

Wyzalek, J. (1999). *Systems Integration Success*. Auerbach: London.

KEY TERMS

Autonomous (Sub)System: A system that decides about its own information input and output requirements.

Complex System: From mathematical perspective a system which is described by differential or difference equations; from informational perspective a system for which information is the main resource and functioning in information rich environment.

General System Theory: Collection of tools, approaches, hypotheses and models that can be used for scientific discovery.

Information Based Integration: Process in which given informational inputs and outputs of autonomous (sub)systems, analyst develops the overall system being designed that meets the desired functions and is interconnected through the flow of information.

Information closure: Boundary defined by information and its sources necessary for autonomous (sub)systems functioning.

Information Engineering: Proper design of information flow, its management, its use, and its maintenance.

Information system: A collection of organised procedures collecting, storing, processing and retrieving data. It provides information to support the organisation.

Intelligence: Ability of a given system to act appropriately (i.e. to increase the probability of the achievement of given aims) in an uncertain and changing environment.

System: A collection of components (subsystems) together with their interrelations.

Integrative Document and Content Management Solutions

Len Asprey
Practical Information Management Solutions Pty Ltd., Australia

Michael Middleton
Queensland University of Technology, Australia

INTRODUCTION

Developments in office automation, which provided multiple end-user authoring applications at the computer desktop, heralded a rapid growth in the production of digital documents and introduced the requirement to manage capture and organization of digital documents, including images. The process of capturing digital documents in managed repositories included metadata to support access and retrieval subsequent to document production (D'Alleyrand, 1989; Ricks, Swafford & Gow, 1992).

The imperatives of documentary support for workflow in enterprises, along with widespread adoption of Web-oriented software on intranets and the Internet World Wide Web (WWW), has given rise to systems that manage the creation, access, routing, and storage of documents, in a more seamless manner for Web presentation. These content management systems are progressively employing document management features such as metadata creation, version control, and renditions (Megill & Schantz, 1999; Wiggins, 2000), along with features for management of content production such as authoring and authorization for internal distribution and publishing (Addey et al., 2002; Boiko, 2002; Hackos, 2002; Nakano, 2002).

If business applications are designed taking into account document and Web content management as integral constructs of enterprise information architecture, then the context of these solutions may be an integrative document and content management (IDCM) model (Asprey & Middleton, 2003). As the name implies, the IDCM model aspires to combine the features of a document management system with the functionality of Web content management. An integrative business and technology framework manages designated documents and their content throughout the continuum of their existence and supports record-keeping requirements.

The IDCM model supports system capabilities for managing digital and physical documents, e-mail, engineering and technical drawings, document images, multimedia, and Web content. These systems may be deployed individually to address a specific requirement. However, due to the volume and varied formats of important documents held in digital format, these systems are often deployed collectively based on a strategic IDCM approach for better managing information assets. An organizational approach to IDCM supports enterprise knowledge strategies by providing the capability to capture, search, and retrieve documented information.

SCOPE

IDCM depends upon effective integration of organizational systems that together are used for managing both digital and physical document types. The scope of this management is across all stages of document lifecycles. It includes provision for distribution of the document content over intranets and the Internet.

Features of enabling IDCM technologies are described in the following section. The technologies may be differentiated into those with core capabilities and supporting technologies.

Core capabilities are: document management; e-mail management; drawing management; document imaging; Web content management; enterprise report management; and workflow. Supporting technologies include: Web services; database management systems; digital signatures; portals; universal interfaces; and network management.

Significant issues that need to be addressed with respect to IDCM solutions include the provision of seamless functionality that may be employed across different capabilities so that currency, integrity, and authority are managed effectively. These in turn must be complemented by user interfaces that provide stylistic consistency and that are augmented by metadata that enhances retrieval capabilities through the supporting technologies.

The following section itemizes the types of features that are required.

SYSTEM FEATURES—CORE TECHNOLOGIES

Document Management

An encompassing approach to document management sees documents within a framework that supports integrity, security, authority, and audit, and that are being managed so that effective descriptions of them are used to support access, presentation, and disposal (Bielawski & Boyle, 1997; Wilkinson et al., 1998). In this context document management applications implement management controls over digital and physical documents. The general capabilities of a document management application are:

- *Document production and capture*—interface with common office productivity software.
- *Classification*—support business classification schemes (e.g., folder structures, document properties).
- *Metadata*—capture of properties that describe document.
- *Check-in/checkout*—maintain document integrity during editing.
- *Version control*—increment versions of document to support integrity.
- *Complex relationships*—manage links and embedded content within digital documents.
- *Security*—implement user/group access permission rights over documents.
- *Document lifecycles*—manage the transition of document states through pre-defined lifecycles.
- *Integrated workflow* to automate review and approvals; controlled distribution of documents.
- *Search and information retrieval*—search metadata or text within documents, or both.
- *Viewing*—view documents in native application or using integrated viewer.

These should be associated with recordkeeping features such as disposal scheduling and archiving.

E-MAIL MANAGEMENT

The growth in e-mail has brought a high demand for solutions that allow enterprises to manage e-mails that have value to the business. The IDCM model offers two types of capabilities:

- *Direct capture*—These applications are often referred to as e-mail archiving applications.
- *End-user capture*—These capabilities are typically offered as a module within document management systems.

Direct capture or archiving facilities intercept incoming and outgoing e-mail. They operate by taking a copy of incoming and outgoing messages that are managed by the e-mail messaging system, and use customized business rules to extract e-mail that may not have a business context. Unwanted e-mail such as spam, or that received from news lists or information bulletins, can be eliminated.

These systems may feature auto-categorization based on metadata such as that contained in e-mail message headers, and possibly also within attachments. Categorization can also occur using the content and context of e-mail by applying techniques such as learning by example from previously processed e-mail. These types of solutions might be valuable for capturing statistical information differentiated by the types of requests made by customers. For example, statistics can aid call centers to monitor turnaround timeframes for responding to e-mail requests, or undertake trend analysis.

Search options include the capabilities to search messages and text attachments. Depending on the capabilities of the system, searches might be invoked from an e-mail client, desktop client application, or Web browser.

Some systems are able to apply rules defined in disposal authorities so that e-mails are purged from the system within a legal framework. In some cases, different retention schedules can be applied to specific categories of e-mail.

End-user capture facilities are adopted by some enterprises to save relevant sent and received e-mails that evidence business transactions into an e-mail management repository, such as a document management system, leaving it up to the user to identify e-mails that need to be saved according to organizational guidelines.

The document management system would need then to integrate effectively with the existing enterprise e-mail client software. This capability would enable end-users to save e-mails and/or attachments to the managed repository, automatically derive metadata from the header of e-mail messages, add custom metadata, and store the e-mail and attachment/s (where appropriate) as a digital record.

DRAWING MANAGEMENT

Many systems for registering or managing drawings have been developed independently of more generic approaches to document management. They may include information systems that enable users to register or index physical drawings in a database, along with generation of transmittals for issue of new documents, and management of the distribution of revisions to drawings and technical documents.

A drawing management system may be differentiated from a registry system in that the software implements automated management controls over the digital drawing objects maintained within a vault-like repository. This capability evolved to support the capture and management of drawings created by Computer Aided Design (CAD) packages. Functionality should include base capabilities such as:

- Integration with CAD tools for capturing electronic drawings.
- Automated features for drawing revision control and revision numbering.
- Management of electronic and hardcopy drawings, technical specifications, and manuals.
- Management of parent-child relationships between multiple drawings.
- Registration and tracking of physical copies of controlled drawings.
- Management of incoming and outgoing transmittals.
- Electronic document review and authorization using integrated workflow.
- Provision of viewing, red line, markup, and annotation functions.
- Maintenance of history logs and audit trails.

Extended capabilities may include automation of drawing numbering, synchronization of digital and physical drawing objects, synchronization of title block and metadata registration and updates, and management of drawing status during engineering change lifecycle transitions.

Drawing management capabilities may be provided by a dedicated drawing management system; as unified functionality within a document management system; as an inbuilt module of an Enterprise Resource Planning (ERP) system, maintenance management system, or similar; or an integral component of a document management application.

DOCUMENT IMAGING

Imaging systems have evolved from the principles of film-based imaging and may now be characterized in two groups for document imaging, these being (a) film-based imaging (micrographics) and (b) digital imaging systems.

In film-based imaging, micrographics technology is used to capture images of physical documents on microfilm, so that the images may subsequently be viewed using a reader, and printed if required. In digital imaging, images of physical documents are captured in a digital file format, with subsequent viewing or printing from the image format.

Digital imaging systems may be differentiated as desktop (ad hoc scanning), workgroup (shared tools in network), or production (high volume, diverse type). IDCM normally implies a workgroup or production environment. Capabilities offered include image manipulation functions such as:

- Capture of hardcopy documents into digital format.
- Capability for scanning and conversion of different sizes, sides (duplex scanning), physical orientation, and physical structure of documents.
- Managing multi-page images as a single entity (e.g., multi-page TIF file).
- Images may be saved to specified file formats. For example, a document might be saved in PDF or JPG for publishing on a Web server, or as a multi-page TIF for viewing/transmission.
- Support for a range of resolution, contrast, threshold, and size settings to meet the diverse requirements of document capture.
- Capture of color and/or grayscale images to suit forms processing and other applications (e.g., colored contour maps).
- Despeckling/deskewing and border removal.
- Multi-level registration capabilities, including batch-, folder-, envelope-, and document-level indexing.

Imaging systems are often integrated with recognition systems to facilitate capture and retrieval. These include technologies for automatically capturing data encoded in barcodes, integration with optical character recognition (OCR and ICR) technologies to enable text information to be extracted from scanned images, and integration with optical mark recognition (OMR).

WEB CONTENT MANAGEMENT

IDCM has the capability to provide a managed environment for the processes associated with publishing Web content. It has been said that a content management system is a concept rather than a product (Browning & Lowndes, 2001). This adds weight to analysis of it within the context of an IDCM model, where documents and their content may be considered more broadly than in terms of Web presentation. Document creation, management, and utilization can thus be undertaken with reference to business requirements and workflow of business processes.

Typically, functionality is characterized in terms of content creation, presentation, and management (Arnold, 2003; Robertson, 2003). Increasingly this functionality is seen to be employed within a unified content strategy for an enterprise (Rockley, Kostur & Manning, 2003).

Content creation functionality includes separation of presentation and content, utilization of elements of documents such as illustrations in different contexts, and continuation of associations between pages after restructuring. Metadata support should also be available, and markup should be transparent to the content creator.

Presentation elements include multiple formats for distribution of internal material such as manuals and business forms over intranets, and for external material such as marketing information and application forms. Other features expected are template availability through style sheets, integration of multiple formats as compound documents, provision of alternative renditions, and personalization of display according to user profiles.

Management features include version control and integrity maintenance among multiple users, and associated security procedures and audit trails. Managed interfaces to other subsystems should provide for dynamic provision of content to pages so that current data can be presented in validated form within compound documents. There should also be utilization of workflow for accommodating distributed users, content review, and approval processes.

These capabilities are shared at least in part with other IDCM systems' functionality. The IDCM environment has the capability to manage Web content within a continuum that includes initial document creation processes, potentially in a distributed environment, through to managed archiving of content.

ENTERPRISE REPORT MANAGEMENT (COLD)

As digital media have been developed, businesses with high volumes of management information reporting have made increasing use of Enterprise Report Management (ERM). These capabilities enable organizations to capture reports from business application databases and store them in a managed repository, to reduce printing, improve information accessibility, and maintain records.

Technologies that provide support for ERM include output reports in a range of formats. Examples of these include text-based digital format (e.g., XML) that is stored and searched via a document management application, and image format such as TIF or JPG that can be captured and accessed via a document management or imaging application. Reports may be captured on optical disk, using a capability known as Computer Output Laser Disk (COLD), which stores digital reports and enables data to be represented with graphical overlays to facilitate interactive communications.

General ERM capabilities are defined as follows:

- Capture of digital report objects to managed repository (document management, imaging, or COLD application).
- Utilization of indexing capabilities for capturing metadata relevant to the report.
- Support for inquiry and retrieval of metadata or report contents (where applicable).
- Management of database growth to support performance.
- Support for repository that can include different data objects.
- Support for document integrity.
- Control of processes through workflow.
- Provision of extraction and use of parts of reports.
- Support for high-volume printing.
- Management of security—user authentication, group and user levels.

WORKFLOW

Workflow management systems are designed to automate and implement controls over a diverse range of business processes, from the initiation of a process through to execution of all tasks, and process closure. The need for transparent interfaces between the workflow management system and IDCM is vital to maintain the integrity of documents or Web content files during their transition through a workflow process.

There are a number of technology options for enterprises that are seeking a workflow management capability. The most suitable workflow engine will depend on the nature and complexity of the requirement and the functionality supported by the workflow technology options. Options for workflow within the context of IDCM are:

- *Messaging/collaboration systems/workflow:* IDCM should support ad hoc and cooperative review and production of documents and reports, and it may be desirable to support integration with electronic forms for recordkeeping purposes.
- *Embedded workflow:* This capability is offered in systems such as document and Web content applications, or in application suites such as ERP systems. The host application provides inbuilt workflow for facilitating document-centric or process-centric modules.
- *Autonomous workflow:* These types are functional without any additional application software, with

the exception of database and message queuing. When used in the context of IDCM, the functionality may support automation of document or Web content review and approval processes.

IDCM ARCHITECTURE

The IDCM system model should feature scalable, flexible, and extensible applications that integrate with the enterprise information architecture, enabling the system to grow with the organization and facilitate knowledge sharing. Some architectural scenarios include:

- Scalability, flexibility, extensibility.
- Intuitive interface, preferably Web based, to facilitate usability, software upgrades, and support.
- Integrate with heterogeneous operating environments (where required).
- Implement three-tier (or "n-tier") client server architecture to facilitate Web client functionality and usability.
- Support distributed computing environment (databases, document/Web content repositories, replication services).
- Support mobile workers—access from remote sites, limited bandwidth.
- Integration with enterprise backup/recovery and business continuity regimes.

REASONS FOR UTILIZING IDCM SOLUTIONS

The business justification for implementing IDCM solutions ranges widely with respect to policy, compliance, and economics. For example, policy initiatives may include support for customer service initiatives, knowledge management (Laugero & Globe 2002), or risk reduction in relation to brand damage. Compliance with legislative requirements may include administrative requirements that support privacy and freedom of information legislation. Economic justification may include support for timely delivery of product to market, reduction in operational costs, continuous process improvement initiatives, and profit maximization strategies.

CRITICAL ISSUES OF DOCUMENT AND CONTENT TECHNOLOGIES

There are critical issues that must be managed, and it is imperative that organizations undertake risk analysis in order to identify risks and develop strategies to mitigate them. Table 1 summarizes some of the critical issues.

Table 1. A summary of critical issues of IDCM technologies

Business Issues	Technology Issues
Executive management lacks resolve Lack of executive management engagement, both at the start and during the project, may impair outcomes.	**Inadequate infrastructure** Client, server, and network architecture needs to be adequate to optimize performance.
Inadequate planning Poor definition of scope and inadequate product and project lifecycle management may impair outcomes.	**Incorrect technology application** Inadequate business definition and failure to examine solution options results in implementation of inappropriate technology solution.
Inadequate specifications Lack of analysis and determination of requirements leads to project complications and inhibits extensibility of applications across enterprise.	**Integrity of metadata** Metadata can be abused in non-validated fields, which may create significant retrievability issues.
Mandated use may cause rejection Document management is often mandated, without appropriate consultation and analysis.	**Security** IDCM solutions contain vital documents and content files, and security should reflect importance.
Lack of process integration Mandated use is often accompanied by failure to integrate capabilities with existing processes, often meaning duplication of effort.	**System incompatibilities** Lack of proven integration capabilities may impact delivery document/content management enabled end-to-end business solutions.

CONCLUSION

The IDCM model supports a business value proposition that aligns enabling systems and technology with an enterprise's strategic, tactical, and operational planning imperatives. The IDCM system architecture provides a range of enabling applications and technologies that support end-to-end business process improvement initiatives and provide a key foundation for knowledge management strategies.

REFERENCES

Addey, D. et al. (2002). *Content management systems*. Birmingham, UK: glasshaus.

Arnold, S.E. (2003). Content management's new realities. *Online, 27*(1), 36-40.

Asprey, L. & Middleton, M. (2003). *Integrative document and content management: Strategies for exploiting enterprise knowledge*. Hershey, PA: Idea Group Publishing.

Bielawski, L. & Boyle, J. (1997). *Electronic document management systems: A user-centered approach for creating, distributing and managing online publications*. Upper Saddle River, NJ: Prentice-Hall.

Boiko, B. (2002). *Content management bible*. New York: Hungry Minds.

Browning, P. & Lowndes, M. (2001). JISC TechWatch report: Content management systems. Retrieved October 10, 2003, from *www.jisc.ac.uk/uploaded_documents/tsw_01-02.pdf*.

D'Alleyrand, M.R. (1989). *Image storage and retrieval systems: A new approach to records management*. New York: Intertext Publications.

Hackos, J. (2002). *Content management for dynamic Web delivery*. New York: John Wiley & Sons.

Laugero, G. & Globe, A. (2002). *Enterprise content services: A practical approach to connecting content management to business strategy*. Boston: Addison-Wesley.

Megill, K.A. & Schantz, H.F. (1999). *Document management: New technologies for the information services manager*. East Grinstead, UK: Bowker-Saur

Nakano, R. (Ed.). (2002). *Web content management: A collaborative approach*. Boston: Addison-Wesley.

Ricks, B.R., Swafford, A.J. & Gow, K.F. (1992). *Information and image management: A records systems approach* (3rd ed.). Cincinnati, OH: South-Western Publishing.

Robertson, J. (2003). So, what is a content management system? Retrieved March 18, 2004, from *www.steptwo.com.au/papers/kmc_what/index.html*.

Rockley, A., Kostur, P. & Manning, S. (2003). *Managing enterprise content: A unified content strategy*. Indianapolis: New Riders.

Wiggins, B. (2000). *Effective document management: Unlocking corporate knowledge*. Aldershot, UK: Gower.

Wilkinson, R. et al. (1998). *Document computing: Technologies for managing electronic document collections*. Boston: Kluwer Academic Publishers.

KEY TERMS

Content Management: Implementation of a managed repository for digital assets such as documents, fragments of documents, images, and multimedia that are published to intranet and Internet WWW sites.

Document Capture: Registration of an object into a document, image, or content repository.

Document Imaging: Scanning and conversion of hardcopy documents to either analogue (film) or digital image format.

Document Management: Implements repository management controls over digital documents via integration with standard desktop authoring tools (word processing, spreadsheets, and other tools) and document library functionality. Registers and tracks physical documents.

Drawing Management System: Implements repository management controls over digital drawings by integration with CAD authoring tools and using document library functionality. Registers and tracks physical drawings.

E-Mail Management: Implements management controls over e-mail and attachments. These controls may be implemented by direct capture (e-mail archiving software) or invoked by the end-user in a document management application.

Recognition Technologies: Technologies such as barcode recognition, optical character recognition (OCR), intelligent character recognition (ICR), and optical mark recognition (OMR) that facilitate document registration and retrieval.

Workflow Software: Tools that deal with the automation of business processes in a managed environment.

Intelligent Agents for Competitive Advantage

Mahesh S. Raisinghani
University of Dallas, USA

John H. Nugent
University of Dallas, USA

INTRODUCTION TO MEASUREMENT AND REPORTING SYSTEMS

Since 1494 with the appearance of the double-entry accounting system, developed by Pacioli, those involved in business have attempted to measure business performance in an organized manner. As many accounting functions are repetitive in nature (payroll, inventory, etc.), accounting was one of the first business disciplines to which early computing technology was applied. Today we see comprehensive enterprise models that have been incorporated into ISO Standards in an attempt to build quality, capability, and uniformity into business enterprise systems. Supporting these models and systems is an effort to also launch the Extensible Business Reporting Language (XBRL), such that metadata models gain uniformity and make business information more readily accessible across systems and enterprises.

This article addresses a concept for transitioning from an "end of period" reporting model to one based on "push agents" delivering to the pertinent manager information that is key to managing the enterprise in near real time in order to gain substantive competitive advantage. The high-level model in Figure 1 demonstrates the suggested movement from an "end of period" model to an automated push agent model, with an intermediate step already utilized in some enterprises, that is a "dynamic query model."

Comparing where we are today in business reporting to that of network reporting, Computer Associates, perhaps the first to offer comprehensive, automated agents that forewarn of impending network trouble introduced some time ago neural agents that measure current states as they change against historical databases of past network activity in order to discern conditions that may be reoccurring and are similar to those that caused problems previously. Such detections may involve likely equipment failure to circuits becoming overloaded, with automated warnings and recommendations as to corrective actions being sent to the appropriate manager via the chosen message system (e-mail, voicemail, paging, etc).

Such "heads-up," automated, near-real-time reporting of impending conditions permits network operators to be proactive in addressing problems before they occur. Such real-time network feedback requires an enterprise to develop "key performance indicators" (KPIs), or key measurements that it wishes to track against to plan and run the enterprise. Here, many products are on the market that address the identification of such KPIs and go by the names of "Cockpit Charts," "Digital Dashboards," and "Balanced Scorecards." The secret of the balanced scorecard and the reason it has gained such wide acceptance is primarily due to the fact that it allows organizations to reach their full potential by putting strategy—the key driver of results today—at the center of the management process in organizations facing uncertain equity markets, an accelerating pace of change, and increased expectations for productivity and results. The comparative characteristics of reporting models are listed in Table 1.

Instead of waiting for end-of-period reports, critical measures can be monitored in near real time as a function of the KPIs assigned to each measure, as seen in Figure 2.

Figure 1. Migration of reporting models for competitive advantage

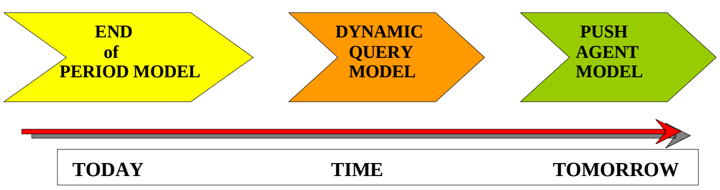
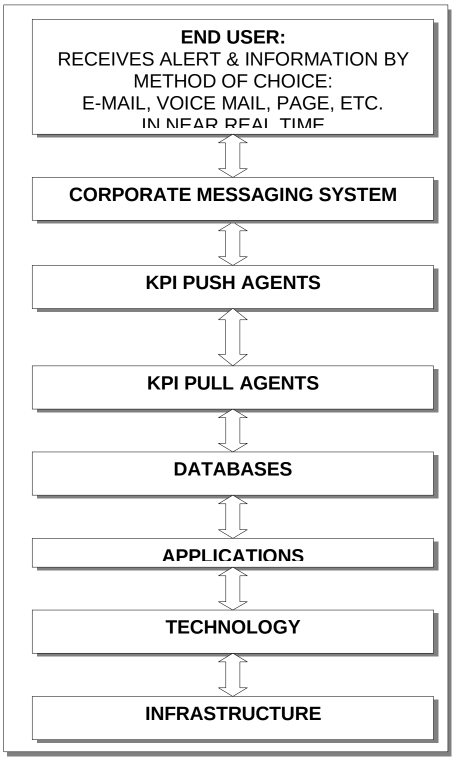


Table 1. Comparative characteristics of reporting models

Type of Reporting	End of Period	Dynamic Query	Push/Pull Agents		Pull/Push Agents
Timeliness of Data	Stale	Near Real Time	Near Real Time		Near Real Time
Potential for Managerial Filtering	Yes	No	No		No
Delivery Choice	No	Not Typically	Yes		Yes
Alerts	No	Not typically	Yes		Yes
Silo Issues	No	Possibly	No		No

Figure 2. Simplified intelligent agent push model



BENEFITS OF EMPLOYING A BUSINESS MANAGEMENT PROCESS INTELLIGENT AGENT TOPOLOGY

Erik Thomsen, Distinguished Scientist at Hyperion Solutions Corporation, defines the term "agent" as:

"...a solution-oriented ensemble of capabilities including natural language processing, autonomous reasoning, proactive computing, discourse modeling, knowledge representation, action-oriented semantics, multimodal interaction, environmental awareness, self awareness, and distributed architectures."

He describes the following five areas where the potential impact of intelligent agents on the logical functionality and physical performance of traditional business analytic systems can be positive (Thomsen, 2002):

- Agents should help move business intelligence (BI) from being application centric to being truly process centric, and provide a single point-of-access to distributed information. This is operationalized in intelligent agent solutions by self-description of individual modules that is made available to the system of agents as a whole, and a user's personal agent being able to query all librarian agents responsible for various data sets in an organization for specific information.
- An active dialog is needed between the software and the user to seek out and learn user wants, and be able to anticipate/predict user wants in the future. Thus in addition to tailoring client layers to individual users and enhancing BI applications with options and preferences, the software/intelligent software agent plays an active role in querying the user.
- As attention shifts to business process as much as data states, the number-centric BI applications will have to provide more integrated text and multimedia handling.
- Intelligent software agents can provide personal analytic "coaching" for higher-level business processes by observing, learning about, and interacting with users. Agent applications within an overall BI/Business Process Management (BPM) framework can be deployed to encode horizontal and domain-specific analytic knowledge.

- The server side of most BI applications is very complex due to the changes in the load patterns/ factors and the wide range of physical tuning options. Intelligent physical optimizers can evaluate their own physical organization and, optionally, interact with an administrator before performing any reorganization.

A survey by Lederer, Mirchandani, and Sims (1998) indicates that businesses adopting Web-based information systems (WISs) responded that the most important benefit of being on the Web was to "enhance competitiveness or create strategic advantage." Intelligent online agents can help simplify complex Web environments acting as artificial secretaries (Maes, 1994). Nwana et al. (1998) define agents as "software entities that have been given sufficient autonomy and intelligence to enable them to carry out specified tasks with little or no human supervision." There is no standard definition for intelligent agents, but they are generally described (Decker, Pannu, Sycara & Williamson, 1997; Specter, 1997; Hendler, 1996) as programs that act on behalf of their human users to perform laborious and routine tasks such as locating and accessing necessary information, resolving inconsistencies in the retrieved information, filtering away irrelevant and unwanted information, and integrating information from heterogeneous information sources.

Business intelligence software using intelligent agents and BI technologies such as data marts/warehouses, data mining, Web site analytics, modeling and predictive tools, and data visualization tools can help analyze customer data to make connections that can boost sales, increase efficiency, and retain customers. The BI server connects to the data sources and provides analytical services to clients, which access the BI server through a Web interface or other desktop client. An application server typically provides the Web interface and often runs on the same server as the BI software.

The key benefits of employing a business management process intelligent agent topology are as follows:

- Timeliness of reports
- Automated measurements against planned goals
- Unfiltered information
- Delivery based on the end-user's preference

The key features that should be analyzed when evaluating BI reporting tools are as follows:

- Analysis: The ease in creating reports and the end-user's ability to manipulate the report once distributed.
- Formats and data sources supported.
- Data access: Preferably provide data-attribute-level security.
- Price: This can be compared based on a given number of users, on a specific hardware configuration.

To enable enterprises to better track and manage cross-departmental performance, BI vendors are rolling out frameworks designed to help enterprises integrate and leverage multiple existing BI systems and analytic capabilities. Theoretically, this enables a powerful enterprise-wide management tool for optimizing performance and profits. However, in practice, there are challenges associated with these frameworks that range from data acquisition, cleansing, and metadata management to aligning models and delivering performance management.

POTENTIAL DRAWBACKS OF EMPLOYING A BUSINESS MANAGEMENT PROCESS INTELLIGENT AGENT TOPOLOGY

The pressures on the management team are intense. Executives often struggle to integrate multiple data sources that can then be analyzed and synthesized for financial reporting. For example, inventory data may be stored in the data warehouse while sales information is housed elsewhere. In addition, executives may want to close their books as often as daily, without having to wait until the end of the month to address a problem that could affect revenue or earnings forecasts.

Business intelligence tools are best suited for allowing executives to drill down into the source of financial reports to review their accuracy. For example, if an employee forgets to book an order in an order management system, but makes the adjustment in the general ledger, a BI tool could start investigating further; if anomalies are found, it is critical to drill into the support data to determine the reason. However, a majority of BI systems would be hard-pressed to catch someone intent on committing fraud if the data was entered correctly in the data capture system.

Enterprise application vendors also are adding analytics to back-end financial systems to help companies better leverage raw transactional data. For example, PeopleSoft has embedded analytics in its financial module designed to report to executives on a daily basis the status of KPIs. Executives have little time to analyze data before senior management inquiries begin, and they need visibility across the internal and extended supply chain to have the control in their organization. With a broader audience, the need for more detailed intelligence will drive analytics that are targeted not only at specific vertical industries, but at specific departments and roles. However the dangers of creating more analytic silos cannot be overemphasized.

It is prudent to remember that pressure to act may result in erroneous actions. Business survival depends on an infrastructure that can adapt to changing market conditions. With companies focused on leveraging existing resources and increasing efficiency, infrastructure is no longer just an operational cost of doing business. Solution providers are paying attention to this growing need for intelligent infrastructure, with companies such as Network Appliance Inc. and Web Methods Inc. partnering with BI vendors to bring new abilities—such as storage analytics and business activity monitoring—to their solutions which can help companies realize the full potential of their resources.

CONCLUSION

This article conceptualizes the transition of intelligent agents utilized in network performance management into the field of business and management. A tiered IA system could be implemented at many levels of management and could be the key for successful, timely knowledge management strategies and successes. Such a system would be timely, unbiased, objective, and should provide significant competitive advantages. Moreover, it could leverage existing assets and provide a single objective measure of employee performance at appraisal time.

Although there has been some progress in increasing the number of users in the area of query and reporting, a lot more progress needs to be made between closing the loop between decision making and operations. IA benefits realized in telecommunications networks, grid computing, and data visualization for exploratory analysis connected to simulations should likewise be achievable in business management processes.

REFERENCES

Burriesci, J., Kestelyn, J. & Young, M.M. (2002). The top 10 trends for 2003. Retrieved December 17, 2002, from *www.intelligententerprise.com/030101/602news1.shtml.*

Computer Associates. (2000). Retrieved June 20, 2002, from *www.cai.com/clientspotlights/neugents.htm.*

Decker, K., Pannu, A., Sycara, K. & Williamson, M. (1997). Designing behaviors for information agents. *Proceedings of Autonomous Agents "97* (pp. 404-412).

Hendler, J. (1996). Intelligent agents: Where AI meets information technology. *IEEE Expert, 11*(6), 20-23.

Lederer, A.L., Mirchandandi, D.A. & Sims, K. (1998). Using WISs to enhance competitiveness. *Communications of the ACM*, *41*(7), 94-95.

Maes, P. (1994). Agents that reduce work and information overload. *Communications of the ACM, 37*(7), 30-40.

Nwana, H., Rosenschein, J., Sandholm, T., Sierra, C., Maes, P. & Guttmann, R. (1998). Agent-mediated electronic commerce: Issues, challenges and some viewpoints. *Proceedings of the 2nd International Conference on Autonomous Agents* (pp. 9-13, 189-196).

Spector, L. (1997). Automatic generation of intelligent agent programs. *IEEE Expert*, *12*(1), 3-4.

Thomsen, E. (2002). Agents uncovered. *Intelligent Enterprise, 5*(15), 45.

KEY TERMS

Data Mart: Database containing data extracted and often summarized from one or more operational systems or from a data warehouse and optimized to support the business analysis needs of a particular unit.

Data Mining: A component of business intelligence decision-support process in which patterns of information in data are discovered through the use of a smart program that automatically searches the database, finds significant patterns and correlations through the use of statistical algorithms, and infers rules from them.

Data Warehouse: A form of data storage geared towards business intelligence. It integrates data from various parts of the company. The data in a data warehouse is read-only and tends to include historical as well as current data so that users can perform trend analysis.

Enterprise Deployment: Term used in the computer industry to describe hardware and software configurations. They are aimed to address the corporation as a whole as opposed to a single department.

ERP: Enterprise resource planning system. Enables the company to integrate data used throughout the organizations in functions such as finance, operations, human resources, and sales. This system extends the pool of information for business intelligence.

Knowledge Management: Application for any business to create, maintain, and share company knowledge. The challenge of capturing collective experience, core values, and expertise of an organization.

Metadata: The information a database or application stores to describe the university's business data and the applications that support it. Refers to the information used to describe a set of data.

Portals: Means entry point. Describes the entry point for users to information available throughout the World Wide Web.

Set-Based Analysis: A recent method that uses groups of sets. It facilitates the examination and comparison of data.

Slice and Dice: Another term for multidimensional analysis. When data has three (or more) dimensions, it can be thought of as being arranged in a cube (or hypercube), with each side representing a dimension. When the data is analyzed, part of the cube can be "sliced" off or "diced" to get to an individual cell.

Intelligent Business Portals

Xue Li
University of Queensland, Australia

INTRODUCTION

Business Portals are now widely used in e-commerce. Intuitively Portals can be regarded as an information gateway for exchanging business information over the Internet. The process of business information exchange requires the services of *security and access control*, *interoperability* over a variety of software of platforms and systems, *interconnectivity* for communications among different data communication networks, and *scalability* for information being sharable on different hardware devices. So, portals can also be regarded as an information carrier for delivering the right information to the right user, at the right time, to the right place, to make the right decisions. A portal is a packaged piece of information with the properties of self-servicing, personalization, and real-time delivery. From a business point of view, a portal is a mobile, self-explanatory, and just-in-time delivered piece of information. In e-commerce, business information is a set of time-stamped transactions that can be triggered by events in business activities. This article will illustrate and explain the architecture of intelligent business portals for Web-enabled business applications.

BACKGROUND

The portal concept was introduced by Merrill Lynch (Shilakes & Tylman 1998). It was estimated that by the year 2002, the portal market value in Business Intelligence will reach US$7 billion(Bergert 2000), in Content Management will reach US$4.5 billion, and in Data Warehousing and Mart will reach US$2.5 billion. This fast growth has been attributed to the World Wide Web on the Internet and the high-speed network infrastructure.

The origination of portals comes from a question on how we could deliver the right information to users. In traditional *pull technology*, individual Web users have to initiate the search operation to find information on the Web, while in *push technology*, information is sent to individual users with or without solicitation. On the other hand, a user may need to interact with a system to provide information (such as filling out a survey form). Thus, there is a need to provide an application-independent mechanism for switching information between information providers and requestors. This requirement has generalized the traditional many-to-one client-server relation into a many-to-many client-server relation. Inevitably, an information explosion is introduced on the Web. Many Web users are frustrated in dealing with the overwhelming information bombardment. To solve the problem, the portal as a packaged piece of information is used for delivering the right information to the right user.

PROPERTIES OF PORTALS

A portal has the properties of self-servicing, personalization, and real-time delivery. For *self-servicing*, a user would be able to use predefined templates to redesign personalized Web pages. For example, an employee in a business organization can be assigned by a predefined Web site according to the business roles played by the individual. The employee can then modify the Web site according to personal interests and preferences. This Web site then is used to conduct the business.

For *personalization*, a user would be able to deliver and receive information that is dedicated to the person. For example, portals are used by university students to select subjects and view personal enrollment information. For *real-time delivery*, a portal is used as a messaging tool to deliver instant messages to an individual.

From a business point of view, a portal provides mobile, self-explanatory, and just-in-time delivered information. Business intelligence is about how to apply knowledge to making business decisions, searching for useful information, or controlling the business processes.

When portals are used as an approach to the fundamental information infrastructure of e-businesses, we need to know how to maximize their usefulness in order to improve our business performance, competitiveness, and viability. Intelligent portals would be portals with knowledge in order to be driven around on an intranet or on the World Wide Web. In other words, by capturing the domain-specific business knowledge, we can deploy portals on the Web and let their behavior be controlled by a knowledge base. In this case, a knowledge management system (Choo,1998; Liebowitz & Wilcox, 1997) would play an important role in an e-business environment.

This article discusses the theoretical issues on the integration of knowledge management systems with portal deployment mechanisms. We will illustrate and explain the technological issues in designing and implementing intelligent business portals for Web-enabled business applications.

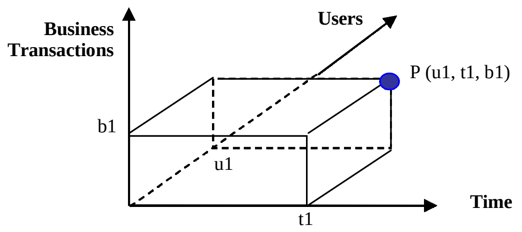
INTELLIGENT PORTALS

Intelligence is the execution of knowledge. When a reasoning mechanism is invoked by a question or a problem, the relevant knowledge is retrieved and a possible answer or solution is then concluded. Intelligence also implies the capability of finding the best solution for a problem. In a business environment, knowledge can be in different forms. It could be a set of if-then production rules for decision-making problems, a set of facts for corporate infrastructure descriptions, or a set of procedural descriptions for the business transactions. Business activities are event driven, so the timely execution of certain business processes is crucial to the success of business. This research area is mainly covered by workflow management (Marin, 2001; WfMC, 1996).

A portal in this context is an information feeder that will satisfy the information needs of different users at different times for different business processes. In Figure 1, it can be seen that a point *P* in the 3-D space is the information about *who is doing what at what time*. Since the business activities can be the predefined workflows (like an application for an insurance policy will go through a step-by-step process to get an approval), a workflow control system should be able to check the information requirements for a particular business process and a particular person. For example, in Figure 1, point *P* may be interpreted as a person "John" on "Monday, 3 Nov, 2003" at "9am" is "placing an order to buy a product".

In an e-business, if a Web site is designed for online ordering, while a user is online, the system should be able to deliver context sensitive information to the order form (e.g., user account number, best sales, etc). Furthermore, a reasoning mechanism may be triggered for context-sensitive reasoning and decision making. In this case, an intelligent portal is a context-sensitive information/service supplier that will accompany the user through the lifetime of the transaction. The 3-D space of the workflow control illustrated in Figure 1 shows the demands of the intelligent portals.

Figure 1. Three-dimensional space for the workflow control



PORTALS VS. AGENTS

Here we need to differentiate intelligent portals from intelligent agents. Intelligent agents (Knoblock & Ambite, 1997) are mobile software programs and are task oriented, while intelligent portals are information carriers and are content based. By using agents, we get things done; by using portals, we know what happened and what information should be supplied. One of the advantages of intelligent portals over intelligent agents concerns the acceptance by general users: agents are the programs to be executed on the clients' machines, while intelligent portals are the pieces of information driven by knowledge to deliver Web services. Consequently, there is no fear from users about potential virus attacks.

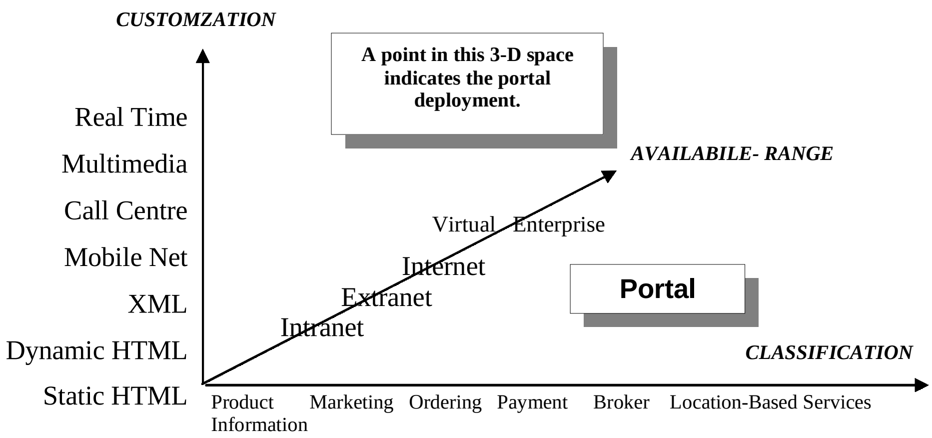
KNOWLEDGE MANAGEMENT WITH PORTALS

In a business environment, knowledge management has many aspects, from low-level, day-to-day business process control to high-level, executive decision making. A knowledge management system should be able to collect relevant knowledge, store knowledge in a sharable enterprise memory, communicate the knowledge with parties, and maintain consistencies. In all these activities, a portal can play an important role within an enterprise, that is, as an information carrier to shift information around the organization.

One important task relating to portals in knowledge management is Workflow Management (Allen, 2001; WfMC, 1996). Workflow management involves:

- *Specification* of process control for business transactions, which concerns data coordination, exception handling, recovery, etc. The workflow specifications provide execution plans.
- *Verification* of the feasibility and correctness of a design, while allowing for re-design and implementations for coping with changes.
- *Execution control* for carrying out business transactions. A Workflow Engine is responsible for execution of the processes. During any given execution, a workflow plan may be applied to many individual users. As a result, many concurrent workflow

Figure 2. Deployment of intelligent business portals



instances will be generated. The Workflow Engine is responsible for concurrency control (process dependencies), exception handling, and recovery.

In Workflow Management, portals can be used as basic constructs to build workflow tasks (Marin, 2001). For example, in processing a procurement request (e.g., purchasing a laptop computer), one may need information about competing offers from different suppliers. In this context, a portal should play a role not unlike a real-life broker who can assess the buyer's requests and evaluate the seller's offers in order to make a deal.

Portals Deployment

Knowledge about portals is presented from three aspects: ontological services, applicability services, and directory services. These three services answer the questions of what portals are, where they are used, and for what they are used. Ontological services are facilities that categorize business information into a searchable, indexed structure with the meaning of the data (e.g., Classification dimension in Figure 2). Availability services provide maps of portals over the Internet for their availability (e.g., Available-Range dimension in Figure 2). Directory services are a view of the information tailored for personalized service (e.g., Customization dimension in Figure 2, as well as location-sensitive services such as those portals for the mobile phone and PDA users. These three aspects are orthogonal in the sense that they make up the key information required for providing business data over the Internet. Figure 2 shows the idea of the deployment of the intelligent portals.

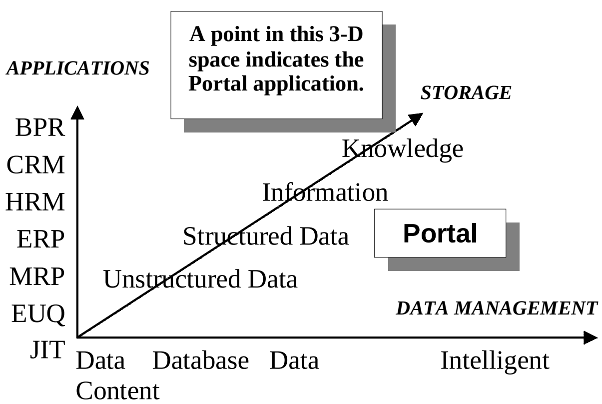
It should be pointed out that the security services and access control for the intelligent portals are embedded in all dimensions of the above discussions.

Portals Development

In Figure 3, we show the intelligent portals in a three-dimensional paradigm for representation purposes. In the Storage dimension, the business data are packaged as structured data, unstructured data, or knowledge in a ready-to-use format. In the Data Management dimension, the business data are the objects managed within databases, data warehouses, intelligent applications, and content-management systems. In the Applications dimension, the business data are the objects in Business Process Restructuring (BPR), Customer Resource Management (CRM), Enterprise Resource Planning (ERP), Material Resource Planning (MRP), or Just-in-time Delivery (JID), etc. applications.

By identifying these three dimensions of business portals, we will be able to design and manage portals in a way that they are "intelligent" – i.e., the portals as

Figure 3. Development of intelligent business portals



packaged business data will be "living" in a business transaction management environment and will be "clever" enough to deliver the business data on the Internet at the right time, to the right places.

In order to use intelligent portals, we need to design the business workflow so that an event-driven workflow engine can be integrated with a portal delivery system.

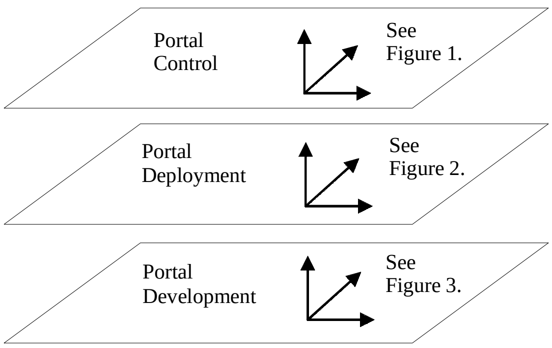
A LAYERED ARCHITECTURE

We identify that intelligent portals are organized in three layers (see Figure 4). From the top, the Portal Control Layer is used to describe the Portal's application logic, which concerns how and when Portals are used in e-Business applications. Figure 1 shows this aspect in a 3-D space. The Portal Control Layer provides information to applications. The middle layer is the Portal Deployment Layer. The Portal Development Layer concerns how and where Portals are deployed. Figure 2 details a 3-D space for portal deployment. At the bottom layer, we discuss portal development. The portal is conceptualized in a 3-D space in order to package data into highly structured and managed pieces of information.

KEY COMPONENT: INFORMATION BROKER

The intelligence of portals relies on the performance of a kernel system component called the Information Broker (see Figure 5). It is responsible for pushing portals to users at the right time. The Information Broker is also capable of searching information for user needs. It is an executor of business workflows that is context sensitive to user requests for information. Its input consists of the User Profile (UP) and the stored information of the Business Workflow (BW). By using UP, the system knows the requirements of individual users. By using BW, the system knows how and when to deliver what information. Figure 5 shows the architecture of a Web site that uses the Information Broker as a key component of portals.

Figure 4. Three layers of intelligent portals



BUILDING INTELLIGENT PORTALS

In order to build intelligent portals, we need to select the right tools to put them together. However, there is currently no consensus on a standard for portals or intelligent portals. In the following discussion, we provide some suggestions from currently available tools.

Figure 6 shows the architecture of the Information Broker – the key component that needs to be built. It is an integrated system that carries out information brokering. The Workflow Engine is a sub-system with knowledge on how portals are used. The Search Engine is a sub-system that knows where to find and collect information, while Portal Management is a sub-system that knows what information is available. The implementation of this component relies on the proper usage of the Web services that can support the distributed information management and the information subscriber/provider architectures.

Portals can be implemented in XML. Many tools provided by major companies (e.g., Microsoft, Oracle, IBM, etc.) are now supporting XML (Finkelstein & Aiken, 2000) used for portals. As that of SQL can be used to represent inference rules, when using XML to represent knowledge, the domain structural knowledge must be specified in a standard Document Type Definition (DTD) as a set of rules to validate the portals. There are many specialized XML in business applications (e.g., FXML, see http://www.finxml.org.). Also, OML/CKML (Kent, 1999) provides a conceptual knowledge framework to specify the

Figure 5. Architecture of a Web site that uses intelligent portals

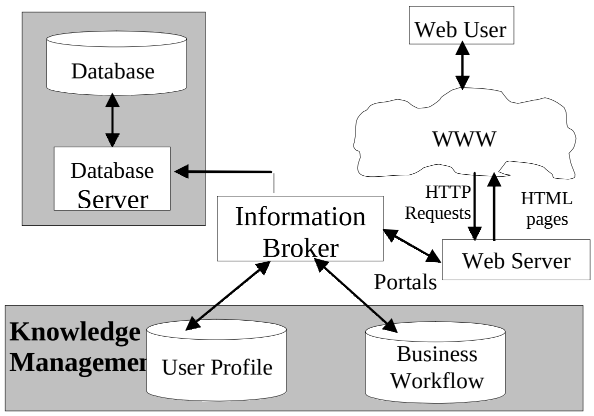
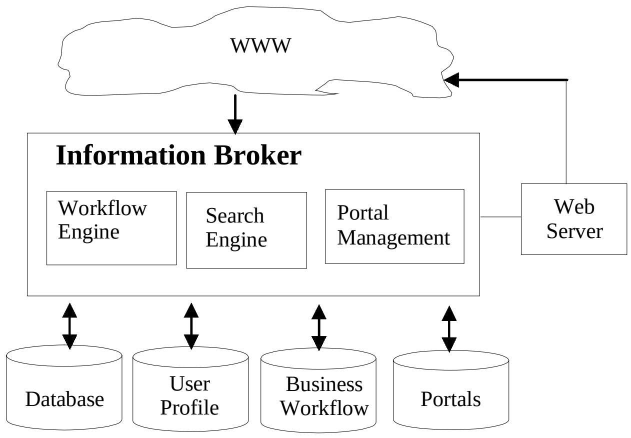


Figure 6. Architecture of information broker



ontological and schematic structures and domain knowledge of data analysis, information flow, and concept lattice (Sowa, 1999). Portals as a content-based approach is becoming a major application of XML.

FUTURE TRENDS

More and more service or content providers are now changing from free services to paid subscriptions. This change requires business portals to play a main role in business applications. As portals are capable of packaging information for its intelligent delivery, the future of this technology can be characterized as follows:

- **Personalized services.** Portals will be more and more used as a vehicle carrying around timely information to users on the World Wide Web. Personalization is a way that different users would be served with the special needs.
- **Self-explainable information services.** With the maturity of Web-enabled business ontology, portals are to be used as a semantic tool to make effective communications among different applications.
- **Uniform interactivities.** XML has provided a language to specify contents, while portals are used to answer the question on how contents are delivered. By using portals, the first class citizen of the WWW would be those well-organized and purposely packaged pieces of information communicating through the protocols such as UDDI, SOAP, and WSDL.
- **Harmonizing business transactions and workflow.** With a large amount of information transmitting on the Internet is now a part of business transactions in situations such as B2B, B2G, etc., portals will be regarded as intelligent because of their ability to be transactional.

The architecture of portals, as illustrated in Figures 5 and 6, will evolve from traditional client-server of many-to-one service pattern to the flexible services patterns such as P2P or many-to-many client-server patterns.

CONCLUSION

Currently the contents on the World Wide Web are evolving from passive hypertext documents to active distributed services (Li, 2000; Zirpins, Weinreich, Bartelt, & Lamersdorf, 2001). By providing self-servicing, personalization, and real-time delivery, portals can actively deliver information to Web users. In this article, we have discussed the following problems:

- What are intelligent business portals?
- What potential do they provide?
- What is the architecture of intelligent portals?
- What are the technical issues in implementing them?

In order to implement intelligent portals, we have discussed a three-layer architecture that reflects the usage of the modern information technology infrastructure. At the development layer, portals are packaged according to the needs. At the deployment layer, portals are allocated to their applications. At the top control layer, portals become knowledgeable and know how, where, and when to deliver their services. The key component in our architecture is the Information Broker who is responsible for implementing the three-layer intelligent portal architecture.

REFERENCES

Allen, R. (2001). Workflow: An introduction. In L. Fischer (ed.), *Workflow Handbook 2001*, p.15-38. WfMC (Workflow Management Coalition), Lighthouse Point, FL.: Future Strategies Inc.

Bergert, S. (2000). Power to the people. *Intelligent Enterprise*, 3(18): 47-51.

Choo, C.W. (1998). *The knowing organization: How organizations use information to construct meaning, create knowledge and make decisions.* Oxford, UK: Oxford University Press.

Finkelstein, C. & Aiken, P. (2000). *Building corporate portals with XML*. New York: McGraw-Hill.

Kent, R.E. (2000). Conceptual knowledge markup language: An introduction. In F. Daoud (ed.), *Netnomics: Economic research and electronic networking, Special issue on information and communication middleware, 2(2), (pp. 139-169).* Kluwer Online, Baltzer Science Publishers, BV. Found at *http://www.kluwer online.com/issn/1385-9587/contents*

Knoblock, C.A. & Ambite, J.L. (1997). Agents for information gathering In J. Bradshaw (ed.), *Software Agents*. pp.1-27. Menlo Park, CA: AAAI/MIT Press.

Li, X. (2000) Engineering issues in Internet commerce. In B. Bignall (Ed.), *Internet commerce and software Agents*, pp. 6-28. Hershey, PA: Idea Group Publishing.

Liebowitz, J. & Wilcox, L. (1997). *Knowledge management and its integrative elements*. Boca Raton, FL: CRC Press.

Marin, M. (2001). The role of workflow in portal environments. In L. Fischer (ed.), *Workflow Handbook 2001*, pp. 79-90. WfMC (Workflow Management Coalition), LIghthouse Point, FL: Future Strategies Inc.

Shilakes, C.C. & Tylman, J. (1998). Enterprise information portals, In-depth report. *Merrill Lynch & Co. Global Securities Research & Economics Group*, Retrieved April 17, 2001, from *http://www.sagemaker.com/company/WhitePapers/eip_indepth.pdf*

Sowa, J. F. (1999). Knowledge representation: Logical, philosophical, and computational foundations. Pacific Grove, CA: Brooks/Cole Publishing Co.

WfMC (1996). Workflow management coalition terminology and glossary (WFMC-TC-1011), Technical Report, Workflow Management Coalition, Brussels.

Zirpins, C., Weinreich, H., Bartelt, A., & Lamersdorf. W. (2001). Advanced concepts for next generation portals. *Proceedings of the 1st Workshop on Web-Based Collaboration (WBC'01)*, IEEE Computer Society, Los Alamos, CA, 501-506.

KEY TERMS

Business Intelligence: Business information systems have transitioned from function-oriented to knowledge-oriented systems. Business intelligence is a study of business rules that are the best in practice. Intelligence is the execution of knowledge. Business intelligence is related to the knowledge acquisition, repository, sharing, and application in the activities of an organization. Business intelligence is becoming a necessity to the most business organizations to be carried out in their business procedures and functions. Business intelligence has now emerged as a sharable commodity embodied in commercial software and no longer something that could only be possessed by a small bunch of experts.

Information Broker: A key software component that supports intelligent portals. Information Broker has ability to coordinate between search engine, workflow engine, and portal management. Without an Information Broker, an individual user's profile would not be managed and consequently would not be serviced by context-sensitive portals. Information Broker may work behind the Web server to provide the personalized services for Web users. It may also work directly with WWW for searching Web services online.

Intelligent Portal: Portal designed for decision-making purposes. Intelligent Portal is not an agent program but a content-based information package that *knows* what it is used for and how it can achieve its goal. In practice, intelligent portals are supported by a software component, namely, *Information Broker* running on application servers over the Internet. The intelligence comes from the mechanisms that support the portals to deliver information at right time to the right place.

Personalization: From the service provider's viewpoint, any service provided to a user should be personalized. Personalization is a process through which the system utilizes the knowledge about a user to tailor the service to suit the online user with a flavor of the individuality. An automated personalization may promote the service provider's image and improve the effectiveness of the system functions.

Portal: A self-addressed information package with the ability to be *pushed* or *pulled* among Internet applications. It is self-contained with meta data for its information structure and presentation. Compared with the concept of *datagram* on the Internet, portal can be regarded as a kind of *information-gram* on the Internet for its properties of self-service, personalization, and real-time delivery. Mostly portals are implemented in XML.

Self-Service: One of the major differences between traditional business information system and current Web-enabled business information systems is that the user does not need to be trained and is assumed to have sophisticated prerequisite knowledge to be able to use the system. A Self-Service system is a new trend of business systems that is Web-enabled and has a user-configurable interface. A system is deployed on the Web with a generic user interface. Self-Service allows online users to select system functions to construct a Web information system in a way that is best suited for the user's information needs.

User Profile: With a large number of customers, a service provider may need to analyze its customers in terms of a list of identified business-critical factors. A user profile is a dynamically maintained data repository that groups user behaviors and classifies them into different categories according to their attributes such as age, gender, profession, etc., so that the user profile may be used as a resource of business intelligence.

Workflow Engine: A software component defined by WfMC (WfMC, 1996). Workflow engine is application independent. When a workflow plan is executed, there may be multiple workflow instances concurrently running at different states. Workflow engine is responsible for resolving the problems such as resource/data dependency, process dependency, live-lock or deadlock, event-trigger handling, etc.

Intelligent Metabusiness

Luiz Antonio Joia
Brazilian School of Public and Business Administration of Getulio Vargas Foundation and Rio de Janeiro State University, Brazil

INTRODUCTION

In 1994, the Sloan School of Management at MIT inaugurated a multi-year research and education initiative called "Inventing the Organizations of the 21st Century", headed by Thomas Malone, Director, Center for Coordination Science. One of the key activities of this initiative has been developing a series of coherent scenarios of possible future organizations. The Scenario Working Group considered a wide variety of possible driving forces, major uncertainties, and logics that might shape 21st century organizations. Two scenarios were then created addressing the size and the *modus-operandi* of the future organizations: "Small Companies, Large Networks", as the one found in Northern Italy (Textile Production in the Prato region of Italy), and "Virtual Countries", as more mergers and acquisitions are turning up worldwide (e.g., Exxon and Mobil) (Laubacher & Malone, 1997).

One of the greatest challenges of the new knowledge economy is to deal with new organizational forms, that is, the ones that challenge traditional notions of structure, coordination and control, such as the companies derived from the "Small Companies, Large Networks" scenario. When all of the tasks and processes of an enterprise are centralized in just one company, it is far from difficult to organize and manage the knowledge accrued from a project. However, a lot of different players can now be involved in major projects. Hence, how is it possible to manage and store the knowledge generated during an enterprise, so as to use it during the current project and not to lose it at all for future projects?

Some very important researchers addressed some features on this issue, such as Badaracco (1991), Bahrami (1992) and Baker (1994), just to name a few.

Notwithstanding being very important in their realm, this research just taps on how to create, deploy, transfer, store and retrieve the intelligence of an enterprise encompassing a lot of different companies, in different places, with different " although important " duties. Therefore, the next logical step includes expanding the research to ongoing and ad-hoc intraorganizational groups. In order to accomplish this, it is paramount to understand how information technology can leverage and strengthen the knowledge links among the players of a major project involving a lot of subcontractors, suppliers, and other firms, namely a metabusiness.

BACKGROUND

Metabusiness

A metabusiness or a relational company is a *quasi*-firm created through digital links among several companies, in such a way that it is almost impossible to know exactly its boundaries (Keen, 1991). A metabusiness is also independent of its organizational structure, as each node has its own structure that can be changed without interfering in other nodes' structures.

"The Organization is its Formal Structure" and "Structure follows Strategy" are two paradigms challenged by metabusinesses that wisely use information technologies.

This overview presents the role and impact of information technologies in three branches of a metabusiness: its degree of connectivity, its degree of sharing and its degree of structuring (Haeckel & Nolan, 1993). According to the latter authors, these three parameters are considered vital to establish the intelligence of a metabusiness and its expertise to manage the involved knowledge.

The connectivity issue addresses the "degree of reach" of the metabusiness, that is, if and how the involved companies are linked within the metabusiness in order to transmit data and information among themselves.

The sharing issue addresses the "degree of range" of the metabusiness, that is, the type of transactions developed within the metabusiness, and the way the companies are working together, in order to set up a workgroup environment.

Finally, the structuring issue deals with the ability that the companies have to extract knowledge from the data and information retrieved and shared by them. As is known, knowledge " either tacit or explicit " is much more than data and information, and according to the Autopoiesis Theory (Maturana & Varela, 1980) is created when a "structural coupling" occurs with the workers (see also, Kim, 1998). This overview shows that this issue is a key point for the success of an enterprise, and the one where the major flaws and drawbacks occurred. People have great difficulty to transform raw data and information into knowledge, as well as tacit to explicit knowledge, notwithstanding several frameworks explaining how this can be processed, as the Knowledge Spiral from Nonaka and Takeuchi (1995). The current educational system hinders workers to learn how to learn, making it difficult for them not to create standardized mental models to deal with new knowledge.

Hence, different technologies such as Electronic Data Interchange (EDI), Electronic Document Management Systems (EDMS), Workflow Systems, Internet/intranet/extranet and mainly Web-Based Instruction (WBI), just to name a few, are integrated to leverage the metabusiness' intelligence.

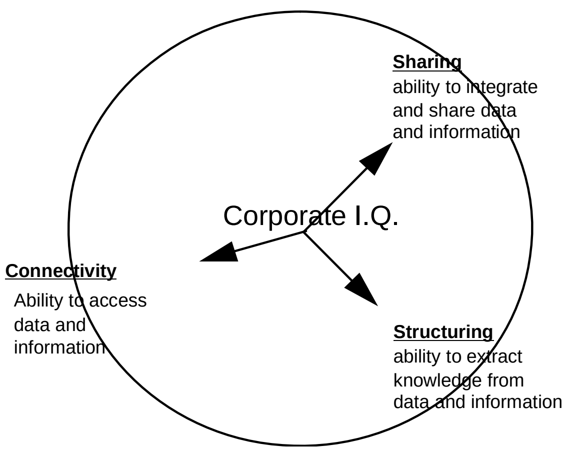
Intelligent Metabusiness

As was already said, metabusiness is a quasi-firm created through digital links among several companies, in such a way that it is almost impossible to know exactly its boundaries (Keen, 1991). This definition matches the "Small Companies, Large Networks" scenario of MIT Scenario Working Group (Laubacher & Malone, 1997). In a metabusiness, the integrator keeps the core competency of the business, outsourcing most of the other productive processes. The integrator is in charge of managing dependencies and restrains among the players and their due processes, coordinating the transactions among the involved partners.

According to Prusak (1997), some trends are forcing companies to be engaged in a metabusiness:

a) the globalization of the economy and the terrific pressure on firms for increased adaptability, innovation and process speed;
b) the awareness of the value of specialized knowledge, as embedded in organizational processes and routines of the nodes of a metabusiness;
c) the awareness of knowledge as a distinct factor of production; and
d) cheap networked computing, which is at last giving us a tool to work and learn with each other.

Figure 1. Corporate I.Q.



During the development of an enterprise, data and information are exchanged among the players compounding the metabusiness. Data and information are not knowledge, although often considered as such. There is great misunderstanding and confusion about the differences between data, information and knowledge.

Data means a set of discrete and objective facts concerning events. Therefore, it can be understood as a structured record of transactions within an organization (Davenport & Prusak, 1998).

Information is data that makes difference and is relevant, or as Peter Drucker says: "information is data with attributes of relevance and purpose" (cited in Davenport & Prusak, 1998, p.4). Normally, information is understood as a message, usually having the format of a document or visual and/or audible messages. Information is, above all, context-based.

Knowledge is linked to the capacity of action (Sveiby, 1997). It is intuitive, therefore hard to be defined. It is linked to the user's values and experience, being strongly connected to pattern recognition, analogies and implicit rules. Most of the time, knowledge within an organization is located both inside employees' heads (tacit knowledge) and in documents (explicit knowledge). This can explain why too much confusion has arisen between document management and knowledge management.

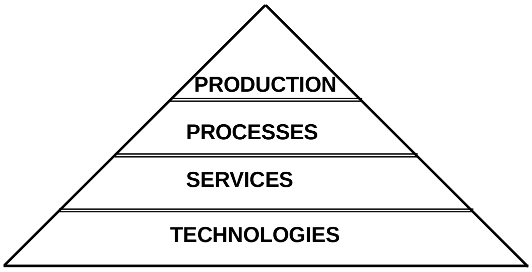
Although it is a generally accepted distinction, doubts have been cast recently over the tacit-explicit dichotomy (Polanyi, 1958). According to the autopoietic epistemology school (Varela, Thompson, & Rosch, 1992), knowledge is a private, personal thing, and so an organization cannot possess it. Hence, knowledge cannot be explicit, only tacit: Explicit knowledge is actually data and/or information which help other people to create their own knowledge through what is known as "structural coupling". However, this overview will accept the tacit-explicit distinction, which will enable us to reach more interesting conclusions.

Then, assuming the tacit-explicit dichotomy, the following mathematical formulas depict what was said (Joia, 1999):

INFORMATION = DATA + $\sum$(Attributes, Relevance, Context)
KNOWLEDGE = INFORMATION + $\sum$(Experience, Values, Patterns, Implicit Rules)

The main question is to know how knowledge can be transformed into metabusiness intelligence. Using the I.Q. metaphor (notwithstanding its flaws), it can be said that the metabusiness I.Q. (Haeckel & Nolan, 1993) can be

Figure 2. Taxonomy of the integrated model



evaluated through the enterprise's capability of connectivity (ability to access data and information through internal link and link with its partners), sharing (data and information shared among its personnel, and its partners) and structuring (ability to extract knowledge from information and raw data). Figure 1 (Joia, 1999) depicts this concept.

AN INTEGRATED MODEL FOR INTELLIGENT METABUSINESS

To integrate all of the technologies in a structured taxonomy for intelligent metabusinesses, it is taken for granted that technologies are important as they can allow services to be deployed, and services are important as they convey to productive processes, according to the following formula: TECHNOLOGIES => SERVICES => PROCESSES => PRODUCTION , as shown in Figure 2, according to Joia (2001).

Therefore, Table 1 (Joia, 2001) can present more deeply the idea shown in Figure 2, relating the Connectivity, Sharing and Structuring issues " the three facets of an intelligent metabusiness " with the needed services and productive processes within a relational enterprise.

Barriers, Causes and Solutions

Notwithstanding the great advances in the connectivity and sharing technologies and services, the same can not be said concerning the structuring technologies and services. Although great advances have been made lately, such as the development of the data mining technology, this realm is the bottleneck for the effective development of the metabusiness' intelligence. Structuring, as said before, is about creating meaning (or knowledge) for lots of data and information. Some research shows that the overload of information does not necessarily convey to adequate knowledge creation (Joia, 1999) and is the main cause of problems in complex enterprises developed in a metabusiness way. The

Table 1. Integrated model for intelligent metabusiness

	CONNECTING	SHARING	STRUCTURING
TECHNOLOGIES	Client-Server Based Intranet, Extranets, LAN, WAN, Satellite Links, Dedicated Links, Fiber-Optics Network	Internet, Groupware, Electronic Data Interchange (EDI), EDMS, Workflow, Video-Conferencing and Distributed Data Base	Web-Based Instruction Systems, Data Mining, Expert System
SERVICES	Physical Link among the players (Star Topology with the Integrator as the Hub)	Data and Information Transactions, Forms Exchange, Documents Life-Cycle Control and Tracking	Education Anytime, Anywhere Just-in-Time Education, Knowledge Acquisition, Pattern Recognition
PROCESSES	PRODUCT INNOVATION; OPERATION MANAGEMENT; PLAYERS' COORDINATION (management of restrains and dependencies); QUALITY CONTROL OF THE NODES OF THE METABUSINESS		

⇩

PRODUCTION

Table 2. Barriers, causes and solutions within an intelligent metabusiness

BARRIERS	CAUSES	SOLUTIONS
Structural		
Focus only on direct man-power and indexes	Obsolete decision criteria	Deep analysis of the costs and benefits involved
Failure to perceive the actual benefits	Lack of measures to intangible benefits	Intangible and tangible productive analysis
High risk for the managers	Reward system not considering innovation	Different reward systems for managers
Lack of coordination and cooperation	Organizational fragmentation	Systems to allow coordination/cooperation
High expectation and hidden costs	Selling of an unreal system	Planning strategic objectives
Human		
To avoid the risk	Fear of change and uncertainty	Communication and involvement
Resistance	Fear of loss of power and status	Board engaged in project implementation
Unplanned decisions and fear of being made redundant	Orientation and Action: lack of patience with planning	Pilot Project planning: long-range objectives
Technical		
Incompatibility of systems	Purchase of different hardware and software platforms	Purchase of only one integrated system; write own system; neutral transfer files

metabusiness is intelligent if and only if their nodes and respective personnel are intelligent too.

Besides, a generic table with the main problems arisen to build and maintain an intelligent metabusiness, their causes and some possible solutions, is shown in Table 2, according to Joia (2001).

FUTURE TRENDS

As it is known, the knowledge can be either explicit or tacit. Also, a metabusiness is intelligent if and only if its nodes also have intelligence. Intelligence is the ability of using knowledge to create new knowledge (Drucker, 1998). But if it is hard to convince the professionals within a company to externalize their knowledge, both because they do not know how to structure it and because of lack of confidence in the firm, what about a lot of professionals working together for the first time and belonging to different companies, geographically dispersed, tied digitally?

Hence, the first trend is the need companies have to create mechanisms that reward their professionals involved in metabusinesses, to externalize their strategic knowledge, as is already done in some management consulting companies, as Bain and Mckinsey (Hansen, Nohria, & Tierney, 1999). It is necessary to develop an environment of trust where this is possible (von Krogh, 1998). According to Venkatraman & Henderson (1998), it is important to develop new ways of managing knowledge workers, as the human resource policies are geared toward production and administrative workers. Hence, it is necessary to create incentives and compensation practices to attract and keep knowledge workers.

A second trend is to have technology used in a more comprehensive way. As McKenney, Zack, and Doherty (1992) realized:

"- managers use electronic mail for efficient communication in well-defined contexts - monitoring task status, coordinating efforts, exchanging factual

information, sending alerts, and broadcasting information.

- managers use face-to-face interaction for defining and discussing problems and solutions, building a shared understanding of the situation, discussing shifting priorities and external pressures, interpreting ambiguous signals, and socialization of members."

Once again, the solution passes through the overcoming of organizational and cultural barriers, rather than technical ones. Also, the technology must increase the employment of the tacit knowledge, as it has been used to leverage and deploy the explicit one.

The third trend is to homogenize the technological platform of all the nodes involved in a metabusiness, defining a "bottom-line" platform, as just one node with an inadequate platform can jeopardize the intelligence of the metabusiness altogether.

Finally, the fourth and more important trend, is to give the employees of an intelligent metabusiness a digital thinking in order they have their structuring skill effectively developed, that is, to give them:

- ability to decodify and structure data and information, in order to transform them in knowledge;
- identify, enunciate and solve problems by themselves;
- have the fundamentals to research, evaluate, criticize and select information among complex knowledge networks;
- be creative;
- have the ability to work in teams, most of time geographically dispersed, and have the ability of self-management;
- ability to work in digital and, often, virtual environments.

So, investments must be made in the development of customized Web-based instruction systems, so as to allow the increasing of the human capital of the firms and by consequence the intellectual capital of the metabusiness, as well as to develop a trust-based environment among the companies and the employees.

CONCLUSION

The metabusiness is an extreme case that challenges our general understanding of the management of knowledge development processes in virtual organizations through the use of technology. The intelligence of a metabusiness depends on its degree of "reach" and "range", or in other words, on its degree of connectivity, sharing and structuring.

Technology, itself, does not create either knowledge or intelligence. Technology is just an enabler to achieve this stage. Notwithstanding that a lot of information technologies are already available to be used to increase the connectivity, sharing and structuring issues of a metabusiness in order to make it intelligent and with a high degree of "reach" and "range", the organizational and cultural obstacles are still very high to be smoothly overcome, and the causes and solutions must be analyzed.

The attitude is dangerous that drives management towards strong investments in information technology, possibly at the expense of investments in human capital. The danger lies on an IT-driven knowledge management strategy that may end up objectifying and calcifying knowledge into static, inert information, thus disregarding altogether the role of tacit knowledge. It can be mentioned that the major problem Volkswagen has faced in its newly built truck plant in Resende, Brazil "regarded as the most advanced automobile production process implemented in the world" has come from problems with the quality of the products/services delivered by its partners. As part of the enterprise, VW built a Training Center in Angra dos Reis, near Resende, close to the factory, where all of the metabusiness' nodes employees are trained before being engaged in the production network. However, the quality problems still remain, mainly due to the lack of employees' structuring skills, which hinders them for creating their own knowledge through the training process. Notwithstanding that the main players of this metabusiness work in the same geographical place, without certain physical boundaries, this innovative project has shown the need of investing in the leverage of the human capital of all the companies, otherwise all of the enterprise can be jeopardized.

The knowledge society with its knowledge workers demands new organizational forms (Laubacher & Malone, 2002), organic, flexible, seemingly almost anarchic, but coordinated enough to allow knowledge to be created, stored, retrieved and reused. Development and improvements in the newly created coordination science (Malone & Crowston, 1994) is paramount to spread intelligent metabusinesses as the "knowledge-based organizational structure" in compliance with the "Small Companies-Large Networks" scenario of the MIT Group for Organizations of the 21st Century.

REFERENCES

Badaracco, J. (1991). Knowledge links. Chapter 5 of *The knowledge link: How firms compete through strategic* *alliances*, (pp.107-128). Boston: Harvard Business School Press.

Bahrami, H. (1992). The emerging flexible organization: Perspectives from Silicon Valley. *California Management Review*, *34*(4).

Baker, W. (1994). Building intelligent networks. Chapter 3 of *Networking smart*. McGraw-Hill, Inc.

Davenport, T.H., & Prusak, L. (1998). *Working knowledge*. Harvard Business School Press.

Drucker, P. (1998). From capitalism to knowledge society. In D. Neef (Ed.), *The knowledge economy*, (pp.15-34). Butterworth-Heinemann.

Haeckel, S., & Nolan, R. (1993). Managing by wire. *Harvard Business Review*, September-October, 122-132.

Hansen, M.T., Nohria, N., & Tierney T. (1999). What's your strategy for managing knowledge? *Harvard Business Review*, March-April, 106-126.

Joia, L.A. (1999). A new model for workers' retraining in Brazil. *Journal of Workplace Learning*, *11*(4), MCB University Press.

Joia, L.A (2001). Information technology for intelligent metabusiness. In Y. Malhotra (Ed.), *Knowledge management and business model innovation*, (pp.95-103). Hershey, PA: Idea Group Publishing.

Keen, P. (1991). *Shaping the future*. Harvard Business School Press.

Kim, D.H. (1998). The link between individual and organizational learning. In D. Klein (Ed.), *The strategic management of intellectual capital*, (pp.41-62). Butterworth-Heinemann.

Laubacher, R., & Malone, T. (1997). Two scenarios for 21st century organizations: Shifting networks of small firms or all-encompassing "virtual countries"? MIT Initiative on Inventing the Organization of the 21st Century, Working Paper 21C WP #001, Sloan School of Management, MIT.

Laubacher, R., & Malone, T. (2002). Retreat of the firm and the rise of guilds: The employment relationship in an age of virtual business. MIT Initiative on Inventing the Organization of the 21st Century, Working Paper 21C WP #033, Sloan School of Management, MIT.

Malone, T., & Crowston, K.G. (1994, March). The interdisciplinary study of coordination. *ACM Computing Surveys*, *26*(1), 87-119.

Maturana, H., & Varela, F.J. (1980). *Autopoiesis and cognition: The realization of the living*. London: Reidl.

McKenney, J.L., Zack, M.H., & Doherty, V.S.(1992). Complementary communication media: A comparison of electronic mail and face-to-face communication in a programming team. Chapter 10 of N. Nohria & R.G. Eccles (Eds.), *Networks and organizations: Structure, form and action*. Harvard Business School Press.

Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. New York: Oxford University Press.

Polanyi, M. (1958). *Personal knowledge*. London: Routledge.

Prusak, L. (1997). Introduction to knowledge in organizations. In L. Prusak (Ed.), *Knowledge in organizations*, (pp.ix-xv). Butterworth-Heinemann.

Sveiby, K.E. (1997). *The new organizational wealth*. Berret-Koehler Publishers, Inc.

Varela, F.J., Thompson E., & Rosch E. (1992). *The embodied mind*. Cambridge, Mass.: MIT Press.

Venkatraman, N., & Henderson, J.C. (1998, Fall). Real strategies for virtual organizing. *Sloan Management Review*, *40*(1), 33-48.

von Krogh, G. (1998, Spring). Care in knowledge creation. *California Management Review*, *40*(3), 133-153.

KEY TERMS

Corporate I.Q.: Enterprise's capability of connectivity (internal link and link with its partners), sharing (data and information shared among its personnel, and its partners) and structuring (ability to extract knowledge from information and raw data).

Data: A set of discrete and objective facts concerning events.

Degree of Range: Type of transactions developed within the metabusiness, and the way the companies are working together, in order to set up a workgroup environment.

Degree of Reach: How the involved companies are linked within the metabusiness in order to transmit data and information among themselves.

Degree of Structuring: Deals with the ability that the companies have to extract knowledge from the data and information retrieved and shared by them.

Digital Thinking: Creation of a new mental model that is based on an assynchronous mental model and on the ability of linking, combining and associating different, or even opposite ideas.

Information: Data with attributes of relevance and purpose.

Intelligent Metabusiness: A metabusiness with a high degree of reach, range and structuring.

Knowledge: User's values and experience, being strongly connected to pattern recognition, analogies and implicit rules.

Metabusiness: A *quasi*-firm created through digital links among several companies, in such a way that it is almost impossible to know exactly its boundaries.

Intelligent Software Agents

Alexa Heucke
Munich University of Applied Sciences, Germany

Georg Peters
Munich University of Applied Sciences, Germany

Roger Tagg
University of South Australia, Australia

INTRODUCTION

An agent, in traditional use of the word, is a person that acts on behalf of another person or group of persons. In software, the term *agent* is broadly used to describe software that carries out a specialist range of tasks, on behalf of either a human user or other pieces of software. Such a concept is not new in computing. Similar things have been said about subroutines, re-usable objects, components and Web services. So what makes agents more than just another computer technology buzzword and research fashion?

BACKGROUND

The idea of intelligent agents in computing goes back several decades. Foner (1993, p. 1) dates the first research on software agents to the late 1950s and early 1960s. However, with the breakthrough of the Internet, intelligent agents have become more intensively researched since the early 1990s. In spite of this long heritage, the uptake of these ideas in practice has been patchy, although the perceived situation may be partly clouded by commercial secrecy considerations. Even today, the many different notions of the term *software agent* suggest that the computing profession has not yet reached a generally accepted understanding of exactly what an agent is.

DEFINITIONS AND CLASSIFICATIONS

According to Jennings et al. (1998, p. 8), "an agent is a computer system, *situated* in some environment that is capable of *flexible autonomous* action in order to meet its design objectives".

- S*ituated:* the agent interacts directly with its environment. It receives inputs from the environment and it performs activities with external effects.
- *Autonomous:* the agent is in charge of its own internal status and actions. Thus, it can perform without explicit interference of a user or other agents.
- *Flexible:* the agent is responsive to its environment and, at the same time, proactive. The agent should show social behaviour and should feature the ability to interact with external entities.

Furthermore, many agents show intelligence, in that they "carry out some set of operations on behalf of a user or another program with some degree of independence or autonomy, and in doing so, employ some knowledge or representation of the user's goals or desire" (Gilbert et al., 1995, p. 2f).

The research literature discusses many different types of agents, carrying out all sorts of functions with what can be termed primary and secondary characteristics. Primary characteristics include autonomy, cooperation and learning, while secondary characteristics include aspects like multi-functionality, goodwill or trustiness.

A typology of software agents has been proposed by Nwana (1996, pp. 7-38):

- Collaborative agents feature a high degree of cooperation and autonomy. They are determined by the idea of distributed artificial intelligence and by the concept of task sharing, cooperation and negotiation between agents.
- Interface agents focus on the characteristics of learning and autonomy. By collaborating with the user and by sharing knowledge with other agents they learn a user's behaviour and are trained to take the initiative to act appropriately.
- Mobile agents are not static but have the ability to travel. This entails non-functional benefits such as freeing local resources, showing more flexibility and enabling an asynchronous work scenario.

- Information or Internet agents emphasise managing enormous amounts of information. Their main task is to know where to search for information, how to retrieve it and how to aggregate it.
- Reactive agents are showing a stimulus-response manner as opposed to acting deliberatively. Since they are based in the physical world and only react to present changes, their behaviour is not predetermined.
- Hybrid agents comprise more than one agent philosophy and benefit from the combination of different architectures.

Wooldridge and Jennings (1995, pp. 24-30) offer a two-way classification, based on contrasting approaches to building agents. They distinguish the following representative architectures:

- Deliberative agent architecture. This classical agent architecture consists of one definite, symbolic world model with all decisions being made on the basis of logical reasoning. Challenges of this approach are the translation of the real world into an accurate model and the establishment of an efficient reasoning.
- Reactive agent architecture. In contrast to the deliberative agent architecture this alternative approach is lacking an explicit and symbolic model of the world as well as extensive reasoning.

Wooldridge and Jennings also allow for hybrid agent architectures that are built as a hierarchy of deliberative and reactive agent architecture layers.

DISCUSSION

Four aspects are of particular interest when trying to understand how agents work and could be successfully employed in applications and environments. Discussed next are: agent knowledge, agent applications, agent standards and multi-agent systems.

Agent Knowledge

To operate autonomously, any software agent must build up a collection of knowledge, typically data and rules that enable it to serve the human or client software it is acting for. According to Maes (1994, p. 2f), an agent's knowledge base should be built up gradually by learning from users and other agents. The key issues are competence and trust. To be competent, the agent must have a knowledge base that is comprehensive and flexible enough to adapt to the user's profile. For an agent to be trusted, a human user must feel comfortable when accepting help from the agent or when delegating tasks to it. Generally, an agent can only learn from its user and other agents if their actions show an iterative pattern. Maes suggests four different ways of training an agent to build up competence: observation and imitation of the user's habits, user feedback, training by examples and training by other agents.

However, Nwana and Ndumu (1999, p. 10) have criticized Maes' approach, claiming that an agent would not only need to know all peculiarities of the deployed operating system but also must understand all tasks its user is engaged in. Furthermore, the agent would need to be capable of gathering the user's intent at any time, thus continuously modelling its user. Nwana and Ndumu identify four main competences for an agent: domain knowledge about the application, a model of its user, strategies for assistance and a catalogue of typical problems that users face in the environment.

Agent Applications

Software agents can be employed in many fields of information technology. One role for agents is to act as an assistant or helper to an individual user who is working with a complex computer system or physical equipment. Examples are:

- Information agents (Davies et al., 1996, pp. 105-108) that help a human researcher in finding the most relevant material, from many sources and possibly through different search engines.
- Decision support agents that help a user assess alternative courses of action; functions include filtering and summarisation of data, optimising algorithms, heuristics and so forth.
- E-mail agents (Maes, 1994, p. 5f), which filter spam, allocate incoming mail to folders, and work out addresses to which outgoing mail should be sent.
- Buying and selling agents, which assist a user in finding good deals in Internet marketplaces or bidding agents, which assist participants in auctions. These agents have characteristics of information agents as well as of decision support agents.

A second group of applications is where the agent acts as a coordinator of activities, or "virtual manager". Any workflow management system could qualify for this category. Other examples include meeting scheduling agents (Kozierok, 1993, p. 5), and dynamic scheduling agents that are able to re-allocate resources to meet the goals of a business process (Lander et al., 1999, p. 1ff).

Delegation agents are another example in this category, although they could also be regarded as individual support.

A third group of applications is where the agent continually monitors data and rules in an organisation, and on that organisation's behalf alerts, or sends messages to, human recipients. Examples are advertising agents, notification agents, recommendation agents and selling agents. Such agents are at work when you receive an e-mail from Amazon.com about a book that might interest you.

Other agents act as a third party between two humans, or pieces of software, that need to co-operate. Examples include brokering agents, negotiation agents, mediation agents and ontology agents (Helal et al., 2001; Pivk & Gams, 2000).

Many humans, computer systems and even other agents depend on one particular specialised task, which is a common agent or subagent, especially useful in an era of information overload. This is a categorisation agent (Segal, 2000, p. 2f). Such an agent has the task of applying, and where necessary building up, a classification structure for incoming data. This structure may be particular to an individual (e.g., for e-mail filtering and filing) or it may be for an organisational unit.

Agent Standards

Intelligent agents are intended in order to function in heterogeneous system environments. To interact smoothly and efficiently in such environments, standardisation is essential.

Although agent technology is relatively immature and many researchers still have their own definition of agents, professional bodies have been developing standards for agents since the late 1990s.

These organisations include (Dickinson, 1997):

- ARPA Knowledge Sharing Effort (KSE)
- Agent Society
- OMG Mobile Agent System Interoperability Facility (MASIF)
- The Foundation for Intelligent Physical Agents (FIPA)

The FIPA and MASIF standards are regarded as of special importance for intelligent agents. While the FIPA Standard has its origins in the intelligent agent community and has been influenced by the KQML (Knowledge Query and Manipulation Language), MASIF deals primarily with agent mobility.

The FIPA 2000 standard (*http:www.fipa.org*) specification deals with mobility and tries to integrate MASIF (Milojicic et al., 1998, pp. 50-67). Therefore this specification bridges the gap between the intelligent and mobile agent communities. FIPA's specification is divided into five main categories: applications, abstract architecture, agent communication, agent management and agent message transport.

Multi-Agent Systems

Much of the recent literature on agents envisages a system with a community of agents that cooperate in some way to achieve an overall set of goals. According to Jennings et al. (1998, p. 9, p. 17f), "Multi-agent systems are ideally suited to representing problems that have multiple problem-solving methods, multiple perspectives and/or multiple problem solving entities". Since each agent has a restricted view of any problem and only limited information, multi-agent systems feature a flexible and advanced infrastructure to solve issues beyond individual capabilities. Thus, the system can benefit from every agent's expert knowledge. Other characteristics of multi-agent systems are decentralised data, asynchronous computation and the lack of a central control system.

A major challenge of multi-agent systems is clearly the means of coordination between agents. Nwana et al. (1997, pp. 33-55) have identified the following key components in such coordination: foreseeable structures of agent interaction; defined agent behaviour and social structures; flexibility and dynamics; and the knowledge and reasoning to utilise the above.

Possible coordination techniques are:

- Organisational structuring. The agents' roles, responsibilities, their communication chains and paths of authority are defined beforehand.
- Contracting. All tasks and resources distributed amongst agents are controlled by a contract net protocol.
- Multi-agent planning approach. Agents decide on a detailed and interlaced plan of all activities and goals. Multi-agent planning can be centralised, with one agent reviewing all individual plans and coordinating them into a single – or a distributed - multi-agent plan.

When agents are interacting in a multi-agent system, they may have to negotiate in order to fulfil their interests. Nwana et al. suggest two different negotiation theories:

- In the *game theory-based negotiation* approach, each agent holds a utility matrix that lists how much a certain interaction or goal is worth. During the negotiation process, which is defined by a negotiation protocol, the parties exchange bids and counter-offers following their strategies.

- In the *plan-based negotiation* theory, each agent schedules its actions individually before all plans are coordinated. This is similar to the multi-agent planning coordination approach, but any agent can play the role of central coordinator.

FUTURE TRENDS

We believe that the area of support for human users in the carrying out of highly heterogeneous workloads represents a promising area for the development of agent applications. The current support for users who work with a mixture of word processing, e-mail, spreadsheets, databases, digital libraries and Web search tools is very primitive. The user has to do most of the work in correlating the different sources, and current tools are poor at learning the user's commonly repeated work patterns. The authors of this article feel that agents are the most promising technology to redress this shortcoming, and have worked on architecture for linking agents with tools such as Groupware and Workflow.

CONCLUSION

In spite of a considerable amount of research, the killer application for intelligent agents is still somewhat elusive. The IT industry has still not reached a consensus about the use of agents now and in the future. Nwana and Ndumu (1999, p. 14) are even more critical, claiming that "not much discernible progress has been made post 1994". The main reason might be that intelligent agent theory integrates some of the most challenging concepts in science: for example, artificial intelligence, data mining or contract theory. Agent technology can be considered as another new demanding application of these concepts and will succeed or fail depending on any progress in these areas. The take-up of agent technology is therefore likely to suffer the same ups and downs that AI has experienced in recent decades. In the longer term, however, there is a large area of opportunity for agents supporting human users of computer systems, which has yet to be fully developed.

REFERENCES

Davies, J., Weeks, R., & Revett, M. (1996). Information agents for the World Wide Web. *BT Technology Journal, 14*(4), 105-114.

Dickinson, I. J. (1997). Agents standards. Retrieved September 21, 2003, from *http://www.hpl.hp.com/techreports/97/HPL-97-156.pdf*. HP Laboratories Bristol, HPL-97-156, Bristol, UK.

Foner, L. (1993). *What's an agent, anyway? A sociological case study*. Retrieved June 3, 2003, from *http://foner.www.media.mit.edu/people/foner/Reports/Julia/Agents—Julia.pdf*

Gilbert, D., Aparicio, M., Atkinson, B., Brady, S., Ciccarino, J., Grosof, B. et al. (1995). *The role of intelligent agents in the information infrastructure.* IBM Report.

Helal, S., Wang, M., & Jagatheesan, A. (2001). Service-centric brokering in dynamic e-business agent communities. *Journal of Electronic Commerce Research.* Baltzar Science Publishers. Retrieved September 10, 2003, from *http://www.harris.cise.ufl.edu/projects/publications/BPtemp148.pdf*

Jennings, N., Sycara, K., & Wooldridge, M. (1998). A roadmap of agent research and development. *Autonomous Agents and Multi-Agent Systems, 1,* 7-38.

Kozierok, R., & Maes, P. (1993). A learning interface agent for scheduling meetings. *Proceedings of the 1st International Conference on Intelligent User Interfaces,* Orlando, FL (pp. 81-88).

Lander, S., Corkill, D., & Rubinstein, Z. (1999). KPM: A tool for intelligent project management and execution. Workshop on Intelligent Workflow and Process Management: 16th International Joint Conference on Artificial Intelligence, Stockholm, Sweden.

Maes, P. (1994). Agents that reduce work and information overflow. *Communications of the ACM, 37*(7), 31-40.

Milojicic, D., Breugst, M., Busse, I., Campell, J., Covaci, S., Friedman, B. et al.(1998). MASIF The OMG Mobile Agent System Interoperability Facility. *Mobile Agents 1998,* 50-67.

Nwana, H. (1996). Software agents: An overview. *The Knowledge Engineering Review, 11*(3), 205-244.

Nwana, H., Lee, L., & Jennings, N. (1997). Co-ordination in multi-agent systems. In H. Nwana & N. Azarmi (Eds.), *Software agents and soft computing. Towards enhancing machine intelligence* (pp. 42-58). Berlin: Springer.

Nwana, H., & Ndumu, D. (1999). A perspective on software agents research. *The Knowledge Engineering Review, 14*(2), 125-142.

Pivk, A., & Gams, M. (2000). E-commerce intelligent agents. *Proceedings of ICTEC 2000*, Dallas, TX (pp 418-429).

Segal, R., & Kephart, J. (2000). Incremental learning in SwiftFile. *Proceedings of the 17th International Conference on Machine Learning,* Stanford, CA (pp. 863-870).

Wooldridge, M. (2002). *An introduction to multiagent systems*. Chichester: John Wiley & Sons.

Wooldridge, M., & Jennings, N. (1995). Intelligent agents: Theory and practice. *The Knowledge Engineering Review*, *10*(2), 115-152.

KEY TERMS

Artificial Intelligence: Computer systems that feature automated human-intelligent, rational behaviour and employ knowledge representation and reasoning methods.

Business Process: A process at the business layer of an organisation. Since the 1990s, the focus of any business reengineering project and one of the central inputs for IT design. It is sometimes also used as a synonym for workflow.

Categorisation: The process of deducing, from the content of an artefact, the potentially multiple ways in which the artefact can be classified for the purpose of later retrieval from a database, library, collection or physical storage system.

Contract Theory: Contract theory deals with aspects of negotiation and contracting between two or more parties.

Data Mining: Integrating statistics, database technology, pattern recognition, and machine learning, to generate additional value and strategic advantages.

Game Theory: Mathematical theory of rational behaviour for situations involving conflicts of interest.

Workflow: The automation of a business process, in whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural rules (WFMC.org) (sometimes also used as a synonym for business process).

Intelligent Software Agents in E-Commerce

Mahesh S. Raisinghani
University of Dallas, USA

Christopher Klassen
University of Dallas, USA

Lawrence L. Schkade
University of Texas at Arlington, USA

INTRODUCTION

Agent technology is one of the most widely discussed topics in information systems and computer science literature. New software products are being introduced each day. A growing number of computer information professionals recognize that there are definite issues surrounding intelligent agent terminology. These must be resolved if agent technology is to continue to develop and establish.

Current research on intelligent agent software technology can be categorized as two main areas: technological and social. In the excitement of emergent technology, people often forget to scrutinize how new technology may impact their lives. The social dimension of technological progress is the driving force and most central concern of technology. Technology is not created for its own sake as a technological imperative. This article critiques the current state of software intelligent agents by examining technological issues and the social implications of intelligent agent software technology.

TECHNOLOGICAL ISSUES

An attempt to arrive at a generally accepted definition is the first hurdle. In order for this term to have any effectiveness, there must first be a universal definition that can be agreed upon and used consistently. Unfortunately, there is none. Many proposals for defining an "intelligent agent" have been put forth, but none has received wide acceptance. Some of these proposals are the following:

- "An agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through effectors." Russell and Norvig (1995)
- "Let us define an agent as a persistent software entity dedicated to a specific purpose. 'Persistent' distinguishes agents from subroutines; agents have their own ideas about how to accomplish tasks, their own agendas. 'Special purpose' distinguishes them from other entire multifunction applications; agents are typically much smaller."
- "An autonomous agent is a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future." Franklin and Graesser (1996)

While these terms attempt to describe characteristics of intelligent agents, no comprehensive and generic description for these agents has gained wide recognition as the definitive description of a software agent. A consensus definition has not yet been achieved. As Franklin and Graesser (1996) indicate, most of the definitions proposed are derived from conceptualizations peculiar to the subjective views of the individuals. It is important to note that it is this intuitive aspect of an "intelligent agent" which makes it difficult to establish a broadly accepted formal definition. Ironically, it facilitates marketing of intelligent agent software technology.

A second reason for a lack of a consensus definition is that much of the agent research is proprietary. Companies that make investments to sponsor such research do not wish to reveal their competitive edge nor give away the value of their work. Standardization of new technology is difficult. Uncertainty will continue until the companies and individuals with the proprietary information recognize that sharing knowledge benefits everyone.

A third reason for the difficulty for the lack of a generally approved definition of an intelligent software agent, and probably the most important reason of the three outlined in this article, is that intelligent agent software does not seem to be qualitatively different from other software. "Is it an agent, or just a program?" Franklin and Graesser (1996) ask and observe, correctly, that all software agents are programs. The authors also state that not all programs are agents. The implication is that some programs are, in fact, agents. If an "intelligent agent" were just an added complex program, the term "intelligent agent" would simply mean that a software program was simply extended, made more composite and possibly more useful than other typical programs. An intelligent agent differs from a procedural program in two ways. First, it is an agent and broadly speaking, it is defined as someone or something that acts. To be able to act, the entity must have a purpose or a goal. A computer program can only perform a prescribed set of instructions. An intelligent software agent has the same capability and is similar to a computer program in this respect.

Computer programs act utilizing a relatively low level of logic. These programs cannot act autonomously. For any entity to act with autonomy there must be concomitant independence and freedom. Procedural computer programs do not have volition, because whatever is written into the program is executed. The key factor is logic bound and a closed program. The term "react" is an inherent limitation of closed computer programs. An agent, in the true sense of the word, initiates action. The several reasons illustrate why some time is required for an acceptable definition of software agent. This process is likely to be somewhat similar to the emergence of the distinction between artificial intelligence, expert systems and decision support systems, which became clear gradually with more widespread usage of this distinctive software.

We do not demean the effort that has been invested into these products. Systems that are based on the detection of patterns in conjunction with explicit user commands and preferences are based on straightforward computational mathematics and logic. Technical challenges exist in the areas of security, connectivity, storage, peer group collaboration, network-based services, user interface, stability and standards (Bantz et al., 2003). Park and Park (2003) propose an agent-based system for merchandise management and verify its application in a duty-free shop, which performs evaluating and selecting merchandise and predicting seasons and building purchase schedules autonomously in place of human merchandise managers under a business-to-business (B2B) electronic commerce (EC) environment. In order to facilitate the agent's intelligent behavior, several analysis tools such as data envelopment analysis (DEA), genetic algorithm (GA), linear regression and rule induction algorithm are incorporated into the system.

E-mail and filters reject messages that do not comply to the user's defined preferences. Help engines and data warehousing tools search for built-in patterns. Patterns are pre-built into the engines, which are limited by the closed logic bounds specified by the designs. News and searching tools have a great potential, albeit they pose a concern. The dilemma is if many users have news searching intelligent agent tools constantly searching for information on the Internet, the Internet may possibly be clogged up by too many of these searching tools. Imagine if one of these intelligent agents had a built-in error (bug) that caused the program to continuously spawn even more agents to search the Internet. Moreover, some intelligent agents searching the Internet for information could get lost and not return with the requested information. Thus, one can see the latent technical threat in employing such ill-designed intelligent agents. These "lost" agents may create severe bottlenecks on the Internet.

Although intelligence means people thinking, it may be possible to replicate the same set of behaviors using computation. This idea was discussed by Turing in the 1940s. In 1950 he proposed a test, now called the Turing Test (TT), for computational intelligence. In the test, a human judgment must be made concerning whether a set of observed behaviors is sufficiently similar to human behaviors that the same word - intelligent - can justifiably be used. Feigenbaum (2003) discuss the challenges for computational intelligence, including: 1. an alternative to TT that tests the facet of quality (the complexity, the depth) of reasoning, 2. building a large knowledge base by reading text, thus reducing knowledge engineering effort, and 3. distilling from the WWW a huge knowledge base, thus reducing the cost of knowledge engineering.

Olin et al. (2001) suggest that although tools in the shape of distributed artificial intelligence will be available and be particularly applicable to the complexities of decision-making in the typical global enterprise, a number of issues will arise in the next few years as intelligent agents become the mainstream enablers of "real-time" enterprise process, such as the need to ensure continuity in the transition phase by careful integration of intelligent agent systems in to a legacy systems.

Van Den Heuvel and Maamar (2003) propose a framework for contract-based support to establish virtual collaboration using loosely coupled and heterogeneous intelligent Web services (IWS) in which contracts encapsulate the control information for IWSs engaged in e-business transactions. Since IWS technology is still in its infancy, several important issues must be addressed before agentified Web services can be successfully deployed at e-marketplaces, such as the integration of IWSs with wrapped legacy systems, the semantic integration of Web services, and the integration of Web services into e-commerce transactions (Van Den Heuvel & Maamar, 2003).

While search engines have become the major decision support tools for the Internet, there is a growing disparity between the image of the World Wide Web stored in search engine repositories and the actual dynamic, distributed nature of Web data. The traditional static methods in which search and retrieval are disjoint seem limited. Menczer (2003) proposes using an adaptive population of intelligent agents mining the Web online at query time and presents a public Web intelligence tool called MySpiders, a threaded multi-agent system designed for information discovery and augmenting search engines with adaptive populations of intelligent search agents for significant competitive advantage. Weippl et al. (2003) describe the manifold security threats that occur in mobile agent systems. For instance, when masquerading, an agent claims the identity of another agent and tries to gain unauthorized access to resources. A possible solution to this security flaw is signing agent code with digital signatures such as those provided by the Java cryptographic extension (JCE 1.2.1). Mobile agents could launch denial of service attacks since too many agents could flood the Java runtime environment. Another relevant issue that needs to be addressed in the future is database intelligent agents' independence of the Java agent platform used for transport agents. Due to the lack of support offered by major Java agent platforms, the unconfined employment of standardized interfaces is not yet possible. Finally the issue of multiple, heterogeneous data warehouse islands that need to be made interoperable is being investigated by using intelligent agents to realize the federation. For instance, Xu et al. (2003) propose a two-layer approach for the formal modeling of logical agent mobility (LAM) using predicate/transition (PrT) nets in which a mobile agent system is viewed as a set of agent spaces and agents could migrate from one space to another.

SOCIAL IMPLICATIONS

The social implications, as with any new technology, include both positive and negative aspects. We are aware that the current literature debate and research on intelligent agent software technology deals in moderation with the topic of the social and ethical implications. Technologies are human artifacts, developed, presumably, for the benefit of humankind and the improvement of the quality of life. New technologies must, however, be tested to determine if the product meets these. Most important is to check if the new technology is significantly different from the existing ones.

One of the major benefits from using intelligent agents is the potential for liberating humans from the tedious task of searching for information on the Internet. The intelligent agent can aid in the search by filtering information that has little or no value. Unfortunately, intelligent agents also have the potential for damage. First, with excessive reliance on intelligent agents, humans can risk an excessive loss of choice. According to Lanier (1999), confining to an artificial world created by a programmer(s) can limit human potential for innovation. Another objection Lanier raises is that human beings degrade themselves by using intelligent agents. When individuals begin to think of computers anthropomorphically, as actually possessing intelligence and autonomy, people will tend to relate to computers as if they were human. The opposite is true too. This is a serious dilemma that must be avoided. An additional technical problem that Lanier (1999) raises is that info-consumers see the world through agents' eyes. The point is that if intelligent agents are used to find useful information, the agents themselves may be manipulated.

FUTURE TRENDS

The technological, social, and ethical issues notwithstanding, the new business environment is characterized not only by a rapid pace of change but also the dynamically discontinuous nature of such change. The issues such as lack of standardization in mobile agents may cause lack of identity traceability due to multiple transfers among networks. The security concerns relate to machine protection without artificially limiting agent access rights. Finally, there are issues surrounding performance and scalability, such as the performance effects that high levels of security would have on the network, as well as the effects of having multiple mobile agents in the same system. The emergence of intelligent mobile/software agents not only will change the way that we communicate across networks, but also have a profound impact on the way that we accomplish many tasks.

CONCLUSION

In conclusion, while intelligent agent technology has the potential for being useful to humankind, many fundamental issues remain unsolved. These problems are technical, social and/or ethical and require careful thought and consideration by developers. This discussion is critical of the current state of intelligent agent software technology and research, in the hope of encouraging developers to be aware of spillover effects. Unfortunately, issues raised by authors such as Lanier (1999) are often thrust aside as being extremist. Intelligent agent software technology has advanced, but greater progress must be made socially and ethically before these agents can be accepted as tools for the improvement of the quality of human life.

REFERENCES

Bantz, D.F., Bisdikian, C., Challener, D., & Karidis, J.P. (2003). Autonomic personal computing. *IBM Systems Journal, Armonk: 42*(1), 165-177.

Feigenbaum, E.A. (2003, January). Some challenges and grand challenges for computational intelligence. *Journal of the Association for Computing Machinery, 50*(1), 32-42. New York: Association for Computing Machinery.

Franklin, S., & Graesser, A. (1996). Is it an agent, or just a program?: A taxonomy for autonomous agents. *Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages.* Springer-Verlag.

Lanier, J. (1999). Agents of alienation. Retrieved May 18, 2000, from *http://www.well.com/ user/ jaron/ agent alien.html*

Menczer, F. (2003, May). Complementing search engines with online Web mining agents. *Decision Support Systems, 35*(2), 195-206.

Olin, J., Greis, N., & Morgan, L. (2001). The enterprise at the edge: Agents to the rescue. *European Management Journal, 19*(5), 489-501.

Park, J.H., & Park, S.C. (2003, June). Agent-based merchandise management in business-to-business electronic commerce. *Decision Support Systems, 35*(3), 311-324.

Russell, S.J., & Norvig, P. (1995). *Artificial intelligence: A modern approach.* Englewood Cliffs, NJ: Prentice Hall.

Smith, D.C., Cypher, A., & Spohrer, J. (1994). KidSim: Programming agents without a programming language. *Communications of the ACM, 37*(7), 55-67.

The IBM Agent. *http://activist.gpl.ibm.com: 81/ WhitePaper/ptc2.htm*

Van Den Heuvel, W.-J., & Zakaria, M. (2003, October). Moving toward a framework to compose intelligent Web services. *Communications of the ACM, 46*(10), 103-114. New York: Association for Computing Machinery.

Vinaja, R., & Sircar, S. (1999). Agents delivering business intelligence. *Handbook of information technology* (pp. 477-490). CRC Press.

Weippl, E., Klug, L., & Essmayr, W. (2003, January-March). A new approach to secure federated information bases using agent technology. *Journal of Database Management, 14*(1), 48-69.

Xu Dianxiang, Y., Jianwen, D.Y., & Ding, J. (2003, January). A formal architectural model for logical agent mobility. *IEEE Transactions on Software Engineering, 29*(1), 31-42. New York.

KEY TERMS

Agent: A program designed to provide specialized and well defined services. Agent can be *static* – executing on the computer where it was installed, or *mobile* – executing on computer nodes in a network.

Autonomous Agent: A system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future.

Bottlenecks: A stage in a process that causes the entire process to slow down or stop.

Data Warehousing: A form of data storage geared towards business intelligence. It integrates data from various parts of the company. The data in a data warehouse are read-only and tend to include historical as well as current data so that users can perform trend analysis.

Decision Support Systems: A specific class of computerized information system that supports business and organizational decision-making activities. DSS is an interactive software-based system that compiles useful information from raw data, documents, personal knowledge, and/or business models to identify and solve problems and make decisions

Expert Systems: A computer system that facilitates solving problems in a given field or application by drawing inference from a knowledge base developed from human expertise. Some expert systems are able to improve their knowledge base and develop new inference rules based on their experience with previous problems.

Intelligent Agent: A program that gathers information or performs some other service without your immediate presence and on some regular schedule.

Turing Test (TT), for computational intelligence: In the test, a human judgment must be made concerning whether a set of observed behaviors is sufficiently similar to human behaviors that the same word - intelligent - can justifiably be used.

Interactivity and Amusement in Electronic Commerce

Yuan Gao
Ramapo College of New Jersey, USA

INTRODUCTION

The use of structural features such as text size, font, graphics, color, animation, video and audio has been widely explored in the traditional media. Similar uses of such features have also been found in the online environment (Rodgers & Thorson, 2000). Factors related to consumer's behavior, attitude, and perceptions regarding Web sites have been examined in academic literature (Chen & Wells, 1999; Coyle & Thorson, 2001; Ducoffe, 1996; Eighmey, 1997; Koufaris, 2002; Koufaris, Kambil, & Labarbera, 2001; Vijayasarathy, 2003). They include the investigation of effects of interactive features on Web site appeal (Ghose & Dou, 1998), and e-store characteristics on site traffic and sales (Lohse & Spiller, 1998). Some experimental studies examined the effects of animation and image maps on perceived telepresence and consumer attitude (e.g., Coyle & Thorson, 2001), and the use of pop-up windows on consumer decision-making processes (Xia & Sudharshan, 2000).

BACKGROUND

The current literature is gradually expanding on the study of using interactive features to provide visitors a more realistic experience, and the exploration of using amusement techniques to enhance entertainment value of a Web site. Recent literature has also extended Resnik and Stern's (1977) content analytical method to examine Web site information content and design features (Ghose & Dou, 1996; Huizingh, 2000; Palmer & Griffith, 1998; Philport & Arbittier, 1997). Mitra (2003) argues that institutions find themselves dwelling at the intersection of the real and the virtual, and are consequently faced with a more complex act than developing a Web site in order to make its appearance in cyberspace attractive. Thus, more aspects of a Web-based communications message, especially attributes related to interactivity, entertainment, and amusement need to be explored.

INTERACTIVITY

An important attribute of using a Web site to disseminate product information and conduct e-commerce is its ability to engage customers in interacting with the Web site. Interactive functions include text-based email inquiries, feedback forms, and multimedia components that attempt to retain customers at the site for longer visits. Entertaining features in the hypermedia include pictures, virtual reality displays of products, multimedia shows, online games, and the use of cartoons and humor. A Web site is a mix of direct selling and advertising with characteristics of both general product display and interactive involvement with customers (Berthon, Pitt, & Watson, 1996). In addition to providing product information, a Web site can engage visitors in dialogues such as inquiries, suggestions, order status tracking, new product proposal, and online problem diagnostics (Ghose & Dou, 1998). This medium affords a rich collection of formats that are available to the marketer in the presentation of products.

One such format is the personalization (e.g., yahoo.com) and customization (e.g., NikeID.com) offered by informational and corporate portals. Such features as intelligent agents, online help functions, recommender systems, custom-tailored news tickers and personalized home pages made it easy for online firms to service specific shopping needs of consumers (Andre & Rist, 2002). Research showed that a company's willingness to customize its products and services had a positive impact on customer trust in the company both online and off-line (Doney & Cannon, 1997; Koufaris & Hampton-Sosa, 2004).

Interactivity is an important dimension of features that distinguish the Internet from other media. Though the concept of interactivity can be conceptualized in many different ways, defining it as a medium characteristic best reflects the exchange between a Web user and the site (Lombard & Snyder-Duch, 2001). According to Steuer (1992), interactivity is "the extent to which users can participate in modifying the form and content of a mediated environment in real time" (p. 84). Interactive functions allow the user to participate actively in the exchange and persuasion process through direct manipulation of the structural elements of a site (Rodgers & Thorson, 2000). Like television, the interactivity of this medium provides the potential to deliver information in an entertaining form.

Hypermedia is multimedia. Animation, video and audio complement traditional text and graphics when used judiciously. Entertainment supports experiential flow of surfers who engage in "shallow, sensory-level, peripheral processing of the executional aspects of the message" (Singh & Dalal, 1999, p.95). In addition to traditional features such as color, size, typeface, product class, appeal type, animation, audio, sound level, sound clarity, and movement, new design elements like vividness, realism, and interactivity need to be considered in the context of the Web (Rodgers & Thorson, 2000). In a study of 651 companies from Yahoo! and Dutch Yellow Pages, Huizingh (2000) finds that larger sites tend to contain more forms of entertainment features such as pictures, jokes, cartoons, games, and video clips than smaller sites.

A picture is worth a thousand words. Certain types of animation used to display a product from a multi-dimensional perspective and activated upon user request derive benefits at least comparable to large pictures that are often available upon user clicks at most online shopping sites. Animated product displays enhance the directness of user product experience, which has been found to produce more confidently held and more enduring attitudes (Smith & Swinyard, 1983). With animation, product detail can be presented from more angles than with still pictures. Graphics, animation, and video clips might be some of the elements that prevent a casual visitor from simply clicking away from a site.

AMUSEMENT

In addition to product display, animation also provides amusement through imparting motion and activity in products and cartoons that could consequently be perceived as lively, active, or spirited. The use of humor has also been widely explored in advertising as well as computer-mediated communications (Morkes, Kernal, & Nass, 1999; Weinberger & Gulas, 1992). Research has shown that humor attracts attention and seems to be most effective in ads promoting low-risk routine purchases. In a series of field studies, Scott, Klein, and Bryant (1990) found that humorous fliers increased attendance at social events like neighborhood picnics and clambakes, though having little effect on attendance of business gatherings.

When a Web surfer is engaged in low-risk message processing, peripheral cues like animation and humor may be used and can succeed in gaining visitor attention as they do in traditional media (Weinberger, Spotts, Campbell, & Parsons, 1995). A more likable interface will also encourage visitors to visit more pages and stay longer, hence nurturing a more positive attitude.

Humor fulfills audience's needs for escapism, diversion, aesthetic enjoyment, and emotional release. Humor's intention is to amuse and entertain people, to make them laugh, and to convey light-hearted enjoyment through "an incongruous comment that is recognized by the receiver as an attempt to amuse and that succeeds at amusing" (Morkes et al., 1999, p.403). In addition to entertainment, humor also helps gain attention and comprehension, especially when humor is related to the products presented (Speck, 1991).

Web site designers for apparent reasons have examined amusement features in a great length. A first-time visitor judges a site by its look: eye-catching graphics and animations, along with navigation buttons and company logos. Software packages intended to enhance Web site appeal have been developed by many vendors, and software technologies such as Flash® and LiveMotion® have been used by many Web sites as entrance pages for the purpose of entertainment and amusement.

RESEARCH ON EFFECTS OF INTERACTIVITY AND AMUSEMENT

The study of interactive and amusement Web techniques and their effects have been conducted in both advertising and information systems research. This section briefly summarizes the research findings reported in the current literature.

Philport and Arbittier (1997) studied the content of over 2000 commercial communications messages across TV, magazines, newspapers, and the Internet, and found no distinguishing characteristics in Web advertising, that is, banner ads were not particularly different in any dimension. This observation suggests to us that we probably should look at an entire Web site in the study of techniques and effects in the hypermedia. Ghose and Dou (1998) found that greater degrees of interactivity, defined by the total count of interactive functions considered in their research, related positively to Web site appeal.

Coyle and Thorson (2001) argue that interactivity and vividness are the two main factors affecting perceived telepresence, and consequently attitude toward the site. In addition, they found that the presence of image maps, audio and animation influence perceived telepresence and attitude toward the site. Higher perceived interactivity of a Web site also leads to a more positive attitude toward the site (Jee & Lee, 2002).

A user profile study found that, in addition to information, structure and design of a Web site are important factors contributing to better perceptions by the visitor (Eighmey, 1997). Relating to site features, Lohse and Spiller (1998) performed a study measuring 32 user interface features at 28 online retail stores against store traffic and sales. They conclude that online store traffic and sales are influenced by the presence of such interface elements as an FAQ page, promotional activities, and better organization of the product menu. Li and Bukovac (1999) find that the size and animation on a banner ad make a difference in viewer responses. Large animated banners were found to be more attention getting than smaller and static banners and thus assist recall of the advertising. A research addressing the impact of different digital retailing interfaces by Westland and Au (1998) finds that virtual reality storefronts increase a consumer's time spent searching for products but do not significantly increase sales.

In a study linking the use of interruption implemented via pop-up windows, Xia and Sudharshan (2000) manipulated the frequency of interruptions and found that interruptions had a negative impact on consumer shopping experiences. Intrusive formats of advertising like interstitials are found to have "backlash risks" in this new medium (Johnson, Slack, & Keane, 1999).

In traditional humor research, Wells, Leavitt, & McConville (1971) find that along with vigor, personal relevance, and irritation, humor is a major perceptual dimension in their profiling of user responses to TV commercials. The adjectives that accounted for this factor include jolly, merry, playful, humorous and amusing. They concluded that humor contributes to perceived entertainment. From a pool of 600 adjectives, Aaker and Bruzzone (1985) identify four factors that explain the majority of variance in predicting informative, enjoyable, annoying, or offensive ads. In particular, the amusing/humorous variable contributed to 37.6% of the variance explained by the factor "entertainment". In an experimental study of computer-mediated communications, humor was found to enhance the likability of the computer interface, as well as to have a positive effect on user cooperation (Morkes et al., 1999). Animated product displays were found to have a significant effect on perceived entertainment, while the use of humor was found to have a marginally significant effect on perceived entertainment (Gao, 2002).

FUTURE TRENDS

Research in Web marketing and user interface design undertakes both observational and experimental approaches. Observational research has provided a high degree of external validity by taking advantage of the existing real-world Web sites (e.g., Chen & Wells, 1999; Eighmey, 1997; Koufaris, 2002; Lohse & Spiller, 1998). On the other hand, experimental studies perform carefully designed and controlled hypermedia-based stimuli in an attempt to answer the questions of what causal effect a certain Web-based technique may have on perceptual, attitudinal, or behavioral outcomes through between-subjects comparisons (e.g., Coyle & Thorson, 2001; Gao, 2002; Li & Bukovac, 1999; Xia & Sudharshan, 2000). Future research may explore various combinations of presentation attributes like interactive features and humor as fixed factors. Researchers should take advantage of the convenience of Web-based survey techniques, and consider non-intrusive means of collecting usage and exploration data through click streams.

CONCLUSION

In summary, this article provides an overview of the current state of Web interactivity and amusement features and their effects reported in past research. This article advocates a research paradigm that goes beyond observational research and user profiling, and suggests potential future directions in the study of the effects of interactive techniques and amusement features in electronic commerce. Doing business online is constantly evolving. It is necessary to take a rigorous and scientific look at the various components that go into this environment, in order to help electronic business grow in a structured, efficient, and effective way.

REFERENCES

Aaker, D.A., & Bruzzone, D.E. (1985). Causes of irritation in advertising. *Journal of Marketing*, *49*, 47-57.

Andre, E., & Rist, T. (2002). From adaptive hypertext to personalized Web companions. *Communications of the ACM*, *45*(5), 43-46.

Berthon, P., Pitt, L.F., & Watson, R.T. (1996). The World Wide Web as an advertising medium: Toward an understanding of conversion efficiency. *Journal of Advertising Research*, *36*(1), 43-54.

Chen, Q., & Wells, W.D. (1999). Attitude toward the site. *Journal of Advertising Research*, *39*(5), 27-38.

Coyle, J.R., & Thorson, E. (2001). The effects of progressive levels of interactivity and vividness in Web marketing sites. *Journal of Advertising*, *30*(3), 65-77.

Doney, P.M., & Cannon, J.P. (1997). An examination of the nature of trust in buyer-seller relationships. *Journal of Marketing*, *61*(4), 35-51.

Ducoffe, R.H. (1996). Advertising value and advertising *on the Web*. Journal of Advertising Research, *36*(5), 21-34.

Eighmey, J. (1997). Profiling user responses to commercial websites. *Journal of Advertising Research*, *37*(3), 59-66.

Gao, Y. (2002). Effects of humor and realism in hypermedia. In M. Khosrow-Pour (Ed.), *Issues and Trends of Information Technology Management in Contemporary Organizations*. Hershey, PA: Idea Group Publishing.

Ghose, S., & Dou, W. (1998). Interactive functions and their impact on the appeal of the Internet presence sites. *Journal of Advertising Research*, *38*(2), 29-43.

Huizingh, E.K.R.E. (2000). The content and design of Websites: An empirical study. *Information & Management*, *37*, 123-134.

Jee, J, & Lee, W.N. (2002). Antecedents and consequences of perceived interactivity: An exploratory study. *Journal of Interactive Advertising*, *3*(1). Retrieved from *www.jiad.org*

Johnson, M., Slack, M., & Keane, P. (1999). Inside the mind of the online consumer — Increasing advertising effectiveness. *Jupiter Research*. Retrieved from *www.jupiter.com*

Koufaris, M. (2002). Applying the technology acceptance model and flow theory to online consumer behavior. *Information Systems Research*, *13*(2), 205-223.

Koufaris, M., & Hampton-Sosa, W. (2004). The development of initial trust in an online company by new customers. *Information & Management*, *41*(3), 377-397.

Koufaris, M., Kambil, M.A., & Labarbera, P.A. (2001). Consumer behavior in Web-based commerce: An empirical study. *International Journal of Electronic Commerce*, *6*(2), 131-154.

Li, H., & Bukovac, J. L. (1999). Cognitive impact of banner ad characteristics: An experimental study. *Journalism and Mass Communication Quarterly*, *76*(2), 341-353.

Lohse, G.L., & Spiller, P. (1998). Electronic shopping. *Communications of the ACM*, *41*(7). 81-86.

Lombard, M., & Snyder-Duch, J. (2001). Interactive advertising and presence: A framework. *Journal of Interactive Advertising*, 1(2). Retrieved from *www.jiad.org*

Mitra, A. (2003). Cybernetic space: Bringing the virtual and real together. *Journal of Interactive Advertising*, 3(2). Retrieved from *www.jiad.org*

Morkes, J., Kernal, H.K., & Nass, C. (1999). Effects of humor in task-oriented human-computer interaction and computer-mediated communication: A direct test of SRCT theory. *Human-Computer Interaction*, 14, 395-435.

Palmer, J.W. & Griffith, D.A. (1998). An emerging model of website design for marketing. *Communications of the ACM*, *41*(3), 45-51.

Philport, J.C., & Arbittier, J. (1997). Advertising: Brand communications styles in established media and the Internet. *Journal of Advertising Research*, *37*(2), 68-76.

Resnik, A., & Stern, B.L. (1977). An analysis of information content in television advertising. *Journal of Marketing*, *41*(1), 50-53.

Rodgers, S., & Thorson, E. (2000). The interactive advertising model: How users perceive and process online ads. *Journal of Interactive Advertising*, 1(1). Retrieved from *www.jiad.org*

Scott, C., Klein, D.M., & Bryant, J. (1990). Consumer response to humor in advertising: A series of field studies using behavioral observation. *Journal of Consumer Research*, 16(1), 498-501.

Singh, S.N., & Dalal, N.P. (1999). Web homepages as advertisements. *Communications of the ACM*, *42*(8), 91-98.

Smith, R.E., & Swinyard, W.R. (1983). Attitude-behavior consistency: The impact of product trial versus advertising. *Journal of Marketing Research*, *20*, 257-267.

Speck, P.S. (1991). The humorous message taxonomy: a framework for the study of humorous ads. In J.H. Leigh & C.R. Martin, Jr. (Eds.), *Current Issues and Research in Advertising. Ann Arbor*: The University of Michigan.

Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of Communication*, *42*(4), 73-93.

Vijayasarathy, L.R. (2003). Psychographic profiling of the online shopper. Journal of Electronic *Commerce in Organizations*, *1*(3), 48-72.

Weinberger, M.G., & Gulas, C.S. (1992). The impact of humor in advertising: A review. *Journal of Advertising*, *21*(4), 23-59.

Weinberger, M.G., Spotts, H., Campbell, L., & Parsons, A.L. (1995). The use and effect of humor in different advertising media. *Journal of Advertising Research*, *35*(2), 44-56.

Wells, W.D., Leavitt, C., & McConville, M. (1971). A reaction profile for TV commercials. Journal of Advertising, *11*(6), 11-17.

Westland, J.C., & Au, G. (1998). A comparison of shopping experience across three competing digital retailing interfaces. *International Journal of Electronic Commerce*, *2*(2), 57-69.

Xia, L., & Sudharshan, D. (2000). An examination of the effects of cognitive interruptions on consumer online decision processes. Paper presented at *The Second Marketing Science and the Internet Conference*, USC, Los Angeles, CA.

KEY TERMS

Amusement: The state of being amused, entertained, or pleased.

Animation: Technique of imparting motion and activity in graphic images of such objects as products and cartoons to create a condition of being alive, active, or spirited.

Attitude Toward the Site (Ast): A Web user's predisposition to respond either favorably or unfavorably to a Web site in a natural exposure situation.

Entertainment: Something that fulfills a visitor's need for aesthetic enjoyment, escapism, diversion, or emotional release.

Humor: Techniques such as cartoons and jokes that are intended to induce laughter or amusement.

Interactive Advertising: Advertising that simulates a one-on-one interaction to give consumers more control over their experience with product information than do traditional media ads.

Interactivity: A characteristic of a medium in which the user can influence the form and content of the mediated presentation or experience.

Web Marketing: The dissemination of information, promotion of products and services, execution of sales transactions, and enhancement of customer support via a company's Web site.

Interface Design Issues for Mobile Commerce

Susy S. Chan
DePaul University, USA

Xiaowen Fang
DePaul University, USA

INTRODUCTION

Effective interface design for mobile handheld devices facilitates user adoption of mobile commerce (m-commerce). Current wireless technology poses many constraints for effective interface design. These constraints include limited connectivity and bandwidth, diverse yet simplistic devices, the dominance of proprietary tools and languages, and the absence of common standards for application development.

The convergence of mobile Internet and wireless communications has not yet resulted in major growth in mobile commerce. Consumer adoption of m-commerce has been slow even in countries such as Finland, which have broadly adopted wireless technology (Anckar & D'Incau, 2002). An international study of mobile handheld devices and services suggests that mobile commerce is at a crossroads (Jarvenpaa, Lang, Takeda & Tuunainen, 2003). The enterprise and business use of wireless technology holds greater promise, but it demands the transformation of business processes and infrastructure. Poor usability of mobile Internet sites and wireless applications for commerce activities stands out as a major obstacle for the adoption of mobile solutions. For example, even with the latest 3G phones in Japan, consumers still find the small screen display and small buttons on these devices difficult to use (Belson, 2002).

BACKGROUND

Mobile Commerce

Mobile commerce broadly refers to the use of wireless technology, particularly handheld mobile devices and mobile Internet, to facilitate transaction, information search, and user task performance in business-to-consumer, business-to-business, and intra-enterprise communications (Chan & Fang, 2003). Researchers have proposed several frameworks for the study of m-commerce. Varshney and Vetter's framework (2001) presents 12 classes of m-commerce applications, ranging from retail and online shopping, auction, mobile office, and entertainment to mobile inventory emphasizing the potential of mobile B2B and intra-enterprise applications. The framework by Kannan, Chang, and Whinston (2001) groups mobile services into goods, services, content for consumer e-commerce, and activities among trading partners.

Waters (2000) proposes two visions for the potential and opportunities of m-commerce. One perspective argues that the mobile, wireless channel should be viewed as an extension of the current e-commerce channel or as part of a company's multi-channel strategies for reaching customers, employees, and partners. The second, more radical view suggests that m-commerce can create markets and business models.

Recent development in m-commerce has substantiated the first perspective. Major e-commerce sites have implemented their mobile Internet sites as an extension of wired e-commerce to support existing customers (Chan & Lam, 2004; Chan et al., 2002). Consumers have shown relatively low willingness to use m-commerce, but adopters of e-commerce are more likely to embrace this new technology (Anckar & D'Incau, 2002). Furthermore, perceived difficulty of use can affect consumers' choice of m-commerce as a distribution channel (Shim, Bekkering & Hall, 2002). These findings suggest that in a multi-channel environment, m-commerce *supplements* e-commerce instead of becoming a *substitute* for e-commerce.

Enterprise and business applications of m-commerce technologies seem to hold greater promise, because it is easier for companies to standardize and customize applications and devices to enhance current work processes. An Ernst & Young study (2001) of the largest companies in Sweden shows that, except for the retail industry sector, most industries have viewed m-commerce as being vital for growth and efficiency strategies, but not necessarily for generating new revenue. However, integrating the wireless platform in an enterprise requires significant structural transformation and process redesign.

Research on Wireless Interface Design

Several recent studies have examined interface design for mobile applications using handheld devices. Researchers have found that direct access methods were more effective for retrieval tasks with small displays (Jones, Marsden, Mohd-Nasir, Boone & Buchanan, 1999). Novice WAP phone users perform better when using links instead of action screens for navigation among cards, and when using lists of links instead of selection screens for single-choice lists (Chittaro & Cin, 2001). Ramsay and Nielsen (2000) note that many WAP usability problems echo issues identified during the early stage of Web site development for desktop computers, and could be alleviated by applying good user interface design. Such design guidelines for WAP applications include: (1) short links and direct access to content, (2) backward navigation on every card, (3) minimal level of menu hierarchy, (4) reduced vertical scrolling, (5) reduced keystrokes, and (6) headlines for each card (Colafigi, Inverard & Martriccian, 2001; Buchanan et al., 2001). Buyukkokten, Garcia-Molina, and Paepcke (2001) have found that a combination of keyword and summary was the best method for Web browsing on PDA-like handheld devices.

Diverse form factors have different interface requirements. The study by Chan et al. (2002) of 10 wireless Web sites across multiple form factors reveals that user tasks for the wireless sites were designed with steps similar to the wired e-commerce sites, and were primarily geared towards experienced users. Many usability problems, such as long download and broken connections, information overload, and excessive horizontal and vertical scrolling, are common to three form factors—WAP phone, wireless PDA, and Pocket PC. Interface design flaws are platform independent, but the more limitations imposed on the form factors, the more acute the design problems become.

Mobile users access information from different sources and often experience a wide range of network connectivity. Context factors have a particular impact on the usability of mobile applications. Based on a usability study conducted in Korea, three use context factors—hand (one or two hands), leg (walking or stopping), and colocation (alone or with others)—may result in different usability problems (Kim, Kim, Lee, Chae & Choi, 2002). Therefore, the user interface design has to consider various use contexts. Researchers also suggest a systems-level usability approach to incorporating hardware, software, "netware," and "bizware" in the design of user-friendly wireless applications (Palen & Salzman, 2002). Perry, O'Hara, Sellen, Brown, and Harper (2001) have identified four factors in "anytime anywhere" information access for mobile work: the role of planning, working in "dead time," accessing remote technological and informational resources, and monitoring the activities of remote colleagues.

Multimodal interfaces are gaining importance. The MobileGuiding project developed in Spain is aimed at building a European interactive guide network on a common, multimodal, and multilingual platform in which contributors will provide leisure information and cultural events in their locations (Aliprandi et al., 2003). Furthermore, there has been a study conducted in Finland that addresses the design and evaluation of a speech-operated calendar application in a mobile usage context (Ronkainen, Kela & Marila, 2003).

MAIN THRUST OF THIS ARTICLE

Five issues are essential to the interface design for mobile commerce applications, including: (a) technology issues, (b) user goals and tasks, (c) content preparation, (d) application development, and (e) the relationship between m- and e-commerce.

Technology Issues

Limitation of Bandwidth

Most mobile communication standards only support data rates that are less than 28.8 kbps. Connections to the wireless service base stations are unstable because signal strength changes from place to place, especially on the move. These constraints limit the amount of information exchanged between device and base station. Indication of the download progress and friendly recovery from broken connections are necessary to help users gain a better sense of control.

Form Factor

Mobile commerce services are accessible through four common platforms: wireless PDA devices using Palm OS, Pocket PCs running Microsoft Windows CE/Pocket PC OS, WAP phone, and two-way pagers. Within the same platform, different form factors may offer different functionalities. A developer should consider the form factor's unique characteristics when developing m-commerce applications.

User Goals and Tasks

Mobile users can spare only limited time and cognitive resources in performing a task. Services that emphasize mobile values, and time-critical and spontaneous needs, add more value for m-commerce users. These mobile services may include the ability to check flight schedules, check stock prices, and submit bids for auction (Anckar & D'Incau, 2002). In addition, mobile tasks that demonstrate a high level of perceived usefulness, playfulness, and security are the ones most likely to be adopted by users (Fang, Chan, Brzezinski & Xu, 2003).

Content Preparation

Constraints in bandwidth and small screen size demand different design guidelines. Most design guidelines for e-commerce (e.g., Nielsen, Farrell, Snyder & Molich, 2000) support the development of rich product information sets and a complete shopping process. In contrast, wireless Web sites have to simplify their content presentation.

Amount of Information

Content adaptation is necessary to convert information for the mobile Web (Zhou & Chan, 2003). However, users should have sufficient, if not rich, information to accomplish the goals for the application.

Navigation

Navigation systems vary from one form factor to another because the design of handheld devices differs. Currently, there is no consensus on which functions or features should be provided by the application, or built into the device itself.

Depth of Site Structure

Since mobile users have limited time for browsing wireless applications, the organization of information is critical. A flatter structure with fewer steps for wireless applications would allow users to review more options in the same step, and to locate the desired information more quickly.

Graphics or Text

Text is a better choice for displaying information on small screen browsers. However, better technology may improve the screen quality of handheld devices to display more complicated graphics. When determining the format of information to present, it is important to consider the form factor, because it may pose additional constraints on the format.

Development Environment

Mobile computing alters the assumption of "fixed" context of use for interface design and usability testing (Johnson, 1998). Traditional means of user interviews or usability testing in a laboratory environment cannot reveal insights into users' activities and mobility in real life. Contextual consideration is critical for gathering information about user requirements. For example, when developing and testing a mobile application for grocery shoppers, user requirement gathering and prototype evaluation should be conducted in a grocery store (Newcomb, Pashley & Stasko, 2003). The method of contextual inquiry can augment user interface design by exploring the versatility of usage patterns and usage context (Väänänen-Vainio-Mattila & Ruuska, 1998). While contextual inquiry may help developers gain a realistic understanding of contextual factors affecting user behaviors in motion, it is difficult to conduct non-obtrusive observations and inquiries. Developers for mobile applications need to consider the application context surrounding the relationship between the mobile device and user goals and tasks.

Relationship Between M-Commerce and E-Commerce

The wireless channel for e-commerce has raised many new questions regarding coordination between interactions with users across multiple channels. Some researchers suggest that because of the "transaction aware" and "location aware" characteristics of the wireless technology, mobile consumers may increase impulse purchases, especially in low-value, low-involvement product categories, such as books and CDs (Kannan et al., 2001). At present, many Web sites have extended the wireless channel to leverage relationships with exiting customers (Chan et al., 2002). The current state of technology and poor usability of mobile Web sites makes it difficult to expand m-commerce as an independent channel. Many analysts believe that the wireless channel is promising for customer relationship management (CRM) because of its ability to: (1) personalize content and services; (2) track consumers or users across media and over time; (3) provide content and service at the point of need; and (4) provide content with highly engaging characteristics (Kannan et al., 2001). The challenge is how to coordinate interface and content across multiple channels so that experienced users and repeat customers can handle multiple media and platforms with satisfaction.

FUTURE TRENDS

Technology Trends

User interface design for mobile commerce will likely be influenced by four trends. First, multiple standards for wireless communication will not be resolved quickly, especially in North America. Second, the high cost of third-generation (3G) technology may delay the availability of broadband technology for complex functionality and content distribution for mobile applications. Third, instead of the convergence of functionalities into a universal mobile handheld device, there may be a variety of communication devices operating in harmony to support users in their everyday lives. Fourth, input and output format may expand to incorporate voice and other formats, as well as expandable keyboards. The introduction of the voice-based interfaces may complement the text-based interface and remedy some of the information input/display problems of the handheld devices. These trends suggest opportunities to conceptualize wireless user interface beyond text-based interaction. The new challenges are to design better multimodal interfaces for inter-device communication in order to simplify tasks for mobile users.

Development Trends

Alternative methods for interface design and evaluation will be necessary to support m-commerce applications development. First, requirement analysis should focus on the context of mobile users' behaviors and tasks. Contextual inquiry and other methods may be developed to facilitate the understanding of interactions between mobility and usability. Second, usability testing should be conducted with an understanding of contextual variables beyond user behavior. Third, mapping form factors, user tasks, data needs, and content across multiple channels and platforms is necessary to synchronize content and coordinate functionality in a distributed system. Fourth, user-centered design guidelines for mobile applications will be important. These trends require a fresh look at current methodology and will help determine new ways of incorporating user interface design and usability testing for distributed wireless application development. The reference framework proposed by Lee and Benbasat (2003) may be useful in this regard. Their framework incorporates seven design elements for m-commerce interface: context, content, community, customization, communication, connection, and commerce.

M-Commerce Business Models

Wireless technology for m-commerce is likely to evolve in two areas. For intra-enterprise and business-to-business uses, wireless technology provides location-aware and mobility-aware solutions for mobile workers. There is a broad range of possibilities for B2B applications because such deployment can be controlled more easily. Content distribution may be integrated with the enterprise systems. Context-based applications, interfaces, functionality, and even devices can be customized according to the mobile tasks and user groups in the B2B context. This approach makes application development, deployment, and integration easier to manage. In contrast, it is far more challenging to manage the design, development, and deployment of wireless applications for customers. Wireless technology's capability for personalization seems to be the strongest argument for m-CRM services to enhance customer retention (Chan & Lam, 2004). A careful mapping of tasks, data, form factors, and the CRM process is essential for user interface design.

CONCLUSIONS

Wireless technology and the mobile Internet continues to evolve. Until the technology matures and bandwidth improves, wireless applications will be geared toward users requiring limited bandwidth, short exchange of data and text, and simple functionality. Two areas of wireless applications, CRM and enterprise efficiency, may reap greater success. Consumer e-commerce Web sites should focus on the selection of tasks that are most suitable for the wireless channel and demonstrate mobile values, especially for experienced users. Such mapping process requires a solid understanding of the CRM strategy, user preferences, and the constraints imposed by a mobile environment. For enterprise adoption, consolidating the wireless platforms and form factors will facilitate interface design. In either case, additional research to improve usability for mobile commerce is essential.

REFERENCES

Aliprandi, C., Athenour, M., Martinez, S.C., & Patsis, N. (2003). MobileGuiding: A European multimodal and multilingual system for ubiquitous access to leisure and cultural contents. In C. Stephanidis & J. Jacko (Eds.), *Proceedings of the 10th International Conference on Human-Computer Interaction* (vol. 2, pp. 3-7). Mahwah, NJ: Lawrence Erlbaum.

Anckar, B., & D'Incau, D. (2002). Value creation in mobile commerce: Findings from a consumer survey. *Journal of Information Technology Theory & Application, 4*(1), 43-64.

Belson, K. (2002, April 22). Japan is slow to accept the latest phones. *The New York Times,* C4.

Buchanan, G., Farrant, S., Jones, M., Thimbleby, H., Marsden, G., & Pazzani, M. (2001). Improving mobile Internet usability. *Proceedings of the 10th International World Wide Web Conference* (pp. 673-680). New York: ACM Press.

Buyukkokten, O., Garcia-Molina, H., & Paepcke, A. (2001). Seeing the whole in parts: Text summarization for Web browsing on handheld devices. *Proceedings of the 10th*

International World Wide Web Conference. New York: ACM Press.

Chan, S., & Fang, X. (2003). Mobile commerce and usability. In K. Siau & E. Lim (Eds.), *Advances in mobile commerce technologies* (pp. 235-257). Hershey, PA: Idea Group Publishing.

Chan, S., Fang, X., Brzezinski, J., Zhou, Y., Xu, S., & Lam, J. (2002). Usability for mobile commerce across multiple form factors. *Journal of Electronic Commerce Research, 3*(3), 187-199.

Chan, S., & Lam, J. (2004). Customer relationship management on Internet and mobile channels: A framework and research direction. In C. Deans (Ed.), *E-commerce and m-commerce technologies*. Hershey, PA: Idea Group Publishing.

Chittaro, L., & Cin, P.D. (2001). Evaluating interface design choices on WAP phones: Single-choice list selection and navigation among cards. In M.D. Dunlop & S.A. Brewster (Eds.), *Proceedings of Mobile HCI 2001: Third International Workshop on Human Computer Interaction with Mobile Devices*.

Colafigli, C., Inverard, P., & Martriccian, R. (2001). Infoparco: An experience in designing an information system accessible through WEB and WAP interfaces. *Proceedings of the 34th Hawaii International Conference on System Science*. Los Alamitos, CA: IEEE Computer Society Press.

Ernst & Young. (2001). *Global online retailing: An Ernst & Young special report*. Gemini Ernst & Young.

Fang, X., Chan, S., Brzezinski, J., & Xu, S. (2003). A study of task characteristics and user intention to use handheld devices for mobile commerce. *Proceedings of the 2nd Annual Workshop on HCI Research in MIS* (pp. 90-94).

Jarvenpaa, S., Lang, K., Takeda, Y., & Tuunainen, V. (2003). Mobile commerce at crossroads. *Communications of the ACM, 46*(12), 41-44.

Johnson, P. (1998). Usability and mobility: Interactions on the move. In C. Johnson (Ed.), *Proceedings of the 1st Workshop on Human Computer Interaction with Mobile Devices*.

Jones, M., Marsden, G., Mohd-Nasir, N., Boone, K., & Buchanan, G. (1999). Improving Web interaction on small displays. *Computer Networks: The International Journal of Distributed Informatique, 31*, 1129-1137.

Kannan, P., Chang, A., & Whinston, A. (2001). Wireless commerce: Marketing issues and possibilities. *Proceedings of the 34th Hawaii International Conference on System Sciences*. Los Alamitos, CA: IEEE Computer Society Press.

Kim, K., Kim, J., Lee, Y., Chae, M., & Choi, Y. (2002). An empirical study of the use contexts and usability problems in mobile Internet. *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*. Los Alamitos, CA: IEEE Computer Society Press.

Lee, Y. ,& Benbasat, I. (2003). Interface design for mobile commerce. *Communications of the ACM, 46*(12), 49-52.

Newcomb, E., Pashley, T., & Stasko, J. (2003). Mobile computing in the retail arena. *Proceedings of the Conference on Human Factors in Computing Systems, 5*(1), 337-344.

Nielsen, J., Farrell, S., Snyder, C., & Molich, R. (2000). *E-commerce user experience: Category pages*. Nielsen Norman Group.

Palen, L. & Salzman, M. (2002). Beyond the handset: Designing for wireless communications usability. *ACM Transactions on Computer-Human Interaction, 9*(2), 125-151.

Perry, M., O'Hara, K., Sellen, A., Brown, B., & Harper, R. (2001). Dealing with mobility: Understanding access anytime, anywhere. *ACM Transactions on Computer-Human Interaction, 8*(4), 323-347.

Ramsey, M., & Nielsen, J. (2000). *WAP usability: Déjà vu: 1994 all over again*. Nielsen Norman Group.

Ronkainen, S., Kela, J., & Marila, J. (2003). Designing a speech operated calendar application for mobile users. In C. Stephanidis & J. Jacko (Eds.), *Proceedings of the 10th International Conference on Human-Computer Interaction* (vol. 2, pp. 258-262). Mahwah, NJ: Lawrence Erlbaum.

Shim, J.P., Bekkering, E., & Hall, L. (2002). Empirical findings on perceived value of mobile commerce as a distributed channel. *Proceedings of the 8th Americas Conference on Information Systems* (pp. 1835-1837).

Varshney, U., & Vetter, R. (2001). A framework for the emerging mobile commerce applications. *Proceedings of the 34th Hawaii International Conference on System Sciences*. Los Alamitos, CA: IEEE Computer Society Press.

Väänänen-Vainio-Mattila, K., & Ruuska, S. (1998). User needs for mobile communication devices: Requirements gathering and analysis through contextual inquiry. In C. Johnson (Ed.), *Proceedings of the 1st Workshop on Human Computer Interaction with Mobile Devices*.

Waters, R. (2000, March 1). Rival views emerge of wireless Internet. *Financial Times FT-IT Review*, 1.

Zhou, Y., & Chan, S. (2003). Adaptive content delivery over the mobile Web. *Proceedings of the 9th Americas Conference on Information Systems* (pp. 2009-2019).

KEY TERMS

Contextual Inquiry: This interface design method employs an ethnographic approach such as observing user activities in a realistic context.

Fixed Context of Use: Traditional user interface design and testing assumes a single domain, with the users always using the same computer to undertake tasks alone or in collaboration with others.

Form Factor: This platform or operating system runs on a handheld device. Major form factors include Palm, Pocket PC, and WAP.

Interface Design: Design of the interactions between humans and computers.

Location-Aware Service: Mobile services that provide information based on a user's location through the support of a global positioning system. Such services include mobile maps, weather, restaurants, and movie directories.

M-CRM: Interactions between a company and its customers for marketing, sales, and support services through the mobile Web and wireless channel.

Multimodal Interface: An interface that communicates with users through multiple modes.

Usability: Usability refers to how well an application is designed for users to perform desired tasks easily and effectively.

International Digital Studies Approach for Examining International Online Interactions

Kirk St.Amant
Texas Tech University, USA

INTRODUCTION

As global access to the Internet increases, so does the potential for miscommunication in international online interactions (IOIs). Unfortunately, many models for examining cross-cultural communication focus on conventional (offline) interactions or settings. As a result, researchers lack a mechanism for examining how cultural factors could affect online discourse.

This article presents an approach—international digital studies—for examining how cultural factors could affect IOIs. The purpose of this approach is to identify points of contention or areas where online media can create conflicts in cultural expectations associated with credibility. Once identified, these points of contention can serve as the subject of future research related to culture and communication.

BACKGROUND

Creating credibility, or ethos, is not a random process. Rather, audiences use certain factors, or *ethos conditions*, to develop a checklist for determining if a presentation is credible (worthy of attention). That is, audiences come to a particular presentation situation thinking, “This individual must do x, y, and z if I am to consider him or her credible/worth listening to.” If all of these expectations are met (can be “checked off”), then the presenter and his or her ideas will be considered credible. If one or more ethos conditions are not met, then audiences will be less likely to view a presenter as credible (see St.Amant, 2002a, for a more in-depth discussion of this concept).

The ethos conditions one expects to encounter can vary from culture to culture, and such differences have been noted at a variety of levels (Campbell, 1998; Tebeaux, 1999; Lewis, 2001). Persons from different cultures, for example, often use different organizational structures (e.g., stated vs. implied conclusions) and different methods of citing sources to establish the credibility of a presentation (Woolever, 2001; Lewis, 2001; Hofstede, 1997). Cultures can also associate different credibility expectations with sentence length. Southern Europeans, for example, associate longer sentences with credible presentations, while Americans view shorter and more direct sentences as being more credible (Ulijn & Strother, 1995). Additionally, the kind of relationship associated with the use of a particular word can cause cross-cultural credibility problems (Li & Koole, 1998; Li, 1999).

Online media complicate cross-cultural interactions by creating conditions that affect credibility expectations. In many cases, online media reduce human interaction to typed words. Typed online messages, however, tend to follow patterns related to spoken discourse. This mix of written and spoken communication creates a new and interesting situation, for recipients of online text messages do not obtain nonverbal identity cues key to communicating in spoken exchanges. The sender of an online message therefore seems faceless and anonymous (Gauntlett, 2000; St.Amant, 2002b).

As a result, notions of authority, identity, and credibility take on new forms in cyberspace. As Fernback (1999) notes, in online exchanges, the markers of credibility—marks that draw others to listen to you—are not, “brawn, money, or political clout,” but are rather “wit, and tenacity, and intelligence” (p. 213). Thus, wit, tenacity, and intelligence become ethos conditions individuals can use to appear more authoritative or more credible than other participants in an online exchange. These factors therefore become *digital ethos conditions*, for individuals come to expect them when assessing the credibility of online presentations. These digital ethos conditions, however, can conflict with the communication expectations of different cultural groups.

Understanding how cultural factors can affect online exchanges can be a complicated and seemingly overwhelming process. Yet, now that more of the world is getting online, it is becoming increasingly important to understand IOI situations so that miscommunications and mistakes can be avoided. (Such culture-related mistakes, moreover, could affect everything from online social exchanges to international outsourcing and international e-commerce activities.) For this reason, researchers can benefit from an approach that helps them focus their analysis of IOIs on a more manageable set of topics. The international digital studies approach is designed to establish such a focus.

MAIN THRUST OF THE ARTICLE

International digital studies is a research approach used to examine how cultural groups differ in their responses to digital ethos conditions. According to this perspective, the objective of the researcher is two-fold:

- First, the researcher must identify *actual* digital ethos conditions—presentation factors that actually contribute to a presenter's credibility in online exchanges. Once isolated, these digital ethos conditions can become variables used to evaluate how different cultures communicate online.
- Second, the researcher must determine if a digital ethos condition is also a factor that varies in relation to cultural communication expectations. That is, researchers need to determine if cultures would differ observably in how they responded to a particular digital ethos condition.

The key to this line of research becomes identifying variables that could affect communication in IOIs. To achieve this objective, researchers must use a two-part literature review involving the fields of Internet studies and intercultural communication.

The purpose of the dual-field literature review is to determine how factors of medium and of culture might create conflicting expectations of ethos conditions in IOIs. To identify these situations, individuals must first survey the research literature in Internet studies in order to identify digital ethos conditions in direct, two-way interactions online. The focus of this review is to isolate behavior resulting from online communication conditions vs. the transfer of communication patterns from more traditional media to an online setting. Name-dropping, for example, can be used to create credibility in both print and online media; the use of emoticons to create credibility, however, is more restricted to online communication.

After researchers identify digital ethos conditions, they must determine if these factors could cause confusion or conflict in cross-cultural exchanges. The goal then becomes evaluating if a particular digital ethos condition is also a *point of contention*—or a situation in which the communication patterns documented in the literature of one field (Internet studies) conflict with patterns noted in the literature of another field (intercultural communication).

To identify points of contention, the second part of the international digital studies process involves a review of the research literature in intercultural communication. In this second review, the researcher would look specifically for indications that digital ethos conditions identified in the initial (Internet studies) literature review relate to findings reported in the intercultural communication literature. If little or no mention of this factor is made, or if this variable appears to cause no real conflict in cross-cultural exchanges, then that variable would be a *weak* point of contention. If, for example, different cultural groups did not react differently to uses of wit (a key ethos condition noted in Internet studies), then uses of wit would be a weak point of contention, for there is little evidence of different cultural behavior related to this digital ethos condition. If, however, the intercultural literature review reports that the ethos condition noted in the Internet studies literature can cause problems in cross-cultural interactions, then that factor would be a *strong* point of contention that could affect IOIs.

Researchers must next determine if a strong point of contention could actually cause problems in IOIs. That is, just because the two-part literature review indicates a particular ethos condition could be a point of contention.

- Does that ethos condition actually affect discourse in IOIs?
- Do reactions to that ethos condition vary along cultural lines in IOI? (e.g., Do some cultures use it more than others? Are some cultures more confused by its use than others?)
- Can researchers develop a ranking system to compare how specific cultural groups vary in relation to uses of and responses to a particular ethos condition?

To answer these questions, researchers must use strong points of contention as the foundation for experiments that test if and how a strong point of contention can affect IOIs. In this way, the international digital studies approach helps researchers identify suitable topics for conducting further research into IOIs.

FUTURE TRENDS

An application of international digital studies indicates that the concept of identity could be a key problem area in future IOIs. For this reason, it is important that researchers understand how aspects and perceptions of identity could cause problems in online exchanges involving individuals from different cultures.

A review of the Internet studies literature reveals identity is a factor that affects discourse in online forums. Many researchers note that, in cyberspace exchanges, identity is neither fixed nor stable; rather, it can easily change because of online media that reduce interactions to typing words (Gauntlett, 2000; St.Amant, 2002b). By reducing identity to texts, online media allow individuals to create their online identity on their own terms (Arnold & Plymire, 2000).

Other researchers note that by limiting identity to texts, online media allow other users to co-opt or to change someone else's online identity by cutting and pasting another person's words into a different message (Warnick, 1998). Cutting and pasting parts of a message allows individuals to separate what was said from who said it (Warnick, 1998). The ability to separate authors from their original online message gives presenters a great deal of liberty, for they can:

- attribute segments of forwarded messages to whomever they wish, thus altering the online identity of someone else;
- create new identities for themselves by co-opting the words or ideas of another;
- alter what another said, creating a new online identity for that author by "putting words in that person's mouth."

In all three cases, altering textual factors allows one individual to change the online identity of another person. The question then becomes: Could this plasticity of online identity result in reactions that vary along cultural lines? To answer this question, the researcher must review the intercultural communication literature in search of information relating to identity.

A review of the intercultural communication literature reveals that a fixed or verifiable identity is often essential to creating credible messages (Weiss, 1998; Ferraro, 2002; Ng & Van Dyne, 2001). In some cultures, one's identity is not based on the claims or the proof that the individual himself or herself presents, but is rather based on the claims of others. For example, in certain cultures, people interact within relatively large and complex social networks. These networks are often formed from long-term relationships developed between individuals over time, or from strong familial ties based on trust and senses of family duty and family honor (Hofstede, 1997; Weiss, 1998; Richmond, 1995).

In social network cultures, the identity of the presenter often determines if others will listen to or ignore information presented by that individual (Hofstede, 1997; Weiss, 1998; Richmond, 1995). The distinction is essentially one of in-group vs. out-group. If one can be identified as a member of the in-group as confirmed by someone else in the social network, then certain behaviors are expected and a certain level of credibility is awarded. Additionally, a special authority—or credibility—is given to the information that "identified" individual presents. If one is viewed as a member of the out-group (his or her identity cannot be confirmed by someone else in the social network), different behaviors are expected and different, more restricted levels of trust and disclosure are granted.

Also, in social network cultures, outsiders (members of a different culture) tend to be viewed with suspicion, and being "heard" or "listened to" often becomes a matter of having the proper introduction (Richmond, 1995; Hofstede, 1997; Scharf & MacMathuna, 1998). In such systems, if a person who is part of the network (has a known identity) says an outsider should be listened to, then that outsider is identified as "credible" by other members of the network. In this case, the identity of the outsider gains credibility by being associated with a particular individual who is known and is trusted by members of that system (Hofstede, 1997; Scharf & MacMathuna, 1998; Weiss, 1998). Success in such a cultural communication system becomes a matter of identity. That is, does the recipient of a message know the identity of the presenter and how do factors of identity affect perceptions of credibility?

Social network perspectives on identity might affect how individuals from such cultures use online communication technologies. For example, when an individual from a social network culture receives an e-mail message that lacks identity cues, will that individual trust the identity of the sender? Moreover, if individuals from such cultures (cultures in which a person's word is often that person's bond) see how easily online messages can be altered and reposted, would those individuals be willing to use online media to introduce new people or present new ideas? And how would resistance to disclosing information in cyberspace be perceived by individuals from cultures in which a fixed identity is not as important?

According to such cultural expectations, the identity of the presenter affects his or her credibility. This perceived credibility affects the degree of access an individual has to certain kinds of information in social network cultures. Persons with a solid and an easy-to-confirm identity are considered credible and gain access to information. Individuals with a limited/text-based identity, however, might be viewed as non-credible and would be limited in the information they could obtain. In such cases, identity = trust/credibility, and lack of identity = doubt.

The limited identity resulting from IOIs could therefore undermine the credibility of some participants in IOIs. This situation could lead to "unexpected" behaviors or reactions from members of social network cultures. (These behaviors might include ignoring important communiqués because the identity, and thus the credibility, of the sender is "suspect.") Such unexpected behavior could, in turn, lead to confusion or even offense.

This two-part literature review reveals instances where cultural perceptions of identity could cause problems in

IOIs. As a result, identity appears to be a strong point of contention that merits further examination. From this point, researchers can use different methods to examine how this strong point of contention might affect discourse patterns in IOIs. For example, some researchers might conduct online case studies to observe how individuals from different cultural groups act and interact in online environments where identity remains restricted to texts. Other researchers might conduct controlled experiments in which different cultural groups interact in situations where identity is known and stable, and in situations where identity is limited and text based.

The purpose of these approaches would be to determine if identity actually affects how cultural groups interact in IOIs. Such research could also be used to establish a comparative system for determining how specific cultures differ according to this particular point of contention. The overall objective would therefore be to determine if a strong point of contention could allow for a relativistic comparison of cultural behavior according to a similar concept. The results of such research could provide important insights that could be used to shape everything from international e-commerce strategies to communication protocols used in international outsourcing projects.

CONCLUSION

The rapid evolution of online communication technologies is constantly changing how people think about space and time. Now, international communication often transpires in seconds or minutes, not days or weeks. This new degree of proximity, however, could lead to an increase in cross-cultural misunderstanding as many aspects of online interactions contradict the communication expectations of certain cultures. This essay has overviewed how the international digital studies approach can serve as a foundation for examining international online interactions. It now becomes the task of researchers to examine culture, technology, and communication in order to explore the true nature of IOIs.

REFERENCES

Arnold, E.A. & Plymire, D.C. (2000). The Cherokee Indians and the Internet. In D. Gauntlett (Ed.), *Web.Studies* (pp. 186-193). New York: Oxford University Press.

Campbell, C.P. (1998). Rhetorical ethos: A bridge between high-context and low-context cultures? In S. Niemeier, C.P. Campbell & R. Dirven (Eds.), *The cultural context in business communication* (pp. 31-47). Philadelphia: John Benjamins.

Fernback, J. (1999). There is a there there: Notes toward a definition of cybercommunity. In S. Jones (Ed.), *Doing Internet research* (pp. 203-220). Thousand Oaks, CA: Sage Publications.

Ferraro, G. (2002). *Global brains: Knowledge and competencies for the 21st century*. Charlotte, NC: Intercultural Associates.

Gauntlett, D. (2000). Web studies: A user's guide. In D. Gauntlett (Ed.), *Web.Studies* (pp. 2-18). New York: Oxford University Press.

Hofstede, G. (1997). *Culture and organizations: Software of the mind*. New York: McGraw-Hill.

Lewis, R. (2003). *The cultural imperative: Global trends in the 21st century*. Yarmouth, ME: Intercultural Press.

Li, X. (1999). *Chinese-Dutch business negotiations: Insights from discourse*. Atlanta, GA: Rodopi.

Li, X. & Koole, T. (1998). Cultural keywords in Chinese-Dutch business negotiations. In S. Niemeier, C.P. Campbell & R. Dirven (Eds.), *The cultural context in business communication* (pp. 185-213). Philadelphia: John Benjamins.

Ng, K.Y. & Van Dyne, L. (2001). Culture and minority influence: Effects on persuasion and originality. In C.K.W. De Dreu & M.K. De Vries (Eds.), *Group consensus and minority influence: Implications for innovation* (pp. 284-306). Malden, MA: Blackwell Publishers.

Richmond, Y. (1995). *From da to yes: Understanding East Europeans*. Yarmouth, ME: Intercultural Press.

St.Amant, K. (2002a). International digital studies: A research approach for examining international online interactions. In E. Buchanan (Ed.), *Virtual research ethics: Issues and controversies* (pp. 317-337). Hershey, PA: Idea Group Publishing.

St.Amant, K. (2002b). When cultures and computers collide. *Journal of Business and Technical Communication, 16*(2) 196-214.

Scharf, W.F. & MacMathuna, S. (1998). Cultural values and Irish economic performance. In S. Niemeier, C.P. Campbell & R. Dirven (Eds.), *The cultural context in business communication* (pp. 145-164). Philadelphia: John Benjamins.

Tebeaux, E. (1999). Designing written business communication along the shifting cultural continuum: The new face of Mexico. *Journal of Business and Technical Communication, 13*, 49-85.

Ulijn, J.M. & Strother, J.B. (1995). *Communicating in business and technology: From psycholinguistic theory to international practice*. Frankfurt, Germany: Peter Lang.

Ulijn, J.M. & Campbell, C.P. (1999). Technical innovations in communication: How to relate technology to business by a culturally reliable human interface. *Proceedings of the 1999 IEEE International Professional Communication Conference* (pp. 109-120). Piscataway, NJ: IEEE Professional Communication Society.

Warnick, B. (1998). Rhetorical criticism of public discourse on the Internet: Theoretical implications. *Rhetoric Society Quarterly, 28,* 73-84.

Weiss, S.E. (1998). Negotiating with foreign business persons: An introduction for Americans with propositions on six cultures. In S. Niemeier, C.P. Campbell & R. Dirven (Eds.), *The cultural context in business communication* (pp. 51-118). Philadelphia: John Benjamins.

Woolever, K.R. (2001). Doing global business in the information age: Rhetorical contrasts in the business and technical professions. In C.G. Paneta (Ed.), *Contrastive rhetoric revisited and redefined* (pp. 47-64). Mahwah, NJ: Lawrence Erlbaum Associates.

KEY TERMS

Cross-Cultural: Situations where individuals from different cultures interact with or exchange information with one another; interchangeable with the term "intercultural."

Digital Ethos Conditions: Factors individuals use to assess the credibility or the worth of an online presentation of information.

Ethos Conditions: Factors individuals use to asses the credibility or the worth of a presentation.

Intercultural: Situations where individuals from different cultures interact with or exchange information with one another; interchangeable with the term "cross-cultural."

International Online Interaction (IOI): Situation in which individuals from two or more cultures use an online medium to interact directly with one another.

Literature Review: Reviewing the research findings reported by a certain field of research; the process usually involves the examination of articles published in the research journals of a particular field.

Point of Contention: A communication factor that is found to be positive (contribute to one's credibility) in the research of one field, but is found to be negative (detract from one's credibility) in the research of another field.

Presentation: The sharing of information with other via spoken, written, or online media.

Internet Abuse and Addiction in the Workplace

Mark Griffiths
Nottingham Trent University, UK

INTRODUCTION

As with the introduction of other mass communication technologies, issues surrounding Internet use, abuse and addiction have surfaced. This article has a number of objectives. It will first introduce readers to the concept of Internet addiction before going on to look at the wider issue of Internet abuse in the workplace. In this section, generic types of Internet abuse will be described, in addition to further examination of the reasons why Internet abuse occurs. The chapter ends with some guidelines and recommendations for employers and human resources departments.

BACKGROUND: INTERNET ADDICTION

There have been a growing number of academic papers about excessive use of the Internet. These can roughly be divided into four categories:

- Studies that compare excessive Internet users with non-excessive users (e.g., Brenner, 1997; Young, 1998)
- Studies that have examined vulnerable groups of excessive Internet use; for example, students (e.g., Nalwa & Anand, 2003; Scherer & Bost, 1997)
- Case studies of excessive Internet users (Catalano, Catalano, Embi & Frankel, 1999; Griffiths, 2000a; Tsai & Lin, 2003; Young, 1996)
- Studies that examine the psychometric properties of excessive Internet use (e.g., Armstrong, Phillips & Saling, 2000; Charlton, 2002; Pratarelli et al., 1999).
- Studies examining the relationship of excessive Internet use with other behaviors; for example, psychiatric problems, depression, loneliness, academic performance and so forth (e.g., Kubey, Lavin & Barrows, 2001; Nie & Ebring, 2000; Shapira, Goldsmith, Keck, Khosla & McElroy, 2000)

Despite the predominance of drug-based definitions of addiction, there is now a growing movement that views a number of behaviors as potentially addictive, including those which do not involve the ingestion of a psychoactive drug (e.g., gambling, computer game playing, exercise, sex, and now the Internet) (Orford, 2001). Research has suggested that social pathologies are beginning to surface in cyberspace. These have been termed "technological addictions" (Griffiths, 1996a) and have been operationally defined as non-chemical (behavioral) addictions that involve excessive human-machine interaction. They can thus be viewed as a subset of behavioral addictions (Marks, 1990) and feature core components of addiction (Brown, 1993; Griffiths, 1996a); that is, salience, mood modification, tolerance, withdrawal, conflict and relapse. Young (1999) claims Internet addiction is a broad term that covers a wide variety of behaviors and impulse control problems. This is categorized by five specific subtypes:

- Cybersexual addiction: compulsive use of adult Web sites for cybersex and cyberporn
- Cyber-relationship addiction: over-involvement in online relationships
- Net compulsions: obsessive online gambling, shopping or day-trading
- Information overload: compulsive Web surfing or database searches.
- Computer addiction: obsessive computer game playing (e.g., Doom, Myst, Solitaire, etc.)

In reply to Young, Griffiths (2000a) has argued that many of these excessive users are not "Internet addicts" but just use the Internet excessively as a medium to fuel other addictions. Put very simply, a gambling addict or a computer game addict who engages in their chosen behavior online is not addicted to the Internet. The Internet is just the place where they engage in the behavior. However, in contrast to this, there are case study reports of individuals who appear to be addicted to the Internet itself (e.g., Young, 1996; 2000b). These are usually people who use Internet chat rooms or play fantasy role playing games - activities that they would not engage in except on the Internet itself. These individuals to some extent are engaged in text-based virtual realities and take on other social personas and social identities as a way of feeling good about themselves.

In these cases, the Internet may provide an alternative reality to the user and allow them feelings of immersion and anonymity that may lead to an altered state of consciousness. This in itself may be highly psychologically and/or physiologically rewarding. There are many factors that make the Internet seductive. It is clear from research in the area of computer-mediated communication that virtual environments have the potential to provide short-term comfort, excitement, and/or distraction (Griffiths, 2000a). These reasons alone provide compelling reasons alone why employees may engage in non-work related Internet use. There are also other reasons that are outlined in more detail in the next section on Internet abuse.

Case study accounts (e.g., Griffiths, 2000b; Tsai & Lin, 2003; Young, 1996) have shown that the Internet can be used to counteract other deficiencies in the person's life (e.g., relationships, lack of friends, physical appearance, disability, coping, etc.). Internet addiction appears to be a bona fide problem to a small minority of people but evidence suggests the problem is so small that few employers take it seriously. It may be that Internet abuse (rather than Internet addiction) is the issue that employers should be more concerned about.

TYPES OF WORKPLACE INTERNET ABUSE

It is clear that the issue of Internet abuse and Internet addiction are related but they are not the same thing. Furthermore, the long-term effects of Internet abuse may have more far-reaching effects for the company that the Internet abuser works for than the individual himself or herself. Abuse also suggests that there may not necessarily be any negative effects for the user other than a decrease in work productivity.

As seen in the previous section, Young (1999) claims Internet addiction is a broad term that covers a wide variety of behaviors and impulse control problems categorized by five specific subtypes. These can be adapted and refined to produce a typology of Internet abuse within the workplace. These are cybersexual Internet abuse, online friendship/relationship abuse, Internet activity abuse, online information abuse, criminal Internet abuse, and miscellaneous Internet abuse. These are examined in more detail below.

- Cybersexual Internet abuse: this involves the abuse of adult Web sites for cybersex and cyberporn during work hours. Such online sexual services include the conventional (e.g., Internet versions of widely available pornographic magazines like Playboy), the not so conventional (Internet versions of very hardcore pornographic magazines) and what can only be described as the bizarre (various discussion groups). There are also pornographic picture libraries (commercial and free-access), videos and video clips, live strip shows, live sex shows and voyeuristic Web-cam sites (Cooper, 2000; Griffiths, 2001).
- Online friendship/relationship abuse: this involves the conducting of an online friendship and/or relationship during work hours. Such a category could also include the use of e-mailing friends and/or engaging in discussion groups, as well as maintenance of online emotional relationships. Such people may also abuse the Internet by using it to explore gender and identity roles by swapping gender or creating other personas and forming online relationships or engaging in cybersex (see above) (Griffiths, 2001; Whitty, 2003).
- Internet activity abuse: this involves the use of the Internet during work hours in which other non-work related activities are done (e.g., online gambling, online shopping, online travel booking, online computer gaming, online day-trading, etc.). This may be one of the most common forms of Internet abuse in the workplace.
- Online information abuse: this involves the abuse of Internet search engines and databases. Typically, this involves individuals who search for work-related information on databases and so forth but who end up wasting hours of time with little relevant information gathered. This may be deliberate work-avoidance but may also be accidental and/or non-intentional. It may also involve people who seek out general educational information, information for self-help/diagnosis (including online therapy) and/or scientific research for non-work purposes.
- Criminal Internet abuse: this involves seeking out individuals who then become victims of sexually-related Internet crime (e.g., online sexual harassment, cyberstalking, paedophilic "grooming" of children). The fact that these types of abuse involve criminal acts may have severe implications for employers.
- Miscellaneous Internet abuse: this involves any activity not found in the above categories such as the digital manipulation of images on the Internet for entertainment and/or masturbatory purposes (e.g., creating celebrity fake photographs where heads of famous people are superimposed onto someone else's naked body) (Griffiths, 2001).

WHY DOES INTERNET ABUSE OCCUR?

There are many factors that make Internet abuse in the workplace seductive. It is clear from research in the area of computer-mediated communication that virtual environments have the potential to provide short-term comfort, excitement, and/or distraction (Griffiths, 2000a). These reasons alone provide compelling reasons why employees may engage in non-work related Internet use. There are also other reasons (opportunity, access, affordability, anonymity, convenience, escape, disinhibition, social acceptance, and longer working hours), which are briefly examined next:

- Opportunity and access – Obvious pre-cursors to potential Internet abuse include both opportunity and access to the Internet. Clearly, the Internet is now commonplace and widespread, and is almost integral to most workplace environments. Given that prevalence of undesirable behaviors is strongly correlated with increased access to the activity, it is not surprising that the development of Internet abuse appears to be increasing across the population.
- Affordability - Given the wide accessibility of the Internet, it is now becoming cheaper and cheaper to use the online services on offer. Furthermore, for almost all employees, Internet access is totally free of charge and the only costs will be time and the financial costs of some particular activities (e.g., online sexual services, online gambling, etc.).
- Anonymity - The anonymity of the Internet allows users to privately engage in their behaviors of choice in the belief that the fear of being caught by their employer is minimal. This anonymity may also provide the user with a greater sense of perceived control over the content, tone, and nature of their online experiences. The anonymity of the Internet often facilitates more honest and open communication with other users and can be an important factor in the development of online relationships that may begin in the workplace. Anonymity may also increase feelings of comfort since there is a decreased ability to look for, and thus detect, signs of insincerity, disapproval, or judgment in facial expression, as would be typical in face-to-face interactions.
- Convenience - Interactive online applications such as e-mail, chat rooms, newsgroups, or role-playing games provide convenient mediums to meet others without having to leave one's work desk. Online abuse will usually occur in the familiar and comfortable environment of home or workplace, thus reducing the feeling of risk and allowing even more adventurous behaviors.
- Escape - For some, the primary reinforcement of particular kinds of Internet abuse (e.g., to engage in an online affair and/or cybersex) is the sexual gratification they experience online. In the case of behaviors like cybersex and online gambling, the experiences online may be reinforced through a subjectively and/or objectively experienced "high". The pursuit of mood-modifying experiences is characteristic of addictions. The mood-modifying experience has the potential to provide an emotional or mental escape and further serves to reinforce the behavior. Abusive and/or excessive involvement in this escapist activity may lead to problems (e.g., online addictions). Online behavior can provide a potent escape from the stresses and strains of real life.
- Disinhibition – Disinhibition is clearly one of the Internet's key appeals, as there is little doubt that the Internet makes people less inhibited (Joinson, 1998). Online users appear to open up more quickly online and reveal themselves emotionally much faster than in the offline world. What might take months or years in an offline relationship may only takes days or weeks online. As some have pointed out (e.g., Cooper & Sportolari, 1997), the perception of trust, intimacy and acceptance has the potential to encourage online users to use these relationships as a primary source of companionship and comfort.
- Social acceptability – The social acceptability and perception of being online has changed over the last 10 years (e.g., the "nerdish" image of the Internet is almost obsolete). It may also be a sign of increased acceptance as young children are exposed to technology earlier and so become used to socializing using computers as tools. For instance, laying the foundations for an online relationship in this way has become far more socially acceptable and will continue to be so. Internet interaction takes away the social isolation that we can all sometimes feel and there are no boundaries of geography, class or nationality.

FUTURE TRENDS: GUIDELINES FOR MANAGERS AND HUMAN RESOURCES DEPARTMENTS

As has been demonstrated, being able to spot someone who is an Internet addict or an Internet abuser can be very difficult. However, there are some practical steps that can be taken to help minimize the potential problem.

- Develop an "Internet Abuse At Work" policy. Many organizations have policies for behaviors such as smoking or drinking alcohol. Employers should develop their own Internet abuse policies by liaison between personnel services and local technology councils and/or health and safety executives.

- Take the issue of Internet abuse/addiction seriously. Internet abuse and addiction in all their varieties are only just being considered as potentially serious occupational issues. Managers, in conjunction with personnel departments, need to ensure they are aware of the issues involved and the potential risks it can bring to both their employees and the whole organization. They also need to be aware that for employees who deal with finances, the consequences of some forms of Internet abuse/addiction (e.g., Internet gambling) for the company can be very great.
- Raise awareness of Internet abuse/addiction issues at work. This can be done through e-mail circulation, leaflets, and posters on general notice boards. Some countries will have national and /or local agencies (e.g., technology councils, health and safety organizations, etc.) that can supply useful educational literature (including posters). Telephone numbers for these organizations can usually be found in most telephone directories.
- Ask employees to be vigilant. Internet abuse/addiction at work can have serious repercussions not only for the individual but also for those employees who befriend Internet abusers and addicts, and the organization itself. Fellow staff need to know the basic signs and symptoms of Internet abuse and addiction. Employee behaviors such as continual use the Internet for non-work purposes might be indicative of an Internet abuse problem.
- Give employees access to diagnostic checklists. Make sure that any literature or poster within the workplace includes a self-diagnostic checklist so that employees can check themselves to see if they might have (or be developing) an Internet problem.
- Monitor Internet use of your staff that you suspect may have problems. Those staff with an Internet-related problem are likely to spend great amounts of time engaged in non-work activities on the Internet. Should an employer suspect such a person, they should get the company's IT specialists to look at their Internet surfing history, as the computer's hard disc will have information about everything they have ever accessed. The fact that specific individuals may be monitored should be outlined in the organization's "Internet Abuse At Work" policy so that employees are aware they may be monitored.
- Check Internet "bookmarks" of your staff. In some jurisdictions across the world, employers can legally access the e-mails and Internet content of their employees. One simple check is to simply look at an employee's list of "bookmarked" Web sites. If they are spending a lot of employment time engaged in non-work activities, many bookmarks will be completely non-work related (e.g., online dating agencies, gambling sites).
- Give support to identified problem users. Most large organizations have counseling services and other forms of support for employees who find themselves in difficulties. In some (but not all) situations, problems associated with Internet use need to be treated sympathetically (and like other more bona fide addictions such as alcoholism). Employee support services must also be educated about the potential problems of Internet abuse and addiction in the workplace.

CONCLUSION

In this chapter, major issues that surround Internet abuse/addiction issues in the workplace have been highlighted. Internet abuse/addiction can clearly be a hidden activity and the growing availability of Internet facilities in the workplace is making it easier for abuse to occur in lots of different forms. Thankfully, it would appear that for most people Internet abuse is not a serious individual problem, although for large companies, small levels of Internet abuse multiplied across the workforce raises serious issues about work productivity. For those whose Internet abuse starts to become more of a problem, it can affect many levels including the individual, their work colleagues and the organization itself.

Managers clearly need to have their awareness of this issue raised, and once this has happened, they need to raise awareness of the issue among the work force. Knowledge of such issues can then be applied individually to organizations in the hope that they can develop an Internet abuse policy in the same way that many organizations have introduced smoking and alcohol policies. Furthermore, employers need to let employees know exactly which behaviors on the Internet are reasonable (e.g., the occasional e-mail to a friend) and those that are unacceptable (e.g., online gaming, cybersex, etc.). Internet abuse has the potential to be a social issue, a health issue *and* an occupational issue and needs to be taken seriously by all those employers who utilize the Internet in their day-to-day business.

REFERENCES

Armstrong, L., Phillips, J.G., & Saling, L. (2000). Potential determinants of heavier Internet usage. *International Journal of Human-Computer Studies, 53*, 537-550.

Brenner, V. (1997). Psychology of computer use: XLVII. Parameters of Internet use, abuse and addiction: The first 90 days of the Internet usage survey. *Psychological Reports, 80,* 879-882.

Brown, R.I.F. (1993). Some contributions of the study of gambling to the study of other addictions. In W.R. Eadington & J.A. Cornelius (Eds.), *Gambling behavior and problem gambling* (pp. 241-272). Reno: University of Nevada Press.

Catalano, G., Catalano, M.C., Embi, C.S., & Frankel, R.L. (1999). Delusions about the Internet. *Southern Medical Journal, 92,* 609-610.

Charlton, J.P. (2002). A factor analytic investigation of computer ‘addiction’ and engagement. *British Journal of Psychology, 93,* 329-344.

Cooper, A. (Ed.). *Cybersex: The dark side of the force* (pp. 5-29). Philadelphia: Brunner Routledge.

Cooper, A., & Sportolari, L. (1997). Romance in cyberspace: Understanding online attraction. *Journal of Sex Education and Therapy, 22,* 7-14.

Griffiths, M.D. (1996a). Behavioural addictions: An issue for everybody? *Journal of Workplace Learning, 8*(3), 19-25.

Griffiths, M.D. (1996b). Internet “addiction”: An issue for clinical psychology? *Clinical Psychology Forum, 97,* 32-36.

Griffiths, M.D. (2000a). Internet addiction: Time to be taken seriously? *Addiction Research, 8,* 413-418.

Griffiths, M.D. (2000b). Does Internet and computer “addiction” exist? Some case study evidence. *CyberPsychology and Behavior, 3,* 211-218.

Griffiths, M.D. (2001). Sex on the Internet: Observations and implications for Internet sex addiction. *Journal of Sex Research, 38,* 333-342.

Joinson, A. (1998). Causes and implications of disinhibited behavior on the Internet. In J. Gackenback (Ed.), *Psychology and the Internet: Intrapersonal, interpersonal, and transpersonal implications* (pp. 43-60). New York: Academic Press.

Kubey, R.W., Lavin, M.J., & Barrows, J.R (2001, June). Internet use and collegiate academic performance decrements: Early findings. *Journal of Communication,* 366-382.

Marks, I. (1990). Non-chemical (behavioural) addictions. *British Journal of Addiction, 85,* 1389-1394.

Nalwa, K., & Anand, A.P. (2003). Internet Addiction in students: A cause of concern. *CyberPsychology and Behavior, 6,* 653-656.

Nie, N.H., & Ebring, L. (2000). *Internet and society: A preliminary report.* Stanford, CA: The Institute for the Quantitative Study of Society.

Orford, J. (2001). *Excessive appetites: A psychological view of the addictions* (2nd ed.). Chichester: Wiley.

Pratarelli, M.E., Browne, B.L., & Johnson, K. (1999). The bits and bytes of computer/Internet addiction: A factor analytic approach. *Behavior Research Methods, Instruments and Computers, 31,* 305-314.

Scherer, K., & Bost, J. (1997, August). *Internet use patterns: Is there Internet dependency on campus?* Paper presented at the 105th Annual Convention of the American Psychological Association, Chicago, IL.

Shapira, N.A., Goldsmith, T.D., Keck, P.E., Khosla, U.M., & McElroy, S.L. (2000). Psychiatric features of individuals with problematic Internet use. *Journal of Affective Disorders, 57,* 267-272.

Tsai, C.-C., & Lin, S.S.J. (2003). Internet addiction of adolescents in Taiwan: An interview study. *CyberPsychology and Behavior, 6,* 649-652.

Whitty, M.T. (2003). Pushing the wrong buttons: Men’s and women’s attitudes toward online and offline infidelity. *CyberPsychology and Behavior, 6,* 569-579.

Young, K. (1996). Psychology of computer use: XL. Addictive use of the Internet: A case that breaks the stereotype. *Psychological Reports, 79,* 899-902.

Young, K. (1998). Internet addiction: The emergence of a new clinical disorder. *CyberPsychology and Behavior, 1,* 237-244.

Young, K. (1999). Internet addiction: Evaluation and treatment. *Student British Medical Journal, 7,* 351-352.

KEY TERMS

Conflict: This refers to the conflicts between the addict and those around them (interpersonal conflict), conflicts with other activities (job, social life, hobbies and interests) or from within the individual themselves (intrapsychic conflict) that are concerned with the particular activity.

Cybersex: The act of computer-mediated sex either in an online or virtual environment. Examples include two consenting adults engaging in an e-mail or real-time chat sex session. The advantages to this are that two people who are at opposite ends of the globe can maintain a relationship.

Internet Addiction: This is a term used to describe excessive Internet use and has been also been referred to as Internet addiction disorder (IAD), Internet addiction syndrome (IAD) and pathological Internet use. As with other addictions, Internet addiction features the core components of other addictive behaviors (salience, mood modification, tolerance, withdrawal, conflict and relapse) and can be defined as a repetitive habit pattern that increases the risk of disease and/or associated personal and social problems. It is often experienced subjectively as "loss of control" and these habit patterns are typically characterized by immediate gratification (short-term rewards), often coupled with delayed, deleterious effects (long-term costs). Attempts to change an addictive behavior (via treatment or by self-initiation) are typically marked by high relapse rates (see also technological addictions).

Mood Modification: This refers to the subjective experiences that people report as a consequence of engaging in the particular activity and can be seen as a coping strategy (i.e., they experience an arousing "buzz" or a "high" or paradoxically, tranquilizing feel of "escape" or "numbing").

Relapse: This is the tendency for repeated reversions to earlier patterns of the particular activity to recur and for even the most extreme patterns typical of the height of the addiction to be quickly restored after many years of abstinence or control.

Salience: This occurs when the particular activity becomes the most important activity in the person's life and dominates their thinking (preoccupations and cognitive distortions), feelings (cravings) and behavior (deterioration of socialized behavior). For instance, even if the person is not actually engaged in the behavior they will be thinking about the next time they will be.

Technological Addictions: These addictions are operationally defined as non-chemical (behavioral) addictions that involve human-machine interaction. They can either be passive (e.g., television) or active (e.g., computer games, Internet), and usually contain inducing and reinforcing features which may contribute to the promotion of addictive tendencies. Technological addictions can be viewed as a subset of behavioral addictions and feature core components of addiction, that is, salience, mood modification, tolerance, withdrawal, conflict and relapse.

Tolerance: This is the process whereby increasing amounts of the particular activity are required to achieve the former effects. For instance, a gambler may have to gradually have to increase the size of the bet to experience a euphoric effect that was initially obtained by a much smaller bet.

Withdrawal Symptoms: These are the unpleasant feeling states and/or physical effects that occur when the particular activity is discontinued or suddenly reduced, for example, the shakes, moodiness, irritability and so forth.

Internet Data Mining Using Statistical Techniques

Kuldeep Kumar
Bond University, Australia

INTRODUCTION

Data mining has emerged as one of the hottest topics in recent years. It is an extraordinarily broad area and is growing in several directions. With the advancement of the Internet and cheap availability of powerful computers, data is flooding the market at a tremendous pace. However, the technology for navigating, exploring, visualising, and summarising large databases is still in its infancy. Internet data mining is the process of collecting, analysing, and decision making while the data is being collected on the Internet. In most of the financial applications, data is updated every second or minute or hour. By using Internet data-mining tools, a decision can be made as soon as data are updated.

The quantity and diversity of data available to make decisions have increased dramatically during the past decade. Large databases are being built to hold and deliver these data. Data mining is defined as the process of seeking interesting or valuable information within large data sets. Some examples of data mining applications in the area of management science are analysis of direct mailing strategies, sales data analysis for customer segmentation, credit card fraud detection, mass customization, and so forth. With the advancement of the Internet and World Wide Web, both management scientists and interested end users can get large data sets for their research from this source. The Web not only contains a vast amount of useful information, but also provides a powerful infrastructure for communication and information sharing. For example, Ma, Liu, and Wong (2000) have developed a system called DS-Web that uses the Web to help data mining. A recent survey on Web-mining research can be seen in the paper by Kosala and Blockeel (2000).

Both statistics and data mining are concerned with drawing inferences from data. The aim of inference may be to understand the patterns of correlation and causal links among the data values (explanation) or making predications for the future data values (generalization). At present, data-mining practitioners and statisticians seem to have different approaches to solving problems of a similar nature. It appears that statisticians and data miners can profit by studying each other's methods and using a judiciously chosen combination of them.

Data-mining techniques can be broadly classified into four areas.

1. **Exploratory Data Analysis (EDA) and Inferential Techniques:** As opposed to traditional hypothesis testing designed to verify an a priori hypothesis about relations between variables, EDA is used to identify systemic relationships between variables when there are no a priori expectations as to the nature of those relations. Computational EDA includes both simple and basic statistics and more advanced, multivariate exploratory techniques designed to identify patterns in multivariate data sets.
2. **Sampling Techniques:** Where an incomplete data set is available, sampling techniques are used to make generalizations about the data. Various considerations need to be accounted for when drawing a sample, not the least of which is any a priori knowledge about the nature of the population.
3. **Neural Networks:** Neural networks are analytical techniques modeled after the process of learning in the cognitive system and the neurological functions of the brain. These techniques are capable of predicting new observations from other observations after executing a process of so-called learning from data. One of the major advantages of neural networks is that they are capable of approximating any continuous function and the researcher does not need to have any hypothesis about the underlying model or even, to some extent, which variables matter.
4. **Decision-Tree Techniques:** These techniques successively split the data into subgroups in order to improve the prediction or classification of the dependent variable. These techniques can handle a large number of independent variables and, being nonparametric in nature, can capture the relationship where traditional statistical techniques fail.

Table 1. Summary of exploratory data analysis techniques

Graphical Representation of Data
- Histogram
- Stem and Leaf Plot
- Box-Cox Plot

Descriptive Statistics
- Measures of Central Tendency (Mean, Median, Mode)
- Measures of Dispersion (Variance, Standard Deviation, Coefficient of Variation)
- Skewness
- Kurtosis

Data-Driven Modeling
- Correlation
- Multiple Regression

Data-Reduction Techniques
- Principal Component Analysis
- Factor Analysis

Classification Techniques
- Discriminant Analysis
- Cluster Analysis

Influential Observations

Forecasting Techniques
- Box- Jenkins Analysis
- Exponential Smoothing Techniques
- State Space Modeling

EXPLORATORY DATA ANALYSIS AND DATA-MINING TECHNIQUES

Keogh and Kasetty (2003) have published a survey on time series data mining. References of these techniques can be seen in Flury (1997), Kumar and Baker (2001), McLachlan (1992), Mendenhall and Sincich (2003), and so forth.

SAMPLING TECHNIQUES FOR DATA MINING

Sampling techniques are essential tools for business data mining. Although there are several sampling techniques, simple random sampling is commonly used in practice. There are quite a few drawbacks of using simple random sampling, and it may not be an appropriate technique in many situations. In this section we have reviewed various sampling techniques with their relative merits and demerits. We have also proposed a new sampling technique.

Sampling techniques play a key role in data mining. These techniques are also widely used in market research, economics, finance, and system analysis and design. Sampling is the process of selecting representative elements of a population from a huge database. When these selected elements are examined closely, it is assumed that the selected sample will reveal useful information about the whole database.

For example, given a huge database of customers, it may be difficult to interview each and every one because of the enormous cost and it will also take lots of time. Obviously, the best way is to draw a small representative sample and interview these customers to assess the preferential pattern of consumers for different types of products, the potential demand for a new product, scope for any diversification in the production schedule, and so forth. Similarly, the system analyst has to make a decision on two key issues: first, the selection of the documents generated by the members of the organization, and second, the fact that a great many employees can be affected by the proposed information system. Which people should the system analyst interview, seek information from via a questionnaire, or observe? Obviously, he or she will need a representative sample.

Rocke and Dai (2003) describe a clustering method for unsupervised classification of objects in large data sets. Grabmeier (2002) has given an overview of clustering analysis techniques from a data-mining point of view. There are quite a few techniques available for selecting a sample, but only the simple random sampling technique is widely used in selecting samples from a database. A summary of these techniques is given in Table 2.

Neural networks are very sophisticated modeling techniques, capable of modeling extremely complex functions. In particular, neural networks are nonlinear and moreover, they are distribution free. For many years, linear modeling had been the commonly used technique in most modeling domains since linear models had well-known optimisation strategies. Where the linear approximation was not valid, the models suffered accordingly.

Neural networks learn by example. The user gathers representative data, and then involves training algorithms to automatically learn the structure of the data. Thus, the user needs only to have knowledge of

- how to select and prepare data,
- how to select an appropriate neural network, and
- how to interpret the results.

Thus, the level of user knowledge needed to successfully apply neural networks is much lower than would be the case using traditional statistical methods.

The type of problems amenable to solution by a neural network is defined by the way they work and the way they are trained. Neural networks work by feeding in some input variables and producing some output variables. They can therefore be used where you have some known information and would like to predict the value of some unknown information. For example, to analyse form in horse racing, the factors to be taken into account might include times, weights, distance suitability, barriers, beating and winning margins, the class of race, and the overall field strength. The output information would be some guide as to the likelihood of success in a particular race (e.g., a prediction of the expected time that a given horse would take to run the race). There are a number of sites where this type of information is available for a fee. The user then has to decide which components of the data they will present to the neural network and the parameters of processing that they want the network to use. Neural networks have been successfully applied across a range of problem domains in areas as diverse as finance, medicine, engineering, geology, and physics. Indeed, anywhere that there are problems of prediction, classification, or control, neural networks are being introduced. This success can be attributed to three key factors.

- **Power:** Neural networks are very sophisticated modeling techniques, capable of modeling extremely complex functions. In particular, neural networks are nonlinear. For many years, linear modeling had been the commonly used technique in most modeling domains since linear models had well-known optimisation strategies. Where the linear approximation was not valid, the models suffered accordingly. Neural networks also keep in check the curse of the dimensionality problem that bedevils attempts to model nonlinear functions with a large number of variables. Quite often, the back propagation method is used for training the neural network.
- **Ease of Use:** Neural networks learn by example. The neural-network user gathers representative data, and then involves training algorithms to automatically learn the structure of the data. Although the user does need to have some heuristic knowledge of how to select and prepare data, how to select an appropriate neural network, and how to interpret the results, the level of user knowledge needed to successfully apply neural networks is much lower than would be the case using more traditional statistical methods.
- **Reasonable Processing Time:** The key performance objective of the ANN will be to achieve the highest possible success rate. A further objective is to achieve this success rate without the ANN processing time being unacceptably slow. These two performance objectives may counteract each

Table 2. A summary of sampling techniques

- Stratified Random Sampling
- Systematic Sampling
- Cluster Sampling
- Two-Stage Sampling
- Sampling Proportional to Size

other but can often be traded off against one another through the design of the ANN.

DECISION TREE TECHNIQUES

Classification and regression tree (CART) techniques were developed by Dan Steinberg. These techniques are robust and easy-to-use decision-tree tools that automatically sift large databases, searching for and isolating significant patterns and relationships in the data. These relationships are then used to generate prediction models for applications such as profiting customers, detecting frauds, and managing credit rules. CART's output can also be used as input to improve prediction accuracy of neural nets. This technique can also handle missing values in the database. CART can also accommodate situations in which some misclassifications are more serious than others. Multivariate adaptive regression splines (MARS), developed by Jerry Friedman, is another variation of CART. It is a high-speed prediction-modeling solution that provides superior forecasting accuracy by deployment of an accurate and easy-to-understand regression model. MARS can find optimal variable transformation and interactions. Every variable selected for entry into the model is checked for a nonlinear response. It can be used for forecasting and also for understanding the prediction model. There are several other decision-tree techniques; for example, SPSS uses a CHAID- (chi-square automatic interaction detection) tree-based statistical method.

FUTURE TRENDS

The flow diagram given can be used for data mining. It illustrates the five steps mentioned in SAS Enterprise Miner.

- **Sample:** The most important thing is to draw a representative sample from a large database. The details of drawing the sample are given in "Sampling Techniques for Data Mining."
- **Explore:** By exploring, we mean calculating various statistics as mentioned in "Exploratory Data Analysis and Data-Mining Techniques."
- **Modify:** This consists of creating new variables or adding some new variables to clarify the data.
- **Model:** This consists of choosing the best available model for estimation or forecasting.
- **Assess:** At this stage we need to make sure that various assumptions inherent in the model are satisfied.

Recent Developments in Data Mining

Data mining has become an essential analytical skill in e-commerce. Business organizations are using data-mining techniques to increase profit. Some authors (e.g., Fayyad, 1997) see data mining as a single step in a larger process that we call knowledge discovery in database. Some other interesting open problems in data mining, for example, finding interesting patterns, spurious relationships, contaminated data, and so forth, is given in the paper by Hand (1998). A new journal named *Data Mining and Knowledge Discovery* covers recent developments in this area. Also, *IEEE Intelligent Systems and their Applications* has a special issue on data mining (March 2000) that gives some important developments in special topics like knowledge design and feature mining. Some recent books on data mining are by Giudice (2003), who uses statistical methods, Nemati and Barko (2003), writing on organizational data mining, Chakrabarti (2002), writing on Web mining, Freitas (2002), whose book is on data mining with evolutionary algorithms, and Kloesgen and Zytkow (2002), writing on data mining and knowledge discovery. Kosala and Blockeel (2000) have given a survey of the research in the area of Web mining. Ma et al. (2000) have presented a system called DS-Web that uses Web to help data mining.

Packages for Data Mining

As mentioned by Hand (1998), because of commercial interest in data mining, a number of software tools have appeared in the market. Some are general tools that are good enough to do statistical analysis like SPSS, SAS,

Table 3. Packages for data mining

- Clementine (SPSS)
- Data Surveyor
- Data Engine
- Datasage
- Data Scope
- Explora
- IBM Visualization Data Explorer
- Model Quest Enterprise
- Nuggets
- Partek
- The Data Mining Suite
- Tooldaiag
- XpertRule

MINITAB, and so forth, while others are special tools built for data mining, for example, EXPLORA (Klosgen, 1996) and Interactive Data Exploration and Analysis Systems (Selfridge, Srivastava, & Wilson, 1996). These packages, for example, can segment market data and analyse the effect of new promotions and advertisements. Two special packages developed by Salford Systems are CART and MARS. Some of the automatic forecasting packages, which do not require specialist knowledge, are AUTOBOX and FORECAST PRO. A review of these packages including some other packages is given in Kumar (1995).

CONCLUSION

Data mining is a very general term that refers to the things that people do to understand and explore data. In this paper we have outlined some of the techniques for data mining. We have not looked at data mining from the point of view of database architecture and data warehouse. Internet data mining is different from ordinary data mining in the sense that data mining is done over the Internet and results keep on changing. With the commercial interest growing in data mining, we feel that there is a good scope for developing an automated Internet data-mining package using the tools mentioned in the paper, which could be easy enough to understand and to make decisions. This will help managers to quickly make a decision as soon as the data is updated on the Internet.

REFERENCES

Chakrabarty, S. (2002). *Mining the Web: Discovering knowledge from hypertext data.* Morgan Kaufmann.

Fayyad, U. (1997). Editorial. *Data Mining and Knowledge Discovery*, *1*, 5-10.

Flury, B. (1997). *A first course in multivariate statistics.* Springer Texts in Statistics.

Freitas, A. A. (2002). *Data mining and knowledge discovery with evolutionary algorithms.* Springer Verlag.

Giudice, P. (2003). *Applied data mining: Statistical methods for business and industry.* John Wiley.

Grabmeier, J., & Rudolp, A. (2002). Techniques of cluster algorithms in data mining. *Data Mining and Knowledge Discovery*, *6*, 303-360.

Hand, D. J. (1998). Data mining: Statistics and more? *The American Statistician, 52*, 112-118.

Keogh, E., & Kasetty, S. (2003). On the need for time series data mining bench marks: A survey and empirical demonstrations. *Data Mining and Knowledge Discovery, 7*, 349-371.

Kloesgen, W., & Zytkow, J. (Eds.). (2002). *Handbook of data mining and knowledge discovery.* Oxford University Press.

Kosala, R., & Blockeel, H. (2000). Web mining research: A survey. *Sigkdd Exploration, 2*, 1-15.

Kumar, K, & Baker, J. (2001). Internet and data mining. In J. D. Haynes (Ed.), *Internet management science: A global perspective* (pp. 236-253). Hershey, PA: Idea Group Publishing.

Kumar, K., & Bhatti, M. I. (1995). Experts systems in forecasting packages: A review. *Science International,* 141-143.

Ma, Y., Bing, L., & Wong, C. K. (2000). Web for data mining: Organizing and interpreting the discovered rules using Web. *Sigkdd Exploration, 2*, 16-23.

McLachlan, G. J. (1992). *Discriminant analysis and statistical pattern recognition.* John Wiley.

Mendenhall, W., & Sincich, T. (1996). *A second course in statistics: Regression analysis* (5th ed.). NJ: Prentice Hall.

Nemati, H. R., & Barko, C. D. (2003). *Organisational data mining: Leveraging enterprise data resources for optimal performance.* Hershey, PA: Idea Group Publishing.

Rocke, D. M., & Dai, J. (2003). Sampling and subsampling for cluster analysis in data mining. *Data Mining and Knowledge Discovery*, *7*, 215-232.

KEY TERMS

Back Propagation: A training method used to calculate the weight in a neural net from the data

Classification Tree: A decision tree that places categorical variables into classes

Data Mining: A process of seeking interesting and valuable information from a large database using a combination of methods

Discriminant Analysis: A multivariate statistical method that separates the data into categories

Internet Data: Data collected through Internet by an organization

MARS: Multivariate adaptive regression splines that are a generalization of decision trees

Neural Network: A complex, nonlinear modeling technique based on a model of the human neuron

Outlier: Data items that did not come from the assumed population of data

Sampling: Part of the data selected to represent the whole population

Time Series Model: A model that forecasts future values of the time series based on past data

Internet Diffusion in the Hospitality Industry

Luiz Augusto Machado Mendes-Filho
Faculdade Natalense para o Desenvolvimento do Rio Grande do Norte, Brazil

Anatália Saraiva Martins Ramos
Universidade Federal do Rio Grande do Norte, Brazil

INTRODUCTION

Tourism is the most important industry in the world in terms of the numbers of employees and its effect on the social and economical development of a region or country. Holjevac (2003) believes that, by the year 2050, tourism will by far be the largest industry worldwide, with 2 billion tourists and US$24 billion in domestic and international receipts. Moreover, the major tourist destinations will be India, China, Indonesia, and Brazil.

The use of information technologies for basic functions—conferences, business meetings in distant places, training, designed routes and airlines, reservations and tickets purchased through computer systems, tourist shops, restaurants—is becoming usual in tourism. All these services have led tourist companies to adopt more updated methods in order to increase competition. Consumers, who are already becoming familiar with new technologies, demand more flexible, interactive, and specialized products and services, bringing new management techniques from the intelligent use of IT used to accomplish tour company business processes (Buhalis, 2000).

The hotels depend progressively on the resources of new information technology to follow and update the tools which allow an efficient development of activities in each section of the company, leading to better results for its management (Mendes-Filho & Ramos, 2003a). To Phillips and Moutinho (1998), information technology (IT) is one of the critical factors of success in the hotel industry.

According to studies and data, the use of technological tools will allow a bigger competitiveness for hotels (Cline, 1999). Technology will be the catalyst of change, a source of growing connectivity and one of the most important factors in distinguishing success among hotel companies. Few issues have greater importance to the business of hospitality than the technological decisions that will be made in the coming years (Buhalis, 2000; Mendes-Filho & Ramos, 2004; Olsen & Connoly, 2000).

The hotel industry is one of the most important kinds of Web commerce. The data shows that all major companies linked to the tourism industry (hotels, agencies, air companies, and rentals) possess some kind of e-commerce activity through the Web (O'Connor, 1999; Scottish Executive, 2000; Werthner & Klein, 1999).

BACKGROUND

Decades ago, before the use of the computer in the accommodation sector, those charged with making reservations performed their service by checking availability tables exposed on the wall or in large, updated, handwritten lists (O'Connor, 1999). The hotels received innumerous telephone calls, letters, and telex from potential clients, sometimes larger than that of the hotel's reception, and worked to select correspondence, type letters, send telegrams, and deal with other demands. The delays were frequent, the cost of correspondence writing went sky-high, and specialized typists were in demand (World Tourism Organization, 2003).

A way found by the American hotel chains to streamline the reservation services was to centralize this function in a main office, serving the consumer better and offering a valuable service to the hotels belonging to that chain. O'Connor (1999) states that the reservation process in hotels in the USA was made even easier with the introduction of free telephone services in the mid-'60s, which permitted potential clients to perform an only call to obtain information or make reservations in any of the hotels of that chain in the world.

Although the reservation area became faster and more efficient, two large costs remained, those of telecommunications (free telephone service payment) and labor costs of the reservation agents necessary to answer the phones. With the increase in trips during the 1960s, the airline companies developed the computer reservation system (CRS), which pressured the hotel sector to develop its own (O'Connor, 1999).

The main focus in hotel and restaurant management has always been the maximization of consumer satisfaction and personalized attention. The use if IT has, at times, seemed incompatible with this objective, and the hotel sector has, in a way, delayed the application of IT in its operations. The technology has been viewed as a hindrance to personalized service because it creates an impersonal, mechanical, and cold environment with the clients.

However, the change of this belief is being changed within the hotel sector. Nowadays, according to Sheldon (1997), the establishments are noticing that IT can bring efficiency to the hotel, besides reducing costs and offering a great potential to increase the levels of personalized service to the clients.

In a survey performed by financial managers of American hotels, all stated that IT increased the hotel's productivity (David, Grabski, & Kasavana, 1996). The motives used to justify this statement were the following: Technology reduces the administration costs, decreases the amount of paperwork between sectors, minimizes operational errors, increases the earnings/profits of the hotel, and makes the reservation management more efficient. This same survey proved that IT is not only used to increase the hotel's productivity but also to improve the service, as well as to offer new services to the guests.

According to Namasivayam, Enz, and Siguaw (2000), IT can also be employed to reach business objectives. American and European hotel executives have plans to use the technology to reduce operational costs, increase sales, improve the service to the client, increase employees' productivity, and increase hotel earnings.

In research performed by Van Hoof et al. (1995), 550 American hotel managers answered questions about their perceptions of the use and implementation of technology in their establishments. Those responding identified the front office (reception and reservations) of the hotel as the sector that can benefit the most from the use of the technology, followed by sales and marketing, accounting, and the food and drink sector. According to Van Hoof et al. , having a quality service is a challenge to the hotel industry, which has high employee turnover indexes, employee salary increase, and low age of the most qualified people. Consequently, technological applications have been developed in hotels to increase this quality in the services and improve the interaction of the hotel employees with the guests.

IMPACTS OF THE INTERNET IN HOTEL INDUSTRY

During the '80s and '90s several authors from companies and universities had already foreseen that as new technologies were increasingly used, hotels could benefit from that in a great range of situations, for example: better qualified services for customers, increased sales and profits, efficiency in operation and integration of hotel sectors, rapid communication, and cost reduction (Laudon & Laudon, 1999).

Technological applications enable information and knowledge to bring a competitive advantage to the future profile of the hotel. The "Information Age" idea is that the most modern companies will build their success upon the amount of knowledge they have about their clients as well as information on their products and services and how they will make a profit in this new environment (Olsen & Connoly, 2000).

With the Internet being used as a means of communication, this brings several advantages or benefits compared to other vehicles. Flecha and Damiani (2000) state that when it comes to the tourist area, the main points are: the new relationship between consumers and companies, marketing for actively participating consumers, the importance of detailed information, self-service application, credibility, and agility of communication.

The use of the Internet and World Wide Web is spreading quickly in most consumer access areas to travel database developments. There are hundreds of thousands of suppliers' homepages, associations, e-news, newsgroups, and chats for the travel and tourism community. This group of technologies provides many opportunities for the industry to interact with its consumers and suppliers. It is also possible that, through information technology, products and services may be personalized according to the tourist's needs and thus may become a differential feature for those who adopt it (Buhalis, 2000; Sheldon, 1997).

The purchase of products and services through the Internet is revolutionizing the world of business and people's lives as well. For some clients it is more comfortable to book an e-ticket through the company home page rather than going to the travel agency (Franco, 2001).

As the Internet began and grew, the use of such technologies at home or work and also the new opportunities that arose from the lower costs in telecommunication equipment made it possible for suppliers to distribute information to their clients and process reservations directly with the clients (O'Connor, 1999).

According to Jeong and Lambert (2001), the Internet has already modified the competitive strategy of some hotels. It is through the Internet that the client can have a "self-understanding" in a service that is being offered to him in a more efficient way. In hotels, check-in processes can already be totally automatic, from the Internet booking until the moment the client takes his keys in an automatic dispenser. The result is that clients can become more informed and willing to have quick answers from the orders online. Though many experts and businessmen agree that the Internet is probably the most important technological tool, it is still relatively new and misused in the hotel industry (Van Hoof & Verbeeten, 1997).

Several authors have identified impediments to the growth of the Internet in the industry and, hence, have reservations about the willingness of hotel operators to adopt the Internet wholeheartedly (Wei et al., 2001). These problems include user-friendliness, the quality and accuracy of information obtained from the Web, and the issue of data security (Wei et al.). Here are other difficulties found by Lituchy and Rail (2000) in their research: problems in updating new information in hotel Web pages, annoyance expressed at inaccessible Webmasters by managers, hard to find hotel Web sites, lack of knowledge on the employees' part of how to use the technology, and the impersonality of the medium. Because of these, consumers have been slow to adopt the Internet as a means of making hotel reservations. Only 4% of reservations are made online (Maselli, 2002).

Namasivayam et al. (2000) summarize that almost 60% of the hotels in their study had few technologies. To Feiertag (2000), a lack of proper training, high turnover rates, and limited financial resources were major barriers to the successful use and implementation of new technologies.

In addition to this, many hotels still believe that conventional means of advertising, such as radio, television, and printed material, are the most effective way of promoting their properties. The share of reservations received through the Internet remains minute compared to reservations received through conventional means, such as phone, fax, or mail (Van Hoof & Combrink, 1998). However, these problems are diminishing with the increase of the number of Internet users. And if customers become accustomed to browsing for rooms and making reservations through the Internet, more and more properties will be forced to get on the Internet as well (Mendes-Filho & Ramos, 2002).

In a specific way, the Internet provides an expansion of hotel services, changing this industry and giving new opportunities to clients, thus being a new channel to be developed. Besides online reservation services, the Internet allows hotels to sell their services and charge them electronically as well as offer new products through the World Wide Web (Blank, 2000; Laudon & Laudon, 1999).

Through the Web the customer can check hotel location, compare rates, see pictures and watch videos, get information about tourist destinations and other facilities, check room availability, and book and confirm reservations for the amount of time he wants to stay, among other services. Hence, the interactivity of the Web provides an ideal medium for distributing accommodations online, consolidating itself as a very adequate platform for bringing information and services to the client in a very straightforward, efficient, and quick way (Flecha & Damiani, 2000; Hotels, 2001).

Marriot, Hospitality Services of America, and Hilton are some of the hospitality industry's members that have successfully used marketing on the Internet to reach new markets, track customers, take online reservations, and offer information about their products and services (Lituchy & Rail, 2000).

FUTURE TRENDS

A hotel chain's success has always depended on excellent services performed by operation, marketing, and human resources sectors. For Withiam (2000) in the 21st century an essential factor will be technological support, making it possible for computers to process information of reservation systems, affinity programs, and marketing data banks.

Improvements in integration, centralized data banks, and the use of Web sites are some of the tendencies in the development of software for hotels (Adams, 2001). Therefore, the connection of a hotel system to the Internet will integrate information of the internal system with the Web site, and this will make a lot of information available to managers. The new systems are being developed with this integration with the Web site.

With the increasing demand of information in the tourist sector, the importance of IT use in this industry will only tend to increase in the future. Therefore, the tourist businesses must understand, incorporate, and use IT strategically to serve the target markets, improve their efficiency, maximize profitability, perfect services, and maintain the profitability in the long term (Buhalis, 2000).

To Olsen and Connoly (2000), the volume of information about the guests collected electronically is too large for the directors to be able to manage without the help of technology. Data warehousing and data mining are technologies that are gaining popularity to analyze information about clients. These technologies may be used to help hotel keepers construct good relationships with their guests, increasing their loyalty to them.

Using the Internet in the hotel industry has good prospects of growth, though in many hotels the use of such technology is still moving slowly. On the other hand, there are some hotels using and steadily setting the trend. It will be an important and strategic issue for businessmen to stimulate such Internet use policies inside the tourist trade so that they become wired to this new reality and can work on even terms with their competitors.

CONCLUSION

The Internet has decreased expenses and enabled small businesses to conduct international business from home (Lituchy & Rail, 2000). Small inns and bed and breakfasts are advertising on the Web and are therefore becoming a presence in the global market. So, they face the likelihood of serving foreign customers that may have different hospitality expectations.

Despite the fact that Internet use is very common, the proportion of reservations received from the Internet is small. The public could still be concerned about issues of

security for financial transactions or could not be satisfied with its inability to synchronize inquiries. A low reservation rate from the Internet may also be partly attributed to the lack of certain relevant information, such as room availability and virtual tours of the property not commonly included in the homepage (Wei et al., 2001).

In general, the Internet does enable tourist companies to increase their competitiveness. IT can improve the efficiency of suppliers and provide tools for the development and delivery of different tourist products (Mendes-Filho & Ramos, 2003b). One of the benefits reached is the reduction of the dependence on the middlemen in the distribution of tourist products. Hotel owners should invest more money in technology besides concentrating more time and attention to subjects in that area. IT affects all aspects of a hotel chain's value, going far beyond sectors and departments. As technology will be intrinsically linked to hotel business, its executives will insert technology in all their strategic decisions for the facility. That implies all the employees (including managers and directors) need to have enough knowledge to extract the potential the technology provides.

REFERENCES

Adams, B. (2001). The PMS picture. *Hotel and Motel Management*, *216*(2), 36-37.

Blank, D. (2000). Internet will shape revenue-management role. *Hotel and Motel Management*, *215*(11), 54-55.

Buhalis, D. (2000). Marketing the competitive destination of the future. *Tourism Management*, *21*(1), 97-116.

Cline, R. (1999). Hospitality 2000—The technology. *Lodging Hospitality*, *55*(7), 18-26.

David, J. S., Grabski, S., & Kasavana, M. (1996). The productivity paradox of hotel-industry technology. *Cornell Hotel and Restaurant Administration Quarterly*, *37*(2), 64-70.

Feiertag, H. (2000). Technology can help salespeople, but it can't replace them. *Hotel and Motel Management*, *215*(14), 22.

Flecha, A. C., & Damiani, W. B. (2000). Avanços da tecnologia da informação: Resultados comparados de sites da indústria hoteleria. *Proceedings of the 20th Production Engineering National Meeting*, (Vol. 1, pp. 153-161).

Franco, C. F., Jr. (2001). *E-business: Tecnologia de informação e negócios na Internet*. São Paulo, Brazil: Atlas.

Holjevac, I. A. (2003). A vision of tourism and the hotel industry in the 21st century. *Hospitality Management*, *22*, 129-134.

Hotels. (2001). Hotels' 2001 worldwide technology survey—Part 1. *Hotels*, *35*(2), 75-85.

Jeong, M., & Lambert, C. (2001). Adaptation of an information quality framework to measure customers' behavioral intentions to use lodging Web sites. *International Journal of Hospitality Management*, *20*(2), 129-146.

Laudon, K. C., & Laudon, J. P. (1999). *Sistemas de informação com Internet* (4th ed.). Rio de Janeiro, Brazil: LTC.

Lituchy, T. R., & Rail, A. (2000). Bed and breakfasts, small inns, and the Internet: The impact of technology on the globalization of small businesses. *Journal of International Marketing*, *8*(2), 86-97.

Maselli, J. (2002, April 22). Hotels take to the Web to battle discounters. *InformationWeek*.

Mendes-Filho, L. A. M., & Ramos, A. S. M. (2002). The Internet adoption in the hotel industry: A multiple cases study in Brazilian hotels. *Proceedings of the 13th Information Resources Management Association International Conference*, (Vol. 1, pp. 209-211).

Mendes-Filho, L. A. M., & Ramos, A. S. M. (2003a). The benefits and difficulties of the Internet use in hotels: The effect of hotel rate on the managers' perception. *Proceedings of the 14th Information Resources Management Association International Conference*, (Vol. 1, pp. 328-330).

Mendes-Filho, L. A. M., & Ramos, A. S. M. (2003b). The perception of managers on the impacts of Internet in Brazilian hotels: An exploratory study. In S. Kamel (Ed.), *Managing globally with information technology* (pp. 244-259). Hershey, PA: Idea Group Publishing.

Mendes-Filho, L. A. M., & Ramos, A. S. M. (2004). The benefits and difficulties of Internet use in hotels and its effects according to the facilities' rank, property size, manager's age and experience. In C. Deans (Ed.), *E-commerce and m-commerce technologies* (pp. 217-239). Hershey, PA: Idea Group Publishing.

Namasivayam, K., Enz, C. A., & Siguaw, J. A. (2000). How wired are we? The selection and use of new technology in

U.S. hotels. *Cornell Hotel and Restaurant Administration Quarterly*, *41*(6), 40-48.

O'Connor, P. (1999). *Electronic information distribution in tourism and hospitality*. Wallingford, UK: CAB International.

Olsen, M. D., & Connoly, D. J. (2000). Experience-based travel. *Cornell Hotel and Restaurant Administration Quarterly*, *41*(1), 30-40.

Phillips, P. A., & Moutinho, L. (1998). *Strategic planning systems in hospitality and tourism*.Wallingford, UK: CAB International.

Scottish Executive. (2000). A new strategy for Scottish tourism. Edinburg. Retrieved July 21, 2002, from *http://www.scotland.gov.uk/library2/doc11/sfst.pdf*

Sheldon, P. (1997). *Tourism information technology*. Wallingford, UK: CAB International.

Van Hoof, H. B., & Combrink, T. E. (1998). U.S. lodging managers and the Internet: Perceptions from the industry. *Cornell Hotel and Restaurant Administration Quarterly*, *39*(2), 46-54.

Van Hoof, H. B., & Verbeeten, M. J. (1997). Vendors receive mixed reviews. *Hotel and Motel Management*, *212*(11), 42.

Van Hoof, H. B., Collins, G. R., Combrink, T. E. & Verbeeten, M. J.(1995). Technology needs and perceptions: An assessment of the U.S.lodging industry. *Cornell Hotel and Restaurant Administration Quarterly*,*36*(5), 64-69.

Wei, S., Ruys, H. F., Van Hoof, H. B. & Combrink, T. E. (2001). Uses of the Internet in the global hotel industry. *Journal of Business Research*,*54*, 235-241.

Werthner, H., & Klein, S. (1999). *Information technology and tourism*: *A challenging relationship*. New York: Springer-Verlag.

Withiam, G. (2000). Carlson's "24K" consumer-centric computer. *Cornell Hotel and Restaurant Administration Quarterly*, *41*(3), 13.

World Tourism Organization (2003). *E-Business para turismo: Guia prático para destinos e empresas turísticas*. Porto Alegre, Brazil: Bookman.

KEY TERMS

Bed and Breakfast: An establishment (as an inn) offering lodging and breakfast.

Computer Reservation System: A computer system that manages the distribution of the tourist products to transportation, lodging, and entertainment companies.

Data Mining: The process of analyzing data to determine relationships undiscovered by previous analyses.

Data Warehouse: A data warehouse is a central repository for all or significant parts of the data that an enterprise's various business systems collect.

E-Ticket: An e-ticket (electronic ticket) is a paperless electronic document used for ticketing passengers, particularly in the commercial airline industry. Virtually all major airlines now use this method of ticketing.

Front Office: The department of the hotel that deals directly with clients. Normally, it involves the reception and the reservation sector of the hotel.

Middleman: A dealer or agent intermediate between the producer of goods and the consumer or retailer.

Internet Support for Knowledge Management Systems

Murray E. Jennex
San Diego State University, USA

INTRODUCTION

Organizations are building and maintaining systems for managing organizational knowledge and memory. Users of these systems may not be at the same location; in many cases they are distributed across large geographical distances and multiple offices. Key to this task is developing an infrastructure that facilitates distributed access and utilization of the retained knowledge and memory. Connectivity and easy to use interfaces are main concerns. Jennex (2000) found that using the Internet as a common communications platform (either as an Intranet or an Extranet) and Web browsers as an interface is a viable, low cost solution. Newell, et al. (1999) found that Intranets not only supported distributed knowledge processes but also enhanced users' abilities to capture and control knowledge. Stenmark (2002) proposes that using a multiple perspective of the Internet— information, awareness, and communication—allows developers to build successful Internet-based knowledge management systems, KMS. This article discusses how the Internet can be effectively used as an infrastructure for knowledge management/organizational memory systems, KMS/OMS.

BACKGROUND

The OMS consists of the processes and information system components used to capture, store, search, retrieve, display, and manipulate knowledge. The KMS consists of the tools and processes used by knowledge workers to interface with the knowledge contained in the OMS. Knowledge is managed and used through a combination of the KMS and OMS. Jennex and Olfman (2002) identified the KMS-OMS model in Figure 1 as a representation of the relationships between the OMS, KMS, and organizational learning. Organizational learning, OL, is identified as a quantifiable improvement in activities, increased available knowledge for decision-making, or sustainable competitive advantage (Cavaleri, 1994; Dodgson, 1993; Easterby-Smith, 1997; Miller, 1996).

There are two approaches to building a KMS as discussed by Hansen et al. (1999), Morrison and Weiser (1996), and Stenmark (2002). These can be described as a process/task approach and the infrastructure/generic approach. The process/task approach focuses on the use of knowledge/OM by participants in a process, task or project in order to improve the effectiveness of that process, task or project. This approach identifies the information and knowledge needs of the process, where they are located, and who needs them. This approach requires the KMS to capture less context, as users are assumed to understand the knowledge that is captured and used.

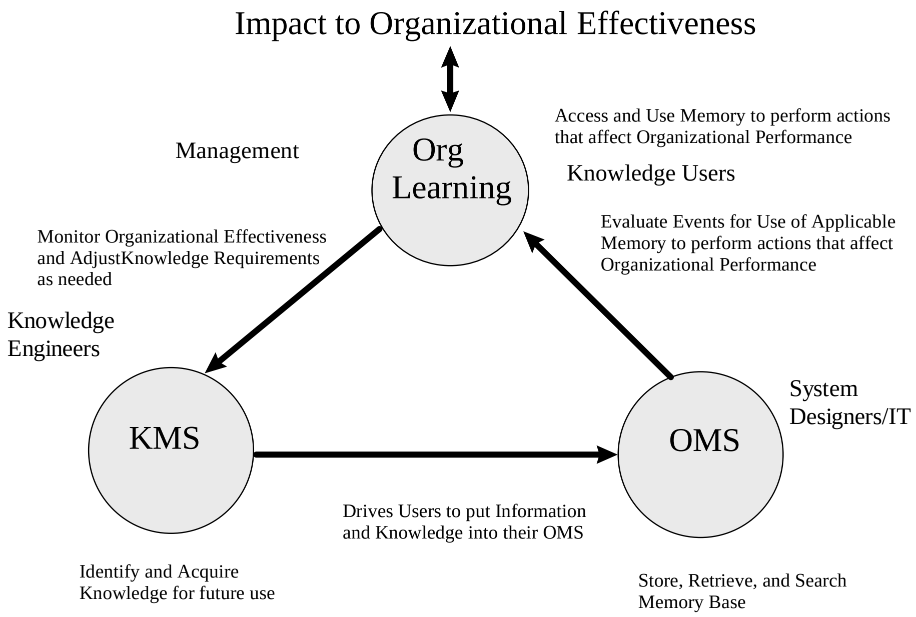
The infrastructure/generic approach focuses on building a system to capture and distribute knowledge/OM for use throughout the organization. Concern is with the capturing of context to explain the captured knowledge and the technical details needed to provide good mnemonic functions associated with the identification, retrieval, and use of knowledge/OM. The approach focuses on network capacity, database structure and organization, and knowledge/information classification.

Both approaches may be used to create a complete KMS. The process/task approach supports specific work activities, while the infrastructure/generic approach integrates organizational knowledge into a single system that can be leveraged over the total organization instead of just a process or project.

Jennex and Olfman (2001) developed a set of design recommendations for enabling KM/OM in systems. The recommendations, Table 1, are based on studies of KMS/OMS success factors. One recommendation calls for use of a common infrastructure. The Internet is suggested for this due to its widespread availability, open architecture, and developed interfaces. This also assists in standardizing software across the organization through the use of browsers and Web applications.

The Internet meets several of these recommendations. It provides a common network that is global. Use of common browsers aids in standardizing software. Ease of use of browsers and in building and maintaining Internet-based systems empowers users (Newell et al., 1999) and simplifies incorporating the KMS into everyday processes. Ease in handling unstructured data as well as databases simplifies knowledge representation, capture, and dissemination. Table 2 lists Internet tools and features that expand the ability of the Internet to serve as the infrastructure for a KMS. Some of these features are expanded in the following.

Figure 1. The Jennex-Olfman KMS-OMS model



Gandon et al. (2000) propose using XML to encode memory and knowledge, and suggest using a multi-agent system that can exploit this technology. The proposed system would have improved search capabilities and would improve the disorganization and poor search capability normally associated with Web pages. Chamberlin et al. (2001) and Robie et al. (1998) discuss using XML query language to search and retrieve XML encoded documents.

Dunlop (2000) proposes using clustering techniques to group people around critical knowledge links. As individual links go dead due to people leaving the organization, the clustered links will provide a linkage to people who are familiar with the knowledge of the departed employee. Lindgren (2002) proposes the use of Competence Visualizer to track skills and competencies of teams and organizations.

Te'eni and Feldman (2001) propose using task-adapted Web sites to facilitate searches. This approach requires the site be used specifically for a KMS. Research has shown that some tailored sites, such as those dedicated to products or communities, have been highly effective.

Eppler (2001), Smolnik and Nastansky (2002), and Abramowicz et al. (2002) use knowledge maps to graphically display knowledge architecture. This technique uses an intranet hypertext clickable map to visually display the architecture of a knowledge domain. Knowledge maps are also known as topic maps and skill maps. Knowledge maps are useful, as they create an easy to use standard graphical interface for the Intranet users and an easily understandable directory to the knowledge.

The use of ontologies and taxonomies to classify and organize knowledge domains is growing. Zhou et al. (2002) propose the use of ROD, rapid ontology development, as a means of developing an ontology for an undeveloped knowledge domain.

FUTURE TRENDS

Although there is strong support for using the Internet as a knowledge infrastructure, there are areas that current research is improving. Chief among these is the difficulty in organizing and searching large quantities of knowledge in varying knowledge formats and structures. Knowledge can be stored as documents, audio, images, databases, and spreadsheets. Lack of standard structure can make organizing knowledge difficult, while the lack of standard terms and naming conventions makes searching difficult. An example is Ernst & Young UK, who in early 2000 had in excess of one million documents in its KMS (Ezingeard et al., 2000). Another concern is the tendency to not to use the system. Jennex and Olfman (2002) fount that voluntary use is enhanced if the system provides near and long-term job benefits, is not too complex, and the organization's culture supports sharing and using knowledge and the system. Other significant issues requiring resolution are summarized in Table 3 and include security, having ad-

Table 1. KMS design recommendations

- Use a common network structure, such as the Internet.
- Add KM/OM skills to the tech support skill set.
- Use high-end PCs and/or clients.
- Standardize hardware and software across the organization.
- Incorporate the KMS into everyday processes and IS.
- Use an enterprise-wide data dictionary to design knowledge base.
- Allocate maintenance resources for KMS.
- Train users on use and content of the KMS.
- Create and implement a KM strategy/process for identifying/maintaining the knowledge base.
- Expand system models/life cycles to include the knowledge process.
- Assess system/process changes for impact to the KMS.
- Automate data capture.
- Design security into the knowledge base.
- Incorporate KM into personnel evaluation processes.
- Implement KMS use/satisfaction metrics.
- Identify organizational culture concerns that could inhibit KMS usage.

equate bandwidth for the expected use, maintaining content in large sites, and system incompatibilities between distributed offices/users.

CONCLUSION

The conclusion is that the Internet is an effective infrastructure for a KMS. However, there are issues associated with using the Internet that KMS designers need to be aware of. Chief among these are knowledge representation and search. Several tools such as knowledge maps, XML, adaptive Web sites, clustering, and examples of effective Internet-based KMSs were discussed that addressed these issues. However, as knowledge bases grow, designers need to be aware of increasing search times as well as a variety of knowledge artifacts. This is perhaps the most important area for future research. Developing ontologies and taxonomies to aid in classifying and structuring knowledge domains is critical.

Maintaining a site is critical. User, organizational, and/or project needs for knowledge change over time, requiring the KMS to change its knowledge content. Also, knowledge has a life cycle and eventually reaches

Table 2. Internet features/technologies that support KMS

- Common Architecture and Interfaces
- Easy to Use Front-end Systems (Browser User Interface)
- Trends Towards Internet-Based Processes
- Back-end Systems that Provide Database Access to Users
- XML wrapping of documents and other data
- Powerful Search Engines
- Virtual Private Networks
- Internet Opportunities Include:
  - Ability to push time-sensitive data quickly to a wide audience
  - To answer frequently asked questions
  - To create 24-hour service
  - To make knowledge additions/updates available quickly
  - To allow feedback from users
  - To service both specialized teams/users and generic users

Table 3. A summary of issues for future trends of Internet-based KMS

Bandwidth Restrictions and Latency
Improving transmission and methods to enable large numbers of users to receive knowledge from the KMS

Organizing Knowledge
Creating standard ontologies and taxonomies to enable standard knowledge structures

Flooding of the Web with content
Improving knowledge acquisition so that content not helpful to the users is captured

Maintenance and integrity of data
Keeping Web maintenance up-to-date and accurate knowledge on the site for users

Exposure Points/Security
Creating secure systems that allow remote access yet keep unauthorized user access out

System incompatibilities
Improving cross-platform compatibility to improve system integrations and access

a point where it is no longer useful. Organizations must allocate resources to update and maintain every KMS.

Securing the KMS is also critical, as knowledge is valuable. KMS/OMS designers need to ensure the security of captured knowledge and use of secure connections is a viable alternative. This may be the greatest impediment to the development of Internet-based KMSs.

The final issue is the tendency of people not to use the computer portion of a KMS. Jennex and Olfman (2002) found that this is a tendency of new members and suggest that this is a matter of context. New members do not understand the context under which the knowledge was created and stored so do not know how to retrieve and use the knowledge. As these members gain experience they gain context and rely more upon the computer and less upon their peers.

REFERENCES

Abramowicz, W., Kowalkiewicz, M., & Zawadzki, P. (2002). Tell me what you know or I'll tell you what you know: Skill map ontology for information technology courseware. *Proceedings of the 2002 Information Resources Management Association International Conference.*

Cavaleri, S. (1994). Soft systems thinking: A pre-condition for organizational learning. *Human Systems Management, 13*(4), 259-267.

Chamberlin, D., Clark, J., Florescu, D., Simon, J., Robie, J., & Stofancscu, M. (2001). *Xquery 1.0: An XML Query Language*. WSC Working Draft 2001. *www.w3.org/TR/xquery/*

Davenport, T.H., & Prusak, L. (1998). *Working knowledge.* Harvard Business School Press.

Dodgson, M. (1993). Organizational learning: A review of some literatures. *Organization Studies, 14*(3), 375-394.

Dunlop, M.D. (2000). Development and evaluation of clustering techniques for finding people. *Proceedings of the 3rd International Conference on Practical Aspects of Knowledge Management, PAKM2000.*

Easterby-Smith, M. (1997). Disciplines of organizational learning: Contributions and Critiques. *Human Relations, 50*(9), 1085-1113.

Eppler, M.J. (2001). Making knowledge visible through intranet knowledge maps: Concepts, elements, cases. *Proceedings of the 34th Hawaii International Conference on System Sciences.* IEEE Computer Society.

Ezingeard, J.-N., Leigh, S., & Chandler-Wilde, R. (2000). Knowledge management at Ernst & Young UK: Getting value through knowledge flows. *Teaching Case, 807-822.*

Gandon, F., Dieng, R., Corby, O., & Giboin, A. (2000). A multi-agent system to support exploiting an XML-based corporate memory. *Proceedings of the 3rd International Conference on Practical Aspects of Knowledge Management, PAKM2000.*

Hansen, M.T., Nohria, N., & Tierney, T. (1999, March-April). What's your strategy for managing knowledge? *Harvard Business Review*, 106-116.

Jennex, M.E. (2000). *Using an intranet to manage knowledge for a virtual project team. Internet-based organizational memory and knowledge management.* Hershey, PA: Idea Group Publishing.

Jennex, M.E., & Olfman, L. (2001). *Development recommendations for knowledge management/ organizational memory systems. Contemporary trends in IS development.* Kluwer.

Jennex, M.E., & Olfman, L. (2002). Organizational memory/ knowledge effects on productivity, a longitudinal study. *Proceedings of the 35th Annual Hawaii International Conference on System Sciences.* IEEE Computer Society.

Jennex, M.E., & Olfman, L. (2003). Organizational memory. In C.W. Holsapple (Ed.), Handbook on knowledge management (pp. 207-234). Berlin, Heidelberg: Springer.

Lindgren, R. (2002). Competence visualizer: Generating competence patterns of organizational groups. *Proceedings of the 35th Annual Hawaii International Conference on System Sciences.* IEEE Computer Society.

Malhotra, Y. (1998). Knowledge management for the new world of business. *www.brint.com/km/whatis.htm*

Miller, D. (1996). A preliminary typology of organizational learning: Synthesizing the literature. *Journal of Management*, *22*(3), 485-505.

Morrison, J., & Weiser, M. (1996). A research framework for empirical studies in organizational memory. *Proceedings of the 29th Annual Hawaii International Conference on System Sciences.* IEEE Computer Society Press.

Newell, S., Scarbrough, H., Swan, J., & Hislop, D. (1999). Intranets and knowledge management: Complex processes and ironic outcomes. *Proceedings of the 32nd Annual Hawaii International Conference on System Sciences.* IEEE Computer Society.

Robie, J., Lapp, J., & Schach, D. (1998). XML Query Language (XQL). *WWW The Query Language Workshop (QL).*

Smolnik, S., & Nastansky, L. (2002). K-discovery: Using topic maps to identify distributed knowledge structures in groupware-based organizational memories. *Proceedings of the 35th Annual Hawaii International Conference on System Sciences.* IEEE Computer Society.

Stein, E.W., & Zwass, V. (1995). Actualizing organizational memory with information systems. *Information Systems Research*, *6*(2), 85-117.

Stenmark, D. (2002). Information vs. knowledge: The role of intranets in knowledge management. *Proceedings of the 35th Annual Hawaii International Conference on System Sciences.* IEEE Computer Society.

Te'eni, D., & Feldman, R. (2001). Performance and satisfaction in adaptive Websites: An experiment on searches within a task-adapted Website. *Journal of the Association for Information Systems*, *2*(3), 1-30.

Zhou, L., Booker, Q.E., & Zhang, D. (2002). ROD – toward rapid ontology development for underdeveloped domains. *Proceedings of the 35th Annual Hawaii International Conference on System Sciences.* IEEE Computer Society.

KEY TERMS

Knowledge: An evolving mix of framed experience, values, contextual information and expert insight that provides a framework for evaluating and incorporating new experiences and information (Davenport & Prusak, 1998).

Knowledge Management: The process established to capture and use knowledge in an organization for the purpose of improving organizational performance (Malhotra, 1998).

Knowledge Management System: The system created for users to interact with the organizational memory system.

Knowledge Map: An Intranet hypertext-clickable map to visually display the architecture of a knowledge domain. Knowledge maps are also known as topic maps and skill maps.

Knowledge Ontology: Common definitions established for captured knowledge, similar to key words, used to capture and express a common context for search, retrieval, and use of knowledge.

Knowledge Taxonomy: The hierarchical organization of knowledge categories within a knowledge management system.

Organizational Learning: The process by which an organization assimilates experiences of its members and uses that experience to modify the organization's potential actions (Jennex & Olfman, 2003).

Organizational Memory: The means by which knowledge from the past is brought to bear on present activities resulting in higher or lower levels of organizational effectiveness (Stein & Zwass, 1995; Walsh & Ungson, 1991).

Organizational Memory System: The system created to capture, store, search, and retrieve knowledge from a repository.

Interoperability in Geospatial Information Systems

Hassan A. Karimi
University of Pittsburgh, USA

Ratchata Peachavanish
University of Pittsburgh, USA

INTRODUCTION

Geospatial information systems (GIS) are an important sector of the information industry, as well as an essential component of the information technology infrastructure (Lo & Yeung, 2002). They are a type of computerized information system specifically designed and used to solve *geospatial* problems, those which are related to locations on the surface of the earth (Longley, Goodchild, Maguire, & Rhind, 2001). The extent of usefulness of GIS has been proven across many diverse applications in many disciplines. They have long been used in traditional application settings, such as land management and natural resources, and have recently become an important element in emerging applications, for example, in ubiquitous mobile computing environments.

Since the mid-1990s, the focus of computing has shifted from stand-alone and locally networked environments to wide-scale, distributed, heterogeneous computing infrastructures. This coupled with the exponential growth in the use of the Internet has enabled and compelled new ways of using GIS. A wide range of GIS applications are now widely available on the Internet for users from anywhere in the world. In addition, the proliferation of wireless and mobile computing technologies, such as cellular phones and Personal Digital Assistants (PDA), has provided new platforms and paved the way for the emergence of new GIS applications. Because of the advances in computing, GIS applications are now designed, implemented, and applied very differently from their predecessors.

Contrary to their past monolithic design and implementation, GIS are now becoming an integral component of many diverse software packages specifically designed for solving problems in different application domains. In addition, current computing trends suggest that future GIS will be multitiered and used in heterogeneous network environments, where computers of different platforms coexist and tasks are performed in a distributed manner. Consequently, future GIS will have to be *interoperable*—they will have to be able to work together in a seamless fashion.

BACKGROUND

Efficient and effective use of GIS to solve geospatial problems generally requires special skills. Today's GIS platforms are mostly designed for workstations or personal computers and provide generic "toolbox" *geoprocessing* operations that can be broadly applied to many problems in different application domains. To utilize these packages, the users must possess certain knowledge and skills. First, they must have knowledge about how real-world objects are represented in GIS; for example, the boundary of a county is represented as a set of points that defines a polygon, and a railroad is represented as a line object. Second, the users must know the range of available geoprocessing operations in GIS and how they are applied to solve geospatial problems. For example, the users must know that in order to determine if a railroad crosses a county boundary, a geometric intersection operation using a polygon and a line that define the county boundary and the railroad, respectively, should be applied. In addition, knowledge about geospatial data sources, geospatial data storage, and methods of obtaining geospatial data are also needed. To solve the problem of county boundaries, the users must have knowledge about the sources of data sets for polygons representing county boundaries. They must also know the format and structure in which the data sets are stored. Lastly, the users must know how a GIS software package operates and how to use it to solve problems. For example, they must know how to operate the ArcInfo GIS software package and be familiar with the methodology in ArcInfo for incorporating data sets into the project, including format conversion, coordinate transformation, and importing procedures. They must also know the specific commands and syntax for invoking geoprocessing operations in ArcInfo as well as the specific behavior of each operation (e.g., Does the "intersect" operation provided by ArcInfo partially, or fully, solve the problem?).

These difficulties in using GIS are due to a number of historical and practical reasons. GIS software packages generally approach geospatial problems in terms of abstract geometrical objects and operations, that is, computations on points, lines, and polygons. This imposes a heavy burden on the users because the first task in solving any problem with GIS is to map real-world problems into an environment where GIS techniques and tools can be used. This task is further complicated by the fact that problems in different application domains are often treated differently in GIS. For example, applications in ecology usually involve large-scale raster data and spatiotemporal analysis for visualization purposes, while applications in urban navigation generally involve a smaller geographic extent and are concerned mainly with real-time decision information. In addition, GIS software packages have historically been developed independently with little regard to data sharing (Goodchild, Egenhofer, & Fegeas, 1998). Different GIS software packages use their own proprietary formats, schemas, and terminologies to represent geospatial data and concepts. This has exacerbated the issue of data use, especially when they are to be shared, requiring manual conversions or availability of import and export tools. This process is often nontrivial and, considering the large volumes of data commonly required in GIS projects, is also very time consuming.

MAIN THRUST OF THE ARTICLE

Information Heterogeneity

The aforementioned issues are related to interoperability. The basis for problems related to interoperability is *information heterogeneity*, which is divided into three levels (Sheth, 1999): *syntactic heterogeneity*, which refers to the differences in formats and data types; s*tructural heterogeneity*, which deals with the differences in data-modeling constructs and schemas; and *semantic heterogeneity*, which refers to the variations of the intended meanings of concepts and terminologies. Table 1 provides examples of information heterogeneity in GIS.

The issues of syntactic and structural heterogeneities have been extensively addressed in the past within the computer and information science discipline. Recently, much research has been focused on addressing the issue of semantic heterogeneity, which is a significant problem in the field of GIS. In general, semantic heterogeneity is a result of different conceptualizations and representations of things in the world and can be distinguished into two types (Bishr, 1998).

Cognitive heterogeneity, which arises when two groups of people from different disciplines conceptualize the same real-world facts differently. As an example, a geologist thinks of hill slopes as areas where soil erosion or landslides can occur, but a tourist manager may think of hill slopes as areas where skiing is possible (Dehn, Gartner, & Kikau, 1999).

Naming heterogeneity, which arises when different names are used for identical concepts of real-world facts. For example, *hill slope* is also known as *valley side*, *mountain flank*, or simply *slope*.

Due to the widespread use of GIS by users both within and across disciplines, semantic heterogeneity in GIS is increasingly becoming an important issue in the GIS community. In the first example illustrated by Lutz, Riedemann, and Probst (2003), the semantic of the *touch* topological operator in the GeoMedia Professional GIS software package is different from that of Oracle 9i Release 2 Spatial (Table 2.). In GeoMedia, two polygons would satisfy the touch operator if their boundaries and/or interiors intersect. In Oracle, on the other hand, two polygons would satisfy the touch operator only if their boundaries, and not their interiors, intersect.

Furthermore, two GIS software packages may use

Table 1. Information heterogeneity in GIS

Information Heterogeneity	**Examples**
Semantic	Different behaviors of the "intersect" operation from different GIS software packages Different interpretations of the word "within" in a user's query
Structural	Different data dictionaries when merging two or more data sets Different metadata standards
Syntactic	Different data formats (e.g., Shapefile, ASCII [American Standard Code for Information Interchange], XML [eXtensible Markup Language])

Table 2. Topological relationships between two polygons and whether they satisfy the "touch" operator invoked in two different GIS software packages

		
GeoMedia Professional GIS	YES	YES
Oracle 9i Release 2 Spatial	YES	NO

different names for the same spatial operation. For example, the operation for aggregating polygons based on an attribute is called *dissolve* in the ArcGIS software package, but may be known by others as a *merge* operation (Figure 1).

These occurrences of semantic heterogeneity can lead to confusion and unexpected outcomes for users who need to deal with multiple GIS platforms or interact with other users who use different GIS platforms. The reconciliation of the differences in semantics must be accomplished by all parties involved in order for them to interoperate.

In another example of semantic heterogeneity, a German motorist in 1998 drove his car into a river after following instructions given by its navigation system. Though there may be other factors that led to the accident, Raubal and Kuhn (2004) hypothesize that the technical factor was that the in-car navigation computer did not make the distinction between a *bridge*, which is a permanent pathway, and a *ferry*, which is a transport carrying cars across a river. Though both a bridge and a ferry are pathways that can be used for route computations, a crucial semantic distinction must be made between them when instructions are given to drivers to account for the nonpermanent nature of the ferry.

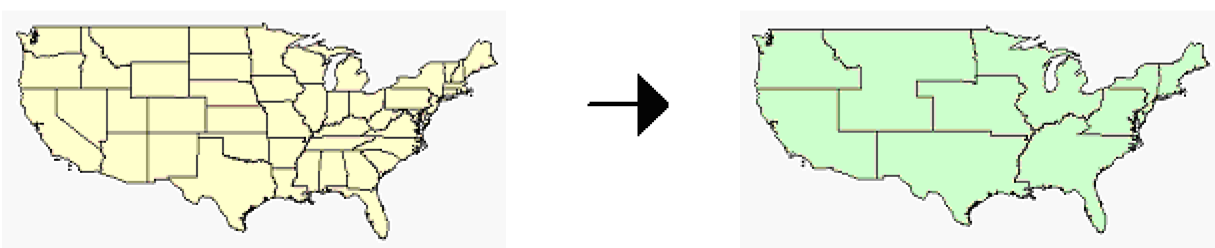
To overcome information heterogeneity in GIS and provide interoperability in GIS platforms, users, and data, there needs to be an agreement among the parties involved.

The Open GIS Consortium Standards

Many standards have been defined for the GIS domain to allow interoperability among different GIS platforms. However, currently the most prominent standard body for GIS is the Open Geospatial Consortium (http://www.opengeospatial.org). OGC is an organization consisting of companies, government agencies, and universities participating in a consensus process to develop publicly available geoprocessing specifications that result in interoperability among diverse GIS platforms. The creation of the OGC was to address interoperability issues among different GIS platforms, particularly when they are used in the Internet environment. The development process of the OGC involves the creation of abstract specifications and implementation specifications.

The purpose of abstract specifications is to create the conceptual foundation that facilitates understanding of real-world geospatial phenomena, and allow for the development of implementation specifications by precisely capturing and stating requirements and knowledge of the abstract geospatial domain. Essentially, abstract specifications describe how "ideal" software should work, and they include, but are not limited to, topics on feature geometry, topology, coordinate reference systems, and geospatial metadata. Abstract specifications mainly concern abstract geospatial objects and concepts applicable to GIS (e.g., point, line, and polygon) and do not specifically address real-world, application-

Figure 1. Operation for aggregating areas based on an attribute may be called differently in different GIS software packages



context concepts and terminologies (e.g., street, river, forest). For example, the topic *Feature Geometry*, which is also a draft international standard (ISO 19107 Spatial Schema; Herring, 2001), is the standard that specifies geometrical and topological objects as well as operations which can be applied on them.

Implementation specifications provide programmers with specific programming rules and advice for implementing interfaces and protocols that enable interoperability between different GIS platforms. They are engineering specifications that implement part of the abstract specification for particular distributed computing platforms. For example, the Web Map Service (WMS) Implementation Specification specifies the interface for providing mapping services over the Web (Beaujardiere, 2002). Another example, the Geography Markup Language (GML), is a language designed to be a general data format for modeling, transporting, and storing geospatial information (Cox, Daisey, Lake, Portele, & Whiteside, 2003). GML is an XML (eXtensible Markup Language) grammar written in XML schema that provides a variety of kinds of objects for describing geography as defined in abstract specifications, including features, coordinate reference systems, geometry, topology, time, and units of measurement.

Although the OGC standards address many aspects of information heterogeneity in geoprocessing, they are designed by GIS experts for use by GIS experts in implementing GIS projects. Missing from the OGC standards, however, is the issue of how to allow non-expert users to realize the potential of GIS by making the task of geospatial problem solving easier through semantic interoperability in the application-domain context.

Ontological-Based GIS

There are two key aspects related to semantic interoperability in GIS. First, there is a need for semantic agreement on geospatial data models (e.g., point, line, polygon) and geoprocessing operations (e.g., buffering, intersection). Second, there is a need for semantic agreement about real-world, application-level geographic objects, concepts, and terminologies used in geospatial problem solving. The OGC standards address the first aspect of semantic interoperability by providing a uniform definition and behavior of abstract geometrical objects and geoprocessing operations. However, it does not address the second aspect of semantic agreement regarding the geographic world and application-domain contexts. As previously discussed, one difficulty in using GIS is the mapping of real-world problems into a form which GIS understand. This is arguably the first and most important task of problem solving in GIS, and current technology does not have a means to automate it.

An approach to address the problem is by incorporating *ontologies* into GIS, which would provide shared bodies of semantic knowledge of the geospatial domain. An ontology is a specification of a conceptualization (Gruber, 1993) that allows parties who agreed to an *ontological commitment* to communicate with one another and share knowledge. It may include a dictionary of terms and a specification of their intended meanings. The concepts defined in an ontology and how they are interrelated collectively impose a structure on the domain and constrain the possible interpretations of terms (Uschold & Jasper, 1999). In the information-system context, ontologies are machine-processable bodies of knowledge. As such, an ontological-based GIS would allow the use of geospatial information based primarily on its meaning (Fonseca, Egenhofer, Agouris, & Camara, 2002).

In an ontological-based GIS, an ontology would include concepts and terminology about an application domain that users can directly relate to and use. For example, an ontology would define terms specific to the application domain of ecology and how they can be interpreted. This ontology can then be used by the users to formulate their geospatial queries that would conform to the knowledge defined in the ontology. Furthermore, the OGC geoprocessing standards can be considered as another distinct ontology that constrains the meaning and behavior of geometrical objects and operations. Since solving geospatial problems using GIS involves mapping real-world queries into geometries and operations, bridging the two ontologies would provide the means for interpretation of geospatial queries by computers.

Recent research efforts on geospatial ontology include cognitive and philosophical aspects on how the real world should be modeled and formalized into ontologies (Mark, Freksa, Hirtle, Lloyd, & Tversky, 1999; Mark, Smith, & Tversky, 1999; Smith & Mark, 1998, 2001), as well as how to use ontologies in GIS (Fonseca et al., 2002; Karimi, Akinci, Boukamp, & Peachavanish, 2003; Kuhn, 2001; Raubal & Kuhn, 2004; Visser, Stuckenschmidt, Schuster, & Vogele, 2002).

FUTURE TRENDS

One of the goals of GIS research, explicitly stated or not, is to advance the technology to a point where it can be used as decision-support systems assisting users in solving a wide variety of problems in many applications. We consider GIS to be decision-support systems when they are easy to use by all users with different backgrounds, able to solve complex problems that otherwise are handled inefficiently, semantically interoperable, and equipped with knowledge and reasoning to provide automated decision-making tasks, especially in real-time applications.

This need is evident by the proliferation of application-specific GIS on the Internet (e.g., Web sites that provide driving directions) and in other distributed environments (e.g., location-based wireless real-time services). To thrive in these heterogeneous, distributed environments, GIS platforms must support interoperability. Additionally, much research is still needed to unlock the potential of GIS to ordinary users for solving complex problems. For instance, semantic integration into GIS through ontologies is a research area that would facilitate interoperability at a higher level, allowing the use of GIS by many users with little GIS background, lowering geoprocessing costs, and increasing the usefulness of GIS in general.

CONCLUSION

GIS have come a long way from being simple stand-alone tools that facilitated digital mapping and primitive geoprocessing to information systems capable of performing sophisticated geoprocessing on stand-alone or distributed platforms. This evolution took over 4 decades and was made possible through advances in computer geometry, database systems, personal computers, Internet, and other techniques and technologies. During the same period, the number of applications that adopted GIS technology increased. Today, numerous applications utilize GIS technology to solve a range of simple to complex problems. However, despite the complex operations current GIS support and the widespread applications in which they are employed, they markedly lack the ability to interoperate due to various historical and practical reasons. Advances in key areas are needed before GIS become more interoperable and accessible to all users, novice or expert, paving the way for the emergence of new applications.

REFERENCES

Beaujardiere, J. (2002). Web map service implementation specification. Retrieved from *http://www.opengis.org/techno/specs/02-058.pdf*

Bishr, Y. (1998). Overcoming the semantic and other barriers to GIS interoperability. *International Journal of Geographical Information Science, 12*(4), 299-314.

Cox, S., Daisey, P., Lake, R., Portele, C., & Whiteside, A. (2003). Open GIS Geography Markup Language (GML) implementation specification. Retrieved from *http://www.opengis.org/techno/documents/02-023r4.pdf*

Dehn, M., Gartner, H., & Kikau, R. (1999). Principles of semantic modeling of landform structures. Paper presented at *GeoComputation '99*, Fredericksburg, VA.

Fonseca, F. T., Egenhofer, M. J., Agouris, P., & Camara, C. (2002). Using ontologies for integrated geographic information systems. *Transactions in GIS*, *6*(3), 231-257.

Goodchild, M. F., Egenhofer, M. J., & Fegeas, R. (1998). *Interoperating GISs: Report of the specialist meeting*. Santa Barbara, CA: National Center for Geographic Information and Analysis, University of California.

Gruber, T. R. (1993). Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human and Computer Studies, 43*(5/6), 907-928.

Herring, J. (2001). *The OpenGIS Abstract Specification, Topic 1: Feature Geometry (ISO 19107 Spatial Schema), version 5* [OGC Document No. 01-101]. *http://www.opengeospatial.org/docs/01-101.pdf*

Karimi, H. A., Akinci, B., Boukamp, F., & Peachavanish, R. (2003). Semantic interoperability in infrastructure systems. Paper presented at the *Fourth Joint Symposium on Information Technology in Civil Engineering*, Nashville, TN.

Kuhn, W. (2001). Ontologies in support of activities in geographical space. *International Journal of Geographical Information Science, 15*(7), 613-631.

Lo, C. P., & Yeung, A. K. W. (2002). *Concepts and techniques of geographic information systems*. Upper Saddle River, NJ: Prentice Hall.

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2001). *Geographic information systems and science*. Chichester, England/New York: Wiley.

Lutz, M., Riedemann, C., & Probst, F. (2003). A classification framework for approaches to achieving semantic interoperability between GI Web services. Paper presented at the *Conference on Spatial Information Theory 2003*, Ittingen, Switzerland.

Mark, D. M., Freksa, C., Hirtle, S., Lloyd, R., & Tversky, B. (1999). Cognitive models of geographical space. *International Journal of Geographical Information Science, 13*(8), 747-774.

Mark, D. M., Smith, B., & Tversky, B. (1999). Ontology and geographic objects: An empirical study of cognitive categorization. In *Lecture notes in computer science: Vol. 1661, Spatial information theory: A theoretical basis for GIS* (pp. 283-298). Berlin/Heidelberg, Germany: Springer-Verlag.

Raubal, M., & Kuhn, W. (2004). Ontology-based task simulation. *Spatial Cognition and Computation, 4*(1), 15-37.

Sheth, A. (1999). Changing focus on interoperability in information systems: From system, syntax, structure to semantics. In M. F. Goodchild, M. J. Egenhofer, R. G. Fegeas, & C. A. Kottman (Eds.), *Interoperating geographic information systems*, (pp.165-180). Norwell, MA: Kluwer.

Smith, B., & Mark, D. M. (1998). *Ontology and geographic kinds*. Paper presented at the International Symposium on Spatial Data Handling (SDH'98), Vancouver, Canada.

Smith, B., & Mark, D. M. (2001). Geographical categories: An ontological investigation. *International Journal of Geographical Information Science, 15*(7), 591-612.

Uschold, M., & Jasper, R. (1999). *A framework for understanding and classifying ontology applications*. Paper presented at the IJCAI'99 Workshop on Ontology and Problem Solving Methods: Lesson Learned and Future Trends, Stockholm, Sweden.

Visser, U., Stuckenschmidt, H., Schuster, G., & Vogele, T. (2002). Ontologies for geographic information processing. *Computers & Geosciences, 28*, 103-117.

KEY TERMS

Geoprocessing: Operations in GIS for integrating, analyzing, computing, and presenting geospatial data.

Geospatial Data: Data representing objects on or near the surface of the earth.

Geospatial Information Systems (GIS): Information systems capable of storing, managing, computing, and displaying geospatial data for solving geospatial problems.

Geospatial Problems: Problems involving geospatial data, objects, and phenomena.

Information Heterogeneity: The differences in syntax, structure, and semantics used in different information systems.

Interoperability: The ability of two or more heterogeneous systems to work together in a seamless manner.

Ontology: A conceptualization and representation of objects and phenomena and the relationships among them in a domain.

Standard: An agreed-upon set of concepts, terminologies, and methodologies by a given community.

Interoperability of Information Systems

Christophe Nicolle
Université de Bourgogne, France

Jean-Claude Simon
Université de Bourgogne, France

Kokou Yétongnon
Université de Bourgogne, France

INTRODUCTION

An information system is a multilevel system characterized by a "data" level, a "behavioral" level, and a "communication" level. The data level represents the data stored by the system. The behavioral level represents management and production processes carried out by the system. The processes can interact with the data level to extract, generate, and store data. The communication level relates to the network used to exchange data and activate processes between geographically distant users or machines.

Information system interoperation has emerged as a central design issue in Web-based information systems to allow data and service sharing among heterogeneous systems. Data heterogeneity stemming from the diversity of data formats or models used to represent and store information in the Web is a major obstacle to information systems interoperability. These data models range from the structured data models [network, relational, object oriented (OO)] found in traditional databases to flat files and emerging Web-oriented semistructured models. Information system interoperability aims at supporting the amalgamation of autonomous heterogeneous systems to create integrated virtual environments or architectures in which information from multiple disparate sources can be accessed in a transparent and efficient manner. As an example of such integrated virtual systems, consider an airline reservation system based on the integration of a group of airlines reservation and ticket sale information systems. The specific airline systems provide various types of fares and special discount trips that can be searched and compared to respond to user queries for finding the best available prices for specified flights.

BACKGROUND

Database interoperability issues have been extensively studied in the past. Several approaches, including database translation, distributed systems, federations, language-based multidatabase, ontology, and mediation, have been proposed to bridge the semantic gaps among heterogeneous information systems.

The database translation approach is a point-to-point solution based on direct data mappings between pairs of information systems. The mappings are used to resolve data discrepancies among the systems (Yan & Ling, 1992). The database translation approach is most appropriate for a small-scale information-processing environment with a reduced number of participants. The number of translators grows with the square of the number of components in the integrated system. For example, consider two information systems IS1 and IS2 in the travel agency example above. The corresponding translators must be placed between the information systems as shown in Figure 1. Information in IS1 is represented by vertical lines, while the information in IS2 is shown as horizontal lines.

In the standardization approach (Figure 2), the information sources use the same model or standard for data representation and communication. The standard model can be a comprehensive metamodel capable of integrating the requirements of the models of the different components (Atzeni & Torlone, 1997). The use of a standard metamodel reduces the number of translators (this number grows linearly with the number of components) to resolve semantic differences. However, the construction of a comprehensive metamodel is difficult; the manipulation of high-level languages is complex; and there are no unified database interfaces. In our example, the travel agencies must define a common model to export their data. A centralized information system can be built to replace the original information systems (IS1, IS2). The global centralized schema is a combination of the data (horizontal and vertical lines) contained in IS1 and IS2.

Federated systems (Figure 3) consist of a set of heterogeneous databases in which federation users can access and manipulate data transparently without knowledge of the data location (Sheth & Larson, 1990). Each federation database includes a federated schema that incorporates the data exported by one or more remote

Figure 1. Database translation approach

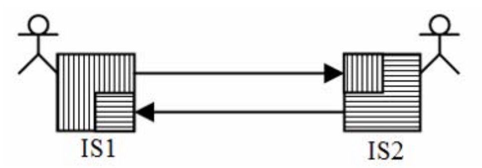


Figure 2. Standardization approach

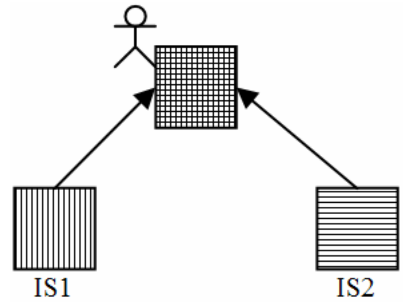


Figure 3. Federated systems

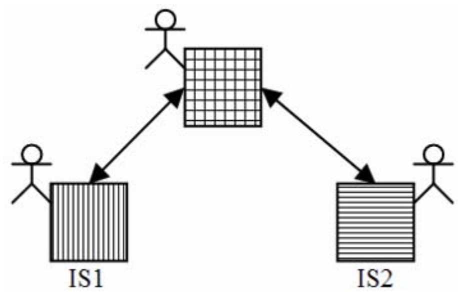


Figure 4. Multibase systems

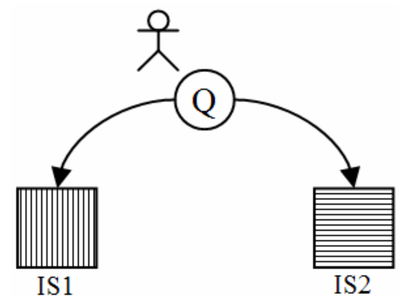


Figure 5. Ontology approach

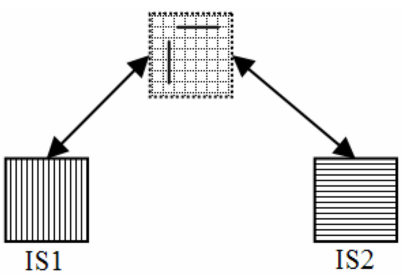
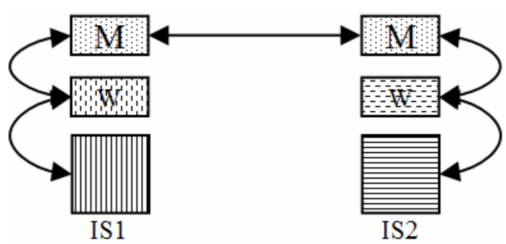


Figure 6. Mediation approach



information systems. There are two types of federations. A tightly coupled federation is based on a global federated schema that combines all participant schemas. The federated schema is constructed and maintained by the federation administrator. A loosely coupled federation includes one or more federated schema that are created by users or the local database administrator. The federated schema incorporates a subset of the schema available in the federation. This approach becomes rapidly complex when the number of translators required becomes large. In our example, the existing information systems are completely operational for local users. Only the shared data are integrated in the federated schema. The federated system is made only of horizontal and vertical lines that IS1 and IS2 want to exchange.

Language-based multibase systems (Figure 4) consist of a loosely connected collection of databases in which a common query language is used to access the contents of the local and remote databases (Keim, Kriegel, & Miethsam, 1994). In this approach, in contrast to the distributed and federated systems, the burden of creating the federated schema is placed on the users, who must discover and understand the semantics of the remote databases. In our example, the various companies have to define a global common language (Q) to query their information systems (IS1, IS2). This solution is well adapted for information systems that are based on the same family of data models and do not require complex query translators.

The ontology-based interoperability approach (Figure 5) uses ontology to provide an explicit conceptualization of the common domain of a collection of information systems (Benslimane, Leclercq, Savonnet, Terrasse, & Yétongnon, 2000). An ontology defines a common vocabulary that can be used by users from different systems. The construction of an ontology for a domain is a difficult task and often requires merging existing overlapping ontologies. The interoperability solutions based on ontology describe the semantics of information rather than their organization or their format. In our example, the companies have to define ontology to capture the semantics of their domain of activity.

The mediation approach (Figure 6) is based on two main components: mediator and wrapper. The mediator is used to create and support an integrated view of data over multiple sources. It provides various services to support query processing. For instance, a mediator can cooperate with other mediators to decompose a query into subqueries and generates an execution plan based on the resources of the cooperating sites. The wrapper is used to map the local databases into a common federation data model. The wrapper component provides the basic data access functions (Garcia-Molina, Hammer, Ireland, Papakonstantinou, Ullman, & Widow, 1995). In our example, a translator, which acts as a wrapper, is placed between the conceptual representation of the mediator and the local description of each information source.

Table 1. Overview of architectures for interoperable information systems

Systems	**Advantages**	**Limits**	**Tools or methods used**	**Levels**
Translation	• Better control of point-to-point translation	• Requires a large number of translators in open environments • Adding a new information system requires 2(*n* - 1) translators	Required *n**(*n* - 1) translators	☑ D ☐ B ☐ C
Standard-ization	• Use of pivot, canonical model or metamodel • Reduce the number of translators	• Definition of a common standard accepted by all IS • The construction of a comprehensive metamodel is difficult	Required 2*n* translators	☑ D ☐ B ☐ C
Federation	• Derived from standardization • Local IS are autonomous	• Use of a global, static federated schema • The construction of an integrated federal schema is difficult • New addition requires redesign of federated schema	Required 2*n* translators	☑ D ☐ B ☐ C
Multi-base	• Used of a single language for many IS	• The common interoperating language does not export local system semantics • Users need to discover and understand the semantics of remote IS	Query based	☑ D ☑ B ☐ C
Ontology	• Semantic-oriented solution	• Extensive ontologies are voluminous • Requires meta-level translation	Semantic	☑ D ☐ B ☐ C
Mediation	• Combine translation and semantic • Local IS are autonomous	• Difficult to construct automatic mediator process	Required 2*n* semantic translators	☑ D ☑ B ☐ C

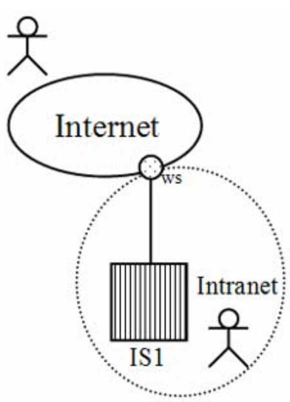
Table 1 summarizes the various architectures for the interoperation of information systems. In this table, a brief presentation of the advantages and limits of each approach is given.

FUTURE TRENDS

As new data models are developed for Web-based information systems, there is a need to extend interoperability solutions to take into account requirements and specifications of the new models. For instance, *XML* (XML, 2004) emerged as an important model for describing and sharing Web-based data. This importance stems from two major factors. First, XML is becoming a de facto data standard supported by many software vendors and applications developers. Second, XML is based on a relatively simple structure that is both user and machine readable and that can be used by nonexpert database administrators. The existing Web technologies are not initially intended to address some of the issues involved in database integration. For instance, the Web-browsing paradigm is efficient for data lookup in a large environment, but it is inadequate for database integration support. To use this paradigm to locate and merge data requires costly applications that are often tailored to specific integration needs. New challenges have arisen from the development of Web-based information systems. One of the challenges is the need to develop Web-oriented tools to support information integration and allow access to local as well as remote information sources.

Recently, Web services (WS) have been proposed as a method to address some of the challenges of Web-based integrated systems. A Web service can be viewed as a set of layers contained in a stack (Figure 7). The layers are dynamically defined following user needs and are called through a set of Internet protocols. The protocols are different than those proposed for various network architectures. However, in all Web service architectures, a base set of protocols is always used (W3C, 2002). This base set is composed of SOAP (SOAP, 2003), WSDL (WSDL, 2003), and UDDI (UDDI, 2002). They allow for the discovery, description, and information exchanges between Web services.

Figure 7. WS approach



SOAP is a mechanism that uses XML for the exchange of structured and typed information between several actors in a decentralized and distributed environment. SOAP does not define the semantics of the application but provides a mechanism for expressing semantics by proposing a modular template and mechanisms for data coding.

WSDL uses XML syntax to describe the methods and parameters of Web services. These parameters include protocols, servers, ports, input and output messages format, and exceptions format. With WSDL, an application using SOAP can autoconfigure the Web services exchanges, masking the majority of the low-level technical details.

UDDI is a Web-based company, world directory, combining “white pages” (information such as name, address, telephone number, and other contact information of a given business), “yellow pages” (information that categorizes businesses). and “green pages” (technical information about the Web services provided by a given business). UDDI allows Web service references by automating all search procedures. Table 2 presents the advantages and limits of Web services.

In our example, a set of Web services can be built from each information system independent from the other information systems. The Web services become a standard interface to access the local information system. These Web services can be used by customers and partners via the Internet and by local users via an intranet. This solution is flexible and reduces the complexity of the heterogeneity problem.

To achieve a Web service architecture, several industrial tools have been developed. Four main actors in the industrial world share the market. The solutions proposed by Microsoft and SUN are language oriented, while the solutions proposed by IBM and BEA are platform oriented.

Microsoft .NET proposes a software platform on which companies can exchange data and services on the Internet based on an ASP model (application provider service). Most Microsoft products can be extended to use Web services developed with the .NET. The philosophy of this solution can be resumed by “one OS, many languages.”

The *SUN J2EE* is developed by the Java Community Process. It is a set of services and specifications containing JDBC (Java database connector), JMS (Java Message Services), JSP (Java Server Pages), EJB (Enterprise Java Beans), etc. J2EE 1.4 includes Web service specifications

Table 2. Web services, new architecture for interoperability

Systems	Advantages	Limits	Tools or methods used	Levels
Web services	• Resolved format level • Resolved process translation • All levels of IS are managed • Normalized solution • Developed by industrials and researchers	• Security mechanism not finalized • Combination of Web services not resolved\	Protocol SOAP, WSDL, UDDI	☑ D ☑ B ☑ C

using an open-source framework called AXIS (used by IBM WebSphere). In response to the Microsoft .NET solution, SUN proposes ONE, which groups the set of SUN Web services propositions. The philosophy of this solution can be resumed by "many OS, one language."

IBM WebSphere is a set of components allowing the creation of interoperable information systems based on Web services. These components include Interchange Server, which allows process integration; MQ Integrator Broker, which allows data integration; MQ Workflow, which allows processes management, etc. IBM WebSphere uses the SUN JAVA language for the development of its Web services.

The *BEA WebLogic Server* is based on the Java Connectors architecture. This tool uses the notion of components and connectors that can be integrated between them. The integration of the connectors is carried out by the Application Integration framework and the Adapter Development Kit. To manage the resulting architecture, business process management is used in coordination with the business-to-business (B2B) integration tool. This tool exploits standards, such as XML, HTTP, or SSL, and semantic solutions, such as RosettaNet, cXML, ebXML, and EDI.

CONCLUSION

For the past 20 years or so, the need to exchange information between various partners pushed researchers to develop architectures for the interoperability of information systems. The proposed architectures have addressed several key interoperability issues, ranging from the resolution of data format heterogeneity using translations-based architecture and the reduction of the number of required translators in standardization-based architecture to the resolution of semantic heterogeneity based on ontology, and the resolution of process heterogeneity with mediation-based architecture.

Nowadays, information systems can be integrated or disassociated depending on the market trends of enterprise mergers. The Web-service-based architecture allows the development of this type of interoperability by proposing a standard data format with XML, a standard communication architecture based on the SOAP protocol, and a standard description of processes using WSDL and UDDI. The next major challenge in the Web service world is to extend Web services to include security, data owner, and semantics.

REFERENCES

Atzeni, P., & Torlone, R. (1997). MDM: A multiple-data-model tool for the management of heterogeneous database schemes. In *Proceedings of the SIGMOD International Conference* (pp. 538–531).

Benslimane, D., Leclercq, E., Savonnet, M., Terrasse, M. N., & Yétongnon, K. (2000). On the definition of generic multi-layered ontologies for urban applications. *International Journal of Computers, Environment and Urban Systems, 24*(2000), 191–214.

Garcia-Molina, H., Hammer, J., Ireland, K., Papakonstantinou, Y., Ullman, J., & Widow, J. (1995, March). Integrating and accessing heterogeneous information sources in TSIMMIS. In *Proceedings of the AAAI Symposium on Information Gathering* (pp. 61–64).

Keim, D. A., Kriegel, H. P., & Miethsam, A. (1994). Query translation supporting the migration of Legacy Database into cooperative information systems. In *Proceedings of the Second International Conference on Cooperative Information Systems* (pp. 203–214).

Sheth, A. P., & Larson, J. A. (1990). Federated database systems for managing distributed heterogeneous, and autonomous databases. *ACM Computing Surveys, 22*(3).

SOAP. (2003). SOAP, Version 1.2, Part 0: Primer, W3C Recommendation. Retrieved June 24, 2003, from *http://www.w3.org/TR/soap12-part0/*

UDDI. (2002). UDDI, Version 3.0, UDDI Spec Technical Committee Specification. Retrieved July 19, 2002, from *http://uddi.org/pubs/uddi_v3.htm*

W3C. (2002). Web services architecture requirements, World Wide Web Consortium (W3C), Working Draft. Retrieved November 14, 2002, from *http://www.w3.org/TR/2002/WD-wsa-reqs-20021114*

WSDL. (2003). Web Services Description Language (WSDL), Version 1.2, Part 1: Core Language W3C Working Draft. Retrieved June 11, 2003, from *http://www.w3.org/TR/wsdl12*

XML. (2004). Extensible Markup Language 1.0 (3rd ed.), W3C Recommendation. Retrieved February 4, 2004, from *http://www.w3.org/TR/REC-xml*

Yan, L. L., & Ling, T. W. (1992). Translating relational schema with constraints into OODB schema. In *Proceedings of the IFIP WG2.6 Database Semantic Conference on Interoperable Database Systems* (DS-5) (pp. 69–85), Lorne, Victoria, Australia, 16–20 November.

KEY TERMS

Interoperability: The ability of heterogeneous software and hardware to communicate and share information.

Ontology: An explicit formal specification of how to represent the objects, concepts, and entities existing in some area of interest and the relationships among them.

SOAP (Simple Object Access Protocol): An XML-based message protocol used to encode information in Web service requests and response messages before sending them over a network. SOAP messages are independent of any operating system or protocol and may be transported using Internet protocols (SMTP, MIME, and HTTP).

UDDI (Universal Description, Discovery, and Integration): A Web-based distributed directory for discovery of Web services offered by companies. It is similar to a traditional phone book's yellow and white pages.

Web Service: A software system designed to support interoperable machine-to-machine interaction over a network. It has an interface described in a machine-readable format (specifically, WSDL). Other systems interact with the Web service in a manner prescribed by its description using SOAP messages, typically conveyed using HTTP with an XML serialization in conjunction with other Web-related standards.

WSDL (Web Services Language Description): An XML-formatted language used to describe a Web service's capabilities as collections of communication endpoints capable of exchanging messages.

XML: A language for creating markup languages. There are two kinds of XML documents: well-formed and valid. The first respects the XML standard for the inclusion and the names of the tags. The second must be well-formed and uses a grammar to define the structure and the types of data described by the document.

Intranet Use and the Emergence of Networks of Practice

Emmanuelle Vaast
Long Island University, USA

INTRODUCTION

Communities of practice (CoPs) are key to today's knowledge management (Schultze & Leidner, 2002; Von Krogh, 2002). Moreover, the capability of exchanging professional knowledge beyond distance has become a strategic asset for innovative firms. How can members of local CoPs exchange knowledge with remote colleagues and create networks of practice (NoPs)? This article contends that the use of information technology (IT), and more specifically, of intranet systems, is especially suited to link local CoPs to an overall network of practice.

BACKGROUND

Communities of practice are social groupings whose members work in the same material context, interact frequently, acquire common knowledge, and experience similar professional concerns (Brown & Duguid, 1991; Lave & Wenger, 1991; Wenger, 1998). Members of CoPs work together and achieve activities that are for some similar and for others complementary (Gherardi & Nicolini, 2000). As they share the same work environment, they have frequent occasions to discuss directly about their job and unusual issues (Orr, 1990). Communities of practice unfold from a shared situation that creates a context favorable to direct encounters, mutual assistance in practice, and collective goals (Iverson & McPhee, 2002). Even though members of a CoP may not spontaneously name their workgroup a community, they usually acknowledge their membership to their occupational group and value its rules and principles.

CoPs display three distinguishing features: mutual engagement, joint enterprise, and common repository (Wenger, 1998).

- *Mutual engagement*: People join a CoP by committing themselves in actions whose meaning is mutually negotiated. Members of a CoP are related to each other through their mutual engagement in social practices.
- *Joint enterprise*: The community exists and provides social support and identity to its members to favor the achievement of common goals. These objectives may be explicit or not, officially defined or not, but members of the community engage themselves to complete them.
- *Common repository*: Over time, shared practices, repeated interactions, and the emergence of a shared culture provide traces of the community. Its members may refer to a common repository to deal with daily or more unusual issues. This repository may be material and concrete (files, forms) or more intangible (routines, specific idioms).

The network of practice extends the notion of CoPs beyond geographical distance. NoPs relate local CoPs whose respective members share occupational competences, job duties, and tasks, but who do not directly interact because of geographical distance (Brown & Duguid, 2000, 2001). As the literature on this notion is extremely recent, the appellation has not been stabilized yet. Some refer to "constellations of practice" or to "virtual communities of practice" (Gherardi & Nicolini, 2000). This article relies on the notion of "network of practice" as the most direct extension of CoPs beyond geographical distance. This phrase also explicitly accounts for the practice foundation of both communities and networks of practice.

People who are not collocated and do not necessarily know each other, but still achieve the same kinds of activities and experiment with similar identification processes belong to an NoP (Vaast, 2004). The relationships among members of an NoP are looser than the ones that characterize CoPs. Members of the NoP can nevertheless exchange on occupational issues. Although each local community displays idiosyncratic features, the overall network is characterized by shared knowledge, culture, and patterns of action. To some extent, the NoP also experiments mutual engagement, joint enterprise, and common repository.

The ways in which local CoPs may get connected to each other and favor the emergence of an NoP are, however, anything but obvious. Given that CoPs rely heavily on the sharing of a material context and on situated recurrent direct interactions, how may these local CoPs get connected into a network of practice?

It has been proposed that specific IS may favor CoPs (Brown, 1998) and may help spread knowledge among communities (Pan & Leidner, 2003). More specifically, the use of intranet systems seems especially suited to relate communities and networks of practice (Vaast, 2004).

INTRANETS, IDEAL TOOLS FOR COPS AND NOPS

Intranets are internal networks based on Web standards that aggregate and integrate various computing applications, such as e-mail, databases, groupware systems, or forums (Bansler, Damsgaard, Scheepers, Havn & Thommesen, 2000; Curry & Stancich, 2000; Ryan, 1998). Since 1995, intranets have represented a major growth area in corporate computing thanks to the availability of standard network technologies like Ethernet, TCP/IP, Web browsers, and servers. They have become increasingly more sophisticated and have integrated dynamic databases and various occupational applications. *Intranets* are private networks that favor flows of information and applications among members of an organization or parts of it (Newell, Scarbrough & Swan, 2001). Specific groups may implement and appropriate their own intranet, and protect it with passwords and various levels of security. Moreover, intranets may easily be customized to various contexts and end-user needs.

Key features of intranets seem appropriate to fulfill the needs of CoPs and NoPs:

Interoperability: Based on universal Web standards, intranets connect local computing networks and unify multiple software systems. Interoperability is useful to connect various local groups, to create room for communication, and to share applications among members of diverse communities.

Cost- and time-efficiency: The wide availability of standard TCP/IP protocols and of other network standards have recently made the implementation of intranets easy, fast, and reasonably priced. Basic intranets only require the availability of one server and of local computers equipped with a browser and connected to the network. Thus, even informal communities may implement and appropriate their own intranet. This ensures that the specific needs of occupational groups are taken into account in the system.

Flexibility: Typical of the new generation of information systems, intranets are also highly flexible and may include multiple applications. IT professionals as well as end-users may thus customize them to take into account the specific needs of their occupational groups. Flexible intranets may also be transformed and enriched over time. As agents become more familiar with the network, they can upgrade or introduce changes. Moreover, the flexibility of intranets makes it possible to adapt them, along with the emergence of an NoP from local CoPs. In particular, as the NoP emerges, communication features (through e-mail, FAQ or forum systems, chat) may become increasingly critical to create and maintain links among local communities.

Privacy: Intranets are private networks. Their design and architecture restrict access to authorized users. For instance, firewalls screen requests to the servers to make sure that they come from acceptable domain names and IP addresses. Mobile users may access the private network thanks to secure logon procedures and authentication certificates. Various levels of confidentiality also ensure that members of CoPs feel that their computing network is to be used only by peers and that outsiders will not intrude into the most private parts of the system (such as the ones that deal with occupational applications). Moreover, access rights and authentication procedures allow for differentiated uses by localized employees. For instance, an intranet may simultaneously present information relevant to all local CoPs and include sub-parts or folders dedicated to specific CoPs. Discriminate access in and among local communities favors the exchange of information and encourages the building of trust throughout the NoP.

User-friendliness: Based on hypertext interfaces and on graphical commands, most intranet systems are intuitive to use. Thus, no matter whether end-users are computer literate or not, they can easily learn how to make good use of the resources the intranet systems provide. As human-computer interactions are made easier and more intuitive, even members of CoPs who are not familiar with computers and computing networks may nevertheless consequently spontaneously appropriate their intranet. User-friendliness also favors end-users' willingness to improve features of the system and to adapt it to fit the communication needs of the NoP.

Two Examples of Intranet Use Creating Links Between CoPs and NoPs

The following examples show how an NoP emerges from the use of an intranet system by members of local CoPs.

Insurance Company Vendors

Thirty-five hundred vendors of an insurance company were geographically dispersed and worked in local teams of about 15 people. Vendors in any one team had many activities in common with vendors in other teams, but traditionally most felt that they were in competition with other teams from the same geographical area. The central headquarters of the company introduced an intranet system dedicated to these professionals. The intranet was not widely used at first, because of limited IT competencies of salespersons and because of the perceived competition among teams. Gradually however, newcomers started to use the intranet and transmitted their expertise in browsing the intranet to old-timers. Members of local teams thus socialized around the intranet. Also, over time, the intranet was used to exchange professional experiences among local teams. It then favored the emergence of a feeling of membership to the same occupation beyond geographical distance.

Buyers of a Railroad Company

Buyers of a railroad company (about 2,500 employees working in local teams) implemented an intranet site dedicated to their work. They directly fully integrated use of an intranet into their daily business. Buyers used the intranet to order supplies and to get information about their shipments. By making it possible for distant local buyers to aggregate their orders, use of the intranet reinforced a pre-existing trend towards the greater centralization of procurement. It also deeply affected work processes and tasks: buyers' practices became more transparent to the other departments of the railroad company. The intranet also made buyers become more aware that, even though they worked in dispersed local services, they all belonged to the same occupational group. Finally, exchanges of experiences and electronic messages resulted in new communications among distant colleagues and favored the exchange of professional experiences and mutual learning.

THE USE OF INTRANETS RELATES COPS AND NOPS

The use of intranets favors the overlapping of local CoPs. When members of various situated CoPs appropriate an intranet system, they become aware that their occupational group is not restricted to their local community. Thanks to common repositories, shared databases, and the possibility to exchange electronically, members of different CoPs start interacting with remote colleagues. Intranet systems allow members of distant communities who share professional concerns to exchange knowledge and experiences with each other. Local communities remain geographically separate and their respective members do not interact directly, but the use of an intranet contributes to the overlapping of local CoPs and to the emergence of NoPs, as it increases the visibility of shared practice and favors more frequent electronic communications.

Furthermore, the availability of an intranet dedicated to a specific occupation favors identification to this occupation thanks to the external and internal recognition of the work accomplished by these professionals. As the intranet is visible (if not entirely open, because of security issues) to the overall organization, it publicizes the occupation throughout the organization. On the other hand, the availability of common information and applications increases the feeling of occupational membership by members of local CoPs and thus also increases the internal recognition of the occupation.

Specific features of intranets strengthen the links among CoPs and favor the emergence of NoPs thanks to the geographical extension of mutual engagement, joint enterprise, and shared repertoire from the community to the network level. Table 1 presents these features.

FUTURE TRENDS

Managers should be especially careful in their attempt to implement and manage intranets to relate CoPs and NoPs. In particular, managers have to deal with three delicate dilemmas: initiative vs. control, sharing vs. competitive emulation, and official vs. emergent processes.

Defining the intranet too tightly and preventing the expression of local initiatives presents obvious drawbacks, as members of CoPs will not appropriate systems that are imposed on them and that they are unable to customize. However, if too much room is left for local initiatives, multiple fragmented intranet sites may be created which do not favor the establishment of links among CoPs. Alternate phases in the management of the intranet may thus prove useful to deal with this dilemma between initiative and control. Initially, light management—as well as relative freedom of implementation and use—favors the appropriation of the intranet. Later, more control encourages expected positive outcomes at the level of the overall NoP.

The sense of community in CoPs and NoPs provides support to their members, but it does not always encourage the search of improved practices or new knowledge. Healthy emulation among members of CoPs thus fosters the continuous improvement of work process and the pursuit of innovation. On the other hand, leaving too much room for competition and for the expression of political struggles in the CoPs or the NoPs is detrimental to the sharing of valuable knowledge. It therefore seems important to ensure that a basic level of knowledge, practices, and rules are established at the network level to make the dialog among local communities possible *and* to provide incentives for competition and sharing throughout the NoP, notably through different sections of the intranet.

Table 1. Three dimensions of CoPs and NoPs favored by intranet use

	Intranet Applications and Features	***Links Between CoPs and NoPs***
Mutual Engagement	**Repertory:** Index of all users of the intranet, of all members of the profession.	Access to all members of the profession beyond geographical distance. Repertory with name, phone number, e-mail, localization, competences.
	Forum: Discussion application where any end-users freely ask questions, answer, and react.	Discussions on occupational topics taking place at the level of the NoP, and not just of the CoP. Complementary to direct conversations.
	FAQ: List of answers to usual questions. Answers made by experts on covered topics.	Members of local groupings have access to competent and validated knowledge from experts from the whole network.
	E-mail system: Link from the intranet to the e-mail system.	Members of local CoPs may electronically get in touch with remote colleagues. Related to the repertory feature.
Joint Enterprise	**Homepage:** First page of the site that provides information on the site and links to applications.	Presentation of the overall purpose of the work of local CoPs and NoPs. General information on the CoPs and the NoP.
	Occupational applications: Migration of these applications to the intranet.	Members of local CoPs accomplish parts of their job by using the intranet. Nourishes sense of commonality of work throughout the NoP.

The official management may be unaware of emergent and improvisational dynamics of change that contribute to the richness of real-world practices. The lack of official support, however, weakens the legitimacy and visibility of local CoPs or of the overall NoP. Coaches at various levels therefore appear as useful intermediaries between informal workgroups and the management. They make it possible to connect CoPs, NoPs, and the official management of the firm. Coaches may also help deal with the two preceding dilemmas. In particular, they may decide when and how to introduce more control or more freedom in the management of the intranet. They may also know how to favor the sense of community and the emulation among members of the network thanks to initiatives on the intranet.

CONCLUSION

Intranet features and use constitute useful links between CoPs and NoPs. They make their users aware that, beyond geographical distance, they belong to an overall occupational network. Intranet use also makes it possible to store and exchange information throughout the network and to improve professional practices.

One would expect these occupational networks to expand even more as technologies such as extranets become more available. The extension of networks of practice from the intra- to the extra-organizational level will undoubtedly present new challenges and opportunities to knowledge management.

REFERENCES

Bansler, J.P., Damsgaard, J., Scheepers, R., Havn, E. & Thommesen, J. (2000). Corporate intranet implementation: Managing emergent technologies and organizational practices. *Journal of the Association for Information Systems, 1*(10), 1-10.

Brown, J.S. (1998). Internet technology in support of the concept of "communities of practice." *Accounting, Management and Information Technologies, 8*(4), 227-236.

Brown, J.S. & Duguid, P. (1991). Organizational learning and communities-of-practice: Toward a unified view of working, learning and innovation. *Organization Science,* 2(1), 40-57.

Brown, J.S. & Duguid, P. (2000). *The social life of information.* Boston: Harvard Business School Press.

Brown, J.S. & Duguid, P. (2001). Knowledge and organization: A social-practice perspective. *Organization Science, 12*(2), 198-213.

Curry, A. & Stancich, L. (2000). The intranet—an intrinsic component of strategic information management? *International Journal of Information Management, 20*(4), 249-268.

Gherardi, S. & Nicolini, D. (2000). The organizational learning of safety in communities of practice. *Journal of Management Inquiry, 9*(1), 7-18.

Iverson, J.O. & McPhee, R.D. (2002). Knowledge management in communities of practice: Being true to the communicative character of knowledge. *Management Communication Quarterly, 16*(2), 259-266.

Lave, J. & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* Cambridge, UK: Cambridge University Press.

Newell, S., Scarbrough, H. & Swan, J. (2001). From global knowledge management to internal electronic fences: Contradictory outcomes of intranet development. *British Journal of Management, 12*(2), 97-111.

Orr, J. (1990). *Talking about machines: An ethnography of a modern job.* Ithaca, NY: Cornell University Press.

Pan, S.L. & Leidner, D.E. (2003). Bridging communities of practice with information technology in pursuit of global knowledge sharing. *Journal of Strategic Information Systems, 12*(1), 71-88.

Ryan, B. (1998). *The corporate intranet: Harness the power of the next-generation intranet.* New York: John Wiley & Sons.

Schultze, U. & Leidner, D.E. (2002). Studying knowledge management in information systems research: Discourses and theoretical assumptions. *MIS Quarterly, 26*(3), 213-242.

Vaast, E. (2004). The use of intranet: The missing link between communities of practice and networks of practice? In P. Hildreth & C. Kimble (Eds.), *Knowledge networks: Innovation through communities of practice* (pp. 216-228). Hershey, PA: Idea Group Publishing.

Von Krogh, G. (2002). The communal resource and information systems. *Journal of Strategic Information Systems, 11*(2), 85-107.

Wenger, E. (1998). *Communities of practice: Learning, meaning and identity*. Cambridge: Cambridge University Press.

KEY TERMS

Browser: A client software program used for searching and viewing various kinds of resources such as information on a Web site or on an intranet.

Computer Literacy: The ability confidently and competently to make good or optimum use of the facilities that computers provide.

Extranet: A private network that uses Internet protocols and the public tele-communications system to share a business's information, data, or operations with external suppliers, vendors, or customers.

Firewall: The set of related programs, located at a network gateway server, that protects the resources of a private network from users from other networks. Basically, a firewall, working closely with a router program, filters all network packets to determine whether to forward them toward their destination. A firewall is often installed away from the rest of the network so that no incoming request can get directly at private network resources.

Flexibility: The ease with which a system or component can be modified for use in applications or environments other than those for which it was originally designed. A flexible system may be transformed by IT professionals and customized by end-users.

Intranet: A private network inside a company or organization that is based on Web standards (i.e., TCP/IP protocols) and offers various applications for members of a specified group.

Server: The computer on a network that is dedicated to a particular purpose, and stores all information and performs the critical functions for that purpose. For example, a Web server would store all files related to a Web site and perform all work necessary for hosting the Web site.

User-Friendliness: Quality of a program or user-interface that is easy to use and appropriate for end-users. With graphical interfaces, in particular, human-computer interaction is made easier and more intuitive for end-users.

Introducing Java to the IT Master's Curriculum

Wendy Lucas
Bentley College, USA

INTRODUCTION

The object-oriented programming paradigm has gained popularity in both industry and academia, and Java is becoming the language of choice. Yet, it can be a difficult language to learn, with many hurdles for novice programmers. This overview describes how Java was successfully introduced as the first programming language in an information technology master's program at Bentley College. Careful consideration was given to a variety of factors, including when to introduce object-oriented concepts, which integrated development environment to use, and how to support students with minimal prior programming experience. The impact of these choices on the learning experience and the factors that led to the successful implementation of Java as a first programming language are described.

BACKGROUND

The Java programming language was developed at Sun Microsystems in 1991 for use in consumer electronics devices, such as television sets and VCRs. It, therefore, needed to be both small and portable. While the language has grown in size with each new release, the Java Virtual Machine continues to ensure implementation-independent code that can be run under a variety of operating systems. Coupling this capability with the ability to run Java applets from Web pages was what first attracted interest in the language when it was initially released in 1995. Since that time, Java has developed into a general-purpose language used throughout enterprise-wide distributed applications.

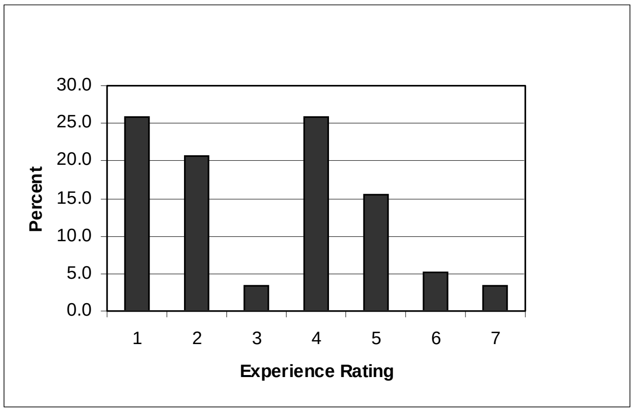
While Java has gained acceptance and widespread use in industry, it has also made inroads into academia. It is now taught extensively in intermediate programming courses and has been widely adopted as the first programming language. Its inclusion in introductory courses, however, has been problematic due to inherent complexities in the language. In comparing Java to C++, for example, it soon becomes clear that many of Java's "simplifications" are not correlated with a simpler learning experience for beginning programmers. Although Java does not support multiple inheritance, a difficult concept for students learning C++, it does allow classes to extend multiple interfaces, which is similar in purpose and complexity. Java does not permit pointer arithmetic and hides pointers from the user, but beginning programmers must understand the concept of references in order to work with objects and arrays. Other problems include: the library documentation is often ambiguous; the encapsulation model is actually more complicated than that of C++; and a large number of methods in the class library throw exceptions that must be caught or passed to a caller. Benander, Benander, and Lin (2003) found that, while professionally employed programmers understand the importance of catching exceptions, many students fail to appreciate Java's exception handling approach. Even capturing keyboard input from the user is difficult, as Java does not provide basic support for such input in non-GUI programs. As a result, authors typically provide their own methods, requiring students to develop a basic understanding of packages, classes, and methods at a very early stage (see, for example, Lewis & Loftus, 2003; Savitch, 2004). One way in which Java is truly simpler than C++ is in providing automatic memory management in the form of a garbage collector.

Given the difficulties new programmers must overcome in learning Java, some educators, such as Collins (2002), have reached the conclusion that it is not reasonable to expose students to this language in their first programming course. Others have sought means for shielding students from some of Java's complexities. Roberts (2001) describes the use of the MiniJava environment, which contains a subset of the standard Java release along with simplifying enhancements that make it easier to use. Another development environment, BlueJ, was specifically designed with teaching in mind (Poplawski, 2001; Sanders, Heeler & Spradling, 2001). It provides an easy-to-use interface with customizable templates for class skeletons, and allows the user to instantiate objects and test methods without having to write a driver program. Kölling and Rosenberg (2001) used this environment during three semesters of an introductory object-oriented programming course. The system supports an "objects first" approach, provides visual representations that help students understand the relationship between classes and objects, and supports student experimentation, thereby promoting frequent and early testing of code. Lewis and Watkins (2001) also found that BlueJ helped them explain and demonstrate object-oriented concepts from the beginning of their introductory programming course within the MSc in Computing program, in which the majority of students were graduates from other disciplines. Barnes (2002) also described an objects-early approach to teaching introductory Java through the use of LEGO® MINDSTORMS™ kits. Providing physical models was found to enhance and support the introductory programming course experience.

The debate between following an object-oriented or procedural paradigm when first teaching Java has persisted from the earliest days of its inclusion in programming courses. Results of a survey of sixty-one students who had taken or were taking a Java programming course showed that while object-based programming was not considered particularly difficult to learn, object-oriented programming concepts were difficult (Madden & Chambers, 2002). A recurring objection to "objects first" is that students must proceed on faith that the concepts and structures, which they make use of early on, will be explained to them and understood later in the course. Based on the premise that students should gradually build upward from primitive data types and control structures in order to free them from the immediate conceptual load associated with abstract data types, Cecchi, Crescenzi, and Innocenti (2003) follow a "structured programming before object-oriented programming" approach in their CS1 course. An automated development tool called JavaMM was used in this course to remove the burden of creating the complex structure of Java programs from the student. A preliminary evaluation indicated that the tool was very useful in improving the success rate of students.

Little research exists on the use of Java as a first programming language in graduate information systems (IS) or information technology (IT) programs. While these programs face a number of the same issues as those

Figure 1. Self-ranking of prior programming experience



mentioned earlier, there are significant differences between IT graduate and undergraduate students. For the former, their first programming course may well be the only one they take. In addition, many of them expect to see environments and applications similar to those they have been exposed to in the workplace. They may also have a broader range of prior programming experience. It was necessary to take these differences into account when designing the course described next.

OBJECT-ORIENTED PROGRAMMING COURSE

The graduate level object-oriented programming course within the master's level IT program was taught for the first time in the fall of 2001 to 58 graduate students in three sections. On a scale from one (*novice/beginner*) to seven (*expert*), the average student prior programming experience level at that time was 3.14 ± 1.78. Figure 1 shows the distribution.

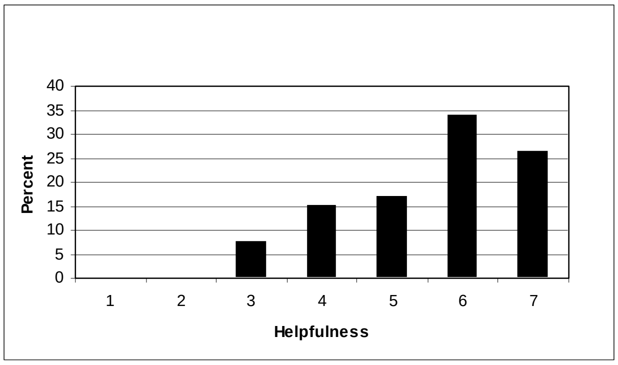
A proactive approach is taken to support students in what, for many, is their first exposure to programming. The following are descriptions of choices that have been made for enhancing the learning experience without adding to student anxiety over learning to program in Java.

Supporting Materials

One critical decision concerned which integrated development environment (IDE) to use. After extensive trials with a number of environments, the Borland® JBuilder IDE was selected due to the strength of its built-in help facilities and its tools for writing, running, and debugging code. While it could be confusing at first, JBuilder is easier to use than most of the professional environments found in the industry (Savitch, 2001). Plus, several students had expressed interest in working with JBuilder because it was being used in their own work environments. JBuilder was installed on the computers in all of the technology classrooms and computer labs, and students were encouraged to download their own copies from the Web.

A review of several textbooks led to the selection of *Computing with Java: Programs, Objects, and Graphics* by Gittleman (2001). This book was chosen for its in-depth coverage of critical course concepts, the clarity of its text, and its frequent and easy to follow code examples. The course Web site provides access to weekly lecture notes, assignments, individualized grades and comments, source code for programming examples from the lecture, and links to other relevant sites, such as Java's class libraries.

Figure 2. Student ranking of JBuilder



Course Organization

The course is taught over a fifteen-week semester, and is organized as weekly, 2-hour and 15-minute lectures with ten programming assignments, a midterm exam, and a final exam. The lecture notes, which are projected during class, allow students to focus on understanding concepts and programming examples, rather than on taking notes. A portion of the lecture is devoted to problem solving, with students writing code both individually and together. Working individually provides an opportunity for the instructor to observe the various approaches being taken and comment on their effectiveness and efficiency. Either the instructor (earlier in the term) or a student (later in the term) will then type a solution into the classroom computer using JBuilder's editor. It is often more educational when a student's code fails to run properly, as the class will then need to determine and correct the errors.

Programming assignments reinforce the concepts covered in the lectures. Assignments are completed on an individual basis, with students encouraged to seek the help of lab assistants for questions involving syntax or the IDE, or the instructor for more complex questions.

The first seven assignments are all one week in length and are primarily logic-intensive. Short, focused assignments force the students to stay up-to-date with the material and to practice what they have been learning. The last three assignments are each two weeks in length, and encourage creativity and investigation of Java's class libraries. Assignments account for 30% of the grade. Midterm and final exams make up 30% and 35%, respectively, and the remaining 5% is for class participation.

Course Content

The content of both lectures and assignments begins at a rudimentary level and increases in complexity throughout the semester. The first few weeks focus on basic programming concepts, including data types, operators, and control structures. Class and object concepts are then introduced using built-in classes and arrays. User-defined classes come next, followed by the object-oriented concepts of inheritance, polymorphism, and encapsulation. Additional topics such as dynamic data structures, exception handling, and file I/O are also covered.

Students are initially guided through Java's class libraries and then, as the weeks progress, are encouraged to explore the libraries to find classes and methods for use in assignments. In the last few weeks, graphics and event handling are taught using applets, which require an understanding of a predefined startup sequence of method calls, system-defined and system-passed arguments, and inheritance that is gained throughout the course.

Survey Results

Mid-way through the fall 2001 semester, the 58 students enrolled in the course completed a course survey. A total of 143 students from six preceding semesters of the programming course taught using Visual Basic (VB) had filled out a similar survey. In the following sections, the results of the Java survey are presented and contrasted to the results from the VB survey, wherever possible.

SUPPORTING MATERIALS

Students were in general pleased with the supporting materials provided for this course. The use of JBuilder was only required for one assignment, yet the majority of students (86%) continued using this IDE throughout the term. Its average rating was 5.57 ± 1.25 on a seven-point

Figure 3. Student ranking of the course Web site

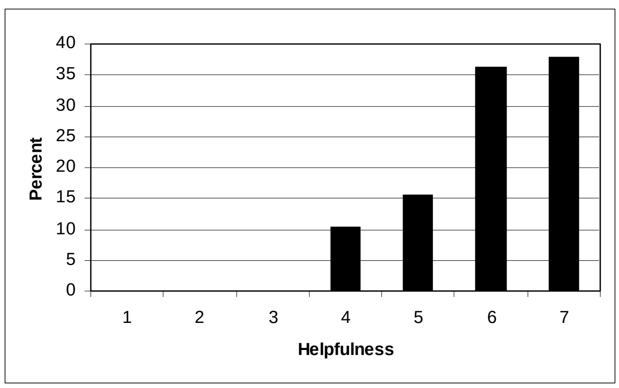
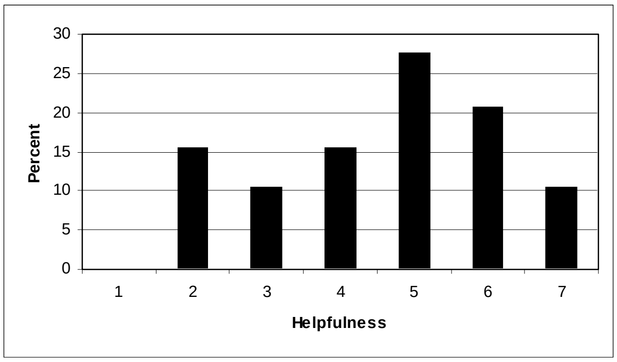


Figure 4. Student ranking of the textbook



scale ranging from one (*no value*) to seven (*high value*). Figure 2 shows the distribution.

The course Web site was highly valued, with an overall helpfulness rating of 6.02 ± 0.98 on the same seven-point scale (see Figure 3). The average rating of the text was lower (4.59 ± 1.57) and showed more variation in responses (see Figure 4). This was still considered a positive finding, however, because students are typically unenthusiastic about technically-oriented textbooks and generally rate them poorly.

Course Content

The average difficulty of the course was rated on a seven-point scale as 4.79 ± 1.14, where *one* represents *not challenging*, and *seven* represents *very demanding*. Figure 5 contrasts the difficulty ratings to those for the prior programming course taught in VB. The average difficulty rating for the VB course was 4.69 ± 1.15. The mean difference between the two groups is not significant.

The mean difference in perceived difficulty for students in the Java course with prior programming experience versus those without is significant at the 0.05 level. Figure 6 shows this relationship, where *little* refers to those who rated their experience as either a one or a two, *somewhat* refers to those in the three-to-five range, and *very* is for a rating of six or seven.

It should be noted that the survey was distributed before many of the more advanced object-oriented concepts had been introduced. As the course proceeded, even those students with prior procedural programming experience had difficulty with some of the object-oriented concepts, based on their performance on assignments and the final exam. This agrees with the findings of Mehic and Hasan (2001) that prior procedural language experience can actually create an obstacle for students learning object-oriented design.

Figure 5. Comparison of VB course difficulty to Java course difficulty

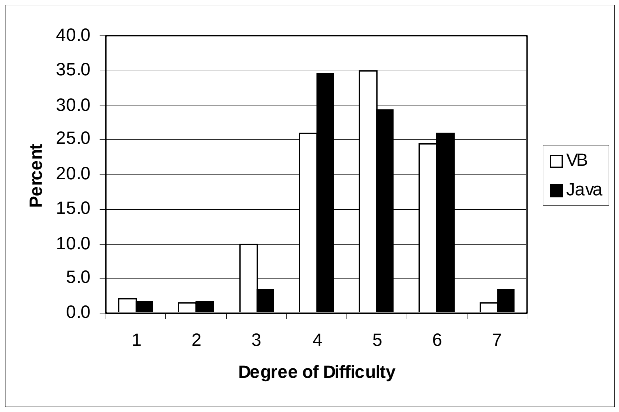


Figure 6. Java course difficulty as a function of prior experience

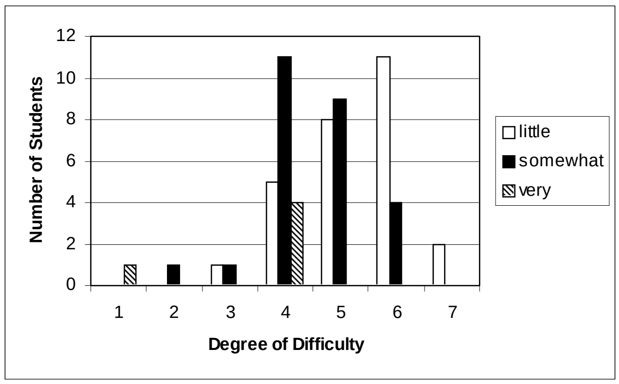
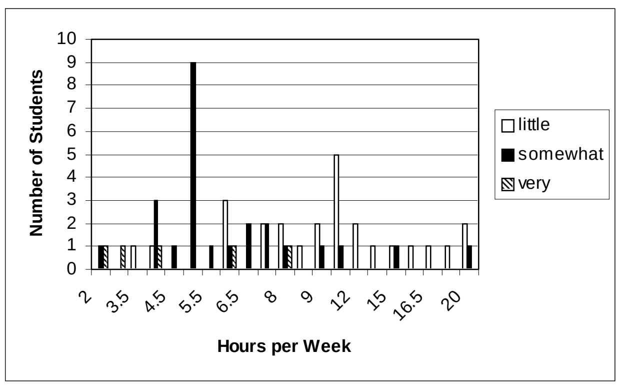


Figure 7. Hours worked in the Java course as a function of prior experience



The average number of hours worked per week was 8.24 ± 4.62, while the average for the VB course had been 6.59 ± 4.35. The mean difference between the two is significant at the 0.05 level. There was also a significant difference between means for those students with prior experience versus those without, as shown in Figure 7.

FUTURE TRENDS

Given the continued popularity of Java in both industry and academia (with the latter evidenced by the decision of the College Board to have the Advanced Placement® Computer Science Examination be administered in Java, beginning with the 2003-2004 academic year), the trend toward teaching Java as the first programming language gives no indication of abating. Nonetheless, the inherent complexities associated with Java will continue to make it a difficult first language to learn. There is, therefore, a continuing need for research into factors that can positively affect the likelihood of success for students in introductory programming courses built upon the Java platform.

CONCLUSIONS

There are several factors that have led to the successful implementation of this Java course. Students particularly enjoy the frequent programming assignments, which reinforce their understanding of course material. Earlier assignments boost confidence in their ability to write functioning code. Those successes help students maintain a positive attitude when the level of complexity increases. The incremental dependence upon Java's class libraries also assists students in learning to navigate the extensive documentation without becoming overwhelmed.

The support structure for the students definitely has a positive impact. JBuilder's debugging environment is very helpful, and the IDE as a whole has been well received. Finally, access to lecture notes, code, and grades from the course Web site is highly appreciated and contributes to the positive view of the course held by most of the students.

In summation, the benefits of frequent programming assignments to reinforce learning, a carefully structured, well thought-out curriculum, and a strong support structure have helped ease the transition to the Java programming language and continue to make this course a successful and vital component of the MSIT curriculum.

REFERENCES

Barnes, D. (2002, February 27-March 3). *Teaching introductory Java through LEGO MINDSTORMS models.* Paper presented at the 33rd SIGCSE Technical Symposium on Computer Science Education, Cincinnati, Kentucky.

Benander, A.C., Benander, B.A., & Lin, M. (2003, Summer). Perceptions of Java – Experienced programmers' perspective. *Journal of Computer Information Systems.*

Cecchi, L., Crescenzi, P., & Innocenti, G. (2003, June). *C : C++ = JavaMM : Java.* Paper presented at the PPPJ2003 International Conference on the Principles and Practice of Programming in Java, Kilkenny, Ireland.

Collins, D. (2002, January). *The suitability of Java as a first programming language.* Paper presented at the Sixth Java & the Internet in the Computing Curriculum Conference, London, England.

Gittleman, A. (2001). *Computing with Java: Programs, objects, graphics* (2nd ed.). El Granada, CA: Scott/Jones, Inc.

Kölling, M., & Rosenberg, J. (2001). *Guidelines for teaching object orientation with Java.* Paper presented at the 6th Annual Conference on Innovation and Technology in Computer Science Education, Canterbury, United Kingdom.

Lewis, J., & Loftus, W. (2003). *Java software solutions: Foundations of program design* (3rd ed.). Addison-Wesley.

Lewis, S.F., & Watkins, M. (2001, January). *Using Java tools to teach Java, the integration of BlueJj and CourseMaster for delivery over the Internet.* Paper presented at the 5th Java in the Computing Curriculum Conference, London, England.

Madden, M., & Chambers, D. (2002). *Evaluation of student attitudes to learning the Java language.* Paper presented at the Principles and Practice of Programming in Java, County Kildare, Ireland.

Mehic, N., & Hasan, Y. (2001). *Challenges in teaching Java technology.* Paper presented at the Informing Science, Krakow, Poland.

Poplawski, D. (2001). *Objects have class: An introduction to programming in Java*. McGraw-Hill.

Roberts, E. (2001). *An overview of MiniJava.* Paper presented at the 32nd SIGCSE Technical Symposium on Computer Science Education, Charlotte, NC.

Sanders, D., Heeler, P., & Spradling, C. (2001). Introduction to BlueJ. *The Journal of Computing in Small Colleges, 16*(3).

Savitch, W. (2001). Using JBuilder. *Java: An introduction to computer science and programming* (2nd ed., p. 1027). Prentice Hall.

Savitch, W. (2004). *Java: An introduction to computer science and programming* (3rd ed.). Prentice Hall.

KEY TERMS

Abstract Windows Toolkit (AWT): Library of classes for writing window interfaces.

Byte-code: This machine-independent code is translated into the machine language of the computer on which it is running.

Encapsulation: Data and actions are packaged together in a class, with the details of the implementation hidden from other classes.

Inheritance: A way of organizing classes so that properties can be defined once and applied to a whole collection of classes, with a general class defining those properties and specialized classes inheriting them.

Instantiation: Creation of an object, or class instance, from a class.

Java Virtual Machine (JVM): A platform-independent execution environment that converts Java byte-code into machine language and then executes it.

Methods: Actions performed by objects.

Polymorphism: One method name can cause different actions to occur, depending on the kind of object performing the action.

IS Implementation in the UK Health Sector

Stuart J. Barnes
Victoria University of Wellington, New Zealand

INTRODUCTION

Since the mid-1980s, the UK public sector has been the subject of wide-ranging reforms involving the introduction of IS and IT. Change has been sought in the ways that services are managed and delivered, the evaluation of the quality of aforesaid services, and in accountability and costing. One of the most predominant of such changes has been the introduction of competition for services, the motivation of which has been to invite efficiency, effectiveness, and related benefits ensuing from the accrual of economies.

Pivotal to such change has been an explosion in the introduction of a variety of information systems to meet such challenges. Focusing on health care, a large part of the work of the health service involves collecting and handling information, from lists of people in the population to medical records (including images such as X-ray pictures), to prescriptions, letters, staffing rosters and huge numbers of administrative forms. Yet until recently, the health service has been woefully backward in its use of the technology to handle information by the standards of private industry.

This has been quickly changing in recent years and by 2003 the National Health Service (NHS) spent £2.8 billion annually on capital in hospitals (Department of Health, 2003a), around 10% of which was for IT. In the last 20 years, IT has added 2% to overall health expenditure (Wanless, 2001). This investment is still small by the standards of the private sector, but is all the more significant when we consider that health care is an industry which has been slow to adopt IT and one which presents some of the biggest IT opportunities (Department of Health, 2002).

BACKGROUND: INFORMATION SYSTEMS IN THE NHS

The implementation of IT in the UK health sector has been fraught with difficulties. In fact, estimates suggest that problems with the first wave of projects in the public sector, from the mid-1980s to mid-1990s, cost over £5 billion (Collins, 1994). There are a number of high-profile examples of IS failure in the NHS, including that of Wessex Regional Health Authority's Regional Information Systems Plan (£63 million), the London Ambulance Service's Computer Aided Dispatch system (£1.1 to £1.5 million), and more recently, various Resource Management Initiative (RMI) Case Mix failures (£1 to £3 million) (Barnes & Targett, 1999).

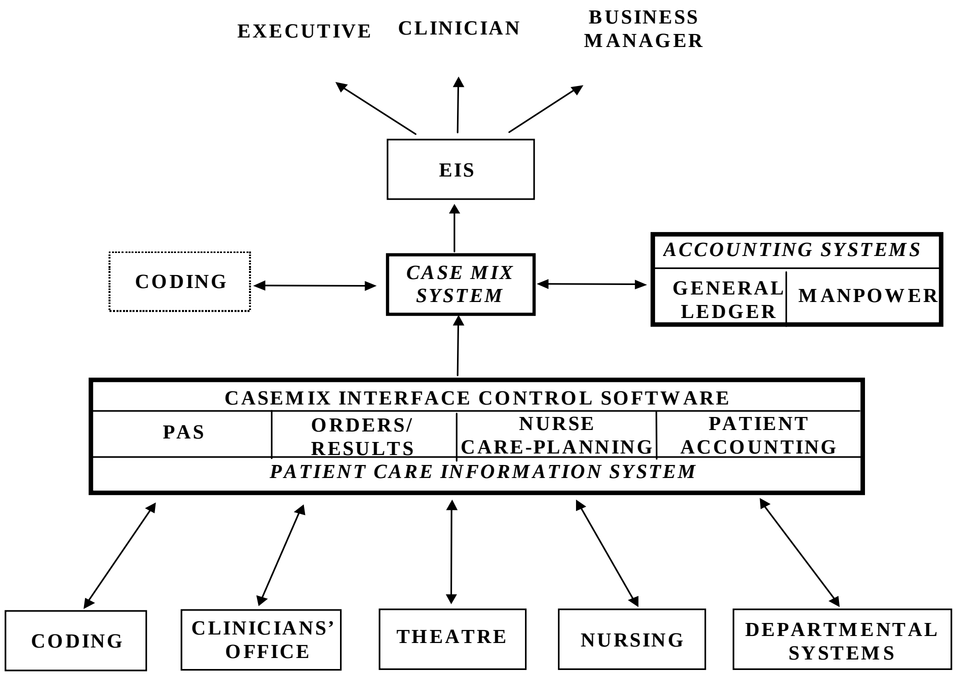
In this section we will focus on RMI and Case Mix as an illustration, since this was one of the first initiatives for wide reaching IT-induced organizational transformation and information integration in the health sector. The systems implemented have been an important support for financial developments in the management of UK hospitals. During 2003, the importance of Case Mix systems has again come to the fore as the NHS has attempted to implement prospective payment systems (PPS) (Department of Health, 2003b).

RMI was a driving force in the move towards information systems and cultural change in the NHS. First announced in 1986, RMI was going to help clinicians and other hospital managers to make better-informed judgments surrounding how the resources they control can be used most effectively (DHSS, 1986). The Initiative was not only aimed at persuading clinicians to own the management process, but to provide them with accurate, up-to-date and relevant information which could be used to cost medical activities and improve patient care. The response to this need for improved information services available to hospital units was the development and implementation of a sophisticated and extensive package of IT referred to as the "Case Mix" information system (CMIS), an idea borrowed from the U.S., with the purpose of clinical and management audit.

Prior to RMI, the introduction of IT in the NHS was patchy and limited. Where systems existed, the technology was very varied, incompatible, archaic, and dependent upon regional computer departments to deliver necessary operational systems. The development of CMIS, with its dependence on data fed from other systems such as the Patient Administration System (PAS), radiology, pathology, theatre and nursing systems, provided a catalyst for the adoption of operational systems throughout the hospital (Barnes, 2001).

CMIS takes a central position in the hospitals' IT infrastructure, as shown in Figure 1 (an illustrative example - specifications may vary depending upon the hospital), providing a tool for collecting and analyzing data from all areas of hospital operations using an execu-

Figure 1. The role of case mix in NHS hospitals



tive information system (EIS). As we can see from Figure 1, there are two main types of data feed: financial and medical. The financial feed consists of pulling data from the general ledger and humanpower systems, particularly standard costs and budgets. This contrasts with the other main feed to CMIS, that of the "patient care information system" - a label given to the array of feeder sub-systems providing information on all aspects of patient treatment and care.

Each of the feeder systems is interfaced with CMIS, so as to provide appropriate data in an acceptable format. Such data are accumulated by CMIS within the care profile sub-system: this stores the actual tests, treatments, costs, number of cases and so on to be compared with expected "ideal" profiles or projected activity levels as drawn from the financial data, enabling financial audit. Regarding clinical audit, CMIS provides the tool for assessment of the professional clinical practices of each clinician.

EXPERIENCES IN INFORMATION SYSTEMS IMPLEMENTATION

In addition to traditional problems of IS implementation, RMI also reveals a number of interesting and contrasting influences on the early implementation of strategic IS in UK hospitals:

- *Central influence* - Regional authorities (RAs) had a substantial influence on the development of systems in a number of areas, for example investment justification, what to procure, objectives, financing, and project management milestones. Importantly, the original "standard" systems were not sensitive to the needs of individual hospitals, affecting stakeholder support and the need for project redefinition. Where relationships with RAs were tenuous, RMI was looked upon with suspicion. Initiatives imposed subsequently changed the shape and direction of the project, and created other priorities within hospitals, while recommendations about clinical coding were never clear. Central influence is, interestingly, both a reason for the existence of the project, and for many of its problems: paradoxically, it is both an enabler and an inhibitor.
- *Project purpose* - The problems in communicating the purpose of CMIS, and in approaching locally sensitive designs, affected the attitudes of stakeholders: the project was very much an imposed directive. Many individuals were unclear about the rationale for CMIS, and this was compounded by the traditional absence of IT within hospitals. While RM was aimed at improving resource allocation, ironically, many saw IT spending as a waste of money as opposed to direct patient care.
- *Clinicians, management and CMIS* - Hospitals are distributed organizations, with a variety of frag-

mented groups and cultures, each with their own defined roles. RMI and CMIS cut a swathe through many of these, particularly clinicians and management, and sat uncomfortably in hospitals. Where the system impinged upon the territory of individuals or groups, friction was encountered. Hospitals, for many years without IT, and with a defined structure and set of roles, found change difficult. While politics occurred in a variety of areas, the clinician-management divide is perhaps the most interesting. Each of these groups found it difficult to adjust to a new culture and structure, and for apparently conflicting reasons: top management was often opposed to devolvement, while clinicians were wary of CMIS as a "tool of management" and a control device. Thus, paradoxically, CMIS was seen both as a tool of centralization and of devolution.

CONCLUSIONS AND FUTURE TRENDS

The complexities surrounding RMI compounded early experiences in strategic IS implementation in the NHS. Evidence suggests that success has been difficult to achieve, and that hospitals found it difficult to achieve any tangible benefits (e.g., see Barnes, 2001; Barnes & Targett, 1999; Lock, 1996). Evaluating investments is made all the more difficult by the fact that published evidence is scarce. Lock (1996) found that only 5% (by value) of all investments were the subject of published assessment, which may be an indicator of the poor value of such investments. Even published studies are far from conclusive: the "official" resource management evaluation conducted by Brunel University found "no measurable patient benefits" in its review of the six pilot sites, while costs were more than double those expected (Heath Economics Research Group, 1991).

One of the key impacts of case mix has been on the financial management of hospitals. The sustainable financial management of hospitals has recently come into question (Audit Commission, 2003). During 2003, the UK Department of Health has outlined the Government's desire to implement PPS, as outlined in its *Payment by Results* White Paper (Department of Health, 2003b). This would allow the trust financial regime (TFR) of hospitals to carry deficits and surpluses, spurring greater efficiency. This has led to some concerns that it may lead to under-investment and reduction in quality (Medicare Payment Advisory Commission, 2003; Rosenberg & Browne, 2001; Topping, 1997). On balance, the evidence from the U.S. and elsewhere seems to suggest that these systems are not detrimental to the quality of health care (Dismuke & Guimaraes, 2002; Shen, 2003).

REFERENCES

Audit Commission. (2003). *Achieving the NHS plan.* HMSO, London.

Barnes, S.J., & Targett, D. (1999). A framework for strategic IS implementation in the UK health sector. *Topics in Health Information Management, 19*(4), 62-74.

Barnes, S.J. (2001). Experiences in SIS implementation in UK healthcare. In T. Spil & R. Stegwee (Eds.), *Strategies of healthcare information systems* (pp. 11-30). Hershey, PA: Idea Group Publishing.

Collins, L.W. (1994). CIOs must look beyond the IS horizon. *Computers in Healthcare, 12*(5), 39-40.

Department of Health. (2002). *Delivering 21st century IT support for the NHS.* HMSO, London.

Department of Health. (2003a). *Implementing payment by results and investment in technology and drugs.* Technical Paper, DoH, London.

Department of Health. (2003b). *Payment by results.* HMSO, London.

DHSS (Department of Health and Social Security). (1986). *Health services management: Resource management (management budgeting) in health authorities.* HMSO, London.

Dismuke, C.E., & Guimaraes, P. (2002). Has the caveat of case-mix based payment influenced the quality of inpatient hospital care in Portugal? *Applied Economics, 34,* 1301-1307.

Health Economics Research Group. (1991). *Final report of the Brunel University evaluation of resource management.* Brunel University, Uxbridge.

Lock, C. (1996). What value do computers provide to NHS hospitals? *British Medical Journal, 312,* 1407-1410.

Medicare Payment Advisory Commission. (2003). *Report to Congress: Medicare payment policy*. US Congress, Washington DC.

Rosenberg, M.A., & Browne, M.J. (2001). The impact of the inpatient prospective payment system and diagnostic-related groups: A survey of the literature. *North American Actuarial Journal, 5*(4), 84-94.

Shen, Y. (2003). The effect of financial pressure on the quality of care in hospitals. *Journal of Health Economics, 22,* 243-269.

Topping, S. (1997). The impact of health care reform on capital acquisition for hospitals. *The Financial Review, 32*(3), 751-778.

KEY TERMS

Case Mix: On admission to a hospital, the individual patient is provided with a variety of services, each of which expends resources. Each encounter or episode has a mix of, for example, tests, treatments, accommodation, and so on. Case mix refers the combination of services and resources used for each patient episode.

Case Mix Information System (CMIS): An information system, fed with data from an array of hospital subsystems, for the principal purpose of clinical and financial audit of patient cases. Further developments to CMIS have seen other functionality added, such as contract planning, quality assurance, waiting lists, and clinical support.

Clinical Audit: The assessment of professional clinical practices. General information routinely provided includes lengths of stay, deaths and re-admissions. More specific information such as drugs administered and operative procedures performed are available via the appropriate feeder systems. The data allow aspects of case mix, clinical management, diagnostic accuracy and patient outcomes to be compared.

Clinical Coding: Categorises diagnoses and procedures for patient episodes according to a detailed clinical standard index. The most common standard is the International Classification of Diseases (ICD), particularly for financial audit, but a UK standard, Read, is very popular with medical staff. These are grouped to provide aggregate information for audit.

Executive Information System (EIS): An IS that provides high-level information tailored to the needs of decision-makers. Typically a graphics-oriented system, in the context of CMIS the EIS aggregates and uses summary data to present information to clinicians, executives and business managers in the form of graphs, histograms, pie charts, tables and so on.

Financial Audit: The budgetary assessment of the cost of hospital care. CMIS collects data on the actual tests, treatments, costs, number of cases and so on. These are then compared with expected "ideal" profiles or projected activity levels as drawn from the financial data, typically based on best practices or historical data.

Resource Management Initiative (RMI): A UK initiative providing clinicians and other hospital managers with information required to use resources to maximum effect, generally by the introduction of new IT. This was aimed at encouraging clinicians to take more interest and involvement in the management of hospital and community units, making them responsible for operational and strategic decisions.

IS Project Management Contemporary Research Challenges

Maggie McPherson
University of Sheffield, UK

INTRODUCTION

Although project management is often said to have its roots in other traditional fields, such as construction, Morris (2002) asserts that modern project management practices have their origins in the 1950s US aerospace agencies. Much has been written about Information System (IS) / Information Technology (IT) project initiatives in both the public and private sectors. In fact, many information systems frequently fall short of their requirements, and are, more often than not, costlier and arrive later than anticipated, if indeed they are completed at all. For instance, according to a report for the Organization for Economic Co-operation and Development (2001), failures of major IT investments and key systems development projects have raised concerns for the achievement of service improvement through information technology. Additionally, it has been argued that failures in IT projects are more common than failures in any other aspect of modern business (Nulden, 1996). The widely-cited Standish Group (1994) study, carried out in the US, classified IT projects as follows:

- **Resolution Type 1 (Project Success)**: The project is completed on-time and on-budget, with all features and functions as initially specified.
- **Resolution Type 2 (Project Challenged)**: The project is completed and operational but over-budget, over the time estimate, and offers fewer features and functions than originally specified.
- **Resolution Type 3 (Project Impaired)**: The project is cancelled at some point during the development cycle.

The report estimated the success rate was only 16.2%, while challenged projects accounted for 52.7%, and impaired projects (cancelled) amounted to 31.1%. Since large complex projects in any area are difficult to organize, it could be said that the level of abstraction required often leads to a lack of understanding between all stakeholders involved with the project. Callahan and Moretton (2001) describe software design as being "in the code". They assert that since it is not visible, it makes it hard to use software design as a focal point for development project coordination and integration, unlike many physical designs which can be made visible to all project participants. As a result of this "invisibility", managing the development of an IS project is arguably more problematic than project management within the manufacturing sector because software development is often a highly conceptual and complex process.

Indeed, a lack of adequate project management knowledge could be said to be a major contributing factor to unsuccessful IS projects. For instance, as project managers should be aware, unless specific objectives and clear-cut end points have been set, it can be difficult to know if a milestone has been reached and indeed if the required end-product has been produced. However, making use of proprietary tools such as Microsoft™ Project is sometimes mistakenly thought of as project management, whereas real project management expertise goes beyond the mere production of Gantt or Pert (Program Evaluation Review Technique) charts, which simply represent project activities in the form of bar charts or flow diagrams. As Mandl-Striegnitz et al. (1998) point out, important project management techniques include estimation of costs and explicit identification of risks. Clearly, there is a need for more in-depth research to gain a better understanding relating to the complex role of project management within the whole IS design and development process. This discussion considers how these problems affect contemporary IS project management research and explores the methodological approaches open to researchers carrying out investigations in this area.

BACKGROUND

In order to better understand the challenges facing researchers of Information Systems Project Management (ISPM), it is necessary to explore what is meant by some of these terms. As stated by the American National Standard for Telecommunications (2000), an IS is "an organized assembly of resources and procedures united and regulated by interaction or interdependence to accomplish a set of specific functions, whether automated or manual, that comprises people, machines, and/or methods organized to collect, process, transmit, and disseminate data that represent user information". In its simplest terms, an IS can be described as a human activity or social system, which may or may not involve the use of computer systems; although, these days the former is more likely. According to Stoner et al. (1994), management can be regarded as a process of planning, organizing, leading and controlling the efforts of staff and other resources in order to achieve organizational goals, and the Association for Project Management (2000) describes a project as a distinct set of coordinated activities "… with definite starting and finishing points, undertaken by an individual or organization to meet specific objectives within defined time, cost and performance parameters". By combining these terms, a definition for ISPM could be said to be the process of managing the creation of an IS through the establishment of project goals; organizing, leading, co-coordinating the efforts of staff processes and tasks; and controlling other resources to achieve a set of agreed objectives.

Since IS projects are frequently comprised of multi-disciplinary teams of people, a definition of what is meant by a team in this particular context is called for. Geddes et al. (1993), regard a team as comprising those individuals who have a significant contribution to make to the successful achievement of the project, whether this is through technical or specialist expertise; sponsorship, political support or sponsorship; or expectation of, and interest in, outcomes. Programmers and associated staff are often selected according to their ability to demonstrate the appropriate technical knowledge, which does not guarantee proficiency in managing successful projects. Despite the emphasis on team leadership ability, senior developers/project managers are often promoted from the programming team, with a continued emphasis on technical expertise (Mandl-Striegnitz et al., 1998).

In reality, IS project managers must not only be able to plan and break activities down into components that can be understood and to control tasks and monitor risks, but must additionally be able to consider people and process issues requiring significant team-building skills. Although IS may be implemented by staff with technical competence, they may well lack the necessary abilities to evaluate organizational contexts and analyze corresponding behaviors.

Nevertheless, since 1994 there has been an improvement in project management outcomes. By 2001, the Standish Group published another report stating that project time and cost overruns had reduced significantly. Although this improvement in project results was confirmed by a UK-based survey (Saur & Cuthbertson, 2004), the authors acknowledged that their sample could have been unrepresentatively experienced, signifying a continued need for further research.

CRITERIA FOR ISPM SUCCESS

Referring to the Standish Group report "Extreme Chaos" (2001), it seems that lessons can be learned from the successes and failures of past projects which warrant further study. From extensive research, the Standish Group identified ten criteria for project success:

1. Executive support
2. User involvement
3. Experienced project managers
4. Clear business objectives
5. Minimize scope
6. Standard software infrastructure
7. Firm basic requirements
8. Formal methodology
9. Reliable estimates
10. Other criteria such as small milestones, proper planning, competent staff and ownership

In the UK based study, Sauer and Cuthbertson (2004) reported a higher project success rate than the US Standish Report (1994). Nevertheless, Sauer and Cuthbertson suggested that in order to continue this general improvement, the following recommendations ought to be adhered to:

- Project managers should:
  - Structure projects into smaller units
  - Select the right team and involving them in decision making
  - Invest time and effort in self-development
- Senior IT managers should:
  - Establish a project management focus in the organization
  - Identify the right person for project management role
  - Create appropriate career paths
  - Be accountable through more effective performance management
- Senior business managers/sponsors should:
  - Develop client understanding of project management
  - Engage more actively with projects for which they have responsibility

Some reasons for the improvements described above were costs being cut, better tools being created to monitor and control processes and, not least, project managers becoming better skilled with better management processes being used, giving rise to optimism for the future of project management. Despite the change for the better as highlighted above, the Standish Group (2001) considered "Nirvana" still to be a long way off, indicating a need for continued research. In order to select more appropriate research approaches to investigate ISPM, it is necessary to explore some of the issues specifically related to this particular field.

INFORMATION SYSTEMS PROJECT MANAGEMENT RESEARCH ISSUES

Prior to the 1950s, computing was primarily associated with scientific applications. Even after computer installation began in business environments for data-processing tasks, associated research had a tendency to be dominated by scientific approaches. Thus, despite the increasing importance of IS within modern businesses, for historical reasons the close association of IS development with IT induced many researchers to consider IS problems using methods from the natural sciences (Baskerville & Wood-Harper, 1996; Garcia & Quek, 1997) that are more suited to science laboratories. As a consequence of this scientific tradition, investigations within the field of IS appear to have had a predominance of both positivist and technical points of view, despite major criticisms (op.cit.) that may these research methods might be not always be wholly appropriate.

Keen (1984) stressed that IS is not simply the installation of a technical system in an organization. He suggested that successful implementation needs to consider institutional issues affecting its use in the ongoing context of jobs, formal and informal structures, as well as personal and group processes. Hughes and Wood-Harper (2000) concurred, rejecting the notion that the process is exclusively technical and rational. In their view, IS research should not merely pay attention to the technology and called for IS development to be understood in its situated context, i.e. to consider the domain, the organizational constraints, the social actors and the politics in situ.

ISPM RESEARCH METHODS

The debate as to whether to adopt quantitative or qualitative research methods is well-documented and will only be touched on lightly here. However, as described by Wilson (2002), one of the most contentious debates is whether to adopt the positivist view of the nature of social reality, in which social facts can be known with certainty and in which laws of cause and effect can be discovered, or whether to apply humanistic approaches which generally see social reality as constructed through social action on the part of people who undertake those acts because they have meaning for them. Conventional "scientific" research can run the risk of being reductionist (Lincoln & Guba, 1985), since complex problems are condensed in order to produce models that can provide a simplified simulation of reality. Bryman (1988) suggests that the basic choice of methodological approach is largely influenced by the type of research question being asked and according to Yin (1994) a researcher's choice of methodological approach depends on the problem at hand and the control that the researcher has over the behavioral events. Given that many researchers now believe IS to be socially constructed in particular contexts, it is thought to be important to extend research methodologies for ISPM problems beyond the positivist paradigm in order to uncover rich qualitative data.

FUTURE TRENDS

Mumford et al. (1985) argued for a methodological pluralism within IS research domains, asserting that scientific proof was being regarded as the only valid method despite the fact that many IS problems were not susceptible to these systematic methods. A decade later, Allen (1995) agreed that research methods from different paradigms can be used simultaneously or consecutively and are equally valid. Despite the historical emphasis on positivist methods in IS research, there is increasing support for developing this type of methodological mix and this is equally applicable to project management research. This is demonstrated by the fact that although the data for the Standish Group (1994) project management research was primarily collected through a survey, focus groups were conducted to augment the survey results.

In fact, Myers (1997) argues that all studies are based on some underlying assumptions about what constitutes "valid" research and it is this which should dictate which research methods are appropriate. Avegerou (2000), writing about alternative reasoning for IS, notes that the development literature argues that developing societies need to recognize the limitations of the validity of techno-economic rationality and that they ought to pursue rationalities stemming from their own value systems. IS professional roles have been based and legitimated mostly on technocratic logic, without an obligation to consider the validity of the requirements, which are normally based on the social context (Avegerou, 2000). Morris (2002) concurs, stating that project management is not a science in the full or proper sense of the word. This is contrary to the view of Khazanchi and Munkvold (2000) who believe that methodological and philosophical diversity do not preclude researchers from making scientific inquiries into the fundamental nature of IS phenomena.

Nevertheless, Morris believes that it is a discipline worthy of theoretical study, with various questions susceptible to the methods of scientific enquiry, whilst other areas will always have a large element of unpredictability. He therefore considers that some knowledge of this field will always be personal and experiential. Following this line of reasoning, prominent researchers such as Galliers (1992) have proposed that an interpretative stance for IS research would be wholly applicable. Correspondingly, other researchers have explored and recommended various interpretivist approaches as being particularly suitable. These methodologies include case study research (Yin, 1994), where the researcher interprets data without direct involvement, and action research (Baskerville, 1999), where the researcher is actively involved.

However, Garcia and Quek (1997) stress the need for further critical awareness, stating that importing methods into the IS field is not a simple task. They warn that without critical awareness, there is a danger of methods becoming stereotyped or distorted. Nonetheless, they supported the use of multiple methods to correspond with the complexity of research investigation which will allow a better understanding of the different aspects involved in the constitution of the object under investigation.

CONCLUSION

If one accepts that IS are socially constructed, as many researchers now appear to do, then it follows that evaluative research of IS projects needs to be situated in contextualized and authentic settings. It would therefore seem highly appropriate to set aside the positivist versus interpretivist debate, since Marcella and Knox (2004) suggest that "…it is only based upon a much fuller and more precise understanding of the complex and multifaceted needs of all users, internal and external, in all functional areas of the institution, that systems will be developed which are truly responsive and which function to meet overall … objectives". With this in mind, it seems reasonable to suggest that it is worthwhile combining diverse research methods, as endorsed by Fitzgerald and Howcroft (1998), with a view to maximizing their complementary strengths. This would seem to be particularly appropriate to address many of the concerns and issues relating to ISPM as highlighted in this discussion.

REFERENCES

Allen, D. (1995). Information systems strategy formation in higher education institutions. *Information Research, 1*(1). Retreived March 14, 2004, from *http://InformationR.net/ir/1-1/paper3.html*

American National Standard for Telecommunications. (2000). *Telecom glossary*. Retreived March 14, 2004, from *http://www.its.bldrdoc.gov/projects/t1glossary2000/*

Association for Project Management. (2000). *Glossary of project management terms*. Retreived March 14, 2004, from *http://www.apm.org.uk/resources/p.htm*

Avegerou, C. (2000). Recognizing alternative rationalities in the deployment of information systems. *The Electronic Journal on Information Systems in Developing Countries*. Retrieved March 10, 2004, from *http://www.ejisdc.org*

Baskerville, R.L. (1999). Investigating information systems with action research., *Communications of the Association for Information Systems*, *2*(19). Retreived March 19, 2004, from *http://www.cis.gsu.edu/~rbaskerv/CAIS_2_19/CAIS_2_19.html*

Baskerville, R.L. & Wood-Harper, A.T. (1996). A critical perspective on action research as a method for information systems research. *Journal of Information Technology*, *11*, 235-246. Retreived March 14, 2004, from *http://taylorandfrancis.metapress.com/media/fcjlvprqvam991h5tw5w/Contributions/5/N/G/E/5NGER4X63FYQE50N.pdf*

Bryman, A. (1988). *Quantity and quality in social research*. London: Unwin Hyman.

Callahan, J. & Moretton, B. (2001). Reducing software product development time. *International Journal of Project Management*, *19*(1) 59-70.

Fitzgerald, B. & Howcroft, D. (1998). Competing dichotomies in IS research and possible strategies for resolution. In R. Hirschheim, M. Newman, & J.I. DeGross, (Eds.). *International Conference on Information Systems* (pp. 155-164). Helsinki, Finland: Association for Information Systems.

Fitzgerald, G., Hirscheim, R., Mumford, E. & Wood-Harper, A. (1985). Information systems research methodology: an introduction to the debate. In E. Mumford, R. Hirscheim, G. Fitzgerald, & A. Wood-Harper, (Eds.). *IS - a doubtful science? Research methods in information systems* (pp. 3-9). North Holland: Elsevier Publishers.

Galliers, R. (1992). Choosing information systems research approaches. In R. Galliers (Ed.), *Information systems research: Issues, methods and practical guidelines* (pp. 144-162). Oxford: Blackwell Scientific.

Garcia, L. & Quek, F. (1997) Qualitative research in information systems: Time to be subjective? In *Proceedings of IFIP WG8.2 Working Conference on 'Information Systems & Qualitative Research' Philadelphia, USA; 31 May-03 June 97*. Chapman and Hall, London. Available online at: *http://is.lse.ac.uk/iswnet/pub/ifip8297.htm* [last accessed 12 August 2004].

Geddes, M., Hastings, C. & Briner, W. (1993). *Project leadership*. Gower.

Hughes, J. & Wood-Harper, T. (2000). An empirical model of the information systems development process: A case study of an automotive manufacturer. *Accounting Forum. 24*(4), 391-406.

Keen, P. (1984). VDT's as agents of change. In J. Bennett, D. Case, J. Sandelin, & M. Smith (Eds.), *Visual display terminals*. London: Prentice Hall.

Khazanchi, D. & Munkvold, B. (2000). Is information systems a science? An inquiry into the nature of the information systems discipline. *The DATA BASE for Advances in Information Systems*, *31*(3) 24-42.

Lincoln, Y. & Guba, E. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage.

Mandl-Striegnitz, P., Drappa, A. & Lichter, H. (1998). Simulating software projects - An approach for teaching project management. In C. Hawkins, M. Ross, G. Staples, & J.B. Tompson (Eds.), *Proceedings of INSPIRE '98 (International conference on Software Process Improvement - Research into Education and training)*,(pp. 87–98). London: University of Sunderland.

Marcella, R. & Knox, K. (2004). Systems for the management of information in a university context: an investigation of user need. *Information Research*, *9*(2). Paper 172. Retreaived March 27, 2004, from *http://InformationR.net/ir/9-2/paper172.html*

Morris, P.W.G. (2002). ICE James Forest lecture: science, objective knowledge, and the theory of project managemet. *ICE Civil Engineering*, *150*(May), 82-90.

Myers, M. D. (1997). Qualitative research in information systems. *MIS Quarterly*, (21:2) 241-242. MISQ Discovery. Archival version, June 1997, Retrieved March 12, 2004, from *http://www.misq.org/discovery/MISQD_isworld/*. MISQ Discovery. Updated version, last modified: *www.qual.auckland.ac.nz*

Nuldén, U. (1996). Escalation in IT projects: Can we afford to quit or do we have to continue.*Information Systems Conference of New Zealand* (pp.136-142). Palmerston North, New Zealand: IEEE Computer Society Press.

OECD. (2001). Management of large public IT projects: Case studies. *Public management service, public management committee report*. In J. Kristensen (Ed.), Organization for Economic Co-operation and Development. Retreived March 12, 2004, from *http://www.olis.oecd.org/olis/2001doc.nsf/LinkTo/PUMA-SBO-RD (2001)1*

Sauer, C. & Cuthbertson, C. (2004). The state of IT project management in the UK 2002 - 2003. *ComputerWeekly.com Ltd*. Retreived March 15, 2004, from *http://www.cw360ms.com/pmsurveyresults/index.asp*

Standish Group. (1994). *The CHAOS Report*. Retreived March 15, 2004, from *http://standishgroup.com/sample_research/chaos_1994_1.php*

Standish Group. (2001). *Extreme CHAOS*. Retreived from *http://www.standishgroup.com/sample_research/PDFpages/extreme_chaos.pdf*

Stoner, J.A.F., Yetton, P.W., Craig, J.F. & Johnston, K.D. (1994). *Management*. Sydney, Australia: Prentice Hall.

Wilson, T. (2002). Information science and research methods. In J. Steinerová, & S. Kimlika (Eds.), Kni•nicná a informacná veda (*Slovak* Library and Information Science), (pp. 63-71). Bratislava, Slovak Republic: Department of Library and Information Science, Comenius University.

Yin, R. (1994). *Case study research, design and methods* (2nd ed.). Newbury Park: Sage Publications.

KEY TERMS

Case Study Research: An in-depth investigation that attempts to captures lessons learned through studying the environment, procedures, results, achievements, and failures of a particular project or set of circumstances.

Development Literature: Literature about impoverished countries of the world that are trying to modernize or to find different ways of supporting their populations.

Focus Group: A small group interview, conducted by a moderator, which is used to discuss one or more issues.

Humanism: A philosophical approach that focuses on human value, thought, and actions.

Information Systems Project Management: The process of managing the creation of an IS through the establishment of project goals; organizing, leading, co-coordinating the efforts of staff processes and tasks; and controlling other resources to achieve a set of agreed objectives.

Interpretivism: A research approach that attempts to reach an understanding of social action in order to arrive at a causal explanation of its course and effects.

Positivism: A belief that natural science, based on observation, comprises the whole of human knowledge.

Project Team: All individuals who have made a significant contribution to make to the successful achievement of the project.

Project: A distinct set of coordinated activities with definite starting and finishing points, undertaken by an individual or organization to meet specific objectives within defined time, cost and performance parameters.

Techno-Economic Rationality: Logical justification for making a connection between technical advances and economic growth.

Isochronous Distributed Multimedia Synchronization

Zhonghua Yang
Nanyang Technological University, Singapore and Southern Yangtze University, China

Yanyan Yang
University of California, Davis, USA

Yaolin Gu
Southern Yangtze University, China

Robert Gay
Nanyang Technological University, Singapore

INTRODUCTION

A multimedia system is characterized by the integrated computer-controlled generation, manipulation, presentation, storage, and communication of independent discrete and continuous media data. The presentation of any data and the synchronization between various kinds of media data are the key issues for this integration (Georganas, Steinmetz, & Nakagawa, 1996). Clearly, multimedia systems have to precisely coordinate the relationships among all media that include temporal and spatial relationships. Temporal relationships are the presentation schedule of media, and spatial relationships are the location arrangements of media. Multimedia synchronization is a process of maintaining these relationships by employing appropriate synchronization mechanisms and algorithms. Multimedia synchronization is traditionally challenging, especially in distributed environments.

Three types of multimedia synchronization can be distinguished: intrastream synchronization, interstream synchronization, and intermedia synchronization (Crowcroft, Handley, & Wakeman, 1999). The approaches used for interstream synchronization can also be used for intermedia synchronization.

The word *synchronization* refers to time. The easiest way of synchronizing between streams at different sites is to use a single time reference. There are several ways to provide this time reference.

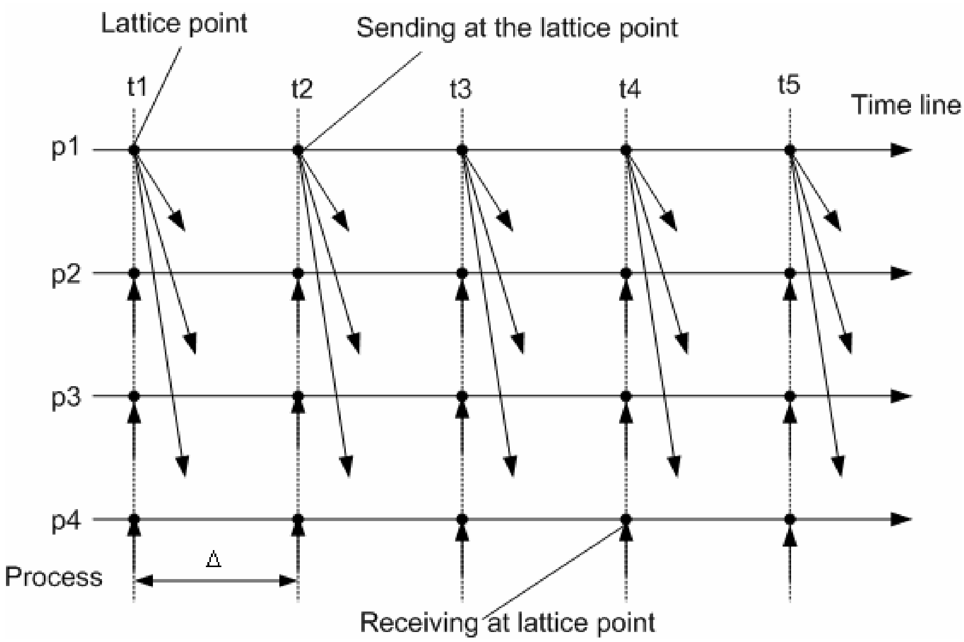
- The network will have a clock serve as a single reference. This approach is used in H.261/ISDN- (integrated services digital network) based systems. A single clock time is propagated around a set of codecs and multipoint control units (MCSs).
- The network deploys a clock-synchronization protocol, such as NTP (the network time protocol; Mills, 1993). The time stamps of media packets will be derived from the globally synchronized clocks. The isochronous synchronization approach as described in this article heavily relies on this time reference.

AN ISOCHRONOUS SYNCHRONIZATION APPROACH

The isochronous synchronization approach employs a clock-driven protocol for achieving multimedia synchronization (any one of three types of synchronization; Yang, Gay, Sun, Siew, & Sattar, 2002). This approach is particularly suitable for distributed collaborative multimedia environments where many-to-many multimedia communication is the basic interaction pattern. In this approach, multimedia synchronization is based on the use of synchronized physical clock time instead of any form of logical clock or sequence numbers, and thus clock synchronization across the distributed system is assumed. A real-time (synchronized) clock is incorporated in the system as a mechanism used for initiating significant events (actions) as a function of real time.

With globally synchronized clocks that satisfy the granularity condition, we can construct an *action lattice* (or *event lattice*; Kopetz, 1992). One dimension of this lattice represents the progression of time, the other dimension is the processes in the system (Figure 1). Processes in the system are designed to execute a simple *clock-driven protocol*, which requires that the events of sending and receiving messages are restricted to only

Figure 1. Lattice structure



occur at the lattice point of the globally synchronized space-time lattice (Figure 1). Thus, whenever an action has to be taken, it has to be delayed until the next lattice point of the event lattice.

This lattice structure greatly simplifies multimedia synchronization and readily maintains the temporal and causal relationship among the media.

The idea behind the clock-driven, isochronous synchronization is very simple and intuitive in that the easiest way to synchronize processes is to get them all to do the same thing at the same time. Using the simple mechanism based on the synchronized clock without requiring complex algorithms, the approach can equally well be applied to various multimedia applications in distributed environments, including live multimedia applications (live teleconferencing and CSCW) and stored media applications.

THE ORDERING PROPERTIES OF THE SYNCHRONIZATION PROTOCOL

In essence, what is really required for distributed multimedia synchronization is *order*; that is, a synchronization protocol must ensure that multimedia messages or streams are sent, delivered, and presented in an order that is consistent with the expected behavior of the distributed multimedia system as a whole. Clearly, multimedia systems have to precisely coordinate the relationships among all media. These relationships include temporal and spatial relationships. Temporal relationships are the presentation schedule of media, and spatial relationships are the location arrangements of media.

There are two specific cases concerning temporal order: *causal order* and *Δ-causal order*. These ordering concepts are derived from a happens-before relation, which is a more fundamental notion in distributed computing. The expression $a \rightarrow b$ is read as "a happens before b" and means that all processes agree that first, event a occurs, then afterward, event b occurs. The happens-before relation can be observed directly in two situations in a distributed environment: (a) If a and b are events in the same process and a occurs before b, then $a \rightarrow b$ is true, and (b) if a is the event of a message being sent by one process and b is the event of the message being received by another process, then $a \rightarrow b$ is also true. Obviously, a message cannot be received before it is sent or even at the same time it is sent since it takes a finite, nonzero amount of time to arrive. Note that happens-before is a transitive relation, so if $a \rightarrow b$ and $b \rightarrow c$, then $a \rightarrow c$.

The notion of causal order, as introduced by Birman and Joseph (1994), states that for any process, the order in which it delivers messages must respect the happens-before relation of the corresponding sending of the messages. More formally, a distributed computation E, all of whose messages is denoted as a set $M(E)$, respects causal order if for any two messages $m1$ and $m2$ and corresponding message-sending events $send(m1)$ and $send(m\_2)$, $send(m1) \rightarrow send(m2)$. If $m1$ and $m2$ have the same destination process, then for the message delivering events, $deliver(m1) \rightarrow deliver(m2)$.

In the definition of causal order, nothing is mentioned about the time at which the messages are delivered; also, it does not prescribe what to do with the cases of message loss and late arrival. However, in a distributed multimedia system, messages have limited *validity time*, after which the messages become useless and are allowed to be discarded. Messages that arrive at its destination within its validity time must be delivered within the expiration of its validity time and in its causal order. This motivates the notion of Δ-causal order, introduced in Yavatkar (1992) and formalized in Baldoni, Mostefaoui, and Raynal (1996). The Δ-causal order is defined as follows. A distributed computation respects Δ-causal order if (a) all messages in $M(E)$ that arrive within a time interval Δ are delivered within Δ and all the others are never delivered (they are lost or discarded), and (b) all delivery events respect causal order. That is, for any two messages $m1$ and $m2$ in $M(E)$ that arrive within Δ we have, if $send(m1) \rightarrow send(m2)$ and $m1$ and $m2$ have the same destination process, then $deliver(m1) \rightarrow deliver(m2)$.

The clock-based isochronous synchronization protocol requires the following ordering property be respected.

- **Same order:** The multimedia messages are delivered to the destinations in the same order.
- **Temporal order:** Different destinations see the different messages in the temporal order.
- **Simultaneity:** Different destinations see the same messages at about the same time.

Note that the temporal order is a prerequisite for the causal order. If and only if the occurrence of an event e_1 has preceded the occurrence of an event e_2 in the domain of real time, it is possible that e_1 has an effect on e_2. On the other hand, if it can be established that e_2 has occurred after e_1, e_2 cannot be the cause of e_1.

The basic mechanism is to use clock values as event time stamps to preserve temporal ordering of events and to achieve synchronization, and clocks in the system must have sufficient granularity or resolution. The granularity g of a synchronized clock is defined as the real-time duration between two consecutive global ticks. Obviously, the temporal order of two or more events, which occur between any two consecutive ticks of the synchronized clock, cannot be reestablished from their time stamps. This is a fundamental limit when using clock time for temporal ordering.

With globally synchronized clocks having sufficient granularity, we can construct an action lattice (or event lattice) as described above; that is, one dimension of this lattice represents the progression of time, and the other dimension is the processes in the system (Figure 1). Processes in the system are designed to execute a simple clock-driven protocol (an isochronous protocol below), which requires that the events of sending and receiving messages are restricted to only occur at the lattice point of the globally synchronized space-time lattice. Thus, whenever an action has to be taken, it has to be delayed until the next lattice point of the event lattice. This delay is the price we have to pay for the simple and intuitive synchronization protocols.

This lattice is a basic mechanism for the isochronous approach to multimedia synchronization. The lattice interval, Δ, is an important design parameter for the synchronization protocols. The following factors will affect how to choose Δ.

- The bounded end-to-end communication delay
- The validity time of multimedia, beyond which the multimedia objects become useless
- The granularity g and precision p of clock synchronization

As a general guideline, Δ must be large enough to accommodate all these factors; it must also be small enough not to unduly delay events.

AN ISOCHRONOUS SYNCHRONIZATION PROTOCOL

We now describe a general clock-driven, isochronous protocol to achieve desired multimedia synchronization. The protocol seeks to guarantee that in a time period, a set of processes will deliver the same messages at the same time and in the same temporal order. Here, "same time" must be understood to be limited by the clock skew (clock-synchronization precision) as much as π, meaning that two processes undertaking to perform the same action at the same time may in fact do so as much as π time units apart.

The protocol executes its events on every clock tick (i.e., at the lattice point), and executes a *No-Op* event by doing nothing if there is no communication event to take place on a clock tick. In practice, operations on a clock tick can be implemented by an interrupt-driven program; for example, the receipt of a message time-stamped T causes the setting of a clock interruption for $T+\Delta$, which in turn will cause the message to be processed at that time. When a process disseminates a message, it will not do so immediately; rather, it will wait until the next tick (i.e., next lattice point) on its clock and then time-stamp the message using its clock reading and send out the message. When the message arrives at its destination, it is not sufficient for the destination process to handle messages that have been received in ascending order by time-stamp. We must ensure that a process delivers messages only if no message with a smaller time stamp can be subsequently received. We say that a message is *stable for p* once no message with a lower time stamp can be delivered to the process *p*. Clearly, a message should be processed only after it becomes stable.

The lattice structure for the protocol operation provides a convenient way of establishing message stability. Here, testing the stability of a message can be accomplished by exploiting the bounds on delivery delays (i.e., the next lattice point) and process clocks. A message time-stamped T by process p will be received by $T+\Delta$ at every other process in the system according to each process' local clock, which is synchronized with the others. The message that arrives later than $T+\Delta$ is considered useless and discarded (because the multimedia message is beyond its validity time).

The isochronous synchronization protocol follows the following rules.

- **Sending Rule:** A process sends out the message at every clock tick; if there is no message to send, the process will do nothing (or you can think of it as sending out a null message) at the clock tick.
- **Stability Rule:** A message is stable at process p_i if the time stamp on the message is T and the clock at p has a value equal or greater than $T+\Delta$.
- **Delivery Rule:** Stable messages are delivered to the application in ascending order by time stamp.
- **Tie-Breaking Rule:** Two messages with the same time stamp from different processes are ordered according to the process ID, which is assumed to be unique.

Note that the ordering properties as required by the synchronization are guaranteed by executing this protocol. Noticeably, the temporal order, causal order, and Δ-causal order are all respected without requiring additional sophisticated algorithms. In executing this protocol, all processes have a consistent behavior toward messages.

The isochronous synchronization approach assumes globally synchronized clocks in distributed systems. The improved NTP (version 3) enjoys synchronization to within a few tens of milliseconds in the global Internet of today; the clock synchronization for LANs (local area networks) can obtain accuracy as high as a few microseconds. The introduction of the Global Positioning System (GPS) in the 1990s has further advanced the clock-synchronization technique (Herring, 1996). GPS has introduced an inexpensive way to obtain accurate information (including time information) using a radio receiver, which consists of nothing more than a GPS receiver and a network interface. Time obtained in this manner is accurate to a few tens of microseconds. Accuracy such as this is adequate for even the most demanding real-time applications. With this development and such an accurate timing source in place, we believe that the clock-based isochronous distributed synchronization, as advocated by Lamport (1984), provides a promising yet simple and intuitive alternative.

ACHIEVING ISOCHRONOUS SYNCHRONIZATION USING RTP/ RTCP

RTP (a transport protocol for real-time applications) is the real-time transport protocol within an Internet integrated-service architecture, which is designed to provide a quality-of-service guarantee beyond the current TCP/IP (transmission-control protocol/Internet protocol) best-effort service model. RTP provides end-to-end network transport functions suitable for applications transmitting real-time data (e.g., audio, video, or simulation data) over multicast or unicast network services. The data transport is augmented by a control protocol (RTCP) to allow monitoring of the data delivery in a manner scaleable to large multicast networks, and to provide minimal control and identification functionality. RTP and RTCP are designed to be independent of the underlying transport and network layers.

The noticeable feature associated with media synchronization is the 32-bit time stamp field in RTP data packets, the 64-bit NTP time stamp field, and the 32-bit RTP timestamp field in RTCP control packets. Although RTP does not mandate running the NTP to provide clock synchronization, running NTP is very useful for synchronizing streams transmitted from separate hosts. The mechanisms incorporated in RTP/RTCP enable the isochronous synchronization protocol.

One of primary functions of RTCP is to provide feedback in RTCP control packets (sender report, SR, and receiver report, RR) on the quality of the data distribution and information for intermedia synchronization. The RTP standard requires that the NTP time stamp (based on synchronized clocks) and corresponding RTP time stamp (based on data-packet sampling) are included in RTCP packets by data senders. This correspondence between the RTP time stamp and NTP time stamp may be used for intra- and intermedia synchronization for sources whose NTP time stamps are synchronized. Using the time-stamp mechanisms in RTP/RTCP and the lattice structure described in this chapter, our isochronous approach can be readily applied to multimedia systems, particularly in distributed many-to-many conferencing environments.

CONCLUSION

In distributed multimedia systems, there exist two approaches to protocol design, event driven and clock driven, and most protocols have taken an event-driven approach. While using physical time based on globally synchronized clocks for obtaining synchronization was advocated a long time ago, the clock-driven approach has not been popular in the distributed-system research community. In conjunction with the lattice structure, the isochronous protocol achieves the required synchronization, which guarantees the temporal order, including causal order and Δ-causal order, without additional sophisticated algorithms for respecting causality. The only assumption of the isochronous protocol is the synchronized clocks in a distributed system and the known bounds of the network communication delays. These assumptions can readily be satisfied with the deployment of modern networks. The mechanisms incorporated in the Internet standard real-time protocols such as RTP/RTCP make the isochronous synchronization approach readily applicable.

REFERENCES

Baldoni, R., Mostefaoui, A., & Raynal, M. (1996). Causal delivery of messages with real-time data in unreliable networks. *Journal of Real-Time Systems*.

Birman, K., & Joseph, T. (1994). Reliable communication in the presence of failure. In K. P. Birman & R. van Renesse (Eds.), *Reliable distributed computing with the Isis toolkit* (pp. 176-200). IEEE CS Press. (Reprinted from *ACM Transactions on Computer Systems, 5*(1), 47-76, February 1987)

Crowcroft, J., Handley, M., & Wakeman, I. (1999). *Internetworking multimedia*. Morgan Kaufmann Publishers.

Georganas, N., Steinmetz, R., & Nakagawa, N. (Eds.). (1996). Synchronization issues in multimedia communications. *IEEE Journal on Selected Areas in Communications, 14*(1).

Herring, T. A. (1996). The Global Positioning System. *Scientific American, 274*(2), 32-38.

Kopetz, H. (1992). Sparse time versus dense time in distributed real-time systems. *Proceedings of the 12th International Conference on Distributed Computing Systems*, 460-467.

Lamport, L. (1984). Using time instead of timeout for fault-tolerant distributed systems. *ACM Transactions on Programming Languages and Systems, 6*(2), 254-280.

Mills, D. L. (1993). Precision synchronization of computer network clocks. *ACM Computer Communications Review, 24*(2), 28-43.

Yang, Z., Gay, R., Sun, C., Siew, C. K., & Sattar, A. (2002). An isochronous approach to multimedia synchronization in distributed environments. In S. M. Rahman (Ed.), *Multimedia networking: Technology, management, and applications* (chap. 16). Hershey, PA: Idea Group Publishing.

Yavatkar, R. (1992). MCP: A protocol for coordination and temporal synchronization in multimedia collaborative applications. *Proceedings of International Conference on Distributed Computing Systems*, 606-613.

KEY TERMS

Clock Synchronization: Physical clocks in a network are synchronized to within certain precision and accuracy. The precision refers to the difference between readings of clocks, and the accuracy refers to the difference between the clock reading and the universal standard time.

Intrastream Synchronization: This, also called playout synchronization, ensures that the receiver plays out the medium a fixed time after it was generated at the source and it experienced variable end-to-end delay. In other words, intrastream synchronization assures that a constant-rate source at the sender again becomes a constant-rate source at the receiver despite delay jitter in the network.

Intermedia Synchronization: This is concerned with maintaining the requirements of the temporal relationships between two or more media. Lip synchronization between video and audio is an example of interstream synchronization where the display of video must synchronize with audio.

Interstream Synchronization: This ensures that all receivers play the same segment of a medium at the same time. Interstream synchronization may be needed in collaborative environments. For example, in a collaborative session, the same media information may be reacted upon by several participants.

Isochronous: The term refers to time-dependent processes where data must be delivered within certain time constraints. For example, multimedia streams require an isochronous transport mechanism to ensure that data is delivered as fast as it is displayed, and to ensure that the audio is synchronized with the video. Isochronous processes can be contrasted with asynchronous processes, which refers to processes in which data streams can be broken by random intervals, and synchronous processes, in which data streams can be delivered only at specific intervals. Isochronous service is not as rigid as synchronous service, but not as lenient as asynchronous service.

Multimedia: This term is used to indicate that the information and data being transferred over the network may be composed of one or more of the following media types: text, images, audio, and video.

NTP: NTP stands for network time protocol, and it is a standard Internet protocol used to synchronize the clocks of computers to some time reference.

Stream: This technique is for transferring data such that it can be processed as a steady and continuous stream. Streaming technologies are becoming increasingly important with the growth of the Internet because most users do not have fast-enough access to download large multimedia files quickly. If the stream is for transferring multimedia data, it is called a *multimedia stream*.

Validity Time: This is a time interval within which the message remains valid, available, and useful to its recipients. After the validity time of a message, the message becomes useless and may be discarded. The notion of validity time is important in multimedia communication.

Isoluminance Contours for Animated Visualization

Marion Cottingham
The University of Western Australia, Australia

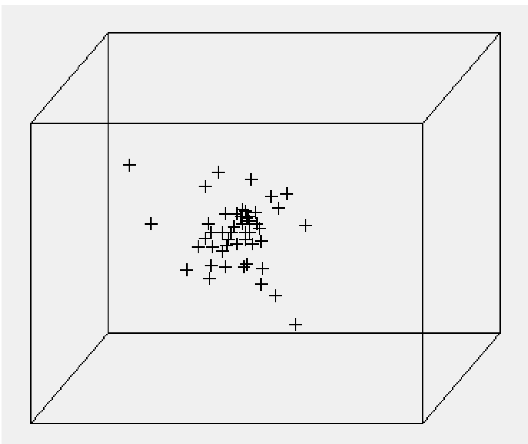
EVOLUTION OF THE ISOLUMINANCE CONTOURS MODEL

Visualization software applications were introduced to enable the sheer volume of data produced by computers worldwide to be viewed in picture form. Visualization makes it easier for the human visual system to pick up trends, patterns, and relationships in the data. The first wave of these applications typically generated simple two-dimensional line graphs, bar charts, and pie charts on expensive graphics workstations.

The second generation enabled three-dimensional images, with single data points typically denoted by dots or crosses, as shown in Figure 1. Axes lines or boxes enclosing the data were typically displayed. The whole image could then be rotated with the data points to give some indication of their orientation to each other. Since the majority of computers were not fast enough to perform the rotation in real-time, there was often some confusion in the picture's interpretation.

Graphics software applications became available, such as OpenGL or DirectX, containing functions for creating simple solid objects such as cubes, cylinders, cones, and spheres, and researchers could now attempt their own visualization. Most of these graphical applications were based on the polyhedral representation and did not provide any control over the number of patches representing each solid. Being able to animate the data points in real-time enhances trends and relationships in the data that are not easily identified in a still image. Since time is crucial in animation, representing solids with the minimum number of patches is of utmost importance.

Figure 1. Data cloud with data points denoted by crosses



The Isoluminance Contour model (IC model) was introduced in 1980 (Cottingham, 1981) as a means of representing these simple solids. It provides realistic results that far outweigh those produced using traditional methods, and in a fraction of the time. The amount of data used to generate an image can be easily adapted to suit each frame making real-time animation possible even on slow PCs.

Generating Images Using Traditional Graphics Techniques

Traditional graphics techniques adopt a divide-and-conquer approach to rendering solid objects. Each solid's surface is subdivided into planar patches that are generated one-by-one to produce the image while balancing speed with realism.

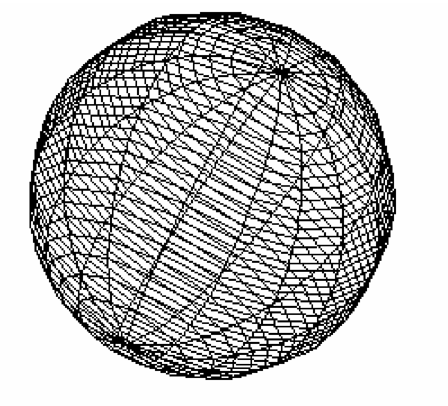
Storage Considerations

Each surface patch consists of four edges, each defined by two vertices specified in three-dimensional space. Topological information defines how edges and vertices are connected. Because surface patches from the simple solids form complete rows and columns, the topology can be implied by the order of storage using the Compressed Data Structure (CDS) (Cottingham, 1985; 1987).

Figure 2 shows a sphere represented by 640 facets, stored in 41 rows of 17 columns and generated by drawing lines to connect vertices stored at adjacent columns and rows. This is termed a wire frame drawing.

For speed, wire-frame drawings with only the edges of patches drawn are often used for quick previewing, but these do not provide much realism. At the next level, patches are flat shaded by calling a polygon fill function. However, in general, only triangular surfaces patches are guaranteed to be planar, so it is common practice for surface patches to be triangulated in an attempt to im-

Figure 2. Wire-frame sphere



prove the degree of realism—this doubles the number of patches. Figure 3 shows a sphere defined by 11 rows and 21 columns of vertices that specify 400 triangular patches or 400 quadrilateral patches.

Isoluminance Contour Model (IC Model)

The Isoluminance Contour (IC) model takes an entirely different approach by representing solids by planar contours that are defined by a collection of perimeter points. Points that share the same orientation to the light source are chosen, therefore, they all share the same constant chromaticity. To generate a contour, the illumination model needs to be called only once before calling a polygon fill function. Contours are rendered in depth order. For speed, you can also simply apply linearly spaced chromaticity values without calling an illumination model at all and select one that achieves the required effect.

Using the Isoluminance Contour model, each contour is defined by a single row of vertices from the CDS array, so there is no need to connect vertices from adjoining rows. Figure 4 shows the planar contours drawn as a wire-frame image and Figure 5 shows the planar contours filled with color.

Using the IC model, hidden surface elimination is automatically performed by generating contours in depth order (Painter's algorithm). The illumination model can be minimized to deciding on a range of color values and evaluating a contour's color according to its distance from the light source. For optimum realism, a complex illumination model can be applied. This would achieve results indistinguishable to those produced using ray casting. Ray casting is considered to give the highest degree of realism of all the traditional rendering techniques available. It generates images pixel by pixel and is renowned for the vast amounts of CPU time it consumes. However, the Isoluminance Contour Model produces identical results, give or take a pixel, yet can produce images in a fraction of the time taken to produce even wire-frame.

Comparison of Image Generation Speeds

Table 1 compares the CPU cycles taken to perform a single image containing a thousand spheres using traditional rendering methods on polyhedra with equivalent images produced using the IC model. For this comparison, each sphere was defined by 1250 patches—25 circular planes defined by 50 vertices, and the image was generated on a slow Pentium 400 PC. Although different hardware configurations will yield different results, the relative proportions of these times will remain constant.

For wire-frame drawings, all the edges for the polyhedral representation were drawn, displaying mainly quad-

Figure 3. Flat-shaded sphere represented by triangulated patches

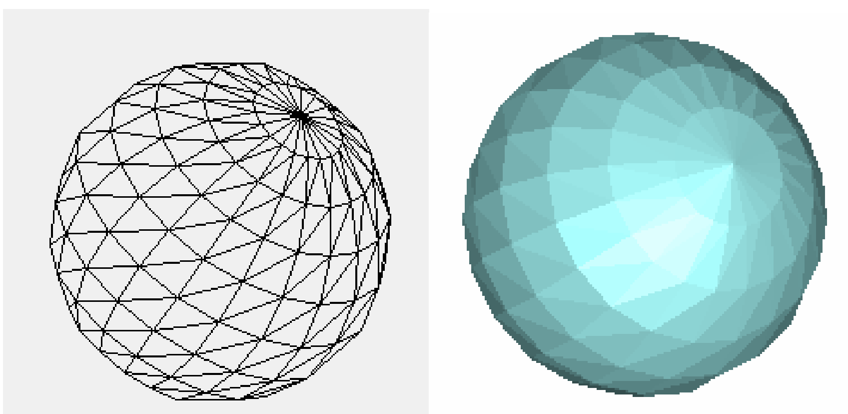


Figure 4. Isoluminance contour represents a sphere by a collection of planar contours

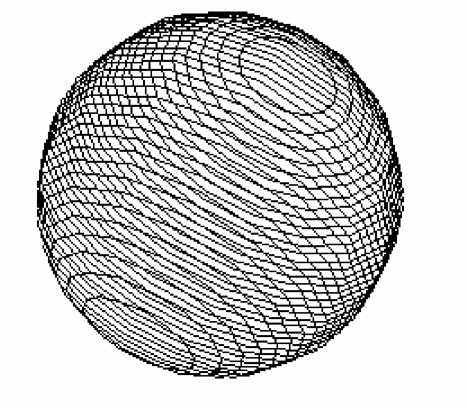
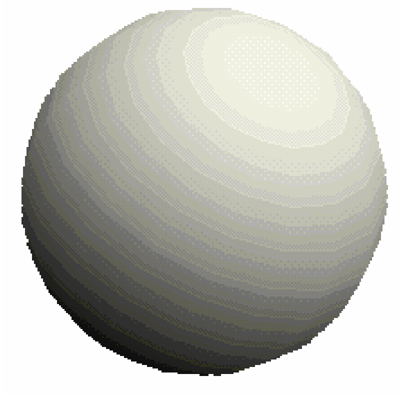


Figure 5. Filled isoluminance contours



rilateral patches; the IC model displays circular planes and avoids drawing the edges connecting the poles, as shown in Figure 4.

Different traditional techniques were applied to render the shaded images, each producing different levels of realism. The IC model used in the flat-shading timing included the specular reflection calculation, which is an add-on when required rather than part of the illumination model. The specular area shown in Figure 6 is the circular umbra that would appear if the light source was at the same angle as the viewing direction; this becomes more elliptical as the light is moved around. A higher level of realism would also require the penumbra to be displayed.

Using the IC model, a realistic image can be produced faster than even the simplest wire-frame image can be rendered by traditional methods. One reason is that only one normal and illumination calculation is required by the IC model for each of the 25 circular planes. In comparison, shading the polyhedral representation requires calculating the normals and illumination values for each of the 1250 vertices. In addition, when Gouraud's smooth-shading method (1971) is used, these illumination values have to be interpolated at every pixel lying inside each patch. The slowest time used Phong's (1975) method that requires the normals for every vertex to be interpolated and the illumination values calculated at every pixel lying inside each patch.

Figure 6 shows the images of one of the spheres rendered for this timing comparison. The top three spheres all use the polyhedral representation. From left to right, the polygons were rendered to produce a full wire-frame drawing, a wire-frame drawing with hidden lines re-

Figure 6. Spheres generated during the timing stage

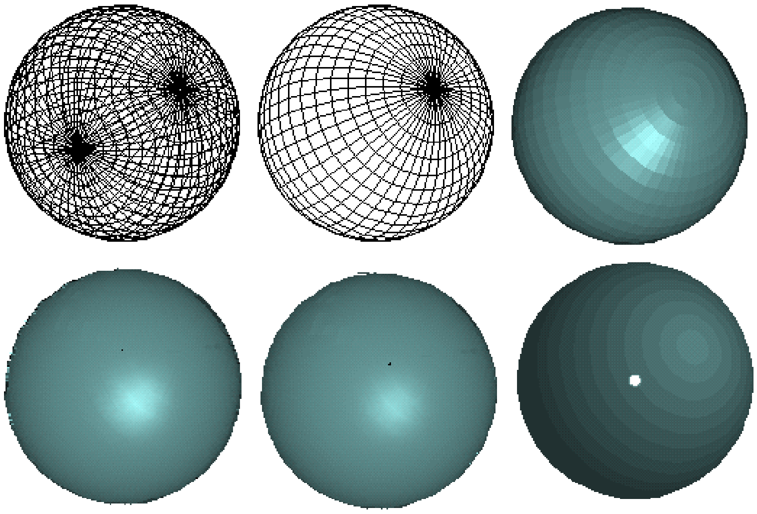


Table 1. CPU times required to render a thousand spheres

	Full Wire frame	Wire Frame Hidden lines removed	Shaded			
			Flat	Triangulated	Gouraud	Phong
Polyhedra	9.33	222	238	262	313	1388
IC model	5.08	5.13	8.07	-	-	-

moved, and a flat-shaded image. The bottom three spheres, from left to right, were rendered using Gouraud's method and Phong's method—both of which smooth-shaded patches from the polyhedra representation, and the IC model with specular reflection.

IC Model for Animated Visualization

Two approaches can be taken when animating data. In the first approach, the data points themselves are translated in three dimensions or rotated around one of the coordinate axes or about a line in three-dimensional space. In the second approach, the viewer moves around inside the data cloud and the data points remain static.

Figure 7 shows the three-dimensional visualization of the data points shown in Figure 1. These represent the evolution of mammals' brains in terms of changes in overall brain size. This data set was supplied courtesy of Professor Heinz Stephan, Max-Planck Institut fur Hirmforschung, Frankfurt. Exploratory univariate and multivariate analyses were applied by Mr. Willem De Winter and Professor Charles Oxnard (Department of Anatomy and Human Biology at The University of Western Australia) to the volumes of 11 non-overlapping brain parts from 921 specimens representing 363 species of various taxonomic groups consisting of primates, insectivores, bats, tree-shrews, and elephant shrews. This screen shot is from a visualization performed by software developed by Paul Merendah as an undergraduate, third-year Computer Graphics unit assignment (School of Computer Science and Software Engineering, The University of Western Australia). The assignment was specified and supervised by the author.

Uniform-sized spheres represent specimens, with the distances between the spheres representing the relative similarities in brain proportions (three dimensions). Perspective projection dictates the screen size of spheres. Species groups are color-coded (fourth dimension), and different levels of saturation and hue define individual relative significance within their animal group (fifth and sixth dimensions).

Figure 7. Visualization of the proportional sizes of the major brain divisions for a wide range of mammals

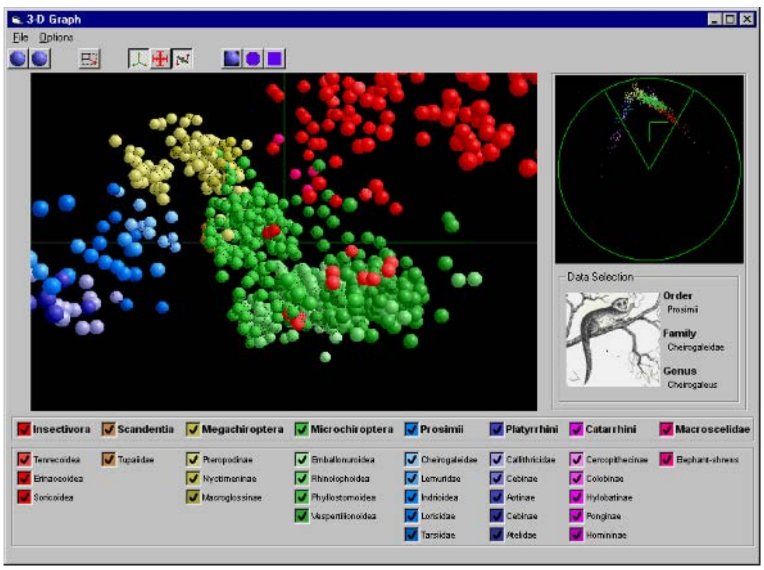
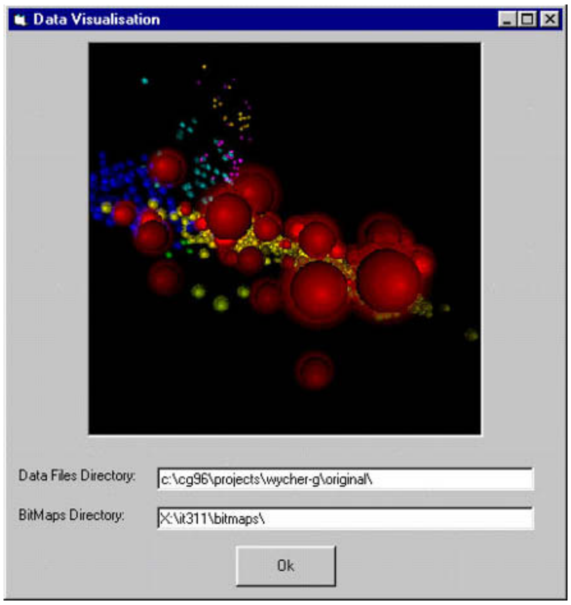


Figure 8. Visualization of the major brain divisions of mammals with different parameters in the illumination model



The data set contained 921 data points (spheres), and with the IC model it was possible to fly around the data cloud in real-time on a Pentium 166 PC. The relative position of the light source to the viewing position was assumed to be constant so that the contours of the spheres did not require reconstruction at any time during the simulation.

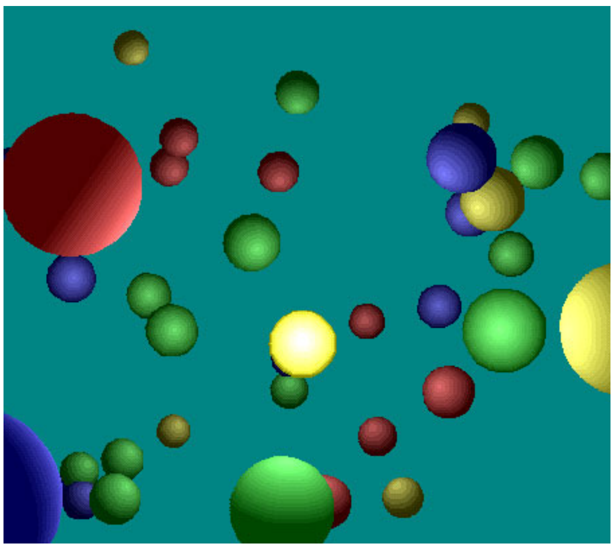
Figure 8 shows the same data set being visualized by software developed by Graham Wycherly in the same project; this software uses a different illumination model.

The ability to fly around the data clouds in real-time on such a slow PC was due to the adaptability of the IC model. The number of contours used for rendering any sphere was based on its screen size. The animation performed in these two student projects enabled even more relationships to be discovered, as before this dynamic visualization it was thought by biologists that the evolution of the brain in mammals was simply a matter of changes in the overall brain size (Finlay & Darlington, 1995). This visualization demonstrated the existence of different phylogenetic trends in the evolution of the mammalian brain and that where species shared a particular lifestyle the brain had evolved converging towards a similar proportional organization (Oxnard and de Winter 1997: 2 journals, de Winter).

Both these projects assumed that the light source was positioned infinitely far away so that all light rays were parallel and all spheres were shaded the same. However, it is possible to remove this assumption and still attain real-time animation if you have a sufficiently powerful PC.

Allowing the position of the light source and the viewing position to move independently means that the light direction vector is no longer constant throughout the scene, therefore, the spheres at different positions in space no longer have identical shading. This makes it necessary to create each sphere from scratch whenever the viewing position or the light source changes. This is still achievable in real-time on a Pentium 400 or faster PC.

Figure 9. Light source positioned in the data cloud in full view of the viewing position



This independence means that the light source itself may become visible. Figure 9 shows the light source (brightest sphere) positioned amongst the data cloud in full view of the viewing position.

You may be required to display a timed series of data points on the screen simultaneously. The visualization shown in Figure 10 displays "bid and ask" stock market data over a five-day period. Professor Philip Brown from the Department of Accounting and Finance at The University of Western Australia supplied the data. The software was developed by Ken Hooi for a third-year Computer Graphics unit assignment that was specified and supervised by the author.

"Asks" are shares for sale at a specified price and "bids" are offers to buy at a stated bidding price. The asking price is usually more than the bidding price. The bids are displayed as blue spheres and asks as red spheres. The software creates an x- and y-axes pair for each day requested by the user and displays the bids and asks not realized for that day. The x-axis depicts the volume of the bid or ask parcels of shares, and the y-axis depicts the bid or ask price per share. When an ask parcel and a bid parcel match in price and volume then a transaction takes place and the relevant spheres disappear off the screen. The dynamic view allows the user to watch while the bid and ask pairs are being processed, so that he/she can view the trades as they happen.

IC Model as an Adaptive Model for Real-Time Animation

Solids are created from templates to avoid having to recalculate the vertices for each one individually. To maintain the level of realism required, a template is defined by enough vertices to render the solid at its largest screen size. Since it would be inefficient to include all these vertices as the solid's screen size diminishes, the number of vertices copied from the sphere template is adapted to reflect the screen area. For example, suppose the viewing position starts close to a sphere solid and moves away, causing the sphere to zoom away into the distance. Initially, all the vertices must be rendered to maximize the smoothness of the sphere's surface. As the sphere moves away and its screen size grows smaller, every second vertex from every second contour can be used, then every third vertex from every third contour, and so on. For sphere solids, the screen size is calculated from the sphere's diameter, which is approximated by (maximum X – minimum X).

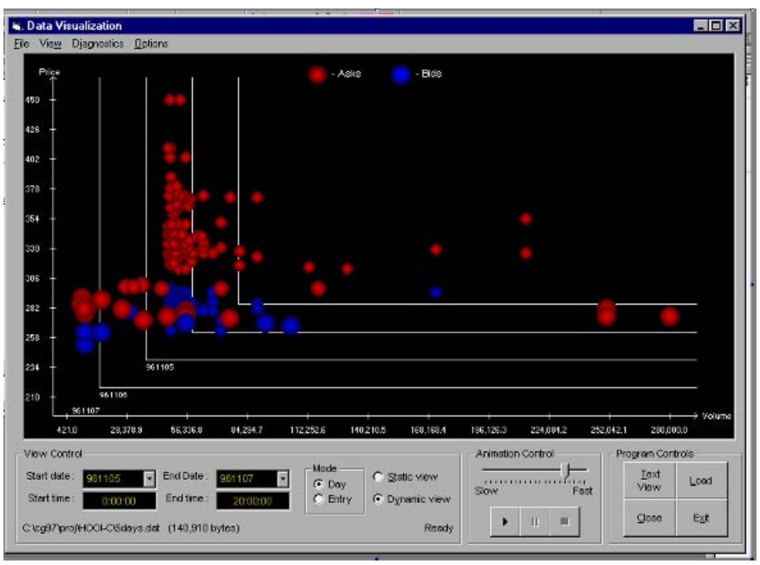
Additional information on the IC model is available from Cottingham (1989, 2004a, 2004b), Cottingham and Conway (1988), and Conway (1991, 1993).

FUTURE TRENDS

First, the Isoluminance Contour model could be generalized by introducing a pre-processing step that takes existing data sets and combines any adjoining patches that have similar normal direction vectors. This would require a more general data structure such as Baumgart's (1975)Winged-Edge Data Structure rather than the simpler CDS.

Second, because of the simplistic nature of the IC model, implementing it in hardware should be reasonably straightforward.

Figure 10. Displaying "bid and ask" stock market data one day per axis



CONCLUSION

The IC model is quick and simple to program as the topology of the solids is implied by the order of storage in a simple two-dimensional array. It eliminates the need for most of the steps required by traditional computer graphics techniques. It is extremely fast and does not require any quick-previewing methods to be developed. The realistic results that are indistinguishable from ray-tracing speak for themselves.

REFERENCES

Baumgart, B.G. (1975). A polyhedron representation for computer vision. *Proceedings of AFIPS National Computer Conference*, May 1975, Stanford University, California, (pp. 589-596).

Conway, D.M. (1991). Constructive solid geometry using the isoluminance contour model. *Computers and Graphics*, 15(3), 341- 347.

Conway, D.M. (1993). *Fast three-dimensional rendering using isoluminance contours.* Doctoral Thesis, Department of Computer Science, Monash University, Australia.

Cottingham, M.S. (1981). *Movies.* Senior Honours Thesis, The University of Glasgow, Scotland, UK.

Cottingham, M.S. (1985). A compressed data structure for surface representation. *Computer Graphics Forum,* 4(3), 217-228.

Cottingham, M.S. (1987). Compressed data structure for rotational sweep method. *Ausgraph'87 Conference Proceedings, Australian Computer Graphics Association,* May 1987 (pp. 387-404).

Cottingham, M.S. (1989). Adaptive data structure for animated (polyhedral) objects. *International Conference on CAD & CG Proceedings*, Beijing, China, August (pp. 722-726). International Academic Publishers.

Cottingham, M.S. (2004a). *Computer Graphics with Visual Basic 6*, Second Edition. Guildford, WA: Vineyard Publishing, Australia.

Cottingham, M.S. (2004b). Isoluminance contours: A fast, simple model for 4D animated visualization. In John DiMarco (Ed.), *Computer graphics and multimedia: Applications, problems, and solutions.* Hershey, PA: Idea Group Publishing.

Cottingham, M.S. & Conway, D.M. (1988). The isoluminance contour model. *AUSGRAPH88 Conference Proceedings*, Melbourne, Victoria. p. 43-50.

Finlay, B.L.(1995). Linked regularities in the development and evolution of mammalian brains. *Science,* 268, 1578-1584.

Gouraud, H. (1971). Computer display of curved surfaces. Tech. Rep. UTEC-CSC-71-113, Dept. of Computer Science, University of Utah, Salt Lake City, Utah.

Oxnard, C.E., & de Windter, W.(2001). Evolutionary radiations and convergencies in the structural organisation of mammalian brains. *Nature. 409*,710-714.

Phong, B.T. (1975). Illumination for computer generated pictures. *Comm. ACM,* 18(6), 311-317.

Winter, W. de & Oxnard, C.E. (2001). Evolutionary radiations and convergencies in the structural organisation of mammalian brains. *Nature,* 409, 710-714.

KEY TERMS

Data Cloud: Collection of data points in space.

Illumination Model: A mathematical formula for calculating the amount of diffuse light, ambient light, and specular reflection that contributes to the color value that will represent a particular surface point on the screen.

Isoluminance Contour: Planar polygon that has a single constant color at all points it encloses.

Real-Time Animation: Frames created at a minimum rate of 20 per second to avoid flicker.

Smooth-Shading Techniques: Interpolative techniques that restore the smoothness of solids whose surfaces are represented by planar polygons.

Viewing Position: The point in three-dimensional space where the user is positioned to view the data. There is also an associated direction where the user is looking.

Visualization: Viewing data in picture form in an attempt to gain a further understanding of the trends and relationships it contains.

Wire-Frame Drawing: Only the lines defining the edges of patches are drawn, which is useful as a quick previewing tool.

Issues and Challenges for IT in Small Business

Stephen Burgess
Victoria University, Australia

INTRODUCTION

This article examines the main issues and challenges facing small business owners and managers in the manner in which they use information technology and Internet technologies within their businesses. The first part of the article concentrates upon the issues. The "Challenges" section examines some of the areas that small business researchers are targeting in their recommendations for how the use of IT in small businesses can be improved. These are areas that small businesses often neglect or, even worse, know nothing about.

BACKGROUND

Issues

This section looks at some of the issues facing researchers as they attempt to investigate the use of IT in small businesses. It starts off by looking at some of the various definitions of small business that are being used. Then barriers to successful use of IT by small businesses are examined, leading into the factors that lead to its successful use.

What is Small?

When studying the use of IT in small business, the range of definitions used to describe "small business" is interesting to say the least. This range can make it extremely difficult for researchers to "match up" different small business studies. A 2003 study by worldwide members of the Information Resources Management Association Special Research Cluster on Small Business and Information Technology[1] found that:

- Definitions of "small business" ranged from less than 20 (Australasia), 50 (Europe) and 100 (North America) employees (with some definitions including requirements for annual turnover and asset levels).
- Definitions of "micro business" ranged from less than 5 to less than 10 employees.
- Definitions of "medium" business ranged up to 200, 250 and 500 employees!

A common acronym used to represent small and medium sized businesses is "SME". There was quite a deal of argument as to whether the term was of any use at all – given the vast differences between small and medium sized businesses. Still, half of the responding members felt there was still a use for the term.

Pollard and Hayne (1998) have identified a number of studies that set the upper limit of employees for a business to be classified as small as 50, 100, 200, 250 or 500!

Business Size is Important!

Why is it so important to consider the size of the business? A number of studies suggest that there is a relationship between the size of a business and its level of adoption of IT (McDonagh & Prothero, 2000). There is also a relationship between the size of a business and the different characteristics it will have that can lead to the successful use of IT (Igbaria et al., 1997; Pollard & Hayne, 1998). As such, research findings based upon traditional "MIS" in larger businesses are not necessarily directly applicable to small businesses.

Barriers and Success Factors

The literature around the area of small business and information technology is rife with what is now a fairly accepted list of "barriers" to the successful implementation of IT in small businesses. These barriers typically include (Igbaria et al., 1997; Management Services, 1997; McDonagh & Prothero, 2000; Pollard & Hayne, 1998):

The cost of IT
- Lack of time to devote to the implementation and maintenance of IT
- A lack of IT knowledge combined with difficulty in finding useful, impartial advice
- Lack of use of external consultants and vendors
- Short-range management perspectives
- A lack of understanding of the benefits that IT can provide, and how to measure those benefits
- A lack of formal planning or control procedures

One advantage that small businesses (especially innovative ones) have is that they are flexible. They are able to preserve labour relationships, bring a "personal touch" to operations, cater to niche markets and they have small capital requirements. The fact that they are under constant pressure can also spur them to be inventive and innovative in their business operations (International Trade Forum, 1999).

Having identified some of the barriers to the successful use of IT, there is also a fairly common list of factors that are listed in the literature that appear to indicate a greater chance of successful implementation of IT in small businesses.

Some of these factors are (Naylor & Williams, 1994; Swartz & Walsh, 1996; Yap & Thong, 1997; Zinatelli et al., 1996):

- The involvement of owner/managers in the implementation of IT
- The involvement of users (employees) in development and installation
- The training of users
- The selection of applications chosen for computerisation
- The use of disciplined planning methodologies in setting up applications
- The number of analytical/strategic (versus transactional) applications being run
- The level of IT expertise within the organisation
- The role of the external environment (especially consultants and vendors)

The idea behind introducing this list (and the barriers listed earlier) is that you will see these points raised again and again in the following sections as some of these and other issues are discussed.

Role of Owner/Managers

One of the key factors leading to successful use of IT in small businesses identified in the previous section was the involvement of small business owner/managers in the IT implementation.

There is some evidence to indicate that managers in small businesses are less likely to know how to use IT effectively or to keep up with the latest trends in IT than their counterparts in larger businesses (Pollard & Hayne, 1988). Igbaria et al. (1997) cite a number of references to support the view that management support can promote the acceptance of IT. They found that the support of management positively affected the perceived ease of use and the perceived usefulness of IT within the small business.

Applications

These days, the vast majority of small businesses in most countries have computers. Small businesses that use computers mainly use them for administrative and operational purposes (such as accounting, budgeting, payroll, inventory control and the like) (Bridge & Peel, 1999; El Louadi, 1998; Pollard & Hayne, 1998). Much of the software that is used by small businesses is purchased “off the shelf” (McDonagh & Prothero, 2000), although there is some evidence to suggest that small businesses with particular (specialised) needs are prepared to invest in customised software (Burgess, 1997).

The literature also suggests that some small businesses are beginning to realise that IT can be used to gain competitive advantage (Pollard & Hayne, 1998).

Location

Why is location important when considering the use of IT in small business? The major answer to this is a combination of resources and distance. The further you are away from resources, the longer it takes and the more it costs to get them. This can particularly be the case with hardware and software purchases, training and support.

Another reason for examining location is **culture.** Some countries, and even different regions within countries, have their own traditions and their own established ways of doing things. This can influence the behaviour of small businesses and the manner in which they use IT.

Developing Countries

Small businesses make up a major portion of businesses in developing countries (in some countries the percentage is higher than in developed countries). One of the major barriers faced by small businesses in developing countries is access to information, especially information used in decision making. Another problem is the lack of data sources from which to obtain the type of information required. Problems with the technological infrastructure of developing countries only exacerbate this (Sawyerr et al., 2000).

Today, the cost of installing a national telecommunications structure has fallen. When combined with reduced costs of international transport it is easier for small businesses in many developing countries to gain access to international markets (International Trade Forum, 1999). Uptake of the Internet in developing countries has been on the increase since 1996, the level of growth (again) being hindered by problems with the telecommunications infrastructure (Gallagher, 1999).

Rural Small Businesses

Some of the problems facing rural small businesses are similar to those facing small businesses in developing countries. One of the benefits that the Internet may provide is remote access to many desired IT resources, such as training (Gallagher, 1999). A year 2000 survey of Australian small businesses revealed that 29% of metropolitan small businesses had a Web site, compared to 20% of rural small businesses. The main reason given by rural small businesses for not having a Web site was that they did not have access to the skills needed to design, build and maintain a Web site (Telstra Corporation & NOIE, 2000).

Industry

There is some relationship between the industry that small businesses are involved in and the types of IT that they use. For instance, professionals and finance/property/business services small businesses are the ones that use the latest IT. The building/ construction industry is typically the least enthusiastic to use IT at all, let alone the latest IT (Burgess, 1997). There is also some evidence that different software is used by different industries. For instance, the building/construction industry uses scheduling, drafting/design and project management software. Manufacturing businesses use software for scheduling, operating machinery and monitoring operations. Wholesale/retail small businesses use point-of-sale software (as well as bar code scanners) (Burgess, 1997).

CHALLENGES

This section examines some of the areas that small business researchers are targeting in their recommendations for how the use of IT in small businesses can be improved. These are areas that small businesses often neglect or, even worse, know nothing about.

Planning

One of the barriers identified earlier that hinder the effective use of IT in small businesses is a lack of formal planning and control methodologies. This relates to a lack of knowledge of how to plan effectively, lack of time and money to seek this knowledge, lack of time to apply it even if they have the knowledge and a lack of understanding that they even need the knowledge! Small businesses are, however, concerned with issues relating to how they can operate more effectively and efficiently and/or how they can grow (El Louadi, 1998). One of the problems is that management practice in small businesses is often based on the short term and is informal and ad hoc. Much of the time is spent "surviving," so that little time can be devoted to examine IT projects (Pollard & Hayne, 1988).

Working Out the Benefits of IT

Another barrier to the successful use of IT in small businesses is a lack of understanding of the benefits that IT can provide, and how to measure those benefits. The most common way used to determine the level of IT success is to measure small business user satisfaction with information technology. Such measures of user satisfaction have one major problem – they are linked with user expectations (Naylor & Williams, 1994). For instance, an owner/manager understanding the strategic benefits that IT can provide may be less satisfied with a simple transactional system than an owner/manager who is unaware of these strategic benefits. This is despite the possibility that they may be reviewing systems that perform in a similar manner. Again, the problem falls back to a lack of proper knowledge about the advantages that IT can provide.

Training

Factors relating to a lack of knowledge of IT or lack of understanding of the benefits of IT have been mentioned a number of times already in this article. Appropriate training in IT has been mentioned earlier as one of the factors leading to successful use of IT in small businesses. Igbaria et al. (1997) found that the amount of training that users had received from other users or IT specialists within the firm had an effect on the perceived usefulness of the system. External training had a positive effect on the perceived ease of use of IT, which they theorise may enhance computer skills and reduce "negative attitude" barriers to the acceptance of IT. A lack of training was a cause of user frustration.

External Information

Small business managers perceive more uncertainty in their environment than their counterparts in larger businesses. Effective management of external information can help them to reduce the level of uncertainty that they feel (El Louadi, 1998).

Vendors and Consultants

In many instances, small businesses have to rely on the IT expertise of vendors and/or consultants because of a lack of internal IT expertise. Igbaria et al. (1997) have found that good external support provided by vendors and/or consultants, such as technical support, training and a harmonious working relationship can reduce the risk of IT failure in small businesses.

There is a view, however, that vendors and consultants do not understand the small and medium business market and that the level of support provided by them is only adequate or less than adequate (Management Services, 2000). Careful selection of vendors and/or consultants is vital.

Government Support

Governments worldwide are beginning to realise the importance of the small business community. The role of government in developing countries has already been touched upon in this article. IT is one of the areas that are the subject of increased government resources, through improved information programs, increased training opportunities and technology support grants and awards.

The Internet and Electronic Commerce

There is no doubt that a rapidly increasing number of small businesses are using the Internet (McDonagh & Prothero, 2000). Gallagher (1999) claims that the level of Internet use is growing faster than any other technology in history! The level of this growth is such that almost half of this book is devoted to chapters that address the use of the Internet or e-commerce in small businesses.

Many small businesses are establishing a presence on the Internet, be it through the development of a Web site or through the use of e-mail. Two primary uses of Internet technologies in small businesses are the use of e-mail for communication and the Internet to gather information (Dandridge & Levenburg, 2000; McDonagh & Prothero, 2000). "Proactive" small businesses were likely to have a Web site and saw it as (Telstra Corporation & NOIE, 2000):

- An essential part of their promotional and advertising mix,
- A means of providing business and product information to clients, and
- A basis for developing an order and payment facility

FUTURE TRENDS

As time goes on, governments are increasingly becoming aware of the importance of small businesses to their economies and are providing increased resources to their support. This has resulted in various support programs for small businesses that have directly or indirectly resulted in improvements in the use of IT by small businesses. It is anticipated that this trend will continue. There is also an increasing awareness of the importance of the efficient and effective use of IT by small businesses by the research community. There continues to be increased amounts of research being carried out in the field, which can only lead to greater understanding. Groups such as the Information Resources Management Association Special Research Cluster on the Use of Information Technology in Small Business reflect the increasing research being carried out.

CONCLUSION

This article has introduced a number of issues related to the use of IT in small businesses. When looking at small business research, it is important to determine what the researcher's view of "small" actually is. This is so that proper comparisons can be made across studies.

Barriers, opportunities, applications and success factors in relation to the use of IT were identified for small businesses as being common areas covered in the literature. The importance of owner/managers, the location of the business and industry involvement were other issues that were discussed.

A number of challenges facing small businesses were also identified. The need for planning, to be able to measure success, to provide effective training and have access to important external sources of information were discussed. Finally, the effect of e-commerce and the Internet on small businesses was introduced.

REFERENCES

Belisle, J.D., & Czinkota, M.R. (1999). Trade must extend to poorer countries. *International Trade Forum, 3,* 11-13.

Bridge, J., & Peel, M.J. (1999, July-September). Research note: A study of computer usage and strategic planning in the SME sector. *International Small Business Journal, 17*(4), 82-87.

Burgess, S. (1997, June). *Information technology and small business: A categorised study of the use of IT in small business*. Detailed Survey Report, Small Business. Victoria, Melbourne.

Dandridge, T., & Levenburg, N.M. (2000, January-March). High tech potential? An exploratory study of very small firm's usage of the Internet. *International Small Business Journal, 18*(2), 81-91.

El Louadi, M. (1998, June). The relationship among organization structure, information technology and information processing in small Canadian firms. *Canadian Journal of Administrative Sciences, 15*(2), 180-199.

Gallagher, P. (1999). E-commerce trends. *International Trade Forum, 2*, 16-18.

Igbaria, M., Zinatelli, N., Cragg, P., & Cavaye, A.L.M. (1997, September). Personal computing acceptance factors in small firms: A structural equation model. *MIS Quarterly, 21*(3), 279-305.

International Trade Forum. (1999). Export strategies for small firms. *1*, 9-12.

Management Services. (1997). Computers fail to click with small businesses. *41*(9), 4.

Management Services. (2000). Nearly half of SMEs believe that the Internet and IT has no impact on them. *44*(10), 6

McDonagh, P., & Prothero, A. (2000). Euroclicking and the Irish SME: Prepared for e-commerce and the single currency? *Irish Marketing Review, 13*(1), 21-33.

Naylor, J.B., & Williams, J. (1994). The successful use of IT in SMEs on Merseyside. *European Journal of Information Systems, 3*(1), 48-56.

Pollard, C.E., & Hayne, S.C. (1998, April-June). The changing faces of information systems issues in small firms. *International Small Business Journal, 16*(3), 70-87.

Sawyerr, O.O., Edbrahimi, B.P., & Thibodeaux, M.S. (2000, June). Executive environmental scanning, information source utilisation, and firm performance: The case of Nigeria. *Journal of Applied Management Studies, 9*(1), 95-115.

Swartz, E., & Walsh, V. (1996). Understanding the process of information management in small firms: Implications for government policy. *19th ISBA National Conference Proceedings*, Birmingham (pp. 387-399).

Telstra Corporation and NOIE (The National Office for the Information Economy). (2000). *Small business index: Survey of computer technology and e-commerce in Australian small and medium businesses*. Melbourne, Australia: Pacific Access Pty Ltd.

Yap, C.-S., & Thong, J.Y.L. (1997). Programme evaluation of a government information technology programme for small businesses. *Journal of Information Technology, 12*, 107-120.

Zinatelli, N., Cragg, P.B., & Cavaye, A.L.M. (1996). End user computing sophistication and success in small firms. *European Journal of Information Systems, 5*, 172-181.

KEY TERMS

Medium Business: This term is used to describe businesses that are too large to be considered as being small and too small to be considered as being large. This somewhat vague description is matched by the varying definitions of medium sized business there are around. In relation to use of IT, medium sized businesses usually exhibit more of the characteristics of larger businesses than smaller ones.

Micro Business: This term is used to describe very small businesses. Many of these operate as family businesses. They form the majority of businesses.

Owner/Manager: In most small businesses the owner/manger is the driving force behind the business, and as such can be the catalyst for change in the business that occurs through or around IT. In some small businesses the positions of owner and manager are separated.

Small Business: Small business can be measured by number of employees, annual turnover and/or assets. It usually represents those businesses with up to 20, 50 or 100 employees (depending upon the region being investigated). This term encompasses micro businesses.

SME: This acronym is used to refer to small and medium sized enterprises as a collective group.

ENDNOTE

1 Refer to http://www.businessandlaw.vu.edu.au/sbirit for the group's Web site.

Issues in Delivering Course Material Via the Web

Karen S. Nantz
Eastern Illinois University, USA

INTRODUCTION

"Education over the Internet is going to be so big it is going to make e-mail usage look like a rounding error."
- John Chambers, Cisco Systems, New York Times, November 17, 1990

Web-based courses (Mesher, 1999) are defined as those where the entire course is taken on the Internet. In some courses, there may be an initial meeting for orientation. Proctored exams may also be given, either from the source of the Web-based course or off site at a testing facility. The Internet-based course becomes a virtual classroom with a syllabus, course materials, chat space, discussion list, and e-mail services (Resmer, 1999). Navarro (2000) provides a further definition: a fully interactive, multimedia approach. 2002 figures suggest that over two million students are taking at least one Web course (Thornton, 1999). According to *E-Learning* (2001), more than 3,000 universities will offer substantial Web courses by 2004.

The Web-enhanced course is a blend with the components of the traditional class, while making some course materials available on a Web site, such as course syllabi, assignments, data files, and test reviews. Additional elements of a Web-enhanced course can include online testing, a course listserver, instructor-student e-mail, and other activities on the Internet. Navarro (2000) defines this type of course as a "digitized text" approach.

Navarro (2000) suggests that faculty are far more likely to start by incorporating Internet components into a traditional course rather than directly offering Web-based courses. These Web-enhanced courses might be considered the transition phase to the new paradigm of Internet-based courses.

BACKGROUND

A 1999 research study showed that 27.3% of the faculty members think they use the Internet for the delivery of course materials, but only 15.6% actually did so. Of this group, the major use was simply the substitution of a Web page for the printed page. Most faculty members (73.8%) updated their sites so infrequently that the sites only served to replicate printed handouts. In a follow-up study at the same university, the number of faculty who used Web pages to enhance their courses showed a decrease from the previous year (Garrett, Lundgren & Nantz, 2000). In the same study, 22% of the faculty are not ever planning to use a Web site for delivery of any portion of their courses. Less than 5% are truly incorporating Web technology into their courses in a meaningful way.

Clearly, there are many positive reasons for using a Web site in a course including greater efficiency in the delivery of materials, providing up-to-the-minute content, enhanced status for the course and faculty, and of course the seemingly inevitable trend to use more technology in education.

Some of the issues inhibiting the use of Web sites include: lack of faculty knowledge of Web page design, HTLM, server sites, and file transfer protocols (Nantz & Lundgren, 1998); perceived need for Web glitz to provide entertainment along with content, such as high interactivity, animation, audio and video streaming; lack of accessibility to Web resources for both faculty and students (Rao & Rao, 1999); sufficient training for faculty (Rupp, 1999), and compensation for cyberprofs who typically spend twice as much time developing and teaching Web-based courses for no extra pay (Navarro, 2000).

A Course Web Site Classification

Web classes can be classified in six different levels. At the top levels are the Internet-based classes—those created and organized to be Web delivered. The middle levels involve a Web class that uses the Internet for delivery of content and communication among the course registrants, but also uses face-to-face meetings for some classes, orientation, and testing. At the lowest level, some course materials are simply presented in a hypertext format that replaces traditional printed handouts. Table 1 shows the classification levels of academic Web pages by typical content and maintenance levels.

The six levels presented above indicate progression from the most basic Web-enhanced course to a course delivered fully on the Internet. Faculty would likely proceed through the levels to reach Level 4 for traditional classes unless limited by resources, expertise, and admin-

Table 1. Classification of academic Web pages

Level	Description	Typical Content	Maintenance Level Required
1	Traditional Course Presentation, Basic-Level Course Materials on Web—Internal Links	Instructor data (name, phone, office hours, e-mail address) Course materials (syllabus, generic schedule, assignments); non-interactive	Low—static pages after initial upload. Low-volume e-mail correspondence.
2	Traditional Course Presentation—Intermediate-Level Course Materials on Web—External Links	All Level 1 Some external links, such as textbook and reference sites; non-interactive.	Low—mostly static pages with occasional updates and checking of external links. Low-volume e-mail correspondence.
3	Traditional Enhanced Course Presentation—Intermediate-Level Course Materials on Web and Web Content Delivery	All Level 2 All traditional course materials posted. Web access in class used for delivery of some course content. Some assignments/requirements involve interaction, e.g., e-mail submissions, listserv postings.	Weekly updates to schedule, FAQ, course materials, notes to students. Medium-volume e-mail correspondence.
4	Traditional Enhanced Course Presentation—Complete Web Content and Materials	All Level 3 Course presentations and lectures dynamically available on Web. Data files, links, programs on Web for students. Forms for student “reply” assignments, course evaluations, etc. Link to course grades.	2-3 times per week. Regular updating of grades. Medium-volume e-mail correspondence.
5	Web-Delivered Course with Orientation and Testing Meetings	All Level 4 plus any additional materials to allow for full Web delivery of course including audio and video augmentation; multimedia CDs. Few or no regular classes—orientation meeting may be necessary. Testing may be proctored off-site or unproctored on the Web.	Daily maintenance and access by instructor. High-level e-mail correspondence. Regular updating of grades and course materials.
6	Virtual Class	All Level 5 plus online testing and orientation. Discussion, chat groups, listserv, e-mail, and other interactive tools. Teleconferencing. No class meetings.	Substantial daily maintenance (average 1-3 hours) by instructor including all course aspects. High-level e-mail correspondence.

istrative factors. Levels 5 and 6 require significant changes in the academic structure and considerable support of the academic computing environment. The following table summarizes the resources that would be involved in the process of moving courses to the Web.

Although Table 2 shows a summary of the typical resources faculty need to develop Web course materials at varying levels, there are other elements that will be just as important in achieving a specific level of Web course expertise. The following list defines show of the issues. For a more comprehensive discussion, see Nantz and Lundgren (2003).

Issues Inhibiting Web-Enhanced Courses and Recommendations

- Be realistic about your own level of expertise and the instructional support you have available. Convert print-based materials to HTML using Word or some other familiar software. Once a comfort level is achieved, incorporate other HTML code using simple programs like Netscape Composer. Cut and paste code from sample Web pages. Extend knowledge to knowledge of common gateway interface

Table 2. Resources involved in moving to the Web

Level	Description	Resources Needed
1	Traditional Course Presentation, Basic-Level Course Materials on Web—Internal Links	Basic computer literacy, Web browsing experience. Course site can be created by the faculty member, professional designers, by use of Web course applications such as Blackboard (2001) or WebCT (2001).
2	Traditional Course Presentation—Intermediate-Level Course Materials on Web—External Links	Experience with preceding level. Web application packages can be extended or with additional training, a general Web development package like MS FrontPage or DreamWeaver can be used.
3	Traditional Enhanced Course Presentation—Intermediate-Level Course Materials on Web and Web Content Delivery	Experience with preceding level. Commitment to regular maintenance. Knowledge of e-mail attachments, listserv maintenance, or other interactive Web applications. Both Web application and general Web development packages can be extended for this level.
4	Traditional Enhanced Course Presentation—Complete Web Content and Materials	Experience with preceding level. Professional Web applications may not be able to accommodate this level without considerable difficulty. Usually requires considerable expertise with general Web development packages and some knowledge of HTML, and programming concepts include Javascript, ASP, and XML.
5	Web-Delivered Course with Orientation and Testing Meetings	All of the above. No additional faculty resources required; academic structural change to allow for registration and other student activities online.
6	Virtual Class	Use of a sophisticated commercial Web course package that allows for secure online testing; considerable administrative support and faculty expertise in the selected package.

(CGI) scripts, Java, or XML (Extensible Markup Language). Use university's instructional support personnel to help you set up simple Web pages that are at your comfort level for maintenance.

- Be realistic about the cost in time and money to maintain course Web sites. The development of a full Web-delivered course may be as high as $115,000 (Navarro, 2000). Marchese (1998) suggests a range of $12,000 to 90,000 per credit hour. Any Web platform provider will charge a licensing fee that may be based on the number of students. The equivalent of a one-hour lecture may require 24 hours for writing, recording, and editing, and up to 162 hours for full multi-media support.
- Be realistic about your access and the students' access to technology. Any administration who wants Web-delivered coursework must provide adequate technology to support it, either through on-campus servers or an off-campus Web host.
- Be realistic about converting paper-based content to Web content. A direct conversion usually doesn't work well. The visual indicators on printed materials (headers, footers, page numbering) don't convert well to HTML. Content and access must be re-evaluated. PowerPoint slides can be posted, but good slide design means key points only. Substantial notes must be provided to expand on the slides. Some students see the notes as ancillary and don't get the depth of content. Text-based sites are seen as boring. An academic course site that simply creates text on a Web page defeats the purpose of using Web pages—of having the ability to create links to interesting sites, to provide graphics, to provide sound and video. Providing hundreds of pages of text is the death knell for a Web class.
- Be realistic about your expectations of a Web-enhanced course. Our research shows that when course content is placed on the Web, student attendance will drop by as much as 50%. Students see printing Web pages as a substitute for class attendance (Lundgren & Lundgren, 1996).
- Be realistic about the stability of the technology. Over 90% of instructors report frequent problems (Navarro, 2000). You can't expect to use transparencies as a backup to an interactive lecture with dynamic linking. Especially with the vulnerability of the Internet and servers to virus and worm threats, there will be times when the Internet is down or so slow that interactivity isn't possible. Plan for technical problems and have a contingency plan that is communicated to students.

- Be realistic about the reward system for incorporating technology. If you are at an institution where research is valued more than teaching, then you may need to forego creating Web content. Navarro (2000) also notes that "cyberprofs" are reporting strong negative reactions from their colleagues. If you are sitting in your office answering student e-mail or creating course content, you don't appear to be teaching.
- Make sure intellectual property and royalty procedures are clearly spelled out. Many faculty do not consider the issue of copyright and intellectual property when course materials are developed (Rueter, 2001). Most faculty believe they own their own course materials. This is often not the case. The issue becomes even muddier when entire courses are delivered on the Web (Levels 5 and 6). A course that you developed could be offered by the university with someone else teaching it, without your consent or knowledge. Earnings from distance learning are viewed quite differently by faculty and administration (Guernsey & Young, 2001).
- Make sure you and your administration agree what "quality" teaching is. Age-old tenets of quality teaching include meaningful discussion, question-and-answer discourse, and significant teacher-student interaction. If faculty develops a Web-enhanced course following the myth of preserving student interaction, they will be quickly mired in Web activities that consume the majority of their time with no observable educational payoff. The less subtle problems that stem from a lack of administrative understanding include difficulties in obtaining resources, especially release time for the initial development of a Web class, lack of understanding about how many hours are needed to run the course when you aren't standing in front of a classroom, e-mail overhead, managing listservs, and problems in obtaining reasonable hardware and software to develop and maintain a Web course.

FUTURE TRENDS AND CONCLUSIONS

As younger, more computer-literate faculty emerges, there will be a slow and steady move toward the incorporation of Internet components into university courses. Many educational pundits believe that we are moving into a new learning paradigm with the integration of technology into our schools (Von Holzen, 2000). This new educational model envisions a complete shift in course delivery from the traditional lecture classroom to on-demand, flexible learning through the use of telecommunications technology or "Just-in-Time" learning. In this paradigm, the faculty will become the designers of interactive course materials. With this new paradigm, many of the issues discussed in this chapter may take care of themselves. In the meantime, faculty who are considering Web-enhanced or Web-delivered courses need to be aware of the issues. But, most of all, learning is more than just content delivery; we need to create learning environments whether they are in the classroom or in cyberspace.

ACKNOWLEDGMENT

This article is based on work originally published by the author and Terry L. Lundgren.

REFERENCES

American Association of University Professors. *Report on distance learning.* Retrieved January 26, 2001, from www.aaup.org/dirptext.htm.

Blackboard. Background retrieved January 18, 2001, from *www.blackboard.com*.

Cooper, L. (1999). Anatomy of an online course. *T.H.E. Journal, 26*(7), 49-51.

E-learning spending and enrollment to grow steadily. (2001). *College Planning and Management, 4*(2), 10.

Frizler, K. (1999). Designing successful Internet assignments. *Syllabus, 12*(6), 52-53.

Garrett, N.A., Lundgren, T. & Nantz, K.S. (2000). Faculty course use of the Internet. *Journal of Computer Information Systems*, (Fall).

Gellner, E. *Java's cyber espresso bar*. Retrieved November 6, 2000, from *www.javas.com/home.asp*.

Guernsey, L. & Young, J.R. *Professors and universities anticipate disputes over the earnings from distance learning*. Retrieved January 24, 2001, from www.chronicle.com/colloquy/98/ownership/background.shtml.

Hayen, R.L., Holmes, M.C. & Cappel, J.J. (1999). Enhancing a graduate MIS course using a discussion web. *Journal of Computer Information Systems, 39*(3), 49-56.

Institute for Higher Education. (1999). *Distance learning in higher education.* Washington, DC.

Kahn, R.L. (1997). Creation and maintenance of a syllabi Web site: A case study. Supplement to *T.H.E. Journal:*

The Internet in Education, (May), 23-26. Retrieved from *www.arsc.sunyit.edu/~com400/res1.html.*

Kubala, T. (1998). Addressing student needs: Teaching on the Internet. *T.H.E. Journal, 25*(8), 71-74.

Lundgren, T., Garrett, N.A. & Lundgren, C. (1999-2000). Student attitudes toward Internet course components. *Journal of Computer Information Systems, 40*(2), 64-68.

Lundgren, T. & Lundgren, C. (1996). College student absenteeism. *Proceedings of the 1996 Delta Pi Epsilon National Research Conference* (pp. 71-78), Little Rock, Arkansas.

Mallard. *Asynchronous learning on the Web.* Retrieved June 18, 2001, from *www.ews.uiuc.edu/Mallard/.*

Marchese, T. (1998). Not-so-distant competitors: How new providers are remaking the postsecondary marketplace. *AAHE Bulletin,* (May).

Mesher, D. (1999). Designing interactivities for Internet learning. *Syllabus, 12*(7), 16-120.

McGraw Hill's PageOut. Retrieved June 11, 2001, from *www.pageout.net*.

Nantz, K.S. & Lundgren, T. (1998). Lecturing with technology. *College Teaching, 46*(2), 53-56.

Navarro, P. (2000). Economics in the classroom. *Journal of Economic Perspectives, 14*(2), 119-132. Retrieved March 1, 2001, from *www.powerofeconomics.com/I.2-Research-Studies.htm*.

Rao, P.V. & Rao, L.M. (1999). Strategies that support instructional technology. *Syllabus, 12*(7), 22-24.

Resmer, M. (1999). IMS: Setting the course for distributed learning. *Syllabus, 12*(7), 10-14.

Rueter, J. *Modular courses and intellectual property rights.* Retrieved January 24, 2001, from *Web.pdx.edu/~rueterj/rlw/modular.htm*.

Rups, P. (1999). Training instructors in new technologies. *T.H.E. Journal, 26*(8), 67-69.

Stevenson, N. (2000). *Distance learning online for dummies.* New York: IDG Books Worldwide (now Hungry Minds).

The Western Mail. (2000). Log on as you fill up at the petrol station. (October 21), 25, Cardiff, Wales.

TeleCampus. *Specializing in online learning.* Retrieved April 2000 from *telecampus.edu.*

Thompson, D. *Intellectual property meets information technology*. Retrieved January 24, 2001, from www.educause.edu/ir/library/html/erm99022.htm.

Thornton, C. (1999). Back to school, Web-style. *PC World, 17*(7), 39-40.

Twigg, C.A. *Who owns online courses and course materials? Intellectual property policies for a new learning environment*. Retrieved January 24, 2001, from www.center.rpi.edu/PewSym/mon02.html.

Von Holzen, R. (2000). A look at the future of higher education. *Syllabus 14*(4), 54-57, 65.

WebCT. *The e-learning link*. Retrieved June 18, 2001, from *www.Webct.com.*

World Lecture Hall. *Academic computing and instructional technology services.* Retrieved April 2000 from *www.utexas.edu/world/lecture.*

Yahoo. *Distance learning.* Retrieved February 2000 from www.yahoo.com/education/distance_learning.

KEY TERMS

Web-Based Course: A course that is delivered entirely by electronic methods, such as the Internet.

Web-Enhanced Course: A traditional course with some electronic enhancements, such as Web pages for course syllabi, data files, and test reviews.

Issues of E-Learning in Third World Countries

Shantha Fernando
University of Moratuwa, Sri Lanka

INTRODUCTION

Around the world, e-learning is becoming popular, especially among higher education institutes (universities). Many highly ranked universities have either already deployed an e-learning system and are fully operational, or they are in a process of deployment where e-learning-based and non e-learning-based educational environments co-exist. It is also possible to find a few virtual universities. The amount of money and effort that has to be spent on e-learning is high. In addition to the initial e-learning system installation costs, there are ongoing maintenance, management and content development costs. Due to the rapid growth in the field of e-learning and the role it plays in today's education systems, those working in the field have begun to introduce standards for different aspects of e-learning. The Open Knowledge Initiative (OKI) which is described as "a collaboration among leading universities and specification and standards organizations to support innovative learning technology in higher education" is an example (OKI, 2003).

Many highly ranked universities use commercial e-learning systems such as BlackBoard, WebCT, e-college, Netschool, etc. Several open source products are available though their usage is not wide spread, although it is expected that collaborative projects such as Sakai will enable large-scale open source products to be introduced to the market. This effort is described on the Sakai website as, "The University of Michigan, Indiana University, MIT, Stanford, the uPortal Consortium, and the Open Knowledge Initiative (OKI) are joining forces to integrate and synchronize their considerable educational software into a modular, pre-integrated collection of open source tools" (OKI, 2003).

BACKGROUND

Many third world countries have become "Transitional Countries". The term "transitional country" has been used in different ways in different times and different contexts. However, today's meaning of a "transitional country" is a country that lies between a developed and a developing country, and has an evolving market economy. Dung (2003) states:

Generally speaking, the expression 'transition' is used, mainly by political scientists, in the context of changes that have followed the fall of regimes, usually when dictatorial regimes have given way to more democratic ones, but this usage has been extended to contexts where previously rigid structures, such as those governing the economy, are giving way to more liberal, market-friendly structures and associated features of liberal democracy.

Third world or transitional countries require sustainable development. Sustainable development of a country is very much dependent on industry, higher education and research, hence university education is vital. The importance of the higher education is stressed in the United Nations Resolution on the Decade of Education For Sustainable Development January 2005 – December 2014 (UN Report, 2002). For a third world country, as De Rebello (2003) puts it, "The university system was seen as being uniquely equipped to lead the way by their special mission in teaching and training the leaders of tomorrow, their experience in transdisciplinary research and by their fundamental nature as engines of knowledge."

CURRENT TRENDS IN INFORMATION TECHNOLOGY IN THIRD WORLD COUNTRIES

IT is becoming a driving force of economy. Realizing its potential, many transitional countries have embarked on projects in collaboration with funding agencies to improve IT services, though their IT infrastructure facilities are not adequate. Many foreign investors start IT based companies in transitional countries. The products are aimed at the US or European market, where the parent companies are based. India, in particular, exemplifies this for the IT sector, and many major IT companies have branches in India. In Sri Lanka, due to the limited market, poor infrastructure and slightly higher labor costs, such foreign investments are limited. However, the level of IT expertise is at a competitive level. Many local IT companies carryout sub-contracts for foreign IT companies. A few companies directly interact with the global market. Realizing the potential, the Sri Lankan government embarked on "e-Sri Lanka move" project to introduce e-governance and to improve e-services within the country, and formed the ICT Agency using World Bank funds (Development Gateway, 2003). Motivated by these initiatives and realizing the importance of e-learning for today's form of higher education, some Sri Lankan universities have deployed e-learning systems as pilot projects and a few others have started exploring the possibility of using e-learning for their university education.

Due to the employment opportunities offered for IT professionals of transitional countries by developed countries, many professional IT programs have been initiated in transitional countries. In Sri Lanka, income generated by foreign employment has now become considerable compared to its other income sources such as garment, tea, rubber, minerals, spices, etc. Though most employment opportunities are labor-oriented, many professional opportunities are in the IT sector. However, this causes "brain drain".

IMPORTANCE OF E-LEARNING FOR HIGHER EDUCATION IN THIRD WORLD COUNTRIES

In order to understand the importance of e-learning, it is important to consider what we mean by e-learning. According to the definition of NCSA's e-learning group (Wentling, T.L. et al., 2000):

E-learning is the acquisition and use of knowledge distributed and facilitated primarily by electronic means. This form of learning currently depends on networks and computers but will likely evolve into systems consisting of a variety of channels (e.g., wireless, satellite), and technologies (e.g., cellular phones, PDA's) as they are developed and adopted. E-learning can take the form of courses as well as modules and smaller learning objects. E-learning may incorporate synchronous or asynchronous access and may be distributed geographically with varied limits of time.

In an abstract form, I would define it as "electronically facilitated, enhanced and managed learning". It can consist of many components or elements of a learning environment of a university system if they can be electronically facilitated, enhanced and managed. Some aspects that could be integrated into an e-learning system to make an impact in a university system, especially in the context of a third world country, are given below.

- Curriculum related aspects – courses and course contents, discussions, library catalogues, etc.
- Academic administration related aspects – registrations, student information, grading, etc.
- Technology infrastructure related aspects – alternative technologies, lab facilities, home use, etc.
- Societal context related aspects – cultural events, forums, activities, etc.
- Industrial collaboration related aspects – industrial expertise and contents, know-how dissemination, guidance to/from industry, etc.

These aspects, when incorporated in an e-learning system, will improve the quality of the higher education, if implemented using strategies and technologies suitable for constrained environments in third world countries. However, deployment of a suitable e-learning system requires a particular educational, administrative and technological environment, and the university educational system will also need to undergo changes. This is where the issues are faced in third world countries. One should not think that the deployment of e-learning is an adaptation to the required educational change. Contrarily, an ability to adapt is a must for the deployment of e-learning.

Bates (2000) states that higher education institutes consider technology-based learning for the following reasons:

- the need to do more with less
- the changing learning needs of society
- the impact of new technologies on teaching and learning (Bates, 2000, p. 8).

Although we observe that mainly the universities in developed countries tend to consider the above reasons, they are applicable to any university. It is in this context that e-learning is becoming attractive. However, when universities in third world countries embark on e-learning-based educational transformations, they face many barriers. In many cases, e-learning cannot be implemented in the way it is done at US or European universities. The approach has to be tailored to the environment, if it is to be a success.

COMMON ISSUES TO BE ADDRESSED

Administrative Issues

Most of the universities in third world countries are traditional universities. Gunn (2000) in his keynote paper states the following:

Perhaps the most critical challenge to traditional universities is develop capacity to change. This calls for major restructuring, removal of unnecessary processes and streamlined administration procedures. Motivation to progress, change and develop is hard found in the current insecure climate. . . The challenge this raises is being able to exploit the resources of commercial interests while maintaining quality and standards of service as a priority area. Ability to achieve the right balance between opposing forces of cost and quality without reducing education to the lowest common factor will be a powerful survival strategy.

In many third world countries university academic administration is stream-lined and rigid. Changes are usually not welcomed. Many fear loosing the value of their jobs if IT strategies are introduced. Many administrative officers have the mentality that the others should come to them to get the work done. While this shows an attitude problem or an inferiority complex, it affects many productive plans.

However, rigid administrative procedures are sometimes required to prevent exploitation and use of facilities for personal advantage.

Some administrative functions can be handled efficiently through e-learning. Typical examples would be student semester and exam registrations, yearly progress archiving, student information management, etc. However, administrative officers such as registrars, examination branch officers, etc, are not comfortable when it is handled entirely by the e-learning system. There is the fear they might loose their job. Another fear is whether they will have any value for the university. A valid concern that is raised is whether the e-learning system is secure enough to protect confidential data and prevent students tampering with data.

IT Infrastructure Issues

IT infrastructure facilities in third world countries are often primitive. While IT infrastructure needs improvement for better interconnectivity of academic institutes, a countryman's concern is food, water supply, clothing, roads and transportation, housing, primary schools, and other essential items for their living. Governments in these countries have to allocate the majority of their funds for the latter and a low priority is given for IT infrastructure. It is not justifiable to allocate huge funds for the improvement of IT infrastructure when the basic needs of the people are not met. The good news is that some form of infrastructure is already available. The solution we propose for the improvement of higher education using e-learning has to consider alternative techniques given this serious limitation. This is not to say that mobile communications and other new inventions are not penetrating the market.

Consider Sri Lanka as a case, every university is interconnected by a university network called LEARN (Lanka Academic and Research Network). Some universities have 2 Mbps E1 links, while the rest have only 128 or 64 kbps links. Very soon the latter will be upgraded, but the maximum would be 2 Mbps in the foreseeable future. The current international bandwidth allocated for the whole university network is below 2 Mbps. This will gradually increase on demand, but on-demand increase implies the presence of congestion. The universities also experience disruption of the telecom services, either due to faults or non-payment of bills. However, within these infrastructure constraints, the majority of universities are able to have an acceptable level of communication for the current IT operations within the country. Web servers are acceptably fast and e-mail is heavily used for communication and collaboration among academics. A few e-learning systems are also operational.

Any e-learning-based solution has to work within these IT infrastructure constraints. Within a university it will work acceptably since many universities have local area networks with either gigabit fibre optics, or fast Ethernet or at least 10 Mbps links. Between universities it will work as long as it does not have heavy content delivery, congesting the links. However, international collaborations through e-learning will not be at the levels required by many e-learning systems in the near future.

Limitation of Equipment

In third world countries, equipment such as servers, routers, cabling, laboratory computers, etc. are usually procured under special university budgets, or grants and loans from funding agencies. It is not possible to expect frequent upgrades to equipment. It is very unlikely that high-end servers with redundant power supplies and disk arrays will be always available for the deployment of an e-learning system and redundancy and backup systems are not a priority. Sometimes valuable information stored in the system may be at stake. However alternative approaches such as weekly or critical time-based backups may be carried out.

Thus, any approach to introducing e-learning has to start with a low-end solution. Once the importance is recognized by the authorities, some form of ongoing support is feasible. Strategic planning is required to get the funds for improving the performance and reliability of the systems gradually.

It is not possible to assume that students will always have access to computers. While a few have their own computers, the majority of the students in transitional countries use common lab facilities to access computers. Labs are open only during working hours and usually scheduled for different groups of students based on assignments and workloads. In most cases, e-learning-based learning activities also need to be planned accordingly. For an example, if an assignment is given with a deadline for the submission through the e-learning system, this deadline has to be flexible in situations such as insufficient computers, labs being not open on demand, workers' strikes which are frequent in many third world countries, long electricity power cuts, etc.

Cost Factors of E-Learning Systems

Most of the commercially developed e-learning systems such as BlackBoard, WebCT, etc, used by US and European universities are extremely expensive for the third world countries to purchase. A monetary grant may be a possibility, but then the question would be maintenance costs, purchase of additional modules to suite the changes as time passes, costs of customizations, etc., if these costs have to be born by the university, which are very high given the limited budgets. Therefore, any grant must include these costs. Otherwise, it will be a waste of funds.

An alternative is to select an open source solution. However, currently it is difficult to find the exact match of an open source solution, or to customize it to a particular university's environment. Projects such as Sakai may help solve this situation in the future, but we have to wait until their collaborative environment is functional. However, there is the concern whether it also will assume state-of-the-art technology infrastructure.

Another alternative is in-house development, however, for this to be a success, continuous employment of developers and good software development approaches with research input from an e-learning perspective is required. Finding developers is not a difficulty in most transitional countries, and the costs for this will be far below the purchase of a commercial e-learning system. To succeed, however, a vision to continue the project, and institutionalized incentives to the people involved, should be in place. While the result of this approach may not be as sophisticated as, or as reliable as, available commercial systems, it is possible to come up with an acceptable solution at a very low cost. In the author's environment, it was possible to get a group of students to start on the development of an e-learning system using research findings. Later, an expert was used to further improve it to be used as a production system. It needs further development, but the advantage is, while the required institutional changes for an e-learning-based education are conveyed to the rest of the faculty, the changes can also be synchronized with the development cycle, as illustrated by Collis and Moonen (2001). Even if a fully fledged e-learning system had been purchased, it would have been a failure due to the faculty being not ready to adapt immediately.

Reliability Issues

Reliability issues have already been mentioned under IT infrastructure issues and limitation of equipment. The following summary is provided to emphasize the issue of reliability.

- Frequent electricity power failures.
- Data communication connectivity failures or disruptions due to non payment of bills.
- Congested links.
- Less emphasis on backup and redundant systems.

Socio-Cultural Issues

In most of the third world countries, especially in South Asian and African continents, socio-cultural setting is very prominent. It affects how people engage in learning activities. Verbal and physical interactions are important and hence total virtual learning environments may not produce good results. This situation may change in the years to come, especially among the urban population. However, socio-cultural aspects cannot be neglected when dealing with education, and it is true also for technology-based education, as described by Gunawardena (1998).

Most of the e-learning systems and available contents are based on popular languages. However, this is not to say that they do not support other languages, but it will require an additional effort to prepare contents in native languages. In many third world countries primary education is done in native languages, although at university level popular languages like English or Spanish may be the medium. This situation can create communication barriers in e-learning-based learning processes.

Many people in third world countries believe that developments in IT will cause many people to lose their jobs. This is a serious social issue. However, there are situations where it is thought to be the other way round. For an example, in e-Sri Lanka move, the government expects that there will be an increase in job opportunities if IT is promoted. For an example, to deploy e-learning in a university environment, additional support staff is required for facilitation, content creation, maintenance, etc.

FUTURE TRENDS AND CONCLUSION

E-learning can play a major role in higher education in third world and transitional countries. It will help improve the higher education, thereby contributing to sustainable development. Using e-learning it is possible to improve curriculum, academic administration, industry collaboration, etc.

Emerging related standards such as Sharable Content Object Reference Model (SCORM, 2003), IEEE Learning Technology Standards Committee (LTSC, 2002) and collaborative work currently being carried out such as OKI (OKI, 2003) will make e-learning more widespread.

However it may not be possible to deploy it in third world countries in the way it is done in the highly ranked universities in the US and European countries. First, the related issues have to be addressed and alternative solutions should be explored. Given suitable alternative solutions, or desirable approaches, e-learning can be a success in many third world and transitional countries.

REFERENCES

Bates, A. W. (2000). *Managing technological change: Strategies for college and university leaders*. San Francisco: Jossey-Bass Publishers.

Collis, B., & Moonen, J. (2001). *Flexible learning in a digital world: Experiences and expectations*. UK: Kogan Page.

De Rebello, D. (2003). What is the role for higher education institutions in the UN decade of education for sustainable development?, *Theme IV, International Conference on Education for a Sustainable Future* (pp. 10-11). Prague, Czech Republic: Charles University, Karolinum.

Development Gateway. (2004). *e-Sri Lanka: Transforming government, business and society* (December 29, 2003). Retrieved March 01, 2004, from *http://www.developmentgateway.com/node/133831/sdm/docview?docid=841120*

Dung, L. T. (2003). Judicial independence in transitional countries, *The Democratic Governance Fellowship Program, United Nations Development Program, Oslo Governance Centre, January 2003* (page 5). [Electronic version]retrieved March 02, 2004, from *http://www.undp.org/oslocentre/docsjuly03/DungTienLuu-v2.pdf*

Gunawardena, C. (1998). Designing collaborative learning environments mediated by computer conferencing: Issues and challenges in the Asian socio-cultural context. *Indian Journal of Open Learning*, *7*(1), 101-119.

Gunn, C. (2000, Decmber). *Identity, control and changing reality*. Keynote paper at ASCILTE Conference, Coffs Harbour. [Electronic version] retrieved June 25, 2003, from *http://www.ascilite.org.au/conferences/coffs00/papers/cathy_gunn_keynote.pdf*

LTSC. (2002). *Learning object metadata*. Learning Object Metadata Working Group, Learning Technology Standards Committee (LTSC), IEEE. Retrieved March 24, 2004, from *http://ltsc.ieee.org/wg12/index.html*

OKI Project. (2003). *Open knowledge initiative*. Retrieved March 02, 2004, from OKI Project web site *http://web.mit.edu/oki/*

Sakai Project. (2003). Retrieved March 02, 2004, from Sakai Project Web site *http://www.sakaiproject.org/*

SCORM. (2003). SCORM overview. Advanced Distributed Learning. Retrieved May 05, 2004, from *http://www.adlnet.org/index.cfm?fuseaction=scormabt*

UN Report. (2002). *World summit on sustainable development: Plan of implementation* (para 117d). Retrieved March 01, 2004, from *http://www.johannesburgsummit.org/html/documents/summit_docs/2309_planfinal.htm*

Wentling, T.L., Waight, C., Gallaher, J., La Fleur, J., Wang, C., & Kanfer, A. (2000, September), *E-learning – A review of literature*. Knowledge and Learning Systems Group, University of Illinois at Urbana-Champaign, NCSA.

KEY TERMS

Academic Administration: Administration procedures or formalities linked with university education, such as registrations for semesters or examinations, progress reviews and monitoring, eligibility formalities, student history records or progress archiving, promotions to levels or years, academic timetables, etc.

E-learning: Electronically facilitated, enhanced and managed learning.

IT Infrastructure: Technological infrastructure that enables the transfer of information.

Learning Environment: Overall university setting in which many educational and administrative processes interact.

Open Source E-learning Systems: E-learning systems developed by the Open Source Community and freely distributed with their own license or a GPL (General Purpose License) to use, modify and distribute together with the source code.

Third World Countries: Countries that are not yet developed.

Transitional Countries: A third world country that is in a transition process based on more liberal, market-friendly structures and associated features of liberal democracy.

Virtual Universities: All the learning and administration activities are done through e-learning and very minimum physical interactions, or no physical interactions at all.

IT Implementation in Small Business

Elaine R. Winston
Hofstra University, USA

Dorothy G. Dologite
Baruch College, City University of New York, USA

INTRODUCTION

The small business owner's dominating role in any information technology (IT) implementation is well documented (Cragg & King, 1993; Doukidis, Smithson, & Lybereas, 1994; Thong & Yap, 1995). A survey of the IT implementation process in small businesses indicates that both the owner's attitude towards IT and management style influence the implementation process and have a direct impact on outcomes (Winston & Dologite, 1999). These characteristics, attitude and style play a key role in directing the small business owner's decisions regarding the technology to implement, the personnel that will manage the implementation, and the supporting managerial polices (Winston & Dologite, 2003).

A more practical and pressing reason for concern with the small business owner's attitude towards IT and management style comes from the Internet. Innovative owners, who are early IT adopters, seize new opportunities that the Internet offers by reaching and selling to new markets (Pratt, 2002). More cautious managers will wait for the pioneers to reorganize as e-businesses and then attempt to learn from them (Pratt, 2002; Winston & Dologite, 2002). Other small business owners will completely avoid investing in Internet technologies (Winston & Dologite, 2002). It is advantageous for the IT practitioner to recognize how the small business owner's attitude and management style affect the adoption of IT and quality of the implementation.

BACKGROUND

Quality of IT Implementation

At an operational level, the results of a high-quality IT implementation process include increased internal organizational performance and efficiency, better customer service, and a higher quality working life within the organization (Doukidis et al., 1994; Julien, 1995). Studies of IT implementations have identified specific outcomes of success and find such indicators as increased profits, greater market share, or improved return on investment (ROI) performance (Byrd, Sambamurthy, & Zmud, 1995; Iivari & Ervasti, 1994). For small organizations, it is the added considerations of a high quality IT implementation that lead to increased profits, greater market share, or improved ROI performance.

Attitude Towards IT

Generally, small business owners with positive attitudes towards IT often have a deep knowledge about IT and find it easy to commit to an IT implementation (Cragg & King, 1993; Nickell & Seado, 1986; Ray, Thomas, & Dye, 1994; Winston & Dologite, 1999). More specifically, a positive attitude will likely help the owner endure the frustration and effort of overcoming the technical and organizational obstacles that frequently confront innovative initiatives (Harrison, Mykytyn, & Riememschneider, 1997; Winston & Dologite, 2002).

Conversely, small business owners with negative attitudes towards IT implementation generally avoid investing in IT because they fail to understand the strategic impact of IT on their business (Cragg & King, 1993; DeYoung & Spence, 2004). Only when faced with the threat of closing do these entrepreneurs seek the short-term solution of implementing IT for survival (Agarwal & Prasad, 1997). Some small business owners who perceive IT implementation as risky become immobilized by indecision and hesitation (Beckers & Schmidt, 2001). Even after approving an investment, owners with a negative or uncertain attitude towards IT often have difficulty sustaining commitment through an implementation. This results in reducing resources allocated to a project, inhibiting organizational changes required to achieve benefits, and preventing the use of IT to its fullest potential.

Management Style

Management style influences various aspects of the implementation process, such as IT strategy, owner involvement, and timing (Winston & Dologite, 2002). Yet, IT studies often do not explore the owner's management style. Perhaps this is because of the confusion that style is connected to personality, and therefore not likely to change. Management style, however, is demonstrated by work role and by patterns of actions (Waterman, Peters, & Phillips, 1995). A small business owner's management style, then, is essentially manageable. Management style may be generally classified as either entrepreneurial or traditional (Jelinek & Litterer, 1995; Senge, 1990; Waterman et al., 1995).

An entrepreneurial policy enhances the capacity of a business to address and reduce any risk and uncertainty connected with a new IT implementation (Jelinek & Litterer, 1995). Employees tend to rely on one another to make the IT implementation successful and to keep others informed of its progress (Cragg & King, 1993; Winston & Dologite, 2003). Procedures that allow employees to have their views considered greatly affect the support and commitment a work force has to an IT implementation (Senge, 1990; Way, 2002). In other words, in organizations with a shared management style, employees accept responsibility for guiding the affairs of the organization.

The small business owner who practices a traditional style typically attempts to reduce the uncertainty surrounding the use of IT by maintaining policies that support bureaucracy, identifying clear boundaries of responsibility, and retaining control of the problem as well as its solution (Herbert & Bradley, 1993; Thong & Yap, 1995). The basis of the traditional management style is to establish order, exercise control and achieve efficiency in the work force. Policies may not exist or be clearly defined for hearing end-users' opinions (Jelinek & Litterer, 1995).

The entrepreneurial and traditional management styles provide small business owners with an approach to actively deal with the complexity of an IT implementation. Furthermore, the management style that is employed may support or moderate an owner's attitude towards IT (Winston & Dologite, 2002).

ATTITUDE/STYLE PROFILES

This section discusses the emergence of distinct attitude/style profiles, which are Positive Entrepreneur (PE), Negative Entrepreneur (NE) and Uncertain-Traditionalist (UT). In a study conducted by the authors, support for classifying small business owners into these three profiles was provided by a consistent pattern of data that describes attitude towards IT and management style (Winston & Dologite, 2002). Alternative attitude/style profiles, such as positive-traditionalist, negative-traditionalist, or uncertain-entrepreneur, were not identified in this study.

Table 1 lists the data clusters and sample supporting statements for the three identified attitude/style profiles. To further demonstrate the profiles, more detailed descriptions and quotes from business owners follow.

Positive-Entrepreneur (PE)

A small business owner classified as a positive-entrepreneur (PE) has a positive attitude towards IT and actively seeks applications of IT to enhance his or her business.

Table 1. Attitude/style profiles of small business owners towards IT implementation

	Positive-Entrepreneur (PE)	**Negative-Entrepreneur (NE)**	**Uncertain-Traditionalist (UT)**
Attitude	**Positive:** Expresses a positive perception about IT, identifies opportunities for innovative use of new IT.	**Negative:** Expresses a negative perception about IT, hesitates to invest in IT, implements IT out of necessity.	**Uncertain:** Expresses indecision about IT, hesitates to invest in IT, is anxious about using IT.
Management Style	**Entrepreneurial:** Identifies the use of policies that encourage collaboration and shared management.	**Entrepreneurial:** Identifies the use of policies that encourage collaboration and shared management.	**Traditional:** Uses policies that support power, authority, and status; demonstrates an inability to adopt tactics to decipher overwhelming data and IT alternatives, as well as conflicting suggestions about them.

The following statements exemplify the proactive strategies of PE owners (study participant numbers appear in parenthesis):

I'm fascinated at how things change because of IT and often investigate how change can impact my business. (PE1)

We put in the latest IT to remain competitive. We know the technology because we develop software for it. (PE3)

It gives me the image that I've got a huge organization behind me. I learned about the IT that was available and I knew exactly what I wanted from it in terms of my business. (PE4)

In addition, PE owners exhibit an entrepreneurial management style by encouraging their employees to become involved in areas well beyond their normal scope of responsibility. They encourage employees to consider such fundamental IT implementation issues as hardware and software purchases and training.

Employees tell us how to do things more effectively. (PE3)

We provide lots of employee training and even let the employees decide when they could use a computer specialist for additional help. (PE1)

I have employees I can trust to make decisions in my business. Even though I am stuck with a key person who limits the amount of technology I can use, what I do have I used—and it made me in the hundred thousands. (PE4)

Negative Entrepreneur (NE)

Small business owners classified as negative entrepreneurs (NE) reflect negative attitudes towards IT and invest in IT primarily to remain in business. As a group, they indicate that a barrier to investing in IT is their lack of knowledge. Eventually, they become like PE owners, when they take actions about IT that characterize an entrepreneurial style; however, they consistently report reactive rather than proactive strategies:

We are purchasing a major IT system that's costing over $100,000, and by the time we pay it off, we will need a new system. The cost of the IT [in relation to its] longevity is ludicrous—what we are buying now will be obsolete in five years, but we have no other choice. No one knew anything about enterprise software, only that we had to purchase a package just to remain competitive in business. (NE2)

Interestingly, all NE owners have developed practices that might assist them to overcome their negative attitudes towards IT. Displaying an entrepreneurial management style and the desire to remain competitive in the marketplace, NE owners rely on young, IT-current employees for IT recommendations. The following statements made by small business owners provide supporting evidence:

We use young employees right out of college to show us how to use some of the new IT required by law. They tell me what we need. (NE1)

Employee suggestions are tremendously important. We asked the employees what they wanted, and their suggestions were taken seriously. (NE2)

Uncertain Traditionalist (UT)

Small business owners classified as uncertain traditionalist (UT) have uncertain attitudes towards IT. They take actions about IT that characterize a traditional style of management. These owners state that they actively learn about IT, but they do not have the capability to interpret all of the information. Consequently, they rely on traditional practices, as evidenced in the following statements:

For us to put in a voice mail system would cost over $5,000 and it's not necessary. We like the personal touch that comes from a person actually picking up the telephone. (UT4)

UT owners are technologically aware; however, information overload appears to aggravate their uncertainty towards IT:

I spent a lot of time and got a lot of information on systems, but I still don't know much about any of them. (UT1)

A different perspective on the uncertain attitudes of small business owners towards IT emerged from case UT3, perhaps because it involved telephony technology implementation. It illustrates how conflict arising in the outside environment creates ambiguity, as illustrated by both incentive and deterrent:

Incentive: We are looking into telephony where we can track telephone calls in the system. When I get a call a list of the last five conversations ... will come up. I can get an idea of the last conversation the customer had with our company and say, "Yes, I know so and so, whom you spoke with on this date is doing this." (UT3)

Deterrent: The technology moves people beyond wanting information in one minute. Now they want it this second. They assume everything is online and that we have instant access to it. (UT3)

In the case of UT3, the customers and their satisfaction drive the business and supply both the incentive and the deterrent for IT implementation. The implementation that resulted, in this case, was of a low quality.

ATTITUDE AND MANAGEMENT STYLE ISSUES

The key issues focus on the interaction between different types of attitudes and management styles. The biggest surprise concerns the negative-entrepreneur type of small business owner. Although an NE owner invested in new IT primarily as a reactive strategy to avoid going out of business, he or she still experienced a high quality IT implementation. This is a revelation because previous literature suggests that owners with negative attitudes towards IT would most likely achieve poor quality IT implementations (Agarwal & Prasad, 1997).

Owners with negative attitudes towards IT may reduce the risk associated with new technology initiatives by using collaborative and shared management policies and practices in their organizations (Winston & Dologite, 2002). They not only depend on employee input, but they also acted on recommendations of employees. This practice was reported to be a significant influence when contending with changes that occurred as a result of an IT implementation.

Conversely, it appears that the uncertain traditionalist type of small business owner is unable to initiate a forward looking approach for guidance through an IT implementation process (Winston & Dologite, 2002). Based on previous research, we would expect that the uncertain owner would seek advice and information from employees, or other more technical sources, about which type of IT is an appropriate investment. It is more likely that uncertain-traditionalist owners became immobilized from an information overload in their pursuit of IT knowledge.

The uncertain-traditionalist owners' preference for a high level of control and bureaucracy inhibits innovation and theoretically reduces the chance for a quality IT implementation. Often, UT owners evaluate IT primarily on a financial basis. They easily calculate the cost of the IT as well as the additional employee time required to maintain it. However, it is difficult for a UT owner to calculate productivity increases that arise from more accurate reports or faster processing time. Such operational benefits of IT are elusive to them. Further, as a result of an overload of IT information, they do not wish to become heavily involved in the IT adoption and implementation process.

Generally, small business owners with a positive attitude toward IT, have a high-quality IT implementation. An interesting aspect surfaces: Positive attitude is coupled with an entrepreneurial management style. This type of owner reports a collection of proactive strategies, such as:

- Engaging the support and input of employees;
- Making a concerted effort to learn about IT;
- Becoming involved in the IT implementation process;
- Using IT to attain competitive advantage for the business; and
- Finding benefits from an IT implementation in operational areas much before they appear in financial areas.

Unlike the uncertain owner, the positive entrepreneur takes advantage of a deep knowledge of the business as well as knowledge of IT to resolve ambiguous issues that might otherwise impede the successful and creative use of IT in the organization (Winston & Dologite, 1999).

FUTURE TRENDS

Future trends concerning the implementation of IT in small business will likely incorporate knowledge management. Owners adapt IT information to their small business with the help of both internal and external collaborators. As the e-commerce environment becomes more complex, the different backgrounds of collaborator and owner may hinder their ability to transfer implicit knowledge of IT among themselves. Therefore, organizations will rely on knowledge management to identify the quality of the sources and type of IT information obtained by an owner. This information will help to reduce the uncertainty surrounding the use of IT in small business.

CONCLUSION

Early in an implementation process, an IT practitioner should recognize a small business owner's attitude/style profile and be aware of how the profile can influence implementation quality. The following suggestions may help IT practitioners improve the likelihood of achieving a high quality IT implementation.

- Consider that a small business owner with a negative attitude toward IT might improve the likelihood of achieving a high quality IT implementation by

Table 2. Summary

Small Business Owner		Quality of IT Implementation	Contributors of IT Implementation Quality
Attitude Towards IT	**Management Style**		
Positive	Entrepreneurial (open, collaborative)	High	Attitude and style
Negative	Entrepreneurial (open, collaborative)	High	Reliance on employees
Uncertain	Traditional (closed, hierarchical)	Low	Bureaucracy and feeling of being overwhelmed

practicing entrepreneurial management policies.

- Work to help change a small business owners' traditional management style, rather than try to change an owners' negative attitude towards IT, which is potentially a more complex issue to deal with.
- Encourage owners with traditional management styles to employ policies that require employee participation and collaboration in IT decisions.
- Help uncertain-traditional small business owners solicit support from employees in their business who understand the IT and how it can be applied to the business.

Table 2 displays an interesting relationship between attitude and management style when considering the achievement of a high or low quality IT implementation. Main contributors to IT implementation quality were also identified, and business owners were classified, as Positive-Entrepreneur, Negative-Entrepreneur, and Uncertain-Traditionalist. Only the Uncertain-Traditionalist experienced a low quality IT implementation.

The results concerning how attitudes towards IT and management style affect the quality of an IT implementation can guide the small business owner and information expert concerned with the design and plan of the implementation process. This also suggests that IS work should increasingly address combined managerial and technical issues. In a small business, the owner's feelings and organizational policies, rather than the technology, may be most important in determining an IT initiative's success.

REFERENCES

Agarwal, R., & Prasad, J. (1997). The role of innovation characteristics and perceived voluntariness in the acceptance of information technologies. *Decision Sciences, 28*(3), 557-582.

Beckers, J.J., & Schmidt, H.G. (2001). The structure of computer anxiety: A six factor model. *Computers in Human Behavior, 17*(1), 35-49.

Byrd, T.A., Sambamurthy, V., & Zmud, R.W. (1995). An examination of IT planning in a large diversified public organization. *Decision Sciences, 26*(1), 49-73.

Cragg, P.B., & King, M. (1993). Small firm computing: Motivators and inhibitors. *MIS Quarterly, 17*(1), 47-60.

DeYoung, C.G., & Spence, I. (2004). Profiling information technology users: En route to dynamic personalization. *Computers in Human Behavior, 20*(1), 55-65.

Doukidis, G.I., Smithson, S., & Lybereas, T. (1994). Trends in information technology in small businesses. *Journal of End User Computing, 6*(4), 15-25.

Harrison, D. A., Mykytyn, P.P., & Riememschneider, C.K. (1997). Executive decisions about adoption of information technology in small business: Theory and empirical tests. *Information Systems Research, 8*(2), 171-195.

Herbert, F.J., & Bradley, J.H. (1993). Expert systems development in small business: A managerial perspective. *Journal of Small Business Management, 31*(3), 23-34.

Iivari, J., & Ervasti, I. (1994). User information satisfaction: IS implementability and effectiveness. *Information and*

Management, *27*(5), 205-220.

Jelinek, M., & Litterer, J.A. (1995). Toward entrepreneurial organizations: Meeting ambiguity with engagement. *Entrepreneurship Theory and Practice, 19*(3), 137-168.

Julien, P. (1995). New technologies and technological information in small businesses. *Journal of Business Venturing*, *10*(6), 459-475.

Nickell, G.S., & Seado, P.C. (1986). The impact of attitudes and experience on small business computer use. *American Journal of Small Business*, *10*(4), 37-47.

Pratt, J.H. (2002). E-Biz: Strategies for small business success. Office of Advocacy, US Small Business Administration. Retrieved on February 15, 2004 from *http://www.sba.gov/advo/research/rs220tot.pdf*

Ray, C.M.H., Thomas, M., & Dye, J.L. (1994). Small business attitudes toward computers. *Journal of End User Computing, 6*(1), 16-25.

Senge, P.M. (1990). *The fifth discipline*. New York: Doubleday.

Thong, J.Y.L., & Yap, C.S. (1995). CEO characteristics, organizational characteristics and information technology adoption in small businesses. *Omega*, *23*(4), 429-442.

Waterman, Jr., R.H., Peters, T., & Phillips, J. (1995). The 7-S framework. In H. Mintzberg, J. Quinn, & J. Voyer (Eds.), *The strategy process,* New Jersey: Prentice Hall.

Way, S.A. (2002). High performance work systems and intermediate indicators of firm performance within the US small business sector. *Journal of Management, 28*(6), 765- 785.

Winston, E.R., & Dologite, D.G. (1999). Achieving IT infusion: A conceptual model for small businesses. *Information Resources Management Journal*, *12*(1), 26-38.

Winston, E.R., & Dologite, D.G. (2002). How does attitude impact IT implementation: A study of small business owners. *Journal of End User Computing, 14*(2), 16-28.

Winston, E.R., & Dologite, D.G. (2003). Attitude and management style matter in IT implementation: A study of small business owners. In M.A. Mahmood (Ed.), *Advanced topics in end user computing, volume 2* (pp.234-262). Hershey, PA: Idea Group Publishing.

KEY TERMS

Entrepreneurial Management Style: An approach that supports interaction and shared decision making with employees and the early adoption of IT. Employees are encouraged to embrace ambiguity as a source of opportunity, present new ideas, and initiate action in new directions.

Implementation Quality: The outcome of an IT installation in an organization, where operational success can be indicated by increased internal organizational performance and efficiency, better customer service, and higher quality working life within the organization, although other traditional indicators are increased profits, greater market share, and improved return on investment performance.

IT Implementation Process: All the activities involved in initiating IT in an organization, from the original inception of an idea to innovate or install IT, to evaluating the success or failure of the IT in an organization.

Management Style: Refers to the way the small business owner delegates authority and responds to new technologies and innovative practices. This is demonstrated by work role and by patterns of actions that influence various aspects of the IT implementation process.

Negative Attitude: A perception about IT that results in behavior that displays discomfort around IT, hesitation to invest in IT, and the implementation of IT out of necessity.

Positive Attitude: A perception about IT that results in behavior that identifies opportunities for innovative uses of new IT.

Traditional Management Style: An approach that maintains dominance and control over employees and relies on a "wait and see" policy for new technology adoption. Employees are encouraged to ignore data that do not fit existing frames of reference, focus only on the job at hand, and comply with established reporting relationships and rules.

Uncertain Attitude: A perception about IT that results in behavior that shows indecision about IT, hesitation to invest in IT, and anxiety about using IT.

IT Industry Success in Finland and New Zealand

Rebecca Watson
Deloitte, Australia

INTRODUCTION

At the beginning of the 21st century, information and communication technologies are creating global markets for goods and services. These technologies are impacting on every aspect of our lives, including how people work, communicate, and entertain themselves. Many economists have started to suggest that we may be entering a new era of greater productivity (without inflation) in the "knowledge economy" of the future.

Given the importance of the IT industry in today's global economy, much recent research has focused on the relative success of small countries in fostering IT industries. This research is important because those countries that can adapt and develop new information based industries will thrive and become significant players in the global economy. Those countries that cannot adapt will suffer and may find themselves as producers of low value products for wealthier nations. This article summarizes the factors of IT industry success in small developed countries, focusing on two such countries, Finland and New Zealand.

BACKGROUND

Most of the previous research in this area has compared a reasonably large number of countries. For example, Blanning et al. (1997) examined the information infrastructure of 12 Asia Pacific nations; Dedrick et al. (1995) examined reasons for the success of IT industries in nine small countries from around the world; and Kraemer, Gurbaxani, and King (1992) discussed the diffusion of computing use in nine Asia Pacific nations. Generally, these studies have examined a small range of factors that impact on either the success of a nation's IT industry or its extent of IT usage.

In contrast, Ein-dor et al. (1997) examined only three small countries – Israel, New Zealand, and Singapore. These three countries were of similar size and economic development; however, they were experiencing differing levels of IT industry success. Ein-dor et al.'s (1997) study has been one of the few pieces of research that has examined only a small number of countries in an in-depth manner.

Ein-dor et al.'s (1997) model was largely based on Grossman and Helpman's (1991) macro-economic theory concerning the relationship between technology development, trade, and growth as applied to small open economies. The latter suggested that growth stemmed from endogenous technological progress, as entrepreneurs introduced innovative (intermediate) products whenever the present value of the stream of operating profits covered the cost of product development. Grossman and Helpman (1991) postulated that the best growth path can be attained with subsidies to both R&D and the production of "intermediates" (those products that are used to produce consumer goods). The second-best growth path can be achieved with subsidies to R&D alone.

In order to study IT industry success in accordance with this theory, Ein-dor et al. (1997) considered four groups of variables. These variables are all frequently quoted in the context of industrial success. The variables they considered were as follows:

1. Controlled variables: country size and economic development.
2. Dependent variables - those that define IT industry success.
3. Exogenous mediating factors.
4. Endogenous mediating factors:
    - Domestic IT use.
    - Firm strategies.
    - Government IT policies.
    - Government education policies.

Watson and Myers (2001) adopted the model used by Ein-dor et al. (1997) to compare just two countries, New Zealand and Finland. In that study, the authors considered the same four groups of variables; however, they replaced "firm strategies" with "level of research and development," because the latter appeared to have more explanatory power. The major factors that were considered in the study by Watson and Myers (2001) are represented graphically in Figure 1.

This model was then used to compare and explain IT industry success in Finland and New Zealand. As will be summarized next, three major factors that impact on the development of a successful IT industry were identified: the extent of government IT promotion, the level of re-

Figure 1. Factors affecting IT industry success (Watson & Myers, 2001)

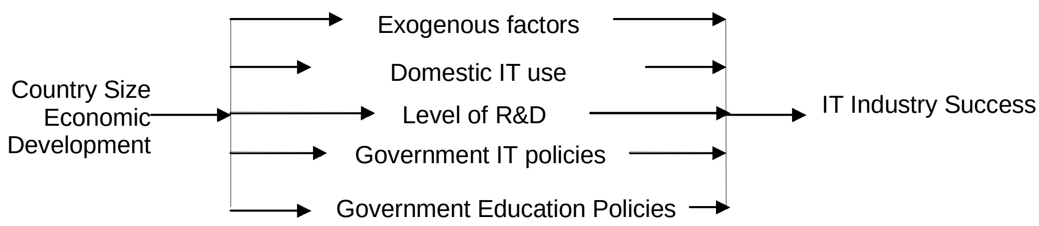


Table 1. IT industry success (1998)

	Finland	New Zealand
IT Industry Development		
IT Industry Sales (US$ billions)	11,087	2,155
IT Sales/GDP	9.0%	4.1%
Number of Firms in IT	4,200	2,500
IT Firms/Largest Firms	4/50	1/50
IT Industry Employment	5.5% (1997)	2.6% (1996)
IT Industry Success		
IT Exports:		
Hardware (US$ millions)	7,255	175
Software (US$ millions)	488	123
Total (US$ millions)	7,743	298
High Technology Exports (US$ millions) –1997	8,797	428
High Technology Exports/Manufactured Exports -1997	26%	11%
Stock Market Listings:		
Domestically listed IT firms	27	3
Internationally listed IT firms	1	0

Sources: (Computerworld, 1999; Deloitte & Touche Consulting Group, 1998; Ein-Dor et al., 1997; Finnfacts, 1999; Helsinki Stock Exchange, 1999; March, 1999; Nygard & Kunnas, 1998; Statistics Finland, 1999a, 1999b; World Bank, 1999).

search and development, and the existence of an education system that produces IT literate graduates.

COUNTRY SIMILARITIES

Dedrick, Goodman, and Kraemer (1995) define a small country as one that has fewer than ten million people. Finland and New Zealand, with populations of 5.2 million and 4.0 million respectively, are thus considered small countries. The physical size of the two countries is also similar, meaning that their population densities are almost identical.

Finland and New Zealand are remarkably alike in terms of economic development. Both countries entered the 20th century with a heavy dependence on commodity products. Forestry has played the same role in Finland's development as agriculture has played in New Zealand's development. The only real difference between the two countries is that Finland has moved away from its dependence on forestry and has embraced new technologies, whereas agriculture is still a major part of New Zealand's economy (Watson & Myers, 2001).

IT INDUSTRY SUCCESS

Finland's IT industry is far more successful than New Zealand's, as can be seen from Table 1.

Table 2. A summary of government IT policies in Finland and New Zealand

	Finland	**New Zealand**
National IT Strategies	• "Finland Towards the Information Society - a National Strategy" developed in the 1990s • Government promotion of information networks and the Internet • Extensive government grants available to the IT industry • Government assisted in the development of Technopolis, a leading science park	• No formal IT strategy • "Growing an Innovative New Zealand (GIF)" framework only released in 2002 • Preference to let the free market reign • All industries supported through wider economic strategies • No special tax incentives and few government grants for the IT industry • Government does not insist on purchasing local IT products where possible • Breaking its free market stance, the government assisted in the development of the Canterbury Technology park
IT Priority	• Development and use of IT receives very high priority • IT strategy work prolific in other areas, including the industrial, educational, cultural and health and welfare sectors	• Historically lower priority than Finland • Only since 2002 has IT become a priority with the release of the GIF framework and the development of an Information and Communication Technology (ICT) Taskforce
Government IT Organizations	• The Technology Department develops IT strategies and policies • Three additional government departments provide advice and stimulate discussion on IT issues (Science and Technology Policy Council, National Information Society Forum and Government Committee for Information Society Issues)	• No New Zealand government IT organization is responsible for setting IT related policies • Government does receive advice from ITAG and the ICT taskforce

Sources: Watson and Myers, 2001; ICT Taskforce, 2003

KEY IT INDUSTRY SUCCESS FACTORS

There are three major factors that have been identified to explain the differing IT industry success between Finland and New Zealand. These include – extent of government IT promotion, level of private sector research and development investment, and existence of an education system that produces IT literate graduates.

Government IT Policies

The majority of recent research has concluded that direct government promotion seems necessary for the development of a successful IT industry (Dedrick et al., 1995; Ein-Dor et al., 1997; Kraemer & Dedrick, 1992; Kraemer et al., 1992). Therefore, as summarized in Table 2, the existence of national IT strategies, the priorities attached to IT, and the role of government IT organizations are considered

Table 3. A summary of R&D in Finland and New Zealand

	Finland	New Zealand
Extent of R&D	• Priority area for the last 20 years • Growth in research spending among the highest of the OECD countries • Proportion of GDP spent on R&D has been steadily rising for the last two decade • Research activity in IT has been growing over the last decade, initially encouraged by increases in government funding • Majority of today's IT R&D funded by the private sector • As a percentage of sales, Finnish companies spend double the amount on R&D than other OECD countries • Nokia alone spends more on R&D than the whole of New Zealand	• New Zealand is more than halfway down the OECD country list of R&D expenditure • Government is the most significant contributor to R&D • Government funded R&D is less than half the OECD average and disproportionately skewed toward the agriculture sector • IT R&D spending is an insignificant percentage of total R&D spend • IT R&D largely funded by the private sector though government funding has increased since 2000
Performance of R&D	• Majority of R&D activities performed by the private sector • Research directed towards commercially viable areas • The Technology Development Center of Finland (TEKES) has been fostering industry oriented R&D since 1983 • TEKES encourages close industry-university interaction • TEKES has encouraged R&D growth in many industries, including IT	• Most R&D performed by the government's Crown Research Institutes (CRI) • CRI focus is on basic research from which commercial applications can be derived • Concentrates on primary production industries • Recent developments are encouraging as CRI funding has been opened up to companies and private researchers
R&D Tax Incentives	• Tax incentives provided to stimulate investment, which have now been discarded • Current tax situation still favorable to R&D investment as companies can deduct current business expenditures on R&D in the year incurred	• R&D tax situation is one of the least favorable of the OECD countries • No tax incentives or subsidies offered • Recent tax improvements for R&D still do not allow all R&D expenditure to be deducted in the year incurred • The requirement to capitalize R&D may lead to an under-reporting of R&D and/or act as a disincentive to such investments

Source: ICT Taskforce 2003; Watson & Myers, 2001 important to the development of a successful IT industry.

Finland and New Zealand are clearly different in terms of government IT promotion. Finland has a high priority national IT strategy that promotes both IT use and industry development. As a result, Finland is both a heavy user of IT and a significant player in the international IT industry. New Zealand has no IT strategy and IT has only recently become a high priority industry for the current labour government (ICT Taskforce, 2003). Whilst this historic lack of government support has not hindered New Zealand's adoption of IT, it has hindered its IT industry success. Government promotion of IT is thus one factor that impacts on the development of a successful IT industry (Watson & Myers, 2001).

Level of Research & Development

Investment in research and development (R&D) fuels innovation and raises the technological sophistication of manufacturing processes. The amount of R&D undertaken in an industry is often an indication of potential growth, productivity gains, and competitive advantage (Dobbeck & Woods, 1994). Total R&D expenditures, sources of R&D funding, and R&D tax incentives are summarized next, as these factors are believed to be associated with success in IT production (Kraemer & Dedrick, 1992).

Finland and New Zealand have vastly different levels of R&D investment. Finland spends an increasingly

Table 4. A summary of education in Finland and New Zealand

	Finland	**New Zealand**
Total Education Expenditure	• One of the highest education expenditures as a percentage of GDP within OECD countries • Finland spends slightly more of its GDP on education than New Zealand	• One of the highest education expenditures as a percentage of GDP within OECD countries
School Education	• Actively promotes IT skills at school level • Extensive IT usage throughout schools • Primary and secondary schools have offered computing since the mid 1980s • Computer literacy is part of today's national curriculum • Every student has access to a computer and every school has fast Web access	• Smaller investment in IT infrastructure than other OECD countries • Limited computer access in primary and secondary schools • Only 55% of primary and 60% of secondary schools have Web access from at least one classroom • IT expenditure in schools increasing and it is expected that all schools will have adequate computer and Web access within the next few years
Tertiary Education	• Around 20 universities and institutes of higher education • Computing established in the higher education sector in 1965 • IT topics taught in over 15 universities which annually graduate over 600 5-year degrees and 40 doctorates • Polytechnic network produces over 2,000 degrees a year in computing and engineering • Five times as many science and technology graduates as law graduates	• Currently nine universities • Most universities and the 25 polytechnics offer IT related degrees or diplomas • Student enrolments in IT related degrees have rapidly increased over the past decade • IT graduates still a fraction of the total number of graduates • The IT industry has complained of a mismatch between IT graduate skills and those required

Source: Watson & Myers, 2001

significant proportion of its GDP on R&D. Most of this R&D is funded and performed by the private sector, ensuring that it is concentrated in growth industries such as IT. Further, R&D continues to be stimulated through favorable tax conditions and grants from TEKES. In New Zealand, R&D spending is extremely low. The government funds and performs most of the research, which tends to be concentrated in agricultural industries. As a result, R&D in the IT industry is minimal. Even with recent improvements, New Zealand's R&D tax situation could be further enhanced to increase private sector R&D. Therefore, a country's level of R&D appears to affect the development of a successful IT industry. Moreover, a high level of private sector R&D investment seems to be important for IT industry success.

Government Education Policies

It is believed that a top quality education system is essential to being successful in the information age. Both Finland and New Zealand consider education important and have well developed education systems. Total education expenditure, quality of school education and the quality of tertiary education is used to compare the education systems of Finland and New Zealand.

Both Finland and New Zealand consider education important and have well-developed education systems, indicating that a reasonable level of education is required for IT industry success. However, the Finnish and New Zealand education systems differ in two main ways. Firstly, Finland has implemented IT and promoted IT use in schools to a greater extent than New Zealand. Secondly, Finland is producing a greater number of IT related graduates. These two findings may even be correlated. Students that are exposed to IT in schools may be more inclined to pursue IT related courses at tertiary level. In any case it appears that educational policies have an impact on IT industry success and suggests that a high degree of IT competence at school and tertiary level is associated with a successful IT industry (Watson & Myers, 2001).

FUTURE TRENDS

These findings have many implications for policy makers in small countries. The findings suggest that governments need to take a proactive role in fostering their IT industry if they are to have any chance of developing information-based industries of their own. Those governments that do support their IT industries are likely to experience strong growth and significant success. Governments that do not support their IT industries may still become sophisticated users of IT; however, it is not expected that they will develop IT industries that are as successful as countries providing government support through policy, education and R&D support. The extent of support required is still a matter for debate, but it is clear that more support is better than none at all.

Further, after witnessing the success of countries that have supported their IT industries, other countries may begin to increase their levels of support in order to compete in the information-based economy. This trend has already been seen in New Zealand. Traditionally, New Zealand governments have preferred to let the free market reign. However, since 2002 a number of initiatives have been implemented to boost the development of New Zealand's IT industry. If these initiatives are successful, it is expected that further initiatives will be developed.

CONCLUSION

Finland and New Zealand have been identified as being similar in terms of country size and economic development. Despite this similarity, the two countries are experiencing differing levels of IT industry development. Finland's IT industry is far more successful than New Zealand's, particularly in terms of the hardware sector.

Though there may be other factors that have contributed to Finland's success, research has suggested three major factors – government IT promotion, high levels of private sector R&D investment, and an education system that produces IT literate graduates - as being important for IT industry success. These findings are consistent with those of Ein-Dor et al. (1997) and support the macroeconomic theory of Grossman and Helpman (1991). It appears that there is an optimal level of government support for IT industries in small, open economies, and that government support for the IT industry (as in Finland) is substantially better than no support at all (as in New Zealand).

Of course, one of the main limitations of these findings is that only two small, developed countries were studied. Whether these findings hold for other small, developed countries requires further research. Despite this limitation, the findings are consistent with earlier work in this area.

REFERENCES

Blanning, R.W., Bui, T.X., & Tan, M. (1997). National information infrastructure in Pacific Asia. *Decision Support Systems*, *21*, 215-227.

Computerworld. (1999). *NZ computer industry directory*

1999. IDG Communications.

Dedrick, J.L., Goodman, S.E., & Kraemer, K.L. (1995). Little engines that could: Computing in small energetic countries. *Communications of the ACM, 38*(5), 21-26.

Deloitte & Touche Consulting Group. (1998). Top 200 New Zealand companies. *Management Magazine, 45*, 74-87

Dobbeck, D., & Woods, W. (1994, June/July). Mapping industrial activity. *OECD Observer, 188,* 19-23.

Ein-Dor, P., Myers, M.D., & Raman, K.S. (1997). Information technology in three small developed countries. *Journal of Management Information Systems*, *13*(4), 61-89.

Finnfacts. (1999). *50 largest Finnish companies*. *www.finnfacts.com/ Ffeng0399/record_profits.htm*

Grossman, G., & Helpman, E. (1991). *Innovation and growth in the global economy*. Cambridge, MA: MIT Press

Helsinki Stock Exchange. (1999). *Listed companies*. *www.hex.fi/eng/listed_companies/*

ICT Taskforce. (2003). *Breaking through the barriers*. New Zealand: Information and Communications Technology Taskforce.

Kraemer, K., & Dedrick, J. (1992). *National technology policy and the development of information industries*. Irvine, CA: Centre for Research on Information Technology and Organizations, University of California.

Kraemer, K.L., Gurbaxani, V., & King, J.L. (1992). Economic development, government policy, and the diffusion of computing in Asia-Pacific countries. *Public Administration Review, 52*(2), 146-156.

March, F. (1999). *Statistics on information technology in New Zealand*. Wellington: Ministry of Commerce.

Nygard, A.M., & Kunnas, T. (1998). *Computer networking hardware/software*. Finland: International Trade Administration.

Statistics Finland. (1999a). *On the road to the Finnish information society - Summary*. *www.stat.fi/tk/yr/tttietoti_en.html*

Statistics Finland. (1999b). *Statistical news*. *www.stat.fi/tk/tp_tiedotteet/v99/ 002ttte.html*

Watson and Myers. (2001). A comparison of IT industry success in Finland and New Zealand. *Journal of Global Information Management, 9*(2), 3-13.

World Bank. (1999). *World development indicators 1999*. *www.worldbank.org/data/wdi/*

KEY TERMS

High Technology: Includes space and aviation, computers and office machinery, electronics and telecommunication equipment, pharmaceuticals, scientific instruments, electrical machines and equipment, chemicals, non- electrical machines, and weapons.

IT Hardware: Includes telecommunications equipment, computing hardware, and computing parts. Excludes embedded hardware in other products.

IT Industry: Includes the IT hardware and IT software industries. Excludes embedded hardware and software in other products (e.g., washing machines).

IT Industry Development: In this context is measured by gross IT sales, IT sales relative to GDP, firms in IT and their relative size, and IT industry employment relative to total employment.

IT Industry Success: In this context is measured by IT software and hardware exports, high technology exports, and IT stock market listings.

IT Software: Includes software applications, computer services, and training. Excludes embedded software in other products.

IT Productivity Impacts in Manufacturing Contexts

Kristina Setzekorn
University of Evansville, USA

Arun Rai
Georgia State University, USA

Arlyn J. Melcher
Southern Illinois University at Carbondale, USA

INTRODUCTION

The fundamental questions of whether and how information technology (IT) contributes to firm performance have been answered in different ways. IT value research findings have been equivocal, with some studies finding negative performance impacts (Berndt & Morrison, 1995; Johansen, Karmarkar, Nanda & Seidmann, 1996), some finding no overall effect (Barua, Kriebel & Mukhopadhyay, 1995; Dos Santos, Peffers & Mauer, 1993; Loveman, 1994; Strassman, 1985, 1990), and some finding positive impacts (Brynjolfsson & Hitt, 1996; Brynjolfsson & Yang, 1997; Hitt & Brynjolfsson, 1996; Mukhopadhyay, Kekre & Kalathur, 1995).

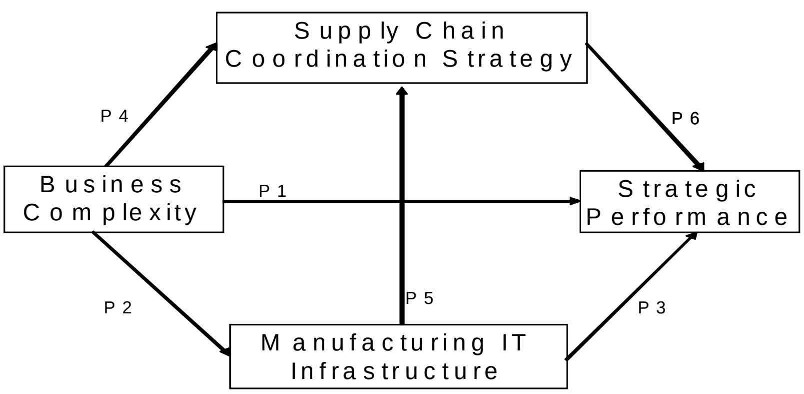
To reconcile these findings, several studies suggest that contextual factors associated with the firm and/or its environment mediate IT's performance effects (c.f., Banker, Kauffman & Morey, 1990; Brynjolfsson & Yang, 1997; Scott-Morton, 1991; Venkatraman, 1991; Weill, 1992). Brynjolfsson and Hitt find that "firm effects" accounted for roughly half the productivity benefits attributed to IT. Firm capabilities leverage investments in IT. They suggest:

"…an interesting extension would be to identify common characteristics of the highly productive firms and thereby examine some of the conventional wisdom regarding management best-practice" (1995, p. 12).

BACKGROUND

This project explores how business complexity, supply chain coordination strategy, and manufacturing IT infrastructure interact to impact inventory productivity. The research model is represented in Figure 1, where theoretical constructs are represented by rectangles, and proposed associations between them are represented by numbers.

Figure 1. Research model



Business complexity describes the degree of coordination difficulty associated with a firm's supplier- and customer-facing processes due to volatility and diversity of its product-market (Holland, 1995). This reflects the transactions cost perspective in which uncertainty results from imperfect foresight and difficulty in solving problems containing multiple variables.

Business complexity has become a crucial consideration, as global competition has pressured firms to diversify their product offerings into markets with which they have little experience, and at the same time, to compete on cost, quality, reliability, and responsiveness. They are attempting to maximize performance on all dimensions, while struggling to cope with product proliferation and heightened customer expectations in unfamiliar product markets.

Manufacturing IT infrastructure describes the extent of IT deployment for manufacturing planning and control (MPC) functions. It is defined as the enabling base of shared IT capabilities which "provides common services to a range of applications" (Broadbent, Weill, O'Brien & Neo, 1997, p. 175), including "…information to efficiently manage the flow of materials…coordinate internal activities with those of suppliers, and communicate with customers about market requirements" (Vollmann, Berry & Whybark, 1992, p. 2). It is hypothesized to mediate business complexity's impact on inventory turnover.

As companies struggle to expand margins and improve operational excellence within their own organizational boundaries, their efforts often earn diminishing marginal returns. Accordingly, organizations have begun to focus on performance optimization across organizational boundaries in the supply chain. IT presents opportunities to streamline and integrate key operations and processes by coordinating distributed activities within and across a firm's boundaries. Consequently, as firms within a supply chain coordinate product designs, demand forecasts, and production among themselves, schedules become more stable and coordination simpler.

Schedule stability improves output efficiency by reducing inventory buffers and their attendant physical costs throughout the supply chain. This confers competitive advantage, in that capital invested in inventory becomes available for other differentiating opportunities. Market mediation costs are also reduced. These include costs associated with lost sales and customer dissatisfaction due to stock outs, as well as lower profits due to product markdowns on unsold goods. In volatile industries, these costs can exceed the manufacturing cost (Fisher, 1997).

Brynjolfsson and Yang (1997) found that firms investing in complementary intangibles such as business process adjustments, training, and interorganizational relationships reap four times the return from their investment in IT as those not making these investments. To the extent that firms leverage IT to improve interorganizational relationships, and to coordinate and integrate their marketing, planning, and production decisions, their inventory productivity and other measures of performance should improve.

The *supply chain coordination strategy* describes the firm's approach to process coordination across firm boundaries. The goal is to improve performance across the entire supply chain. It measures the extent to which the firm has integrated its activities with those in the larger business network so that efficiencies are maximized across the whole business network—not just the individual function or firm (Venkatraman, 1991). This construct is measured on a continuum anchored by transaction on one end and partnership on the other (c.f., Henderson, 1990; Malone, Yates & Benjamin, 1987).

Inventory productivity is the dependent variable, calculated as annual sales revenue/value of average total inventory (Vollmann et al., 1992). This is consistent with Brynjolfsson and Yang's (1997) calculation of productivity, and is regarded as a useful measure of the physical efficiency of a firm's supply chain (Fisher, 1997). This key measure addresses the physical movement of goods and the efficiency of this process. Inventory productivity also reflects firms' effectiveness in responding to market context. Whether the market dictates responsiveness, or low-cost strategies, or both, annual sales revenue reflects success in meeting market demands.

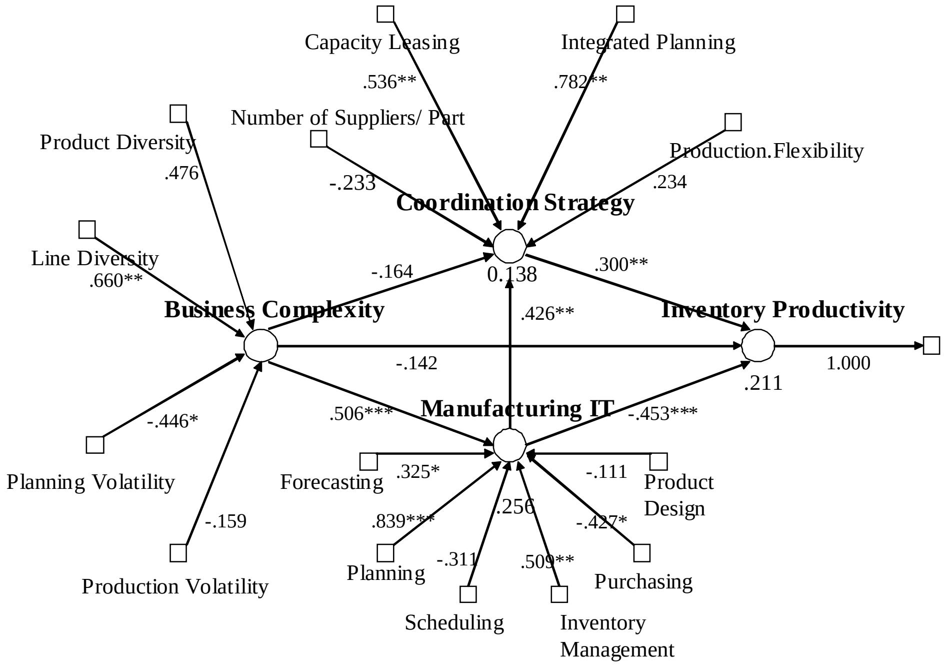
The reduction of physical costs associated with production, transportation, and inventory, and of market mediation costs associated with stock-outs and markdowns, has become a strategic necessity. To the extent that IT can be used to accurately forecast demand; efficiently plan, schedule, and accomplish production; and purchase and manage inventory in the supply chain, it can support improved performance.

MEDIATION EFFECTS OF SUPPLY CHAIN COORDINATION STRATEGY AND MANUFACTURING IT

Our research question asks, "How do manufacturing IT infrastructure and supply chain coordination strategy mediate the relationship between business complexity and inventory productivity?"

Previous empirical studies have generally attempted to measure generic IT dollar investments and their relationships to productivity, profitability, or firm valuation. These rely on aggregated data regarding multiple applications from a variety of industries. We're aware of none that include both complexity and supply chain coordina-

Figure 2. PLS model results



tion strategy in the study of manufacturing IT infrastructure and its effect on inventory turnover.

A few theoretical studies (Das, Zahra & Warkentin, 1991; Holland & Lockett, 1997) have suggested that these variables' fit is important for firm performance. Few of the relationships have been tested empirically. Also, since our study controls for confounding factors, and our partial least squares (PLS) method of analysis is more precise than first-generation multivariate methods, we are able to distinguish these mediation effects, where previous studies have not.

This research was able to empirically test theoretical frameworks (c.f., Das et al., 1991; Holland & Lockett, 1997), suggesting that the degree of fit between an organization's complexity level, its supply chain coordination strategy, and its manufacturing IT infrastructure influences its performance. Figure 2 shows the PLS model results. Firms with longer planning horizons and more product lines deployed more IT for forecasting, planning, and inventory management. This deployment alone was associated with lower levels of inventory productivity; however, it was associated with higher levels of integrated planning and capacity leasing, which were associated with higher inventory productivity.

This result empirically supports theorists who postulate that IT alone is insufficient for success. Complementary factors, especially those that are tacit—that is, intangible and inimitable—also need to be adjusted in order to reap IT's promised payoff (c.f., Brynjolfsson & Yang, 1997; Clemons & Row, 1992; Davidow & Malone, 1992; Milgrom & Roberts, 1990; Rockart & Short, 1989; Short & Venkatraman, 1992; Venkatraman, 1991). This complementarity between integrated planning, capacity leasing, and elements of manufacturing IT infrastructure probably enhances inter-organizational "sense-and-respond capabilities" to improve performance, as evidenced by coordination strategy's positive and significant relationship with inventory turnover. Our results also support Bhaskaran's (1998) suggestion that supply chain cooperation and coordination improve forecast accuracy and schedule stability, which improves productivity.

Our study adds an additional theoretical dimension, as previous empirical IT value studies have measured highly aggregated IT dollar investments and their relationships to financial performance measures. We have instead measured IT deployment (which necessarily includes some investment and adoption decisions related to complementary intangibles such as training, change management, and process reengineering) for a particular use in a single industry over a two-year period. This focus has allowed us to control for many confounding factors. Further, our PLS method of analysis allowed us to distinguish the mediation effects of coordination strategy and manufacturing IT infrastructure, where previous studies may not have. Thus, our deployment measures have included some intangible elements and excluded some extraneous factors.

FUTURE TRENDS

To increase our understanding, IT value studies should consider more contextual variables and model them more realistically. The PLS method is a good analysis tool for this type of work (Chin, 1998).

We plan to extend the study to measure competitive advantage as indicated unit cost of manufacturing, product quality, throughput speed, delivery speed and reliability, flexibility, and design time. This would be especially important in contexts where inventory efficiency is secondary to market responsiveness. Additionally, we plan to extend the study of these constructs to a more modern IT paradigm, that of B2B e-commerce such as in the auto industry, where supply chain coordination strategy and business complexity are paramount concerns.

CONCLUSION

In this age of e-business, managers must beware of generalizing theoretical prescriptions too broadly. Managers must analyze IT deployment in terms of desired, sometimes conflicting, performance impacts; using relevant, sometimes non-financial, measures; and in terms of complementary investments, some of which are intangible. We found that firms with more product lines deployed more forecasting, planning, and inventory management IT; but without complementary deployment of integrated planning and capacity leasing, they experienced significantly lower inventory productivity. However, with these complementary investments, they reaped significantly better inventory productivity.

Where inventory productivity is critical, deployment of these technologies must be coupled with cooperative supply chain coordination strategies. Managers must consider complementary investments and strategic adjustments for tangible IT investments to confer advantage.

REFERENCES

Banker, R.D., Kauffman, R.J. & Morey, R.C. (1990). Measuring gains in operational efficiency from information technology. *Journal of Management Information Systems, 7*(2), 29-54.

Barua, A., Kriebel, C. & Mukhopadhyay, T. (1995). Information technology and business value: An analytic and empirical investigation. *Information Systems Research, 6*(1), 1-24.

Berndt, E.R. & Morrison, C.J. (1995). High-tech capital formation and economic performance in U.S. manufacturing industries: An exploratory analysis. *Journal of Econometrics*, *65*, 9-43.

Bhaskaran, S. (1998). Simulation analysis of a manufacturing supply chain. *Decision Sciences, 29*(3), 633-657.

Broadbent, M., Weill, P., O'Brien, T. & Neo, B.S. (1996). Firm context and patterns of IT infrastructure capability. In A. Srinivasan, S. Jarvenpaa & J.I. DeGross (Eds.), *Proceedings of the 17th International Conference on Information Systems* (pp. 174-194), Cleveland, Ohio.

Brynjolfsson, E. & Hitt, L. (1995). Information technology as a factor of production: The role of differences among firms. *Economics of Innovation and New Technology, 3*(4), 183-200.

Brynjolfsson, E. & Hitt, L. (1996). Paradox lost? Firm-level evidence on the returns to information systems spending. *Management Science, 42*(4), 541-558.

Brynjolfsson, E. & Yang, S. (1996). Information technology and productivity: A literature review. *Advances in Computers*, *43*, 179-215.

Brynjolfsson, E. & Yang, S. (1997). The intangible benefits and costs of computer investments: Evidence from financial markets. In E.R. McLean, R.J. Welke & J.I. DeGross, (Eds.), *Proceedings of the 18th International Conference on Information Systems*, Atlanta, Georgia.

Chin, W. (1998). The partial least squares approach to structural equation modeling. In G.A. Marcoulides (Ed.), *Modern research methods* (pp. 295-336). Mahwah, NJ: Lawrence Erlbaum.

Clemons, E.K. & Row, M.C. (1992). Information technology and industrial cooperation: The changing economics of coordination and ownership. *Journal of Management Information Systems, 9*(2), 9-28.

Das, S.R.,. Zahra, S.A. & Warkentin, M.E. (1991). Integrating the content and process of strategic MIS planning with competitive strategy. *Decision Sciences, 22,* 953-983.

Davidow, W.H. & Malone, M.S. (1992). *The virtual corporation.* New York: Harper Collins.

Dos Santos, B.L., Peffers K.G. & Mauer, D.C. (1993). The impact of information technology investment announcements on the market value of the firm. *Information Systems Research*, *4*(1), 1-23.

Fisher, M. (1997). What is the right supply chain for your product? *Harvard Business Review*, (March-April), 105-116.

Henderson, J. (1990). Plugging into strategic partnerships: The critical IS connection. *Sloan Management Review*, (Spring), 7-18.

Hitt, L.M. & Brynjolfsson, E. (1996). Productivity, business profitability, and consumer surplus: Three different measures of information technology value. *MIS Quarterly, 20*(2), 121-142.

Holland, C. (1995). Cooperative supply chain management: The impact of inter-organizational information systems. *Journal of Strategic Information Systems, 4*(2), 117-133.

Holland, C. & Lockett, G. (1997). Mixed mode network structures: The strategic use of electronic communication by organizations. *Organization Science, 8*(5), 475-488.

Johansen, J., Karmarkar, U., Nanda, D. & Seidmann, A. (1996). Computer-integrated manufacturing. *Journal of Management Information Systems, 12*(4), 59-82.

Loveman, G.W. (1994). An assessment of the productivity impact of information technologies. In T.J. Allen & M.S. Scott Morton (Eds.), *Information technology and the corporation of the 1990s: Research studies* (pp. 84-110). Oxford: Oxford University Press.

Malone, T.W., Yates, J. & Benjamin, R.I. (1987). Electronic markets and electronic hierarchies. *Communications of the ACM, 30*, 484-497.

Milgrom, P. & Roberts, J. (1990). The economics of modern manufacturing: Technology, strategy and organization. *The American Economic Review, 80*(3), 511-528.

Mukhopadhyay, T., Kekre, S. & Kalathur, S. (1995). Business value of information technology: A study of electronic data interchange. *MIS Quarterly, 19*(2), 137-156.

Rockart, J.F. & Short, J.E. (1989). IT in the 1990s: Managing organizational interdependence. *Sloan Management Review*, (Winter), 7-16.

Scott-Morton, M. (Ed.). (1991). *The corporation of the 1990s*. New York: Oxford University Press.

Strassmann, P.A. (1990). *The business value of computers: An executive's guide*. New Canaan, CT: Information Economics Press.

Strassmann, P.A. (1985). *Information payoff: The transformation of work in the electronic age*. New York: The Free Press.

Venkatraman, N. (1991). IT-induced business reconfiguration: The new strategic management challenge. In M.S. Scott-Morton (Ed.), *The corporation of the 1990s* (pp. 122-158). New York: Oxford University Press.

Vollmann, T.E., Berry, W. & Whybark, D.C. (1992). *Manufacturing planning and control systems*. Boston: Richard D. Irwin, Inc.

Weill, P. (1992) The relationship between investment in information technology and firm performance: A study of the valve manufacturing sector. *Information Systems Research, 3*(4), 307-333.

KEY TERMS

Business Complexity: Degree of difficulty associated with supplier- and customer-facing processes. Incorporates diversity and volatility aspects.

Electronic Hierarchies: Markets that entail few suppliers, as an intermediate step from hierarchical, ownership strategies toward electronic markets.

Electronic Markets: Markets characterized by infinite numbers of competing suppliers, selling completely specifiable products.

Manufacturing IT Infrastructure: Enabling base of shared IT capabilities which "provides common services to a range of applications" (Broadbent et al., 1997, p. 175), including "…information to efficiently manage the flow of materials…coordinate internal activities with those of suppliers, and communicate with customers about market requirements" (Vollmann et al., 1992, p. 2).

Market Mediation Costs: Costs associated with lost sales and customer dissatisfaction due to stock-outs, as well as lower profits due to product markdowns on unsold goods.

Physical Costs: Costs associated with production, transportation, and inventory.

Productivity Paradox: Term coined to describe the decline in productivity growth that began in the 1970s, just as IT investment began to dramatically increase. Labor productivity growth slowed from 2.5% per year from 1953 to 1968, to 0.7% per year from 1973 to 1979. Multifactor productivity growth also fell from 1.75% a year to 0.32% over this timeframe. Concurrent with these declines, office computers and machines capital rose from 0.5% of all producers' durable equipment in the 1960s to 12% in 1993 (Brynjolfsson & Yang, 1996).

Supply Chain Coordination Strategy: Extent to which the firm has integrated its activities with those in the larger business network. "Choices made for coordinating economic activity with trading partners and includes…degree to which the relationship with trading partner…reflects a long-term commitment, a sense of mutual cooperation, shared risk and benefits, and other qualities consistent with concepts and theories of participatory decision making" (Henderson, 1990, p. 8). Measured on continuum anchored by partnership and transaction coordination strategies.

APPENDIX I: STUDY VARIABLES

Business Complexity

Business Complexity—Firm's response to the scale and difficulty of supplier and customer facing processes in the marketplace (Holland & Lockett, 1997).
(Less Complex ↔ More Complex)

A) Product diversity: Number of products (Less Complex ↔ More Complex)
   1. Number of products for which sales forecasts are developed (Few ↔ Many)
   2. Number of products for which production plans are developed (Few↔Many)

B) Product line diversity: Number of product lines (Less Complex ↔ More Complex)
   1. Number of product lines produced (Few ↔ Many)
   2. Number of product lines forecast (Few ↔Many)
   3. Number of product lines in production plan (Few↔Many)

C) Planning volatility: (Less Complex↔More Complex)
   1. Reverse of months into future company's sales forecast extends (Few↔Many)
   2. Reverse of months into future company's production plan extends (Few↔Many)

D) Production volatility: (Less Complex ↔More Complex)
   1. % orders for which customer schedule changes occur after the start of production (Low↔High)
   2. % orders for which engineering or design changes occur after the start of production (Low↔High)
   3. % incoming material rejected (Low↔High)

IT Infrastructure

Manufacturing IT Infrastructure—The enabling base of shared IT capabilities which "provides common services to a range of applications" (Broadbent et al., 1997, p. 175), including "…information to efficiently manage the flow of materials…coordinate internal activities with those of suppliers, and communicate with customers about market requirements" (Vollmann et al., 1992, p. 2). (Less Extensive↔More Extensive)

Extent of IT deployment for:

A. sales forecasting (Low↔High)
B. production planning (Low↔High)
C. production scheduling (Low↔High)
D. inventory management (Low↔High)
E. purchasing (Low↔High)
F. product design (Low↔High)

Supply Chain Coordination Strategy

Supply Chain Coordination Strategy: "Choices made for coordinating economic activity with trading partners and includes…degree to which the relationship with trading partner…reflects a long-term commitment, a sense of mutual cooperation, shared risk and benefits, and other qualities consistent with concepts and theories of participatory decision making" (Henderson, 1990, p. 8). (Transaction ↔ Partnership)

A. Integrated planning: (Low↔High)
   1. Extent to which firm invests in JIT, MRP, TQM, and supplier partnership programs (Low↔High)
   2. Extent to which sales forecast is used for subcontracting decisions (Low↔High)
   3. Extent to which customers' future plans are considered in the development of the company's production plan (Low↔High)

B. Capacity leasing: (Low↔High)
   How often company responds to capacity imbalance by
   1. leasing temporary capacity from others (Low↔High) or by
   2. leasing excess capacity to others (Low↔High)

C. Reverse of number of suppliers/part: (Low↔High)

D. Production flexibility: (Low↔High)
   Extent to which the firm considers
   1. set-up time (Low↔High) and
   2. set-up cost (Low↔High)
   in determining the size of its production lots

Inventory Productivity

Annual Sales Revenue/Average Inventory Investment

IT Supporting Strategy Formulation

Jan Achterbergh
Radboud University of Nijmegen, The Netherlands

INTRODUCTION

This overview approaches information and communication technology (ICT) for competitive intelligence from the perspective of strategy formulation. It provides an ICT architecture for supporting the knowledge processes producing relevant knowledge for strategy formulation.

To determine what this architecture looks like, we first examine the process of strategy formulation and determine the knowledge required in the process of strategy formulation. To this purpose, we use Beer's viable system model (VSM). Second, we model the knowledge processes in which the intelligence relevant for the process of strategy formulation is produced and processed. Given these two elements, we describe an ICT architecture supporting the knowledge processes producing the knowledge needed for the strategic process.

BACKGROUND: STRATEGY FORMULATION, A VIABLE SYSTEM PERSPECTIVE

Strategy formulation aims at developing and selecting goals and plans securing the adaptation of the organization to its environment. These goals and plans may refer to specific product-market-technology combinations (PMCs) for which the organization hypothesizes that they ensure a stable relation with its environment. The process of strategy formulation needs to generate such goals and plans, needs to reflect upon their appropriateness, and needs to select certain goals and plans to guide the behavior of the organization. This is a continuous process. Goals and plans can be seen as hypotheses about what will work as a means to adapt and survive. Therefore, they should be monitored constantly and revised if necessary. In short, strategy formulation is a continuous contribution to maintaining organizational viability.

Although many authors deal with the process of strategy formulation, we choose the viable system model of Beer (1979, 1981, 1985) to define this process more closely. We select the VSM because Beer explicitly unfolds the functions required for the viable realization and adaptation of an organization's strategy.

To explain what these functions entail, it is useful to divide them into two groups: functions contributing to the *realization* of the organization's strategy and functions contributing to its *adaptation*.

The first group deals with the realization of the organization's strategy. It consists of three functions. Function 1 comprises the organization's primary activities constituting its "raison d'être" (Espejo, Schumann, Schwaninger, & Billello, 1996, p. 110). Function 2 (coordination) coordinates interdependencies between these primary activities. The third function is called the control function. It ensures the synergy of and cohesion between the primary activities by specifying their goals and controlling their performance.

To illustrate these functions, consider Energeco, a company servicing its environment with eco-energy. Function 1 of Energeco consists of three primary activities: supplying solar, tidal, and wind energy. To give an example of the coordination function, suppose that specialists in high-voltage energy are a shared resource between Energeco's business units. Also suppose that there is no coordination between these business units. In this case, the allocation of high-voltage specialists to a project in the business unit Solar Energy may require a revision of the allocation of these same specialists to a project in the business unit Wind Energy. Without a function supporting the coordination of these interdependencies, the business units Solar Energy and Wind Energy may become entangled in a process that oscillates between allocating and revising the allocation of these specialists to projects. It is the task of Function 2 to coordinate these interdependencies. The control function's task is to translate the identity and mission of the viable system (for Energeco, supplying eco-energy) into goals for the primary activities (in this example, supplying wind, solar, and tidal energy) and to control the realization of these goals.

The second group deals with the adaptation of the organization's strategy. It consists of control (Function 3), intelligence (Function 4), and policy (Function 5). Intelligence scans the organization's relevant environment and generates and proposes plans for adaptation. In the example of Energeco, developments in production technology may introduce the possibility of cost-effective, large-scale production of eco-energy from biomass. Intelligence should pick up these developments, assess them, and if relevant, translate them into proposals for innovation. Because of its knowledge of the potentials for change of the primary activities, control (Function 3) reviews the feasibility of the plans proposed by intelligence. For instance, it may object to the plans proposed by intelligence because they require a change posing a risk to the performance of the primary activities.

Discussion about the relevance and feasibility of the proposals for adaptation between intelligence and control should produce finalized plans for adaptation. It is the task of the policy function to balance the discussion between intelligence and control and to consolidate the finalized proposal in the organization's strategy. For instance, in the discussion between intelligence and control about the feasibility of the adoption of large-scale production of eco-energy from biomass, the policy function should ensure that control and intelligence are equally represented in the discussion. By opting for the production of energy from biomass, the policy function consolidates producing eco-energy from biomass as a new goal for Energeco. Figure 1 depicts the process of strategy formulation in terms of the VSM functions and activities.

To contribute to the strategy-formulation process, control, intelligence, and policy require knowledge about particular domains. Table 1 provides an overview of the knowledge required by each function to contribute to the process of strategy formulation.

Given the overview of functions involved in the strategy-formulation process, their relations, and the knowledge required by these functions to contribute to the process of strategy formulation, it is now possible to look into the knowledge processes needed to produce this knowledge and the ICT architecture supporting these knowledge processes.

KNOWLEDGE PROCESSES CONTRIBUTING TO STRATEGY FORMULATION

The question for this section is by means of what processes knowledge in the knowledge domains should be produced and processed so that the process of strategy formulation can take place. To answer this question, we first need to specify what these knowledge processes are. Then we need to link these processes to the knowledge required by control, intelligence, and policy to contribute to the strategy-formulation process.

We distinguish four relevant processes for producing and processing knowledge: generating (G), sharing (S), retaining (R), and applying (A) knowledge (cf. Achterbergh & Vriens, 2002; Bukowitz & Williams, 1999; Davenport & Prusak, 1998).

These four knowledge processes can now be linked to the process of strategy formulation, as formulated according to the VSM. According to the VSM, the functions intelligence, control, and policy contribute to strategy formulation. This contribution involves the *application* of knowledge in the knowledge domains to arrive at the four core products of strategy formulation: proposals for innovation, their reviews, the finalized plans for innovation, and their consolidation. The knowledge applied by each function is *generated* either by that function or by one of the other functions of the VSM. In the latter case, knowledge must be *shared* between functions. Applying, generating, and sharing knowledge requires the *retention* or *storage* of knowledge.

Figure 1. The process of strategy formulation according to the VSM

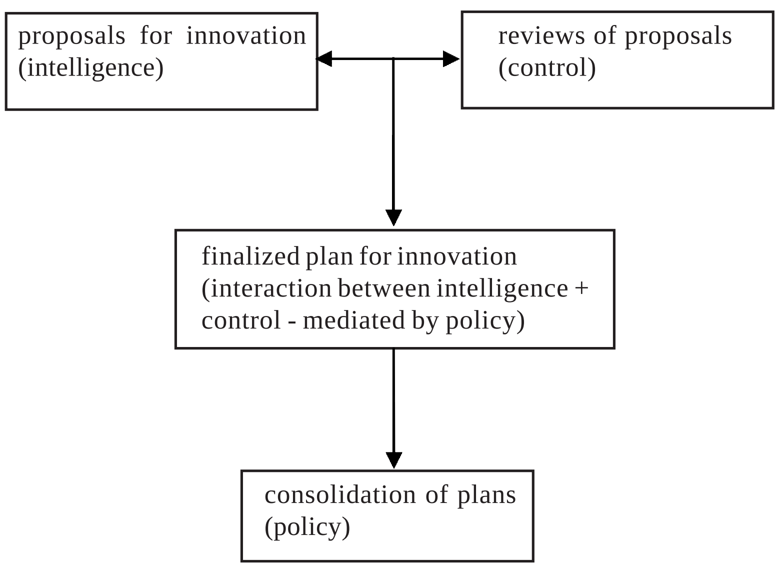


Table 1. Knowledge required by each function to contribute to the strategy formulation process

Function	Related domains of knowledge
F3: Function 3 (control)	For reviewing F4 proposals Organizational goals Proposals for innovation made by F4 Desired goals for F1 based on proposals for innovation Expected performance of the primary activities (goals for F1 activities) Gap between desired and current goals for F1 Required capacity for reorganization of F1 activities Modus operandi of F1 activities Actual capacity for reorganization of F1 activities Gap between required and actual capacity for reorganization Review of proposals for innovation Finalized plans for adaptation of organizational goals (a joint F3 and F4 product) Regulatory measures to counter the imbalance between F3 and F4 (see Function 5)
F4: Function 4 (intelligence)	Organizational goals Goals set by performance and modus operandi of F1 activities Developments in the relevant environment of the organization Reviews by F3 of proposals for innovation Regulatory measures to counter the imbalance between F3 and F4 (see Function 5) Finalized plans for adaptation of organizational goals (a joint F3 and F4 product)
F5: Function 5 (policy)	For balancing purposes Norms for balance between F3 and F4 Proposals by F4 and their reviews by F3 (relative contribution of F3 and F4 to the discussion on adaptation) Actual (im)balance between F3 and F4 Causes of imbalance between F3 and F4 Experiences with regulatory measures to counter the imbalance between F3 and F4 Regulatory measures to counter the imbalance between F3 and F4 For consolidation purposes Finalized plans for adaptation of organizational goals (a joint F3 and F4 product) Organizational goals

Table 2 provides an overview of the relation between the five functions in the VSM, the knowledge domains, and the application and generation of knowledge in these domains. Based on this table, it is possible to draw conclusions about sharing and retaining knowledge. In the table we only included the relevant knowledge for strategy formulation. However, some of this knowledge is generated by Function 1; this is the reason of its inclusion in the table.

The first column of Table 2 summarizes the knowledge domains listed in Table 1. In this column, we eliminated all redundant entries. Columns 2 to 5 indicate whether knowledge in a specific knowledge domain is generated and/or applied by a specific function.

Given the link between the knowledge processes, the functions contributing to the strategy-formulation process, and the knowledge required by them, it is now possible to outline an ICT architecture that can support the generating, retaining, sharing, and applying of this knowledge by these functions.

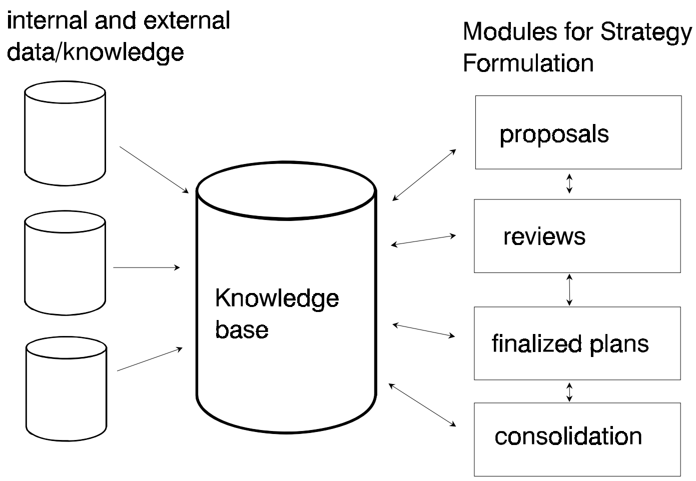
AN ICT ARCHITECTURE SUPPORTING KNOWLEDGE PROCESSES NEEDED FOR STRATEGY FORMULATION

Knowledge from several knowledge domains specified in Table 2 should be generated, stored, shared, and applied to take the steps in the process of strategy formulation: formulating proposals for innovation, reviewing them, making finalized plans for innovation, and consolidating them. We use these steps in the process of strategy formulation as a point of departure for outlining an ICT architecture (cf. Laudon & Laudon, 1997; Tan, 2003; Turban, McLean, & Wetherbe, 2002) for an information system supporting this process. In the literature, ICT architectures are presented as conceptual models, specifying (at a general level) the parts of an ICT infrastructure (applications, databases, technological ICT elements) and their relations. In this chapter we focus on

Table 2. Functions, knowledge domains, and knowledge processes for strategy formulation

Knowledge domains	F1	F3	F4	F5
Goals set by performance and modus operandi of the primary activities in F1	G,A	A	A	
Organizational goals	A	A	A	G,A
Proposals for innovation made by F4		A	G,A	A
Desired goals for F1 based on proposals for innovation		G,A		
Gap between desired and current goals of F1		G,A		
Required capacity for reorganization of F1 activities		G,A		
Actual capacity for reorganization of F1 activities		G,A		
Gap between required and actual capacity for reorganization of F1 activities		G,A		
Reviews by F3 of proposals for innovation		G,A	A	A
Finalized plans for adaptation of organizational goals (a joint F3 and F4 product)		G,A	G,A	A
Regulatory measures to counter the imbalance between F3 and F4		A	A	G,A
Developments in the relevant environment of the organization			G,A	
Norms for balance between F3 and F4				G,A
Actual imbalance between F3 and F4				G,A
Causes of imbalance between F3 and F4				G,A
Experiences with regulatory measures to counter the imbalance between F3 and F4				G,A

Figure 2. Outline of an architecture of an information system supporting strategy formulation



the application and databases parts. An outline of the architecture is presented in Figure 2.

This architecture consists, ideally, of five modules and knowledge and/or databases. The modules (at the right in Figure 2) are applications helping to generate the products of the process of strategy formulation. With the help of these modules, the knowledge from the knowledge domains is applied to produce the proposals, reviews, and (consolidated) plans. The architecture further consists of a central knowledge base in which the knowledge in the knowledge domains necessary for strategy formulation (see Table 2) is stored. This central knowledge base in turn may receive knowledge from other internal and external knowledge and/or databases. Below, we discuss the modules and knowledge bases and their relation to relevant knowledge processes in the course of strategy formulation in more detail.

1. The proposal module
   The main product of this module is a list of innovation proposals and their justification. To produce this list, one should have access to the knowledge in the relevant knowledge domains. To generate this knowledge, the module should have access to external and internal information. For instance, it may have access to a data warehouse by means of a front-end tool, or it may have access to external online databases. Furthermore, the module may have access to a database consisting of previously re-

jected or accepted proposals. The proposals for innovation produced with this module are stored in the central knowledge base.

2. The review module
The input for this module consists of the proposals for innovation. The output is a list of accepted and rejected proposals and the reasons for their acceptance or rejection. To make this list, the module should apply the knowledge in the central knowledge base. This knowledge may be available or may have to be generated. To generate the knowledge, access to several internal and external databases may be required. Also, (external) data on the results of the current PMCs may be input for rejecting or accepting innovations. The review module may benefit from a database with (a classification of) reasons for acceptance or rejection.

3. The finalized-plans module
This module is mainly a means for sharing proposals for innovation (and their reviews) in order to arrive at a finalized plan. It overarches the proposal and review module. By means of this module, results of the review module are shared and applied to revise the proposals (with the aid of the proposal module). The revised proposals are, in turn, used to produce new reviews (with the aid of the review module) and so forth. This module should (a) facilitate sharing proposals and (b) ensure the finalization of an innovation plan. To these ends, this module should support sharing knowledge about

   • the rules for interaction (such as discussion format and deadlines),
   • criteria for imbalance in the discussion,
   • a monitoring function regarding the imbalance,
   • rules and incentives for countering this imbalance, and
   • an overview of the history of the discussion (as well as an overview of previous discussion).

   Implementation could be by means of intranet applications (e.g., an internal discussion site).

4. The consolidation module
This module has as its output the consolidation of (a specific selection of) the innovations on the finalized list of innovations. To make this selection, the arguments used in the previous modules should be scanned and valued. Its main goal is to share the results of the strategy-formulation process with relevant parties in the organization. It should enable sharing knowledge about (a) the selected innovations, (b) the reasons for their selection, and (c) their consequences for the current way of doing business. The process of sharing may benefit from a database with (previously successful) communication formats that can be a part of the consolidation module.

5. The central knowledge base
The central knowledge base consists of all the knowledge in the knowledge domains relevant for strategy formulation. The knowledge base stores the knowledge produced in the modules and supports these modules by servicing them with knowledge relevant to their processes.

Above, an ICT architecture is outlined for an information system supporting strategy formulation. It shows how support should be focused on the products of strategy formulation. Moreover, the focus of the support is on the four knowledge processes involved in the production of proposals, reviews, plans, and consolidations. That is, the application of knowledge leads to proposals for innovation, reviews of these proposals, finalized plans, and consolidation of selected finalized plans. For these products, knowledge from the knowledge domains should be generated, stored, and/or shared. This knowledge is (partly) stored in the knowledge base. The knowledge may be generated by using the four modules and/or by using internal or external databases. Furthermore, knowledge from the knowledge domains may be shared by using connections between the modules.

The description of the architecture specifies the functionality of the different modules in it and how they should be connected. These specifications can be used to select or build the ICT tools to realize the architecture and the knowledge processes it supports.

FUTURE TRENDS

Given the particular outline of the proposed ICT architecture supporting the strategy formulation process, it is possible to link up with current trends that may enhance its performance.

- Developments in the technology for integrating databases (e.g., data warehouse technology) may support the intelligence function in the proposal phase by facilitating the integration and analysis of internal and external knowledge required for strategy formulation.
- Currently, data warehouses are often organized to fit the format of the Balanced Business Scorecard (Kaplan & Norton, 2001). This scorecard is primarily

geared to strategy implementation. The format of the VSM, its related knowledge domains, and steps for formulating strategy may be used to organize data warehouses to fit the requirements for strategy adaptation (Achterbergh, Beeres, & Vriens, 2003).

- Proposing and reviewing proposals for adaptation may be enhanced by the application of computer-aided techniques such as gaming, system dynamics, scenario analysis, and group model building.
- By systematically linking strategy formulation to knowledge management, it becomes possible to enhance the quality of the knowledge processes related to strategy formulation by using acquired insights on improving infrastructures for knowledge management.

CONCLUSION

In this overview, we design an ICT architecture supporting strategy formulation on the basis of the viable system model. By applying the viable system model to the strategy-formulation process, it becomes possible to identify the functions required for strategy formulation, the relations between these functions, and the knowledge required by them.

By identifying the knowledge processes producing and processing this knowledge, and by linking these processes to the functions and the knowledge they require to contribute to the strategy-formulation process, it becomes possible to outline an ICT architecture supporting the processes of generating, retaining, sharing, and applying the knowledge needed for strategy formulation. This architecture consists of five modules dedicated to proposing, reviewing, finalizing, and consolidating strategy changes and related knowledge databases containing knowledge in the knowledge domains required for strategy formulation.

REFERENCES

Achterbergh, J. M. I. M., Beeres, R., & Vriens, D. (2003). Does the balanced scorecard support organizational viability? *Kybernetes, 32*(9/10), 1387-1404.

Achterbergh, J. M. I. M., & Vriens, D. (2002). Managing viable knowledge. *Systems Research and Behavioral Science*, *19*, 223-241.

Beer, S. (1979). *The heart of enterprise*. Chichester, England: Wiley.

Beer, S. (1981). *Brain of the firm*. Chichester, England: Wiley.

Beer, S. (1985). *Diagnosing the system*. Chichester, England: Wiley.

Bukowitz, W. R., & Williams, R. L. (1999). *The knowledge management fieldbook*. Edinburgh, Scotland: Pearson.

Davenport, T. H., & Prusak, L. (1998). *Working knowledge*. Boston: Harvard Business School Press.

Espejo, R., Schumann, W., Schwaninger, M., & Billello, U. (1996). *Organizational transformation and learning*. New York: Wiley.

Kaplan, R., & Norton, D. (2001). *The strategy-focused organization: How balanced scorecard companies thrive in the new business environment*. Boston: Harvard Business School Press.

Laudon, K. C., & Laudon, J. P. (1997). *Management information systems* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Tan, D. S. (2003). *Van informatiemanagement naar informatie-infrastructuurmanagement*. Leiderdorp: Lansa.

Turban, E., McLean, E., & Wetherbe, J. C. (2002). *Information technology for management* (3rd ed.). New York: Wiley.

KEY TERMS

ICT: Information and communication technology. ICT can be used to indicate the organization’s technological infrastructure (comprising of all hardware, software, and telecommunications technology) and to indicate one or more specific collections of hardware, software, and telecommunications technology (i.e., one or more ICT applications).

ICT Architecture: The ICT architecture provides a conceptual model, specifying (at a general level) the parts of an ICT infrastructure (applications, databases, technological ICT elements) and their relations. In this chapter we concentrate on the application and databases parts.

Knowledge Domain: the knowledge related to defining, recognizing, and solving a specific problem.

Knowledge Processes: In the literature, one often finds four knowledge processes: (a) generating knowledge, (b) sharing knowledge, (c) storing knowledge, and (d) applying knowledge.

Strategy: In the literature, many definitions are given. A possible definition is the desired portfolio of product-market-technology combinations of an organization.

Strategy Formulation: The process by means of which the desired portfolio of product-market-technology combinations is defined and updated. This process can be modeled using the viable system model consisting of four steps: defining proposals for innovation, reviewing these proposals, finalizing proposals, and consolidating finalized proposals.

Viable System Model: This model is developed by Beer (1979, 1981) and specifies the necessary and sufficient functions organizations should possess to maintain a separate existence in their environment.

Viability: Viability is the ability of a system "to maintain a separate existence." Most organizations are continuously trying to maintain their viability.

Java 2 Micro Edition for Wireless Enterprise

Kin Choong Yow
Nanyang Technological University, Singapore

Nadia N. Moertiyoso
Hewlett-Packard Singapore, Singapore

INTRODUCTION

For the last couple of years, the wireless industry has been experiencing tremendous growth. Wireless devices have become more intelligent and are providing a new notion of communication. It is now possible to conduct business using the wireless network that will greatly improve the speed and quality of the business. Examples of business decisions across the wireless Internet include:

- A sales manager will be able to browse and download the latest price and stock availability of products during the journey to a client meeting.
- Support engineers, couriers, and delivery services will be able to manage their schedules better by organizing their orders on their cell phones. The headquarters will also be able to monitor their workers in the field.

This article will discuss the issues concerning the development of wireless applications using Java 2 Micro Edition (J2ME). This article will also explain what J2ME is and discuss the guidelines and technical aspects to implement wireless enterprise applications using J2ME.

BACKGROUND

Introduction to Java 2 Micro Edition

Java 2 Micro Edition (J2ME), from Sun Microsystems, is designed to run on consumer devices and electronic appliances, including wireless devices such as cell phones and Palm PDAs (Raju, 2000). J2ME provides a way to exploit the processing power on the mobile device better by running the code on the device itself. Therefore, it provides better network implementation, better graphical user interface, and local database management.

The Java virtual machine for consumer devices is known as the Kilo Virtual Machine (KVM), which is a complete Java runtime environment for small devices. KVM was designed to be small, with a static memory footprint of 40 – 80 kilobytes. KVM is derived from a research system called Spotless at Sun Microsystems Laboratories.

J2ME Profiles and Configurations

J2ME is divided into *configurations* and *profiles* (Ashri et.al., 2001). Configurations are specifications that detail a virtual machine and a base set of APIs that can be used with a certain class of device. A profile builds on a configuration but adds more specific APIs to make a complete environment for building applications. While a configuration describes a JVM and a basic set of APIs, it does not by itself specify enough detail to build complete applications.

Configurations

Mobile devices such as cell phones, pagers, organizers, etc., are diverse in form, functionality, and features. For these reasons, the J2ME supports minimal configurations of the Java Virtual Machine and APIs that capture the essential capabilities of each kind of device (Feng & Zhu, 2001). J2ME configuration defines a set of horizontal APIs for a family of products that have similar requirements.

Profiles

A profile is a set of vertical APIs that reside on top of the configuration to provide domain specific capabilities. Currently, there is one profile defined and implemented, the Mobile Information Device Profile (MIDP) (Sun Microsystems, 2001). Other profiles, which are in the works, include the PDA profile and the RMI profile.

Table 1 shows the current and anticipated configurations, profiles, and APIs.

Comparison of J2ME versus WAP in Enterprise Application

Phone.com (now Open Wave), a merged venture with Software.com, first introduced the Wireless Application Protocol (WAP) in 1995. WAP is an open standard that

Figure 1. Java technology map

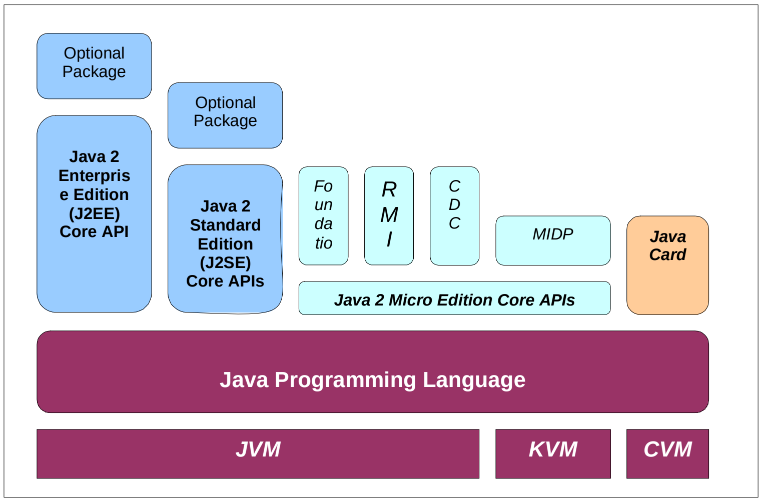


Table 1. J2ME profiles and APIs

Configurations		
JSR 30	CLDC 1.0	Connected, Limited Device Configuration
JSR 139	CLDC 1.1	Connected, Limited Device Configuration 1.1
JSR 36	CDC	Connected Device Configuration
Profiles		
JSR 37	MIDP 1.0	Mobile Information Device Profile
JSR 118	MIDP 2.0	Mobile Information Device Profile 2.0
JSR 75	PDAP	PDA Profile
JSR 46	FP	Foundation Profile
JSR 129	PBP	Personal Basis Profile
JSR 62	PP	Personal Profile
JSR 134		Game Profile
APIs		
JSR 66		RMI Optional Package
JSR 80		Java USB API
JSR 82		Java APIs for Bluetooth
JSR 120		Wireless Messaging API
JSR 135	MMAPI	Mobile Media API
JSR 169		JDBC Optional Package for CDC/FP
JSR 179		Location API for J2ME
JSR 180		SIP API for J2ME
JSR 184		Mobile 3D Graphics API for J2ME

enables easy delivery of information and services to mobile users. WAP is intended to address the need to access the Internet from handheld devices such as cell phones and PDAs. It defines a set of standard components that enables communication between mobile terminals and network servers, including the WAP programming model. WAP has received wide support in the wireless industry.

J2ME and WAP are complementary technologies. J2ME is not going to replace WAP. WAP is comparable to HTML and Web browsers on desktop computers, while J2ME is comparable to desktop Java applications. However, in the enterprise application situation, J2ME is found to be superior to WAP.

Limitations of WAP

WAP is limited in some ways, including:

- Network disconnections: Mobile device communication over wireless networks tend to suffer from greater occurrences of disconnections due to a variety of problems, such as limited cell coverage, loss of signal, limited battery life, etc. Current frameworks only provide partial support for disconnected operations.
- Constant airtime: WAP requires constant airtime in order to operate. When combined with the previous point, this may cause disruptions in the application's usage. This constant airtime requirement is also economically inefficient.
- Lack of security: The infrastructure of WAP forces the information to be passed into a WAP Gateway before it is presented to a cell phone. Therefore, there is a security loophole in the architecture, since the WAP Gateway is usually controlled by a third party (network carrier). In addition, security issues may arise during the translation of pages from the server to the WAP Gateway.
- Primitive user interface: The user interface for WAP-enabled sites is primitive when compared to traditional Web sites. Devices that might be capable of displaying a richer graphical user interface (GUI) through higher screen resolution are not being exploited.

Benefits of J2ME

J2ME has the following benefits:

- Local Computing Power: Browser-based applications can often be an adequate solution for constructing software for devices, allowing Web access with minimal configuration. However, it is more valuable to be able to exploit the device's resources more by executing code on the device's processor and using the device's memory to run the application locally.
- Multithreading: J2ME allows an application to use multithreading. This provides a mean for application developers to create a more complex logic.
- User personalization: Browser-based applications often lack sophistication in user interface compared to applications that take greater advantage of local operating systems. This sophistication can be leveraged by J2ME with better manipulation of device-side services.
- Offline operation: J2ME provides a solution that can run on or off the network, something that a browser solution can do only in a very unsophisticated way with local browser page caching.
- Lower network usage and server load: In a WAP-based solution, the server is responsible for generating display markup. This requires a round-trip every time the interface changes. In contrast, a MIDP client's interface is contained within the device, so it can operate even when disconnected.
- Security: J2ME offers an enhanced security model. It provides connection end-to-end from client device to company servers and does not require gateways such as in WAP.
- Connectivity with distributed system: J2ME provides a better way to talk to the back-end system. Applications written in J2ME are able to talk in XML, a language that is suitable for cross-platform and distributed system applications (Castro, 2001).

WIRELESS ENTERPRISE APPLICATION DESIGN USING J2ME

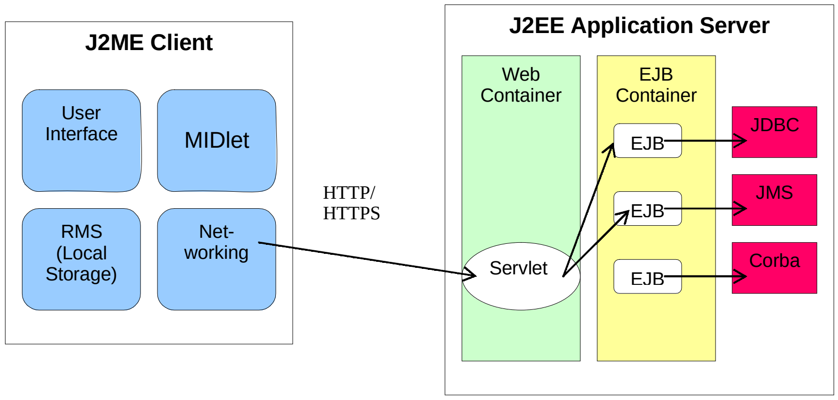
Overall Enterprise Application Architecture

An enterprise application usually does not consist of the wireless module alone. A wireless application usually acts as an interface module that connects to an existing or running system. This section examines the overall enterprise architecture, including the wireless component.

Figure 2 shows the overall logical architecture of a typical Java wireless enterprise application, spanning a J2ME device and a J2EE application server. The use of J2EE for the back-end system is, of course, optional. The J2ME MIDP is flexible enough to talk to any back-end system that supports HTTP transport.

The logical architecture of an enterprise application

Figure 2. High-level architecture of a Java wireless enterprise application



serving wireless clients is similar to that of a canonical J2EE application (Girdley et al., 2002):

- An application client implemented using MIDP, or a *MIDlet* client, provides the user interface on the mobile device. This MIDlet communicates with a Java servlet, usually via HTTP, and over a secure channel (HTTPS) when necessary.
- The servlet interprets requests from the MIDlet, and in turn, dispatches client requests to EJB components (Adatia et al., 2001). When the requests are fulfilled, the servlet generates a response for the MIDlet.
- The EJB components, or *enterprise java beans*, encapsulate the application's business logic. An EJB container provides standard services such as transactions, security, and resource management so that developers can concentrate on the business logic.
- The servlet and EJB components may use additional APIs to access enterprise information and services. For example, they may use the JDBC API to access a relational database, or the JavaMail API to send e-mails to a user.

Developers can write networked application clients that connect to wireless services using standard networking protocols. MIDP requires all devices to support HTTP, the same protocol that Web browsers use (Mahmoud, 2000).

The physical architecture of an entire enterprise application (e.g., Customer Support System) is shown in Figure 3. Client entities (e.g., Support Engineers) access the system through the wireless modules while Helpdesk, System Administrators, and Managers access the system through the Web modules. The back-end system uses Web servers, application servers, and database servers.

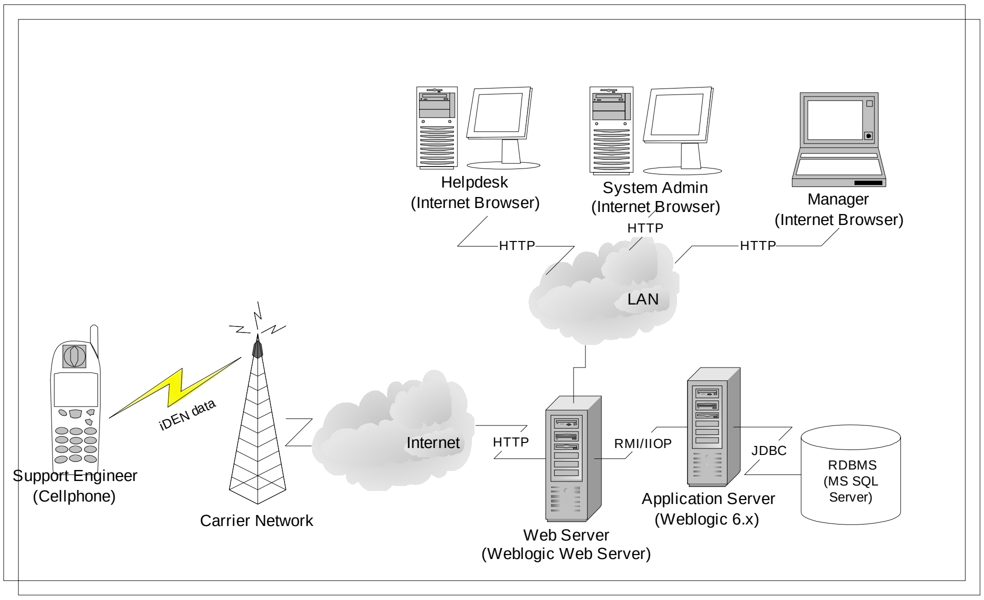
The wireless module (J2ME) runs in a J2ME-enabled cell phone on GPRS or other similar packet data network. It accesses the back-end system using HTTP (Hypertext Transfer Protocol).

Considerations in Designing the J2ME Wireless Module

Enterprise applications usually impose a requirement on the client devices to communicate with the servers to query and manipulate data. In a wireless scenario, the wireless application presents new problems to database query management due to the mobility of the users, power and resource restrictions, and the limited bandwidth of the wireless connection (Bukhres et al., 2002). These challenges are the major issues in designing the architecture of the application and are summarized below:

1. Limited bandwidth: the wireless network operates at a much lower rate than a wired network.
2. Network disconnection: the wireless network is also more susceptible to network disconnections compared to a wired network. The framework must therefore provide a way to ensure that data in the mobile devices will not be lost due to this.
3. Limited computing power and memory resources: mobile devices have much lower computing power than a standalone PC. Therefore, operations that require a lot of computing power must be distributed to other nodes in the system.
4. Working online and off-line: There are many reasons that require the application to be able to work off-line. Some of these reasons include network

Figure 3. Architecture of enterprise applications



interruptions and weak signals. One of the more economical reasons is to save airtime charging.

5. Synchronization and data consistency: The framework must also provide a way to maintain the consistency of the data in the mobile device and the server.

M-Commerce Considerations in the J2ME Application

The J2ME module can easily be extended to transfer financial transaction information on the move (m-commerce). Bank account information and the instruction to transfer funds, etc., are all treated as data to be sent over the wireless network.

However, as in the case of electronic commerce (e-commerce), the security of the transaction is the primary concern (Claessens et al., 2002). The following general security requirements apply to m-commerce applications:

- Confidentiality: Need to ensure that only authorized entities have access to the content of the exchanged information, e.g., an eavesdropper should not be able to find out what transactions a particular user is executing.
- Entity authentication: Client entities should be sure that they are communicating with the real back-end entity, before sending sensitive information to it while the back-end entity should also know the identity of the client entity before processing its transactions.
- Data authentication: This involves data origin authentication and data integrity. The back-end (at least) should be able to detect manipulations (e.g., insertion, deletion, and substitution) and replay of data by unauthorized parties.
- Non-repudiation: This is to prevent an entity from denying previous commitments or actions, e.g., the back-end entity should be able to prove to a third party that a client entity performed a certain transaction, in case the client entity denies having performed it.

To achieve the above requirements, cryptography techniques (Jon, 2001) should be used. Cryptography allows information to be sent in a secure form such that only the intended recipient is able to retrieve the information. Practically all modern cryptographic systems make use of a key to encrypt the information. There are two key-based cryptography: symmetric key encryption and public key encryption:

1. Symmetric key encryption: Symmetric algorithms use the same key to both encrypt and decrypt the plaintext. Hence, both the sender and the receiver need to have a copy of the same key. The key is often called a secret key, as it can only be known by the communicating parties in order for the ciphertext to be secure.

2. Public key encryption: Public key cryptography, also known as asymmetric cryptography, involves a pair of keys called the public key and private key (Benantar, 2002). The encryption key (public key) is used to encode a message, but this message can only be decoded by using the corresponding decryption key (private key). Conversely, a message encrypted using the private key can only be decrypted by the public key. This system allows the recipient to make the public key widely available, so that anyone can send an encrypted message to the recipient.

In J2ME, crypto-APIs such as the Rivest-Shamir-Adleman (RSA) algorithm (RSA Security, 2002) is already available. This greatly reduces the time for developing secure m-commerce applications.

FUTURE TRENDS

The trend is moving increasingly to allow developers to enhance the look and feel of the mobile device. In the future, new APIs may be added such as:

- Phone-centric APIs
  - External Display Access (Caller ID Display)
  - RF Enabling
  - Call Receiving (Answer an incoming phone call with Caller ID)
  - Call Initiation (Place outgoing calls)
- Data Manipulation APIs
  - Gzip/ZIP compression and decompression
  - Secure File
- Graphics APIs
  - MIDP 2.0 Gaming
  - 3D API
  - Additional Image Formats
- JPEG
- WBMP
- PNG w/Alpha Transparency

CONCLUSION

In this article, we have described the use of Java 2 Micro Edition for developing wireless applications. We have given a brief description of the features of J2ME and a comparison with WAP on the suitability of developing Enterprise Applications. We have also outlined the various considerations in developing wireless enterprise using J2ME and their associated m-commerce considerations.

A complete enterprise model consisting of wireless modules, back-end modules, and Web-site modules can easily be developed using J2ME. Cryptographic APIs are already available in J2ME to build secure m-commerce applications. J2ME is an obvious choice because it is portable and designed to run on consumer devices.

REFERENCES

Adatia et al. (2001). *Professional EJB*. UK: Wrox Press.

Ashri et al. (2001). *Professional Java Mobile Programming*. UK: Wrox Press.

Benantar, M. (2002). *Introduction to Public Key Infrastructure for the Internet*. Singapore: Prentice Hall PTR.

Bukhres et al. (2002). A Proposed Mobile Architecture for Distributed Database Environment. Retrieved from the World Wide Web at: http://citeseer.nj.nec.com/43790.html

Castro, E. (2001). *XML for the World Wide Web*. USA: Peachpit Press.

Claessens, J., Dem, V., Cock, D. H., Preneel B., & Vandewalle, J. (2002). On the security of today's online electronic banking systems. *Computer Security and Industrial Cryptography (COSIC), 21(3),* 257-269.

Feng, Y., & Zhu, J. (2001). *Wireless Java Programming with J2ME*. USA: SAMS.

Girdley, M., Woollen, R., & Emerson, S.L. (2002). *J2EE Applications and BEA WebLogic Server*. New Jersey: Prentice Hall.

Mahmoud, Q. (2000). MIDP Network Programming using HTTP and the Connection Framework. *Java Wireless Developer*. Retrieved November 2001 from the World Wide Web at: *http://wireless.java.sun.com/midp/articles/network/*

Morton, S. & Bukhres, O. (1997). Utilizing Mobile Computing in the Wishard Memorial Hospital Ambulatory Service. *In 12th ACM Symposium on Applied Computing (ACM SAC'97)*.

Raju, S. (2000). *Java Programming for Wireless devices using J2ME – CLDC/MIDP APIs*. Sun Microsystems.

Sun Microsystems. (2001). *MIDP APIs for Wireless Applications: A Brief Tour for Software Developers*. Sun Microsystem.

Sun Microsystems. (2002). Introduction to Wireless Java. *Wireless Java Technology*. Retrieved March 2002 from the World Wide Web at: *http://wireless.java.sun.com/getstart/*

KEY TERMS

Enterprise Applications: Enterprise Applications provides computer application support for the full life cycle of business application development and system support. Involves in-depth understanding of business processes and workflow.

Interoperability: The property of software applications that will enable diverse applications to interact and share data seamlessly.

Virtual Machine: A software system that enables a collection of heterogeneous computers to be used as a coherent and flexible concurrent computational resource.

Application Programming Interfaces (APIs): Programming tools that provide developers with a simple, consistent mechanism for extending the functionality of an application and for accessing existing computing systems.

Memory Footprint: The memory usage during the execution of an application. It depends on what kind of processor you intend to use and what components of the system you include.

Multithreading: Programs written to execute in a parallel manner, rather than a serial or purely sequential one, in order to take advantage of multiple processors.

Distributed System: A system made up of components that may be obtained from a number of different sources, which together work as a single distributed system providing the run-time infrastructure supporting today's networked computer applications.

Journalism Online in Peru

Antonio Díaz-Andrade
Escuela de Administración de Negocios para Graduados - ESAN, Peru

INTRODUCTION

Online journalism dates back to the end of the 1970s, when Knight-Ridder launched an initiative to develop a video-text service in the United States, which it later dropped, in 1986, after realizing enormous losses. In 1988, Knight-Ridder bought *Dialog Information Services, Inc*.; only a year later, the first signs of success appeared. By the end of the 1980s, Gannet launched a daily news piece in text format. In 1992, *The Chicago Tribune* became the world's first daily to launch an electronic version of its newspaper. In 1993, Knight-Ridder started publishing what would eventually become one of the paradigms of electronic journalism, the *San Jose Mercury Center*. By 1994, the major newspapers in the United States offered readers an online version (Díaz & Meso, 1998). Now, Internet users can read newspapers, listen to the radio, and watch TV from anywhere, anytime (McClung, 2001).

BACKGROUND

Electronic journalism stands out for four features: it uses multimedia resources; it breaks down the sequence of information through hypertext and hypermedia; it breaks down periodicity, because it allows for information to be accessed at any time; and it is interactive (Díaz & Meso, 1998; Cohen, 2002). Interactivity gives an opportunity for readers to disagree with the publishers or contribute to the writers' work by incorporating readers' opinions (Giles, 2000; Cohen, 2002).

Many sites relating to printed media have evolved from pages that initially were simply reproductions of the printed version into a dynamic environment that is constantly updated (Pack, 2001). In this sense, Internet journalism principles should not be different from those governing traditional channels. Digital-era journalists must find the way to reach a global audience without compromising the values of truthfulness, of reliable sources, and of independence (Pavlik, 2000).

Nowadays, the Internet opens the door to new, purely virtual media, although they must still walk a long way before they gain enough credibility. Additionally, they must rise to the challenge of a huge infrastructure that has already been put in place by consolidated newsgroups and overcome limited access to capital (Giles, 2000; Uy, 2001). Purely virtual media must carefully analyze their projects' feasibility. The collapse of Taiwan's virtual *Tomorrow Times* information medium—despite 1.8 million daily page visits—poses a question about the viability of sustaining the high cost of generating news (Ling & Guyot, 2001), when no journalistic infrastructure has been developed in the physical world. So, online journalists must reconcile money-making and news interests (Cohen, 2002).

Although the cost of preparing a news feature has remained almost unaltered, and although presenting news in a multimedia format (Fulton, 2000) may slightly increase costs, it is clearly cheaper to prepare a page and make it available to millions of people around the world over the Internet than to distribute it on a printed format (Small, 2000 and Singer, 2001).

Finally, given the phenomenon of ethnocentricity (Priess, 2000), publishers may find it convenient to specialize on reporting about a reality that is best known to them (Singer, 2001).

PERUVIAN ONLINE JOURNALISM

To evaluate Web journalism in Peru, the most important print, broadcast, and purely virtual media were analyzed. The criteria to select the media were their readership or rating in the physical world and their number of visits in the virtual world.

The following printed media were chosen:

- *El Comercio* is Lima's best recalled and most widely read daily (Apoyo Opinión y Mercado, 2003). It has a reputation for being the most truthful, entertaining, and the best at covering local political news events (Apoyo.com, 2001). It is read by an average 574,700 people in Lima (Instituto Cuánto, 2001).
- *Gestión* was Peru's first written medium to produce an online version. Its printed edition, targeting the business community, reaches an average 27,500 readers in Lima (Instituto Cuánto, 2001).
- *La República* was Peru's second printed medium to publish an electronic version. It is regarded as the second most truthful and is also mentioned as the second most reliable daily in providing local political news coverage (Apoyo.com, 2001). It is read by

an average 171,300 readers in Lima (Instituto Cuánto, 2001). At this time, it is regarded as the second most widely read online Peruvian daily (Apoyo Opinión y Mercado, 2003).

Radio broadcasters in this study comprise those that broadcast over the Web but focus on news programming:

- *Radioprogramas del Perú*, *RPP Noticias*, enjoys a spontaneous recall rate of 27% among all radio broadcasters and 62% among news radio broadcasters. Its annual average half-hour audience in Lima reaches 95,100 listeners (Insituto Cuánto, 2001). It is one of the most widely heard radio broadcasts (Apoyo Opinión y Mercado, 2003).
- *CPN Radio*, *Radio Cadena Peruana de Noticias*, has an annual average half-hour audience of 22,900 listeners in Lima (Insituto Cuánto, 2001). It is the second most widely heard news radio broadcast in Peru (Apoyo Opinión y Mercado, 2003).

Rather than displaying full and updated information, Peruvian television networks on the Web underscore programming information. Their Web sites are closer to an institutional page than to an information page, so they were not included in this study.

Of the three Peruvian information media lacking a print, radio, or television matching part, only www.peru.com and www.primerapagina.com.pe participated in this study; the latter plummeted shortly after this research; the third did not answer.

A preliminary evaluation of the selected media Web sites was conducted preceding in-depth interviews with their managers; a subsequent examination on the media Web sites complemented the analyses.

In the following pages, a brief explanation is provided about the origins, main goals, content, audience, and income sources of the studied Peruvian Internet media.

El Comercio

Founded in Lima in 1839, it published the company's institutional page on the Internet in 1996 and in 1997 started publishing the www.elcomercioperu.com.pe electronic daily; in May 2001, it organized an independent area charged with the Web publication. News is updated on the Web using information provided by the daily's reporters and input from news agencies. Its Web site gets an average 11 million monthly hits, mostly from Peruvian expatriates. Information sections in the daily's electronic version include national and world affairs, business, entertainment, politics, and sports; its most widely read pages are the front page and local soccer news.

The revenues of www.elcomercioperu.com.pe come from advertising by some announcers, sale of news to the telephone operators who distribute the news to their cellular telephone customers, as well as preparation of private circulation newsletters for other companies. The daily's electronic version has created a portal providing, among other services, the www.ec-store.com.pe virtual shop as an additional source of revenue. At the same time, an Internet public kiosks network was put in place around the country. Additionally, *El Comercio* offers free e-mail and a PDA news service.

El Comercio's management has made a decision to make the newspaper a Latin American leader. Comprised in their strategy to reach their goal is the effort to provide Internet-based news services.

Gestión

Gestión was Peru's first written medium to produce an online version when it launched its electronic site www.gestion.com.pe in September 1996. An economics, finance, and business newspaper, *Gestión* first circulated in 1990. Its specialized reports go hand in hand with information about political events and world affairs.

From the daily's Web site, it is possible to access other media run by the corporation: *CPN Radio* (www.cpnradio.com.pe) information broadcast station and its *Gestión Médica* (www.gestion.com.pe/GM) weekly health publication. Most of its visitors are businessmen from abroad who have business and interests in Peru. Articles on political issues are the most frequently visited, followed by economic reports. *Gestión*'s contents are prepared with materials gathered by its reporters, together with news from news agencies. Information in www.gestión.com.pe is posted on the Web and updated by *CPN Radio* personnel. *Gestión* also offers a free e-mail news service twice a day.

La República

Founded in Lima in 1981, it has as its objective to inform and become an opinion maker for the Peruvian public. Since 1995, it uses satellites to publish local editions in other cities around Peru. In October 1996, *La República* launched its electronic version (www.larepublica.com.pe).

Since its Web site launching, the number of opinion letters it gets through e-mails from readers outside Peru has grown steadily. Some visitors at www.larepublica.com.pe have become the newspaper's foreign correspondents who "add a Peruvian flavor" to international reports. Foreign readers account for 70% of the electronic edition's readership. All of *La República*, including its supplements, is published on its Web site;

the most visited pages cover current political affairs and sports. News is not updated during the day.

Radioprogramas del Perú

Initially conceived in 1963 as an entertainment radio broadcaster, in 1979 it became a round-the-clock news broadcaster. At the end of 1999, the www.rpp.com.pe site was added to the original product. At this time, *RPP Internet* gets an average 12 million monthly visits; about 40% of them originate in Peru. At its Web site, visitors find the latest news in the "news by the minute" section, as well as local affairs, political, sports, world affairs, finance, cultural, and entertainment sections. There is also the option to participate in opinion surveys and a discussion forum concerning current affairs as well as to access the audio-on-demand option. Likewise, the direct *RPP Noticias* audio broadcast can be heard on www.rpp.com.pe. Additionally, wireless application protocol (WAP) technology enables telephone operators' clients to access news.

It is widely held that launching this service has had a positive influence on RPP Group's image building by creating among both listeners and announcers a perception of leadership, while strengthening the Group's competitive standing in terms of its capacity to reach a wider target audience that comprises both radio listeners and Web users.

CPN Radio

This news radio chain has been owned by Corporación Gestión since 1998. Motivated by positioning itself and creating an image before the public vis-à-vis *Radioprogramas del Perú*, *CPN Radio*, through an agreement with Terra Networks, launched the www.terra.com.pe/cpn site in June 2000. In September 2001, a new page was launched at www.cpnradio.com.pe. *CPN Radio* and *Gestión* newspaper share their general information sources; so, to launch the Internet project, professional journalists within the organization were reassigned. Most visitors to www.cpnradio.com.pe originate in Peru, and these visitors can also listen to radio programming in real time. In its publication, www.cpnradio.com.pe includes headlines and the political, economic, financial, business, world affairs, show business, sports, and culture sections; political features are most often followed. It also has a "last minute news" flash section, complete radio programming, and links to articles written by the broadcaster's columnists.

www.peru.com

Interlatin Corporation—a Peruvian company holding similar domains (www.colombia.com, www.bolivia.com, and www.futbolargentino.com)—owns this site that seeks to develop portals that provide information and specific services adapted to individual countries. The site www.peru.com has been operational since 1998; its name became the company's main tool in luring traffic toward its site. The site gets around 16 million visits monthly, mostly of Peruvian expatriates, to its different news sections dedicated to delivering information about Peru and offering other services that include a browser, chat rooms, free e-mail addresses, job searches, a communication media directory, messaging to cellular telephones, access to music radio broadcasts, and electronic postcards.

Interlatin Corporation has two main sources of revenue: its www.iquiero.com virtual shop and its travel agency www.peru.com/travel. Additionally, www.peru.com accepts advertising and provides Web page design services; likewise, through agreements with telephone operators, it sells news to be distributed among their users.

Becoming an Internet news leader is the goal of www.peru.com. Its journalists recognize that they do not have the entire journalistic infrastructure available to traditional media and, therefore, resort to their own inventiveness in finding and publishing the most recent information.

FUTURE TRENDS

Some of the Peruvian media moved onto the Internet seem more clearly determined to explore new options to increase their revenues and earnings, while others simply wish to expand their readership or number of listeners by drawing international audiences. However, they all seem persuaded that there is no turning back, and they have entered the Web never to escape from it. In contrast, those media that do not have a traditional media foundation have no choice but to yield satisfactory economic results that will allow them to survive. News pieces and reports prepared for traditional media are the raw material for the articles published by conventional media on the Web. Exclusively virtual media had to draw a press team out of the blue and have made them responsible for searching news and then publishing their reports.

Initial plans had to be modified, as Internet media identified difficulties and opportunities along the way. Purely virtual Peruvian media are hard put to make any short-term inroads into traditional channels. Instead, they must try and develop business lines that can sustain the work of their news teams. Using Internet resources to offer a varied and interesting range of content options as well as to create virtual communities that may be served with alternative products is an option deserving consideration.

Likewise, increasing participation of Peruvian information groups on the Web opens the possibility for them to reach a larger number of users and thus develop new businesses in the field of information. Notwithstanding, the Peruvian media must acknowledge the restrictions that apply to the Internet in Peru.

Finally, when easy-to-use, flexible, and portable devices are sufficiently developed and widespread, all featuring multimedia capabilities, newspeople will have to be ready to meet the ever faster pace of information needs in modern societies. In the meantime, the number of Peruvians that go to the Internet to stay abreast of events increases constantly.

CONCLUSION

A common feature among all aforementioned media is that the most visited sections are the Peruvian political current affairs and sports pages. In almost all media, most visits come from abroad, likely because of the strong familial, cultural, and social ties of the Peruvian expatriates, which constitute an expression of ethnocentricity, as was explained previously. Nonofficial figures show that Peruvian expatriates may reach 2.5 million. In the same way, it should be noticed that those living overseas are likely to have better conditions for accessing the Web than those in Peru, where the information and communication infrastructure is still poor, although it shows one of the highest rates of public access to the Internet in Latin America, thanks to the phenomenon of the "cabinas públicas" (OSIPTEL, 2003; Salvador, Sherry, & Urrutia)

Using multimedia resources is another characteristic of online journalism. All of the studied news media use text and photographs; www.peru.com, *CPN Radio*, and *Radioprogramas del Perú* broadcast directly over the Web, while the latter is the only one that offers audio on demand. *El Comercio's* Web site and www.peru.com present some news through video.

Except for the Internet media owned by Corporación Gestión, *Gestión* daily, and *CPN Radio*, all the analyzed media make available advertisement space for commercial announcers on their Web sites. However, it is clear that advertising, no matter how high the traffic through the site, does not constitute, in itself, a sufficiently sound source of revenue to cover operating costs.

REFERENCES

Apoyo Opinión y Mercado. (2003, April). *Hábitos y actitudes hacia la prensa escrita*. Lima.

Apoyo Opinión y Mercado. (2003, April). *Hábitos y actitudes hacia la radio*. Lima.

Apoyo.com. (2001, April). *Actitudes hacia la prensa escrita*. Retrieved December 3, 2001, from *http://www.apoyo.com/infor_util/inv_mercados/igm_prensa_042001.html*

Cohen, E. L. (2002, December). Online journalism as market-driven journalism. *Journal of Broadcasting & Electronic Media, 46*(4), 532–548.

Díaz, J., & Meso, K. (1998). Desarrollo del periodismo electrónico. *El profesional de la información, 7*(12), 4–11.

Fulton, K. (2000). News isn't always journalism. *Columbia Journalism Review, 39*(2), 30–35.

Giles, B. (2000). Journalism in the era of the web. *Nieman Reports, 54*(4), 3.

Instituto Cuánto. (2001, September). *Anuario estadístico: Perú en números 2001*. Lima.

Ling, C., & Guyot, E. (2001, February 23). Taiwan newspaper closes web edition, cites funding woes. *Wall Street Journal*.

McClung, S. (2001). College radio station web sites: Perceptions of value and use. *Journalism and Mass Communication Educator*, *56*(1), 62–73.

OSIPTEL. (2003). *Indicadores de Internet*. Retrieved November 1, 2003, from http://www.osiptel.gob.pe/Index.ASP?T=P&P=2642

Pack, T. (2001, September–October). All the news that's fit to digitally print. *Link-up, 18*(5), 16.

Pavlik, J. (2000). The journalist: A disappearing species in the online world? *The UNESCO Courier, 53*(2), 34.

Priess, F. (2000). Los medios de comunicación en los conflictos armados. In A. Cacua & F. Priess (Eds.), *Ética y responsabilidad: reflexiones para periodistas*. Bogotá: Editora Guadalupe.

Salvador, T., Sherry, J. W., & Urrutia, A. E. Less cyber, more café: Design implications for easing the digital divide with locally social cyber cafes.

Singer, J. B. (2001). The metro wide web: Changes in newspapers's gatekeeping role online. *Journalism and Mass Communication Quarterly, 78*(1), 65–80.

Small, J. (2000). Economics 101 of Internet news. *Nieman Reports, 54*(4), 41–42.

Uy, E. (2001). Reporters on superhighway meet roadblocks. *News Media and the Law, 25*(3), 49–50.

KEY TERMS

Direct Audio Broadcasting: The possibility of listening to the transmission of radio broadcast programming directly from the Web.

Electronic Journalism: The use of every Internet resource to disseminate information on an ongoing, updating basis.

Ethnocentric Content: Pieces of information provided by local journalists of particular interest for locals who find sufficient elements for understanding it. On the contrary, local and specific content could be strange to overseas readers.

Information and Communication Technology Indicators (ICTs): Indicators that show the extent of the development of a information and communication infrastructure. They include fixed telephone density, mobile telephone density, personal computer density, Internet access, and host density, usually expressed as a ratio of inhabitants.

Internet Public Kiosk: A booth that provides Internet access in return for a payment on a time basis.

Journalism: The job of compiling and disseminating information.

Multimedia: Combined use of several media, such as text, audio, and images. The Internet provides the ability to use one, some, or all of them to deliver content.

Purely Virtual Media: Those media that do not have any presence in the physical world, neither printed nor broadcasted.

K

Knowledge Discovery Solutions for Intelligent Enterprises

Nilmini Wickramasinghe
Cleveland State University, USA

Sushil K. Sharma
Ball State University, USA

INTRODUCTION

The exponential increase in information—primarily due to the electronic capture of data and its storage in vast data warehouses—has created a demand for analyzing the vast amount of data generated by today's organizations so that enterprises can respond quickly to fast changing markets. These applications not only involve the analysis of the data but also require sophisticated tools for analysis. Knowledge discovery technologies are the new technologies that help to analyze data and find relationships from data to finding reasons behind observable patterns. Such new discoveries can have profound impact on designing business strategies. With the massive increase in data being collected and the demands of a new breed of intelligent applications like customer relationship management, demand planning and predictive forecasting, the knowledge discovery technologies have become necessities to providing high performance and feature rich intelligent application servers for intelligent enterprises. The new knowledge based economy entirely depends upon information technology, knowledge sharing, as well as intellectual capital and knowledge management.

Knowledge management (KM) tools and technologies are the systems that integrate various legacy systems, databases, ERP systems, and data warehouse to help facilitate an organization's knowledge discovery process. Integrating all of these with advanced decision support and online real-time events would enable an organization to understand customers better and devise business strategies accordingly. Creating a competitive edge is the goal of all organizations employing knowledge discovery for decision support. They need to constantly seek information that will enable better decisions that in turn generate greater revenues, or reduce costs, or increase product quality and customer service. Knowledge discovery provides unique benefits over alternative decision support techniques, as it uncovers relationships and rules, not just data. These hidden relationships and rules exist empirically in the data because they have been derived from the way the business and its market work.

The following is a brief synopsis of the major tools and major considerations required to enable an organization to go through the key processes of knowledge sharing, knowledge distribution, knowledge creation, knowledge capture and codification as well as embracing effective knowledge management (KM).

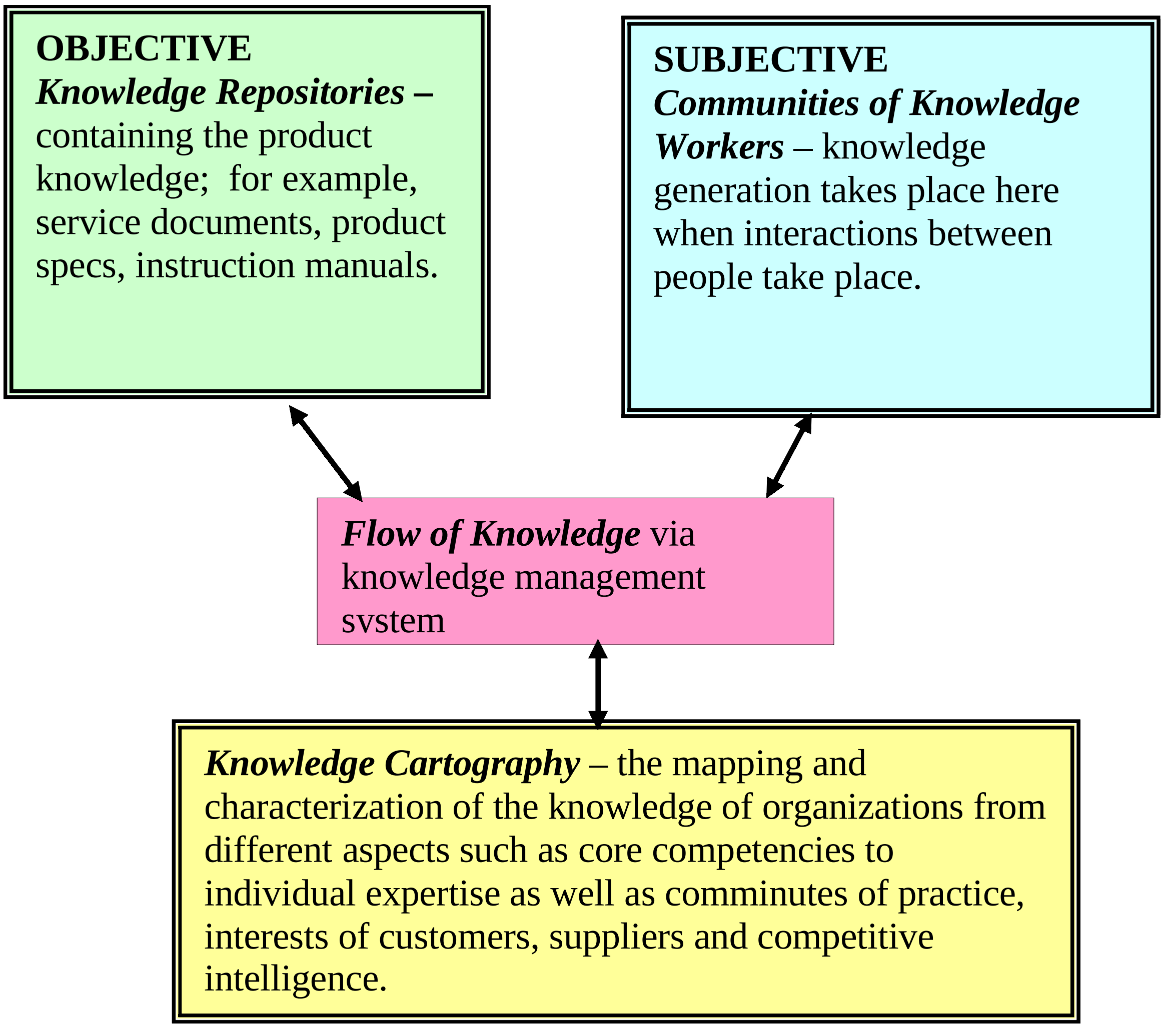
ESTABLISHMENT OF KM INFRASTRUCTURE

The KM infrastructure, in terms of tools and technologies (hardware as well as software) should be established so that knowledge can be created from any new events or activity on a continual basis. This is the most important component of a learning organization. The entire new know-how or new knowledge can only be created for exchange if the KM infrastructure is established effectively. The KM infrastructure will have a repository of knowledge, and distribution systems to distribute the knowledge to the members of organization and a facilitator system for the creation of new knowledge. A knowledge-based infrastructure will foster the creation of knowledge, and provide an integrated system to share and diffuse the knowledge in the organization (Srikantaiah & Koenig, 2000).

KNOWLEDGE ARCHITECTURE

Architecture, specifically the information technology architecture, is an integrated set of technical choices used to guide an organization in satisfying its business needs

Figure 1. The knowledge architecture (Wickramasinghe & Mills, 2001)



(Weil & Broadbent, 1998). Underlying the knowledge architecture (refer to Figure 1) is the recognition of the binary nature of knowledge; namely its objective and subjective components. What we realize when we analyze the knowledge architecture closely is that knowledge is not a clearly defined, easily identifiable phenomenon; rather it has many forms, which makes managing it even more challenging.

The knowledge architecture then recognizes these two different yet key aspects of knowledge; namely, knowledge as an object and a subject, and provides the blueprints for an all-encompassing knowledge management system (KMS). The pivotal function underlined by the knowledge architecture is the flow of knowledge. The flow of knowledge is fundamentally enabled (or not) by the knowledge management system.

KNOWLEDGE MANAGEMENT SYSTEMS

Given the importance of knowledge, systems are being developed and implemented in organizations that aim to facilitate the sharing and integration of knowledge; that is, support and facilitate the flow of knowledge. Such systems are called knowledge management systems (KMS) as distinct from transaction processing systems (TPS), management information systems (MIS), decision support systems (DSS) and executive information systems (EIS) (Alavi, 1999). For example, Cap Gemini Ernst & Young, KPMG and Acenture all have implemented KMS (Wickramsinghe, 1999). In fact, the large consulting companies were some of the first organizations to realize the benefits of knowledge management and plunge into the knowledge management abyss. These companies treat knowledge management with the same high priority as they do strategy formulation, an illustration of how important knowledge management is viewed in practice (Wickramasinghe, 1999). Essentially, these knowledge management systems use combinations of the following technologies: the Internet, intranets, extranets, browsers, data warehouses, data filters, data mining, client server, multimedia, groupware and software agents to systematically facilitate and enable the capturing, storing, and dissemination of knowledge across the organization (Alavi, 1999; Davenport & Prusak, 1998; Kanter, 1999). Unlike other types of information systems, knowledge management systems can vary dramatically across organizations. This is appropriate if we consider that each organization's intellectual assets, intangibles and knowledge should be to a large extent unique and thus systems enabling their management should in fact differ.

KNOWLEDGE MANAGEMENT TOOLS AND TECHNIQUES

KM tools and techniques are defined by their social and community role in the organization in 1) the facilitation of knowledge sharing and socialization of knowledge (production of organizational knowledge); 2) the conversion of information into knowledge through easy access, opportunities of internalization and learning (supported by the right work environment and culture); and 3) the conversion of tacit knowledge into "explicit knowledge" or information, for purposes of efficient and systematic storage, retrieval, wider sharing and application. The most useful KM tools and techniques can be grouped as those that capture and codify knowledge and those that share and distribute knowledge.

a) **Capture and codify knowledge**
   There are various tools that can be used to capture and codify knowledge (Sharma et al., 2004). These include databases, various types of artificial intelligence systems including expert systems, neural networks, fuzzy logic, genetic algorithms and intelligent or software agents.
   i) *Databases*. Databases store structured information and assist in the storing and sharing of knowledge. Knowledge can be acquired from the relationships that exist among different tables in a database. For example, the relationship that might exist between a customer table and a product table could show those products that are producing adequate margins, providing decision-makers with strategic marketing knowledge. Many different relations can exist and are only limited by the human imagination. These relational databases help users to make knowledgeable decisions, which is a goal of knowledge management. Discrete, structured information still is managed best by a database management system. However, the quest for a universal user interface has led to the requirement for access to existing database information through a Web browser.
   ii) *Case-Based Reasoning Applications.* Case-based reasoning (CBR) applications combine narratives and knowledge codification to assist in problem solving. Descriptions and facts about processes and solutions to problems are recorded and categorized. When a problem is encountered, queries or searches point to the solution. CBR applications store limited knowledge from individuals who have encountered a problem and found the solution and are useful in transferring this knowledge to others.
   iii) *Expert Systems.* Expert systems represent the knowledge of experts and typically query and guide users during a decision making process.

They focus on specific processes and typically lead the user, step by step, toward a solution. The level of knowledge required to operate these applications is usually not as high as for CBR applications. Expert systems have not been as successful as CBR in commercial applications but can still be used to teach knowledge management.

iv) *Using I-net Agents - Creating Individual Views from Unstructured Content.* The world of human communication and information has long been too voluminous and complex for any one individual to monitor and track. Agents and I-net standards are the building blocks that make individual customization of information possible in the unstructured environment of I-nets. Agents will begin to specialize and become much more than today's general purpose search engines and "push" technologies.

Two complimentary technologies have emerged that allow us to coordinate, communicate and even organize information, without rigid, one-size-fits-all structures. The first is the Internet/Web technologies that are referred as I-net technology and the second is the evolution of software agents. Together, these technologies are the new-age building blocks for robust information architectures, designed to help information consumers find what they are looking for in the way that they want to find it. The Web and software agents make it possible to build sophisticated, well performing information brokers designed to deliver content from multiple sources to each individual, in the individual's specific context and under the individual's own control. The software agents supported with I-net infrastructure can be highly effective tools for individualizing the organization and management of distributed information.

b) **Systems to share and distribute knowledge**

Computer networks provide an effective medium for the communication and development of knowledge management. The Internet and organizational intranets are used as a basic infrastructure for knowledge management. Intranets are rapidly becoming the primary information infrastructure for enterprises. An intranet is basically a platform based on Internet principles accessible only to members of an organization/community. The intranet can provide the platform for a safe and secured information management system within the organization, help people to collaborate as virtual teams, crossing boundaries of geography and time. While the Internet is an open-access platform, the intranet, however, is restricted to members of a community/organization through multi-layered security controls. The same platform can be extended to an outer ring (e.g., dealer networks, registered customers, online members, etc.), with limited accessibility, as an extranet. The extranet can be a meaningful platform for knowledge generation and sharing, in building relationships, and in enhancing the quality and effectiveness of service/support. The systems that are used for share and distribute knowledge could be: group collaboration systems; groupware, intranets, extranets and Internet, office systems; word processing, desktop publishing, Web publishing and so forth.

HOW TO BECOME A KNOWLEDGE-BASED ENTERPRISE

Just implementing a knowledge management system does not make an organization a knowledge based business. For an organization to become a knowledge-based business several aspects must be considered. An organization that values knowledge must integrate knowledge into its business strategy and sell it as a key part of its products and services. To do this requires a strong commitment to knowledge management directed from the top of the organization. Furthermore, the knowledge architecture should be designed that is appropriate to the specific organization given its industry and the activities, products or services it may provide. From the knowledge architecture it is important to consider the organization's structure as well as its culture. Do the structure and culture support a knowledge-sharing environment or perhaps a more team focussed, sharing culture needs to be fostered (Newell et al., 2002). Then it is necessary to consider the processes of generating, representing, accessing and transferring knowledge throughout the organization and the technology that is required to enable this. Finally, a knowledge based business should also enable organizational learning to take place so that the knowledge that is captured is always updated and current and the organization is continually improving and refining its product or service.

K

CONCLUSION

This discussion has attempted to provide a synopsis of the major technologies for knowledge management. We did this by firstly discussing the knowledge management infrastructure. Then we analyzed the knowledge architecture, highlighting the subjective and objective aspects of KM. From this we discussed knowledge management systems. Finally we discussed how to become a knowledge-based business and that by adopting these techniques and strategies organizations will be able to truly embrace knowledge discovery solutions and thereby maximize their knowledge assets.

REFERENCES

Alavi, M., & Leidner, D. (1999). Knowledge management systems: Issues, challenges and benefits. *Communications of the Association for Information Systems, 1,* paper #5

Davenport, T., & Prusak, L. (1998). *Working knowledge.* Boston: Harvard Business School Press.

Kanter, J. (1999, Fall). Knowledge management practically speaking. *Information Systems Management.*

Newell, S., Robertson, M., Scarbrough, H., & Swan, J. (2002). *Managing knowledge work.* New York: Palgrave.

Sharma, S., Gupta, J., & Wickramasinghe, N. (2004). Information technology assessment for knowledge management. In J.N.D. Gupta & S.K. Sharma (Eds.), *Creating knowledge based organization* (pp. 29-4). Hershey, PA: Idea Group Publishing.

Srikantaiah, T.K., & Koenig, M.E.D. (2000). ASIS Monograph Series. *Information Today.*

Weill, P., & Broadbent, M. (1998). *Leveraging the new infrastructure.* Cambridge: Harvard Business School Press.

Wickramasinghe, N. (1999). Knowledge management systems: Hidden and the manifest. Grant proposal submitted to the Faculty of Economics & Commerce, The University of Melbourne, Australia.

Wickramasinghe, N., & Mills Knowledge Management Systems. *A healthcare imitative with lessons for us all.* 9th ECIS 2001Bled Slovenia.

KEY TERMS

KM Infrastructure: The tools and technologies (the specific tools required to capture and codify organizational knowledge, specific tools required to share and distribute organizational knowledge) that are required to support and facilitate KM in the organization. KM tools and technologies are the systems that integrate various legacy systems, databases, ERP systems, and data warehouse to help organizations to create and use KM systems in the organization.

Knowledge: Knowledge is more comprehensive than data or information. It is a mix of experience, values, contextual information, expert insights, and grounded intuition that actively enables performance, problem solving, decision-making, learning, and teaching.

Knowledge Architecture: The blueprints of subjective and objective knowledge, its flows and cartography of knowledge within the organization.

Knowledge Assets: The knowledge regarding markets, products, technologies, processes and organizations that a business owns or needs to own and which enable its business processes to generate profits, add value, and so forth.

Knowledge-Based Enterprises: Knowledge-based enterprises are those enterprises that derive the most value — from intellectual rather than physical assets. Knowledge-based enterprise is a firm that is fully embracing knowledge management and committed to fostering continuous learning.

Knowledge Management (KM): KM is the process through which organizations generate value from their intellectual and knowledge-based assets. Most often, generating value from such assets involves sharing them among employees, departments and even with other companies in an effort to devise best practices. KM is a newly emerging interdisciplinary business approach that involves utilizing people, processes and technologies to create, store and transfer knowledge.

Knowledge Discovery Using Heuristics

Alina Lazar
Youngstown State University, USA

INTRODUCTION

Uninformed or blind search, which processes and evaluates all nodes of a search space in the worst case, is not realistic for extracting knowledge from large data sets because of time constraints that are closely related to the dimension of the data. Generally, the search space increases exponentially with problem size, thereby limiting the size of problems that can realistically be solved using exact techniques such as exhaustive search. An alternative solution is represented by heuristic techniques, which can provide much help in areas where classical search methods failed.

The word "heuristic" comes from Greek and means "to know," "to find," "to discover" or "to guide an investigation". Specifically, "Heuristics are techniques which seek good (near-optimal) solutions at a reasonable computational cost without being able to guarantee either feasibility or optimality, or even in many cases to state how close to optimality a particular feasible solution is" (Russell & Norvig, 1995).

Heuristic refers to any techniques that improve the average-case performance on a problem-solving task but do not necessarily improve the worst-case performance. Heuristic techniques search the problem space "intelligently" using knowledge of previously tried solutions to guide the search into fruitful areas of the search space. Often, search spaces are so large that only heuristic search can produce a solution in reasonable time. These techniques improve the efficiency of a search process, sometimes by sacrificing the completeness or the optimality of the solution. Heuristics are estimates of the distance remaining to the goal, estimates computed based on the domain knowledge.

The domain knowledge provides help to heuristics in guiding the search and can be represented in a variety of knowledge formats. These formats include patterns, networks, trees, graphs, version spaces, rule sets, equations, and contingency tables. With regard to heuristics, there are a number of generic approaches such as greedy, A* search, tabu search, simulating annealing, and population-based heuristics. The heuristic methods can be applied to a wide class of problems in optimization, classification, statistics, recognition, planning and design.

Of special interest is the integration of heuristic search principles with the dynamic processes in which data become available in successive stages, or where data and inputs are subject to uncertainties or with large-scale data sets. The integration is a vehicle to generate data driven hypotheses.

The kind of knowledge produced, and the heuristic search algorithm selected, will reflect the nature of the data analysis task. The hypotheses are being represented as sets of decision rules and the extracted rules will be represented in terms of rough sets. Rough sets were selected because of the nature of our data sets.

From a mathematical point of view, the problems can be formulated in terms of the well-known, minimal set cover problem, which is a combinatorial optimization problem.

Traditional methods for combinatorial optimization problems are not appropriate here for several reasons. These methods are NP-hard in the worst case and would be costly to use given the size of the data sets. Also, since large data sets are dynamical in nature, adding new data would require running the traditional combinatorial approach again.

The techniques used to solve these difficult optimization problems have slowly evolved from constructive methods, like uniformed search, to local search techniques and to population-based algorithms. Our research goal was to use blend population-based algorithms with methods dealing with uncertainty in order to induce rules from large data sets.

BACKGROUND

Population-based heuristic methods are iterative solution techniques that handle a population of individuals who are evolving according to a given search strategy. At each iteration, periods of self-adaptation (mutations) alternate with periods of cooperation (crossover), and periods of competition (selection). The population-based heuristic search (Conrad, 1978) is dependent on the following components: the knowledge representation for the specific problem to solve and the search strategy or the evolution process. The adaptability of an individual represents its ability to survive in an uncertain environment. Artificial intelligence researchers have explored different ways to represent uncertainty (Russell & Norvig, 1995): belief networks, default reasoning, Dempster-Shafer theory, fuzzy sets theory, rough sets theory.

The learning task will require a representation that explicitly deals with uncertainty. The evolutionary learning methods that are employed must be able to work with such a representation. In this article we look first at basic ways to represent uncertainty in developing rules. And, then we will investigate how that uncertain knowledge can be used to direct evolutionary search and learning.

Uncertainty, as well as evolution, is a part of nature. When humans describe complex environments, they use linguistic descriptors of real-world circumstances, which are often not precise, but rather “fuzzy”. The theory of fuzzy sets (Zadeh, 1965) provides an effective method of describing the behavior of a system that is too complex to be handled with the classical precise mathematical analysis.

The theory of rough sets (Pawlak, 1991) emerged as another mathematical approach for dealing with uncertainty that arises from inexact, noisy or incomplete information. Fuzzy sets theory assumes that the membership of the objects in some set is defined as a degree ranging over the interval [0,1]. Rough sets theory focuses on the ambiguity caused by the limited distinction between objects in a given domain.

Fuzzy sets have been employed to represent rules generated by evolutionary learning systems. Using fuzzy concepts, Hu and Tzeng (2003) tried to overcome the limitations of the conventional rule-based classifier system (Holland, 1975) when representing continuous variables.

Likewise, fuzzy functions have been used to describe and update knowledge in cultural algorithms. First, Reynolds (1994) employed a fuzzy acceptance and influence function in the solution of real-valued constrained optimization problems. Following the same idea, Zhu designed a fully fuzzy cultural algorithm (Reynolds & Zhu, 2001) that included a fuzzy knowledge representation scheme in order to deal with the continuous variables in the belief space, as well as a fuzzy acceptance and influence function. All these approaches were tested on real-values function optimization problems. Jin (2000) used a fuzzy knowledge representation for normative knowledge in the belief space of cultural algorithms to solve the real-valued constrained function optimization.

The design of a fuzzy representation system is not an easy job, because the membership functions should be carefully chosen, and the procedures that use these functions should be specified precisely. The problem is to optimize the fuzzy membership functions for a problem and to find optimum plans related to the fuzzy performance measures. It is a natural approach to use heuristics (i.e., evolutionary algorithms) to solve this task.

Another approach to represent uncertainty is with rough sets. Rough sets are based on equivalence relations and set approximations, and the algorithms for computing rough set properties are combinatorial in nature. Wroblewski (1995) implemented a genetic algorithm for computing reducts, based on permutation code as well as a “greedy” algorithm. Another approach for building reducts is described by Vinterbo (2000) and it is based on the set cover problem, in particular on finding minimal hitting sets using a classical genetic algorithm. Finding a minimal set of decision rules or a satisfactory set is an NP-complete problem. Agotnes (1999) used a genetic algorithm to build an optimal set of decision rules, where the fitness function was based on the quality of each rule. Decision rules extracted via rough set theory could be used to represent hard coded information from neural networks (Lazar & Sethi, 1999).

Evolution can be defined in one word, “adaptation” in an uncertain environment. Nature has a robust way of dealing with the adaptation of organisms to all kinds of changes and to evolve successful organisms. According to the principles of natural selection, the organisms that have a good performance in a given environment survive and reproduce, whereas the others die off. After reproduction, a new generation of offspring derived from the members of the previous generation is formed. The selection of parents from these offspring is often based upon fitness. Changes in the environment will affect the population of organisms through the random mutations. Mayr said, “Evolution is a dynamic, two-step process of random variation and selection” (Fogel, 1995). Using examples from natural systems and theories of adaptive behavior, researchers have been trying to build heuristic evolutionary learning systems.

Evolutionary algorithms are heuristic optimization methods inspired from natural evolution processes. Currently there are three basic population-only mechanisms that model evolution: genetic algorithms, evolution strategies and evolutionary programming. Each of the methods models the evolution of a population of individuals at a different scale and applies election and reproduction operators to find an individual that is fit with regard of the fitness function. The genetic algorithm models evolution at the gene scale, but evolutionary strategies and evolutionary programming model evolution at the species level.

The cultural algorithms (Reynolds, 1994) approach adds another level to the evolutionary process inspired from the human societies and cultural evolution. It adds to the population space, belief space. The belief space will be a collection of symbolic knowledge that will be used to guide the evolution of the population.

Besides the rule-based methods, decision trees (Ziarko, 2003) are well known for their inductive learning capabilities. Any decision tree can be reformulated as a set of rules. One of the problems related to the decision trees is finding the smallest decision tree. Simple heuristics can solve the problem. Researchers have tried to integrate genetic algorithms with decision tree learning in order to solve complex classification problems. Bala (1997) applied this methodology for difficult visual recognition problems involving satellite and facial image data. Other researchers combined the genetic algorithms or evolutionary strategies with neural networks.

Reynolds (2000) investigated the use of cultural algorithms to guide decision tree learning. The data were taken from a real-world archeological database, with a collection of sites found in Valley of Oaxaca, Mexico. The problem was to localize the sites that present evidence of warfare as opposed to those that did not. Evolution-based techniques were used to mine a large-scale spatial data set describing the interactions of agents over several occupational periods in the ancient valley of Oaxaca, Mexico. Specifically, we wanted to extract from the data set spatial constraints on the interaction of agents in each temporal period.

One of the major questions was how to represent the constraint knowledge. Popular data mining methods such as decision trees work well with data collected in a quantitative manner. However, the conditions under which the surface survey data were collected here introduced some uncertainty into the data. Would a representation that explicitly incorporated uncertainty into its structure produce a more efficient representation of the constraints here than one that did not? This is important since the complexity of the constraint set will impact the complexity of the simulation that uses those rules.

We use genetic algorithms to guide the search for a collection of rough set rules to describe constraints on the location of particular types of warfare in the Valley. Since warfare was a major factor in the social evolution in the Valley, the constraints reflecting its spatial and temporal patterning are important ingredients in the model. The rules generated are compared with those produced by a decision tree (Reynolds, 2000) algorithm. In each of the phases examined, the best rule set that used the rough set representation always had fewer conditions in it, and the average rule length was less than that for the decision tree approach in every case but one. In that case they were equal. The differences were most marked in those periods where the warfare patterns were most complex. It was suggested that the differences reflect the inclusion of noise factors as explicit terms in the decision tree representation and their exclusion in the rough sets approach.

Table 1. Comparison between decision trees and rough set rule induction

	Decision Trees	**Rough Set Rules**
Advantages	Easy to understand	Very expressive Modular knowledge Good with missing data They handle imprecise data
Disadvantages	May be difficult to use with continuous data They look at simple combinations of attributes They need to break numeric fields into fixed ranges Not very good with inexact data Not flexible No way to handle missing data Cannot easily approach large data sets May have over fitting Less accurate predictions	Can be memory intensive Can be computationally intensive

A comparison (Table 1) of two decision systems demonstrates that the rough set approach has fewer disadvantages than for the decision tree approach.

In addition, the rough set approach needs to evaluate fewer conditions relative to the inconclusive ones than the decision tree approach. These differences, it is argued, result from the explicit consideration of uncertainty into a period that is more complex and more prone to the introduction of such uncertainty than previous periods.

FUTURE TRENDS

The focus of the comparisons here was on the syntactic or structural differences in the decision systems produced. In future work a comparison of the semantic differences will be accomplished by using the approaches to produce alternative ontologies in the agent-based simulation and assess the differences that are produced. In other words, do the syntactic differences reflect semantic differences in simulation model performance? And, what impact does the use of uncertainty to represent ontological knowledge of the agents have on the basic simulation results?

CONCLUSION

Genetic algorithms, as population-based algorithms, are good vehicles in which to build meta-level heuristics to guide the search more efficiently. That knowledge, rough sets concepts, or rules, can be employed to direct the evolutionary search. The rules can reflect spatial and temporal patterns that will guide the generation of new candidate search objects by the evolutionary engine. The spatial and temporal continuity of the data will facilitate this process.

REFERENCES

Agotnes, T. (1999). *Filtering large propositional rule sets while retaining classifier performance*. Master's thesis. Norwegian University of Science and Technology, Department of Computer and Information Science.

Bala, J., Jong, K.D., Huang, J., Vafaie, H., & Wechsler, H. (1997). Using learning to facilitate the evolution of features for recognizing visual concepts. *Evolutionary Computation, 4*(3), 297-311.

Conrad, M. (1978). Evolution of adaptive landscape. In R. Heim & G. Palm (Eds.), *Theoretical approaches to complex systems, Springer Lecture Notes in Biomathematics* (vol. 21, pp. 147-169). Springer-Verlag.

Fogel, D.B. (1995). *Evolutionary computation - toward a new philosophy of machine learning*. IEEE Press.

Holland, J.H. (1975). *Adaptation in natural and artificial systems*. Ann Arbor, MI: University of Michigan Press.

Hu, Y.C., & Tzeng, G.H. (2003). Elicitation of classification rules by fuzzy data mining. *Engineering Applications of Artificial Intelligence, 16*(7-8), 709-716.

Jin, X., & Reynolds, R.G. (2000). Using knowledge-based systems with hierarchical architecture to guide evolutionary search. *International Journal of Artificial Intelligence Tools, 9,* 27-44.

Lazar, A., & Reynolds, R.G. (2002). Evolution-based learning of ontological knowledge for a large-scale multi-agent simulation. *The 4th International Workshop on Frontiers in Evolutionary Algorithms (FEA 2002),* Research Triangle Park, NC.

Lazar, A., & Sethi, I.K. (1999). Decision rule extraction from trained neural networks using rough sets. In C.H. Dagli, A.L. Buczak & J. Ghosh (Eds.), *Intelligent engineering systems through artificial neural networks* (vol. 9, pp. 493-498). New York: ASME Press.

Pawlak, Z. (1991). *Rough sets - theoretical aspects of reasoning about data.* Kluwer Academic Publishers.

Reynolds, R.G. (1994). An introduction to cultural algorithms. In A.V. Sebald & L.J. Fogel (Eds.), *Proceedings of the 3rd Annual Conference on Evolutionary Programming*, River Edge, NJ (pp. 131-139). World Scientific Publishing.

Reynolds, R.G. (2000). The impact of raiding on settlement patterns in the northern valley of Oaxaca: An approach using decision trees. In T. Kohler & G. Gummerman (Eds.), *Dynamics in human and primate societies* (pp. 251-274). *Oxford University Press.*

Reynolds, R.G., & Zhu, S.N. (2001). Knowledge-based function optimization using fuzzy cultural algorithms with evolutionary programming. *IEEE Transactions on Systems Man and Cybernetics, Part-B 31*(1), 1-18.

Russell, S.J., & Norvig, P. (1995). *Artificial intelligence—a modern approach.* Upper Saddle River, NJ: Prentice Hall.

Vinterbo, S., & Øhrn, A. (2000). Approximate minimal hitting sets and rule templates. *International Journal of Approximate Reasoning, 25*(2), 123–143.

Wroblewski, J. (1995). Finding minimal reducts using genetic algorithms. *Proceedings of Second International Joint Conference on Information Science* (pp. 186–189).

Zadeh, L. (1965). Fuzzy sets. *Information and Control*, *8*, 338-353.

Ziarko, W. (2003). Acquisition of hierarchy-structured probabilistic decision tables and rules from data. *Expert Systems*, *20*(5), 305-310.

KEY TERMS

Cultural Algorithms: Another kind of evolutionary algorithm that adds a belief space to the usual population space from the genetic algorithms in order to improve the search. It is inspired from human societies and cultural evolution.

Evolutionary Computation: Computer-based problem solving systems that use computational models of evolutionary processes as the key elements in design and implementation.

Evolutionary Strategies: Evolutionary algorithms, devoted to the parametric optimization.

Fuzzy Set Theory: Fuzzy set theory replaces the two-valued set-membership function with a real-valued function; that is, membership is treated as a probability, or as a degree of truthfulness.

Genetic Algorithms: An evolutionary algorithm that generates each individual from some encoded form known as a “chromosome” or “genome”. Chromosomes are combined or mutated to breed new individuals. “Crossover,” the kind of recombination of chromosomes found in sexual reproduction in nature, is often also used in GAs. Here, an offspring’s chromosome is created by joining segments chosen alternately from each of two parents’ chromosomes, which are of fixed length.

Heuristics: A rule of thumb, simplification, or educated guess that reduces or limits the search for solutions in domains that are difficult and poorly understood. Unlike algorithms, heuristics do not guarantee optimal, or even feasible, solutions and are often used with no theoretical guarantee.

Knowledge Discovery: In data sets, is the process of identifying valid, novel, potentially useful, and ultimately understandable patterns/models in data.

Rough Set Theory: Rough set theory is a new mathematical tool to deal with vagueness and uncertainty. Any vague concept is replaced by a pair of precise concepts - called the lower and the upper approximation of the vague concept. The lower approximation consists of all objects that surely belong to the concept and the upper approximation contains all objects that possibly belong to the concept.

K

Knowledge Exchange in Electronic Networks of Practice

Robin Teigland
Stockholm School of Economics, Sweden

Molly Wasko
Florida State University, USA

INTRODUCTION

Communities of practice (CoPs) are regarded as essential building blocks of the knowledge economy, and organizations are promoting them as sources of competitive advantage and facilitators of organizational learning. Within organizations, CoPs have traditionally emerged through the mutual engagement in work by individuals who are either physically co-located or who frequently meet face-to-face (Orr, 1996; Wenger, 1998). In an effort to replicate traditional CoPs online and in response to hypercompetitive conditions and increasing complexity, numerous organizations have implemented online networks to facilitate knowledge sharing. We refer to these online social structures focused on knowledge exchange as "electronic networks of practice" (ENoPs).

Although prior researchers have used the term "community" to describe these structures (i.e., electronic community of practice, electronic community, or virtual community), following Brown & Duguid (2000) we use the term "network of practice" to distinguish these social structures from "communities of practice." Networks of practice refer to social structures that link similar individuals engaged in a shared practice, but who may never get to know one another or meet face-to-face. These networks typically consist of weak ties where individuals coordinate through third parties such as professional associations or indirect ties such as newsletters, Web sites, bulletin boards, and listservs (Brown & Duguid, 2000). Thus, we adopt the term network rather than community to distinguish between collectives characterized by sparsely connected weak, indirect ties and collectives where members are connected through frequent face-to-face interactions and direct personal ties. We add the term "electronic" to highlight that communication and coordination within this type of network of practice occurs through asynchronous computer-mediated communication, such as bulletin boards, listservs, etc., and that the focal network structure exists solely in electronic space. The purpose of the network is to facilitate knowledge exchange around a specific practice, and the network structure consists of the aggregation of ties between individuals that are created when individuals post and respond to messages. More precisely, we define an electronic network of practice as *a self-organizing, open activity system focused on a shared practice that exists through computer-mediated communication*. These four defining characteristics are essential for understanding how individuals communicate, coordinate, and interact in these networks.

While traditional, face-to-face CoPs within organizations have received considerable attention, we know much less about ENoPs and knowledge exchange supported by them. Initial research suggests that ENoP participation provides access to useful sources of technical advice for organizational members (Constant, Sproull & Kiesler, 1996). However, there is ample evidence that simply investing in information technologies does not directly enhance knowledge sharing. In fact, researchers estimate that 50-70% of knowledge management projects fail to meet expectations, and they attribute these failure rates to an over-reliance on information technology (Ambrosio, 2000). Thus, a key question for researchers and managers is how to turn an empty electronic space into a vital, active forum devoted to knowledge exchange.

The goal of this chapter is to provide a discussion of ENoPs for researchers and managers interested in studying and supporting these networks. We begin by presenting the characteristics that define ENoPs before discussing two questions related to individual ENoP participation: (1) why do people participate and help others in ENoPs and (2) does participation result in positive knowledge outcomes? Finally, we present and discuss findings from a recent study in a global consulting organization that investigated these questions.

BACKGROUND: DEFINING ELECTRONIC NETWORKS OF PRACTICE

While ENoPs are similar to CoPs in that they are collectives where individuals working on similar problems self-organize to help each other and share perspectives about their practice, they differ in terms of the primary means of communication. In CoPs, mutual engagement typically occurs through physically co-located, face-to-face interactions. However, in ENoPs, individuals mutually engage through asynchronous, text-based computer-mediated communication, such as bulletin boards, listservs, and newsgroups. By posting messages to the ENoP, individuals requiring help may quickly reach out to other participants who then provide valuable knowledge and insight in response. Participants may also share personal experiences and discuss relevant practice issues (Wasko & Faraj, 2000). This posting and responding to messages is recorded like a conversation between participants, representing active mutual engagement in problem solving. However, unlike face-to-face interactions in CoPs where participants perceive various social and visual cues and have access to immediate feedback, in electronic communication these cues are filtered out. As a result, ENoPs represent a lean medium of exchange and the technology impacts how knowledge is actually exchanged between participants (Daft & Lengel, 1986). This mutual engagement also distinguishes ENoPs from more static forms of electronic knowledge exchange, such as document repositories and other databases.

A second characteristic of ENoPs is that the technology creates a weak structural link between an ubiquitous, unlimited number of like-minded "strangers." Participation is open to anyone with a connection anywhere in the world, and as a result, constraints due to size are eliminated. Thus, knowledge exchange occurs between people regardless of personal acquaintance, familiarity, and location. Knowledge seekers are not limited to asking only others whom they personally know or are able to identify, thus increasing the likelihood of connecting with someone willing and able to help. Additionally, membership is fluid, making it difficult to create and enforce boundaries. This sharply contrasts with the tightly knit relationships between specific members that typify CoP structures. Also, this characteristic separates ENoPs from virtual teams, where members are designated and assigned.

Third, ENoP participation is voluntary. Individuals choose whether or not they want to participate, as well as how often - ranging from simply lurking to becoming an active participant. Additionally, individuals have choices about how they participate, deciding whether or not to post questions, replies, or both. Finally, individuals voluntarily determine what they want to contribute and what knowledge they are willing to disclose as well as the length of the messages they contribute, thus influencing the quality and helpfulness of the knowledge exchanged. Because participation is voluntary, a knowledge seeker has no control over who responds to their questions or who uses their responses. This sharply contrasts with CoPs where people typically know one another and interact over time, creating expectations of obligation and reciprocity that are enforceable through social sanctions. This voluntary participation further distinguishes ENoPs from virtual teams, where participants are expected to coordinate efforts to deliver a specific outcome.

Finally, mutual engagement in ENoPs is typically archived and available to all participants in the network. This creates an online repository of knowledge that can be accessed later by any interested individual, regardless of his or her ENoP tenure or participation in the original engagement. This contrasts with CoPs where access to advice is limited to whom you know, and knowledge is exchanged between seeker and provider without necessarily being made available to other members of the CoP.

CURRENT ENOP RESEARCH THEMES

ENoP research to date tends to be limited to an investigation of individual motives behind participation and knowledge sharing in these networks (e.g., Lakhani and von Hippel, 2000; Wasko and Faraj, 2000). Studies have revealed that individuals share knowledge with "strangers" due to expectations of returns for themselves (e.g., increased reputation, enjoyment, etc.) as well as for the network (e.g., advancing the community). Furthermore, research by Lakhani & von Hippel indicates that individuals make discretionary choices regarding their willingness to share knowledge and help others in ENoPs. However, while ENoP research is increasing, there is little research other than Constant et al.'s 1996 study specifically focusing on *intra-organizational* ENoPs. Thus, one question to ask is, why do individuals participate and help others in intra-organizational EnoPs?

Secondly, numerous organizations are attempting to support or even construct various forms of ENoPs within and between their organizations (e.g., Wenger et al., 2002). Similar to other knowledge management initiatives implemented to enhance organizational performance, management's hope is that these efforts will positively affect individual behavior in the workplace and thus ultimately drive increases in firm performance (Davenport & Prusak, 1998; Wenger et al., 2002). However, there is a dearth of solid academic empirical support for this positive relationship, with researchers paying little systematic attention to the relationship to performance at any level (Teigland, 2003). Thus, a second question is, does participation in intra-organizational ENoPs result in positive knowledge outcomes?

Below we further develop these two questions. Additionally, we provide some preliminary findings to these questions based on a study conducted in the Nordic operations (Denmark, Finland, Norway, and Sweden) of Cap Gemini (performed prior to the merger of Cap Gemini and Ernst & Young Consulting). Cap Gemini had numerous ENoPs to enhance the company's knowledge management activities. We chose one ENoP, referred to as the NCN MS Community. This ENoP had 345 members throughout the Nordic countries and was designed for individuals who worked with applying Microsoft products in their Cap Gemini responsibilities. It was based on listserv (mailing list) technology, with members using the listserv when they needed help performing their work tasks.

Why Do People Participate and Help Others in Intra-Organizational ENoPs?

Researchers have proposed that one theoretical lens with which to investigate ENoPs is the theory of collective action and public goods (Fulk, Flanagin, Kalman, Monge & Ryan, 1996; Monge, Fulk, Kalman, Flanagin, Parnassa & Rumsey, 1998). A public good, for example a public park, is a resource that is created only if a collective, or a group of individuals, contributes to its production. However, a public good cannot be withheld from any member of the collective, even if he or she does not participate in the production or maintenance of the good (Olson, 1965; Samuelson, 1954). With public goods, the optimal individual decision is to enjoy the public good without contributing anything to its creation or maintenance and to simply free-ride on the efforts of others. However, if everyone were to act rationally and decide not to contribute, then the good would never be created and everyone would be worse off.

ENoPs are a type of collective in which the knowledge exchanged and created is the collective's public good. As discussed above, mutual engagement in ENoPs is open and voluntary. Participation typically results in the creation of a knowledge repository of archived messages that is available to all individuals regardless of their original participation. This begs the question then – why would individuals invest their valuable time and effort helping strangers in an ENoP if it is in their best interest not to do so?

Our Cap Gemini findings suggest that people who help others in this ENoP are not acting irrationally. Rather, our findings suggest that through ENoP participation, individuals are able to improve their own level of technical competence, helping them to remain competitive in relation to their technical skills. For example, through the ENoP, individuals were able to access help and advice not available locally and individuals learned by receiving help and information related to their work tasks. Additionally, they felt that participation enabled them to keep up-to-date with technical developments and to know who was actively working in different areas. Second, a key dynamic underlying knowledge exchange in this ENoP was a strong norm of reciprocity between the members, or a strong sense of paying back. In other words, to receive help from the ENoP, individuals felt obligated to help others in return. Finally, another key motive underlying why people participated was related to organizational identification or a strong sense of organizational citizenship. Respondents felt that helping others in the ENoP was part of their job at Cap Gemini. Through their participation, they were able to improve the level of technical competence of the ENoP as a whole and ultimately Cap Gemini's competitiveness.

Does Intra-Organizational ENoP Participation Affect Knowledge Outcomes?

The second question deals with the relationship between intra-organizational ENoP participation and knowledge outcomes and, in particular, whether ENoPs exhibit the same degree of continuous incremental innovation as CoPs. As mentioned above, CoPs are generally characterized by rich, face-to-face exchange through person-to-person interactions. Mutual engagement between individuals in CoPs creates boundaries around the shared practice within which the community's knowledge is embedded, and tacit knowledge is shared relatively easily between individuals within the community, often without ever being made explicit. These tightly knit social structures facilitate the creation of a shared identity through the development of a common language, social capital (such as norms, trust, and obligations), boundaries, and social controls, resulting in strong social ties between individuals. These characteristics have been argued as essential for the continuous incremental improvements in the community's practice and the reason why CoPs are centers for learning and innovation within organizations (Brown & Duguid, 1991, 1998; Wenger, 1998).

In contrast, the ability of ENoP members to develop a shared identity and common language through narration, collaboration, and social construction is hampered due to the communication occurring through text-based, asynchronous, computer-mediated communication. However, ENoPs have a greater reach than CoPs, supporting the creation of weak electronic "bridging ties" between an unlimited number of like-minded others. Due to this extensive reach, individuals benefit from ENoPs, gaining access to new information, expertise, and ideas that are often not available locally. As such, the weak tie relationships in ENoPs potentially increase an individual's access to greater resources and advice than those locally available. Thus, one question to ask is whether this extended reach results in positive knowledge outcomes. In other words, are weak electronic links, like their strong tie counterparts, also useful for supporting knowledge sharing and innovation?

Our Cap Gemini research provided some interesting results. First, respondents indicated that ENoP participation had indeed helped them. Participation greatly improved the speed with which respondents were able to solve their problems, and individuals were able to learn and receive new insights from the ENoP. Statistical analysis of the survey responses also suggested that the level of ENoP participation was more important for supporting positive knowledge outcomes in terms of knowledge acquisition and contribution than the length of time an individual had participated. This implies that ENoP newcomers can reap the same benefits of participation as long standing participants. Additionally, both knowledge acquisition from and knowledge contribution to the ENoP were positively related to individual performance. However, tenure in the ENoP was not associated with higher rates of participation nor with individual performance. Finally, and of considerable interest, the results suggested that individuals who relied more on their co-located colleagues and less on the ENoP for help or advice with their work tasks reported no associations with knowledge acquisition or knowledge contribution. In fact, the survey results indicated that *reliance on co-located colleagues is associated with lower levels of self-reported individual performance*.

FUTURE TRENDS AND CONCLUSION

In conclusion, this discussion of ENoPs for researchers and managers began by presenting ENoP characteristics before developing two questions related to intra-organizational ENoP participation and discussing these based on a study of a successful ENoP at Cap Gemini. Our results suggest that ENoPs may support the complex interactions necessary for the sharing of knowledge between individuals, thus facilitating continuous incremental innovation and positive knowledge outcomes. It appears that individuals value accessing new insights and ideas through weak electronic links that transcend their strong tie networks.

Additionally, while research on CoPs and face-to-face interactions with colleagues has provided evidence of knowledge exchange, learning, and innovation, it has also been suggested that tightly knit CoPs may lead to the "not invented here" syndrome or the resistance to new ideas not locally developed. Our finding that individuals who participate in ENoPs outperform colleagues who primarily rely on co-located colleagues for knowledge and advice supports the latter argument. This suggests that knowledge in tightly knit CoPs may be largely redundant, thus providing little additional information over what an individual may already know and impeding the development of new and creative ideas (Granovetter, 1983). Therefore, managers concerned with improving knowledge exchange should note that the highly efficient structures that support knowledge integration in CoPs may evolve into core rigidities and competency traps – inappropriate knowledge sets that preserve the status quo and limit new insights (Leonard-Barton, 1992; Levitt & March, 1988). Our findings suggest that one way to alleviate this concern is to use ENoPs to create electronic bridging links between strong tie communities to enhance the development of new ideas and innovations.

In terms of limitations, our research provided little explanation of how ENoP social contexts develop over time, what factors influence development, and how motives and social context interact and change over time.

Thus, future research could examine ENoP dynamics as well as how individual participation in various ENoPs changes over time. Future research could also compare the dynamics of intra-organizational and inter-organizational ENoPs. Finally, the relationship between ENoP participation and the characteristics of the communication technology supporting the ENoP, e.g., bulletin board, listserv, etc., should be investigated.

REFERENCES

Ambrosio, J. (2000). Knowledge management mistakes. *Computerworld*, 34(27), 44.

Brown, J.S., & Duguid, P. (1991). Organizational learning and communities of practice. *Organization Science*, 2(1), 40-57.

Brown, J.S., & Duguid, P. (1998). Organizing knowledge. *California Management Review*, 40(3), 90-111.

Brown, J. S., & Duguid, P. (2000). *The social life of information.* Boston, MA: Harvard Business School.

Constant, D., Sproull, L., & Kiesler, S. (1996). The kindness of strangers: The usefulness of electronic weak ties for technical advice. *Organization Science*, 7(2), 119-135.

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science*, 32(5), 355-366.

Fulk, J., Flanagin, A. J., Kalman, M.E., Monge, P.R., & Ryan, T. (1996). Connective and communal public goods in interactive communication systems. *Communication Theory*, 6(1), 60-87.

Granovetter, M. (1983). The strength of weak ties: A network theory revisited. *Sociological Theory*, 1, 201-233.

Lakhani, K., & von Hippel, E. (2000). How open source software works: Free user-to-user assistance. Paper presented at The 3rd Intangibles Conference. Knowledge: Management, Measurement and Organization, Stern School of Business, NYU.

Leonard-Barton, D. (1992, Summer). Core capabilities and core rigidities: A paradox in managing new product development. *Strategic Management Journal*, 13 (Special Issue), 111-125.

Levitt, B., & March, J. G. (1988). Organizational learning. *Annual Review of Sociology*, 14, 319-340.

Monge, P. R., Fulk, J., Kalman, M.E., Flanagin, A.J., Parnassa, C., & Rumsey, S. (1998). Production of collective action in alliance-based interorganizational communication and information systems. *Organization Science*, 9(3), 411-433.

Olson, M. (1965). *The logic of collective action.* Cambridge, MA: Harvard University Press.

Orr, J. (1996). *Talking about machines: An ethnography of a modern job.* Ithaca, NY: ILR Press.

Samuelson, P. A. (1954). The pure theory of public expenditure. *Review of Economics and Statistics*, 36(4), 387-389.

Teigland, R. (2003). *Knowledge networking: Structure and performance in networks of practice.* Published Doctoral Dissertation. Stockholm: Stockholm School of Economics.

Wasko, M., & Faraj, S. (2000). It is what one does: Why people participate and help others in electronic communities of practice. *Journal of Strategic Information Systems*, 9(2-3), 155-173.

Wenger, E. (1998). *Communities of practice.* Cambridge, UK: Cambridge University Press.

KEY TERMS

Bulletin Board: A discussion forum, similar to that of Usenet newsgroups, in which questions and responses are connected in a “thread,” resembling a conversation.

Collective Action: The voluntary cooperation of a group of individuals that typically involves the production of a public or semi-public good.

Community of Practice: An emergent group of a limited number of individuals (generally in the same location) informally and contextually bound through close face-to-face interactions in the pursuit of a common enterprise.

Electronic Network of Practice: An emergent group of an unlimited number of dispersed individuals working on similar tasks using a similar competence whose communication channel is purely electronic.

Free-riding: The act of enjoying the public good without contributing anything to its creation or maintenance.

Listserv: Mailing lists that forward e-mail messages to everyone who has subscribed to the list. Members seldom know each other due to the large size of these lists.

Lurking: The act of reading interactions between electronic network of practice participants but seldom if ever publicly contributing anything.

Public Good: Resources from which all individuals in the collective may benefit regardless of whether they have contributed to providing the good.

Knowledge Management and Social Learning

K

Irena Ali
DSTO Fernhill, Australia

Leoni Warne
DSTO Fernhill, Australia

INTRODUCTION

There are probably as many variations of knowledge management definitions as there are practitioners and researchers in the discipline. Complete consensus in such a group would be a surprising finding. This is because the two words are loaded with pre-existing meanings that do not always sit comfortably in juxtaposition, so what it means to "manage knowledge" is difficult to ascertain, and hence comes to mean different things to different people.

We do know, however, that knowledge exists in the minds of individuals and is generated and shaped through interaction with others. In an organizational setting, knowledge management must, *at the very least*, be about how knowledge is acquired, constructed, transferred, and otherwise shared with other members of the organization, in a way that seeks to achieve the organization's objectives. Put another way, knowledge management seeks to harness the power of individuals by supporting them with information technologies and other tools, with the broad aim of enhancing the *learning capability* of individuals, groups, and, in turn, organizations.

BACKGROUND

In this article, we examine both theoretical and practical socio-cultural aspects of knowledge management based on years of research by the authors in a large and diverse organization. The study involved numerous functional settings of the organization and the researchers used qualitative and quantitative methodology to gather data. Elements required to build an organizational culture that supports knowledge management are discussed. Unless otherwise specified, words in double quotes in the text are direct quotes from personnel in research settings.

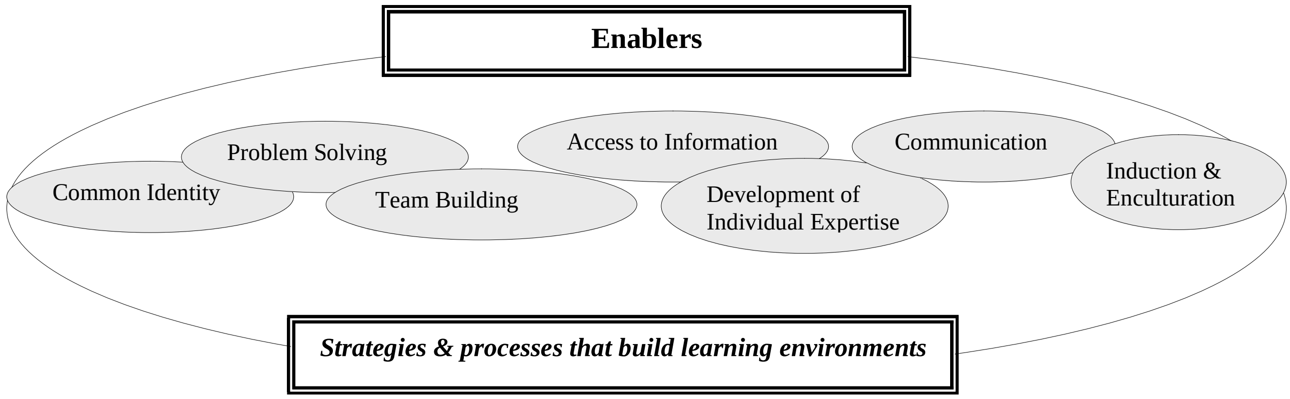
MAIN THRUST OF THE ARTICLE

The research team identified seven basic categories that constitute enabling processes and strategies to facilitate social learning: common identity; problem solving; team building; access to information; development of individual expertise; communication; and induction and enculturation (see Figure 1).

Common identity: a common ground/understanding to which many people/groups can subscribe, and requires a shift from seeing oneself as separate to seeing oneself as connected to and part of an organization unit. Based on our research, motivators impacting on *common identity* are: goal alignment, cultural identity, gendered identity, language, morale, and workplace design (spatial and physical design).

- Doney, Cannon et al. (1998) discuss the relationship between goal alignment and group cohesiveness, claiming that the extent of group cohesiveness relies on the extent to which a team's goals are clear and accepted and also on the degree to which all members adopt team behaviors.
- The term cultural identity refers to a member's sense of self in relation to the specific "tribe" and "tradition" to which they belong and how this distinctiveness applies in their workplace. Cultural identity is another important motivator for social learning because, like common identity, it impacts on the extent to which staff feels that they are part of the system or alienated from it.
- Gendered identity relates specifically to one's sense of self, which is imbued with the social, cultural and historical constructions surrounding femininity and masculinity. Gender identity, because of its relationship with common identity, was also seen to impact on social learning.
- Language is another important factor fundamental to the overall social learning processes. By reflect-

Figure 1: Constructs enabling social learning



ing the social and political relationship between various members, language can impact on common identity. Language is also important in terms of creating a shared understanding among workers and their relationship to the wider organization. "Words are bullets. Never, never use imprecise language." Thus learning the specific work-related language is of central importance to broader social learning development, and is an important outcome of the enculturation process.

- Morale has been a significant focus in the overall study because the research team found evidence of low morale being coupled with higher levels of alienation towards senior management. Such alienation has obvious implications for the broader understanding of a common identity and thus for social learning.
- Workplace design and proximity also threatens common identity when staff are not working in the same location. "[Building X] and us. We don't see them. There is not any spirit that we are belonging to one branch. I have more to do with [a specific area] than anything else and I've made some good contacts in there,... who I sit around with."

Problem solving: a core activity. It fosters social learning, because each problem represents an opportunity to generate knowledge. Motivators associated with this enabler are: networking, perceptions of the organization, systemic understanding, and time for inquiry and reflection.

- An individual's personal and social networks are an important means of acquiring, propagating, and sharing knowledge. As Davenport and Prusack (1998) claim, when those who are in a position of "know-how" share their expertise they contribute to problem solving. Personal networks were seen to function as channels supporting both "information pull" and "information push".
- Individual and shared perceptions of the organization, and how they operate, provide an essential backdrop to problem solving within an organizational context. These perceptions may consist of deeply ingrained assumptions, generalizations, or even pictures or images that influence how people understand their organizational world and how they should act within it (Senge, 1992). The importance of these perceptions cannot be stressed enough, because they directly influence the construction of individuals' knowledge and understandings that they draw upon in their day-to-day-activities.
- Effective problem solving often requires a systemic understanding of organizational and inter-organizational issues. Systemic understanding requires a holistic view of an organization and its interrelationships, an understanding of the fabric of relationships and the likely effect of interrelated actions (Senge, 1992).
- Inquiry and reflection together are a powerful means of enhancing social learning and knowledge creation. Inquiries, or questions, are triggered by problems that require solutions or explanation. Reflec-

tion allows time for examination, contemplation and, often, resolution of the inquiries. To use a common metaphor, it is perhaps the best means for distinguishing between the forests and the trees of everyday working life.

Team building: working together and understanding what each member is trying to do. Team building was seen to be essential for effective social learning and problem solving. Motivators associated are: leadership, team-based morale, performance management, public recognition and reward systems, use of humour, and workplace design.

- In general, the caliber of leadership within the settings studied was to be admired. The leaders and managers were innovative and they motivated and developed their staff, mainly by demonstrating that staff are highly valued and by acknowledging expertise and knowledge regardless of their pay or position. Another team building issue that emerged was that people were appreciative of informal "drop-ins" by senior managers inquiring how they were doing. This "roving management" was said to contribute to better cohesion of teams, to promote system thinking, to help focus on overall goals, and to facilitate communication and feedback.
- "Team spirit" and "team cohesiveness" are both important values within the work culture and work ethic; nonetheless, there was nothing uniform about this in the settings studied. Some teams did not see the significance of their particular tasks to the overall goals of the organization. However, good examples of teamwork and team spirit were also evident. There were instances where teamwork was well integrated into daily work and where people worked collaboratively. Such teams were goal-oriented and were not only teams in structure but in spirit and were led by a leader who saw his/her role as serving team members rather than just having the position of a leader.
- For many employees, the performance cycle is annual and the outcome of a performance report often determines the prospects of one's career progression. Some felt somewhat uneasy as their performance evaluation was due relatively early into their posting cycle. A well-planned performance appraisal system should help to make equitable and unbiased decisions regarding staff selection, placement, development and training (Wood, 1989). Researchers were told that there was often a lack of clear communication about performance expectations. Also, an annual performance appraisal appears to be too long to wait for recognition of good work and too late to correct a performance problem. Morgan (1989) and Wood (1989) explain that to maximize positive results, the appraisal process should be two ways: it should facilitate and coach staff in doing their jobs effectively; and it should be frequent and informal.
- It was observed that humor was used for smoothing discussions that were becoming heated and to stop the conflict from escalating while also enabling the conflicting subordinates to save face. At meetings, humor was used to assist in uniting people around common themes and to make criticism palatable.
- One way of increasing team and individual morale is to publicly acknowledge outstanding work. Making employees feel appreciated, and saying, "Thank you, we know that you are a good employee, we value you and your work", is a big factor in motivation (Mitchell, 2000). Key informants stated that public recognition of good work was scarce and that a written or verbal word of praise, a pat on the back often means more, for example, than a pay raise – "praise is better than money" and praise is needed at all levels.
- Workplace design was seen to have impact on social learning. Staff located at small isolated outposts were at risk of feeling isolated and did not identify strongly with the parent organization. As stated earlier, out-posted staff identified more with the workplace with which they were based than their branch where they affiliated. This was further exacerbated by the fact that they often felt excluded by their colleagues.

Access to information: the easy availability of corporate information in whatever format was observed to effect knowledge acquisition and generation of new knowledge and social learning. Motivators associated are: record keeping, networking, meetings, and information technology (IT) infrastructure.

- The researchers observed that general familiarity with records keeping procedures was quite poor. Some people have developed their own personal records keeping systems but there was little uniformity in these and no adherence to file naming conventions and standards. As some informants stated: "I believe that physical files in the … are no longer managed well because their management has been farmed out to outside bodies." Or "I think we have problems with passing on information in the organization as a whole. We just don't do it very well." The issue of electronic records, particularly e-mail messages containing evidence of business transactions, posed problems not only in the setting studied but also in the whole of organization.

- Personal networks from previous postings as well as newly acquired contacts in the new environment play a vital role in knowledge construction and acquisition. New knowledge often begins with the individual and through conversations people discover what they know, what others know and in the process of sharing, new knowledge is created. Knowledge sharing depends on the quality of conversations, formal or informal, that people have. Webber (1993) aptly describes it: "Conversations – not rank, title, or the trappings of power – determine who is literally and figuratively 'in the loop' and who is not."
- Meetings are another means of accessing information and those that were observed varied significantly in format and the protocols in place. At the tactical headquarters, meetings that were mission-related provided excellent opportunities for learning. Strict protocols were observed at these briefings, for example, allowing participants to discuss errors or problems encountered during missions without assigning blame or shame to individuals. There were few equivalent meetings at the strategic headquarters, other than some induction sessions and briefings, and it appeared that learning how to do one's job was not given quite the same priority.
- The researchers observed that information access due to failings in the IT infrastructure inhibited access to information within the strategic settings. Another issue that caused problems was the difficulty in finding information on the shared drive. Since there was no specific person responsible for maintaining the shared drive and for naming folders, it was left to the discretion of the document originator where information would be stored.

Development of individual expertise: the acquisition and development of expertise was seen as an integral part of social learning. Motivators associated with this enabler are: career trajectories, professional currency, professional training, postings and promotion, and mentoring.

- A career trajectory describes the positions, roles and experience that individuals have accumulated, up to and including the position they currently hold. While not excluding personal experiences outside of a work or training context, a well-designed career trajectory generally equips an individual with the skills, experience, maturity and personal networks needed to successfully fill a particular posting.
- The term professional currency has a somewhat different meaning within different environments. However, professional currency promotes social learning in the same way that appropriate career trajectories do so " by providing a foundation for the generation of new knowledge.
- Appropriate professional training is a significant component of the development of individual expertise and, therefore, a fundamental for generating new knowledge. Training courses are important for furthering individuals' expertise, as well as for forming the personal networks that subsequently develop. However, in times of budgetary constraints, training money is often the first to go, with damaging consequences for the organization's ability to learn and manage their knowledge.
- Mentoring is regarded as an effective method of assisting the development of individual expertise; especially for junior staff, a degree of informal mentoring was seen to be built into elements of the training program in some of the settings studied. In terms of developing a career trajectory, the knowledge acquired through mentoring may also be important when individuals want to prepare themselves for specific roles in the future.

Communication: essential to effective learning within an organization and to effective social learning. Motivators associated with this enabler are: overall communication climate, formal and informal information flows, time for inquiry and reflection, use of humor, language, and workplace design.

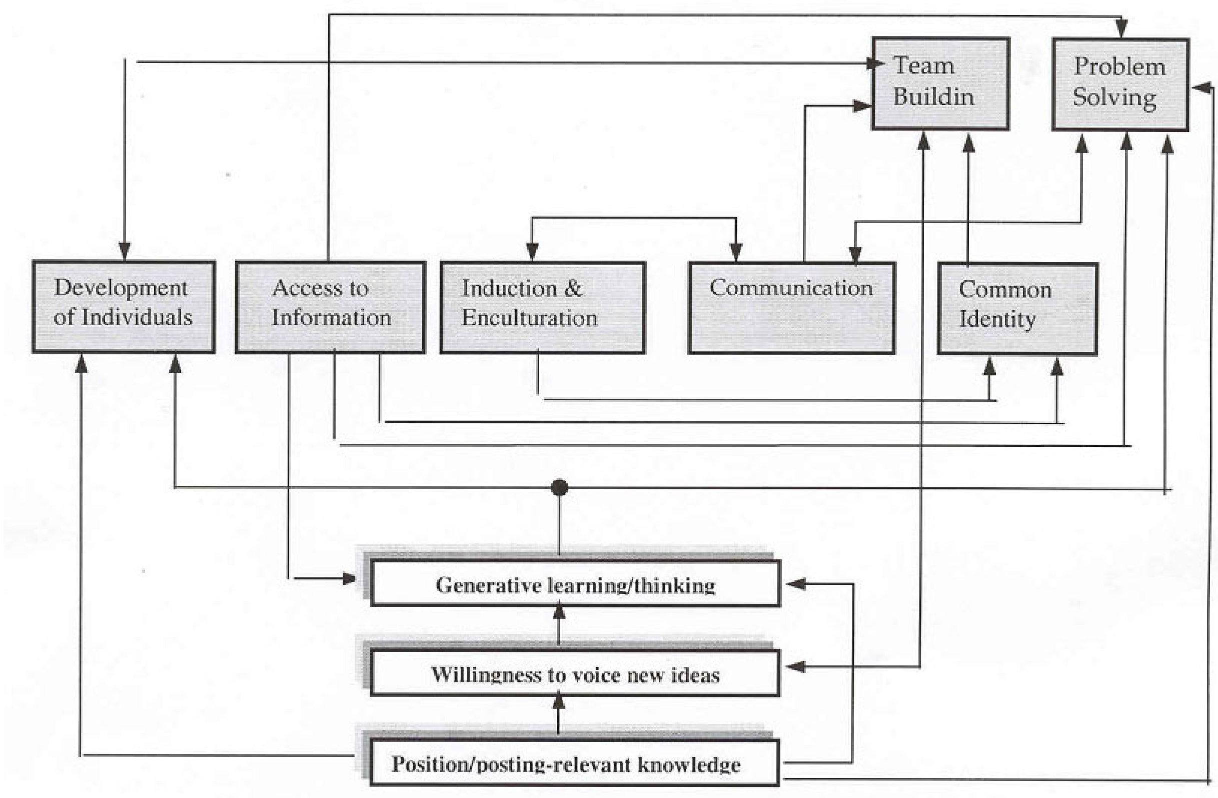
- Supportive communication climates are being positively linked to open and free exchange of information and constructive conflict management. Characteristics of a supportive communication climate include a culture of sharing knowledge, treating each other with respect, and generally behaving in a cooperative manner. Research has established the link between supportive organizational communication climates and generative learning (Bokeno, 2000; Ruppel, 2000) and with higher levels of organizational commitment (Guzley, 1992).
- An important element of generative learning is for organizational members to be able to engage in dialogue which is open and is based on inquiry and reflection. A supportive communication climate is a prerequisite for such dialogue, and it requires learning how to recognize defensive patterns of interaction that undermine learning (Senge, 1992).

- The issue of workplace design and its impact on teams, network building, and on accessing information arose repeatedly during the study. Physical location and proximity to each other had the potential to promote the transfer of pertinent knowledge. The point was made that in addition to more quickly obtaining answers to questions about particular tasks, an open plan workplace enabled one to tap into pertinent knowledge by overhearing others' conversations. Hutchins (1996) uses the term "horizon of observation" to describe the area of the task environment which can be seen, and is therefore available as a context for learning by team members.

Induction and enculturation: facilitates social learning by providing a foundation upon which an individual can become fully productive. Issues associated with this enabler are: timeliness and comprehensiveness of the process, buddy/mentoring system, handovers and information packages, and training.

- Good induction is more than just an introduction to a new job and workmates; it is a way of helping people find their feet. Attitudes and expectations are shaped during the early days of new employment, and work satisfaction is linked to well-timed and conducted work orientation (Dunford, 1992; George & Cole, 1992). The interviews clearly indicated a relationship between meaningful and timely induction and subsequent job satisfaction. An interesting finding was that those who were not properly inducted or enculturated into the organization saw no need and responsibility to actually prepare any form of handover for anyone who may take over their position in the future.
- Although highly desirable, it was not always feasible to conduct an induction program at the beginning of a new posting cycle. In the interim, a "buddy" or "mentoring" system could fill in the gap. A "buddy" would be an experienced workmate who could be available to answer questions and assist the orientation of new members during the initial few weeks. Some interviewees said that having a buddy when they started was invaluable to settling into a new job and to effective learning.

Figure 2. Enabling processes and their impact on social learning



- The researchers were repeatedly told that early training is an important part of effective induction and enculturation. It is an opportunity to learn the explicit knowledge that is taught as part of formal training. It is also an opportunity to be exposed to the attitude and cultural perceptions of colleagues and peers.

These factors enabling social learning identified from our data are by no means exhaustive, however, based on the available data, the research team could see a relationship between these enablers and social and generative learning. Figure 2 depicts these relationships and their impact on social learning.

CONCLUSION

Therefore, organizations seeking to improve information sharing and knowledge generation need to develop a greater awareness of the processes and strategies of organizational learning. Organizational knowledge is distributed across functional groups, and its generation and continual existence is dependant on the overall communication climate which is embedded in the organizational culture. This study indicates that information sharing and subsequent knowledge generation would be successful when interactive environments are cultivated before other, for example, technology-based solutions are implemented.

BIBLIOGRAPHY

The information in this article is largely derived from the following sources:

Agostino, K. (1998). The making of warriors: Men, identity and military culture. *JIGS: Australian Masculinities*, *3*(2).

Ali, I., Pascoe, C., & Warne, L. (2002). Interactions of organizational culture and collaboration in working and learning. *Educational Technology and Society*, *5*(2): 60-69. ISSN 1436-4522.

Argyris, C. (1973). *On organisations of the future*. Beverley Hills, CA: Sage.

Drucker, P.F. (1999, October). Beyond the information revolution. *The Atlantic Monthly*.

Enneking, N.E. (1998). Managing email: Working toward an effective solution. *Records Management Quarterly*, *32*(3), 24-43.

Ganzel, R. (1998). Elements of a great orientation. *Training*, *35*(3), 56.

Warne, L., Agostino, K., Ali, I., Pascoe, C., & Bopping, D. (2002). The knowledge edge: Knowledge management and social learning in military settings. In A. Gunasekaran, O. Khalil & S.M. Rahman (Eds.), *Knowledge and information technology management: Human and social perspectives* (pp. 324-353). Hershey, PA: Idea Group Publishing, ISBN 1-59140-032-5.

REFERENCES

Bokeno, R.M. (2000). Dialogic mentoring. *Management Communication Quarterly*, *14*, 237 270.

Cooke, R. (1998). Welcome aboard. *Credit Union Management*, *21*(7), 46-47.

Davenport, T.H. & Prusack, L. (1998). *Working knowledge: How organisations manage what they know*, Harvard Business School Press.

Doney, P.M., Cannon, J.P., et al. (1998). Understanding the influence of national culture on the development of trust. *Academy of Management Review*, *23*(3), 601-623.

Drucker, P.F. (1999, October). Beyond the information revolution. *The Atlantic Monthly*.

Dunford, R.W. (1992). *Organisational behaviour: An organisational analysis perspective*. Sydney: Addison Wesley.

George, C.S., & Cole, K. (1992). *Supervision in action: The art of managing*. Sydney: Prentice Hall.

Guzley, R.M. (1992). Organizational climate and communication climate: Predictors of commitment to the organization. *Management Communication Quarterly*, *5*(4), 379-402.

Hutchins, E. (1996). *Cognition in the wild*. Cambridge, Mass: MIT Press.

Mitchell, S. (2000). Be bold and discover the power of praise. East Roseville: Simon & Schuster.

Morgan, T. (1989). Performance management - the missing link in strategic management and planning. In D.C. Corbett (Ed.), *Public sector policies for the 1990's* (pp. 243-250). Melbourne, Public Sector Management Institute, Faculty of Economics and Politics, Monash University.

Ruppel, P.C.a.S.J.H. (2000). The relationship of communication, ethical work climate, and trust to commitment and innovation. *Journal of Business Ethics*, *25*, 313-328.

Senge, P.M. (1992). *The fifth discipline: The art & practice of the learning organisation.* Australia: Random House.

Webber, A.M. (1993, Jan-Feb). What's so new about the new economy? *Harvard Business Review*, 24-42.

Wood, R. (1989). Performance appraisal in the reform of public sector management practices. In D.C. Corbett (Ed.), *Public sector policies for the 1990's* (pp. 225-242). Melbourne, Public Sector Management Institute, Faculty of Economics and Politics, Monash University.

KEY TERMS

Career Trajectory: Describes the positions, roles and experience that individuals have accumulated, up to and including the position they currently hold.

Common Identity: A common ground/understanding to which many people/groups can subscribe, and requires a shift from seeing oneself as separate to seeing oneself as connected to and part of an organizational unit.

Communication Climate: Extend to which there is an open and free exchange of information, transparency of decision-making, and how constructively conflict is managed.

Knowledge: An understanding gained through experience or learning: The sum, or a subset, of what has been perceived and discovered by an individual. Knowledge exists in the minds of individuals and is generated and shaped through interaction with others.

Knowledge Management: In an organizational setting, it must, *at the very least*, be about how knowledge is acquired, constructed, transferred, and otherwise shared with other members of the organization, in a way that seeks to achieve the organization's objectives.

Social Learning: Learning occurring in or by a cultural cluster and includes procedures for transmitting knowledge and practices across different work situations/settings and time.

Systemic Understanding: A holistic view of an organization and its inter-relationships, an understanding of the fabric of relationships and the likely effect of interrelated actions.

Knowledge Management on the Web

Ruidong Zhang
University of Wisconsin-Eau Claire, USA

INTRODUCTION

The importance of knowledge management has been recognized both in academia and in practice. In recent years, corporations have started talking about knowledge management, organizational learning, organizational memory, and computerized support. A few years ago, Microsoft®'s awareness of knowledge management and corporate memory was demonstrated by Bill Gates through his keynote speeches in the second and third Microsoft's CEO summits that attracted quite a few CEOs and other corporate executives from Fortune 1000 companies. Gates (1998) outlined his vision through a term he coined "digital nervous system," which is an integrated electronic network that can give people the information they need to solve business and customer problems. An effective digital nervous system should include access to the Internet, reliable e-mail, a powerful database, and excellent line-of-business applications, and should transform three major elements of any business: the relationships to customers and business partners—the e-commerce element; the information flow and relationships among workers within a company—the knowledge management element; and the internal business processes—the business operations element. The recent release of Windows® Tablet PC® edition is an example of a Microsoft tool that supports the concept of digital nervous system.

Even though knowledge management as a conscious practice is still young (Hansen et al., 1999), using information technology to support knowledge management is being explored and is well under way in many organizations. The Web technologies are not only changing the landscape of competition and the ways of doing business but also the ways of organizing, distributing, and retrieving information. Web-based technology is making effective knowledge management a reality, and Web-based knowledge management systems have been developed and deployed.

Currently, Web-based technology is enabling the management of knowledge at the document management level, in contrast to the traditional record-level data management. The record-level data management is basically the focus of traditional database management systems. The document level is higher than the record level. For example, we generally handle daily problems through communicating with each other by using documents and exchanging ideas or perspectives about an issue, rather than dealing with database fields or records. Document-level information management is generally viewed as a lower level of knowledge management.

In this chapter, Web-based knowledge management is explored. Four representative types of Web-based knowledge management models are identified and studied. The study of these models would shed light on the effective management of organizational knowledge, what should be contained in a knowledge management system, the levels of knowledge management support, and how knowledge management support systems can be technically implemented. This chapter is organized as follows. In the next section, some theoretical issues about knowledge management are reviewed. Then, it is justified why Web technology is an enabling technology to the effective knowledge management and why Web-based knowledge management is desirable. Then, the four types of Web-based knowledge management models are discussed and compared. Finally, the conclusion section summarizes the results of this chapter and discusses future directions of Web-based knowledge management.

BACKGROUND

Traditional Information Systems versus Knowledge Management Systems

Traditional information systems were developed to capture data about daily business transactions (transaction-processing systems), and to access, process, and analyze those internal and external data to generate meaningful information to support management [management information system (MIS), decision support system (DSS), or enterprise integration system (EIS)]. These traditional systems help make an organization operate smoothly. However, they were developed at a time when the importance of knowledge management was not recognized.

They all emphasize quantitative data processing and analysis. But an effective organization does not rely on quantitative analysis alone to deal with its problems. The nonquantitative side, such as knowledge creation and management, mental models, document sharing, human communications, information exchange, and meaning making, play a great role in an organization's growth and development. Thus, the nonquantitative areas also need to be supported. Knowledge management systems are supposed to fulfill this role. In other words, knowledge management systems should complement traditional systems in providing nonquantitative side support. A difficult task is to define what needs to be contained in the knowledge management system. A lot of existing studies provide only theoretic suggestions. A study described and discussed 10 knowledge management frameworks (Holsapple & Joshi, 1999). These frameworks are generally concentrated on conceptual knowledge creation or knowledge-building activities. They may be useful in deciding what functions a knowledge management system should eventually provide, but they fall short in suggesting what should be contained in a knowledge management system and how such a system may be implemented. In this chapter, the study of four types of Web-based knowledge management models should provide some practical advice about the content of a knowledge management system.

Knowledge versus Information versus Data and Knowledge Management

Commonly agreed, data is often defined as the raw facts, and information as the processed data. Davenport and Prusak (1998) defined knowledge as "a combination of experience, values, contextual information, and expert insight; and knowledge provides a framework for evaluating and incorporating new experiences and information." On the other hand, Applehans and coresearchers (1999) defined knowledge as the ability to turn information and data into effective action, which brings desirable outcomes for an organization. In this chapter, we follow a compromised approach to define knowledge: it is about the application of data and information for a given task so that the given task can be effectively performed.

The traditional view about the relationship between knowledge, information, and data is that knowledge is above data and information; data is a prerequisite for information; and information is a prerequisite for knowledge. This theory can be simply illustrated by the following diagram:

Data $\rightarrow$ Information $\rightarrow$ Knowledge

The second view, also a different view, is called a reversed knowledge hierarchy, which suggests that we first need to have the knowledge about what information we want, and then we will know what data to look for (Tuomi, 1999). In other words, data emerge only after we have information, and that information emerges only after we already have knowledge. This view can be simply illustrated by the following diagram:

Knowledge $\rightarrow$ Information $\rightarrow$ Data

The third view is also possible. We have a large amount of data collected but fail to make use of the data to create information because of our lack of relevant knowledge. An historical example is at point, which is about the making of the everyday weather map. For a long time in history, weather data were collected, and there were rich data available. But the usefulness was limited when these data are not combined with a map (Monmonier, 1999). Once we have the relevant knowledge about how to process data, how to visualize data, the boring data start making sense and generating meanings. In this situation, the knowledge is the catalyst that transforms data into information. This process can be simply illustrated by the following diagram:

Data $\rightarrow$ Knowledge $\rightarrow$ Information

These differing logics and understandings about the relationships and the sequences between data, information, and knowledge are meaningful, depending on the context. The development of the traditional information systems (TPS, MIS, DSS) basically follows the traditional view of data, information, and knowledge, while scientific research and inference statistics generally follow the second view (the reverse hierarchy) on data, information, and knowledge. In terms of knowledge management, the third view bears more impacts and implications. This third view suggests a fundamental fact in many organizations—there are a lot of documents (data) accumulated over years, and we need to know what to do to turn them into information to support a given task. All knowledge is tacit in nature and largely resides in the human mind, and to articulate knowledge is to create information about knowledge (Stenmark, 2002). The document is one place where information about knowledge can be found. Another place is the human mind, and the identification of the right human mind (expert) for a given task must be a responsibility of knowledge management. This help explains why the awareness of corporations to identify and utilize inside expertise for a given task has increased in recent years.

To summarize, knowledge management is concerned about how documents (data) may be organized, associated, and retrieved so that meaningful information can be produced for a certain task. This also includes identifying experts whose expertise can immediately help a given task. In other words, the following understanding about knowledge management is used in this chapter: Knowledge management is about meaningful organization of data (documents) and people who have expertise about a task, so that for a given task, meaningful information can be uncovered, associated, and retrieved, and experts with matching expertise can be identified.

THE WEB TECHNOLOGY

The content of a knowledge management system is not created by one individual. The content collection and the access of the content is a collective behavior. Therefore, the technological infrastructure installed must be able to facilitate the collective behavior of knowledge management. In a study where 31 knowledge management projects in 24 companies were examined, eight factors were identified to characterize a successful project (Davenport et al., 1998). One of these factors was the use of a technology infrastructure that includes common technologies for desktop computing and communications. Web technology has provided a common technological infrastructure to support the collective knowledge management, which is justified by the following observations in a separate study conducted by Zhang and Chen (1997):

- Web-based technology uses standard transmission-control protocol/Internet protocol (TCP/IP), which is mature and is supported by almost every vendor (Panko, 1997; Telleen, 1996; Strom, 1995).
- Information can be collected, retrieved, and shared through popular browsers like Netscape Navigator and Internet Explorer.
- A home page can be quickly developed, deployed, and shared.
- There are languages specially developed for Web-based applications, such as Java or VB script. Java applets can be embedded in Web home pages. The applets are executed on the client's PC and make it possible to develop interactive home pages with instant user responses and with multimedia features.
- With Web technologies, an organization enjoys platform independence or cross-platform operation. JAVA applets do not have to be rewritten to work with PC browsers, Macintosh browsers, and UNIX browsers (Panko, 1997).

In addition to the common technical infrastructure feature offered by Web technology, there are several major reasons why Web-based knowledge management is desirable:

1. The basic unit of knowledge is at the document level, which is equivalent to the level at which human beings normally communicate. Documents are usually created to deal with particular issues, and we live our everyday lives by dealing with issues. Different from an expert system, a document-based knowledge management system cannot automatically derive solutions. Instead, its usefulness lies in its large repository of classified documents, its multi-indexed powerful searching capabilities, the links between documents, the links within a document, and the potential of including other advanced features (e.g., animation). The interpretation of the documents provided by a knowledge management system largely lies with the users. The function of a document-based knowledge management system is largely to support relevant information for a task.
2. The intranet, which is based on Web technology, is the driver for new business applications. As one study has found, corporate intranets and the Internet have made the process of finding the right expert for a given task more feasible than ever before (Yiman & Kobsa, 2000). Another study has shown that intranets can provide useful and people-inclusive knowledge management environments (Stenmark, 2002).
3. The association between documents and tasks can be easily established by creating hypertext links. Hypertext links can be created between documents and within a document. Hypertext links make explicit the meaningful documents relevant for a given task.
4. The collective behavior of knowledge management can be supported. A Web site can be easily configured to allow multiple users or contributors to edit existing documents or add new documents. When talking about the knowledge management architecture, Morey (1999) suggested that successful knowledge management architecture must have the following characteristics: be available, be accurate, be effective, and be accessible. Web-based technology has made it possible to have effective knowledge management architecture.

THE WEB-BASED KNOWLEDGE MANAGEMENT MODELS

Four Web-based knowledge management models are identified in this section. These models represent the current level of Web-based knowledge management. Nonetheless, these four types of models may not represent all of the Web-based knowledge management models. These four representative models are as follows:

- Library model
- Attachment/association model
- Directory model
- Press center model

Library Model

This model enables content-based document searches. Under this model, a large collection of documents is established. Both the attributes and the content of a document are indexed, in contrast to the traditional method where only the attributes of a document are indexed. The attributes of documents may include title, subject, author name(s), publication (creation) date, number of pages, and so on. Under this model, powerful search functions are provided, where not only these attributes (title, subject, etc.) are searched, but also the contents of documents are searched.

An example was provided by the ITKnowledge.com Web site, which is a large repository of information technology (IT)-related books. The contents of the books are fully available. The chapters in a book are hypertext-linked, and a book is essentially a set of hypertext documents. Not only the attributes of the books are classified and indexed and can be searched easily, but also the chapter titles (the content) are indexed by keywords and can be searched. This makes it possible to find a document with attributes (title, subject) that do not meet a search criterion but that may contain chapters that are relevant to the search criterion. In Figure 1, the ITKnowledge.com search screen is displayed. In Figure 2, a search result is returned. As can be seen in Figure 2, a chapter in the book *Handbook of Data Management* that is relevant to the search criterion "knowledge management" has been returned. If the book is only indexed by its title, this search result should not have been possible.

Attachment/Association Model

Under this model, information is organized around topics. If we search for a particular topic, all information associated with the topic will be returned. New information can be attached to a topic once it becomes available. In fact, anyone at any time can attach new information to a topic. The attachment creation is an ongoing process.

An example of this model is Amazon.com, which is also an example of successful e-commerce. Bookselling is its major business. Amazon.com maintains a large database of book titles. To help sell a book faster and to make users understand a book better, all information relevant to a book title is stored and organized around a book title. Therefore, essentially, the book database is more than just a collection of book attributes. Amazon.com provides users more than book titles. Suppose we want to find books that are relevant to "knowledge management."

After a title that matches the search criterion is returned, other relevant titles, and what other customers often purchase together with the current title, are returned

Figure 1. IT Knowledge.com expert search screen

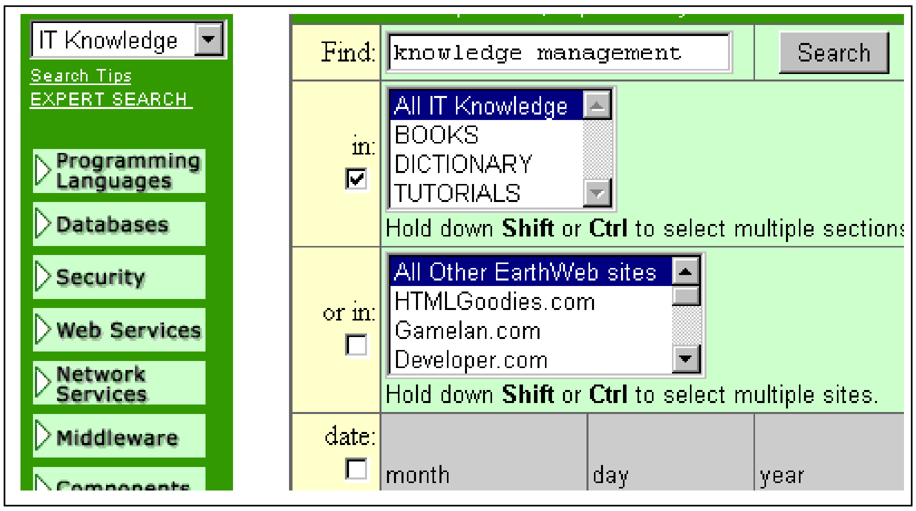
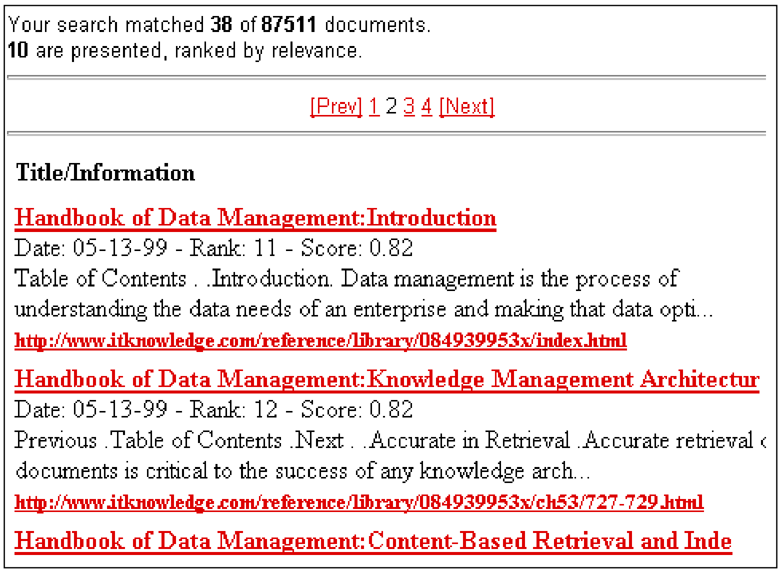


Figure 2. Search results returned



Your search matched **38** of **87511** documents.
10 are presented, ranked by relevance.

[Prev] 1 2 3 4 [Next]

Title/Information

Handbook of Data Management:Introduction
Date: 05-13-99 - Rank: 11 - Score: 0.82
Table of Contents . .Introduction. Data management is the process of understanding the data needs of an enterprise and making that data opti...
http://www.itknowledge.com/reference/library/084939953x/index.html

Handbook of Data Management:Knowledge Management Architectur
Date: 05-13-99 - Rank: 12 - Score: 0.82
Previous .Table of Contents .Next . .Accurate in Retrieval .Accurate retrieval o documents is critical to the success of any knowledge arch...
http://www.itknowledge.com/reference/library/084939953x/ch53/727-729.html

Handbook of Data Management:Content-Based Retrieval and Inde

Figure 3. The identification of expertise by topic

RTG Expertises (by Topic)

[Industry Groups] [Functional Groups] [Political & Soci
[Emerging Economies -- Countries & Related Topics] [T
[Biotechnology & Environmental Issues] [Law] [Employe

Industry Groups

Accounting, Aerospace, Agricultural Production/services, Amusement and Recreati
Policies, Astronomy, Auctions, Automotive, Auditing, Banking, Chemical, Chemica
Construction, Defense, Educational Services, Electronics, Energy, Engineering, Fina
Health Services, High Technology Industries, Housing and Mortgage Markets, Hot
Brokering, Insurance, Legal Services, Lumber, Manufacturing, Metal Mining, Motic
Products, Petroleum, Primary Metal Industries, Printing and Publishing, Radio and T
Business, Social Services, Sports, Telecommunications, Textile Mill Products, Toba

and displayed. In addition, one can learn more about a book by reading the reviews attached to or associated with the book and other customers' comments attached to the book. If a reader wants, he or she can write comments about a book and easily attach his or her comments to the existing pool of information about this book.

Directory Model

Under this model, the experts in different areas are identified, and a directory of experts is created. The areas of expertise of these experts are classified and organized. A directory of experts, together with their areas of expertise, is provided.

A representative example of this model is the Round Table Group (round.table.com). Round Table Group (RTG) was founded in 1994 with a vision of being a virtual consulting firm where business leaders, management consultants, and litigation attorneys could shop for answers to critical questions from world-class thinkers, anywhere in the world, in Internet time (STVP, 1999). According to an STVP case study, RTG has formed a worldwide network of over 3,000 Round Table scholars—professors, researchers in well-known think tanks, and other experts. Essentially, RTG's most valuable competitive asset is its directory of professors from around the world who were available to consult with clients on demand. RTG provides answers-on-demand services to its clients. In Figure 3 below, the classified area of expertise by topic is displayed.

Press Center Model

Under this model, any information that can possibly be collected, including news, relevant articles, solution providers, publications, and discussions about a task is collected. A representative model is Kmworld.com, where a rich collection of relevant information about knowledge management can be found. If one is interested in knowledge management, a good starting point would be to visit a site like Kmworld.com. One could be overwhelmed by the information available, such as discussions, news, and solution providers. But, by spending some time and effort, one could figure out what is going on with a task, the current status, the issues, and history information on some issues.

A Comparison of the Models

The four types of Web-based knowledge management models as discussed above reflect the current level of Web-based knowledge management, which can be summarized as the content-based information retrieval and topic-oriented information organization.

These four models share the following common characteristics:

- Are basically document-based and content-indexed
- Have powerful searching capabilities
- Focus on a specific application area (task-specific: IT knowledge, bookselling)
- Offer anytime, anywhere access
- Have potential to have more advanced features (e.g., multimedia, animation)
- All have an external orientation and are customer focused

On the other hand, the four models discussed above also have unique characteristics, which are indicated in Table 1.

FUTURE TRENDS AND CONCLUSIONS

Current Web-based knowledge management is essentially at the document management level—a lower level of knowledge management, but indeed beyond pure document management (pure classified collection of documents in a central file cabinet) to allow content-based retrieval, distributed access, and topic-oriented information organization and association. In this chapter, four different Web sites that represent four different models of Web-based knowledge management are discussed. But it is believed that a Web site could employ multiple models at the same time. For example, a Web site can be constructed based on the library model while supporting the association and attachment model. As a matter of fact, ITKnowledge.com is planning to include the association and attachment model in its site. It is expected that a sophisticated Web site will provide a menu of knowledge models for its users to choose from so that the potential of a Web site can be fully utilized.

A study, conducted by Hansen, Nohria, and Tierney (1999), discussed two strategies of knowledge management, based on the authors' studies on the practice of knowledge management in consulting firms, with computer vendors, and with health care providers. One strategy is called the personalization strategy, where knowledge is closely tied to the person who developed it and is shared mainly through direct person-to-person contacts. The computer information systems are used mainly to help people communicate knowledge, not to store it. Other examples like the use of groupware products to support the knowledge management process also fall into this category. This may be called the communication-based knowledge management level.

Table 1. Key characteristics of Web-based knowledge management models

Model	Key Characteristic
Library	Content-based document retrieval
Association/attachment	Topic-oriented information organization/attachment
Directory	A well-organized directory of experts that can be searched for a given task
Press Center	A clearinghouse with rich organized information

Another strategy, from the same study conducted by Hansen, Nohria and Tierney (1999), is called codification strategy, where knowledge (actually, a variety of documents) is codified and stored in databases to be used by anyone in the company. The knowledge is extracted from documents and may take the forms of interview guides, work schedules, benchmark data, or market segmentation analysis. These knowledge forms are codified and stored into a database with search capabilities. This is may be called the document-based knowledge management level.

The existing Web-based knowledge management, as evidenced by the four types of models identified and discussed in this chapter, basically corresponds to the knowledge management levels (especially the second). However, these levels are considered as lower levels of knowledge management, because they do not consider the support of meaning interpretation or sense making, they are generally text based, and they do not challenge whether the knowledge inherent in the document is accurately transmitted (knowledge transfer). But, after all, good steps on the right track have been made.

As future research directions about Web-based knowledge management, the higher levels of knowledge management may need to be addressed. Other functional areas should get involved, such as sales and marketing, customer support, and research and development. Other models about Web-based knowledge management may also need to be identified. Automatic and intelligent knowledge extraction and retrieval (knowledge agents) should also be studied for Web-based knowledge management.

REFERENCES

Applehans, W., Globe, A., & Laugero, G. (1999). *Managing knowledge: A practical Web-based approach*. Reading, MA: Addison-Wesley.

Davenport, T. H., De Long, D., & Beers, M. (1998). Successful knowledge management projects. *Sloan Management Review, 39*(2).

Davenport, T., & Prusak, L. (1998). *Working knowledge: How organizations manage what they know*. Cambridge, MA: Harvard Business School Press.

Gates, B. (1998, 1999). Keynote speech for the Microsoft's second and third annual CEO summit meeting, Seattle, WA. Retrieved September 26, 2003, from http://www.microsoft.com/billgates/speeches.asp

Hansen, M. T., Nohria, N., & Tierney, T. (1999, March–April). What's your strategy for managing knowledge. *Harvard Business Review*.

Holsapple, C. W., & Joshi, K. D. (1999). Description and analysis of existing knowledge management frameworks. In *Proceedings of the 32nd Hawaii International Conference on System Sciences*.

Monmonier, M. S. (1999). *Air apparent : How meteorologists learned to map, predict, and dramatize weather*. Chicago, IL: University of Chicago Press.

Morey, D. (1999). Knowledge management architecture. In *Handbook of data management*. Boca Raton, FL: CRC Press.

Panko, R. R. (1997). *Business data communications*. New York: Prentice Hall.

Stenmark, D. (2002). Information vs. knowledge: The role of intranet in knowledge management. In *Proceedings of the 35th Hawaii International Conference on System Sciences*.

Stewart, T. A. (1997). *Intellectual capital: The new wealth of organizations*. New York: Doubleday/Currency.

Strom, D. (1995). Creating private intranets: Challenges and prospects for IS. Retrieved from *http://www.attachmate.com/Intranet*

STVP (Stanford Technology Ventures Program). (1999, February 28). Round Table group at a crossroads: Market maker or market driven?

Telleen, S. L. (1996). *The intranet architecture: Managing information in the new paradigm*. Amdahl Corporation.

Tuomi, I. (1999). Data is more than knowledge. In *Proceedings of the 32nd Hawaii International Conference on System Sciences*.

Yiman, D., & Kobsa, A. (2000). DEMOIR: A hybrid architecture for expertise modeling and recommender systems. In *Proceedings of the IEEE Ninth International Workshops*, June 14–16.

Zhang, R., & Chen, J. (1997). An intranet architecture to support organizational learning. In *Proceedings of 1997 DSI Conference*.

KEY TERMS

Data, Information, Knowledge: Data is often defined as the raw facts, and information as the processed data. Knowledge is about the application of data and information for a given task so that the given task can be effectively performed.

Knowledge management: This is concerned about how documents (data) may be organized, associated, and retrieved so that meaningful information can be produced for a certain task. This also includes identifying experts whose expertise may help a given task.

Web-based knowledge management: This is about using Web technologies to support knowledge management processes.

Web-based knowledge management model: This is about how knowledge may be collected, accessed, and presented on the Web platform. This chapter identifies four Web-based knowledge management models.

Web-technology: This refers to the combination of the Internet communication protocols (TCP/IP), networks, Web server (e.g., IIS), and client (e.g., Opera) software, and the languages that can be used to develop Web applications.

Knowledge Management Systems Acceptance

Fredrik Ericsson
Örebro University, Sweden

Anders Avdic
Örebro University, Sweden

INTRODUCTION

Knowledge management is a set of systematic actions that organizations can take to obtain the greatest value from the knowledge available to it (Davenport & Prusak, 1998). Systematic means that knowledge management is made up of intentional actions in an organizational context. Value means that knowledge management is measured according to how knowledge management projects contribute to increased organizational ability (see for example Prieto & Gutiérrez, 2001; see Goldkuhl & Braf, 2002, on the subject of organizational ability). The motivation for knowledge management is that the key to competitive advantage for organizations in today's business world is organizations' ability to manage knowledge (Nonaka & Takeuchi, 1995; Davenport & Prusak, 1998). Knowledge management as an intentional and value-adding action is not easy to accomplish in practice (Scarbrough & Swan, 1999). Scarbrough and Swan (1999) present several case studies in knowledge management, successful and unsuccessful in their respective knowledge management projects. A major point and lessons learned from the case studies is that prevalent approaches in knowledge management overstate technology and understate how technology is implemented and applied.

To succeed with knowledge management, encompassing development of information technology-based information system, some requirements have to be fulfilled. An important aspect in the development process is system acceptance. Implementation is at large a process of acceptance. Implementation is the process where the system becomes an integrated part of the users' or workers' work practice. Therefore implementation is essential to make a knowledge management project successful in order attain an increased organizational ability and to succeed with knowledge management.

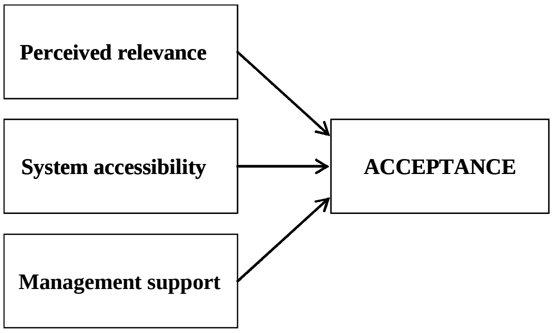
ISSUES OF KNOWLEDGE MANAGEMENT—SYSTEMS AND ACCEPTANCE

In this section we provide broad definitions and discussion of the topics to support our positions on the topics of knowledge management and systems acceptance.

MANAGING KNOWLEDGE

Work in knowledge management has a tendency to omit social or technological aspects by taking on one of two perspectives on knowledge management, the anthropocentric or the technocratic view (Sveiby, 2001; Swan, 1999). The anthropocentric and the technocratic views represent two contradictory views on knowledge management and can be summarized as technology can or technology cannot. The gap between the anthropocentric and technocratic view depends on a difference of opinions concerning the notion of knowledge. The technocratic view conceives knowledge to be some organized collection of data and information, and the anthropocentric view conceives knowledge to reside in humans, not in the collection (Churchman, 1971; Meredith & Burstein, 2000). Our conception of knowledge is that of the anthropocentric view. Taking on an anthropocentric view on knowledge management does not mean that we discard knowledge management technologies; we rather take on a balanced view on the subject. Information technology can support knowledge management in an organization through a number of different technological components, for example intranets, extranets, data warehouses, and database management systems (Borghoff & Pareschi, 1998; Tiwana, 2000; Ericsson & Avdic, 2002). The point in taking on an anthropocentric view of knowledge management is not to lose sight of the knower who gives meaning to the information and data found in IT-based knowledge management systems.

Figure 1. Requirements of Acceptance Model (Ericsson & Avdic, 2003)



KNOWLEDGE MANAGEMENT SYSTEMS

Information systems can include either operative or directive and decision support information (Langefors, 1966; Yourdon, 1989). Operative systems provide system users with information necessary in workers' daily work, while directive and decision support systems provide system users with information that improves the quality of decisions workers make in daily work. Knowledge managements systems are systems developed to manage knowledge directly or indirectly to give support for an improved quality of a decision made in workers daily work, and as an extension, an increased organizational ability. A knowledge management system typically includes directive information, for example in guiding a user's choice in a specific work situation. Such systems are often optional in the sense that users can deliberately refrain from using the system and/or refrain from taking the directed action. Accordingly, user acceptance is crucial for the degree of usage of knowledge management systems.

ACCEPTANCE OF TECHNOLOGICAL SYSTEMS

Technology acceptance has been subject of research by, for example, Davis, Bagozzi, and Warshav (1989), who developed the well-known Technology Acceptance Model (TAM) and later a revised version of the original model, TAM2 (Venkatesh & Davis, 2000). TAM is an explanative model explaining user behavior of computer technologies by focusing on perceived ease of use, perceived usefulness, attitude towards use, and behavioral intentions as determinants of user behavior. TAM2 is an extension of the original model including external factors related to perceived usefulness.

The framework for system acceptance, Requirements of Acceptance Model (RAM) have some resemblances with TAM and the later TAM2. RAM is in comparison with TAM descriptive in nature. Workers' work practice is treated as an integrated element of RAM, compared with not being treated as a determinant of system use in the original TAM and as an external factor in TAM2. Further, RAM covers acceptance of knowledge management systems, and TAM/TAM2 cover a broad range of computer technologies. RAM systematically acknowledges factors important in implementation of knowledge management systems to gain acceptance of such systems.

REQUIREMENTS OF THE ACCEPTANCE MODEL

We perceive acceptance to be a function of perceived relevance, systems accessibility, and management support. Together these elements constitute our framework RAM. In this section we present the requirements of acceptance in RAM. The Requirements of Acceptance Model is illustrated in Figure 1.

PERCEIVED RELEVANCE

The workers, who are to use the system, must perceive the knowledge management system as relevant. Since it is possible for workers to work without using the system, it has to be obvious that usage of the system implies adding value to the work result. An additional aspect of relevance related to perceived relevance is how the system should be integrated in running work, that is, to make the system an integrated part of the workers' work practice.

In summary, perceived relevance is about workers, who are to use the system, perceiving the system as (Ericsson & Avdic, 2003)

- adding value to the work results; and
- being integrated in running work.

ACCESSIBILITY

To obtain acceptance of knowledge management systems, accessibility has to be satisfactory. It must be accessible to the workers who are to use the system. Accessibility is a question of who is to be the user (type of workers concerning organizational position), what action and work the system is to support (daily work, product development, innovation, etc.), where users get access to the system (the physical access), when the system is ready to use, and how the system's interface fulfills the goal of the system.

In summary, systems accessibility is about (Ericsson & Avdic, 2003):

- knowing who the user is;
- systematizing the actions workers perform in the work practice the system is to support;
- deciding the system's physical access;
- securing a certain degree of usage before the system is put into operation; and
- ensuring the system's design meets the goals of the system.

MANAGEMENT SUPPORT

Management support is vital according to many models on information systems development, especially when the system is a directive/decision support system (Yourdon, 1989). Knowledge management systems are typically directive systems, and workers have a choice in deciding whether to use the system or not. Management support is important to stress the value for workers to use the system and to make conditions for workers to do so.

DEVELOPMENT IS A PROCESS OF ACCEPTANCE

There must be a fit between workers' work practice and technology to get acceptance of knowledge management systems. The technology used to create a knowledge management system must fit the actions workers perform in their work practice. On an overall level there must be a fit between technology and actions performed by individual workers, and between individual workers and the organization as a whole, thus forming a coherent whole. It is in the development of knowledge management systems that the requirements of acceptance are fulfilled. A common conception concerning information systems development is that it constitutes analysis, design, construction, and implementation of information systems (Hirschheim, Klein & Lyytinen, 1996).

The groundwork for acceptance is made during the design, but foremost when implementing the system. Workers who are to use the system should be engaged at an early stage of the development process. The point of including workers at an early stage is to acquaint users with the system and the purpose of the system. Further, this is an opportunity for workers to influence the system's design and content. The most prominent aspect addressed when involving workers at an early stage is that of choosing and determining the meaning of crucial concepts managed by the system. Crucial concepts managed by the system are the knowledge represented in the system, and by determining concepts, knowledge represented in the system takes on a systematized character. Further, by involving the workers in the process of choosing and determining the meaning of crucial concepts managed by the system, the knowledge represented in the system does not lose its origin or meaning. The point is to keep the knowledge represented in the system within a frame of understanding or meaning, as perceived by workers. A knowledge management systems should be seen as a tool developed to support workers in learning and acquiring knowledge about actions taking place at work. This requires closeness between how concepts are perceived by

Table 1. Summary of RAM (Ericsson & Avdic, 2003)

Perceived relevance—Workers, who are to use the system, have to perceive the system as:
- Adding value to work results
- Being integrated in running work

Systems accessibility—System accessibility is about:
- Knowing who the user is
- Systematizing actions workers perform in the work practice the system is to support
- Deciding the physical location where users get physical access to the system
- Securing usage of the system before it is put into operation
- The systems' design must meet up to the goals of the system

Management support—Fundamental because management authorizes development of systems workers and how such concepts are represented in a system.

FUTURE TRENDS

Research on technology acceptance (i.e., Davis et al., 1989; Venkatesh & Davis, 2000) has focused on user behavior of computer technologies. RAM is developed for and is used to assess acceptance of knowledge management systems. Acceptance has not been a crucial issue within the knowledge management area. A problem with knowledge management systems is that they work in theory, but seldom in practice (Wickramasinghe, 2003). A contributing factor to that picture may very well be that of having overlooked usage-related problems connected to knowledge management systems. In that sense, knowledge management systems acceptance can be expected to be an area for further research in the future.

CONCLUSION

Acceptance of knowledge management systems is a function of perceived relevance, systems accessibility, and management support. Together these elements constitute our framework RAM. RAM is summarized in Table 1.

The Requirements of Acceptance Model point towards several important aspects concerning relevance, accessibility, and support. The groundwork for system acceptance is the development process. Development is very much a process of acceptance as a process of developing the system itself. Through requirements of acceptance, knowledge management systems can remain and continue to be a contributing factor for the organization's ability to do business.

REFERENCES

Borghoff, U.M. & Pareschi, R. (Eds.). (1998). *Information technology for knowledge management*. Berlin, Heidelberg: Springer-Verlag.

Churchman, C.W. (1971). *The design of enquiring systems: Basic concepts of systems and organization*. New York: Basic Books.

Davenport, T. & Prusak, L. (1998). *Working knowledge: How organizations manage what they know*. Boston: Harvard Business School.

Davis, F.F., Bagozzi, R.P. & Warshaw, P.R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science, 35*(8), 982-1003.

Ericsson, F. & Avdic, A. (2002). Information technology and knowledge acquisition in manufacturing companies: A Scandinavian perspective. In E. Coakes, D. Willis & S. Clarke (Eds.), *knowledge management in the socio-technical world. The grafitti continues*. London: Springer-Verlag.

Ericsson, F. & Avdic, A. (2003). Knowledge management systems acceptance. In E. Coakes (Ed.), *Knowledge management: Current issues & challenges* (pp. 39-51). Hershey, PA: Idea Group Publishing.

Goldkuhl, G. & Braf, E. (2002). Organisational ability: Constituents and congruencies. In E. Coakes, D. Willis & S. Clarke (Eds.), *Knowledge management in the socio-technical world. The grafitti continues* (pp. 30-42). London: Springer-Verlag.

Hirschheim, R., Klein, H.K. & Lyytinen, K. (1996). Exploring the intellectual structures of information systems development: A social action theoretic analysis. *Accounting, Management & Information Technology, 6*(1/2), 1-64.

Langefors, B. (1966). *Theoretical analysis of information systems*. Lund: Studentlitteratur.

Meredith, R. & Burstein, F. (2000). Getting the message across with communicative knowledge management. *Proceedings of the Australian Conference on Knowledge Management and Intelligent Decision Support* (ACKMID'2000) (pp. 43-55). Melbourne: Australian Scholarly Publishers.

Nonaka, I. & Takeuchi, H. (1995). *The knowledge creating company: How Japanese companies create the dynamics of innovation*. New York: Oxford University Press.

Prieto, I.M. & Gutiérrez, E.R. (2001). A contingency perspective of learning and knowledge management in organizations. In D. Remenyi (Ed.), *Proceedings of the 2nd European Conference on Knowledge Management* (pp. 487-502). Slovenia: Bled School of Management.

Scarbrough, J. & Swan, J. (Eds.). (1999). *Case studies in knowledge management*. London: Institute of Personnel and Development.

Sveiby, K.-E. (2001, April). *What is knowledge management?* Retrieved June 28, 2002, from *www.sveiby.com.au*.

Swan, J. (1999). Introduction. In J. Scarbrough & J. Swan (Eds.), *Case studies in knowledge management*. London: Institute of Personnel and Development.

Tiwana, A. (2000). *The knowledge management toolkit. Practical techniques for building a knowledge management system*. Upper Saddle River, NJ: Prentice-Hall.

Venkatesh, V. & Davis, F.D. (2000). A theoretical extension of the Technology Acceptance Model: Four longitudinal field studies. *Management Science, 46*, 86-204.

Yordon, E. (1989). *Modern structured analysis*. Englewood Cliffs, NJ: Prentice-Hall.

Wickramasinghe, N. (2003). Do we practice what we preach? Are knowledge management systems in practice truly reflective of knowledge management systems in theory? *Business Process Management Journal, 9*(3), 295-316.

KEY TERMS

Anthropocentric View of Knowledge: Knowledge resides in humans.

Information Systems Development: Constitutes analysis, design, construction, and implementation of information systems.

Knowledge: Knowledge is personal and talked about and may thus be public and shared among a group of people who have a common frame of reference, providing means for people to make sense of and apply knowledge in practice.

Knowledge Management: The name given to the set of systematic actions that an organization can take to obtain the greatest value from the knowledge available to it.

Knowledge Management Systems: Typically, directive systems developed to manage knowledge directly or indirectly to give support for an improved quality of a decision made in workers' daily work, and as an extension, an increased organizational ability.

Perceived Relevance: Workers who are to use the system perceive the system as adding value to the work results and being integrated in running work.

Systems Acceptance: A function of perceived relevance, systems accessibility, and management support.

Systems Accessibility/Development: Knowing who the user is, systematizing the actions workers perform in their work practice the system is to support, deciding the system's physical location, securing a certain degree of usage before the system is put into operation, and ensuring the system's design meets the goals of the system.

Technocratic View of Knowledge: Knowledge is an organized collection of data and information.

Knowledge, IT, and the Firm

Petros A. M. Gelepithis
Kingston University, UK

INTRODUCTION

No company has ever existed or will ever exist without knowledge. Still, it was only recently that knowledge started being heralded as the way forward (Drucker, 1993; Itami, 1987; Toffler, 1990). This may explain why in the business world, knowledge management (KM) is still perceived in two substantially different senses: (a) as synonymous to information management (e.g., Dempsey, 1999; Vernon, 1999) and (b) as a distinct area of study and practice dealing with the management of knowledge (e.g., Newing, 1999; Zack, 2003). In contrast, the academic world sees knowledge and information as related but fundamentally distinct. Furthermore, the vast majority of both of these communities has focused on the managerial or social aspect of KM (see, for example, Birkinshaw & Sheehan, 2002; Davenport & Glaser, 2002; Davenport, Thomas, & Cantrell, 2002; Gupta & Govindarjan, 2000). The *nature* of knowledge and its implications for management have been largely ignored. The limited work considering knowledge issues falls into the four categories below.

1. Knowledge is self-explainable and, therefore, in need of no further consideration (e.g., Newing, 1999).
2. Knowledge is self-explainable and classifiable into several commonsense categories (see, e.g., Quinn, Baruch, & Zien, 1997; Savage as cited in Skyrme, 1999).
3. Knowledge is attempted to be explained or defined without taking into account the vast relevant work done in epistemology and cognitive science. Primary examples are Borghoff and Pareschi (1998) and Davenport and Prusak (1998).
4. Organisational knowledge creation is a social interaction between tacit knowledge and explicit knowledge (Nonaka, 1991; Nonaka & Takeuchi, 1995).

The first three attempt to deal with the nature of knowledge and give the impression that there are no problems in an area beset with significant issues, whereas the fourth pays serious attention to the fundamental issue of knowledge creation.

Nonaka and Takeuchi's (1995) theory consists of two interacting knowledge spirals. The epistemological one is based on the distinction between tacit and explicit knowledge; the ontological one is based on the widely accepted distinction between the individual and the organisation. Their important contribution is an excellently written expansion of their working hypothesis called "knowledge conversion," namely, "human knowledge is created and expanded through social interaction between tacit knowledge and explicit knowledge" (p. 61). This important "dichotomy" is one of seven that form the basis of their theory, specifically, (a) tacit/explicit, (b) body/mind, (c) individual/organisation, (d) top-down/bottom-up, (e) bureaucracy/task force, (f) relay/rugby, and (g) East/West. Knowledge conversion comprises four modes: socialization (from tacit to tacit), externalisation (from tacit to explicit), combination (from explicit to explicit), and internalisation (from explicit to tacit). These four modes "constitute the 'engine' of the entire knowledge-creation process" (p. 57).

There are four weaknesses in their approach. First, their working hypothesis is characterised by a coarse grain size: Tacit and explicit knowledge are left unanalysed. As a consequence, no actual mechanisms for knowledge creation are proposed. Second, despite an impressively long index on knowledge, the issue of the *nature* of knowledge is ignored. Third, their synthesis of the seven dichotomies, although a highly welcome attempt in bridging unnecessary gaps, it is still biased since their underlying "model of knowledge creation favors the Japanese view" (Nonaka & Takeuchi, 1995, p. 237). Finally, their wide-ranging literature review fails to take into account—or refute—the most widely accepted metaphysical position, namely, physicalism.

The next section presents the results of a unified theory of mind (Gelepithis, 1984, 1989, 1991, 1997, 2002, 2004, in press) that are relevant to the issues of knowledge creation and the nature of knowledge, and form the background to the future-trends section.

BACKGROUND

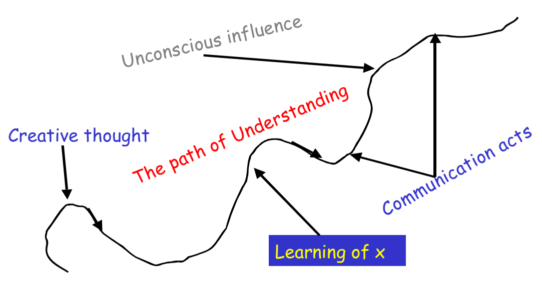
In accordance with contemporary physicalism (a position accepted by the majority of scientists and philosophers), individual human knowledge should, primarily, be seen as neurally realisable. When externalised, in the form of written language, individual human knowledge becomes fossilised. It becomes a snapshot and loses its capacity to initiate near-immediate action. Furthermore, and most crucially, it may be entirely meaningless to virtually all other humans. As a matter of fact, after the passage of some time, it may become unintelligible even to the very human who first externalised it. In other words, externally represented human knowledge becomes information.

It follows that organisational knowledge (sometimes known as organisational memory) may refer to either knowledge or information. Specifically, it may refer to employed people and their individual knowledge, or it may refer to an organisation's information. The latter needs to be interpreted by a human to be useful in any way. Interpreted information becomes internalised, possibly assimilated, and subsequently may trigger, or be involved in, knowledge creation. The central mechanism responsible for human knowledge creation is the process of understanding. It is an invariant neurophysiological process with many significant contributors to its cause as the Figure 1 illustrates. This complex system of interacting processes I call the knowledge nexus.

Through successive instantiations of the process of understanding over a period of time—quite often on the same topic—humans accumulate knowledge. The end result of understanding embodies aspects of individual human knowledge. The knowledge of human H at time t is the *system of understandings* that H has reached by that time. Eventually, an overall complex system of premises and primitives is developed comprising the axiomatic base of H's knowledge-based action. To move away from the subjectivity of individual human knowledge, communication is required. It contributes the component of breadth and intersubjective agreement characterising collective human knowledge.

Because the knowledge nexus is neurally realisable, it is opaque and most likely never fully presentable. Nevertheless, through language, the axiomatic base is externally representable and potentially formalisable. This interdependence and foregoing analysis should have made clear the inseparable link between biology and human language, that is, the inseparability of the subjective and the objective.

Figure 1. Aspects of the proces of understanding



The next section draws upon the background already presented to outline two major directions concerning (a) information and communication technologies (ICT) and (b) the economy.

FUTURE TRENDS

Before embarking on a sketch of the two major future trends, it is worth emphasising that it is the development of the appropriate environment that would nurture employees' knowledge nexus, which is important for business innovation. Successful management will be that which can realise both the near-unlimited scope and the fundamental limits of that nexus. It is in this respect the management of a company that is significant rather than futile attempts to manage knowledge.

ICT and, increasingly, artificial intelligence (AI) are at the forefront of developments providing aids for addressing problems associated with the knowledge nexus. The majority of such problems include the identification, acquisition, and sharing of information, as well as the creation of knowledge. Books and edited collections addressing aspects of information-related or knowledge nexus problems abound (e.g., Borghoff & Pareschi, 1998; Hlupic, 2003; Quinn et al., 1997; Skyrme, 1999). What is much less widely available is discussion of the scope and limits of technology.

The scope of AI and ICT technology is enormous. The following three types of problems cover an extremely large space with far-reaching potential consequences for our society.

- Overall integration of information sources and tools.
- Identification of appropriately specified information through the use of search engines.
- Formalisation of certain aspects of human knowledge through R&D in knowledge representation and reasoning.

In all these cases, advanced technology can be a very significant help. It has to be realised, though, that AI and ICT systems cannot, on their own, either create or share knowledge. This is a point that is very often overlooked with serious negative consequences. Several people have noted that despite the increasing use of artificial aids, the human remains in the loop (see, for instance, Cross & Baird, 2000; Senge & Carstedt, 2001). Actually, no future technology can possibly take the human(s) out of the loop of human knowledge creation. Let us briefly see the reason for this intrinsic limit.

AI and ICT systems may be distinguished into two possible categories: human based and machine based primitives (Gelepithis, 2001). The latter is not yet available. The former type of AI and ICT systems, whether creative or not, are automated axiomatic systems with inbuilt procedures for drawing consequences from their human premises. At worst, they are ad hoc systems capable of providing a cost-effective and, quite often, enhanced solution with respect to their human counterparts. A better solution, nevertheless, does not constitute creation of new knowledge. In the best-case scenario, they are capable of producing all possible consequences, some of which are bound to be new. In other words, "creative" AI systems can produce *new consequences* of existing knowledge. They are incapable, though, of producing *novel* human knowledge. Such knowledge involves *evaluation* of human premises and that, in turn, requires the full spectrum of human primitives, a task shown to be technologically impossible (Gelepithis, 1991). Finally, the highly desirable and potentially fruitful capability of current AI and ICT systems to produce new consequences of existing knowledge has to be ultimately evaluated by humans.

It should be noted that although humans are the only creators of novel human knowledge, ICT and, in particular, AI can facilitate the creation of such knowledge through the development of tools enabling (a) increased connectivity within the semantic system of humans and (b) increased and enhanced human-human communication. The latter should be what Skyrme (1999) calls knowledge networking. It is along these lines that a brighter future for business innovation lies.

As with technology, the intrinsic limit of economy is derived from the nature of knowledge and its associated mechanisms of creation, as sketched in the previous section. In the case of the economy, the limit concerns the scope of the notion of market. Specifically, the notion of a knowledge market is not realisable.

Claims for or against the possibility of a knowledge market are few and far between. They also vary in how convincing they are. At one extreme, reasoning is reduced to blind faith: "Given time, market forces will undoubtedly take care of the situation" (Burton-Jones, 1999, p. 221). At another extreme, Gamble and Blackwell (2001, p. 185) remark, "It is almost axiomatic in management that what you cannot measure you cannot manage." Unfortunately, the next step they go to is to create a metric for a "knowledge management initiative." It is unclear whether a "shifting the goalposts" approach is being practiced or whether an implicit belief against the possibility of a knowledge market is being held.

A more substantial argument is put forward by Davenport and Prusak (1998). Their starting point is that knowledge markets exist not as "pure" markets (i.e., markets that operate solely in economic terms). They claim that a price system for a knowledge market revolves around the notions of money, reciprocity, reputation, altruism, and trust. In a nutshell, they recognise that money is far from adequate and believe that reciprocity, reputation, and altruism (these three in diminishing degree), as well as their combinations, constitute the substitute for money. Since transactions involving reciprocity, reputation, and altruism do not constitute payment in the traditional, monetary sense of the term, they are forced to introduce the notion of trust as the necessary factor for the workings of such a market. But trust is one of the goods (the other two are loyalty and truth telling), which "cannot, in principle, be taken as commodities in the market sense" (Arrow, 1974, p. 22).

One may take Arrow's principle as a basis for an argument against the possibility of a knowledge market. It is nevertheless an argument based on extrapolation and henceforth not fully convincing. Our knowledge-based argument for the impossibility of a knowledge market is summarised below.

Markets require two elements for their existence: (a) an object to be transacted among their players and (b) a price system as its mechanism for the transactions.

Some may argue that knowledge markets have existed at least since humans started exchanging goods they produced themselves. For, any good produced—or procedure for making goods—was bound to involve the use or description of some knowledge. Nevertheless, what was traded was the good or procedure, not the knowledge involved.

What about information such as a piece of advice, a report, or an academic paper? Could such information not be transacted? No; since the meaning of some information depends on the receiver, its price cannot be fixed. Put differently, since the interpretation of information depends on the individual, its value depends on the potential client. Therefore, no common price can be assigned, no metric can be designed. The following two examples should clarify our point. First, what should be the value of the 21st-century equivalent of Newton's laws? Second, how much should one pay for that imperceptible, yet consciously made, grimace that reveals a no-purchase of company X?

True, for thousands of years now, people have been able to sell information to someone able to use it. Nevertheless, the circle of potential buyers and sellers was and will continue to be extremely limited because information is only subjectively evaluative. Expert advice is valuable only to a small number of people, quite often to a single individual only. A knowledge market is not viable because individual human knowledge is neurally realisable, and collective human knowledge is only individually interpretable and assessable (see the background section). It follows that our resources will be better spent in creating a knowledge environment rather than the chimera of a knowledge market. Sustainable business innovation requires near-continuous knowledge creation, and that requires in turn a knowledge environment. This is a difficult but necessary objective. It is difficult because knowledge has no airtight compartments.

CONCLUSIONS

Knowledge management is both mistaken and indispensable. It is mistaken because no one can manage something that takes place inside another one's mind. It is indispensable because it brought to managers' consciousness what they always relied upon: human knowledge.

Innovations (both minor and major) will come with improvements of search engines and further formalisation of the formalisable part of collective human knowledge. Breakthroughs are more likely to come from innovative hybrid designs (i.e., novel human-AI systems interactions). Both require humans in an environment that can nurture their knowledge nexus. Managers should act accordingly.

Finally, because of the strong push for economic, political, and cultural globalisation, one issue becomes of paramount importance. How could the particular cultures of all the peoples be synthesised into an Earthian culture rather than be obliterated and replaced by the monolithic pursuit of profit? Companies should understand the key cybernetic principle: variety requires variety. Innovation requires variety, too.

REFERENCES

Arrow, K. (1974). *The limits of organization.* New York: W. W. Norton & Company Inc.

Birkinshaw, J., & Sheehan, T. (2002). Managing the knowledge life cycle. *MIT Sloan Management Review,* 75-83.

Borghoff, U. M., & Pareschi, R. (Eds.). (1998). *Information technology for knowledge management*. New York: Springer.

Burton-Jones, A. (1999). *Knowledge capitalism: Business, work, and learning in the new economy*. Oxford: Oxford University Press.

Cross, R., & Baird, L. (2000). Technology is not enough: Improving performance by building organizational memory. *Sloan Management Review, 41*(3), 69-78.

Davenport, T. H., & Glaser, J. (2002, July). Just-in-time delivery comes to knowledge management. *Harvard Business Review,* 107-111.

Davenport, T. H., & Prusak, L. (1998). *Working knowledge: How organizations manage what they know*. Boston, MA: Harvard Business School Press.

Davenport, T. H., Thomas, R. J., & Cantrell, S. (2002, Fall). The mysterious art and science of knowledge-worker performance. *MIT Sloan Management Review*, 23-30.

Dempsey, M. (1999). The need for vigilance in the information battle. In *Knowledge Management*, Financial Times Survey, p. II, November 10, 1999.

Drucker, P. F. (1993). *Post-capitalist society*. New York: Harper Collins.

Gamble, P. R., & Blackwell, J. (2001). *Knowledge management: A state of the art guide.* Kogan Page.

Gelepithis, P. A. M. (1984). *On the foundations of artificial intelligence and human cognition.* PhD thesis, Department of Cybernetics, Brunel University, England.

Gelepithis, P. A. M. (1989). Knowledge, truth, time, and topological spaces. *Proceedings of the 12th International Congress on Cybernetics*, 247-256.

Gelepithis, P. A. M. (1991). The possibility of machine intelligence and the impossibility of human-machine communication. *Cybernetica*, *XXXIV*(4), 255-268.

Gelepithis, P. A. M. (1997). A rudimentary theory of information: Consequences for information science and information systems. *World Futures, 49*, 263-274.

Gelepithis, P. A. M. (2001). *Intelligent systems: Knowledge representation, social and psychological impact of artificial intelligence*. Athens, Greece: Stamoulis Editions.

Gelepithis, P. A. M. (2002, October). An axiomatic approach to the study of mind. *Res-Systemica, 2*. Retrieved from http://www.afscet.asso.fr/resSystemica/

Gelepithis, P. A. M. (2004). Remarks on the foundations of cybernetics and cognitive science. *Kybernetes, 33*(9/10).

Gelepithis, P. A. M. (in press). A new theory of human knowledge and some of its consequences. *AI & Society: The Journal of Human and Machine Intelligence.*

Gupta, A. K., & Govindarajan, V. (2000). Knowledge management's social dimension: Lessons from Nucor Steel. *Sloan Management Review, 42*(1), 71-80.

Hlupic, V. (Ed.). (2003). *Knowledge and business process management.* Hershey, PA: Idea Group Publishing.

Itami, H. (1987). *Mobilizing invisible assets.* Boston, MA: Harvard University Press.

Newing, R. (1999). Connecting people: Both through IT and face-to-face. In *Knowledge Management,* Financial Times Survey, p. II, November 10, 1999.

Nonaka, I. (1991). The knowledge-creating company. In *Harvard Business Review on Knowledge Management* (pp. 21-45). Boston, MA: Harvard Business School Press.

Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation.* Oxford, UK: Oxford University Press.

Quinn, J. B., Baruch, J. J., & Zien, K. A. (1997). *Innovation explosion: Using intellect and software to revolutionize growth strategies.* New York: The Free Press.

Senge, P. M., & Carstedt, G. (2001). Innovating our way to the next industrial revolution. *MIT Sloan Management Review, 42*(2), 24-38.

Skyrme, D. J. (1999). *Knowledge networking: Creating the collaborative enterprise*. Oxford, England: Butterworth Heinemann.

Toffler, A. (1990). *Powershift: Knowledge, wealth and violence at the edge of the 21st century*. New York: Bantam Books.

Vernon, M. (1999). Enhancing links to the customer. In *Knowledge Management,* Financial Times Survey, p. X, November 10, 1999.

Zack, M. H. (2003, Summer). Rethinking the knowledge-based organization. *MIT Sloan Management Review*, 67-71.

KEY TERMS

Artificial Intelligence: The study of the principles of intelligence (scientific objective), and the design and build of intelligent machines like robots (engineering objective).

Cognitive Science: The multidisciplinary study of intelligent systems (natural, artificial, and hybrid). The disciplines that currently comprise cognitive science are (in alphabetical order) anthropology, artificial intelligence, education, linguistics, neuroscience, philosophy, and psychology.

Collective Human Knowledge: Shared human knowledge that has been linguistically represented.

Communication: H1 communicates with H2 on a topic T if, and only if, (a) H1 understands T (symbol: U(H1 T)), (b) H2 understands T (symbol: U(H2 T)), (c) U(H1 T) is presentable to and understood by H2, and (d) U(H2 T) is presentable to and understood by H1.

Epistemology: The study of knowledge. It is a vast area that has been studied for more than 2,500 years by the greatest minds in philosophy and, increasingly, by scientists in disciplines like psychology, neuroscience, and cognitive science.

End Result of Understanding: An entity E has understood something, S, if, and only if, E can present S in terms of a system of its own primitives (i.e., self-explainable notions).

Information: Potentially meaningful (material) patterns. For a summary presentation of the major views on information along with a rudimentary theory of information and some of its consequences (see Gelepithis, 1997).

Physicalism: The thesis that everything in the universe is physical. Applied to the mind, physicalism asserts that (a) all entities in the universe are physical entities and their combinations (ontological physicalism) and (b) all properties of any entity in the universe are physical properties (property physicalism).

Shared Human Knowledge on a Topic T at a Particular Time t: The end result of all communication processes among the members of a community on the topic T by the time t.

Knowledge-Based Support Environment

Karen Neville
National University of Ireland, Cork, Ireland

Philip Powell
University of Bath, UK

INTRODUCTION

Organizations and universities alike depend on the collection of the data pertaining to the purpose (Curtis, 1999) of the domain in which they operate. Internally, each functional part of the organization works with data collected from the different types of systems used (Laudon & Laudon, 2000). Organizations, therefore, use technology to collect and store data (Whitten, Bentley, & Barlow, 1994) to be processed by the rules formulated to produce valuable information (Connelly, Begg & Strachan, 1996) and eventually knowledge. Universities, too, collect data, processes them, and endow them with relevance and importance (Drucker, 1993). Most organizations use knowledge, for example, regarding their target audience, to gain a competitive advantage. Knowledge and knowledge workers are theoretically the "products" produced by universities. However, they face the same dilemma as the majority of firms, that is, too much data and information but not enough knowledge. Information can be described as explicit knowledge, the significance of which is that information has meaning and it is clearly understood. Knowledge is regarded as volumes of relevant information but, importantly, in addition to experience (tacit knowledge) in the form of an expert (Avison & Fitzgerald, 1995). An expert, to be effective, must use extensively both formal (quantitative) and informal (qualitative) information in decision making. Knowledge is regarded as a strategic asset and therefore the creation of which is often an enterprise-wide goal. Alavi and Leidner (1999) argued that the importance of knowledge is based on the hypothesis that the barriers to the transfer and duplication of knowledge award it with enormous strategic importance. Universities, with the technological capability necessary, are developing systems that can collect and manage knowledge. The combination or integration along with the capability to combine an expert's experience in the form of a system is regarded as a strategic tool. Systems capable of combining both explicit and tacit knowledge are referred to as knowledge management systems (KMSs). Research in this area is not very detailed due to the fact that organizations, not universities, have only been implementing the systems in the last few years. These systems are used to acquire and manage knowledge and distribute it among the different functional units as well as with any external collaborating groups. The idea of disseminating knowledge is not a new concept, be it in education or in industry. Like the classroom, the traditional approach, such as paper-based knowledge sharing, and the virtual are used, depending on factors such as the number of students or the type of decisions to be made. An organization creates a knowledge base to reduce the level of experience needed by managers and to improve the effectiveness of their decisions (Peterson & Davie, 1996). Industry invests an enormous amount of capital in the training of its employees and therefore in the creation of so called "experts in the field"; a "true" knowledge base will allow the acquisition of the experience of experts to reduce the loss of investment should the employee leave (Curtis, 1999).

BACKGROUND

Nonaka (1995) define knowledge as "just true belief." Knowledge is regarded, in this information-driven economy (Drucker, 1993), as power or a source of competitive advantage (Barua, 1996; Drucker, 1993; Grant, 1996; Laudon & Laudon, 2000). Powell et al. (2001) and Casey (1995) describe knowledge as a combination of both information and expertise. Knowledge is acquired or created when an individual, with expertise in a field, uses relevant information productively (Hertog & Huizenga, 2000). The training and the experience that academics amass over the years (knowledge) allow them to both teach and collaborate to produce additional knowledge. Therefore, a knowledge base support environment (KBSE) in this case can be described as a dynamic repository of existing learning and processing systems such as discussion forums, virtual libraries, and research to allow academics and students to retrieve knowledge (either tacit or explicit) based on individual profiles. The possibilities of such a system are limited only by constraints imposed by the university in question, such as technological or managerial support (Neville, 2000). Innovative universities could use this implementation for a number of reasons, specifically to keep staff and students abreast of research and emerging technologies in their fields (Khan, 1997). Designing the system requires a thorough investigation into the use of the Web as a medium for delivery (Driscoll, 1998; McCormack & Jones, 1997; Ritchie & Hoffman, 1996). The designer must be aware of the attributes of the

WWW (World Wide Web) and the principles of instructional design to create a meaningful learning environment (Gagne, Briggs, & Wagner, 1988; Driscoll, 1998). The Web-based classroom is viewed, as already stated, as an innovative approach to teaching (Relan & Gillani, 1997). It, like the traditional method, requires careful planning to be both effective and beneficial (Dick & Reiser, 1989). A Web-based classroom must do more than just distribute information, it should include resources such as discussion forums to support collaboration between learners and ultimately it should also support the needs of both the novice and advanced learner (McCormack et al., 1997; Sherry, 1996; Willis, 1995). A KBSE is composed of a number of components that are integral to the success of the environment (Banathy, 1994): (a) a student mentoring system to support both full- and part-time learners, (b) an exam domain to test both practical and cognitive abilities, (c) a virtual library to allow easy access to conference papers and journals, and (d) the knowledge base facilitated by an agent to integrate all of the components and automate the retrieval of information for the end users.

THE SYSTEM

The KBSE can be developed to support a (generic) university environment. As well as providing a support system for both students (postgraduate or undergraduate) and staff, the system produces a large amount of reports for managing and expanding research within the test department. The environment tests students' problem-solving skills with "real-world" simulations and Multiple Choice Questions (MCQs) providing feedback to both lecturers and students. The environment will grow and change as both staff and students collaborate to add and extract material from the system. Duplication of work by staff can be dramatically reduced, freeing staff to concentrate on other tasks. The environment itself can be used by the university in training and in the management and creation of knowledge. The system will enable or automate four of the many components that constitute a university, and therefore support virtual learning and research. The four components of the system are as follows.

STUDENT MENTORING SYSTEM

Web-based mentoring systems (WBMSs) can be described as learning-delivery environments in which the WWW is its medium of delivery (Crossman, 1997; Driscoll, 1998). Due to the increase in student numbers, there is a need for greater student support, which can be provided through the Web. The mentoring component of the system allows students to log in and view lecture and tutorial material. In addition, a discussion forum will enable both the mentors (lecturers and tutors) and students to exchange ideas and add to the environment, eliminating constraints such as time and location, making the knowledge base available to each type of student. The mentoring component of the system will do the following.

- allow lecturers and tutors to update the content segments of the Web site (for example, course homepages or online reading lists) through a Web browser on or off campus
- provide 24-hour online support to students
- facilitate group collaboration, for example, in discussion forums
- allow students to have positive input into courses
- provide students with the ability to add to the environment through discussion forums, link, and papers enable anonymous feedback and questioning, for example, with feedback forms

EXAM DOMAIN

Students attend tutorials and demonstrations for practical subjects, for example, programming languages and computer networking. However, due to security systems in place to protect network resources (for example, workstations and servers), students' access rights are restricted. Therefore, written exams are used in universities to test practical skills when industry itself tests the student's practical ability rather than the student's ability to memorise material. A domain (server) with user accounts allocated to test material will enable lecturers to fully evaluate the skills gained through practical work. The exam domain will do the following.

- give students and lecturers the opportunity to both test and evaluate skills in a simulated environment where they can assume roles such as a network or database administrator without risk to departmental resources
- house written exam material for IT and other courses
- reduce the duplication of course materials (MCQs and research) on the part of teaching staff
- enable students to assess their understanding of course material and prepare for summer assessment
- supply lecturers and instructors with case examples to expand students understanding of a particular topic

VIRTUAL LIBRARY

Every college department and individual lecturers archive material in the form of journals and books related to specific topics. However, few use systems to track departmental and individual repositories to facilitate research and budgeting. This component will do the following.

- track research material within departments
- automate the lending of research material to staff and students
- enable inventory and budget reporting
- assist lecturers in managing their personal resources (journals, books, and proceedings)
- eliminate duplicate purchases of material
- reduce the workload of administrative and academic staff

THE KNOWLEDGE BASE

The knowledge base (KB; see Figure 1) is the integrating component of the environment, which acts both as a repository and as the channel enabling the ongoing process of knowledge creation. It is intended to extend and compliment the other parts of the system as well as add the ability for users to automate the task of retrieving research and technical papers relevant to their areas of interest. The KB will generate profiles of users relating to their areas of interest. The KB will intermittently build indexes based on the content collected from the different components, and the system will notify the users when items of interest are added based on the different profiles. The KB has the following functionalities.

- searches out users' particular areas of interest on the Web
- automatically provides users with links and updates of online material
- profiles users areas of research from which the agent can learn more about what information the users would find helpful in the pursuit of their specific research interests
- builds a knowledge base of research, as defined by the individual users, to support research acquisition and collaboration

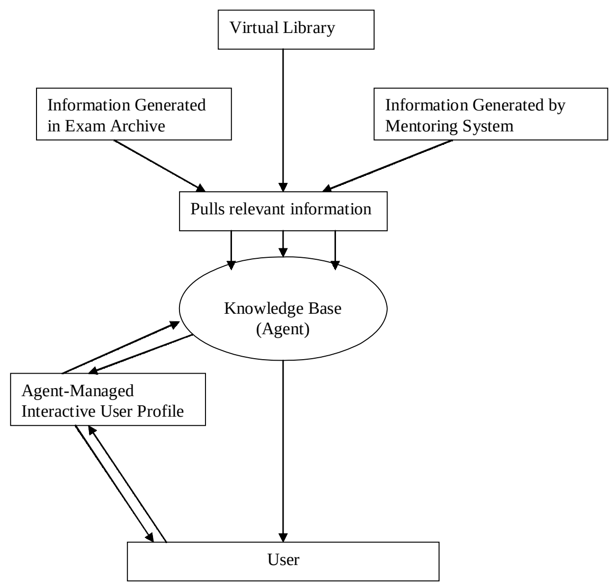
FUTURE TRENDS

Evaluation is the final stage in the development of a KBSE. However, it is an important component of the process. Evaluation is defined as "the value or merit of something" (Driscoll, 1998; McCormack & Jones, 1997). KBSE is a relatively new phenomenon, therefore the evaluation stage of the development life cycle will provide valuable insight into the success and failures of both the development of the system and the value of the system as a knowledge-management tool (Driscoll; McCormack & Jones). The data gathered during this phase will aid in the development of future environments. Evaluation should not be viewed simply as a process of measurement and estimation but a method of learning about the end product, the learner's acceptance of the system, and the effectiveness in creating knowledge. Rowntree (1990) states the following.

1. Evaluation is not an assessment.
2. The unexpected should not be ignored.
3. Evaluation is a planned, systematic, and open endeavour.

Evaluation and assessment are often confused; however, the processes refer to different levels of the investigation into the success of the environment. Evaluation involves all of the factors that influence the acceptance of the system. However, assessment measures the level of understanding of a learner. The developer can perform a number of different evaluations, each of which seeks to elicit specific information. The developer can evaluate the platform used to deliver the system to determine if it provided adequate support. The second type merely involves determining if the system addressed the goals and objectives of the research. Evaluating the impact of the environment on academics and students alike provides the developer or researcher with valuable data regarding its use. Questionnaires are the easiest method used in the evaluation process. The researcher can design a questionnaire to elicit the desired information regarding the system. The questionnaire provides a number of advantages, such as anonymity, which usually provides genuine responses. It enables the researcher

Figure 1. A knowledge base support environment (KBSE)



to collect a large amount of data in a simple format. It is easy to administer to the target group under review, usually by post, and it is relatively easy to analyse. However, the technique cannot be utilised as the only method of evaluation; follow-up interviews are necessary. Semistructured interviews allow the researcher to obtain an elaboration of issues identified as a result of the questionnaires. However, interviews tend to collect large quantities of data that is difficult to analyse. Evaluation provides rich data if designed properly. It is an ongoing process that should be an integral component of the methodology used to develop the system. The use of knowledge-based environments is increasing in importance. Therefore, a suitable method of evaluation is necessary to collect information regarding both the effectiveness and the ineffectiveness of the different components and the system as a whole to continue to add to the value of this knowledge approach. The virtual support environment will initially be offered to a sample of end users (both staff and students) who will manage and use the system for a period of time before the environment is made available to all potential users.

CONCLUSION

Pressures of the information age are forcing organizations and universities to turn to knowledge environments to provide a strategic advantage in this global economy. Effective support environments help create and maintain the knowledge base of a university or organization. There are numerous technologies available to support educational needs. Technologies such as distributed desktop training and computer-based training offer a number of advantages to both universities and organizations. However, these methods can be expensive to maintain as they become out of date relatively quickly. Therefore, more and more universities are turning to KBs to support their knowledge workers. Academics and students alike can share information and access resources over the World Wide Web without restrictions such as geographic location, time, or the platform used. The utilisation of the Web as an educational tool is not an innovative approach to the needs of the learners, however, the use of the Web and software agents as facilitating technologies in the creation of knowledge is.

Educators have discovered that through the use of learning networks and environments, students can interact with their peers to create and improve group work in distance education and it provides the learners with access to online resources and relevant information. Education is built around the use of textbooks, classrooms, and assignments. A KBSE makes the interactive textbook a reality. A network, itself, is defined as a shared space. Telephone and satellite signals form a vast Web or network of computers to enable communication to anywhere in the world. KBs can, therefore, be used by trainers to create a learning environment. Educators and students from any location can share their knowledge and collaborate with each other. Learning in a traditional classroom can be intellectually engaging, as it produces a competitive and collaborative environment. Web-based learning allows "educators and students alike to perform learning related tasks" (McCormack & Jones, 1997). The development of KBSEs requires careful investigation into the requirements of the problem case. The developer must consider factors such as the method employed by the student to learn, incentives to ensure use, the identification of goals and objectives, and the different roles that are needed to support this new approach in training and ease of use. However, the advantages far outweigh the limitations, which through careful planning can be reduced.

REFERENCES

Alavi, M., & Leidner, D. (1999). *Knowledge management systems: Issues, challenges, and benefits, 1*(7). Communications for the Association of Information Systems.

Avison, D. E., & Fitzgerald, G. (1995). *Information systems development: Methodologies, techniques and tools* (2nd ed.). McGraw-Hill Companies.

Banathy, B. (1994). Designing educational systems: Creating our future in a changing world. In C. M. Reigeluth & R. J. Garfinkle (Eds.), *Systemic change in education* (pp. 27-34). Englewood Cliffs, NJ: Educational Technology Publications.

Barua, M. K. (1996). *An empirical study of network and individual performance in distributed design groups.*

Casey, C. (1995). Exploiting expert systems for business. *Executive Business Review*, 2, 16 pages.

Connolly, T., Begg, C., & Strachan, A. (1996). *Database systems: A practical approach to design, implementation and management.* Addison-Wesley.

Crossman, D. (1997). *The evolution of the World Wide Web as an emerging instructional technology tool*. Englewood Cliffs, NJ: Educational Technology Publications.

Curtis, G. (1999). *Business information systems: Analysis, design and practice* (3rd ed.). Addison-Wesley.

Dick, W., & Reiser, R. (1989). *Planning effective instruction*. Englewood Cliffs, NJ: Prentice-Hall.

Driscoll, M. (1998). *Web-based training: Using technology to design adult learning experiences*. Jossey-Bass, Pfeiffer.

Drucker, P. (1993). *Post-capitalist society*. New York: HarperCollins.

Gagne, R. M, Briggs, L. J., & Wagner, W. W. (1988). *Principles of instructional design* (3rd ed.). New York: Holt Reinbank Winston.

Grant, R. M. (1996). The resource-based theory of competitive advantage: Implications for strategy formulation. *California Management Review*.

Hertog, J. F., & Huizenga, E. (2000). The knowledge enterprise. *Implementation of Intelligent Business Strategies*.

Khan, B. (1997). *Web-based instruction.* Englewood Cliffs, NJ: Educational Technology Publications.

Laudon, K. C., & Laudon, J. P. (2000). *Management information systems: Organization and technology in the networked enterprise.*

McCormack, C., & Jones, D. (1997). *Building a Web-based education system.* New York: Wiley.

Neville, K. (2000). A Web-based training (WBT) system development framework: A case study. *Business Information Technology Management (BIT) 2000, 10th Annual Conference*, Manchester, UK.

Nonaka, I. (1995). *The knowledge creating company: How Japanese companies create the dynamics of innovation.* Oxford, England: Oxford University Press.

Peterson, L. L., & Davie, B. S. (1996). *Computer networks: A systems approach.* Morgan Kaufmann Publishers.

Powell, P., Loebbecke, C., & Levy, M. (2001). SMEs, cooperation and knowledge sharing: The IS role. *Global cooperation in the new millennium.* ECIS.

Relan, A., & Gillani, B.B. (1997). Web-based information and the traditional classroom: Similarities and differences. In B.H. Khan (Ed.), *Web-based instruction (p.45).* Englewood Cliffs, NJ: Educational Technology Publications.

Ritchie, D. C., & Hoffman. (1996). Incorporating instructional design principles with the World Wide Web. In C. M. Reigeluth & R. J. Garfinkle (Eds.), *Systemic change in education.* Englewood Cliffs, NJ: Educational Technology Publications.

Rowntree, D. (1990). *Teaching through self instruction.* London: Kogan Page.

Sherry, L. (1996). Raising the prestige of online articles. *Interom, 43*(7), 25-43.

Whitten, J. L., Bentley, L. D., & Barlow, V. M. (1994). *Systems analysis and design methods* (3rd ed.). Burr Ridge, IL: Irwin.

Willis, J. (1995). A recursive, reflective, instructional design model based on constructivist-interpretivist theory. *Educational Technology, 35*(6), 5-23.

KEY TERMS

Agent: Automates a task of, for example, retrieving research and technical papers relevant to end-user areas of interest.

CBT (Computer-Based Training): Is training delivered to employees or students on a computer, providing training on something like word processing or on a programming language such as Java.

DDT (Distributed Desktop Training): Allows the student (employee) to avail of both audio and video applications to enhance training by providing interactive communication between the student and the trainer without being limited by distance.

Explicit Knowledge: Information that has specific meaning and that can be easily and clearly understood.

KBSE (Knowledge-Based Support Environment): A dynamic repository of existing learning and processing systems such as discussion forums, virtual libraries, and research to allow academics and students to retrieve knowledge (either tacit or explicit) based on individual profiles.

Knowledge Web: The use of electronic linkages among different teaching and learning communities to facilitate information acquisition and knowledge building.

Mentoring: A method of teaching that has been used for hundreds of years; this design is incorporated into learning networks to develop more effective learning practices and provide additional support to the learner.

Learning Networks: Computer networks that are used for both educational and training purposes. They consist of groups of people working together, online, to educate both themselves and others. The hardware and software that forms the network system is the only limitation to these groups.

Tacit Knowledge: That which refers to knowledge gained through an individual's own experiences.

Web-Based Classroom: The application of a selection of intellectually stimulating lessons implemented within a creative and collaborative learning environment that utilises the resources of the World Wide Web.

Leader–Facilitated Relationship Building in Virtual Teams

David J. Pauleen
Victoria University of Wellington, New Zealand

INTRODUCTION

How do virtual team leaders assess and respond to boundary crossing issues when building relationships with virtual team members? Virtual teams are a new phenomenon, defined as groups of people working on a common task or project from distributed locations using information and communications technology (ICT). With rapid advances in ICT allowing alternatives to face-to-face communication, virtual teams are playing an increasingly important role in organizations. Due to their global coverage, virtual teams are often assigned critical organizational tasks such as multi-national product launches, negotiating global mergers and acquisitions, and managing strategic alliances (Maznevski & Chudoba, 2000). Their use, however, has outpaced the understanding of their unique dynamics and characteristics (Cramton & Webber, 2000).

Virtual team leadership remains one of the least understood and most poorly supported elements in virtual teams. Virtual team leaders are often the nexus of a virtual team, facilitating communications, establishing team processes, and taking responsibility for task completion (Duarte & Tennant-Snyder, 1999), and doing so across multiple boundaries. Recent research (Kayworth & Leidner, 2001-2002) has begun to look at virtual leadership issues and suggests that the trend toward virtual work groups necessitates further inquiry into the role and nature of virtual team leadership.

This article begins by briefly looking at the key concepts of virtual team leadership, relationship building and boundary crossing. Then, drawing upon the author's research, it examines the complexity inherent in building relationship across boundaries, and concludes with suggestions on how virtual team leaders can mediate this complexity.

BACKGROUND

Virtual Team Leadership

There has been extensive research on leadership in collocated teams and groups. Typically, leadership can be viewed in a number of ways, from a structured authoritative role to the ability of individuals to intrinsically or extrinsically motivate followers. It is generally agreed that leadership involves social influence and the use of communication activities in motivating teams to achieve goals. Barge proposes leadership as mediation in order to overcome the variety of task and relational problems that may be encountered by a group and explains that leadership "entails devising a system of helping the group get its work done, that is simultaneously stable and flexible and assists in managing the information shared among members and between the group and its external audience" (Barge, 1996, p. 319).

A key leadership skill in Barge's concept of leadership as mediation is that of relational management, which refers to the ability of leaders to "coordinate and construct interpersonal relations that allow an appropriate balance of cohesion, unity, and task motivation with a group" (Barge, 1996, p. 325). Cohesive teams tend to perform better and are more motivated to complete tasks. Of concern here is how team leaders can coordinate and construct interpersonal relations in a virtual environment to overcome the difficulty of multiple boundaries that do not exist in traditional collocated teams.

The Importance of Relationship Building in Virtual Teams

The link between team effectiveness and team member relationships is an important but underdeveloped area of study in virtual teams. Usually defined implicitly rather than explicitly, relationships develop over time through a negotiation process between those involved (Catell, 1948). While face-to-face meetings are the preferred way to build relationships and to deal with sensitive and complex situations, it is possible with the skillful and thoughtful application of virtual communication channels to effectively lead a completely virtual team. Research has found that computer-mediated teams do share relational information and are likely to develop relational links over time (Chidambaram, 1966; Warkentin, Sayeed & Hightower, 1997).

The role of the team leader is to move the team towards its objectives by encouraging collaboration. This is done through a sustained process of relationship building, idea generation, prioritisation and selection. The particular challenge to virtual team leaders is to manage this process through ICT. In virtual team research stronger relational links have been associated with higher task performance, more effective information exchange, enhanced creativity and motivation, increased morale, and better decisions (Warkentin & Beranek, 1999; Warkentin et al., 1997). The building of relationships with virtual team members has been shown to be a fundamental concern of virtual team leaders (Pauleen, 2003-04).

Boundary Crossing in Virtual Teams

Boundary crossing is a defining characteristic of virtual teams. Contemporary organizations have highly permeable boundaries allowing substantial communication across boundaries (Manev & Sorenson, 2001). Boundary crossing is an important organizational activity that enhances the flow of information from the external environment. The role and activities of virtual teams leaders make them natural and strategic boundary crossers.

While traditional co-located teams may have members from different functions and cultures, sophisticated new synchronous and asynchronous ICT make it ever easier to form teams consisting of members from different functions, offices, organizations, countries and cultures. Furthermore, virtual teams must function across time and distance, often with team members having never met. These conditions present significant challenges to team leaders and members, team processes and ultimately team outcomes. Because virtual teams are still relatively new, outdated organizational HR and IT policies, which do not support virtual team performance, may be compounding the challenges (Jackson, 1999; Vickery, Clark & Carlson, 1999).

Boundary crossing in virtual teams can affect relationship-building efforts. Maznevski and Chudoba (2000) showed that deliberately addressing relationship building to develop shared views and trust across all types of boundaries could help virtual team performance. The more boundaries between leaders and team members at the start of a virtual team, the more likely higher levels of relationship with team members as well as more intensive relationship-building strategies will be needed.

MAIN THRUST OF THE CHAPTER

The Effects of Boundary Crossing on Relationship Building

The practical effect of working across distance means that teams can and do comprise members from different departments, head and branch offices, and organizations, as well as different countries and cultures. Indeed, access to different organizational, functional and cultural perspectives is a key reason for using virtual teams. These differences represent important conditions that team leaders will probably need to assess and accommodate before commencing a virtual team. According to team leaders, the development of personal relationships between themselves and team members is an important prerequisite in establishing and maintaining virtual working relationships across three conceptual boundary-crossing categories: (1) Organizational Boundary Crossing, (2) Cultural/Language Boundary Crossing and (3) Time/Distance Boundary Crossing (Pauleen, 2003-04). While organizational and cultural/language barriers exist in co-located teams, they are more likely to be found in virtual teams and to a have a more significant impact. Time and distance boundaries are unique to virtual teams.

Organizational Boundary Crossing

Organizational boundary crossing includes intra- and inter-organizational boundaries. Different functions, departments, and organizations may have diverse work cultures as manifested by deeply held core beliefs and assumptions (Kayworth & Leidner, 2000). Wiesenfeld, Raghuram and Garud (1998) suggested that organizational identification would be the psychological tie that binds virtual workers together into an organization, preventing workers from thinking of themselves as independent contractors, operating autonomously.

A strong organizational culture might influence the level of relationship building necessary in a team composed of members from within the same organization, even if they are located in different countries. Strong organizational cultures are exemplified by institution-based trust relationships (Nandhakumar, 1999; van der Smagt, 2000) and an anticipation of future association (Pauleen, 2003-04). The degree of relationship building necessary and the strategy for going about it are likely to be quite different when a team starts with a strong intra-organizational culture. Conversely, virtual teams with members from different organizations will need to be aware of and navigate the different organizational cultures.

Another aspect of organizational boundary crossing is the particular preferences of certain organizations for certain technologies, for example, communication channels such as e-mail or voice mail when leaving messages. Team leaders may experience difficulties trying to agree on common communication platforms with team members outside of the organization.

Cultural/Language Boundary Crossing

Cultural/language boundary crossing is another critical area. Cultural/language boundary crossing will most likely take place in global virtual teams, though it may also be a factor in national or even local virtual teams (Pauleen, 2003-04). The key point is whether a team leader is working with a team member from another nationality or ethnic culture. The effects of culture in team settings can be profound, and include, among other important issues, how individuals relate to each other (Kayworth & Leidner, 2000). Misinterpretations or distortions may occur as team members and team leaders interpret communications through their own cultural programming (Lewis, 1996), a challenge that is greatly complicated when attempted through ICT. In all cases, team leaders need to assess the impact of cultural differences.

In some cases, there may be a strong cultural preference for the use of face-to-face communication to build relationships. Only after a certain level of comfort is attained can ICT be used with any effectiveness. This strategy supports Hall's (1976) theory of high and low context cultures, which states that for some cultures communication is more about context then the actual verbal message. In high-context cultures, messages have little meaning without an understanding of the surrounding context, which may include the backgrounds of the people involved, previous decisions, and the history of the relationship. People from low-context cultures prefer more objective and fact-based information. The message itself is sufficient.

Team leaders will need to consider the degree of personal relationship necessary to get the working relationship underway, as well as the use of appropriate communication channels along with appropriate messages delivered in an appropriate manner (Pauleen, 2003-04).

Time and Distance Boundary Crossing

Time and distance boundaries most obviously distinguish virtual teams from co-located teams. The effect of distance on relationship building strategies is proportional to how far the team leader and team members are from each other. The further away, the more difficult the use of face-to-face communication, which could be problematic in situations where face-to-face communication is the best or maybe only option.

The effect of time on relationship building strategies concerns the challenge of working across time zones. This may have little impact on the degree of relationship building that may be necessary, but a large effect on creating strategies to build relationships, as the time differences can restrict the kinds of ICT, particularly synchronous channels such as telephone and videoconferencing, available to the team leader. Probably not an issue that will make or break a virtual team of professionals, it is one of the conditions that must be carefully and fairly assessed by the team leader before creating relationship-building strategies.

If asynchronous communication channels are used, such as e-mail, the problem of pacing communication exchanges can become a serious consideration. Response times between team leaders and team members may differ, constraining communications, causing uncertainty and negatively impacting trust (Jarvenpaa & Leidner, 1999;Warkentin & Beranek, 1999). Time lags due to technical infrastructure and technological breakdowns, if not understood by the people involved, can cause the team leader or team member to attribute non-communication to lack of manners or conscientiousness, which can then seriously affect relationships (Cramton & Webber, 2000).

Problems associated with crossing time and distance have the potential to greatly disrupt relationship building in a virtual team, particularly with inexperienced team members. It is necessary for the team leader to carefully assess these potential obstacles before creating relationship-building strategies, as well as anticipate the problems that may be caused by time and distance.

Challenges and Solutions for Team Leaders and Organizations

This discussion points to significant differences in co-located and virtual teams in leadership-led relationship building across boundaries. The greater number and variety of boundaries and their deeper impact pose special challenges that virtual team leaders will need to mediate. Table 1 summarizes the key communication challenges faced by virtual team leaders. Two of these challenges are discussed. First the need for virtual team leaders to expand and hone their repertoire of skills in handling virtual communication channels and second, what organizations can do to improve their teams' virtual communication skills.

The strategic use of communication channels is one critical skill that virtual team leaders will need to hone. Table 2 illustrates ways in which virtual teams leaders need to mediate cultural/language boundaries by consciously selecting the most appropriate communication channels. Higher context cultures will tend to require media rich channels such as phone and video conferencing to build relationships, at least until the development of a sufficient level of trust. In contrast, lower context cultures, which are more task-oriented, will tend to be more tolerant of a wider range of communication channels. Indeed, being task-oriented, building relationships might be secondary to getting started on the task. Knowing the

Table 1. Mediating complexity - Communication challenges when building relationships with virtual team members across boundaries (Pauleen & Rajasingham, 2004)

	Types of Boundaries Crossed		
	Organizational	**Cultural/Language**	**Time & Space**
Co-located Teams	Shared organizational culture supports relationship building, although functional culture can pose challenges.	Nonverbal cues can be understood by experienced leaders. Non-native speakers	The ability to regularly meet face-to-face supports relationship building across cultures of all types.
Virtual Teams	Differing organizational policies inhibit communication, as do organizational preferences for the use of different communication channels and ICT infrastructure. Lack of situational knowledge of team members can cause misunderstandings and mis-attribution, leading to potential difficulties.	Lack of nonverbal cues is difficult to overcome with most available communication channels. Cultural preferences for certain communication channels, often face-to-face (see Table 2 for more detailed analysis of the use of channels across cultures)	Building relationships across time and space requires concerted efforts and more time than in co-located contexts. Arranging synchronous meetings or even phone calls across time zones is often problematic. Asynchronous channels face the problem of pacing communications. Dealing with "silence" is a particularly difficult challenge.

Table 2. Mediating complexity - Guidelines for using communication channels to build relationships in virtual teams across cultural/language barriers (Pauleen & Rajasingham, 2004)

Preferred Communication Channels			
		Native Speakers	Non-native Speakers
High Context Cultures (relationship-oriented) (tend toward formality)	Media rich, Synchronous, Face-to-face, phone, video/audio	Media rich, Synchronous, asynchronous	Face-to-face with interpreter, Asynchronous written with editor
Low Context Cultures (task-oriented) (Tend toward informality with notable exceptions, e.g., Germans)	Flexible, as above, plus e-mail, fax, computer conferencing (online synchronous written chat, asynchronous discussion boards)	All channels – synchronous and asynchronous	All channels with translator, interpreter or editor as required

cultural composition of the team and team members' prior experiences working virtually and across boundaries in general are part of the complex web of factors that leaders need to determine and then mediate if they are to successfully build relationships with team members and ultimately complete team tasks.

The second challenge concerns organizational support structures to improve both virtual team leader and member skills. Organizations need to be willing to provide training and support to develop effective boundary crossing behaviors among virtual team members. Virtual team processes and dynamics are different from those of co-located teams and require special skills focusing on networking and establishing links across boundaries (Duarte & Tennant-Snyder, 1999). Team members, and particularly team leaders, will often need to play multiple roles as negotiators (with customers), network and coalition builders (with other teams), lobbyists (with top management), and motivators (of team members). To be effective, team leaders will need training in boundary crossing, networking and relationship building skills (Yan & Louis, 1999).

FUTURE TRENDS

There has been a pressing need for rigorous conceptual and empirical work to examine factors that influence virtual teams (Pare & Dube, 1999), and it is only in the most recent literature that there have been systematic attempts to look at how virtual team leadership can support virtual team success. This article has briefly looked at issues related to team leader-facilitated relationship building across boundaries. Its results suggest directions for future research and practice, including the development of virtual leadership mediation and communication skills.

Realizing effective leadership in group, team and multicultural interorganizational communications across diverse perspectives and global virtual environments presents new, real and compelling challenges to team leaders, but these challenges also present unparalleled opportunities for teams to explore new perspectives, approaches and ideas (Adler, 2002). Understanding these challenges and developing effective leadership and organizational processes to meet them presents opportunities for both practitioners and researchers.

CONCLUSION

Crossing organizational, cultural and time and distance boundaries requires training, experience and organizational support. These can help team leaders determine how to work across boundaries that may be present in their team. In addition to its effects on building relationships, boundary-crossing differences can affect team processes and performance in many ways. Ignoring them is an invitation to team failure. Leaders must learn to mediate the increased complexity inherent in virtual teams. A starting point for leaders is to approach the complexity introduced by boundary crossing by asking these two questions:

1) What are the boundary crossing influences of this situation?
2) How can they be understood and worked with so that a good people-oriented environment of assurance and trust can be maintained and productivity enhanced?

Perhaps the ultimate challenge for team leaders, particularly in long-term or on-going virtual teams, is to work to merge the individual cultures – functional, organizational, national, and so forth - of the team members into a team culture.

REFERENCES

Adler, N. (2002). *International dimensions of organizational behavior*. Cincinnati: South-Western.

Barge, J.K. (1996). Leadership skills and the dialectics of leadership in group decision making. In R.Y. Hirokawa & M.S. Poole (Eds.), *Communication and group decision making* (pp. 301-342). Thousand Oaks, CA: SAGE.

Catell, R. (1948). Concepts and methods in the measurement of group syntality. *Psychological Review, 55*, 48-63.

Chidambaram, L. (1996). Relational development in computer supported groups. *Management Information Systems Quarterly, 20*(2), 142-165.

Cramton, C., & Webber, S. (2000). Attribution in distributed work groups. In P. Hinds & S. Kiesler (Eds.), *Distributed work: New research on working across distance using technology* (pp. 191-212). Cambridge, MA: MIT Press.

Duarte, N., & Tennant Snyder, N. (1999). *Mastering virtual teams: Strategies, tools, and techniques that succeed.* San Francisco: Jossey-Bass Publishers.

Hall, E.T. (1976). *Beyond culture.* New York: Doubleday.

Jackson, P.J. (1999). Organizational change and virtual teams: Strategic and operational integration. *Information Systems Journal, 9*(4), 313-332.

Jarvenpaa, S.L., & Leidner, D.E. (1999). Communication and trust in global virtual teams. *Organizational Science, 10*(6), 791-815.

Kayworth, T., & Leidner, D. (2000). The global virtual manager: A prescription for success. *European Management Journal, 18*(2), 183-194.

Kayworth, T., & Leidner, D. (2001-2002). Leadership effectiveness in global virtual teams. *Journal of Management Information Systems, 18*(3), 7-40.

Lewis, R.D. (1996). *When cultures collide: Managing successfully across cultures.* London: Nicholas Brealey Publishing.

Manev, I.M., & Sorenson, W.B. (2001). Balancing ties: Boundary spanning and influence in the organization's extended network of communication. *The Journal of Business Communication, 38*(2), 183-205.

Maznevski, M.L., & Chudoba, K.M. (2000). Bridging space over time: Global virtual team dynamics and effectiveness. *Organization Science, 11*(5), 473-492.

Nandhakumar, J. (1999). Virtual teams and lost proximity: Consequences on trust relationships. In P. Jackson (Ed.), *Virtual working: Social and organizational dynamics* (pp. 46-56). London: Routledge.

Pare, G., & Dube, L. (1999). Virtual teams: An exploratory study of key challenges and strategies. *Proceedings of the Twentieth International Conference on Information Systems,* Charlotte, NC (pp. 479-483).

Pauleen, D. (2003-2004). An inductively derived model of leader-initiated relationship building with virtual team members. *Journal of Management Information Systems, 20*(3), 227-256.

Pauleen, D., & Rajasingham, L. (2004). Mediating complexity: Facilitating relationship building in start-up virtual teams. In D. Pauleen (Ed.), *Virtual teams: Projects, protocols and processes* (pp. 255-279). Hershey, PA: Idea Group Publishing.

Van der Smagt, T. (2000). Enhancing virtual teams: Social relations v. communication technology. *Industrial Management and Data Systems, 100*(4), 148-156.

Vickery, C.M., Clark, T.D., & Carlson, J.R. (1999). Virtual positions: An examination of structure and performance in ad hoc workgroups. *Information Systems Journal, 9*(4), 291-312.

Warkentin, M., & Beranek, P.M. (1999). Training to improve virtual team communication. *Information Systems Journal, 9*(4), 271-289.

Warkentin, M.E., Sayeed, L., & Hightower, R. (1997). Virtual teams versus face-to-face teams: An exploratory study of a Web-based conference system. *Decision Sciences, 28*(4), 975-996.

Wiesenfeld, B.M., Raghuram, S., & Garud, R. (1998). Communication patterns as determinants of organizational identification in a virtual organization. *Journal of Computer Mediated Communication, 3*(4).

Yan, A., & Louis, M.R. (1999). The migration of organizational functions to the work unit level: Buffering, spanning, and bringing up boundaries. *Human Relations, 52*(1), 25-47.

KEY TERMS

Asynchronous Communication Channels: Communication channels that support communication that usually requires a period of time to pass between communicative transactions. These channels include e-mail, discussion boards, fax, and so forth.

Boundary Crossing: Virtual teams are often characterized by their boundary spanning attributes; that is, they usually cross time and distance, and often include different national (ethnic), organizational and functional cultures.

Personal Relationships: The kind of relationship between people exemplified by shared understanding, mutual trust and social bonding. Communication in personal relationships is initially directed toward the exchange of personal information and later toward the sharing of mutual experiences.

Synchronous Communication Channels: Communication channels that allow real-time interaction. These include telephone, video conferencing, chat, and of course, face-to-face communication.

Virtual Team: A given number of people at distributed locations communicating and working to some degree via information and communication technologies on a set project or task, which may be of a limited or unlimited duration. Face-to-face meetings at the start-up of the team or at regular intervals are possible in a virtual team.

Virtual Team Leader: This is the person who functions as the hub of the team, holding it together. In the literature, this person may be termed a team facilitator, (virtual) project manager, coordinator or coach depending on the nuances of the role, the perspective of the researcher and organizational terminology. The team leader responsibilities may include all or some of the following: selecting team members; setting team tasks and team member roles; ensuring project or task completion; liaising with stakeholders and clients; establishing communication and team protocols, facilitating interpersonal and team communication, handling conflict, and managing technology and in general ensuring effective participation of all the team members.

Working Relationships: The kind of relationship exemplified by people who work together toward the completion of work-based tasks. It involves communication related to sharing information, coordinating tasks, meeting timelines, and so forth.

Leapfrogging an IT Sector

Eileen M. Trauth
The Pennsylvania State University, USA

INTRODUCTION

Accompanying the global spread of the post-industrial society (Bell, 1973) are nations who see economic opportunity deriving from the development of an information economy to support it (Porat, 1977). But while advanced industrialized nations moved gradually from industrial to post-industrial work over a period of decades, newly industrializing countries are "leapfrogging" directly from agrarian to information-intensive work in a matter of years. Given this rapid labor force transformation, a critical consideration in the development of a global information sector is the development and management of information technology (IT) workers.

Ireland is an appropriate country for examination of this leapfrog phenomenon because it was one of the earliest examples of this phenomenon, having developed its information sector rapidly and successfully through inward investment by multinational firms during the 1970s to the 1990s. Thus, this case offers the point of view of both an advanced industrialized or "first wave" country and of a "second wave" country that is taking an alternate path into the information economy by rapidly moving directly from an agrarian or partially-developed industrial economy into an information economy. Since Ireland was one of the earliest examples of "leapfrogging", the Irish case has lessons applicable to other contexts (Trauth, 2000).

BACKGROUND

Ireland's rapid transformation from a poor, agrarian society to a robust information society fueled by its information economy was the result of policy initiatives, cultural compatibility with IT work and adaptive responses to opportunities and crises. Ireland's policy of economic development through inward investment was a direct reversal of the preceding policy of cultural and political sovereignty achieved largely through economic isolationism. But a combination of high emigration and high unemployment signaled the need for change (Trauth, 2001).

The multinational firms brought direct benefits through the jobs that kept people in Ireland and away from unemployment, and indirect benefit through the foreign investment that would provide both jobs and a new business climate. These outside influences were expected to help Ireland more quickly develop an indigenous entrepreneurial capacity. The long-term benefits would be the spillover effects from the development of technical and business expertise.

Ireland provided attractive economic incentives in the forms of tax relief and grants for equipping their factories, and training the work forces. These were the necessary conditions for establishing the multinational IT sector in Ireland. But the sufficient conditions were a societal infrastructure supportive of IT work and a qualified labor force to do it. Today, Ireland's software industry has emerged as a strong contender for multinational sites, along with Israel, India (Heeks, 1996) and Eastern Europe (Heavin, Fitzgerald, & Trauth, 2003). Ireland's software sector employs 30,000 people in both indigenous and multinational operations and creates revenues in excess of Euro 10 billion (Flood et al., 2002).

The Irish case offers two important sets of human resource issues. The first set relates to ensuring a supply of appropriately qualified IT workers. The second set relates to managing IT workers in a cross-cultural environment. To the extent that Ireland's experiences are typical of other second wave countries, the lessons learned apply to indigenous and multinational managers as well as government policy makers in other counties.

ENSURING A SUPPLY OF QUALIFIED IT WORKERS

(Re)Designing Societal Structures to Support IT Work

Among the societal infrastructures that were adapted to support the emerging IT sector, the most important was the educational infrastructure (Clancy, 1988). Irish policy makers recognized that the well-educated Irish population was a powerful resource that could be leveraged to support the emerging information sector. But there were two serious issues to overcome. The first was establishing equality of access to education. This was accomplished in 1968 when secondary education became state-funded. The other issue was enabling potential IT workers to acquire the specialized skills and knowledge for work in this sector. In the 1960's, the traditional university was not oriented toward vocational education much less vocational education of a technical nature. Consequently, in the 1970's and 1980's, two new universities were established and the existing universities were adapted to incorporate business and IT skills into their curricula. Technical colleges were also established. Evening, adult-oriented programs were established for workers to develop their skills and employment prospects. Finally, the government-sponsored IT training programs for those with university degrees or who had been made redundant in other fields (Trauth, 1993).

It was also necessary to maintain alignment between the particular skill sets being developed in the schools and training programs, and the available types of jobs. An unintended consequence of Ireland's educational success was that Irish IT workers became a desired human resource in other countries. In response, industrial policy assessments recommended a closer match between the educational plans of the universities and the employment opportunities available in the country (Industrial Policy Review Group, 1992).

Addressing Barriers to a Wider Participation in IT work

There were barriers to full participation in Ireland's IT sector with respect to age, gender and social class. The perception of IT as a young person's field was reinforced by the extremely young population of the country. This age divide was exacerbated by the prior educational policy limiting access to secondary and, therefore, higher education. Thus, in the early 1990's, those over the age of thirty-five, those who had not had access to free secondary education, were fewer in number in the IT labor force.

Another type of barrier relates to gender. While women found the IT sector better for women than traditional industries, banking and the civil service, there was still a tension between opportunity and restriction. IT was a new industry without established patterns of gendered work. However, there was also an acknowledged stereotype of IT work as a male activity and recognition by both men and women that women were not full participants in the IT sector. The reasons have been typically linked to a culture of large families in which child rearing was a woman's responsibility (Kvasny & Trauth, 2002; Trauth, 1995).

The final type of barrier relates to social class. Despite Ireland's historic disdain for rigid social class categories, there is evidence of social class barriers in the information sector. The absence of free secondary education was a barrier to poor and working class individuals. Further, in family settings without a history of or value placed upon education, there can be pressure on young people to enter the work force as soon as possible in order to add to the family income. Other evidence of attitudinal barriers, coming from members of the middle class, was the importance of having the "correct" accent and address in order to secure employment in indigenous IT firms.

Managing IT Workers in a Cross-Cultural Work Environment

The second set of issues relates to managing IT workers in a cross-cultural work environment in which the "first wave" nation's culture is embedded in the corporate culture of the multinational firm (Trauth, 1996). A firm's corporate culture " its values, management style, method of operations and work environment" reflects the national culture in which it developed. Thus, the multinational IT workplace was a cross-cultural mix of American and Irish cultures, the IT culture, and the particular corporate culture of the firm. While the Irish workers welcomed the American management style and corporate culture, there was also tension over how far – and in what direction – the cultural influence should go. Not surprisingly, the American managers favored the American culture, believing that the multinational firms ought to have a significant cultural impact. On the other hand, Irish human resource managers argued for tailoring the corporate culture to the particular national context. The viewpoint was that while the multinationals were bringing certain values and attitudes to the workplace, there was also a significant contribution to be made by the Irish culture. But along with resistance to the multinational influence was the recognition that importing another work culture was part of the plan. By bringing in multinational firms Ireland would be able to import a well-established work ethic that would have taken considerably longer to develop if done indigenously.

FUTURE TRENDS

Managing IT workers in a cross-cultural environment requires the acknowledgment that two different national cultures are involved when a multinational IT firm sets up operations in a country. Both of these cultures have positive contributions to offer the workplace. Multinational managers should strive to understand work patterns and attitudes of the host country culture that, while different, may nevertheless be productive. They should also understand that introducing a corporate culture means introducing a different national culture. At the same time, the host country must recognize that when the intent is to import expertise in order to quickly introduce an IT sector, one side effect will be the changes in national culture.

This has been referred to in the literature as *situating culture* (Weisinger & Salipante, 2000; Weisinger & Trauth, 2002, 2003). According to this theoretical framework, culture is a socially negotiated, dynamic, practical, and locally situated process. Hence, it is a view of culture as "doing" (which places emphasis on the actual behaviors of people) rather than "thinking" (which places emphasis on shared cognitive schemas). This view sees the interaction among the group members' different cultures as being situated in a particular context. Thus, by situating culture in a particular context, a manager may be better able to comprehend the emergence of unique local cultural processes that reflect distinct socially negotiated realities and workplace practices.

By recognizing that there is more than one way to achieve a management goal, multinational managers can develop procedures and management approaches that exploit the best features of the host country. In the presence of two distinct cultures, it is also necessary to acknowledge that cultural influence goes in both directions. It is natural that the home culture of the multinational IT firm will influence the society it enters. But it is also natural that the workers will influence the corporate culture to make it compatible with their own. By building a permeable wall, by the open exchange of values and norms, both cultures can be enriched.

For example, American managers in Ireland learned that a human relations problem could be diffused by utilizing an Irish cultural institution: the pub. By meeting in a setting that, according to cultural norms, conveys equal standing, a manager could more easily work out the problem with the employee. Clearly, this management approach fits well with Irish culture but may not suit another cultural context. Likewise, what works best in Japan or in America may not be what works best in Ireland. Another example is the approach taken to knowledge management. Knowledge management refers to the capturing and recording of the tacit organizational knowledge that has built up over time and that resides within the employees of a firm. One of the main objectives of engaging in knowledge management is to improve the flow of organizational knowledge to facilitate organizational learning of new employees and cross-training. In Ireland, managers can address some aspects of this goal by leveraging certain cultural characteristics instead of investing in expensive equipment. The cultural characteristic of interest in others and sociability leads naturally to employees knowing more about their colleague's work than might be the case in another culture that places less emphasis on sociability (Trauth, 2000). Hence, cross-training is a natural outcome of this interaction. Simply, by recognizing that this is occurring, management can reap knowledge transfer benefits.

CONCLUSION

The case of Ireland represents an early example of the issues associated with what is currently recognized as a trend toward global outsourcing. Because it was a leader in positioning itself as a destination for global outsourcing, Ireland's experiences offer insights for both nations and multinational companies interested in outsourcing. One insight is that a qualified labor force must be available. Because of the specialized skills required for employment in the information sector, education is the key to attracting outsourcing. But in order to ensure that all citizens are able to participate in this employment sector, there must be equal educational opportunity. In Ireland, and elsewhere, members of certain groups experience barriers that inhibit full and equal participation in the information economy. In order to ensure that the society can sustain its IT sector, the educational infrastructure needs to be aligned with the skills and knowledge required for IT work, and the educational plans of the universities should be coordinated with the employment objectives of the firms. Finally, management consideration needs to be given to the mix of cultures that is present when a multinational firm from one country sets up operations in another. In this new cultural context, both cultures come together to produce a new, *situated* culture. The public policy, human resource and management issues that are raised in this case reveal some of the issues that must be addressed as more and more countries "leapfrog" an IT sector in order to reap the benefits of global outsourcing. These issues must be considered from two perspectives: that of the firms engaged in outsourcing, and that of countries inviting firms to its shores.

REFERENCES

Bell, D. (1973). *The coming of post-industrial society: A venture in social forecasting.* New York: Basic Books.

Clancy, P. (1988). *Who goes to college: A second national survey of participation in higher education*. Dublin: Higher Education Authority.

Flood, P., Heffernan, M., Farrell, J., MacCurtin, S., O'Hara, T., O'Regan, P., & Carroll, C. (2002). *Managing knowledge-based organizations: Top management teams and innovation in the indigenous software industry*. Dublin: Blackhall Publishing.

Heavin, C., Fitzgerald, B., & Trauth, E.M. (2003). Factors influencing Ireland's software industry: Lessons for economic development through IT. In M. Korpela & R. Montealeagre (Eds.), *Information systems perspectives* *and challenges in the context of globalization.* Boston: Kluwer Academic Publishers (pp.235-252).

Heeks, R. (1996). *India's software industry*. New Delhi: Sage Publications.

Industrial Policy Review Group. (1992). *A time for change: Industrial policy for the 1990s.* Dublin: Ministry for Industry and Commerce.

Kvasny, L., & Trauth, E.M. (2002). The digital divide at work and home: Discourses about power and underrepresented groups in the information society. In E. Wynn, M.D. Myers, & E.A. Whitley (Eds.), *Global and organizational discourse about information technology*. Boston: Kluwer Academic Publishers (pp.273-291).

Porat, M. (1977). *Information economy: Definition and measurement.* Washington, D.C.: Office of Telecommunications.

Trauth, E.M. (1993). Educating IT professionals for work in Ireland: An emerging post-industrial country. In M. Khosrowpour & K. Loch (Eds.), *Global information technology education: Issues and trends*. Harrisburg, PA: Idea Group Publishing (pp.205-233).

Trauth, E.M. (1995). Women in Ireland's information economy: Voices from inside. *Eire Ireland, 30*(3), 133-150.

Trauth, E.M. (1996). Impact of an imported IT sector: Lessons from Ireland. In E.M. Roche & M.J. Blaine (Eds.), *Information technology development and policy: Theoretical perspectives and practical challenges*. Aldershot, UK: Avebury Publishing Ltd (pp.245-261).

Trauth, E.M. (2000). *The culture of an information economy: Influences and impacts in the Republic of Ireland*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Trauth, E.M. (2001). Mapping information-sector work to the workforce: The lessons from Ireland, *Communications of the ACM, Special Issue on The Global IT Workforce, 44*(7), 74-75.

Weisinger, J.Y., & Salipante, P. (2000). Cultural knowing as practicing: Extending our conceptions of culture. *Journal of Management Inquiry, 9*(4), 376-390.

Weisinger, J.Y., & Trauth, E.M. (2002). Situating culture in the global information sector. *Information Technology and People*, *15*(4), 306-320.

Weisinger, J.Y., & Trauth, E.M. (2003). The importance of situating culture in cross-cultural IT management. *IEEE Transactions on Engineering Management, Special Issue on Cross-cultural IT Management, 50*(1), 26-30.

KEY TERMS

Cross Cultural IT Management: Managing the IT function and its personnel in a globally distributed setting.

Global Outsourcing: The trend towards directing outsourcing—contracting with other firms to perform non-critical functions for a business—toward countries with low workforce costs.

Information Economy: That portion of the national economy that is based upon information processing and related activities (Porat, 1977).

Information Sector: A component of the information economy. The *primary information sector* includes those who develop hardware, software and information systems. The *secondary information sector* includes those engaged in information processing activities in the course of doing work related to some other primary activity such as insurance claims processing.

Information Society: A societal transformation in which information is the key resource.

Information Technology (IT) Worker: One who works in the primary or secondary information sector.

Post-industrial Society: A society in which knowledge replaces capital as the key economic resource and the predominant type of work is in the service sector (Bell, 1973).

Learnability

Philip Duchastel
Information Design Atelier, Canada

INTRODUCTION

Learnability is not exactly a new concept in information technology, nor in cognitive science. Learnability has been a key concept of usability (Folmer & Bosch, 2004) in the area of software system design, where it relates to such issues as consistency, familiarity and simplicity. It has also been a traditional concept in linguistics in relation to the ease of language learning (McCarthy, 2001) and in machine learning (Valiant, 2000).

The concept of learnability has recently been repurposed within the field of instructional technology (Duchastel, 2003), building on the concept of usability in Web site design (Nielsen, 2000), and it is that learnability that is considered here. Learnability in this new sense concerns how learnable some piece of instruction is. It deals with a facet of educational resources.

The basic question is this: What makes the content of an instructional site (or of some resource) learnable? Take any one of the many thousands of online learning courses currently available on the Web and ask yourself: Does this course seem difficult to learn (assuming you have the proper background for it)? What would improve it? What would the ideal online course in this area look like? These questions all underlie the learnability of the course.

What then is learnability? Could we say that it is defined by successful learning? That would mean that students who study the course thoroughly learn its content, as evidenced on a good test for instance. Or could we say that a main criterion is ease of learning? Meaning that students experience good intellectual flow and enjoy the course.

Both of these factors, success in learning and enjoyment of learning, can be considered criteria of learnability. Are there others? That is the issue of learnability.

The skeptic will immediately insist that learning takes place within a learner and that it is that locus that mainly determines learnability – that is, the curiosity, intelligence, motivation and persistence of the learner. These are what make or break learning. The teaching materials can only go so far, the learner has to make a go of it, make it succeed.

While there is some truth to that view, it is certainly not the full picture, nor the most useful picture. Consider traditional usability in Web sites or software products. There too, the user plays a role. If he is dull-witted, or perhaps too pressed for time (showing a lack of interest), or just resistant to learning the basics (jumping in and thrashing around – as often happens), there is little scope for success no matter how usable the site or program may have been made. But we do not give up on usability in Web site creation because of that.

The point is designers do not blame the user for incompetence, for ill-will or for the lack of success of their site or program. They maximize usability, realizing well enough that usability is certainly contextual. The same applies, as it should, to learnability: success in learning can be maximized through the product, over and beyond context issues, or in spite of them.

The product view of instruction is an important one, one that is emphasized here. An alternate view, much more widespread, is a process one: learning is a process, and so is instruction in the sense of manipulating the situation so as to facilitate learning. This is why the immense amount of research on learning and education over the past century has not dealt explicitly with learnability.

The process view is not to be denigrated, but a product view can incorporate processes and has definite design advantages. Learnability is best considered in this light.

LEADING QUESTIONS

The challenge before us is to identify those features of excellent learning materials. What makes something learnable? Very learnable, most learnable?

But first, why is it so difficult to pinpoint these features? What are the deep issues underlying learnability? There are three of them we need to consider. They are learning, design, and curriculum. Each is difficult in its own right and learnability involves considering them jointly – hence the magnitude of the challenge.

The first deep question is what is learning? The field of learning has long been a core issue in psychology and numerous theories of learning have been put forth in answer (Kearsley, 2004). The issue is far from settled, as practitioners such as educators well know. There is acknowledgment of different kinds of learning, with different factors at play, but no large agreement on these or on the overall picture.

The second deep question concerns teaching. How do you design for learning? There are general principles that have evolved over time, codified broadly in what is known as the field of instructional design (Reigeluth, 1999). But here too, there is hardly agreement. All design theorists will subscribe to general systems principles like those found in software design or in HCI. All subscribe to the value of usability testing, the trying out of the materials designed with sample students in order to verify the strength of the design and capture any ways of improvement. But given divergences in views of learning, it is natural that hard disagreements will occur here too, in how to design for learning.

The third deep question concerns what to teach - the content. That was what led educators to determine and discuss taxonomies of learning objectives half a century ago (Bloom, 1956) and why this issue remains at the heart of much debate in education (Egan, 1997).

At first thought, you might think that this is an outside issue. That first, we decide what to teach, then only after that, how to teach it, how to design it. Or we might think that teachers and curriculum specialists, or professors and institutions, determine the content "to be covered". That learnability applies to any content, whatever it is determined it should be. But that overlooks the crucial notion that the *what* and the *how* of learning are inextricably linked (Carroll, 1990), just as in communication more generally. An instructional designer must fashion the content as much as the process, in the same way an information designer fashions information well beyond the graphic design aspect. Both are information architects, but that is not yet widely recognized, which creates difficulties for the acceptance of learnability.

In the next sections, I will address these leading questions by introducing some simple models that synthesize them in a nutshell. This remains a very cursory look at the issues, but nevertheless shows the direction in which they can be further explored, as is done in Duchastel (2003).

LEARNING - THE CIM MODEL

At its most general, learning is the process of internalizing information in memory, making that information available later on when needed. But learning the names of the bones in the body and learning the principles of acoustics are rather different forms of learning. We learn them in different ways. What are the commonalties? What are the differences?

There are three types of learning, conveniently contrasted in what we can call the CIM model. CIM stands for Comprehension, Interest, Memorizing, these being the three factors involved in the learning process.

Comprehension is based on our ability to reason, to fit things together, to see how they all work together. Comprehension is the process of generating internal models of the world in all its workings, large and small. We comprehend when we see how things fit together, how it all makes sense. Understanding is a process of rational model building.

Interest, the second element in CIM, is the attentional factor in learning. If something stands out from its context, it will be more easily remembered, as will things that are extremely vivid or of great personal importance. More often, we try to learn things that are only of mild interest and then, if attention wanders, learning suffers. Interest has the function of keeping us on task.

The third element, memorizing, handles things that do not fit well together, that have no basis in rationality. For instance, the name "cochlea" to represent one of the components of the ear is quite arbitrary to us – there is no reason for it [no reason that we know]. It is [to us] purely arbitrary and no amount of reasoning will assist in "understanding" it. We just have to associate the name and the component.

DESIGN - THE MOCAF MODEL

Based on the CIM model, we can see that there will be three types of elements that are needed within an instructional product: models, cases and facts. Combining these (and any product would have all three) leads to the acronym MoCaF for the design model appropriate for the creation of highly learnable instructional products.

Models are the tools of understanding; they are what lead to comprehension. Cases are the illustrative materials that instantiate the models in particular settings. They are the main means of grabbing and holding attention. As for facts, they are just the basics that need to be brutely memorized.

Models are what drive comprehension. The aim of design in this area is to create models that embody the disparate elements of content while synthesizing them in an artifact [the model] that clearly communicates and is easily learned. Models show how elements relate to one another; they capture relationships and interactions.

The craft of developing models is one of establishing the underlying structures in a field [content expertise is essential here] and of then representing those structures in synthetic form that facilitate communication and understanding (Wurman, 2000).

Cases are the illustrative material in instructional content. They embody the living problems and the living application of the models. They range from simple examples to complex case studies. Of particular interest are those relatively complex cases that mirror difficult real-life settings, such as those used in problem-based medical education or in business education.

Cases are multi-functional in an instructional application. At least three functions can be served:

1. To illustrate the content of a model, instantiating it and situating it in real life.
2. To provide practice to the student in applying knowledge.
3. To test the student's knowledge (either self-testing for monitoring purposes or formal testing for assessment and validation purposes).

Facts are those ill-fitting elements of knowledge that are considered important to know and that hang out there on their own, only incidentally attached to some model or other.

Facts are simple to state, for instance in a textbook or in a presentation. But that does not ensure they will be learned. While simple to state, facts are hard to enrobe in a context that will make them easily learnable. Practice or an eventful context is needed. There are means to accomplish this in instructional terms, such as through games, problems, contests, high-impact media, and so forth. Often, though, when these are not developed, the student is left pretty much to his or her own devices for rehearsing the facts to be learned. This is not an optimal situation.

Learning involves interacting with information. And so it is the design of that information that is crucial to learnability. We are dealing here with the content of the instructional product, that content being modeled through design into a certain form that makes it understandable, interesting and memorizable.

Models, cases, facts are all basically information content of particular sorts, information with which the learner will interact during learning. To a very large extent, then, instructional design is mostly a matter of information design, a notion that needs to become widely recognized. Even the more recent notion of interaction design [often applied to Web site design or to exhibition design] is largely a matter of information design involving models (structures), cases (events), and facts (impressions). In sum, learnability must focus primarily on the content to be included in an instructional product, that is, on information design. That is the key contribution of the learnability approach to design and its central usefulness in the practical design of knowledge artifacts.

FUTURE TRENDS

Perhaps the greatest trend emerging in the future with respect to learnability will be the continuing merging of instructional design into information design. As access to information becomes more ready, we will likely see a reduction of our need to memorize arbitrary information beyond the frequently used or crucial to know kind. Our external memory supports will fill the need for the less needed information.

This merging of the two traditionally distinct design worlds (information and instruction) is particularly informative for the learner-control issue in education. Adult learners like to have more control (or like to think they do) over what they learn, how they learn it and when they learn. They operate more in an access mode than in a traditionally receptive educational mode.

Well-designed information/instruction products will facilitate this approach, being used at times for informational purposes and at other times for instructional purposes. Informal learning (outside of academic structures) and formal instruction both involve learning, both involve interacting with information, both profit from good information design.

The design of e-learning materials also might offer more means of controlling interaction than does for instance the design of textbook or other printed materials. This may or may not be an advantage, depending on a whole host of factors, such as maturity of the learners, prior knowledge of the learners and other context factors. But it does raise once again the general issue of content vs. process.

This philosophical issue remains a challenging one, as well as a thorn in any attempt to devise an overarching theory of learning and instruction. One facet of this issue, and I will conclude with it, is why we speak of the learnability of instructional products. After all, we learn also a great deal from interacting with the world at large, not just with artifacts.

The way to come to grips with this issue is to adopt a wide conception of information, as does the field of semiotics. Information goes far beyond the written word, and beyond the world of illustration too. Information is structure that lies within the world around us, both in its structural elements and its processes (Duchastel, 2002). Some of these are found elements, others are designed artifacts, ones the design of which we can control. This is where we can affect the learnability of a product or of a structured process.

REFERENCES

Bloom, B. (Ed.) (1956). *Taxonomy of educational objectives, Handbook 1: Cognitive domain*. New York: Addison-Wesley.

Carroll, J. (1990). *The Nurnberg funnel*. Cambridge, MA: The MIT Press.

Duchastel, P. (2002). Information interaction. *Proceedings of the Third International Cyberspace Conference on Ergonomics*. Retrieved on March 24, 2004, CD-ROM available through *http://cyberg.wits.ac.za/backg' 2005.html*

Duchastel, P. (2003). Learnability. In C. Ghaoui (Ed.), *Usability evaluation of online learning programs* (pp.299-312). Hershey, PA: Idea Group Publishing.

Egan, K. (1997). *The educated mind.* Chicago: U. of Chicago Press.

Folmer, E., & Bosch, J. (2004). Architecting for usability: A survey. Journal of Systems and Software, *70*, 61-78.

Kearsley, G. (2004). *Explorations in learning & instruction: The theory into practice database*. Retrieved on March 24, 2004 from *http://tip.psychology.org/*

McCarthy, J. (2001). Optimal language learning. *Trends in Cognitive Sciences*, *5*, 132-133.

Nielsen, J. (2000). *Designing Web usability*. Indianapolis: New Riders Publishing.

Reigeluth, C. (1999) (Ed.). *Instructional design theories and models: A new paradigm of instructional theory, Volume II*. Mahwah, NJ: Erlbaum.

Wurman, R.S. (2000). *Information anxiety 2*. Indianapolis, IN: Que Publishing.

Valiant, L. (2000). Robust logics. *Artificial Intelligence, 117*, 231-253.

KEY TERMS

Information Design: A similar soft technology applied to information more broadly for the purpose of successful access to information.

Instructional Design: The soft technology of organizing learning materials and events so that instruction will be most successful.

Interaction Design: A similar soft technology focusing on the processes of interacting with information, particularly in high-impact or strongly emotive contexts.

Learning: The processes used by organisms, including humans, to augment their knowledge base or their skill set for the purposes of better adaptation to their milieu.

Machine Learning: The processes used to fine-tune a program's performance or to augment its knowledge and functionality.

Online Course: A Web-based instructional program that organizes the learning of a student in a particular subject. Not all learning materials need be online and much of an online course involves dynamic interactions with other participants.

Usability: The ease with which a user can accomplish a desired task within a Web site. One also talks of a site being user-friendly.

Learning 3D Face-Animation Model

L

Zhen Wen
University of Illinois at Urbana-Champaign,USA

Pengyu Hong
Harvard University, USA

Jilin Tu
University of Illinois at Urbana-Champaign,USA

Thomas S. Huang
University of Illinois at Urbana-Champaign, USA

INTRODUCTION

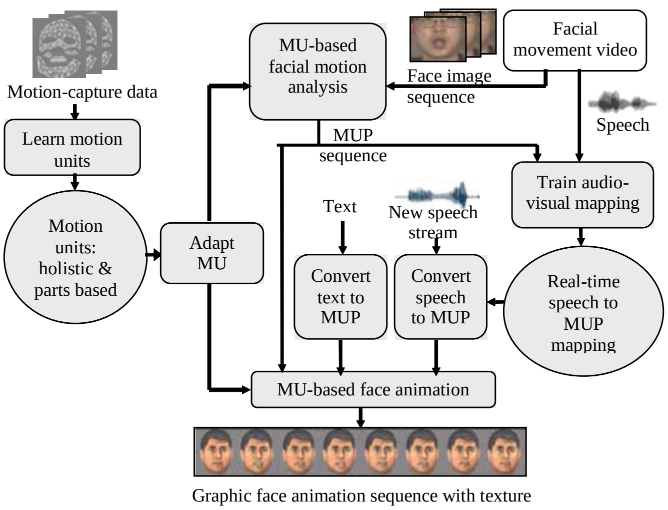
A synthetic human face is useful for visualizing information related to the human face. The applications include visual telecommunication (Aizawa & Huang, 1995), virtual environments and synthetic agents (Pandzic, Ostermann, & Millen, 1999), and computer-aided education.

One of the key issues of 3D face analysis (tracking) and synthesis (animation) is to model facial deformation. The facial deformation is complex, which often includes subtle expressional variations to which people are very sensitive. Therefore, traditional models usually require extensive manual adjustment. Recently, the advance of motion-capture techniques sparked data-driven methods (e.g., Guenter et al., 1998). They achieve realistic animation by using real face-motion data to drive 3D face animation. However, the basic data-driven methods are inherently cumbersome because they require a large amount of data.

More recently, machine learning techniques have been used to learn *compact* and *flexible* face-deformation models from motion-capture data. The learned models have been shown to be useful for realistic face-motion synthesis and efficient face-motion analysis. A unified framework on facial deformation analysis and synthesis is demanded to address in a systematic way the following problems: (a) how to learn a compact 3D face-deformation model from data, and (b) how to use the model for robust facial-motion analysis and flexible animation.

In this article, we present a unified machine-learning-based framework on facial deformation modeling, and facial motion analysis and synthesis. The framework is illustrated in Figure 1. In this framework, we first learn from

Figure 1. Machine-learning-based framework for facial deformation modeling, and facial motion analysis and synthesis



extensive 3D facial motion-capture data a compact set of *Motion units* (MUs), which are chosen as the quantitative visual representation of facial deformation. Then, arbitrary facial deformation can be approximated by a linear combination of MUs, weighted by coefficients called *motion unit parameters* (MUPs). Based on interpolation, the MUs can be adapted to the face model with new geometry topology. MU representation is used in both robust facial motion analysis and efficient synthesis. We also utilize MUs to learn the correlation between speech and facial motion. A real-time audio-to-visual mapping is learned using an artificial neural network (ANN). Experimental results show that our framework achieved natural face animation and robust nonrigid tracking.

Figure 2. (a) Generic model in iFACE; (b) Personalized face model based on the Cyberware scanner data; (c) Feature points defined on a generic model for MU adaptation

(a) (b) (c)

BACKGROUND

Facial Deformation Modeling

Representative 3D spatial facial deformation models include free-form interpolation models (e.g., affine functions, splines, radial basis functions), parameterized models (Parke, 1974), physics-based models (Waters, 1987), and more recently, machine-learning-based models. Because of the high complexity of natural motion, these models usually need extensive manual adjustments to achieve plausible results. To approximate the space of facial deformation using simpler units, people (Tao, 1998) proposed to describe arbitrary facial deformation as a combination of action units (AUs) based on the facial action coding system (FACS; Ekman & Friesen, 1977). Because AUs are only defined qualitatively, they are usually manually customized for computation. Recently, people turned to apply machine learning techniques to learn models from data (Hong, Wen, & Huang, 2002; Kshirsagar, Molet, & Thalmann, 2001; Reveret & Essa, 2001).

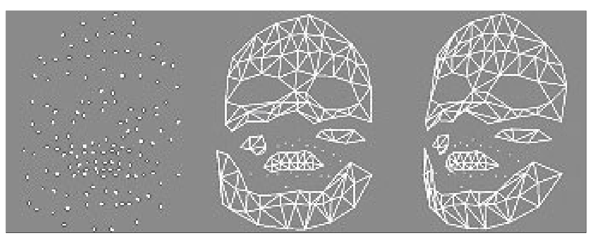
Facial Motion Analysis

Human facial motion analysis is the key component for many applications, such as model-based very-low-bit-rate video coding, audiovisual speech recognition, and expression recognition. High-level knowledge of facial deformation must be used to constrain the possible deformed facial shapes. For 3D facial motion tracking, people have used various 3D deformable model spaces, such as the 3D parametric model (DeCarlo, 1998), B-spline surface (Eisert, Wiegand, & Girod, 2000), and FACS-based models (Tao, 1998). These models, however, are usually manually defined, thus may not capture characteristics of the real facial motion well. Therefore, people have recently proposed to use subspace models trained from real motion data (Reveret & Essa, 2001}.

Facial Motion Synthesis

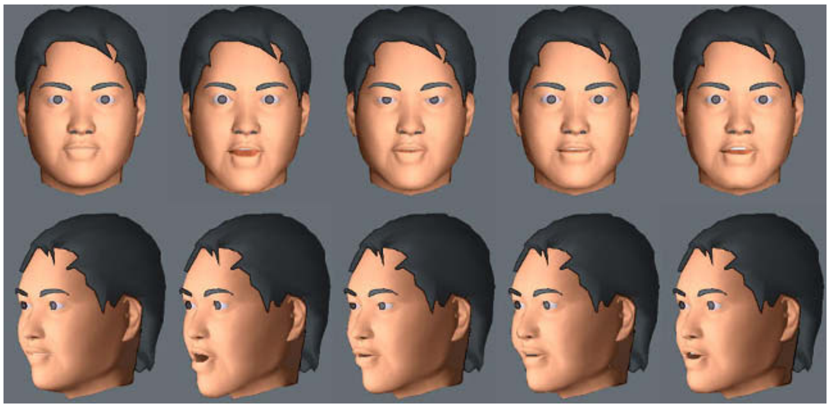
In this article, we focus on real-time-speech face animation. The core issue is the audio-to-visual mapping. HMM-based methods (Brand, 1999) utilize long-term contextual information to generate smooth motion trajectory, but they can only be used in off-line scenarios. For real-time mapping, people proposed various methods such as vector quantization (VQ; Morishima & Harashima, 1991), the Gaussian mixture model (GMM), and ANN (Morishima & Harashima). To use short-time contextual information, people used the concatenated audio feature over a short time window (Massaro et al., 1999) or a time-delay neural network (TDNN; Lavagetto, 1995).

Figure 3. Markers; (a) markers are shown as white dots, (b and c) the mesh is showin in two different viewpoints



(a) (b) (c)

Figure 4. Neutral and deformed faces corresponding to the first four MUs. The top row is a frontal view and the bottom row is a side view.



LEARNING 3D FACE-DEFORMATION MODEL

In this section, we introduce our facial deformation models. We use iFACE (Hong et al., 2002) for MU-based animation; iFACE is illustrated in Figure 2.

The Motion-Capture Database

We use motion-capture data from Guenter et al. (1998). 3D facial motion is captured at the positions of the 153 markers on the participant's face. Figure 3 shows an example of the markers. For better visualization, a mesh is built on those markers (Figure 3b and c).

Learning Holistic Linear Subspace

We try to learn the optimal linear bases, MUs, whose linear combination can approximate any facial deformation. Using MUs, a facial shape $\vec{s}$ can be represented by

$$\vec{s} = \vec{s}_0 + \left(\sum_{i=1}^{M} c_i \vec{e}_i + \vec{e}_0\right) , \qquad (1)$$

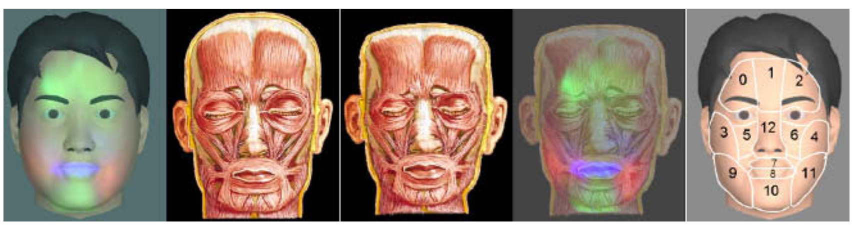
where $\vec{s}_0$ denotes the facial shape without deformation, $\vec{e}_0$ is the mean facial deformation, $\{\vec{e}_0, \vec{e}_1, \ldots, \vec{e}_M\}$ is the MU set, and $\{c_o, c_1, \ldots, c_M\}$ is the MU parameter set.

Principal component analysis (PCA) is applied to learning MUs from the facial deformation of the database. The mean facial deformation and the first seven eigenvectors are selected as the MUs, which capture 93.2% of the facial deformation variance. The first four MUs are visualized by an animated face model in Figure 4.

Learning Parts-Based Linear Subspace

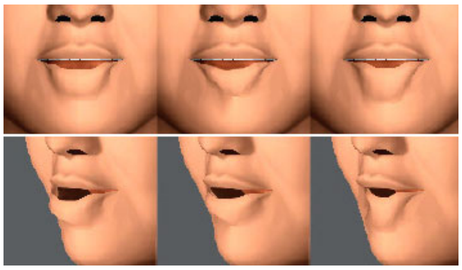
Because the facial motion is localized, it is possible to decompose the complex facial motion into parts. The

Figure 5. (a) NMF learned parts overlaid on the generic face model; (b) the facial muscle distribution; (c) the aligned facial muscle distribution; (d) the parts overlaid on muscle distribution; (e) the final parts decomposition.



(a) (b) (c) (d) (e)

Figure 6. Three lower lip shapes deformed by three of the lower lip parts-based MUs respectively. The top row is the frontal view and the bottom row is the side view.



decomposition helps reduce the complexity in deformation modeling, and improve the analysis robustness and the synthesis flexibility. The decomposition can be done manually using prior knowledge of facial muscle (Tao, 1998). However, it may not be optimal for the linear model used because of the high nonlinearity of facial motion. Parts-based learning techniques, such as nonnegative matrix factorization (NMF; Lee & Seung, 1999), provide a way to help design *parts-based facial deformation models*. We present a parts-based face-deformation model. In the model, each part corresponds to a facial region where facial motion is mostly generated by local muscles. The part decomposition is done using NMF. Next, the motion of each part is modeled by PCA. Then, the overall facial deformation is approximated by summing up the deformation in each part. Figure 5 shows the parts-decomposition results.

The learned parts-based MUs give more flexibility in local facial deformation analysis and synthesis. Figure 6 shows some local deformation in lower lips, induced by one of the learned parts-based MUs. These locally deformed shapes are difficult to approximate using holistic MUs. For each local deformation shown in Figure 6, more than 100 holistic MUs are needed to achieve 90% reconstruction accuracy. Therefore, we can have more flexibility in using parts-based MUs. In face animation, people often want to animate a local region separately, which can be easily achieved by adjusting MUPs of parts-based MUs separately. In face tracking, people may use parts-based MUs to track only regions of interest.

MU Adaptation

The learned MUs are based on the motion-capture data of particular subjects. They need to be adapted to the face model with new geometry and topology. We call this process *MU adaptation*, whose first step fits MUs to a face model with different geometry. It is done by moving the markers of the learned MUs to their corresponding positions on the new face. We interactively build the correspondence of facial feature points shown in Figure 2c via a GUI (graphical user interface). Then, warping is used to interpolate the remaining correspondence. The second step is to derive movements of facial surface points that are not sampled by markers in MUs. It could be done using the radial basis function (RBF) in a similar way to Marschner, Guenter, & Raghupathy (2000).

MODEL-BASED FACIAL MOTION ANALYSIS

In existing 3D nonrigid face-tracking algorithms using the 3D facial deformation model, the subspace spanned by the AUs is used as constraints of low-level image motion. However, the AUs are usually manually designed. For these approaches, our automatically learned MUs can be used in place of the manually designed AUs. We choose to use the learned MUs in the 3D nonrigid face-tracking system in Tao (1998) because it has been shown to be robust and in real time. Details of facial motion estimation can be found in Tao. In the original system, AU is manu-

Figure 7. Typical tracked frames and corresponding animated face models. (a) The input frames; (b) the tracking results visualized by yellow mesh; (c) the front views of the synthetic face animated using tracking results; (d) the side views of the synthetic face. In each row, the first image corresponds to a neutral face.

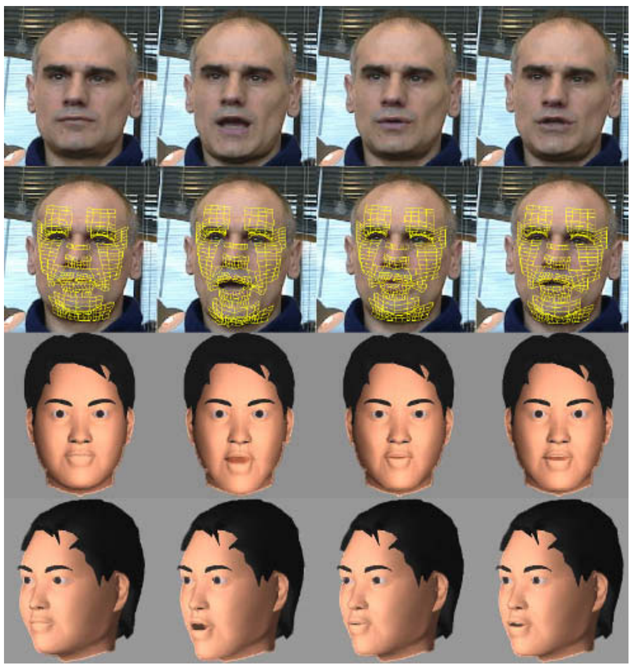


Figure 8. (a) The synthesized face motion; (b) the reconstructed video frame with synthesized face motion; (c) the reconstructed video frame using H.26L codec.

ally designed using Bezier volume and represented by the displacements of vertices of a face-surface mesh. We derive AUs from the learned holistic MUs using the MU adaptation process. The system is implemented to run on a 2.2-GHz Pentium 4 processor. It works at 14 Hz for nonrigid tracking for image size 640×480. The estimated motion can be directly used to animated face models. Figure 7 shows some typical frames that it tracked, along with the animated face model to visualize the results.

The tracking algorithm can be used in model-based face video coding (Tu et al., 2003). We track and encode the face area using model-based coding, and encode the residual in the face area and the background using traditional waveform-based coding method H.26L. This hybrid approach improves the robustness of the model-based method at the expense of increased bit rate. Eisert et al. (2000) proposed a similar hybrid coding technique using different model-based tracking. We capture and code videos of 352×240 at 30 Hz. At the same low bit rate (18 kbits/s), we compare this hybrid coding with the H.26L JM 4.2 reference software. Figure 8 shows three snapshots of a video. Our hybrid coding has 2 dB higher PSNR around the facial area and has much higher visual quality. Our tracking system could also be used in audiovisual speech recognition, emotion recognition, and medical applications related to facial motion disorders such as facial paralysis.

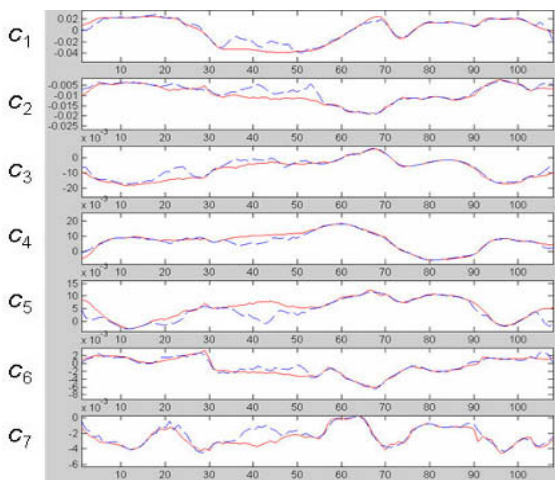
REAL-TIME SPEECH-DRIVEN 3D FACE ANIMATION

We use the facial motion-capture database used for learning MUs along with its audio track for learning audio-to-visual mapping. For each 33-ms window, we calculate the holistic MUPs as the visual features, and 12 Mel-frequency cepstrum coefficients (MFCCs; Rabiner & Juang, 1993) as the audio features. To include contextual information, the audio feature vectors of frames *t*-3, *t*-2, *t*-1, *t*, *t*+1, *t*+2, and *t*+3, are concatenated as the final audio feature vector of frame *t*.

The training audiovisual data is divided into 21 groups, where one is for silence and the remaining groups are generated using the *k*-means algorithm. Then, the audio features of each group are modeled by a Gaussian model. After that, a three-layer perceptron is trained for each group to map the audio features to the visual features. For estimation, we first classify an audio vector into one of the audio feature groups using GMM. We then select the corresponding neural network to map the audio feature to MUPs, which can be used in Equation (1) to synthesize the facial shape. For each group, 80% of the data is randomly selected for training and 20% for testing. A typical estimation result is shown in Figure 9. The horizontal axes represent time. The vertical axes represent the magnitude of the MUPs. The solid red trajectory is the original MUPs, and the dashed blue trajectory is the estimation results.

We reconstruct the facial deformation using the estimated MUPs. For both the ground truth and the estimated results, we divide the deformation of each marker by its maximum absolute displacement in the ground-truth data. To evaluate the performance, we calculate the Pearson

Figure 9. Compare the estimated MUPs with the original MUPs. The content of the corresponding speech track is, "A bird flew on lighthearted wing."



product-moment correlation coefficients (R) and the mean square error (MSE) using the normalized deformations. The Pearson product-moment correlation ($0.0 \le R \le 1.0$) measures how good the global match is between two signal sequences. The coefficient R between the ground truth $\{\vec{d}_n\}$ and the estimated data $\{\vec{d}_n'\}$ is calculated by

$$R = \frac{tr(E[(\vec{d}_n - \vec{\mu}_1)(\vec{d}_n' - \vec{\mu}_2)^T])}{\sqrt{tr(E[(\vec{d}_n - \vec{\mu}_1)(\vec{d}_n - \vec{\mu}_1)^T])tr(E[(\vec{d}_n' - \vec{\mu}_2)(\vec{d}_n' - \vec{\mu}_2)^T])}}, \quad (2)$$

where $\vec{\mu}_1 = E[\vec{d}_n]$ and $\vec{\mu}_2 = E[\vec{d}_n']$. In our experiment, $R = 0.952$ and MSE = 0.0069 for training data, and $R = 0.946$ and MSE = 0.0075 for testing data. Figure 10 shows a typical animation sequence.

Our real-time speech-driven animation can be used in real-time two-way communication scenarios such as videophone. On the other hand, existing off-line speech-driven animation can be used in one-way communication scenarios, such as broadcasting. Our approach deals with the mapping of both vowels and consonants, thus it is more accurate than real-time approaches with only vowel mapping. Compared to real-time approaches using only one neural network for all audio features, our local ANN mapping (i.e., one neural network for each audio feature group) is more efficient because each ANN is much simpler.

Human Emotion Perception Study

The synthetic talking face can be evaluated by comparing the influence of the synthetic face and real face on human emotion perception. We choose three facial expressions, (a) neutral, (b) smile, and (c) sad, and three audio tracks, (a) "It is normal," (b) "It is good," and (c) "It is bad." The information associated with the sentences is (a) neutral, (b) positive, and (c) negative. We capture video and generate animation for all nine combinations of expressions and sentences. Sixteen untrained people participated in the experiments.

The first experiment investigates human emotion perception based on visual-only stimuli. The subjects are asked to infer the emotional states based on the animation sequences without audio. The emotion inference results in terms of the number of the subjects are shown in Table 1. As shown, the effectiveness of the synthetic talking face is comparable with that of the real face.

The second and third experiments are designed to compare the influence of the synthetic face on bimodal human emotion perception and that of the real face. In the second experiment, the participants are asked to infer the emotion from the synthetic talking face animation with audio. The third experiment is the same except that the participants observe the real face instead. In each of the experiments, the audiovisual stimuli are presented in two groups. In the first group, audio content and visual information represent the same kind of information (e.g., positive text with smile expression). In the second group, the relationship is the opposite. The results are combined in Table 2. If the audio content and the facial expressions represent the same kind of information, the human perception is enhanced. Otherwise, it confuses human participants. An example is shown in the fifth and sixth columns

Table 1. Emotion inference based on visual-only stimuli. The S column is the synthetic face, the R column the real face.

		Facial Expression					
		Neutral		Smile		Sad	
		S	R	S	R	S	R
Emotion	Neutral	16	16	4	3	2	0
	Happy	0	0	12	13	0	0
	Sad	0	0	0	0	14	16

Table 2. Emotion inference results agreed with facial expressions. The inference is based on both audio and visual stimuli. The S column is the synthetic face, the R column the real face.

		Facial Expression			
		Smile		Sad	
		S	R	S	R
Audiovisual Relation	Same	15	16	16	16
	Opposite	2	3	10	12

Figure 10. Typical animation frames. Temporal order is from left to right, top to bottom.

of Table 2. The audio content is positive while the facial expression is sad. Ten participants report "sad" for the synthetic face with sad expression. The number increases to 12 if the real face is used. Overall, the experiments show that the effectiveness of the synthetic face is comparable with that of the real face, though it is slightly weaker.

FUTURE TRENDS

Future research directions include investigating systematic ways of adapting learned models for new people, and capturing appearance variations in motion-capture data for subtle yet perceptually important facial deformation.

CONCLUSION

This article presents a unified framework for learning compact facial deformation models from data, and applying the models to facial motion analysis and synthesis. This framework uses a 3D facial motion-capture database to learn compact holistic and parts-based facial deformation models called MUs. The MUs are used to approximate arbitrary facial deformation. The learned models are used in 3D facial motion analysis and real-time speech-driven face animation. The experiments demonstrate that robust nonrigid face tracking and flexible, natural face animation can be achieved based on the learned models.

ACKNOWLEDGMENT

This work was supported in part by NSF Grant CDA 96-24386 and IIS-00-85980. We thank Dr. Brian Guenter, Heung-Yeung Shum, and Yong Rui of Microsoft Research for the face-motion data.

REFERENCES

Aizawa, K., & Huang, T. S. (1995). Model-based image coding. *Proceedings of IEEE, 83*, 259-271.

Brand, M. (1999). Voice puppetry. *Proceedings of SIGGRAPH 1999*, 21-28.

DeCarlo, D. (1998). *Generation, estimation and tracking of faces*. PhD thesis, University of Pennsylvania.

Eisert, P., Wiegand, T., & Girod, B. (2000). Model-aided coding: A new approach to incorporate facial animation into motion-compensated video coding. *IEEE Transactions on Circuits and Systems for Video Technology, 10*(3), 344-358.

Ekman, P., & Friesen, W. V. (1977). *Facial action coding system*. Consulting Psychologists Press Inc.

Guenter, B., et al. (1998). Making faces. *Proceedings of SIGGRAPH 1998*, 55-66.

Hong, P., Wen, Z., & Huang, T. S. (2002). Real-time speech driven expressive synthetic talking faces using neural networks. *IEEE Transactions on Neural Networks, 13*(4), 916-927.

Kshirsagar, S., Molet, T., & Thalmann, N. M. (2001). Principal components of expressive speech animation. *Proceedings of Computer Graphics International*, 38-44.

Lavagetto, F. (1995). Converting speech into lip movements: A multimedia telephone for hard of hearing people. *IEEE Transactions on Rehabilitation Engineering*, 90-102.

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature, 401*, 788-791.

Marschner, S. R., Guenter, B., & Raghupathy, S. (2000). Modeling and rendering for realistic facial animation. *Proceedings of Workshop on Rendering*, 231-242.

Massaro, D. W., et al. (1999). Picture my voice: Audio to visual speech synthesis using artificial neural networks. *Proceedings of Audio-Visual Speech Processing*, 133-138.

Morishima, S., & Harashima, H. (1991). A media conversion from speech to facial image for intelligent man-machine interface. *IEEE Journal on Selected Areas in Communications, 4*, 594-599.

Pandzic, I., Ostermann, J., & Millen, D. (1999). User evaluation: Synthetic talking faces for interactive services. *The Visual Computer*, *15*, 330-340.

Parke, F. I. (1974). *A parametric model of human Faces*. PhD thesis, University of Utah.

Rabiner, L., & Juang, B. H. (1993). *Fundamentals of speech recognition*. Prentice Hall.

Reveret, L., & Essa, I. (2001). Visual coding and tracking of speech related facial motion. *Proceedings of Workshop on Cues in Communication*.

Tao, H. (1998). *Non-rigid motion modeling and analysis in video sequence for realistic facial animation*. PhD thesis, University of Illinois, Urbana-Champaign, IL.

Tu, J., et al. (2003). Coding face at very low bit rate via visual face tracking. *Proceedings of Picture Coding Symposium*, 301-304.

Waters, K. (1987). A muscle model for animating three-dimensional facial expression. *Computer Graphics, 22*(4), 17-24.

KEY TERMS

Artificial Neural Networks: Type of machine learning paradigm. They simulate the densely interconnected, parallel structure of the mammalian brain. They have been shown to be powerful tools for function approximation.

Facial Deformation Model: Model that explains the nonrigid motions of human faces. The nonrigid motions are usually caused by speech and facial expressions.

Facial Motion Analysis: Procedure of estimating facial motion parameters. It can also be called "face tracing." It can be used to extract human face motion information from video, which is useful input for intelligent video surveillance and human-computer interaction.

Facial Motion Synthesis: Procedure of creating synthetic face animations. Examples include text-driven face animation and speech-driven face animation. It can be used as an avatar-based visual interface for human-computer interaction.

Machine Learning Techniques: Techniques that can automatically improve computational models based on experiences

Motion Capture: Techniques that measure complex motions. One type of motion-capture techniques places markers on the target object and tracks the motions of these markers.

Motion Unit Model: Facial deformation model we propose. It is based on measurement of real facial motions. Motion units (MUs) are a set of vectors whose linear combinations can be used to approximate arbitrary facial shapes.

Learning Portals as New Academic Spaces[1]

Katy Campbell
University of Alberta, Canada

INTRODUCTION

Many functional definitions emphasize a portal as an integrated system providing a gateway to organized data (c.f., Batson, 2000; Copeland, 2001; Eisler, 2001; Looney & Lyman, 2000). However, a learning portal may go beyond the information management function to provide important mechanisms for reaching out to new populations of learners and engaging them in new ways to facilitate learning and development. Beyond serving as a gateway and an organizer, a portal can provide access to a broader range of contemporary information and learning resources (experts, teachers, researchers, mentors), encourage enriched interaction with those resources and with other learners anywhere in the world, and support new models of teaching, learning and research. Ultimately, a collaborative, community-based process of designing and implementing a portal may support institutions in reorienting towards a user-centered learning community.

BACKGROUND

Portals and a Transformed Learning Environment

Universities are seeking ways to manage emerging areas of research and discipline specialization, learner profiles, and partnerships with learning providers that challenge the autonomy of the single-source institution. The public has expressed strong interest in alternative methods for delivering, supporting, and facilitating learning — any time, any place, any pace — required in new knowledge-intensive environments and enabled by converging information and communication technologies. Therefore, the decision to implement a campus portal for enhanced learning opportunities must address issues of equity and access, flexibility, innovation, personalization, credibility, quality, transparency, and transferability within the framework of evolving institutional goals and strategies.

Both Campbell (2001) and Batson (2000) contend that *commercial portals* are built on different values and assumptions than those of the academic community, and pursue different goals and purposes. Erhmann (2000) identifies service provision, flexibility and responsiveness of instruction, the enrichment and extension of academic communities, attracting and retaining students and staff, fostering universal, frequent use of computing communications, and sustainability.

A learning portal expands on traditional academic space, which has traditionally been defined as physical infrastructure with related resource structures that shapes the nature of the interactions that occur within it (Batson, 2000). This space has an important socialization function: Members of the community know how to speak and act within these spaces, understand power relationships by the way these spaces organize interactions (e.g., rows of desks with a lectern at the front of the room) and, once acculturated, can subvert the purposes of these spaces. The nature of teaching and learning has been entirely defined by a familiar landscape, the physical classroom, where learning events were structured by place and time and format.

This landscape has fundamentally changed. Faculty have old maps and must redefine their relationships with learners, with new ways of representing knowledge, with research colleagues, and with external communities such as the corporate world. Learners demand customized learning experiences that are flexible, authentic, and relevant, have no brand-loyalty and expect program mobility. This is a challenge to administrators whose management strategy focuses on internal factors like time-definite program completion (e.g., the 4-year undergraduate degree).

FUTURE TRENDS

Although institutions have ranged themselves along an academic space continuum from primarily face-to-face to primarily virtual, most have settled on a technology-enhanced, or blended approach to learning and access. Employing alternative forms of instructional and delivery models, this approach includes: synchronous tools and environments such as classroom lectures, audio and videoconferencing, and data conferencing; and asynchronous tools such as computer-mediated conferencing and other communications systems, learning management systems, and print and digital media. Much of the learning content and interactions can be stored as learning objects and extended and reused in digital repositories. This approach fundamentally realigns and redefines institutional infrastructure to be more learner-centric and open in design and support and include extended information services. It also has a significant social effect on the academic community, raising questions about academic freedom, intellectual property rights management, and the nature of knowledge discovery, representation, and stewardship.

Learning portals can provide the functionality of consumer systems, and at the same time, support the social, cultural, and political goals of HE. While more or less resisting the culture of the corporation, universities nevertheless have begun to adopt the concept of portals as learning storefronts (Galant, 2000). Yet, in order to respect HE values of knowledge creation and dissemination for the greater social good, these portals must go beyond the functional requirements and gateway view of commercial portals.

Gilbert (2000) and Eisler (2000) identify major categories into which a variety of portal features and functions can be organized: gateways to information, points of access for constituent groups, and community/learning hubs. A synthesis of public reports identifies the range of stakeholders that should be involved in this task, and their values and functional requirements. Principles for the new portal-as-learning-environment include:

- Inclusiveness: the portal design must support diverse communities including learners who are: older professionals, at a distance, challenged, at different life cycles, learning outside of formal structures, and those with alternative languages, cultural, and perceptual needs; both present and virtual faculty; multidisciplinary teams of researchers; local and international academic, business, and political partners, and others.
- Integration: learning management systems such as Blackboardä and WebCTä have begun to develop and refine enterprise systems that integrate instructional, delivery, and administrative systems. These portals have evolved from a teaching/learning orientation and reflect institutional movement towards a seamless, multi-purpose, integrated learning environment.
- Learner-centeredness: portal design is based on the interrelated concepts of customization and personalization, reflecting learning environments in which learners can build learning portfolios based on their circumstances, experiences, and current needs. Traditionally, institutional Web sites have been owner-centric.
- Accessibility: the new economy implies that the intellectual resources of the university should be packaged and made available to a global community. Portals identify, organize, and represent these resources in ways that make them easy to retrieve, use, and reuse (see, for example, MIT's Open Knowledge Initiative, or OKI).
- Flexibility: for many reasons, including changes in professional accreditation, a globally mobile workforce, new and emerging professions, and life events, individuals will search for opportunities to time-shift, place-shift, and construct individual programs from many providers. A well-designed learning portal will be scalable and act as a gateway to these opportunities.
- Transparency: a learning portal makes the institution's strategic directions visible to the community. Learners, external research communities, the private sector, and others construct their own "footprint". They can search for all of the services they need, and deal directly with the systems that facilitate their interactions with the environment. Portals can help the community discover and promulgate best practices.
- Accountability: as the learning and support environment becomes more transparent, and as learning opportunities become more available and flexible, community members will expect to be able to evaluate the services and resources to which they have access. As rich information hubs, learning portals can make the institution's quality framework apparent and available for querying.
- Expanded and blended learning communities: a learning portal manages transparent and reliable communication tools, which increase access to resources and social learning communities. These tools are easily accessible from the portal and can therefore include and support group members from different institutions, organizations, regions, and cultures. These communities broaden and enrich the learning environment and enhance inclusiveness. Looney and Lyman (2000) believe that the value of a learning portal is that "it can be used to engage constituent groups, *empower* them with access to information resources and communication tools, and ultimately *retain* them by providing a more encompassing sense of membership in an academic community" (p.33).
- Flattened structures: virtual academic spaces do not support status clues to the same extent as traditional spaces. For example, a physical campus contains buildings with classrooms, labs, information resources, and administrative offices. Very of-

ten administrative offices are grouped around department chairs and the dean's office – there is an inner sanctum access to which is constrained by furniture, walls, workstations, and staff. Appointments must be made. Classrooms sometimes have permanently fixed seating arrangements (e.g., tiers) facing a lectern. Resources are under lock and key. There are barriers. These physical constraints disappear with online access. Implications include democratizing interactions within, and external to, the institution.

- Collaboration: a campus learning portal will fundamentally change the way universities treat its intellectual capital, increasing opportunities for collaborative work on campus, nationally, and internationally. It is critical to involve the owners of this capital in the design of the portal environment. As faculty members, support staff, librarians, learners, administrators, alumni, the public, and partner institutions engage each other a deeper, transformed understanding of the whole knowledge management enterprise will emerge (Beller & Or, 1999).

CONCLUSION

The learning portal can be designed to include a wide range of information, communication, and development tools (Alharti, Bourne, Dawant, & Mayadas, 2000; Eisler, 2000; Erhmann, 2000; Gold, 2001; Moore, 2002; Paadre,

Figure 1. Resources and services for instructors, students, and the public

Resources and Services for Instructors, Students and the Public	
Tools	**Example**
• an array of interactive multi-media tools • extended elements of traditional library services • increasingly rich interlinked libraries of both traditional and electronic resources • access to extra-curricular virtual events • a broadband-enhanced learning object repository • online advising • uniformity of single campus-wide interface linking all courses	A School of Business develops a portal that includes a large database of digital resources for courses in economics, organizational behavior, accounting and finance, human resources management and management and information systems. Resources include case studies (text and video), simulations (video), interactive demonstrations (e.g. graphs in which variables can be changed), exemplars of student projects, historical artifacts, etc. Faculty and students can search for and download these resources using a custom search engine, and can also share or "deposit" new resources, extend existing materials with quizzes; commentary, additional activities. The database (or repository) links directly to other repositories like MERLOT. Users are invited to review resources, similar to Amazon.com's system. The repository also links to the University's library system which is increasing its full-text journal holdings online and is also acquiring digital resources such as archived speeches. Users can customize their portals by choosing among options when they first logon. For example, there are several student-led clubs that hold regular meetings, sponsor guest speakers and events, and offer scholarships. A user may select which of these clubs to include on their portal so that they receive only those announcements on the homepage. The School's IT department has created a library of templates that faculty may use/customize for their own courses. The templates include case-based, collaborative projects, role-plays and other methods – faculty upload text or media files into the templates in categories including "course objectives", "course readings", etc. The templates can be used with a campus-wide LMS. The IT team has created a library of graphics, banners, assessment activities, etc.

Figure 2. Learning environments and tools

Learning Environments and Tools	
Tools	**Example**
• intelligent agents, such as "tutors" • course space • access to learning support systems • interactive discussion spaces, open to the world • an integrated suite of tools for instructional designers, content authors, instructors and learners • flexible delivery platforms	One university is developing a learning portal for faculty called the Faculty Toolkit. Designed to support faculty professional development, the portal is based on a dynamic content management system and includes tools and resources for faculty interested in redesigning their courses or learning activities. The management system will contain examples of course designs; access to peer mentors and their stories and observations; links to research on teaching and learning in HE; templates, checklists, guides, and information resources; scheduled conferences, workshops and institutes, both virtual and campus-based; moderated discussion forums; team-based activities; access to online consultations with instructional designers; evaluation tools, and more. Faculties will be able to design their own interfaces to sit on top of the database. For example, the Faculty of Education is creating a building metaphor with offices, libraries, and classrooms.

Heiki, & King, 2000). These tools could be divided into categories such as those included in Figures 1-3.

REFERENCES

Alharti, M., Bourne, J., Dawant, M., & Mayadas, F. (2000, November). Web portals for online learning. Paper presented at the *6th International Conference on ALN*.

Batson, T. (2000, November). Campus portals and faculty development. Paper presented at *Syllabus 2000: New Dimensions in Educational Technology Conference* held in Boston, Massachusetts.

Beller, M., & Or, E. (1999, October). On the transition from traditional and open university models to a virtual university model. Paper presented at the *5th International Conference on ALN*, Maryland.

Figure 3. Research and administrative support

Research and Administrative Support	
Tools	**Example**
• publishing tools • links to the student information system • a capacity for individual users to customize and organize personal resources external to the campus • horizontal links among departments on campus and vertical links to national academic fields	One Faculty of Adult and Continuing Education has implemented a portal for internal and external communities – including the professional associations with which their programs are aligned. For example, the Association of Professional Engineers can select only those activities, program announcements, course changes etc that relate directly to their professional development needs, and include the link on their own site. When their members logon to the APE site, they link directly only to relevant information. For Faculty staff, access is provided to online research tools such as qualitative analysis software, style guides, and tutorials on proposal writing. The Faculty maintains its own student registration system but the portal allows it to tie in to the campus' administrative system for record-coordination, financial services, etc.

Campbell, J. (2001). The case for creating a scholars' portal to the Web: A white paper. *Portal: Libraries and the Academy*, *1*(1), 15-21.

Copeland, L. (2001, June). Ford launches massive corporate portal. Retrieved November 7, 2003 from the World Wide Web at *http://www.computerworld.com/cwi/story/0,1199,NAV47_STO61399,00.html*

Eisler, D.L. (2000, September). Campus portals: Supportive mechanisms for university communication, collaboration, and organizational change. *Syllabus Magazine*, *14*(1). Retrieved November 5, 2003 from the World Wide Web at *http://www.syllabus.com/syllabusmagazine/sep00_fea.html*

Erhmann, S. (2000, November). Evaluating campus portals - Key ideas. *Syllabus 2000: New Dimensions in Educational Technology Conference,* Boston, MASS.

Galant, N. (2000, November). The Portal for online objects in learning (POOL): An advanced eLearning solution. Paper presented at *TeleLearning NCE 5th Annual Conference,* Toronto, ON.

Gilbert, S. (2000, August). Portal decisions demand collaboration: can portals support it? The TLT Group. Retrieved November 8, 2003 from the World Wide Web at *http://www.tltgroup.org/gilbert/SyllabusCol2.htm*

Gold, S. (2001). A constructivist approach to online training for online teachers. *Journal of Asynchronous Learning Networks*, *5*(1), 35-57.

Looney, M., & Lyman, P. (2000, July/August). Portals in higher education: What are they, and what is their potential? *EDUCAUSE Review.*

Moore, J.C. (2002). *Elements of quality: Synthesis of the August 2002 seminar*. Needham, MA: The Sloan Consortium.

Paadre, H., & King, S. (2000). Electronic community and portals. Holy Cross College. Retrieved November 7, 2003 from the World Wide Web at *http://www.mis2.udel.edu/ja-sig/holycross.doc*

KEY TERMS

Accessibility: Refers to problems encountered by Internet users with perceptual and cognitive challenges, physical conditions or other factors such as geographical location, sociocultural, political and economic issues, language, and so forth, which influence their use of the Web.

Blended Learning: Defined broadly, blended learning is the integration of classroom face-to-face learning with online or technology-supported learning, including a range of pedagogical approaches and delivery systems. Strategic applications of blended learning have shown achievement of learning gains while tackling other problems faced by our universities, most notably the pressures of increasing class sizes, and limitations in funding, classroom space, and learning support.

Inclusiveness: In this context, inclusiveness is a both a value and a design process. Inclusive environments are purposefully designed to address accessibility challenges and to make resources and services available to the broadest (and most diverse) possible audience. An approach based on inclusiveness is also known as Universal Instructional Design (UID), and User-Centered Design (UCD).

Learning Portals: Integrate information, administrative, communication, research, teaching, and learning support systems with global networks of resources and services. Learning portals are typically designed to increase flexibility and accessibility to institutional resources and to encourage interaction and engagement with diverse communities of users.

User-centered Design (UCD): A client or user-centered philosophy in which the individual is at the center of an iterative design process encompassing multiple factors in an interaction between user and information product. UCD considers cognitive, sociocultural, political, and technological factors involved in defining user tasks and goals that drive the design and development of software, Web sites, information systems and processes – anything with which people interact. UCD is concerned with the usefulness, usability, desirability, legibility, learnability, accessibility and meaningfulness of an information product.

ENDNOTE

[1] This article is based on: Campbell, K. & Aucoin, R. (2003). Values-based design of learning portals as new academic spaces. In M. Sheehan & Ali Jafari (Eds.), *Designing portals: Opportunities and challenges* (pp.148-170). Hershey, PA: Idea Group Publishing.

Learning Systems Engineering

Valentina Plekhanova
University of Sunderland, UK

BACKGROUND

Traditionally multi-agent learning is considered as the intersection of two subfields of artificial intelligence: multi-agent systems and machine learning. Conventional machine learning involves a single agent that is trying to maximise some utility function without any awareness of existence of other agents in the environment (Mitchell, 1997). Meanwhile, multi-agent systems consider mechanisms for the interaction of autonomous agents. Learning system is defined as a system where an agent learns to interact with other agents (e.g., Clouse, 1996; Crites & Barto, 1998; Parsons, Wooldridge & Amgoud, 2003). There are two problems that agents need to overcome in order to interact with each other to reach their individual or shared goals: since agents can be available/unavailable (i.e., they might appear and/or disappear at any time), they must be able to find each other, and they must be able to interact (Jennings, Sycara & Wooldridge, 1998).

Contemporary approaches to the modelling of learning systems in a multi-agent setting do not analyse nature of learning/cognitive tasks and quality of agents' resources that have impact on the formation of multi-agent system and its learning performance.

It is recognised that in most cognitively driven tasks, consideration of agents' resource quality and their management may provide considerable improvement of performance process. However, most existing process models and conventional resource management approaches do not consider cognitive processes and agents' resource quality (e.g., Norman et al., 2003). Instead they overemphasise the technical components, resource existence/availability problems. For this reason, their practical utilisation is restricted to those applications where agents' resources are not a critical variable. Formal representation and incorporation of cognitive processes in modelling frameworks is seen as very challenging for systems engineering research.

Therefore, future work in engineering the learning processes in cognitive system is considered with an emphasis on cognitive processes and knowledge/skills of cognitive agents as a resource in performance processes. There are many issues that need new and further research in engineering cognitive processes in learning system. New/novel directions in the fields of systems engineering, machine learning, knowledge engineering, and mathematical theories should be outlined to lead to the development of formal methods for the modelling and engineering of learning systems. This article describes a framework for formalisation and engineering the cognitive processes, which is based on applications of computational methods. The proposed work studies cognitive processes, and considers a cognitive system as a multi-agents system.

This project brings together work in systems engineering, knowledge engineering and machine learning for modelling cognitive systems and cognitive processes. A synthesis of formal methods and heuristic approaches to engineering tasks is used for the evaluation, comparison, analysis, evolution and improvement of cognitive processes.

In order to define learning processes, cognitive processes are engineered via a study of knowledge capabilities of cognitive systems. We are not interested in chaotic activities and interactions between cognitive agents (since cognitive tasks require self-managing activities/work), nor interested in detailed tasks descriptions, detailed steps of tasks performance and internal pathways of thoughts. Rather, we are interested in how available knowledge/skills of cognitive agents satisfy required knowledge/skills for the performance of the cognitive tasks.

The proposed research addresses the problem of cognitive system formation with respect to the given cognitive tasks and considers the cognitive agent's capabilities and compatibilities factors as critical variables, because these factors have an impact on the formation of cognitive systems, the quality of performance processes and applications of different learning methods.

It is recognised that different initial knowledge capabilities of the cognitive system define different performance and require different hybrid learning methods. This work studies how cognitive agents utilise their knowledge for learning the cognitive tasks. Learning methods lead the cognitive agent to the solution of cognitive tasks. The proposed research considers a learning method as a guide to the successful performance. That is, initial knowledge capabilities of cognitive agents are correlated with learning methods that define cognitive processes. An analysis of impact of different cognitive processes on the performance of cognitive agents is provided.

This work ensures support for a solution to resource-based problems in knowledge integration and scheduling of cognitive processes to form a capable cognitive system for learning the required tasks.

AIMS AND OBJECTIVES

The aims of the project are to develop a formal method for the modelling and engineering of cognitive processes. Capability and compatibility factors have an impact on the formation of cognitive systems, the performance processes and define different learning methods. Therefore this work studies cognitive processes and knowledge capabilities of cognitive systems to ensure the required level of the learning and performance of the cognitive systems. In order to support the formation of a cognitive system that will be capable of learning the required tasks within the given constrains, this work addresses problems of the knowledge integration and scheduling for cognitive system modelling, taking into account critical capability and compatibility factors. Study of learning conditions in cognitive systems defines an important task of the proposed project.

The individual measurable objectives are:

- Evaluation of knowledge integration and scheduling approaches in cognitive systems.
- Evaluation of existing machine learning approaches in cognitive systems.
- Determination of the impact of capability and compatibility factors on the formation of cognitive systems.
- Development of knowledge integration metrics.
- Development of knowledge integration models for the formation of the cognitive systems.
- Development of scheduling models for learning of cognitive systems.

METHODOLOGY AND JUSTIFICATION

In order to identify the best learning processes we analyse the cognitive processes. A scenario for engineering the cognitive processes is based on the following steps (Figure 1).

The methodology of the proposed project is based on the following new theoretical basis (Plekhanova, 2003).

- Profile Theory and Machine Learning: For formal modelling of complex systems we utilise the profile theory (Plekhanova, 1999a). A profile is considered as a method for describing and registering multifaceted properties of objects. There are important practical applications of the profile theory (Plekhanova, 2000a, 2000b). For instance, internal properties of the system elements such as capability and compatibility factors are critical variables in modelling, design, integration, development and management of most modern complex systems and their structure. Table 1 provides a comparison of capability and compatibility problems of technical and soft systems.

Figure 1. A scenario for engineering the cognitive processes

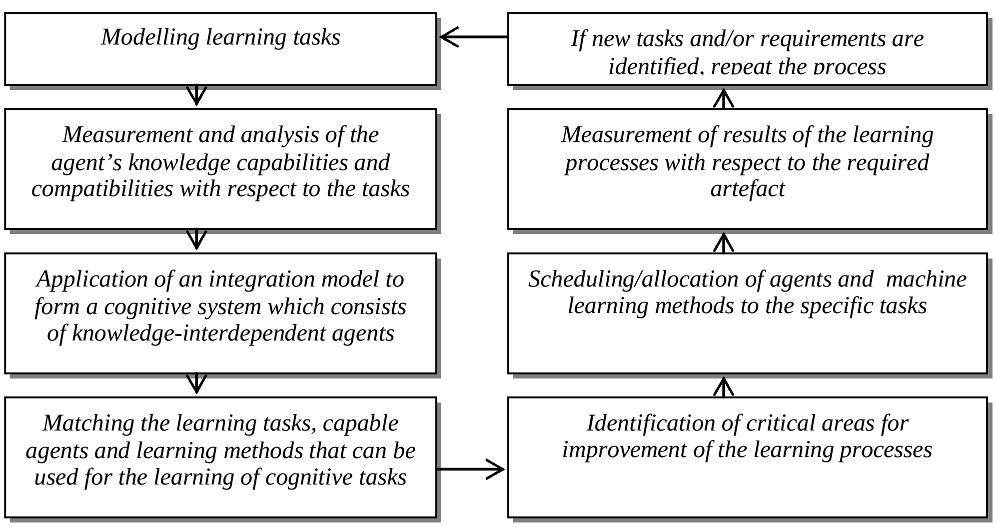


Table 1. A comparison of capability and compatibility problems

Technical Systems	Soft Systems
In *technical systems* the internal characteristics of technical elements are described in specifications, standards and formal documents (i.e., are known a priori) from which it is not difficult to conclude whether combinations of elements are capable and compatible or not and whether they can be used for technical system design, development and construction. Each capability and compatibility factor can be represented by one characteristic.	There are a number of real-world examples when an object factor cannot be described by just one characteristic. For example, systems such as human resources, software, information systems, and cognitive agents, where the internal multifaceted properties can be changed with time, and capability and compatibility factors cannot be defined by one characteristic alone, and cannot be explicitly measured. These systems are termed *soft systems*. At the present time there are problems in the formal definition of specifications/standards and metrics that allow one to determine the capability and compatibility of such complex systems. For this reason, heuristic approaches are used for soft complex system modelling.

Profile theory is used for formal modelling of cognitive agents/systems since existing mathematical theories are limited. In particular, contemporary mathematical theories describe objects, where each internal factor is represented by one meaningful piece (e.g., set theory - an element) or two pieces of meaningful information (e.g., fuzzy set theory - an element and a membership function).

Knowledge factors are considered as basic internal factors in the modelling of cognitive agents, since agents must have particular knowledge capabilities to perform and learn their tasks. In a description of the knowledge of cognitive agents we identify the importance/weight of the factor for the performance of the task; time or factor existence/non-existence; and other specific internal multifaceted properties, for example the property (level, grade, degree) of the factor.

In particular, knowledge of the cognitive agent is described by a set of knowledge factors; each factor is defined by multiple characteristics. A set of such factors forms a knowledge profile (Plekhanova, 1999a). Each factor is represented by qualitative and quantitative information. Quantitative description of the *i*th knowledge factor is defined by an indicator characteristic, property, and weight. In a simple way, a profile can be defined as follows (see Figure 2):

The profile theory is used for formalisation of cognitive systems and cognitive processes, and for the identification of critical areas in learning performance where improvement should be taken. In particular, engineering the cognitive processes is considered to provide improvement of learning process by means of integrating adaptive machine learning into the profile theory. In order to model cognitive processes the profile theory is combined with machine learning methods, which are applied to the initial available knowledge capabilities of the cognitive system to define learning methods. (It is expected that different initial knowledge capabilities of the cognitive system require different hybrid learning methods.)

Machine learning methods are used for formalisation and modelling of learning processes via applications of the profiles. It allows consideration of dynamics in learning processes (i.e., modelling of the *i*th profile factor $e_i(t) = <\varepsilon_i(t),\ v_i(t),\ w_i(t)>$ in the profile b). We should analyse existing machine learning methods, match them to learning tasks with relevance to available knowledge capabilities of cognitive agents and consider cognitive processes. A profile is considered as a model for the description of cognitive processes. That is, a new machine learning method will be developed and incorporated into an engineering framework for cognitive processes.

This research considers knowledge factors as critical variables in learning processes and addresses problems in the formation of a cognitive system that can be capable of learning. In particular, a cognitive system is defined by knowledge-interrelated agents, their flexible cognitive structure and cognitive processes. A teacher is defined as a learning oracle. Soft factors may be defined as a “noise”

Figure 2. Definition of the profile

A profile b is defined as a set of factors $b_1, b_2, ..., b_n$: $b = \{ b_i ,\ i = \overline{1, n} \}$, where the ith factor b_i is represented by a pair $b_i = (\iota_i , e_i)$ with

- n - a number of factors
- ι_i - an identification of the ith factor, that is, a name or label or type of the ith factor
- e_i - the 3-tuple of the ith factor as the Cartesian product: $e_i = <\varepsilon_i ,\ v_i ,\ w_i >$, where
    - o ε_i - indicator characteristic, which indicates the factor presence in the description of a cognitive agent, the existence of certain conditions; for example, ε_i may represent a binary case; a number of times of factor utilisation; or may be defined as a time characteristic $\varepsilon_i = \varepsilon_i(t)$
    - o v_i - property of the ith factor: $v_i \geq 0$; v_i can be defined as a function of time $v_i = v_i(t)$
    - o w_i - weight of a factor which defines either the factor importance or the factor priority: $w_i \geq 0$; w_i can be also considered as a function of time $w_i = w_i(t)$.

in data modelling for the training sets in machine learning.

This work addresses the problems of knowledge integration of the cognitive system in order to provide a better learning performance. A challenge for learning is to ensure the existence of a desired level of performance of a cognitive system. There is a need to make a formal analysis of the available knowledge of cognitive agents in order to ensure the learning of the tasks at a desired performance level while utilising the available knowledge capabilities effectively and efficiently.

This research deals with the problem of agent allocation in a cognitive system. This problem addresses not only task scheduling as in traditional approaches but also scheduling machine learning methods and knowledge of cognitive agents. The proposed project will develop a new scheduling approach where the agent allocation problem has specific emphasis on the following aspects: cognitive agents are allocated to tasks according to their multiple knowledge capabilities; the agent's knowledge capabilities must satisfy the particular combination of knowledge required for a task; agents of the cognitive systems should be compatible with each other (Plekhanova, 1999b); and learning methods are relevant to the available knowledge capabilities of the cognitive agents and system. The consideration of all these aspects defines a problem of knowledge integration in cognitive systems. This work will use formal methods for an integration of cognitive capabilities and compatibilities, and for an analysis of how system capabilities satisfy the learning of the tasks.

Capability and compatibility factors have considerable impact on the process of system integration. An integration model encompasses integration criteria, priorities of the knowledge profiles and knowledge integration goals. Knowledge integration goals are the improvement of available knowledge or generation of new/novel knowledge for better learning performance.

This work addresses problems of effectiveness of the learning processes, their convergence (Vapnik, 1998), stability and accuracy.

Therefore, the adventure in this research is that cognitive processes will be incorporated into multi-agent system development by a synthesis of systems engineering with knowledge engineering and machine learning methods. The combination of machine learning methods with profile theory will provide a more flexible adaptive framework for cognitive tasks performance. That is, the proposed method for the modelling and engineering of cognitive systems and cognitive processes can be used in systems engineering and machine learning for a formalisation of cognitive processes, cognitive systems, and capability and compatibility aspects.

NOVELTY

The proposed project is particularly novel in its approach to learning processes that incorporate a synthesis of systems engineering, knowledge engineering and machine learning methods. There are no formal methods for knowledge integration and scheduling for learning of cognitive systems where capability and compatibility factors are critical variables. Existing machine learning approaches do not address scheduling problems in learning methods. We will develop a new scheduling approach where we consider scheduling machine learning methods and knowledge of cognitive agents versus task scheduling in traditional approaches. New machine learning methods will be developed and incorporated into an engineering framework for cognitive processes. Moreover, the proposed project brings together work in cognitive systems, systems engineering, knowledge engineering and machine learning for the modelling of cognitive processes.

The proposed project is timely because of the availability of new formal methods for engineering cognitive systems. The work is highly topical at present as demonstrated by a large interest in academia and great needs of industry, in particular, in:

- Software/Systems Engineering: Project resource capability and compatibility aspects have become the focus of performance process improvement. However, most contemporary approaches to the formation of project resources (Norman et al., 2003) do not examine their capability and compatibility factors. There is a need to develop evaluation techniques for people's capability, resource capability and compatibility in order to provide support for effective solutions to resource integration management in cognitive systems (Plekhanova, 2002). In particular, methods are of particular merit that incorporate a comparison of cognitive processes, resource capabilities and compatibilities.
- Scheduling: Contemporary approaches to resource scheduling are based on the detailed description of tasks assuming that a resource pool is given and defined by a manager and resources are capable of performing any project task. Existing resource scheduling methods address the issues of resource availability and utilisation, and are not concerned with the capability and compatibility of project resources. Furthermore, in traditional scheduling approaches, the objectives for the allocation of limited resources are to determine the allocation of resources that maximise total benefits subject to the limited resource availability. Contemporary approaches to resource allocation are founded on the assumption that different tasks require equal capability resources, and only one skill is involved. Hence, they cannot be successfully used for software projects where different software tasks require changing different sets of multiple knowledge and skill capabilities in an overall system (Plekhanova, 1998).
- Software Tools for Resource Scheduling: There are many scheduling tools that provide different approaches: event-oriented (PERT), activity-oriented (CPM), actions-oriented (TASKey PERSONAL), or offer a wide variety of scheduling options (SAP). Nevertheless, there are no tools that support an analysis of resource capabilities/compatibilities and their impact on project scheduling (Plekhanova, 2000c). Most existing tools (Microsoft Project, SAP, Up and Running) have facilities for entering new resources, but do not deal with an analysis of cognitive processes, and resource quality based on which resources can be added to the resource pool. Therefore, the existing scheduling tools cannot be effectively used for management of processes where resources are a critical variable.
- Theory/Tools in Machine Learning: Existing machine learning techniques (e.g., Boosting (Schapire, 1999), Lazy Learning (Aha, 1997), Neural Nets, Decision Tree Learning (Quinlan, 1990; Utgoff, 1989), Support Vector Machine (Vapnik, 1998), Reinforcement Learning (Sutton & Barto, 1998)), and contemporary machine learning tools (e.g., WEKA, AutoClass, mySVM) have not yet been examined in terms of agents' capability/compatibility and scheduling problems.

There is a direct relationship between the representation and the learning mechanisms. In many cases the underlying representations in machine learning have been of limited structure (e.g., vectors, trees, networks). A hybrid integration of various machine learning mechanisms for engineering of structured objects is novel and will be examined in this project in the context of the profile theory.

BENEFICIARIES

- Engineering the Complex Systems: Research in engineering of complex systems will provide insight into new methods and approaches to learning in cognitive systems. Research in machine learning will deliver adaptiveness to knowledge integration and scheduling of learning methods. Scientists in cognitive systems research will receive a formal method for modelling of cognitive processes. By developing integration metrics using the profile theory we can provide analysis, development, integration, modelling and management of complex systems and their elements where weight, time and other internal multifaceted properties are critical variables. Further development of the profile theory will establish a new branch in mathematics and extend its applications.
- Industry**:** New evaluation techniques could provide support for a solution to the resource-based problems in cognitive processes in software and IT

projects such as team formation and integration in connection with process tasks.

The application of a new approach could provide learning organisations with:

- superior management of resource capabilities and compatibilities;
- streamlining of process development through better management of project resources and tasks;
- increased opportunities for organisations to implement process improvement based on the constructive criticism derived from self analysis.

It is apparent that there is a worldwide interest in the application of this research. Since most modern processes are cognitively driven our method can be used for the formal modelling of cognitive systems. It is important for the future competitiveness of the software and IT industry to employ a scientific (vs. heuristic) approach to the engineering of cognitive processes.

- Technology: The profile-based approach assures a virtual prototyping of system development within different environment settings. An important application of this approach is that it gives the means of providing systemic methods of study, analysis, prediction, improvement, control and management of a system development. Moreover, this technology demonstrates a modelling flexibility that permits one to represent a fine-granularity of system components as well as to generate different system models of a wide diversity of system development processes. Thus, any traditional system model becomes a special case of the capability- and compatibility-based modelling framework.

Formal modelling of the capability and compatibility of cognitive systems ensures the automation in system modelling. It leads to development of new technologies in system modelling. Some of the enhancements that we intend to offer through this method are to provide support for development and engineering of new knowledge capabilities of cognitive systems, that is, innovative technologies.

REFERENCES

Aha, D. (Ed.). (1997). *Lazy learning.* Dordrecht: Kluwer Academic Publisher.

Clouse, J.A. (1996). Learning from an automated training agent. In G. Weiß & S. Sen (Eds.), *Adaptation and learning in multiagent systems.* Berlin: Springer Verlag.

Crites, R., & Barto, A. (1998). Elevator group control using multiple reinforcement learning agents. *Machine Learning, 33,* 235-262.

Jennings, N.R., Sycara, K., & Wooldridge, M.A. (1998). Roadmap of agent research and development. In N.R. Jennings, K. Sycara & M. Georgeff (Eds.), *Autonomous Agents and Multi-Agent Systems Journal, 1*(1), 7-38.Boston: Kluwer Academic Publishers.

Mitchell, T.M. (1997). *Machine learning*. Boston: WCB/McGraw-Hill.

Norman, T.J., Preece, A., Chalmers, S., Jennings, N.R., Luck, M., Dang, V.D., Nguyen, T. D., Deora, V., Shao, J., Gray, A., & Fiddian, N. (2003). CONOISE: Agent-based formation of virtual organisations. *Proceedings of the 23rd SGAI International Conference on Innovative Techniques and Applications of AI*, Cambridge, UK (pp. 353-366).

Parsons, S., Wooldridge, M., & Amgoud L. (2003). Properties and complexity of some formal inter-agent dialogues. Journal of Logic & Computation, *13*(3), 347-376.

Plekhanova, V. (1998). On project management scheduling where human resource is a critical variable. *Proceedings of the Sixth European Workshop on Software Process Technology, Lecture Notes in Computer Science Series* (pp. 116-121). London: Springer-Verlag.

Plekhanova, V. (1999a). *A capability- and compatibility-based approach to software process modelling.* Unpublished doctoral thesis. Macquarie University, Sydney, Australia and the Institute of Information Technologies and Applied Mathematics, Russian Academy of Sciences.

Plekhanova, V. (1999b). Capability and compatibility measurement in software process improvement. *Proceedings of the 2nd European Software Measurement Conference*, Amsterdam, Netherlands, Technological Institute Publications, Antwerp, Belgium (pp. 179-188).

Plekhanova, V. (2000a). Profile theory and its applications. *International Conference on Information Society on the 21st Century: Emerging Technologies and New Challenges,* The University of Aizu, Fukushima, Japan (pp. 237-240).

Plekhanova, V. (2000b). Applications of the profile theory to software engineering and knowledge engineering. *Proceedings of the Twelfth International Conference on*

Software Engineering and Knowledge Engineering, Chicago, USA (pp. 133-141).

Plekhanova, V. (2000c). On the compatibility of contemporary project management tools with software project management. *Proceedings of the 4th World Multiconference on Systemics, Cybernetics,* Orlando, Florida, USA (vol. I, pp. 71-76).

Plekhanova, V. (2002). Concurrent engineering: Cognitive systems and knowledge integration. *Proceedings of the 9th European Concurrent Engineering Conference,* Modena, Italy (pp. 26-31). A Publication of SCS Europe (Society for Computer Simulation).

Plekhanova, V. (2003). Learning systems and their engineering: A project proposal. In J. Peckham & S. Lloy (Eds.), *Practicing software engineering in the 21st century* (pp. 164-177). Hershey, PA: Idea Group Publishing.

Quinlan, J.R. (1990). Probabilistic decision trees. In In Y. Kodratoff & R.S. Michalski (Eds.), *Machine learning: An artificial intelligence approach* (vol. 3, pp. 140-152). California: Morgan Kaufmann Publishers, Inc.

Schapire, R. (1999). Theoretical views of boosting and applications. In O. Watanabe & T. Yokomori (Eds.), *Proceedings of the Tenth International Conference on Algorithmic Learning Theory* (pp. 13-25).

Sutton, R.S., & Barto, A.G. (1998). *Reinforcement learning: An introduction.* Cambridge: MIT Press.

Utgoff, P.E. (1989). Incremental induction of decision trees. *Machine Learning, 4,* 161-186.

Vapnik, V.N. (1998). *Statistical learning theory*. Chichester: Wiley.

KEY TERMS

Agent: A complex system constituting elements that are individual performers, which can be described by their interrelationships, knowledge/skill, performance and constraints factors.

Agent's Compatibility: A capability of agent to work with other agents without adaptation, adjustment and modification.

Cognitive Process:

- The performance of some composite cognitive activity.
- A set of connected series of cognitive activities intended to reach a goal. Cognitive activities can be considered as a function of their embodied experience.

Cognitive System: A complex system that learns and develops knowledge. It can be a human, a group, an organization, an agent, a computer, or some combination. It can provide computational representations of human cognitive processes to augment the cognitive capacities of human agents.

Complex System: A collection of interrelated elements organised to accomplish a specific function or a set of functions. Complexity can be considered in terms of a number of elements and/or complexity of relationships.

Machine Learning: The ability of a machine to improve its performance based on previous results.

Multi-Agent System: A set of interrelated agents that work together to perform tasks.

Legal Expert Systems in Adminstrative Organizations

Jörgen S. Svensson
University of Twente, The Netherlands

INTRODUCTION

The term *expert system* comes from the world of artificial intelligence. Originally, it comprised the idea that computer programs can be devised to solve complex problems of decision making, as well as, or even better than human experts. Although in some technical domains this ambitious goal is still valid, it is generally relaxed for applications in legal and administrative domains. Here the term *expert system* – or *knowledge-based system* – refers to a category of computer programs that use coded knowledge to help solve problems of decision making. One simple, everyday example is a computer program that helps a tax payer fill in his tax returns and informs him about the implications his answers will have in terms of the amount of tax to be paid. A second, quite different example is a system which, based on a textual summary of a case at hand, can help a legal professional in finding applicable case law.

BACKGROUND

At the core of an expert system is a so-called knowledge base, a formal model of knowledge that is used to make inferences to arrive at relevant decisions. This knowledge base can have different forms and in general a distinction is made between rule-based and case-based systems. In a rule-based expert system, reasoning is based largely on production rules, that is, hundreds or even thousands of coded rules in the form of IF-THEN statements. For example, IF a client's total capital is below • 10.000 THEN the client is exempted from paying wealth tax. By combining many such rules, very complex inferences can be made. In a case-based system, the knowledge base consists of a larger number of coded case descriptions and the system bases its inferences on automated procedures (usually of a mathematical or statistical kind) to compare the case at hand with the different cases in the knowledge base (e.g., Brüninghaus & Ashley, 2003).

In the early years of legal expert systems, it was assumed that the development of such systems was particularly useful for more complex legal reasoning tasks, in situations where a human expert would be faced with personal limitations in knowledge and skills (Bench-Capon, 1991). Nowadays, however, legal expert systems are more often regarded as important tools to control administrative processes and to enhance administrative efficiency, that is, the use of expert systems in public and private bureaucracies to process large numbers of cases in a standardized and controllable manner.

Four important types of applications of expert systems in administrative organizations can be distinguished:

- Knowledge-based information services for citizens/clients: while in the Internet era many organizations have started to put all kinds of rules and statutes on their Web sites, research shows that most people find it very hard to apply such written rules to their personal situation. In such circumstances expert system technology may help to personalize the information. Banks, for example, may add expert system modules to their Web sites to help the customer understand which of the different mortgage plans is most suitable in his particular situation (Stranieri et al., 2001).
- Knowledge-based acquisition of citizen/client information: in combination with the above application, expert systems may be used to help gather information from citizens and organizations, so that their cases can be processed more easily. An example is the Taxis system in Greece, which helps companies in submitting their VAT declarations in a correct manner (Tsiavos et al., 2002).
- Semi-automated decision making in street-level bureaucracies: in many government agencies, the application of legal statutes is the responsibility of so-called street level bureaucrats, that is, lower-level staff with at most limited legal training. Expert systems can be used to support these bureaucrats in making formally correct decisions. An example is the use of expert systems by the Australian Department of Veterans Affairs in determining individual entitlements to disability pensions (Johnson, 2000).
- Automated application of legal rules in high-volume administrations: some administrative agencies are responsible for the repetitive application of the same regulation to very large numbers of cases.

Here expert systems may be used for fully automated processing of the bulk of these cases. An example is the fully automated fining of speeding offenders in the Netherlands (Zouridis, 2000).

EVALUATION STUDIES

The use of legal expert systems in public administration has been the topic of intense debate between enthusiasts on the one hand and skeptics on the other.

When, in the 1970s and –1980s the idea of expert systems proliferated and the opportunities in the legal domain became evident, scientists and entrepreneurs from very diverse backgrounds started experimenting in this field and built the first examples of legal expert systems – to investigate and demonstrate what could be done. Evaluation studies such as that by Nieuwenhuis (1989) were carried out in carefully controlled laboratory settings and led to a heightened enthusiasm and sometimes to oversimplified claims: the suggestion that expert systems could be designed that were superior to human decision makers and that would solve all existing problems of bureaucratic administration.

This, of course, antagonized many scholars in the fields of law and public administration, who rightly criticized the experiments for their subjectivity and limited validity and who argued that there were serious limits to what machines could and should do in administrative practice. Thus, as the stories of really intelligent expert systems proliferated, this in some cases only fed existing computer *angst*: the Terminator type of vision that we were entering a future in which we would trust machines with decisions of life and death.

In recent years, however, various examples of expert systems have entered the market and the opportunities for administrative and legal scientists to do their own objective and critical research has increased dramatically. As a result of this some interesting case studies have begun to appear (e.g., Groothuis, 2004; Smith, 1994).

These new evaluation studies of real applications used in administrative practice indicate that also among legal scientists and students of public administration, opinion is shifting in favor of the use of such systems. Today, even more skeptical observers acknowledge that legal expert systems are actually being used and in some cases really provide opportunities to improve legal decision making in public administration (both in terms of legal quality and efficiency). Particularly where street level bureaucracies are expected to work with a multitude of formal rules, an expert system can clearly help improve quality and efficiency by systematizing decision making and automating the application of complex tables and calculations.

CRITICAL ISSUES OF LEGAL EXPERT SYSTEMS IN PUBLIC ADMINISTRATION

There are, however, still various practical, ethical and philosophical issues related to the broader introduction of expert systems in public administration, and the most important of these are the following.

Expert Systems Can Contain Errors

When legal expert systems are used to make serious decisions about people's lives, it is essential that they are reliable; that they indeed produce correct decisions that conform to the formal rules and the legal status quo. This, however, may be impossible to guarantee. As legal expert systems are generally used in rather complex domains with ever-changing regulations, the question is if and how these systems can be kept formally correct, reliable and up-to-date. It is clear that the issue of expert system validation requires very serious attention, not in the least because the translation of administrative regulations into formal computer rules is not always as straightforward as it may seem. In addition to this, there is still a need for adequate design methodologies and validation tools.

Garbage in is Garbage Out

Even if expert systems are formally correct, they will produce erroneous results when fed with inappropriate data. As the experiments by Nieuwenhuis (1989) showed, users of expert systems may make unintended errors when entering relevant case data, but they may also try to manipulate such systems intentionally. This incorrect use may, of course, seriously influence the conclusions reached by these systems

Expert Systems Cannot Deal with "Hard Cases"

In legal science, special attention is paid to the distinction between clear and hard cases. Simply put, a clear case is one in which there is not much discussion about what the problem is and what the decision should be. With a hard case, however, there are one or more of the following problems involved: (a) the characteristics of the case are not easily matched to existing rules, (b) the existing rules do not deliver a clear conclusion and/or, (c) application of the existing rules leads to unacceptable results.

Expert systems in principle do nothing more than apply fixed rules to case data provided by the user. They have no real understanding of the case and therefore treat all cases as clear, even if, to any human observer, it is evident that they are not (Leenes, 1999; Smith, 1994). This implies that even when a formally correct legal expert system is fed with formally correct data, it may still produce outcomes that are not applicable and need correction.

Expert Systems Require Knowledgeable and Critical Usage

Given the previous critical issues, it is clear that expert systems should be used with care, by administrative staff who have been trained in their application, who understand their possibilities and limitations and who may intervene in expert system conclusions when needed.

There is however some evidence that the use of expert systems in public administration leads to a process of deskilling among personnel. Because "the expert system knows all", street-level bureaucrats may lose their ability to critically evaluate decisions themselves and may start to depend too much on the correctness of the system. This can lead to uncritical application of the system's results and even to neglecting responsibilities for tasks the system does not support (Dijkstra, 2002; Groothuis & Svensson, 2000).

In her recent dissertation study, Groothuis (2004) argues that the co-operation between a computer system and its human user is actually the weakest link in the practical application of expert systems. In this human-computer interaction various problems may arise as a result of false assumptions and misunderstandings.

FUTURE TRENDS

In recent years there has been an increase in the application of legal expert systems in administrative organizations. This increase seems to be connected to the increasing domination of formal rules. Especially in public administration, the provision of services is bound to large numbers of formal rules and regulations, which generally evolve over time to become more and more complex. This provides an excellent two-sided argument for the introduction of expert systems (Svensson, 2002). On the one hand, regulations often become so obscure and complex that it may be hard for human decision makers to administer them correctly by hand. On the other, these formal regulations provide an excellent basis for building rule-based experts systems.

In addition to this there are several other factors that stimulate the future take-up of legal expert systems, such as:

- The increased focus in public administration on efficiency, on decision quality and on auditable decision processes;
- The increased use of the Internet as a vehicle for instant personalized advice on legal and bureaucratic issues;
- The increased availability of personal data in electronic form, which can be processed by means of legal expert systems.

CONCLUSION

Expert systems are important tools for administering complex regulations to individual cases, and administrative organizations increasingly use these systems. There are however several critical issues connected to their application, which means that the development and implementation of these systems has to be handled with care. Since some administrative decisions can really affect people's lives, legal expert systems should be developed using excellent methodologies and when finished, they should be carefully validated and evaluated (Hall et al., 2003). They should be transparent in their functioning, and before implementation, the administrative staff has to be trained in their application and to be informed about their possibilities and limitations. Also there is a need for clear independent appeal procedures for people who are confronted with the decisions made by these systems.

REFERENCES

Bench-Capon, T. (Ed.). (1991). *Knowledge-based systems and legal applications.* London: Academic Press.

Breuker, J., Leenes, R.E., & Winkels, R. (2000). *Legal knowledge and information systems.* Amsterdam: IOS-Press.

Brüninghaus, S., & Ashley, K.D. (2003). Predicting outcomes of case-based legal arguments. In *Proceedings of the 9th International Conference on Artificial Intelligence and Law, ICAIL 2003* (pp. 233-242). Edinburgh, Scotland, UK: ACM.

Dijkstra, J.J. (2000). User interaction with legal knowledge-bases systems. In J. Breuker, R.E. Leenes & R. Winkels (Eds.), *Legal knowledge and information systems* (pp. 11-22). Amsterdam: IOS-Press.

Groothuis, M.M. (2004). *Beschikken en digitaliseren. Over normering van de elektronische overheid.* Doctoral dissertation. Den Haag, Sdu.

Groothuis, M.M., & Svensson, J.S. (2000). Expert system support and juridical quality. In J. Breuker, R.E. Leenes & R. Winkels (Eds.), *Legal knowledge and information systems* (pp. 1-10). Amsterdam: IOS-Press.

Hall, M.J.J., Hall, R., & Zeleznikow, J. (2003). A process for evaluating legal knowledge-based systems based upon the context criteria contingency-guidelines framework. In *Proceedings of the 9th International Conference on Artificial Intelligence and Law, ICAIL 2003* (pp. 23-27). Edinburgh, Scotland, UK: ACM.

Johnson, P. (1999). *Knowledge management, knowledge-based systems and the transformation of government.* Conference paper. Australian Human Resources Institute and the Public Service & Merit Protection Commission.

Leenes, R.E. (1999). *Hercules of Karneades: Hard cases in recht en rechtsinformatica.* Enschede: Twente University Press.

Nieuwenhuis, M.A. (1989). *TESSEC: Een expertsysteem voor de Algemene Bijstandswet.* Deventer: Kluwer.

Smith, T. (1994). *Legal expert systems: Discussion of theoretical assumptions.* Utrecht, University of Utrecht.

Stranieri, A., Yearwood, J., & Zeleznikow, J. (2001). Tools for World Wide Web based legal decision support systems. In *Proceedings of the 8th International Conference on Artificial Intelligence and Law, ICAIL-2001* (pp. 206-214). St. Louis: ACM.

Svensson, J.S. (2002). The use of legal expert systems in administrative decision making. In Å. Grönlund (Ed.), *Electronic government: Design, applications and management*. Hershey, PA: Idea Group Publishing.

Tsiavos, P., Smithson, S., & Kotyvos, S. (2002). A path of discontinuity: The TAXIS case as a transition from e-government to e-regulation. In T. Bench-Capon, A. Dakalopulu & R. Winkels (Eds.), *Legal knowledge and information systems* (pp. 53-62). Amsterdam: IOS-Press.

Zouridis, S. (2000). *Digitale disciplinering: Over ICT, organisatie, wetgeving en het automatiseren van beschikkingen.* Delft, Eburon.

KEY TERMS

Case-Based Expert System: An expert system that uses modeled representations of previous cases and decisions to make inferences about new cases.

Clear Case: A case that experts agree can be solved in an acceptable manner by simply applying the existing legal rules to it.

Expert System: Computer program that uses coded knowledge to (help) solve problems of decision making of some complexity.

Hard Case: A case that cannot be solved by simply applying formal rules, either because: (a) the characteristics of the case are not easily matched to the formal rules, (b) the formal rules do not deliver clear conclusions and/or (c) applying the formal legal rules leads to unacceptable results.

Rule-Based Expert System: An expert system that applies a knowledge base consisting of if-then statements that can be combined to make more complex inferences.

Transparency: Systems supporting decision making in administrative organizations should be transparent to users so that it is clear to all actors involved (street-level bureaucrats, clients, legal and administrative supervisors, appeal boards) how certain conclusions have been reached.

Validation: The process of determining whether an expert system to be used in an administrative process is correct. This involves a combination of methods of which the most important are: letting domain experts verify the knowledge in the knowledge base, using automated procedures to check the system for consistency and applying the system to a larger number of test cases.

Legal Issues of Virtual Organizations

Claudia Cevenini
CIRSFID, University of Bologna, Italy

INTRODUCTION

In the present economic context, organizations, especially of small and medium dimensions, can draw a substantial advantage by collaborating and setting up flexible, temporary ICT-enabled networks.

Identifying the legal issues relevant for virtual organizations can provide a knowledge basis to regulate their activities, thus providing support for their creation and management.

BACKGROUND

The concept of virtual organization (VO) finds its origins in the United States in the early 1990s, when some authors start to give it a first theoretical outline.

Since then, a certain scientific debate has opened, and several attempts to define and concretise it have been made. The most active research sectors with respect to this appear to be business and computer science. Until recently, however, the legal research has substantially disregarded VOs, with few exceptions.

REGULATING VOS

VOs are far from being a consolidated reality; being fluid and flexible structures, they continually evolve over time and are difficult to grasp.

The starting point for their regulation is to provide a definition for the purposes of legal research: "VO's are ICT-enabled collaborations between legally independent subjects aimed at the joint provision of goods or services, where each partner contributes to specific activities. They do not aim at achieving an autonomous legal status but appear as one organization towards third parties."[1]

As a second step, a wide range of legal issues concerning them can be identified. By developing a legal taxonomy, it is possible to aggregate legal problems in major research areas. This makes it possible to focus on those issues most connected with the particular structure and nature of the VO.

The third step is to examine the identified issues in the light of the applicable legal framework at national and international level, considering the nationality of the partners and their reciprocal agreements.

A TAXONOMY OF VO-RELATED LEGAL ISSUES

Hereinafter, a synthesis of the most relevant issues is presented.

Identity and Nationality of the Virtual Organization

The VO does not embody a formal institution separate from its partners, although it may appear as a separate, autonomous entity.

National legal orders will tend to consider the VO as a structure without legal personality and, consequently, also without nationality, provided that the partners do not opt to formally adhere to a company type as foreseen by the national law of one of them.

Role of the Virtual Organization Broker

A VO can be set up and managed without the intervention of a VO broker. This, however, would imply higher coordination costs, more complex negotiations, and a slower speed of action.

The legal status of the VO broker depends upon its actual role and activities in the VO. The broker will be subject to and have to comply with the applicable legal framework set for the legal structure it has opted for in the state in which its head office is located, as well as with the state- and contract-based rules applicable to the same VO.

Virtual Organization Framework Agreement

The VO framework agreement is a set of rules aimed at governing the internal relationships between the partners of a VO.

It has to be signed before the beginning of any activity and is generally drafted with the support of the broker, who may propose business templates on the basis of which the detailed final provisions can be negotiated with the partners.

The absence of a clear agreement would possibly lead to difficulties in the management of the operation stage and, later, to possible disputes between the partners.

Contracting with Third Parties

Having no legal personality, the VO cannot directly close contracts with third parties. Therefore, if its members are to enter into contractual relationships, this will not be feasible for the VO as a separate subject; agreements can only be closed between third parties and some or all the individual members.

Once a partner—or a group of partners—has been selected, the other members can grant to it the power to act in their name and on their behalf to the purpose of closing contracts binding for all of them, or to take care of other jural acts, as it happens with mandates.

The Resolution in Disputes

The involvement in a lengthy dispute resolution procedure can cause severe economic damage or even disrupt a temporary entity like the VO. For this reason, before the final framework agreement between the partners is signed and, later, before the signing of every agreement with third parties, attention has to be placed upon developing and agreeing upon adequate dispute resolution mechanisms, which may range from legal actions before courts, to arbitration and mediation.

Liability Issues

It would be difficult to configure a liability on the VO as such, as there are no legal instruments to construct it. It thus appears more feasible to identify a liability on the individual partners. All of them may be held jointly and severally liable for the damages that can be imputed to the VO. Whenever one partner is held liable, those who have been sued without having contributed to causing the damage which is the object of the claim can resort to an internal redress.

Intellectual Property Rights

The VO activities are based on the reciprocal disclosure of relevant data and possibly on the sharing of immaterial goods. These can, in some cases, enjoy a precise legal protection, as happens with copyright, software, patents, and databases.

This applies both to the data and goods to which the individual partners are already entitled, as well as to the outcome of their collaboration. In the former case, the legitimate owner can grant to the other partners a right of economic exploitation, for example, through licensing contracts. As to the latter case, specific agreements are to be clearly set.

If the individual contributions are not to be clearly identified, it can be assumed that all partners will be entitled to the data and goods produced by the VO and of the relative rights, on the basis of a coownership.

Data Protection

All the different activities of the VO imply the processing of personal data. Within the present data protection framework, attention has to be placed, in particular, on a series of elements, such as the processing through automatic means, the disclosure of data to third parties, and the transfer of data between European Union (EU) and non-EU countries. The VO partners will actually carry out most processing with ICT tools and may need to process data originally collected by one or more of them and perform cross-border data transfers.

Competition Law

Should a VO achieve a substantial dimension in terms of its overall turnover, attention shall be placed by the partners to its compliance with rules on antitrust and the protection of competition applicable to the partners. Specific procedures may be imposed in order to get the authorization of antitrust bodies, as well as to verify the law-abidingness of the collaboration.

ICT-Related Issues

The nature of ICT-enabled entities possessed by VOs requires their compliance with the applicable rules on the use of specific technology tools, for example, with reference to security, electronic signatures, e-commerce, or teleworking. Should certain ICT-based interactions not be specifically regulated, reference shall be made to analogically applicable norms.

FUTURE TRENDS

The present economic and legal scenario does not point toward the drafting of an ad hoc legislation for VOs. However, the growing relevance of collaborative entities, such as industrial districts, coupled with the strong support of the techno-legal scientific community, will make VOs known to a wider audience and stimulate targeted initiatives.

CONCLUSION

A coherent and certain legal framework applicable to regulate VOs is presently still absent, and a clear qualification of their legal identity by the legislator is missing. Besides, their international character would require a level of harmonisation among contrasting rules that does not always appear easy to achieve, while their massive use of rapidly evolving information and communication technologies contrasts with the slow law-making process.

This makes it extremely difficult to envisage a state-based regulation for them, or the drafting of international treaties or agreements, at least in the short term.

This enhances the fundamental role of normative tools, such as codes of conduct and best practices, and especially of contractual agreements and intraorganisational rules drafted by the VO partners, developed on the basis of a clear identification of the relevant legal issues.

REFERENCES

Berwanger, E. (1999). The legal classification of virtual corporation according to German law. In Seiber, P., Griese, J. (Eds.), *Proceedings of the 2nd VoNet Workshop*, Simowa Verlag, Bern (pp. 158-170).

Cevenini, C. (2003). *Virtual enterprises: Legal issues of online collaberation between undertakings*. Milan: Giuffre, 79-80.

Conaway Stilson, A. E. (1997). The agile virtual corporation. *Delaware Journal of Corporate Law*, 22, 497.

Cousy, H., Van Schoubroeck, C., Droshout, D., & Windey B. (2001). *Virtual enterprise legal issues taxonomy*. Public deliverable D 03, ALIVE working group on Advanced Legal Issues in Virtual Enterprise.

Davidow, W. H., Malone, M. S. (1992). *The virtual corporation—Structuring and revitalizing the corporation for the 21st century*. New York: Harper.

Scholz, C. (1994). Virtuelle unternehmen—Faszination mit Rechtlichen Folgen. *jur-pc Zeitschrift*, (12), 2927-2935.

Sommerlad, K. W. (1996). Virtuelle Unternehmen—Juristisches Niemandsland*? Office Management*, (7–8), 22-23.

Van Schoubroeck, C., Cousy, H., Droshout, D. and Windey, B. (2001). *Virtual enterprise legal issue taxonomy*. In B. Stanford-Smith & E. Chiozza (Eds.), *E-work and e-commerce*. Novel solutions and practices for a global networked economy, I, IOS Press, Amsterdam, 609-615.

KEY TERMS

Arbitration: A private form of conflict resolution. The litigating parties may voluntarily submit a dispute to one or more independent, neutral experts (arbitrators), who decide upon the case similarly to a court, generally upon a shorter period of time. Arbitrations are usually regulated by law.

Intellectual Property Rights: Copyright and connected rights that include, *inter alia*, the right of copying, modifying, and distributing the protected work.

Legal Order: The set of legal norms that make up the legal system of a particular country.

Legal Personality: The capacity of a legal person (e.g., a corporation) to be holder of rights and duties.

Mediation: An alternative dispute resolution method. The litigating parties may voluntarily submit a dispute to a neutral, independent mediator. This latter does not issue a decision but supports the parties in finding a mutually agreed upon solution.

Personal Data: Any information concerning a natural or legal person that can identify it.

Virtual Organization: An ICT-enabled collaboration between legally independent subjects aimed at the joint provision of goods or services, where each partner contributes to specific activities. It does not aim at achieving an autonomous legal status but appears as one organization toward third parties.

VO Broker: A subject who acts as an intermediary for the setting up of the virtual organization by identifying a possible business opportunity, contacting the potential partners, and proposing agreement templates.

ENDNOTE

1 Cevenini, C. (2003). Virtual Enterprises. Legal Issues of the Online Collaberation between Undertakings. Milan, Guiffre, 79-80

Liability for System and Data Quality

Robert D. Sprague
University of Wyoming, USA

INTRODUCTION

Accompanying the proliferation of computers in almost every facet of life is an underlying risk to financial well-being related to computer system and data quality. The viability of a business often depends upon the continual and reliable operation of its computer system. The consequences of low-quality computer hardware and software are not infrequent or insubstantial. A recent report from the U.S. Department of Commerce states that total U.S. software sales in 2000 were approximately $180 billion (RTI, 2002). The report estimated that the lack of an adequate software testing infrastructure costs U.S. software users over $38 billion per year, principally through error avoidance and mitigation activities.

In just about every instance, a computer system is acquired, software is used, or data are accessed through a contract-based commercial transaction. Therefore, whether a computer, software, or data vendor will be held liable for defects or errors depends on the language of the contract itself and the law of contracts. Most contracts impose scant liability for vendors. Although aggrieved users have pursued other avenues of relief through non-contract-based legal theories, they have met with little success. For computer, software, and data users, the Latin maxim of *caveat emptor* still applies—let the buyer beware.

BACKGROUND

In the United States, commercial transactions are governed by contract law; a collection of rules of law which provides a basis for predicting whether a court will permit a certain contract provision to be enforced. Most courts and legal commentators now agree that in transactions in which an ultimate software product is to be delivered, the software is classified as a "good" and the associated contracts are governed by the Uniform Commercial Code (UCC or "Code") (Horovitz, 1985; Rodau, 1986). Article 2 of the UCC applies to contracts for the sale of goods. Technically, most software is acquired through a license rather than a sale. As far as the courts are concerned, however, software licenses are treated as ordinary contracts accompanying the sale of products, and therefore are governed by the UCC (ProCD, 1996).

Aggrieved computer and software users, and particularly users of inaccurate information, have sought redress through non-contract-based theories based on a separate body of law known as torts. However, without being able to show some form of injury to a person or destruction of physical property, these claims generally fail.

SOFTWARE CONTRACTS

Application of the UCC to software contracts means that the Code's provisions relating to warranties, consequential damages, and disclaimers of liability will also apply. Warranties can provide legal protection to the buyer regarding the quality of the goods (such as computer hardware or software) that are the subject of the contract. The UCC, however, also allows the seller to limit or exclude any or all of the warranties. When a vendor does provide a limited warranty—for example, that a hard disk drive will function properly for one year after purchase or that a software program will carry out its basic instructions with minimal errors—it is usually accompanied by a limitation of remedies. By contract, the purchaser agrees that in the event the product does not live up to its represented quality, the purchaser's remedies (and, hence, the vendor's liability) will be limited by the terms of the contract's limitation of remedies clause—usually a repair or replace option for defective hardware or a refund of the purchase price of software.

Often a computer hardware or software customer suffers damages beyond the cost of the hardware or software if there is a problem. The customer's business may be effectively shut down, or severely curtailed, if the computer system is not functioning properly. This type of damages is known as consequential damages—occurring indirectly as a result of the problem with the product. The UCC provides that consequential damages may be limited or excluded unless the limitation or exclusion is unconscionable. Limitation of consequential damages for injury to the person in the case of consumer goods is, on its face, unconscionable, but a limitation of damages where the loss is commercial is not.

The exclusion of consequential damages in commercial transactions under the UCC is considered merely an allocation of unknown or indeterminable risks and is almost always upheld. It has become standard practice in the computer industry, particularly for software publishers, to provide very limited express warranties, disclaim all other warranties, and severely limit remedies in the event of a breach of contract. These disclaimers and limitations of remedies are invariably enforced by the courts.

The initial premise under which disclaimers and limitations are allowed is that the parties negotiate the terms of the underlying agreement. In today's software transaction, however, "bargaining," at least on the part of the buyer, often consists of clicking on an "I Agree" button during software installation. Most software is acquired through a shrinkwrap (or clickwrap or browsewrap) agreement. Despite the lack of bargaining power on the part of the purchasers, in nearly all cases, shrinkwrap agreements have been held enforceable. The result is that a vast amount of software is acquired through a form contract with no opportunity to bargain for warranties or remedies.

The historic context of contracts was that the parties negotiated the terms of the agreement. In particular, it was believed that the purchaser was willing to assume some risk that the acquired product may not meet all expectations—through limited warranties and remedies—in exchange for a lower price. In the modern computer-related contracting scenario, however, the purchaser assumes all the risk. Despite theoretical underpinnings surrounding the creation of the UCC to provide buyers and sellers equal protections (Alces, 1999), the reality is that modern contracting practices provide no protection against the substantial losses businesses may encounter, and have encountered, when they cannot operate their business because data are lost or the computer system fails to operate properly.

TORT-BASED FORMS OF LIABILITY

Since software is generally considered a product for contracting purposes, it is logical to consider whether a vendor would be held liable for damages resulting from defective hardware or software under the doctrine of products liability. Products liability law protects those who suffer injuries as a result of a defective product. The law imposes upon a vendor strict liability for placing a defective product in the stream of commerce. This liability applies regardless of the amount of care exercised by the vendor in the preparation and sale of the product.

However, the vendor of a defective product is only strictly liable for certain types of damages suffered by the product's users: personal injury or damage to other physical property. To date, there has been no reported successful products liability lawsuit regarding defective computer hardware or software. All courts that have directly addressed the issue of whether products liability applies to defective computer hardware or software have ruled against application on the basis that the damages sustained are categorized as *economic loss*—a remedy not available under products liability law.

The "economic loss rule" provides that where no person or other property is damaged, the resulting loss is purely economic (East River, 1986). A few computer purchasers have argued that data lost due to defective hardware or a defective computer system constitutes damage to "other property." The courts that have considered this argument have uniformly rejected it. The courts consider the data as integrated into the computer system (Transport Corporation of America, 1994).

Even when the software vendor knows of a defect in the software and fails to notify the user, if the damages suffered are limited to economic losses, a tort claim of negligence will also fail (Hou-Tex, 2000). The economic loss rule provides a substantial barrier to any recovery for defective computer hardware or software, regardless of whether the purchaser pursues a claim for products liability or negligence.

LIABILITY FOR PUBLISHING INACCURATE DATA

The Internet has revolutionized the mass delivery of information. However, information was delivered online long before the commercialization of the Internet. A critical dimension of online information delivery is the speed at which information is disseminated, not necessarily the extent of that dissemination. Due to the speed of dissemination, however, information delivered online is not always accurate. The legal issue that arises is whether someone who claims damages as a result of an inaccuracy can hold the electronic publisher of that information liable for those damages. As a general rule, they cannot.

Historically, the courts have generally not held publishers liable for inaccuracies in the information they have published. There is an almost absolute immunity for publishers of information they do not author (Birmingham, 1992). Where the publisher does not author or guarantee the contents of the publication, it has no duty to investigate and warn its readers of the accuracy of the contents of its publications.

Courts have, however, carved out an exception when the information involved is highly technical in nature. For example, some courts have held a publisher liable for injuries suffered as a result of inaccurate aeronautical charts, considering the charts highly technical tools (Saloomey, 1983). Likewise, one legal commentator has argued that producers of Geographic Information System (GIS) mapping and database analysis may be under a similar threat of liability for inaccuracies (Phillips, 1999). However, as with "traditional" products liability and negligence laws, without a physical injury or damage to other physical property, a publisher will not be held liable for inaccurate information, regardless of the type of information published.

The laws relating to liability for inaccurate information are somewhat different when the publisher is also the author of the information. An author who provides false information in the course of its business as part of a commercial transaction is liable for loss suffered by the recipient's justifiable reliance upon the information, if the author fails to exercise reasonable care or competence in obtaining or communicating the information (Restatement, 1977).

However, before a publisher/author of inaccurate information can be held liable for damages resulting from that inaccurate publication, a special relationship must exist between the publisher and reader—the publisher must owe a specific duty to the reader. The fact that the subscriber obtains the information through an online subscription does not, in itself, create a special relationship.

FUTURE TRENDS

Time and again, courts recognize that the parties to a commercial transaction allocate risk, even when there has been no negotiation and the parties do not enjoy equal bargaining power. The result is that all of the risk for defective computer hardware/software or inaccurate data has been shifted to the purchaser. As long as the purchaser bears all the costs of these defects or inaccuracies, the vendor has no incentive to increase or guarantee quality. If the law was to impose more risk upon vendors, would quality improve?

Microsoft has been the subject of much criticism for producing shoddy software, particularly software with blatant security vulnerabilities (Menn, 2002). The issue of defective software has become so severe that commentators (Computer Science and Telecommunications Board, 2002; Sager & Greene, 2002) and legislators (Menn, 2002) have called for new standards of liability for software publishers. In early 2002, Microsoft Chairman Bill Gates published a company-wide memo outlining a strategic direction for Microsoft to produce software that is available, reliable, and secure (Schneier & Shostack, 2002). In the memo, Bill Gates is quoted as saying, "We must lead the industry to a whole new level of Trustworthiness in computing." Bill Gates is also quoted in the memo as saying, "So now, when we face a choice between adding features and resolving security issues, we need to choose security."

If Bill Gates and Microsoft are reacting to the threat of additional liability for the lack of quality in their software, this supports the contention that computer hardware, software, and data vendors, if faced with the potential of being held liable for the lack of quality in their products, will begin to assume greater responsibility for minimizing defects in hardware/software or inaccuracies in data.

The pressure continues. In late 2003, a class-action lawsuit was filed in California against Microsoft, claiming that Microsoft should bear some of the financial responsibility for its software's vulnerability to computer viruses (Krolicki & Stevenson, 2003).

CONCLUSION

The law allows computer system and data vendors, by contract, to disclaim any guarantee of quality, as well as prohibit any recovery for damages suffered by the purchaser where there is a lack of quality. The courts have consistently upheld such contract provisions. In addition, the type of damages usually suffered by disappointed purchasers or users are economic in nature—exactly the type of damages that are not recoverable in a tort action. The result is that between the contract and the courts, purchasers and users who suffer damages as a result of defective computer hardware/software or inaccurate data simply have no meaningful recourse. *Caveat emptor!*

REFERENCES

Alces, P. (1999). W(h)ither warranty: The b(l)oom of products liability theory in cases of deficient software design. *California Law Review, 87*, 269-304.

Birmingham v. Fodor's Travel Publications, Inc., et al., 833 P.2d 70. (Haw. 1992).

Computer Science and Telecommunications Board. (2002). *Cybersecurity today and tomorrow: Pay now or pay later.* [Prepublication Edition]. Retrieved January 16, 2002, from *books.nap.edu/html/cybersecurity*

East River S. S. Corp. v. Transamerica Delaval, 476 U.S. 858. (1986).

Horovitz, B. (1985). Computer software as a good under the Uniform Commercial Code: Taking a byte out of the intangibility myth. *Boston University Law Review, 65*, 129-164.

Hou-Tex, Inc. v. Landmark Graphics, 26 S.W.3d 103. (Tex. Ct. App. 2000).

Krolicki, K., & Stevenson, R. (2003). Microsoft faces class action over virus crashes. *Reuters*, (October 2). Retrieved October 3, 2003, from *biz.yahoo.com/rc/031002/tech_microsoft_security_3.html*

Menn, J. (2002, January 14). Security flaws may be pitfall for Microsoft. *Los Angeles Times*. Retrieved January 14, 2002, from *www.latimes.com/business/la-000003463jan14.story?coll=la-headlines-business-manual*

Phillips, J. (1999). Information liability: The possible chilling effect of tort claims against producers of Geographic Information Systems data. *Florida State University Law Review, 26*, 743-777.

ProCD, Inc. v. Zeidenberg, 86 F.3d 1447. (7th Cir. 1996).

Restatement (Second) of Torts. (1977). § 552.

Rodau, A. (1986). Computer software: Does Article 2 of the Uniform Commercial Code apply? *Emory Law Journal, 35*, 853-920.

RTI. (2002, May). *The economic impacts of inadequate infrastructure for software testing*. U.S. Department of Commerce, National Institute of Standards & Technology. Retrieved June 18, 2002, from *www.nist.gov/director/prog-ofc/report02-3.pdf*

Sager, I., & Greene, J. (2002, March 18). The best way to make software secure: Liability. *BusinessWeek*, 61.

Saloomey v. Jeppesen & Co., et al., 707 F.2d 671. (2d Cir. 1983).

Schneier, B., & Shostack, A. (2002, January 24). Results, not resolutions: A guide to judging Microsoft's security progress. *SecurityFocus Online*. Retrieved January 22, 2002, from *online.securityfocus.com/news/315*

Transport Corporation of America, Inc. v. International Business Machines, Inc. and Innovative Computing Corporation, 30 F.3d 953. (8th Cir. 1994).

KEY TERMS

Browsewrap Agreement: Generally pertains to accessing information on a Web page. A notice is placed on the Web site informing the user that continued use of (browsing) the Web site constitutes acceptance of a license agreement (the terms of which are usually made available by the user selecting a link on the Web site).

Clickwrap Agreement: Applies to software acquired without any packaging (e.g., when it is copied to a computer ("downloaded") from a Web site or is pre-loaded on a computer). When the buyer installs the software, a dialog box is displayed containing the license agreement. The user is instructed to select a button to accept the terms of the license agreement and complete the installation.

Negligence: A tort theory that applies to conduct that falls below the standard established by law for the protection of others against unreasonable risk of harm. An action for negligence must establish that: (1) one party (the first party) owed a duty of care to a second party; (2) the first party's lack of care (negligence) breached that duty of care; and (3) the first party's breach of the duty of care proximately caused the damages suffered by the second party.

Shrinkwrap Agreement: Derives its name from software contained in a box wrapped in plastic (shrinkwrapped). The agreement pertaining to the software is either printed on the box and readable through the plastic, or there is a notice that the agreement is contained within the box. In either event, opening the box and installing and using the enclosed software is considered acceptance of the terms of the agreement. The shrinkwrap agreement has evolved into the "clickwrap" agreement, as well as the "browsewrap" agreement.

Tort: A civil wrong committed by another. Tort law recognizes that individuals or businesses may owe a duty of care to others with whom they have contact. If that duty of care is breached, and the breach results in damages, the party that owed the duty of care may be liable for the resulting damages.

Unconscionability: Generally involves a lack of meaningful choice on the part of one of the parties to a contract together with terms that are unreasonably favorable to the other party. Unconscionability also refers to a contract that is so unfair or one-sided that a judge cannot in good conscious enforce the contract.

Uniform Commercial Code (UCC, or "Code"): A uniform collection of laws relating to commercial transactions. Specific provisions of the Code which relate to the sale of goods (Article 2) have been enacted by all of the states.

Warranty: A representation by a vendor regarding the quality of the product.

Library Management and Organizational Change

Ardis Hanson
University of South Florida, USA

Phyllis Ruscella
University of South Florida , USA

Kathy Arsenault
University of South Florida St. Petersburg, USA

Joan Pelland
New College of Florida, USA

Derrie Perez
University of South Florida, USA

Beverly Shattuck
Hinks and Elaine Shimberg Health Sciences Center Library, USA

INTRODUCTION

As academic libraries continue to develop a virtual presence, they must cope with the rapidly evolving rate of change and respond proactively to their changing environment in order to take advantage of the opportunities for increasing their visibility, restructuring to meet the needs of their users, and achieving their objective of remaining the preeminent source of information within the academy.

COMPONENTS OF ORGANIZATIONAL CHANGE

The literature reflects four major areas in organizational change: the *structural* (or planning) aspect of change, the *cultural* aspect of change, the *personal* (or *individual human*) reaction to change, and the *political* aspect of change. Lewin's (1958) fundamental description of structural change has been incorporated into many process-oriented models of organizational change developed for organizations to better understand and direct the process of systemic change (Schein, 1987). Since cultural change often affects organizational identity, managers should focus on communication, leadership, and emotional components of change (Caldwell, 2003). Personal transition stages of individuals in the change process include the release of individual identity, ambiguity, and establishing a new beginning (Miller, 2002). The political aspect of change is reflected by Bolman and Deal (2003) in the use of reframing through the use of organizational metaphors as one of the skills of the "new" manager.

MANAGING CHANGE

Effective and successful organizational change incorporates and manages perspectives and models concurrently. As managers develop infrastructures (processes and values, and organizational capabilities) that are more effective (Christensen & Overdorf, 2000) or change institutional identity to assist in the creation of processes and values (Newman & Chaharbagi, 2000), it is important to understand why change does not happen. Mutually reinforcing barriers to implementing change are: top-down or *laissez-faire* management style; unclear strategy and conflicting priorities; an ineffective senior management team; poor vertical communication; poor coordination across functions; and inadequate down-the-line leadership skills and development (Beer & Eisenstat, 2000).

Although libraries of colleges and universities are changing faster than their respective parent institutions, library administrators must still address patron needs, provide services and handle service provision networks, initiate collaborative arrangements, improve staff skills and abilities, and enhance the image of the library (how the organization is regarded by important oversight bodies), and cope with new ways of funding and performing services (Spies, 2000).

Implementing Large Scale Programmatic Change

Innovation, defined as new ways of thinking of, generating, and coping with change (Jarratt, 1999), often means dramatic organizational change, such as cross-functional teams and the creation of new functional units and programmatic areas (Bishop, 1999). Implementing these innovations involves managing both the strategies and the elements of the organization that will have to be changed to enable the organization to anticipate, respond to, and shape future challenges (Morton, Brookes, Smart, Backhouse,& Burns, 2004). Even short-term change initiatives that focus on costs and/or changing established working practices have immediate and inevitable impact on the organization may be highly traumatic for staff (Hailey, 1998). Therefore, one needs to "marshal knowledge about changes, organizations, and corporate behavior—so that corrective actions may be undertaken to bring back balance and relative stability" (Geisler,1997, p.4) .

The judicious use of human resource interventions, the maintenance of organizational identity, and the supportive actions of its line managers help staff through the process of change (Hailey, 1998). More important than a manager's commitment to change is the commitment of line managers to people management. According to Hailey (1998), commitment to the management of people by supervisors ensures that staff members are counseled on a regular basis, both formally and informally; that their personal career development is discussed (with or without vertical career opportunities); and that they receive regular feedback on their performance. If these things are already in place, department heads and supervisors can facilitate change within their departments or units. Middle managers, who see themselves as change agents, are critical to encouraging adaptive change by staff. By doing "real" work themselves, demonstrating optimum skills, and encouraging and assisting staff to do their best possible work (Lamsa & Savolainen, 2000), managers can focus on a few key measures in critical areas and promote the belief that they are accountable for their work.

Performance Measurement

Performance measurement systems can create an essential feedback and learning mechanism in support of key management decisions, especially when criteria measures institutionally focused performance. A successful system also functions as a communication and information system, particularly for senior staff and administration. Basically, performance assessment should allow: up-to-date job descriptions that will ensure staff membersknow what is expected of them; continuing communication between supervisors and staff; recognition of staff for doing well; and staff development processes (Lubans, 1999). Effective people management assists line managers in handling change management.

What are core competencies in existing and potential staff as multiple assignments and opportunities emerge within the library? The literature on change management emphasizes three core competencies: a demonstration of interpersonal competence, personal integrity, and the capacity to think systemically and in an integrated way about how work systems and people need to collaborate. If these skills are present (or can be developed) within existing staff, staff may obtain a variety of professional competencies for managing individual and team projects, as well as system-level initiatives.

Managing Structural Change

Marketing, business, and annual operational plans, with formal objectives for individual staff, follow strategic planning. Wide variations in management practice indicate the need for major improvements, particularly in terms of adopting a strategic approach to the planning and delivery of library and information services. A clear framework of strategic objectives and priorities, formulated through a participative planning process, facilitates delegation of decisionmaking and resource allocation, allowing quick, flexible responses to identified customer needs. As libraries become more "virtual." academic library administrators must determine if the existing management and structure is both responsive to the changing user needs and utilizing technology to its best advantage (Spies, 2000).

Managing Cultural Change

To manage change successfully, library directors must choose the appropriate change path and design its implementation to suit their own situation. This requires an understanding of key internal organizational features, such as staff identity, aspects of the organization they wish to preserve, the degree to which the organization as a whole is aware of the need to change, and the level of capability for change possessed at all levels (Oliver& Roos, 2003). The culture shifts created during virtual library projects are similar to the changes libraries encountered during the mid-1980s when they purchased library management systems to automate circulation, acquisitions, and cataloging functions. At that time, library administration assured its faculty and staff of their continued importance and value within the then new organizational structure and to see innovation as a positive aspect of change.

Managing and Motivating Personal Change (Human Resources)

From a management perspective, the establishment of the team process can be an overall success. With autonomy within the teams and empowerment to tackle the team's charges, and, if necessary, to alter or change them to fit the parameters of the work, team members experience better communication and satisfaction. Further, teams that include both paraprofessional and professional staff lessen the artificial "class" structure often found in academic libraries.

Managing the Political Aspect of Change

From the perspective of the political, change processes are considered to be predominantly bargaining, consciousnessraising, persuasion, influence and power, and social movements (Bolman & Deal, 1991). The political aspect of change requires the reframing of organizations, which is impacted by the structural, cultural, and personal components of change. Combining models can create points of convergence, thereby creating a common ground on which to successfully implement and move through a change process. The distinctive characteristics and contexts of higher education require administrators to creatively balance internal (library) and external (university) forces and to often serendipitously leverage change (Nozero & Vaughan, 2000).

Quality Control

A strategic plan represents how the values, purpose, and operating principles in an organization are connected to its vision and strategy. Strategic objectives must be tied to the everyday operating environment and be measured through well-reasoned, logical performance criteria. (APQC, 1999). Sea changes in a strategic plan often dramatically impact technology-dependent library projects; yet, smaller issues also impact teams and individual work assignments. The library administrator must monitor the place of the team within the organization as the goals and structure of the organization change. To stay aligned, teams need to talk to one another and to the organization. By setting a clear direction, the organization also sets the boundaries within which teams work (Forrester & Drexler, 1999). Furthermore, teams require both individual and team accountability.

Sometimes, leaders of change also must be managers of change. Building job assignments and evaluation capacities into team positions is difficult and requires the use of concrete performance measures. However, without the accountability for staff time (within both regular and team assignments), organizations cannot have an accurate picture of the time, staff, and effort actually involved in projects or other aspects of organizational change. This is particularly true in the case of "volunteer" projects, which tend not to build in or use the necessary administrative tools to measure work performed outside one's normal job duties. Without mechanisms in place to ensure performance or delivery of product, it is extremely difficult, if not impossible, to track actual hours spent, review concrete workflow processes, or have product created. Further, it also makes it difficult to place "volunteer" tasks in relation to ongoing library duties, e.g., supervisors and department heads may feel that "volunteer" library work infringes upon actual duties.

Recommendations

There are six major lessons that administrators need to learn when working with organizational redesign or with team-based organizations. First, an organization's clarity of vision is critical and is subject to change based upon external forces, such as changes in the mission, vision, and values of the university. As the institution's priorities evolve, it is critical that administrators reevaluate and update a working document to ensure mission congruence.

Second, management should not underestimate the power of personal relationships. One of the most important aspects of inter-organizational networking is creating and sustaining the personal relationships among the parties (Lamsa & Savolainen, 2000). For a team to be effective, high-trust mutual relationships need to be developed, i.e., each member believes the others to be honest, capable, and committed to shared goals.

Third, do not let a team take itself too seriously; an innovation is not an ideology. Ironically, those who initially position themselves as change agents may aggressively resist the eventual integration of their project.

Fourth, the rest of the organization should not be ignored. It is important to avoid resentment by making sure that other groups are equitably considered for rewards and recognition, as well as challenged by meaningful assignments. Edwards and Walton (2000) indicate that a number of factors (including perception, limited resources, departmentalization and specialization, nature of work activities, role conflict, inequitable treatment, violation of territory, and environmental change) are major sources of conflict in academic libraries.

Fifth, do not let the rest of the library ignore the team. One of the "meaningful challenges" should be achieving the skills to fulfill the team's mission, e.g., all collection development librarians should work with virtual resources, all catalogers with metadata, and all reference librarians with innovative services and bibliographic instruction.

Finally, do not leave department/unit managers out of the loop, and make sure that goals are clear and clearly evaluated. Without measurable goals, there can be no team. An organization's precision and accuracy in marking progress should be clearly communicated to managers, staff, and teams. With clear communication of expected goals and outcomes, all members of an organization can focus on accountability, evaluating how well goals are achieved, and by specifying exactly who is responsible for what.

CONCLUSION

There are potential advantages in organizational restructuring to achieve more effective collaboration in planning and delivering information services by libraries. The hierarchical, "top-down" management style of the past is rapidly giving way to a system where employees take responsibility for their own actions and leadership comes from employee teams (Pierce & Kleiner, 2000). It does not mean that leadership always makes decisions at the lower levels, but rather that they oversee the decisions that are made and evaluate their congruence with the direction of the organization. This has a direct effect on the organizational composition. Flatter and more flexible structures are emerging, moving away from traditional structures to multi-skilled, multi-tasked, and cross-organizational teams that more effectively tailor services and resources to particular patron requirements. Further, a strong leadership emphasis on team goals, clear expectations from team leaders, attention to team development, and an emphasis on coaching and challenging (rather than directing) are critical for successful change management when utilizing a team-based structure.

FUTURE ISSUES

In the rapidly evolving environment of academic libraries, Collier and Esteban (2000) see library leadership as being the systematic capability diffused throughout the organization to encourage creativity, and generating processes and practices that translate into organizational learning. Administrators exercise this through influence and intention, openness and communication, and autonomy and accountability. Successful leadership of libraries requires the capacity to see change as a legitimate and positive aspect for organizational learning and development.

However, change requires a more thoughtful approach; major projects can and will have a significant impact on the organizational culture and structure. Nearly two decades ago, Soudek (1983) formally defined the relationship between the organizational climate and professional behavior of academic librarians. A good organizational climate is high in individual autonomy, low in job structure, high in reward and recognition of achievement (personal or organizational), and high in consideration, warmth, and support (Soudek, p. 337). Successful organizational change should include these measures as outcomes of the change process.

Almost a decade ago, Pritchard (1995) emphasized that as the roles for librarians change to accommodate new technology and user needs, it is important to remember that these challenges are evolving out of those roles that librarians have always done: connecting users to information; designing and managing complex, interconnected systems and organizations; selecting and organizing information resources; teaching and consulting; creating logical and intuitive insights about information; and formulating and articulating information policy. These basic services are still current today.

Finally, the key tasks for academic library administrators are to articulate future directions, based on a vision shared by all stakeholders and informed by ongoing environmental appraisal; to secure the financial and other resources required to achieve agreed-upon goals; and to inspire and support colleagues as partners in exciting collaborative ventures (Smith, 2003). The information environment of the 21st Century offers libraries the opportunity to play a central role in the academic community, but it will require bold and confident leadership along the way.

REFERENCES

American Productivity & Quality Center (APQC) (1999). Strategic planning: What works ... and what doesn't: Presentations from APQC's third knowledge management symposium: *Knowledge Management: Lessons from the Leading Edge*, Williamsburg, VA (1998, October 19-23). Retrieved December 12, 2002 from *http://www.apqc.org/free/whitepapers/ dispWhitePaper.cfm?ProductID=672*

Beer, M. & Eisenstat, R.A. (2000). The silent killers of strategy implementation and learning. *Sloan Management Review, 41*(4):29-40.

Bishop, S.K. (1999). Cross-functional project teams in functionally aligned organizations. *Project Management Journal, 30*(3):6-12.

Bolman, L.G., & Deal, T.E. (1991). *Reframing organizations: Artistry, choice and leadership*. San Francisco: Jossey-Bass.

Bolman, L.G. & Deal, T.E. (2003). *Reframing organizations: Artistry, choice and leadership* (3rd ed.). San Francisco, CA: Jossey-Bass.

Caldwell, R. (2003). Models of change agency: A fourfold classification. *British Journal of Management, 14*(2): 131-142.

Christensen, C.M. & Overdorf, M. (2000). Meeting the challenge of disruptive change. *Harvard Business Review, 78*(2): 67-76.

Collier, J. & Esteban, R. (2000). Systematic leadership: Ethical and effective. *Leadership & Organizational Development Journal, 21*(4): 207-215.

Edwards, C. & Walton, G. (2000). Change and conflict in the academic library. *Library Management, 21*(1): 35-41.

Forrester, R. & Drexler, A. B. (1999). A model for team-based organizational performance. (Themes: Teams and New Product Development). *The Academy of Management Executive, 13*(3): 36-50.

Geisler, E. (1997). *Managing the aftermath of radical corporate change*. Westport, CT: Quorum.

Hailey, V.H. (1998). Transforming your organisation through people management. *Credit Control, 19*(8): 25-33.

Jarratt, A. (1999). Managing diversity and innovation in a complex organization. *International Journal of Technology Management, 17*(1,2):5-16.

Lamsa, A.-M. & Savolainen, T. (2000). The nature of managerial commitment to strategic change. *Leadership and Organizational Development Journal, 21(2)*: 297-306.

Lewin, K. (1958). Group decision and social change. In E.E. Maccoby, T.M. Newcomb, & E.L. Hartley (Eds.), *Readings in social psychology*, (pp. 197-211). New York: Holt, Rinehart & Winston.Lubans, J. (1999). “I’ve closed my eyes to the cold hard truth I’m seeing” - Making performance appraisal work. *Library Administration and Management, 13*(2): 87-99.

Miller, D. (2002). Successful change leaders: What makes them? What do they do that is different? *Journal of Change Management, 2*(4): 359-368.

Morton, S.C, Brookes, N. J., Smart, P.K., Backhouse, C. J., & Burns, N.D. (2004). Managing the informal organization: Conceptual model. *International Journal of Productivity and Performance Management, 53*(3): 214-232.

Newman, V. & Chaharbagi, K. (2000). The study and practice of leadership. *Journal of Knowledge Management, 4*(1): 64-74.

Nozero, V.A. & Vaughan, J. (2000). Utilization of process improvement to manage change in an academic library. *Journal of Academic Librarianship, 26*(6): 416-421.

Oliver, D. & Roos, J. (2003). Dealing with the unexpected: Critical incidents in the LEGO Mindstorms team. *Human Relations, 56*(9): 1057-1082.

Pierce, T.N. & Kleiner, B.H. (2000). Changes affecting leadership and its importance in organizations. *Management Research News, 23*(7-8): 5-9.

Pritchard, S. M. (1995). Today’s librarians: The ultimate moveable type. *Moveable Type: The Newsletter of the Mark O. Hatfield Library, Willamette University, 2*(2), 3-4. [Electronic Resource]. Retrieved October 12, 2003, from http://library.willamette.edu/publications/movtyp/spring1995/sarah.

Schein, E.H. (1987). *Process consultation volume 2: Lessons for managers and consultants*. Reading, MA: Addison-Wesley.

Smith, A. (2003). New-model scholarship: How will it survive? Washington, DC: Council on Library and Information Resources. [Electronic Resource]. Retrieved August 8, 2003, from: *http://www.clir.org/pubs/reports/pub114/pub114.pdf*.

Soudek, M. (1983). Organizational climate and professional behavior of academic librarians. *Journal of Academic Librarianship, 8*(6): 334-338.

Spies, P. B. (2000). Libraries, leadership, and the future. *Library Management, 21*(3): 123-127.

KEY TERMS

Change: To give a different position, course, or direction to an activity, process, or business.

Change Management: Successfully implementing a new position, course, or direction for individuals, processes, or products.

Performance Assessment: The measurement of an activity, process, or product to see if it meets an established objective or benchmark.

Trust Relationship: A relationship in which each party is assured of the reliability of the character, ability, strength, or truth of someone.

Vision: Discernment or foresight to create a path toward excellence, often used in regard to a future state of desired activity, production, or being for individuals, as well as for corporate entities.

Life Cycle of ERP Systems

Cesar Alexandre de Souza
University São Judas Tadeu (USJT), Brazil

Ronaldo Zwicker
University of São Paulo (USP), Brazil

INTRODUCTION

The 90's witnessed an impressive growth of Enterprise Resource Planning (ERP) systems in the market of corporate IT solutions. For instance, O´Leary (2000) reports that a single ERP system (SAP´s R/3) is used by more than 60% of the multinational firms. Among the explanations for this phenomenon are the competitive pressures suffered by the companies that have forced them to seek alternatives for cost reduction, differentiation of products and services and integration of their business processes. The ERP systems evolved exploiting the need for quick deployment of integrated systems to meet these new business requirements, while companies were (and still are) under pressure to outsource all the activities that are not embraced by their core business.

The ERP systems are introduced in companies following some well-defined stages. In this context, their life cycle encompasses the stages of decision, selection, implementation, stabilization and utilization. This chapter presents aspects involved in each stage of this life cycle, based on the referenced bibliography.

BACKGROUND

Enterprise Resource Planning (ERP) systems are integrated information systems acquired as commercial software packages with the aim of supporting most of the operations of a company. Markus and Tanis (2000) define them as commercial packages that enable the integration of data coming from transactions-oriented information systems and from the various business processes throughout the entire organization. Although companies may internally develop systems with similar characteristics, the term ERP is associated to commercial packages. Examples of ERP systems found on the market are the R/3 of the German company SAP and the PeopleSoft EnterpriseOne of the American Peoplesoft. Some authors present and describe characteristics that, if taken as a whole, permit to differentiate the ERP systems from systems developed within the companies and from other types of commercial packages (Souza & Zwicker, 2001; Markus & Tanis, 2000). These characteristics may be summarized as:

- ERP systems are commercial software packages;
- They include standard models of business processes;
- They are integrated information systems and use a corporate data base;
- They have a large functional scope;
- They require adjustment procedures to be used in a given company.

When deciding to use ERP systems, companies hope to achieve manifold benefits, like the integration of business processes, the increase of possibilities to control the company's operations, the technological updating, IT cost reduction and access to quality information in real time for decision taking (spread over the complete production chain). However, there are also problems to be considered. Table 1 synthesizes the benefits and difficulties of the ERP systems mentioned by many authors (Lozinsky, 1998; Hecht, 1997; Bancroft, Seip & Sprengel, 1998; Davenport, 1998; Stedman, 1998; & Cole-Gomolski, 1998), and relates them to the characteristics of ERP systems.

ERP SYSTEMS LIFE CYCLE MODEL

The life cycle of information systems represents the various stages through which a project of development and utilization of information systems passes through. In its traditional form, the systems development life cycle encompasses project definition, system study, design, programming, installation and post-implementation stages (Laudon & Laudon, 2001). Two different approaches for system development are the *waterfall* model, where stages are executed sequentially and only once for each system, and *prototyping,* where the stages are repeated, refining an initial solution (prototype) until it can be used as the definition for the system to be built or as the system itself.

Table 1. Benefits and difficulties of ERP systems

Characteristics	*Benefits Sought*	*Possible Difficulties*
Commercial Package	- IT costs reduction - Focus on the company's core activities - Technological updating - Backlog reduction	- Supplier dependence - Lack of knowledge on the package - Time for learning of interfaces not specifically developed for the company - Loss of former systems functionalities - Costs regarding the continued adjustment - Excess of screens and fields to be typed in - Lack of adequate managerial reports
Best Practice Business Models	- Knowledge on best practices - Access to other companies' experience - Reengineering of processes	- Need to adjust the company to the package - Need to change the business procedures - Need of consulting for implementation - Resistance to change
Integrated System	- Greater control on the company's operation - Real time access to data and information - Elimination of interfaces between isolated systems - Improvement of information quality - Synchronization between activities of the value chain enhancing global planning of the company	- Higher implementation complexity and costs - Difficulty to update the system as it requires agreement among various departments - One module not available may interrupt the functioning of the others - Resistance due to increase of work in the areas responsible for data input - Resistance due to increase of demands to the areas responsible for data input
Corporate Data Base	- Standardization of information and data definitions - Elimination of discrepancies between information of different departments - Information quality improvement - Access to information for the whole company - Cultural change to a view of dissemination of information from the departments to the entire company	- Cultural change of the view of "owner of the information" to that of "responsible for the information" - Responsibilities attribution on files shared between areas - Overload of the data base causing performance problems
Great Functional Scope	- Maintenance elimination of multiple systems - Standardization of practices - Reduction of training costs - Interaction with a single supplier	- Dependence upon a single supplier - If the system fails the entire company may stop - Support difficulties in the stabilization phase

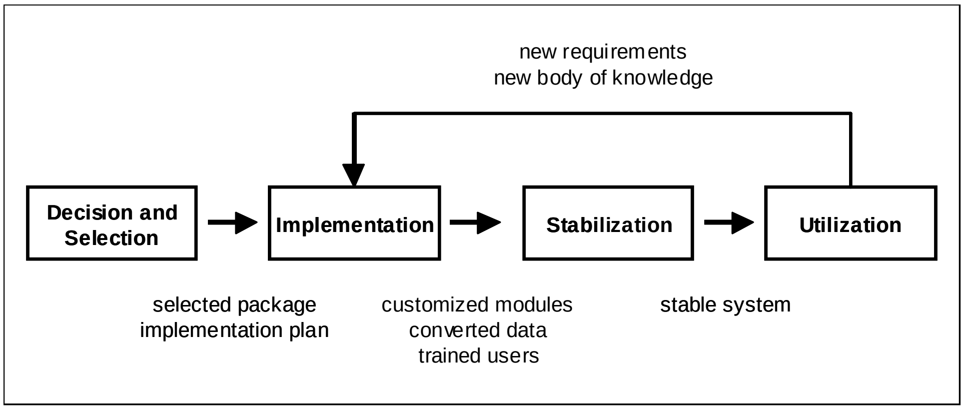
In the case of use of commercial software packages, it is, prewritten and precoded application software programs commercially available for sale or lease, the stages must be considered in a different manner. For instance, in the system study stage the focus is not on obtaining a detailed system specification from the users for programming the system, but instead, verifying the functionality of the many choices available from vendors, against a set of requisites from the users that will guide system adaptation or customization.

Like any commercial software package, ERP systems exhibit differences in their life cycle in relation to traditional systems development projects. But because of their large functional scope and their integration between its various modules, these differences are deepened. Some authors present models for the life cycle of ERP systems (Esteves & Pastor, 1999; Markus & Tanis, 2000; Souza & Zwicker, 2001). Souza and Zwicker´s model is shown in figure 1 and includes the stages of decision and selection, implementation, stabilization and utilization and is used next as a framework to present aspects involved in each stage of the life cycle.

(a) Decision and Selection

At the decision and selection stage the company decides to implement an ERP system as an IT solution and chooses the supplier. A series of issues must be taken into account at this stage. For instance, Davenport (1998) analyzes the decision from the point of view of the compatibility

Figure 1. ERP systems life cycle model (Souza & Zwicker, 2001)



between the organization and the characteristics of the ERP systems. Hecht (1997) presents criteria that may help in this choice: adjustment of the package's functionality to the requisites of the company, the technical architecture of the product, the implementation costs, the quality of post-sales support, the financial health of the supplier, and his view of the future. The main product of this stage is a detailed implementation plan, where the modules to be implemented, the implementation approach, the project schedule, the implementation team and responsibilities are defined.

(b) Implementation

Implementation comprises the second stage of the ERP systems' life cycle, although the term is also used to represent the full life cycle. The implementation of an ERP system may be defined as the process by which the system's modules are put into operation within a company. Implementation entails adjustment of the business process to the system, eventual customizing of the system, the loading or conversion of initial data, hardware and software configuration, training of the users and managers and the offer for support and help. This stage encompasses the tasks ranging from the end of implementation plan's concept to the beginning of the operation.

The implementation stage is reported to be the most critical of all (Bingi, Sharma & Godla, 1999). Difficulties are mainly due to the organizational changes that imply in changes of the tasks and responsibilities of individuals and departments and transformations in the relationships among the different departments. In an ERP system implementation it is generally pursued the optimization of the global processes of the company, which may cause as a counterpart changes in the activities in most of the departments involved. The need of intense participation and commitment of the company's top management and the requirement of permanent communication among the involved units is brought about by the size and complexity of this change and of the conflicts it may generate among those involved.

The process of adjustment of the ERP system to the company's processes is part of the implementation stage and is achieved by the adjustment of parameters or software customizing (development of programs to modify or complement the existing functions). At the implementation stage, the decision on how the start of the operation of the ERP system (the "go-live") will take place is also important. All modules may start operating: in all divisions or plants of the company simultaneously (*big bang)* or in one division or plant after the other (*small bangs)*. The start may also occur in *phases* (one or some modules start operating in one division or plant after the other, also called *roll-out*). The approach used to implement an ERP system is an important decision in its implementation project as it greatly affects the configuration of the system, the allocation of resources, and the management of the project and its risks. It will also play a decisive role at all the stages of the ERP system's life cycle.

The integration of processes pursued by ERP systems entails difficulties for the implementation stage. These difficulties are related to three types of changes in the way people do their work (Souza & Zwicker, 2001):

1) The integration aspect transfers to departments that produce the information the responsibility to insert it properly. This includes data used by other departments only (for instance, typing of an accounting bill in a production entry) and as a consequence, the users feel that their tasks are increased.

2) Information must be recorded into the system at the best-suited moment for the process and not at the best-suited moment for a specific department. Thus, there is a need to change the order and form in which tasks are carried out and other departments begin to demand the information they rely upon.
3) The activities of a department become transparent to all others and this has the inconvenience to require "explanations" for everything it does.

However, once these shortcomings are overcome, many users feel that they grew as professionals, since they start to have a broader view of the company's activities and to have a clearer perception of their role and importance in business processes. Training of end users for the work at an integrated system, taking into account the quoted aspects, is an important consideration for the success of the implementation process.

(c) Stabilization

In the first moments after the beginning of the ERP system operation there is a critical stage for the project's success: the stabilization. At this moment, the ERP system, that until then was only an abstraction, gains reality and starts to be part of the company and of the people's daily life. This is when the highest amount of energy, be it managerial or technical, is required. It is a stage in which problems, that could not have been easily detected at the implementation stage, become apparent. This is a particularly critical stage, as the company is already relying upon the system for its activities and which causes major pressure for the speedy solution of problems. The length of this period depends on the company and takes about eight weeks (Zwicker & Souza, 2004).

Two critical aspects may be highlighted at this stage: difficulties of the end users and problems of the ERP system (in programs and their adjustment to the company). Regardless of the fact that users have been trained in functions of the new system, they operate slowly in the early moments as they have doubts and feel insecure about the appropriateness of their actions. Besides the difficulties of adjustment to the functions of the new system, there is the issue of the cultural adjustment of people to the requirements of work in an integrated environment. While users face these difficulties, the detection of errors in customized programs and parameter settings also take place, hindering the normal operation. The simultaneous occurrence of use difficulties and system faults, associated with the recent familiarity of the project team to the new system, make the identification of the problems real causes even more difficult at this stage.

The exact characterization of this stage is related to the operation-starting mode chosen by the company. If operation of the ERP system started by means of a *big bang*, the stage of stabilization can be clearly distinguished from those of implementation and utilization. However, in companies that implement the modules in phases, or even in *small-bangs*, the stabilization stage is less characterized and merges with the implementation stage of the remaining modules. It can be stated that the stabilization stage in the case of implementation by phases starts with the operation of the first module and ends only when the last module implemented, in the last locality of the company, becomes stabilized. This longer implementation and stabilization time in general entails loss of focus of the project and may be viewed as a risk factor for the implementation in phases (for further details on the influence of ERP *go-live* approach on its life cycle, see Zwicker & Souza, 2004).

(d) Utilization

Finally, at the fourth stage the system starts to belong to the day-by-day operations. This does not mean that all its possibilities of usage are known, or that they are properly equated. Orlikovski & Hofman (1997) report on the difficulty within a company to know, beforehand, all use possibilities of the new information technologies. This knowledge is only achieved after a certain period of continued use of the technology, through the ideas that emerge during the utilization process. Therefore, the stage of utilization feeds back the stage of implementation with new possibilities and needs that may be solved through new modules, parameter adjustments, or software customizing. In the case of implementation in phases, the already implemented modules may impose restrictions upon new modules caused by already defined parameters or customizations. Indeed, new modules should not imply in changes of already configured and operational modules, usually a difficult task.

It is also observed that some time after the beginning of the operation, the implementation of customizations or new modules becomes more difficult. New modules not included in the original plans or improvements in the ERP system exhibit difficulties, characterizing that the new situation is decidedly refrozen. As a rule, these difficulties are associated to changes in priorities of the company and of the IT department, to the appearance of new projects and to the difficulty to bring together again, all departments and users needed for the implementation of the wished changes. Therefore, there are implicit difficulties for the realization of the continued adjustment recommended by the above referenced authors. According to the authors' model, knowledge of the possibilities and functionalities of ERP systems is only consolidated after the operation start, however, it was noted that companies have difficulties to implement new ideas and solutions at that stage. A possible explanation for this discrepancy may be the fact that the case analyzed by the authors dealt with a collaborative computer system that, although of great importance for the company under study, had less functional scope than an ERP system. Apparently, the high functionality of ERP systems and the impact they cause on the organization constitute inflexible barriers against the continued adjustment stage.

Markus and Tanis (2000) point out that it is only in this stage that the organization is finally able to ascertain the benefits (if any) of its investment in the ERP system, due to continuous business improvement, additional user skill building, and post-implementation benefits assessment. Esteves and Pastor (1999) also point out an interesting phase of the life cycle, the "retirement" stage, in which the ERP system would be substituted for other solution or solutions. What is being observed is that companies go through great efforts and costs each time a new version upgrade is needed (McMahon, 2004). Every time the decision to upgrade or not is posed to the organization it is also an opportunity to analyzing other ERP vendors or technologies (Kremers & van Dissel, 2000).

FUTURE TRENDS

The introduction of an ERP system changes the focus of the IT area. Normally the area evolves from a technical approach to a business processes approach. It is also apparent that the role of the users in the management of an ERP system is much more important and fundamental than in the case of internally developed systems.

There are also new challenges at the day-by-day management of IT in an ERP system context. For instance, maybe there are relationship problems with the supplier that demand the attention of the IT manager. Some of this type of difficulties apparently stem from the continued need of system updating that cannot always be strictly accomplished when the supplier sends new patches and releases. Indeed, the IT areas adopt alternative procedures to keep their updating efforts at levels consistent with their possibilities. Apparently, companies that possess ERP systems continue to be subject to the need of carrying out substantial upgrades requiring time and significant resources. At the time of internally developed systems, those aspects possibly were much less significant (Zwicker & Souza, 2004). A more detailed contextual analysis of this myriad of details opens an interesting possibility for future research in the ambit of the ERP utilization stage.

CONCLUSION

During the second half of the nineties, the implementation of ERP systems was one of the central points of attention regarding IT utilization in companies. History shows that implementation of ERP systems is not a simple matter since there are some reported failures (see Jesitus, 1997; Barker & Frolick, 2003).

Research on the subject, developed as from the end of the last decade, studied mainly the factors governing successful implementation and showed that a process of cultural change is involved. One critical factor for success is to avoid that the endeavor be handled as an information technology project (Willcocks & Sykes, 2000). Dedication and involvement of top management, strong participation of users and change management were aspects considered essential for the success of these implementation projects (Bingi, Sharma, & Godla, 1999).

This chapter presented a model for ERP systems life cycle that tries to encompass the complexities involved in implementing ERP systems in companies. To achieve this purpose, several aspects relating to the decision and selection, implementation, stabilization and utilization stages were presented.

The recommendation for companies that are deciding for ERP systems utilization and for companies that are already in the implementation stage is the careful analysis of the difficulties associated which each ERP systems´ life cycle stage. With a better knowledge about these difficulties, the process can be improved though better planning and better action on the inherent difficulties of such an organizational change.

REFERENCES

Bancroft, N. H., Seip, H. & Sprengel, A. (1998). *Implementing SAP R/3: How to introduce a large system into a large organization*, 20(2), Greenwich: Manning.

Barker, T. & Frolick, M. (2003, Fall). ERP implementation failure: A case study. *Information Systems Management*, 20(2), 43-49.

Bingi, P., Sharma, M. & Godla, J. (1999). Critical issues affecting an ERP implementation. *Information Systems Management*, *16*(3), 7-44.

Cole-Gomolski, B. (1998, September 9). ERP! Excuse us as we digest our new system. *Computerworld*, 32(38), 1-2.

Davenport, T. H. (1998, July/August). Putting the enterprise into the enterprise system. *Harvard Business Review*, 76(4), 121-131.

Esteves, J. M. & Pastor, J. A. (1999). An ERP life-cycle-based research agenda. *First International Workshop in Enterprise Management and Resource Planning: Methods, Tools and Architectures – EMRPS'99*, Venice, Italy.

Hecht, B. (1997, March). Chose the right ERP software. *Datamation*, 43(3), 56-58.

Jesitus, J. (1997, November 3). Broken Promises? *IndustryWeek*.

Kremers M. & van Dissel, H. (2000). ERP system migrations. *Communications of the ACM, 43*(4).

Laudon, K. C. & Laudon, J. P. (2001). *Management information systems* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Lozinsky, S. (1998). *Enterprise-wide software solutions: Integration strategies and practices*. Addison-Wesley.

Markus, M. L. & Tanis, C. (2000). The enterprise system experience: From adoption to success. In R. Zmud (Ed.), *Framing the domains of IT research: Glimpsing the future through the past.* Cincinnati: Pinnaflex.

McMahon, S. (2004, June 7). Beating the clock on ERP upgrades. *Datamation*.Retrieved from *http://itmanagement.earthweb.com/erp/article.php/3364581*

O'Leary, D. E. (2000). *Enterprise resource planning systems: Systems, life cycle, electronic commerce and risk.* Cambridge: Cambridge University Press.

Orlikovski, W. J. & Hofman, J. D. (1997, Winter). An improvisational model for change management: the case of groupware technologies. *Sloan Management Review*, 38(2), 11-21.

Souza, C. A. & Zwicker, R. (2001). ERP systems' life cycle: findings and recommendations from a multiple-case study in Brazilian companies. *Proceedings of the 2001 Conference of Business Association of Latin American Studies - BALAS 2001*, San Diego.

Stedman, C. (1998, February 2). ERP user interfaces drive workers nuts. *Computerworld*, 32(44), 1-2.

Willcoks, L. P. & Sykes, R. (2000). The role of the CIO and IT function in ERP. *Communications of the ACM, 43*(4), 33-38.

Zwicker, R. & Souza, C. A. (2004). SAP R/3 Implementation approaches: A study in Brazilian companies. In L.K. Lau (Ed.), *Managing Business with SAP: planning, implementation and evaluation.* Hershey, PA: Idea Group Publishing.

KEY TERMS

Big-Bang Approach: Implementing all modules of an ERP system in all locations or plants of the company simultaneously.

ERP Decision and Selection Stage: Stage at which the company decides to implement an ERP system and chooses the supplier.

ERP Implementation Stage: Stage of an ERP project at which the ERP system's modules are put into operation.

ERP Stabilization Stage: The first weeks after the beginning of an ERP system operation in the company.

ERP Systems: Integrated information systems purchased as commercial software packages with the aim of supporting most operations of a company.

ERP Systems' Life Cycle: The various stages through which a project of introducing an ERP system in a company passes through.

ERP Utilization Stage: Stage of an ERP project at which the system starts to belong to the day-by-day operations of the company.

Phased Approach: Implementing the modules on an ERP system in sequential steps, comprising one or more modules in one or more locations or plants, until the system is completely installed. Also called roll out.

System Development Life Cycle: The various stages through which a project of development and utilization of information systems passes.

Limited-Perspective Bias in Technology Projects

Jo Ellen Moore
Southern Illinois University Edwardsville, USA

Lisa A. Burke
Louisiana State University Shreveport, USA

INTRODUCTION

Limited-perspective bias is a human tendency to overestimate the completeness of what we know and to act on our own (limited) perspective of what is important (Moore & Burke, 2004a). In organizations, each person possesses a view that evolves from what he or she experiences and observes on a daily basis. Given one's location or role in an organization, these views often differ and can affect one's perspective on a problem or situation (O'Reilly & Pondy, 1979). Each individual perspective, however, is by and large "valid" in that it represents a distinct cognitive reality within that firm.

But for truly effective decision making to occur, on both the day-to-day level and strategic level, the individual perspectives must integrate to create a fuller and more accurate view of situations and issues. In other words, limited-perspective bias often needs to be managed, and efforts to counter the bias are particularly crucial in dynamic organizations that depend on empowered employees to make decisions and figure out the right things to do. Left untended, limited-perspective bias (LPB) can lead an individual to formulate decisions and take actions that are ineffective, possibly even detrimental to the organization.

The purpose of this article is to summarize the present formulation of the limited-perspective bias construct, indicating why it is highly pertinent in information systems and technology (IS/IT). The occurrence of LPB in technology projects is then discussed and directions for future research are extended.

BACKGROUND: THE NATURE OF LPB AND ITS OCCURRENCE IN IS/IT

The existence of cognitive biases is established in the psychology and organizational behavior literature. For example, researchers have identified biases of consistency (Janis & Mann, 1977), escalating commitment (e.g., Staw & Fox, 1977), representativeness (e.g., Nisbett & Ross, 1980), and the fundamental attribution error (Ross, 1977). Limited-perspective bias (Moore & Burke, 2004a) is a newly conceptualized type of cognitive bias, presently positioned for further theoretical development and empirical investigation. Within the three-stage model of construct evolution (Reichers & Schneider, 1990), limited-perspective bias is in the "introduction and elaboration" stage, where a concept is discovered and efforts are made to present the new concept to fellow researchers and practitioners and to legitimize and refine it.

In providing a conceptual foundation for LPB, Moore and Burke (2004a) propose five situational factors that contribute to its occurrence: interdependence, uncertainty, ambiguity, deadline-driven work pace, and role incompatibility. All five of these characteristics are common in IS/IT organizational environments. Endeavors to apply technology dictate an interdependence among IS/IT and business personnel. For example, in technology implementations, a number of individual perspectives (e.g., the business user's perspective, technology infrastructure specialist's perspective, application developer's perspective, among others) are pertinent and necessary for the effort to culminate in an effective and useable solution. Such interdependence begs for management of LPB, as these single, specialized perspectives must be integrated and eventually merged in order to produce a proper end product for the organization within the time frame that it is needed.

Uncertainty and ambiguity are also proposed to contribute to occurrences of LPB in information systems and technology. Efforts that involve new technologies, new tools, or new methodologies, as well as projects that attempt to apply technology in novel ways, are inherently laced with uncertainty and ambiguity. LPB can aggravate and increase the difficulties encountered when working through such unknowns by restricting one's understanding of the situation, reducing one's resources for resolving the problem or issue, and ultimately causing one to head in a less than optimal direction.

A frenzied and deadline-driven work pace ("we need it yesterday") also contributes to LPB, as prior research suggests that time pressure adversely impacts information search and processing. Time pressure can prevent a person from recognizing that limited-perspective bias is happening and, even if a person recognizes the possibility of this bias, looming deadlines can prevent one from taking time to gather and assimilate information from perspectives other than one's own. Furthermore, managing the aforementioned factors of interdependence, uncertainty, and ambiguity takes time, so the frenzied work pace in which IS/IT so often operates compounds the likelihood that those factors lead to negative effects of LPB.

Finally, role incompatibility is intuitively associated with LPB. Whetten and Cameron (1998) identify role incompatibility as a source of conflict that is greatly exacerbated by resource scarcities, which are seemingly pervasive in organizations today. For example, in technology application projects, the development team is nearly always focused on producing a fully featured and quality product, believing that time and budget constraints should change to accommodate functionality (Dobson, 2001), but, upper management is likely focused on budgetary issues and getting the product in place as soon as possible because of the business process improvements associated with it. Although these two parties undoubtedly experience role incompatibility (e.g., management focusing on the time constraint and saying we need it now, the development team focusing on the functionality of the system and saying they need more time), if both parties can overcome their limited-perspective bias, role incompatibility can dissolve and true organizational priorities can be determined to better direct efforts of both parties.

MAIN THRUST OF THE ARTICLE: EFFECTS OF LPB ON TECHNOLOGY PROJECTS

Because technology implementation projects tend to possess elements of interdependence, uncertainty, ambiguity, challenging deadlines, and role incompatibility, they provide fertile ground for limited-perspective bias and its harmful effects. An example of how LPB can operate in a technology project is found in the area of project status reporting:

Application development and technical teams were hard at work in a banking institution to develop and implement the firm's first web-based application. The initial target date to deploy the application was June 15th. According to plan, the technical team had the infrastructure in place by that date. This technical platform included a contract outsourcing the web-hosting function for a monthly fee of $30,000, and the contract was signed to begin the services (and corresponding payments) on June 1st.
As June neared, however, the development team delayed the deployment date to September 15th. To meet that date, the new application had to be migrated to the pre-production environment for final user testing on September 1st. In the last week of August (mere days prior to final user testing and less than three weeks from the go-live date), the development team reported they were not ready. The team later delivered new completion estimates, which dictated a major postponement to December 15th. (Moore & Burke, 2004b)

All five proposed antecedents to LPB seem to be present in this project. Interdependence between the development and technical teams is apparent, and ambiguities and uncertainties were likely encountered by the developers in constructing this first-of-its-kind application for the organization. Clearly, the developers were operating under time pressure, striving to meet deadlines that became impossible to meet. They likely faced role incompatibility as well, since management undoubtedly was eager to have this application deployed.

LPB is observable in the project status reporting behavior of the development team. Considering they were so far from product completion, they probably knew for some time that the delivery date was not going to be met, yet they chose not to communicate this to management and other stakeholders. Many ramifications of LPB ultimately resulted, but one is clearly quantifiable from this short description of the project – the company could have saved $180,000 on the Web hosting contract had the development team reported the true status of their project.

In choosing to avoid revealing the true project status, the perspective of the development team likely included concern about negative effects that an inability to bring the product in on time might have on their work reputations, the anticipated unpleasantness of delivering bad news, and perhaps a hope that they could get things back on schedule before anyone needed to know the project was in trouble. Unfortunately, though, the development team operated on a *limited* perspective of what was important and what could be done.

The development team failed to perceive how their decisions and actions (such as not accurately reporting their progress) could adversely affect other important organizational functioning. They also operated on a limited perspective of what could be done to correct the situation and bring the project back under control. Whitten (1995) recognizes that many severe problems that occur in projects are not solvable by the individuals or specific group encountering them. LPB prevented the develop-

Table 1. Recommendations to minimize LPB (from Moore & Burke, 2004a)

General practices:
✓ Recognize that LPB occurs, in yourself and in others.
✓ Actively share your unique perspective, by routinely providing reasons for your decisions and actions.
✓ Seek out other valuable perspectives in an effort to broaden your own.
When faced with a specific dilemma or problem:
✓ Ask "what is best for the company?"
✓ Focus on what is right, rather than who is right.
✓ Remember our human tendency to underestimate situational factors others may be facing and inquire about those.

ment team from considering the ability of others in the organization to make necessary adjustments to support their efforts, as well as the ability of others in the organization to assist in remedies (e.g., by contributing resources or making adjustments in specifications or design).

In addition to illustrating the existence of LPB antecedents and consequences in technology projects, this example shows how easy it is for LPB to occur; that is, just how "human" it is. Because LPB is a natural tendency, deliberate and proactive effort is required to prevent it from occurring. Table 1 provides a summary of general recommendations from Moore and Burke (2004a) for minimizing LPB.

FUTURE TRENDS: RESEARCH OPPORTUNITIES

Because the limited-perspective bias construct is in the early stages of development, it provides fertile ground for exciting and timely investigation. Researchers are encouraged to form operationalizations of LPB and, when a reliable and valid measure is established, to begin studies to empirically define its nomological net. As a starting point, the five antecedents proposed by Moore and Burke (2004a) should be examined to determine the extent to which they contribute to LPB.

An additional angle to LPB that will be interesting to explore is that of national culture. The initial conceptualization of LPB appears to be grounded in the U.S. national culture. One would expect individualistic cultures, such as that of the U.S., to be more vulnerable to LPB, as workers in such cultures are apt to make decisions by themselves without checking with colleagues or soliciting input from others (Greenberg, 2002). Comparative analyses involving cultures that differ on dimensions such as individualism-collectivism are needed to determine the extent to which national culture influences LPB.

Researchers are also encouraged to examine LPB within the bounded (or limited) rationality decision-making framework (e.g., March & Simon, 1993). The theory of bounded rationality contends that although decision makers try to be rational, they are constrained by limited cognitive capabilities and incomplete information. They do not consider all alternatives or all consequences, and relevant information is frequently not sought. Of particular interest would be the examination of LPB in regard to the limitations that have been identified as contributors to bounded rationality: attention, memory, comprehension, and communication. March (1994) notes that division of labor to mobilize and utilize specialized talents (common in technology application efforts) encourages differentiation of knowledge and leads individuals to use different frameworks for understanding and simplifying their worlds. One manifestation of these different frameworks may be LPB.

Finally, efforts are encouraged to identify mechanisms we can employ to minimize LPB in technology projects and organizations. For example, what specific practices can individuals adopt to make the sharing of their perspective a routine and efficient part of what they do in the course of their workday? And what mechanisms can an individual put in place to ensure that he or she receives a steady flow of other perspectives deemed to be key to his or her position in an organization or project?

CONCLUSION

Laurence J. Peter (1982) may have said it best in the September 24th entry of *Peter's Almanac*: "Some problems are so complex that you have to be highly intelligent and well informed to be undecided about them." In today's world of new and complex problems, alleviating limited-perspective bias may be essential in arriving at the best solutions.

REFERENCES

Dobson, M.S. (2001). *Project management for the technical professional*. Newtown Square, PA: Project Management Institute, Inc.

Greenberg, J. (2002). *Managing behavior in organizations.* Upper Saddle River, NJ: Prentice Hall.

Janis, I., & Mann, I. (1977). *Decision making: A psychological analysis of conflict, choice, and commitment.* NY: The Free Press.

March, J.G. (1994). *A primer on decision making: How decisions happen.* NY: The Free Press.

March, J.G., & Simon, H.A. (1993). *Organizations* (2nd ed.). Oxford: Blackwell.

Moore, J.E., & Burke, L.A. (2004a). Managing limited-perspective bias in IT. In M. Igbaria & C. Shayo (Eds.), *Strategies for managing IS/IT personnel* (pp. 248-263). Hershey, PA: Idea Group Publishing.

Moore, J.E., & Burke, L.A. (2004b). Reluctance to report reality in troubled technology projects. In M. Igbaria & C. Shayo (Eds.), *Strategies for managing IS/IT personnel* (pp. 282-299). Hershey, PA: Idea Group Publishing.

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings.* Englewood Cliffs, NJ: Prentice Hall.

O'Reilly, C., & Pondy, L. (1979). Organizational communication. In S. Kerr (Ed.), *Organizational behavior.* Columbus, OH: Grid.

Peter, L.J. (1982). *Peter's almanac.* NY: Morrow.

Reichers, A.E., & Schneider, B. (1990). Climate and culture: An evolution of constructs. In B. Schneider (Ed.), *Organizational climate and culture* (pp. 5-30). San Francisco, CA: Jossey-Bass Publishers.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (vol. 10, pp. 173-220). NY: Academic Press.

Staw, B.M., & Fox, F.V. (1977). Escalation: Some determinants of commitment to previously chosen course of action. *Human Relations*, *30,* 431-450.

Whetten, D.A., & Cameron, K.S. (1998). *Developing management skills.* Reading, MA: Addison-Wesley.

Whitten, N. (1995). *Managing software development projects: Formula for success.* NY: John Wiley & Sons, Inc.

KEY TERMS

Ambiguity: Something difficult to interpret, difficult to understand, unclear. A situational factor proposed to contribute to the occurrence of LPB.

Bounded Rationality: A theory of individual decision-making that contends decisions are not made in a purely rational manner, due to limited cognitive capabilities and incomplete information.

Frenzied and Deadline-Driven Work Pace: A work environment characterized by persistent time constriction and pressure to meet challenging deadlines. A situational factor proposed to contribute to the occurrence of LPB.

Interdependence: Mutual dependence between parties in an organizational setting. A situational factor proposed to contribute to the occurrence of LPB.

Limited-Perspective Bias (LPB): A cognitive bias that reflects a human tendency to overestimate the completeness of what we know and to act on our own (limited) perspective of what is important and what can be done.

Role Incompatibility: A situation in which each party perceives its concerns or needs to be mutually exclusive (or antagonistic) to the other party's concerns or needs; that is, role incompatibility exists when one perceives his or her needs cannot be met if the other party's needs are met. A situational factor proposed to contribute to the occurrence of LPB.

Uncertainty: Something unknown, unreliable, or changeable. A situational factor proposed to contribute to the occurrence of LPB.

Macromedia Flash on the Client and the Server

Robert Barone
Buffalo State College, USA

INTRODUCTION

The Internet and more specifically the World Wide Web is a global communication environment. Either via a personal or commercial Web site, we seek to express ourselves as individuals or as organizations through this technological forum. We desire to provide an experience that will bring visitors back to our site. Providing rich multimedia content is the contemporary paradigm in achieving these ends. The software company Macromedia has addressed this need with their commercial product known as Flash. Macromedia Flash is used to create interactive multimedia environments in the client. There are also server-based technologies that communicate via Flash directly or indirectly.

BACKGROUND

Macromedia Flash has been evolving at a considerable pace. Up to Flash 4.0, the user interface was menu driven. From version 5.0 on, the user interface is palette oriented. Flash MX is version 6.0. MX designates a suite of programs. The Flash version at the time of this writing is MX 2004 (version 7.2). Components, introduced in Flash MX, allow the rapid building of user interfaces (analogous to HTML form elements). Flash MX 2004 introduces a new component architecture, called version 2 (or simply v2) with features not found in the previous (version 1 or v1) Flash MX component architecture (Statler, 2003). MX 2004 sees a branching of Flash into two separate applications, Flash MX 2004, incorporating the standard elements of the application, and Flash MX 2004 Professional, an extended version. A history of Flash is available at *www.macromedia.com/macromedia/events/john_gay/index.html*.

FLASH AND THE CLIENT

Flash is much more than a graphic arts program. This application allows the designer and the developer to create a true and unique graphical user interface (GUI). This GUI can be precisely programmed to support end user interactivity. Flash has found tremendous treatment in the World Wide Web, although it is not limited to this arena.

Flash employs vector graphics as the primary medium in content delivery. Vector graphics are described by mathematical formulas. These formulas are solved by the video processor of the client computer and a raster image is rendered. Raster graphics are bitmap images. They are composed of pixels that require a large number of bits to convey the information. Bitmap file sizes are inherently larger even when lossy compression techniques like JPEG (Joint Photographic Experts Group) (Knuckles & Yuen, 2005), a standard graphic file format, are utilized. Vector graphics are scalable and retain information when enlarged, although they suffer from aliasing (Bardzell & Bardzell, 2004). Raster graphics are not scalable and become unintelligible when enlarged. Flash makes use of streaming. The information contained in the streaming wave file is displayed in the browser as it is received. The browser does not wait for the entire contents of the file to be downloaded first. These two features, vector graphics and streaming, combine to make Flash a successful medium for providing multimedia content even with low bandwidth connections. Furthermore, the programming aspect of Flash further enhances and extends the Web sites usability by bringing interaction to the forefront.

Concepts in Flash are hierarchical in organization. Drawing on the stage with any of the tools creates a drawing object. These are the basic building blocks of a Flash movie. Objects can be selected, copied, grouped, transformed, skewed, stacked, deleted, and so forth. Essentially any designer's demands can be met with the drawing object. Flash also supports importing of images.

The next level of hierarchy is the symbol. If the desire is to motion animate (motion tween) the drawing object, it must first be converted to a symbol. Interestingly shape morphing (shape tweening) is performed on a drawing object. There is a subtle importance here which cannot be underestimated. Once a symbol is created, it is automatically placed inside a container known as the library. If the designer wishes to re-use the symbol, it is dragged and dropped from the library onto the stage. Multiple occurrences of the symbol on the stage are referred to as instances of the symbol. An instance on the stage may be thought of as pointing to the symbol in the library. Since the symbol is only stored once and reused many times in the form of an instance, file size remains at a minimum.

Another advantage is file-to-file reusability of symbols. A library from one Flash file can be opened in a completely separate file and the symbols reused.

There are three symbol types, *movie clip, button,* and *graphic*. Movie clips are self-contained Flash movies that can be treated as instances. Once on the stage they cycle through their animation sequences independently of the main movie. Movie clips are also programmable. Buttons maintain an up state, an over state, a down state, and a hit state. The first three states can be animated and buttons are also programmable. Graphic symbols are simply used to create motion tweens and are not programmable in the sense that an action cannot be directly attributed to one.

The stage also maintains a hierarchy in the form of layers. Think of the stage as a deck of playing cards (a typical analogy). Each card in the deck is a layer—the card at the top being the highest in the hierarchy and those below it are lower. Typically, one animation per layer is the rule. It is not enough to recognize that a symbol must be on the stage to work with. Identifying the corresponding layer it resides in is just as important.

The timeline is the backbone of a Flash movie. A celluloid film strip is composed of still images that are rapidly moved across a movie projector to give the illusion of continuous movement. The same is true of a Flash movie. Each layer has a timeline that is subdivided into frames. A symbol is placed into a layer. Movements of symbols on the stage are broken across a series of frames in the layers timeline. As the software play head moves form one frame to the next, the symbol is given the illusion of movement. When the play head reaches the last frame, it loops back to the beginning and plays over unless a programmable command is given to interrupt the process. Consequently, it is seen that certain symbols can be programmed via the timeline, directly via their instance, as well as the actions and objects that can be attributed in a given situation.

A Flash movie is created in a file with extension *.fla*. This file can create many file types. For Web delivery, the common types are *.html* (hypertext markup language) and streaming wave file *.swf* (streaming wave file or small Web file). The client's browser requests the html from the server. Once the browser has loaded the html file, the browser calls the streaming wave file from the server.

Flash incorporates a programming language know as ActionScript, which was introduced in Flash version 5.0 (Mohler, 2001). ActionScript, like many other scripting languages, is considered loosely typed and prototype based (Hall & Wan, 2003). This means that variables do not have to be declared before they are used and that the data type stored in a variable may change depending upon how program logic flows. Flash MX 2004 introduces ActionScript 2.0, whose distinguishing features are the ability to define a true class, case sensitivity of all identifiers, and strong (or strict) data typing (Statler, 2003). The developer has the ability to choose the version of ActionScript and Flash Player the movie will run in. ActionScript is ECMA (European Computer Manufacturers Association) compliant (Williams, 2002).

With ActionScript, the developer is handed an object toolbox to use. The Flash architects have designed an immense collection of functions and objects which can be utilized to manipulate instances of symbols. Once the concept of using an object in a program is understood, what remains is becoming familiar with the myriad of object methods and properties. It should also be noted that ActionScript and JavaScript can communicate on the client side. Typical examples are pop-up windows.

The end result, the Flash movie, is essentially a unique stand-alone miniature GUI application that executes on a desktop or in a browser window delivered via disk or network.

FLASH AND THE SERVER

Communication to and from a Flash movie, especially from the client back to the server, is possible. One method is to use middleware. Middleware is a server side script that resides and executes on a server computer. Examples of server side scripting environments are Active Server Pages, ASP, by Microsoft (now ASP.NET); ColdFusion by Macromedia; Java Server Pages, JSP, by SUN; and Personal Home Page, PHP, an open source technology. Perl (Practical Extraction and Report Language) (Castro, 2001) was one of the first languages used to build a Common Gateway Interface, CGI, for server side interactivity.

The idea here parallels the concept of creating an ordinary FORM tag in a traditional HTML (hypertext markup language) page and using GET or POST method to send the data collected in the FORM to a server side script that processes it. This is where Flash will make use of program-controlled text boxes and components to produce HTML FORM behavior.

The server side script will process the data and respond in many different fashions. One way is to create an HTML page and send it back to the browser. Another is to store the data in a database manager first and then send a page back to the client. There are many variations of theme depending upon the Web application.

An important consideration (Chambers, 2002) when using a Flash movie to interface with middleware is the following: a Flash movie is loaded into the browser once and can handle many data transfers to and from the middleware. Middleware can build content for a Flash movie without building a new Flash movie (Muller, 2003). An HTML file loaded into the browser interacts with the middleware, and the middleware must dynamically generate an entirely new HTML page to be retuned to the browser for every data transfer.

Flash MX also allows the passing of data between client and server using XML (extensible markup language) (Jacobson & Jacobson, 2002; Weiss, 2004). Flash can be programmed to recognize and parse XML data into meaningful graphical content in the movie. Flash MX 2004 Professional supplies a tailored XML connector component that provides the application with access to an external data source that returns or receives XML (Webster, 2003). Macromedia has also designed their own markup language. The Macromedia Flex Markup Language, MXML, is an XML-based markup language that also incorporates the logic ability of ActionScript (Coenraets, 2003). XML lays out user interface components while ActionScript handles the interaction.

Flash Remoting is basically the Flash Client–Middleware–Data Source model, where the data source can be a database, legacy system, or some other messaging system. However, it is a stand-alone product utilizing a proprietary protocol known as Action Media Format, AMF. Flash Remoting is primarily used in enterprise-level applications (Webster, 2003).

Flash can integrate with Web services (Weiss, 2004), which can provide the same functionality as middleware, but return information in the form of an XML document termed a Web Service Description Language, WDSL. Web services are usually free and are becoming more prominent. Google and Amazon are common examples. Flash MX 2004 Professional supplies Web service connector components.

Macromedia Generator was a server-based tool that accepted data and dynamically returned a Flash movie to the client. It has been phased out due to middleware and Macromedia's current design philosophies. Macromedia has introduced Flash Communication Server MX. This is a server side technology that combines streaming with real-time animation, messaging, audio, and video (www.macromedia.com/macromedia/proom/pr/2002/flash_com_mx.html).

Flash MX Communication Server also allows Remoting (Lozbin, 2002). In this scenario the Communication Server is positioned between the Flash client and the middleware. This provides a more streamlined, less complex, optimized enterprise application.

Finally, Macromedia has introduced Central (Lynch, 2003), an application that runs on the desktop and allows the managing of information obtained from distributed sources outside the browser. It is mentioned here because Central provides Flash developers with a pre-built infrastructure for creating and distributing applications. Macromedia has provided a SDK (software development kit) for this purpose (www.macromedia.com/support/centraldev/).

FUTURE TRENDS

With wireless becoming more affordable and accepted, we may see more of an effort to offer higher quality Flash movies and interactivity on handheld wireless devices. Macromedia has worked tirelessly in making Flash communicate with other Web technologies. We can always expect this trend to continue. With the introduction of Central, we hope to see more desktop interactivity and development. Finally, incorporating video into a Flash movie is another important aspect of the application. We can expect more video intensive development for Flash.

CONCLUSION

The vastness of this application cannot be understated. A person or team needs to be devoted almost exclusively to this application if the full potential is to be realized. It is an exciting cutting-edge application that combines all elements of the Web into a content-rich multimedia environment. With greater emphasis being placed on content as well as presentation, Flash has kept pace and provides some benefits not realized in static environments. As the Web continues to evolve and expand, new challenges and issues will arise that Flash must expand to incorporate.

REFERENCES

Bardzell, S. & Bardzell, J. (2004). *Studio MX 2004, training from the source*. Berkeley, CA: Macromedia Press.

Castro, E. (2001). *Perl and CGI for the World Wide Web* (2nd edition). Berkeley, CA: Peachpit Press.

Chambers, M. (2002). *An overview of client/server interaction using Macromedia Flash and databases*. Retrieved November 28, 2003, from *www.macromedia.com/devnet/mx/flash/articles/flash_databases.html*.

Coenraets, C. (2003). *An overview of MXML, the Macromedia flex markup language*. Retrieved November 28, 2003, from *www.macromedia.com/devnet/flex/ articles/paradigm.html*.

Hall, B. & Wan, S. (2003). *Object-oriented programming with ActionScript* (2nd edition). Boston: New Riders.

Jacobson, D. & Jacobson, J. (2002). *Flash and XML, a developer's guide*. Boston: Addison-Wesley.

Knuckles, C. & Yuen, D. (2005). *Web applications, concepts & real-world design*. New York: John Wiley & Sons.

Lozbin S. (2002). *Understanding Macromedia Flash communication server remoting*. Retrieved November 28, 2003, from *www.macromedia.com/devnet/mx/flashcom/articles/coding_guidelines.html*.

Lynch, K. (2003). *Internet applications a new frontier opportunity*. Retrieved November 29, 2003, from *www.macromedia.com/software/central/whitepaper/central_wp.pdf*.

Mohler, J. (2001). *Flash 5 graphics, animation, and interactivity*. Albany: OnWord.

Muller, R. (2003). *Developing Web sites with Macromedia Flash MX*. Boston: Thompson.

Statler, T. (2003). *Migrating from Macromedia Flash MX to Flash MX 2004*. Retrieved November 28, 2003, from *www.macromedia.com/devnet/mx/flash/articles/migrate_flashmx2004.html*.

Webster, S. (2003). *Choosing between XML, Web services, and remoting for rich Internet applications*. Retrieved November 28, 2003, from *www.macromedia.com/devnet/mx/flash/articles/ria_dataservices.html*.

Weiss, N. (2004). *Flash MX Professional 2004 for server geeks*. Boston: New Riders.

Williams, M. (2002). *ActionScript coding standards*. Retrieved November 29, 2003, from *www.macromedia.com/devnet/mx/flash/whitepapers actionscript_standards.pdf*.

KEY TERMS

Components: Small pre-defined program snippets within Flash that allow reuse of elements. Examples are buttons, pull-down menus, text boxes, and text areas.

European Computer Manufacturers Association (ECMA): Standardizes information and communication systems.

Flash Player: A multiple-platform client that Web users must download and install (a browser plug-in) in order to view and interact with Macromedia Flash content.

Middleware: A server side scripting technology used to facilitate data transfer and communication between a client and a data source.

Software Development Kit (SDK): A set of development tools that allows a developer to create applications for a certain software package, hardware platform, or operating system.

Tween: A term coined at Walt Disney. The master animator creates the key animation frames of a sequence and the apprentice creates all the in between (or tween) frames.

Web Service: Data provided over the Internet, usually in a WSDL (Web service description language), meant as communication between computer software.

Management Considerations for B2B Online Exchanges

Norm Archer
McMaster University, Canada

INTRODUCTION

Information systems that link businesses for the purpose of inter-organizational transfer of business transaction information (inter-organizational information systems, or IOIS) have been in use since the 1970s (Lankford & Riggs, 1996). Early systems relied on private networks, using electronic data interchange (EDI) or United Nations EDIFACT standards for format and content of transaction messages. Due to their cost and complexity, the use of these systems was confined primarily to large companies, but low-cost Internet commercialization has led to much more widespread adoption of IOIS. Systems using the Internet and the World Wide Web are commonly referred to as B2B (business-to-business) systems, supporting B2B electronic commerce.

Technological innovations have led to several forms of B2B Internet implementations, often in the form of online exchanges. These are virtual marketplaces where buyers and sellers exchange information about prices, products, and service offerings, and negotiate business transactions. In addition to substituting proprietary lines of communication, emerging technologies and public networks have also facilitated new business models and new forms of interaction and collaboration, in areas such as collaborative product engineering or joint offerings of complex, modularized products. During the years 1999-2001 a number of online exchanges were introduced, but many of these failed (Gallaugher & Ramanathan, 2002), due mainly to an inability to attract participating business partners. Those that have survived are often owned by companies or consortia that are also exchange customers or suppliers.

The objective of this overview is to describe the evolution and the characteristics of B2B Internet implementations, and to discuss management considerations, the evaluation and adoption of B2B applications, and the technical infrastructure supporting these systems. We also indicate some of the open issues that remain as the technology and its adoption continues to evolve.

BACKGROUND

Although there are many classification schemes available for online exchanges (Choudhury, 1997; Kaplan & Sawhney, 2000), we will use a more generic and functional focus, with three categories: sell-side, buy-side, and neutral/market-type applications (Archer & Gebauer, 2001). Early B2B sell-side applications featured online catalogs, made available to the Internet community by distributors and manufacturers, often complemented by features such as shopping baskets and payment functionality. Many now provide customized and secure views of the data, based on business rules from contract agreements with individual customers. In some cases, buying processes of the customers are supported, including features such as approval routing and reporting. While some sophisticated applications exist to support collaborative forecasting or the configuration of complex products, many sell-side systems handle only the simpler transactions, such as maintenance, repair, and operation (MRO) supplies. Recently, more advanced features have become more widely available, such as CPFR (collaborative planning, forecasting, and replenishment) to support joint initiatives between customer and supplier (Holmstrom, Framling, Kaipia & Saranen, 2002).

Buy-side applications support procurement, moving order processes closer to the end user, and alleviating structured workloads in functional departments such as purchasing and accounts payable. For smaller companies, an affordable alternative is to work through hosted solutions, using Internet browsers to access procurement functionality provided by a third-party vendor or application service provider (ASP). Some applications provide functionality beyond the automation of highly structured procurement processes, including production tendering, and multi-step generation requests for proposals, as they are relevant for the procurement of freelance and management services. Interfacing purchasing systems to internal systems such as enterprise resource planning systems (ERP) makes it possible to automate many transactions, thus greatly increasing processing speed and reducing costs. Buy-side solutions that involve long-term inter-organizational relationships are typically set up by the purchasing organization, which then controls catalog content, data format, and back-end system functionality. Benefits include a reduction in maverick buying, and freeing purchasing and accounts payable personnel from clerical work to handle more strategic tasks. Suppliers typically benefit from long-term relationships, and in many cases the relationships between the buyer and its suppliers were in place before the buy-side operation was established.

The third group of applications, often referred to as B2B electronic markets or hubs, can either bring together multiple buyers and sellers on an ad hoc basis involving various types of auctions, or support more permanent relationships (a many-to-many relationship, equivalent to IOIS). Those that have been more successful are likely to have been sponsored by a consortium (e.g., GlobalNetXchange, in the retail industry, sponsored by buying organizations, and Global Healthcare Exchange in the health care industry, sponsored by selling organizations). They may feature auctions, electronic catalogs, and auxiliary value-added functions, such as industry news and online forums. The initiator typically controls the catalog content, aggregates supplier input, and provides additional functionality and standardized data access to buyers. These marketplaces may eliminate the need for market participants to link directly to their business partners, circumventing costly value-added EDI network services. Their business models typically include service charges based on transaction volume and setup costs. They provide a standard for suppliers to deliver catalog content, increase flexibility if they support access to suppliers and customers outside pre-established relationships, and create customer value through competitive pressure. Participation in such marketplace solutions may also provide a low-cost alternative for SMEs (small and medium enterprises).

MANAGEMENT CONSIDERATIONS

A market assumes an intermediary role that supports trade between buyers and suppliers, including (Bailey & Bakos, 1997): a) matching buyers and sellers, b) ensuring trust among participants by maintaining a neutral position, c) facilitating market operations by supporting certain transaction phases, and d) aggregating buyer demand and seller information. Supporting the marketplace through an electronic exchange has characteristics of (Bakos, 1991): 1) cost reductions, 2) increased benefits with the number of participants, 3) potential switching costs, 4) capital investments but economies of scale and scope, and 5) significant uncertainties in benefits. Many of the management issues of B2B electronic commerce systems relate to the need to coordinate decisions and processes among multiple firms, often through differences in business processes, information systems, business models, and organizational cultures.

Early transaction cost theory recognized markets and hierarchies as the two main methods of governance for coordinating flows of goods and services. Markets such as stock exchanges coordinate the flow through supply and demand forces, with price as the main coordination vehicle. Hierarchies such as production networks consist of predetermined relationships among customers and suppliers, and rely on managerial decisions to coordinate flows. There are many intermediate forms of governance, such as network organizations and strategic alliances (Gulati, 1998). A common theme among all these governance structures is collaboration among the participants, but the level of collaboration varies. These levels can be described as cooperation, coordination, and collaboration (Winer & Ray, 1994). In cooperation, there is little sharing of goods, services, or expertise; coordination requires mutual planning and open communication among participants, who share resources; collaboration involves deeply synergistic efforts that benefit all parties. Collaboration at different levels between buyers and sellers is emphasized by online exchanges, but this can also take place among buyers and among sellers (Wang & Archer, 2004).

In recent years there has been a "move to the middle," with a growth of outsourcing arrangements and more cooperative, integrated long-term inter-organizational relationships with a relatively small number of preferred suppliers (Clemons, Reddi & Row, 1993). Distribution of market power is often an overriding factor. For example, auto manufacturers, as a concentrated industry, will be likely to adopt an approach that involves long-term collaborative relationships among business partners rather than the short-term market-driven relationships that traditionally characterized this industry. On the other hand, companies in fragmented industries such as construction are characterized by short-term relationships and low levels of trust, where transactions such as online procurement are more likely to be through B2B tendering and auctions (Stein, Hawking & Wyld, 2003).

EVALUATION AND ADOPTION

The task of evaluating an electronic exchange becomes difficult when network effects are taken into account (benefits from participating are usually positively related to the number of participants). As a result of complications such as these (strategic necessity, dependence on the commitment of business partners, additional risk, external effects, etc.), the evaluation of an electronic exchange is much more complex than for systems deployed within organizations (Gebauer & Buxmann, 2000). B2B market mechanisms focus on four factors that favor one market mechanism over another: degree of fragmentation, asset specificity, complexity of product description, and complexity of value assessment (Mahadevan, 2003). These have a significant impact on the choice of an appropriate market mechanism for B2B interactions.

From the organizational perspective of setting up a successful B2B application, there are initiators and (potential) participants. Initiators bear the majority of the cost and risk, but on the other hand also enjoy the majority of the benefits, and they typically decide on technology infrastructure, type of systems used, corporate identity, representation of partners, and selection of participants. Success of the system depends on the participation of a critical mass of business partners. Supplier participation considerations in a buy-side solution, for example, include investments necessary to prepare and upload catalog data, integration with back-end systems, training of staff, and adjustments of business processes. Depending on individual arrangements, benefits include reduced time and costs for order processing, improved customer service, increased customer reach in a globalized marketplace, and an increase in revenues from long-term and trusted customer relationships. Neutral intermediaries in such markets face a difficult balancing task, as they have to be careful to satisfy suppliers as well as buyers, and a business model must be chosen that will attract the desired participants.

Although B2B exchanges have received a great deal of attention in the press and among researchers, their rate of adoption by business has not been high. While their aggregate transaction growth rate is higher, they are still (in 2004) outranked by at least a factor of 10 in transaction volume by EDI installations, which are still firmly in place in many large corporations (Jakovljevic, 2004). In addition, by utilizing EDI over the Internet, companies and organizations also benefit from the ability to facilitate a seamless bridging between XML and EDI that can now coexist on the same infrastructure and use common protocols to handle electronic procurement, invoicing, and logistics information. With the need to electronically exchange volume-intensive catalogue and product specification information, organizations can significantly reduce the high cost of this exchange by using Internet EDI. Although EDI is able to support only text content, the combination with XML provides support for images and graphics (Hamdar, 2002).

SMEs can and often do handle B2B transactions through e-commerce solutions without fully automated transaction management systems, including hosted procurement applications. Supply-side solutions with Web access can also be used as parallel and partially automated channels for larger businesses that wish to deal with small suppliers or customers. In practice, SMEs execute small numbers of transactions and may not wish to make the investment in resources, training, and internal integration required to link to their business partners (Archer, Wang & Kang, 2003). Motivations for joining online exchanges include (Gebauer & Raupp, 2000): a) coercion (through market power), b) long-term commitment to business relationships and reduction of associated uncertainty, c) subsidies to support system installations for potential partners, and d) general system improvements that result in improved efficiencies and effectiveness. SMEs that link to online exchanges are most likely to be motivated through pressure from their larger partners and by long-term commitments. Many use alternative interactions utilizing a combination of manual and online functions. For example, a medium-sized value-added retailer might use ad hoc purchasing procedures such as searching the Web for catalogue information on major suppliers, and then use the telephone to negotiate prices and delivery schedules (Archer et al., 2003).

TECHNICAL INFRASTRUCTURE

Software products to support B2B interactions are continuing to mature as more complex functions are added, such as collaborative planning, forecasting, replenishment, negotiation and decision support, and procurement and asset management of complex and highly customizable items and systems (Paul et al., 2003). Linking data from many different sources, including legacy systems, through Web services (Iyer, Freedman, Gaynor & Wyner, 2003) is still in its infancy, but supported by diffusion of industry standard Extensible Markup Language (XML). Technical issues are complicated by the critical role of security and confidentiality in inter-organizational settings, particularly when using public networks such as the Internet as compared to the private networks that were traditionally used for EDI implementations. Meeting these needs may require significant investments in software, training, business process reengineering, technical support, and time, all of which favor larger organizations.

Complexity of transactions is an important adoption factor, and is determined by factors such as the number of sub-processes and organizational units that are involved, as well as their possible interactions, interdependencies and relationships with the process environment (Gebauer & Buxmann, 2000). This in turn depends on the type of goods or services. Acquiring indirect or non-production supplies and services is the least complex type of transaction, followed by direct goods, and capital goods and other types of ad hoc purchases tend to be the most complex. A high degree of automation is economically viable only for high volume, less complex transactions. As complexity increases and volume decreases, human intervention is more likely to be needed to handle exceptions and ad hoc transactions.

B2B applications can have major impacts on interorganizational business processes, depending on the level of IOIS integration required (Stelzer, 2001). After planning and designing the system business model and infrastructure, a careful plan of how to implement it, how to train employees, and how to adapt business processes is the next step towards a successful project (Archer & Gebauer, 2001). Partner adoption, catalog management, and integration with a heterogeneous system of back-end applications are frequently listed as major stumbling blocks. For example, an organization could start out by reengineering and then automating an inefficient process that causes long lead times and possibly frequent complaints, such as management approval of end-user requests. As a next step, putting together an online catalog that contains the offerings of preferred suppliers can be useful as a first step to reduce "maverick" buying outside pre-established contracts. While the exact steps will depend on the situation within the individual firm, the stepwise approach will also allow frequent adjustments during the project planning and implementation process, including the addition of new requirements. The adoption of a B2B e-commerce solution is a strategic company decision and it is important to evaluate the potential overall impact of this innovation on the firm before proceeding (Pant & Hsu, 1996), as it may require substantial reengineering before it will be effective (Maull, Childe, Smart & Bennett, 1995).

FUTURE TRENDS

There is little doubt that rapid growth will continue in the relative value of B2B transactions handled through electronic commerce solutions, especially as they become less costly and easier to implement in SMEs. However, most of the growth in such offerings is likely to be in sell-side or buy-side online exchanges, as there has been significant supplier resistance to participating in neutral market-type exchanges. Although these exchanges offer the greatest theoretical benefit because of the potential for standardized linkages among participating companies, and the collaborative functionalities offered by such systems, this does not outweigh resistance from suppliers who see declining profit margins due to transaction cost payments and competitive price bidding. Meanwhile, EDI systems continue to link many large companies, due to their reluctance to give up related investments or to change business processes to accommodate the newer solutions.

CONCLUSIONS

A clear understanding of the possibilities of emerging technologies is crucial to take advantage of new opportunities in the B2B marketplace. There have been failures in this environment due to a lack of consideration of the wide range of technical, managerial, and economic issues involved. No widely adopted frameworks have been developed to assist in the choice of level of integration or in reengineering boundary spanning business processes. Although technology continues to develop, it is still immature in many areas. In particular, the integration with current IT infrastructures is often extremely complex and difficult to justify for medium to low transaction rates. Meanwhile, B2B applications continue to evolve, changing the rules of the game in subtle ways, but providing fruitful areas for research and new developments.

REFERENCES

Archer, N., & Gebauer, J. (2001). B2B applications to support business transactions: Overview and management considerations. In M. Warkentin (Ed.), *Business-to-business electronic commerce: Challenges and solutions* (pp. 19-44). Hershey, PA: Idea Group Publishing.

Archer, N., Wang, S., & Kang, C. (2003). *Barriers to Canadian SME adoption of Internet solutions for procurement and supply chain interactions*. MeRC Working Paper #5. Hamilton, ON: McMaster eBusiness Research Centre.

Bailey, J., & Bakos, Y. (1997). An exploratory study of the emerging role of electronic intermediaries. *International Journal of Electronic Commerce, 1*(3), 7-20.

Bakos, J.Y. (1991, September). A strategic analysis of electronic marketplaces. *MIS Quarterly, 15*, 295-310.

Choudhury, V. (1997). Strategic choices in the development of interorganizational information systems. *Information Systems Research, 8*(1), 1-24.

Clemons, E.K., Reddi, S.P., & Row, M.C. (1993). The impact of information technology on the organization of economic activity: The "move to the middle" hypothesis. *Journal of Management Information Systems, 10*(2), 9-35.

Gallaugher, J.M., & Ramanathan, S.C. (2002). Online exchanges and beyond: Issues and challenges in crafting successful B2B marketplaces. In M. Warkentin (Ed.), *Business to business electronic commerce: Challenges and solutions* (chap. III). Hershey, PA: Idea Group Publishing.

Gebauer, J., & Buxmann, P. (2000). Assessing the value of interorganizational systems to support business transactions. *International Journal of Electronic Commerce, 4*(4), 61-82.

Gebauer, J., & Raupp, M. (2000). Zwischenbetriebliche elektronische katalogsysteme: Netwerkstrategische gestaltungoptionen und erfolgsfaktoren (Interorganizational electronic catalogs: Strategic options and success factors) - in German. *Informatik Forschung und Entwicklun, 15,* 215-225.

Gulati, R. (1998). Alliances and networks. *Strategic Management Journal*, *19*(4), 293-317.

Hamdar, M. (2002, June). Catch-up: SMEs can benefit from Internet EDI strategy. *Purchasing B2B*.

Holmstrom, J., Framling, K., Kaipia, R., & Saranen, J. (2002). Collaborative planning forecasting and replenishment: New solutions needed for mass collaboration. *Supply Chain Management, 7*(3), 136-145.

Iyer, B., Freedman, J., Gaynor, M., & Wyner, G. (2003). Web services: Enabling dynamic business networks. *Communications of the Association for Information Systems, 11*, 525-554.

Jakovljevic, P.J. (2004, March 4). EDI versus XML: Working in tandem rather than competing? *TechnologyEvaluation.com, 4*.

Kaplan, S., & Sawhney, M. (2000). E-hubs: The new B2B marketplaces. *Harvard Business Review*, *78*(3), 97-100.

Lankford, W.M., & Riggs, W.E. (1996). Electronic data interchange: Where are we today? *Journal of Systems Management*, *47*(2), 58-62.

Mahadevan, B. (2003). Making sense of emerging market structures in B2B e-commerce. *California Management Review*, *46*(1), 86.

Maull, R.S., Childe, S.J., Smart, P.A., & Bennett, J. (1995). Current issues in business process reengineering. *International Journal of Operations and Production Management, 15*(11), 37-52.

Pant, S., & Hsu, C. (1996). Business on the Web: Strategies and economics. *Computer Networks and ISDN Systems, 28,* 1481-1492.

Paul, J., Withanachchi, S., Mockler, R.J., Gartenfeld, M.E., Bistline, W., & Dologite, D.G. (2003). Enabling B2B marketplaces: The case of GE Global Exchange Services. In M. Khosrow-Pour (Ed.), *Annals of cases on information technology* (vol. 5, pp. 464-486). Hershey, PA: Idea Group Publishing.

Stein, A., Hawking, P., & Wyld, D.C. (2003). The 20% solution?: A case study on the efficacy of reverse auctions. *Management Research News, 26,* 1-20.

Stelzer, D. (2001). *Success factors of electronic marketplaces: A model-based approach*. Ilmenau, Germany: Technische Universitat Ilmenau.

Wang, S., & Archer, N. (2004). Supporting collaboration in business-to-business electronic marketplaces (in press). *Information Systems and e-Business Management*.

Winer, M., & Ray, K. (1994). *Collaboration handbook: Creating, sustaining, and enjoying the journey*. Saint Paul, MN: Amherst H. Wilder Foundation.

KEY TERMS

Application Service Provider (ASP): An ASP is a service company that can support and relieve a firm from the daunting challenges of finding, hiring, inspiring and training technical personnel to manage an application in-house. An ASP provides software applications on a pay-per-use or service basis via the Internet and leased lines.

Collaborative Planning, Forecasting, and Replenishment (CPFR): CPFR is a global, open, and neutral business process standard for value chain partners to coordinate the various activities of purchasing, production planning, demand forecasting, and inventory replenishment, in order to reduce the variance between supply and demand and share the benefits of a more efficient and effective supply chain.

Electronic Data Interchange (EDI): A standard used to govern the formatting and transfer of transaction data between different companies, using networks such as the Internet. As more companies are linking to the Internet, EDI is becoming increasingly important as an easy mechanism for companies to share transaction information on buying, selling, and trading. ANSI (American National Standards Institute) has approved a set of EDI standards known as the X12 standards. Although not yet a global standard, because of EDIFACT, a standard developed by the United Nations and used primarily in non-North American countries, negotiations are underway to combine the two into a worldwide standard.

Enterprise Resource Planning (ERP): A business management system that can integrate all facets of the business, including planning, manufacturing, sales, and marketing, through a common database. As the ERP methodology has become more popular, software applications have been developed to help business managers implement ERP in business activities such as inventory control, order tracking, customer service, finance and human resources.

Extensible Markup Language (XML): Document type definitions that can be used to specify or describe various types of objects. When a set of these is used on the Web to describe product information, it is referred to as cXML or commerce XML. It works as a meta-language that defines necessary information about a product, and standards are being developed for cXML in a number of industries, performing a function similar to that of EDI for non-Web-based systems. It will help to standardize the exchange of Web catalog content and to define request/response processes for secure electronic transactions over the Internet. The processes include purchase orders, change orders, acknowledgments, status updates, ship notifications and payment transactions.

Inter-Organizational Information System (IOIS): (Sometimes referred to as an IOS). An automated information system, built around computer and communication technology, which is shared by two or more companies. It facilitates the creation, storage, transformation, and transmission of information across a company's organizational boundaries to its business partners.

Maintenance, Repair, and Operations (MRO): Supplies and services purchased for use internally in the company, often referred to as indirect or non-production supplies and services (such as office supplies, computer equipment and repairs, cleaning supplies, etc.). These tend to be low unit cost, low volume, and off-the-shelf purchases.

Small and Medium Enterprise (SME): The definition of small and medium enterprises varies from country to country. If the definition is based on number of employees, SMEs in the U.S. have from 1 to 499 employees. The dividing line between a small and medium business is variously defined as being either 50 or 100 employees.

Management of Cognitive and Affective Trust to Support Collaboration

Diane H. Sonnenwald
Göteborg University & University College of Borås, Sweden

INTRODUCTION

Cognitive trust focuses on judgments of competence and reliability, and affective trust focuses on interpersonal bonds among individuals and institutions. Both cognitive and affective trusts play an integral role in organizations and institutions that rely on collaboration among individual members to achieve their goals and realize their vision.

Collaboration is increasingly important in the knowledge-based economy, as well as in scientific research and development where no one individual has all the prerequisite knowledge and resources to solve complex problems, develop sophisticated products and services, or complete multi-faceted work tasks. Collaboration is not possible without cognitive or affective trust. Yet cognitive and affective trust may be more difficult to manage in organizations and teams that are geographically distributed (i.e., not physically collocated), because mechanisms, such as informal face-to-face interactions and observations that typically are used in building and maintaining trust, are not universally present. Previous research has shown that when organizations are geographically distributed, trust among members is negatively impacted (Handy, 1995; Jarvenpaa & Leidner, 1995; Rocco et al., 2001).

BACKGROUND

Definitions of Trust and Distrust

There are many definitions of trust arising from different disciplinary perspectives. When synthesizing these definitions, Rosseau et al. (1998) found that scholars fundamentally agree that trust is a "psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another" (p. 395). Trust involves risk (the probability of loss) and interdependence (reliance on others).

Distrust can be defined in opposite terms (i.e., negative expectations of the intentions or behavior of another) (Lewicki et al., 1998; Sztompka, 1999). It involves a lack of risk and no dependence on others. Trust and distrust can exist simultaneously in individuals (Lewicki et al., 1998). They can be conceptualized as a continuum with high trust to high distrust as endpoints; that is, a continuum from high trust to low trust to low distrust to high distrust.

Evolution of Trust and Distrust

Feelings of trust and distrust can change over time (Jones & George, 1998; McKnight et al., 1998). These changes occur as a result of observation of and reflection on behavior (Whitener et al., 1998). That is, trust and distrust are not behaviors, but psychological conditions that influence an individual's behavior.

An individual's behavior influences others' behaviors, both of which may be assessed by the individual (Figure 1) (Sonnenwald, 2003). This assessment is often based on prior experiences, knowledge of the context in which the behaviors occurred, and beliefs. The results of the assessment influence perceptions of trust and distrust, and future assessments (through the modification or reenforcement of prior experiences, knowledge of the context, and beliefs). Thus, trust and distrust shape one's own behavior and others' behavior, whose assessment in turn shapes trust and distrust.

Two Types of Trust and Distrust: Cognitive and Affective

Two types of trust—cognitive and affective—have been identified as important in organizations (McAllister, 1995; Rocco et al., 2001). Cognitive trust focuses on judgments of competence and reliability. Can a co-worker complete a task? Will the results be of sufficient quality? Will the task be completed on time? These are issues that comprise cognitive trust and distrust. The more strongly one believes the answers to these types of questions are affirmative, the stronger is one's cognitive trust. The more strongly one believes the answers to these types of questions are negative, the stronger is one's cognitive distrust.

Affective trust focuses on interpersonal bonds among individuals and institutions, including perceptions of colleagues' motivations, intentions, ethics, and citizenship. Affective trust typically emerges from repeated interactions among individuals, and experiences of reciprocated interpersonal care and concern (Rosseau et al., 1998). It is also referred to as emotional trust (Rocco et al., 2001) and relational trust (Rosseau et al., 1998). It can be "the grease that turns the wheel" (Sonnenwald, 1996).

Interaction among Cognitive and Affective Trust and Distrust

Cognitive trust and distrust may exist in conjunction with affective trust and distrust (Table 1). High cognitive and affective trust typically yields tightly coupled collaboration in which tasks and ideas are openly and frequently shared. Scientists talk of friendship and of liking each other when affective and cognitive trust is high. Risk and vulnerability caused by collaboration is perceived as low.

In comparison, high affective distrust and high cognitive distrust can be sufficient to dissuade individuals from collaborating at all. No friendship exists or develops, and individuals may proactively limit their interaction with others they cognitively and affectively distrust. Collaboration and interaction is perceived as high risk with a high degree of vulnerability.

EVERYDAY MANAGEMENT OF TRUST AND DISTRUST

A trust-distrust match between cognitive and affective trust can yield problematic situations that require explicit management. Feelings of high cognitive distrust and high affective trust will serve to limit collaboration. Primarily, non-critical or unimportant tasks will be given to individuals that are cognitively distrusted. However, friendship as a result of affective trust may exist or emerge. Controls to monitor task completion and support task completion efforts may be used. For example, mentoring and training may be employed to help a friend who is not cognitively trusted.

Feelings of high cognitive trust and high affective distrust can result in competitive collaboration, which can be managed through discussions that identify issues and perceptions. Specific data should be presented and goodwill expressed to counter perceptions. Solutions include changes in work plans and information, and equipment sharing. Controls to monitor and constrain task or work activities can be employed to manage affective distrust. The saying, "Keep friends close and enemies closer" appears applicable in these types of situations. Professional relationships may exist or emerge, but friendship and the perception that the collaboration or interaction is fun may never emerge. Affective distrust can be reduced or accommodated, but may not disappear for some time.

THE ROLE OF INFRASTRUCTURE IN MANAGING TRUST

Organizational Structure

Cognitive and affective trust throughout an organization or team are implicitly encouraged when management exhibits high levels of cognitive and affective trust towards each other and members of the organization. The example provided by leadership sets expectations for others.

To encourage trust and to help resolve issues regarding trust, a management team that includes a site coor-

Figure 1. The evolutionary nature of trust and distrust

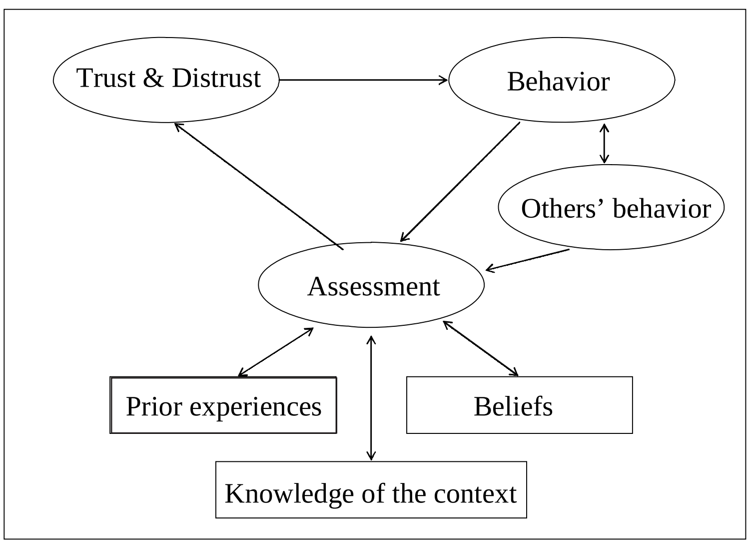


Table 1. Relationships among cognitive and interpersonal trust and distrust

		Cognitive: Trust	Cognitive: Distrust
Affective	Trust	Tightly-coupled collaboration Frequent collaboration Friendship Low risk	Limited collaboration Limited, non-critical task responsibility Controls to monitor & support efforts Friendship
Affective	Distrust	Competitive collaboration Controls to monitor & constrain activities "Keep friends close & enemies closer" Professional relationship	No collaboration No friendship Limited interaction High risk

dinator for each participating location, as well as coordinators for specific activities, can be established. Site coordinators can handle location-specific administrative issues, ranging from reserving a videoconference room for weekly meetings to distributing allocated budget funds. The participation of representatives from each physical location provides ongoing dialog about challenges, progress, perceptions, and ways of working at each location, which are important for building and maintaining cognitive trust. It facilitates learning about different ways of working and collaborative problem solving when members from different locations suggest how practices at their location may solve problems at other locations.

Similarly, coordinators for specific types of activities, domains, or areas of specialization help to manage cognitive trust among these diverse domains. For example, such coordinators can provide feedback to other team members and coordinators regarding the validity and reliability of the research or work methods proposed by members in the same domain, helping to establish cognitive trust.

Power

Boulding (1990) describes three types of organizational power: destructive, economic, and integrative. Destructive power—the power to destroy things—can be used as a prelude to production, where things are destroyed or altered to make way for production and for carrying out a threat. An example of destructive organizational power is the firing of employees who are seen as resisting change in an organization. Economic power is used in all organizations. It involves the creation and acquisition of economic goods, including intellectual property, through production, exchange, taxation, or theft. Integrative power involves the capacity to build organizations, inspire loyalty, bind people together, and develop legitimacy. It has a productive and destructive aspect. In a negative sense, it can create enemies and alienate people. All organizations have some integrative power or they could not survive. Some, however, rely on integrative power more than others; these include religious organizations, political movements, volunteer organizations, and clubs.

Successful organizations and teams that must rely on collaboration among distributed members appear to use a combination of integrative, economic, and destructive power; however, a primary type of power used in successful teams and organizations appears to be integrative. For example, organizations and teams can use integrative power to develop their vision, mission, and goals. Integrative power is also a mechanism to increase cognitive trust. For example, Hart and Saunders (1997) discuss the use of integrative or persuasive power in building trust across organizations in the context of electronic data interchange adoption. Tucker and Panteli (2003) found the use of integrative power positively related to trust in a study of 18 virtual teams within a global, high-tech company.

In any organization, destructive or economic power is used when members do not meet expectations or keep commitments (i.e., when cognitive distrust emerges). Such decisions, however, may be best reached through the use of integrative power and are based on cognitive trust among team members.

Information and Communications Technology

When a team or organization is geographically distributed, it must utilize information and communications technology (I&CT) as a mechanism to realize its vision and mission, or to incur expensive monetary and temporal travel costs. Typically, this has meant using traditional information and communications technology, such as the telephone, fax, file transfer, mail, and e-mail in ways typical of other R&D organizations and scientific disciplines (Daft & Lengel, 1984). It has also meant using newer technologies, such as video conferencing and Web pages, in innovative ways to support the organization's vision, facilitate collaboration, and manage cognitive and affective trust and distrust. Tucker and Panteli (2003) found that distributed teams with a high degree of trust among team members regularly communicated using synchronous IC&T, such as telephony and video conferencing.

New social protocols to compensate for constraints imposed by video-conferencing technology, and operations protocols to help reduce technical problems are needed and can be developed and implemented by working with team members and technical staff (Sonnenwald et al., 2002). For example, one drawback to the video-conference meetings is their inherently formal nature, if they are held in a specially equipped meeting room or if they require members to prepare for the meeting in new ways, such as creating PowerPoint™ slides, instead of writing on a whiteboard. This formality can negatively impact affective trust.

Several things can reduce the formality and increase the interactive nature of these meetings. Managers and key members can introduce informal aspects into their presentations (e.g., use the drawing features of the electronic board to modify their slides in real time). In addition, new practices of having non-work communication before a presentation can be initiated. Interpersonal communication has also been shown to increase affective trust among distributed team members (Rocco et al., 2000) and facilitate collaboration (Sonnenwald, 1996), a byproduct of cognitive trust.

Asynchronous technology may also facilitate trust building through documenting and sharing general information. For example, a Web site can be created to share news, expectations, and resources among team or organization members and to communicate information about the team to stakeholders. The content of the Web site should evolve to meet the needs of the team. Examples of information that could be included in a Web site include the group's vision statement, contact information, annual reports, call for proposals, virtual tours of lab facilities, meeting schedules, member directory, personal member Web pages, a news bulletin with press releases and announcements of awards and other recognitions received by members, and forms to be used by members (e.g., a confidentiality agreement). This type of content can help form a shared identity and affective trust across distances (Rocco et al, 2000).

The Web site could also contains pointers to resources that provide work, career, and personal assistance to members, such as information about suppliers, conferences, job interview processes, and apartment hunting services. This type of information supports a general anonymous mentoring function, allowing members to anonymously find information to assist in their careers and personal life. Thus, these pages have the possibility of positively influencing feelings of cognitive and affective trust toward the team or organization.

FUTURE TRENDS

Additional research is needed to increase our understanding of trust across distances. Issues to investigate include the sustainability of trust over time across distances, and the long-term consequences of distrust. Emerging IC&T should also be evaluated with respect to its roles, both positive and negative, on the development of trust. IC&T that provides a higher degree of situation awareness (Sonnenwald et al., 2004) may help facilitate the development of trust across distances.

CONCLUSION

Trust management is important in any organization, but it is especially important in organizations and teams that are geographically distributed and dependent on collaboration among members in different locations. For example, when organizations need to address large, complex, and challenging problems in which collaboration among experts (irrespective of discipline, department or institution affiliation) is required, cognitive trust and affective trust are both necessary to create a shared working understanding and new knowledge.

Typical trust mechanisms (i.e., informal face-to-face interactions and observations) to build and maintain trust are not inherently present in these organizations. Infrastructure that explicitly supports the creation and maintenance of trust appears vital. Participation in management by representatives at each physical location, early and continuing dialog between organization members and management, and use of integrative power, are three infrastructure mechanisms to manage cognitive trust.

Information and communications technology is a necessity in any geographically distributed organization today, and yet it can limit the creation and maintenance of cognitive and affective trust. However, interactive video-conferencing and Web sites that share organizational information and news, in conjunction with changes in practices, may help overcome the inherent limitations of I&CT to manage trust. Future I&CTs that incorporate advanced features (e.g., opportunistic floor control that allows synchronous execution of commands) and provide higher quality and control of video- and audio-conferencing, as well as better integration with other commonly used applications, hold the promise of facilitating collaboration to a greater extent than today's technology (Sonnenwald et al., 2004). However, it should be noted that face-to-face interaction is currently recommended to augment interaction mediated by technology (Handy, 1995; Olson & Olson, 2001; Rocco, et al., 2000), and a lack of face-to-face interaction may limit the growth of affective trust.

Tightly coupled collaboration appears to only emerge in situations where highly cognitive and affective trust simultaneously exists. No collaboration will emerge in situations where highly cognitive and affective distrust exists. Limited collaboration emerges when affective trust and cognitive distrust exist concurrently. Non-critical work tasks may be given to individuals in these situations, and controls to monitor and support task completion may be utilized. In comparison, competitive collaboration emerges when cognitive trust and affective distrust exist concurrently. Controls to constrain work activities may emerge to manage affective distrust. Limited and competitive collaboration may be manageable, but neither situation is ideal.

ACKNOWLEDGEMENTS

This chapter is based on work supported by the STC Program of the National Science Foundation under Agreement No. CHE-9876674.

REFERENCES

Boulding, K. (1989). *The three faces of power*. Newbury Park, CA: Sage.

Daft, R.L., & Lengel, R.H. (1984). Information richness: A new approach to managerial behavior and organizational design. In L.L. Cummings & B.M. Shaw (Eds.), *Research in organizational behavior* (pp. 191-233). Greenwich, CT: JAI Press.

Duarte, D., & Snyder, N. (1999). *Mastering virtual teams.* San Francisco: Jossey-Bass Publishers.

Handy, C. (1995, May-June). Trust and the virtual organization. *Harvard Business Review*, 43-50.

Hart, P., & Saunders, C. (1997). Power and trust: Critical factors in the adoption and use of electronic data interchange. *Organization Science*, *8*(1), 23-42.

Jarvenpaa, S., & Leidner, D. (1999). Communication and trust in global virtual teams. *Organization Science, 10*(6), 791-815.

Jones, G., & George, J. (1998). The experience and evolution of trust: Implications for cooperation and teamwork. *Academy of Management Review, 23*(3), 531-546.

Kanter, R. (1994). Collaborative advantage: The art of alliances. *Harvard Business Review, 72*(4), 96-109.

Lewicki, R., McAllister, D., & Bies, R. (1998). Trust and distrust: New relationships and realities. *Academy of Management Review, 23*(3), 438-458.

McAllister, D. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, *38*(1), 24-59.

McKnight, D., Cummings, L., & Chervany, N. (1998). Initial trust formation in new organizational relationships. *Academy of Management Review, 23*(3), 479-490.

Olson, G.M., & Olson, J.S. (2000). Distance matters. *Human-Computer Interaction*, *15*(2-3), 139-178.

Orlikowski, W. (1993), Learning from notes: Organizational issues in groupware implementation. *The Information Societ*y, *9*(3), 237-250.

Rocco, E., Finholt, T., Hofer, E., & Herbsleb, J. (2000). Designing as if trust mattered. *CREW Technical Report*, University of Michigan.

Rocco, E., Finholt, T., Hofer, E., & Herbsleb, J. (2001). Out of sight, short of trust. *Proceedings of Founding Conference of the European Academy of Management*, Barcelona, Spain.

Rousseau, D., Sitkin, S., Burt, R., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review, 23*(3), 393-404.

Schroeder, R., & Axelson, A. (2000). Trust in the core: A study of long-term users of activeworlds. *Proceedings of Digital Borderlands, a Cybercultural Symposium*, Norrkoping, Sweden.

Sonnenwald, D.H. (1996). Communication roles that support collaboration during the design process. *Design Studies*, *17*, 277-301.

Sonnenwald, D.H. (2003). Managing cognitive and affective trust in the conceptual R&D organization. In M. Iivonen & M. Houtari (Eds.), *Trust in knowledge management and systems in organizations* (pp. 82-106). Hershey, PA: Idea Publishing.

Sonnenwald, D.H., Maglaughlin, K.L., & Whitton, M.C. (2004). Designing to support situational awareness across distances: An example from a scientific collaboratory. *Information Processing and Management, (40)* 6, 189-1001.

Sonnenwald, D.H., Solomon, P., Hara, N., Bolliger, R., & Cox, T. (2002). Collaboration in the large: Using video conferencing to facilitate large group interaction. In A. Gunasekaran, O. Khalil, & S. Rahman (Eds.), *Knowledge and information technology in 21st century organizations: Human and social perspectives* (pp. 115-136). Hershey, PA: Idea Group Publishing.

Sztompka, P. (1999). *Trust: A sociological theory*. Cambridge, UK: University Press.

Teasley, S., Covi, L., Krishnan, M.S., & Olson, J.S. (2000). How does radical collocation help a team succeed? *Proceedings of the Computer Supported Cooperative Work Conference* (pp. 339-346). New York: ACM Press, 339-346.

Tucker, R., & Panteli, N. (2003). Back to basics: Sharing goals and developing trust in global virtual teams. In M. Korpela, R. Montealegre, & A. Poulymenakou (Eds.), *Organizational information systems in the context of globalization*. London: Kluwer.

Whitener, E., Bridt, S., Korsgaard, M., & Werner, J. (1998). Managers as initiators of trust: An exchange relationship framework for understanding managerial trustworthy behavior. *Academy of Management Review*, *23*(3), 513-530.

KEY TERMS

Trust: A "psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another" (Rosseau et al., 1998, p. 395).

Distrust: Negative expectations of the intentions or behavior of another (Lewicki et al., 1998; Sztompka, 1999).

Cognitive trust: Judgments regarding a person's competence and reliability.

Affective trust: Interpersonal bonds among individuals and institutions, including perceptions of a person's motivation, intentions, ethics, and citizenship.

Collaboration: Human behavior that facilitates the sharing of meaning and completion of tasks with respect to a mutually shared goal; takes place in social or work settings.

Integrative power: The capacity to build organizations, to inspire loyalty, to bind people together, and to develop legitimacy (Boulding, 1990).

Destructive power: The power to destroy things. It can be used as a prelude to production in which things are destroyed or altered to make way for production, and for carrying out a threat (Boulding, 1990).

Economic power: The creation and acquisition of economic goods (including intellectual property) through production, exchange, taxation, or theft (Boulding, 1990).

Managing Hierarchies and Taxonomies in Relational Databases

Ido Millet
Penn State Erie, USA

INTRODUCTION

The need to maintain classification and retrieval mechanisms that rely on concept hierarchies is as old as language itself. Familiar examples include the Dewey decimal classification system used in libraries and the system for classifying life forms developed in the 1700s by Carolus Linnaeus. A more recent example is Yahoo's subject taxonomy.

Information technology has led to an explosive growth in digital documents, records, multi-media files, and Web sites. To facilitate end-user access to these resources, topic hierarchies are frequently maintained to allow intuitive navigation and searching for resources related to specific categories. Maintaining such document, Web site, and knowledge taxonomies within relational databases is not trivial since the management of any hierarchical data in relational databases poses significant challenges (Millet, 2001). Taxonomies pose an additional challenge due to the typical need to classify a single document, concept, or Web site under multiple topics and due to the typical reliance on intelligent keys (Millet, 2003).

While, according to some views, non-relational database technologies and dynamic classification schemes may offer better ways for achieving our objectives, the use of relational database technology and concept taxonomies remains a contemporary practice and challenge.

BACKGROUND

Consider a document database where each document is classified into a hierarchy of topics shown in Figure 1.

Since each document may belong to more than one parent topic, we cannot record the data for this hierarchy by specifying in each topic record the topic above it. Figure 2 shows a data model for this situation. Note that the classify table allows us to assign a single document to multiple topics. If instead of a topic hierarchy, we need to assign each subtopic to more than one parent topic, we would insert a topic assignment table to represent such taxonomies.

To demonstrate the difficulty of hierarchical data retrieval against the normalized data model in Figure 2, consider the following requests:

- Show a list of all documents (at all levels) under Topic 1.
- Show how many documents (at all levels) are classified under each topic at level 1 of the hierarchy.

Using SQL, we can easily join each topic to all the documents associated with it via the classify records. However, we cannot easily identify the documents indirectly associated with a topic via its subtopics (at all levels). This difficulty in locating parent or child nodes at any given level is at the heart of the problem.

SQL-BASED SOLUTIONS

A request to show how many documents belong to each main topic, including all subtopics below it, can be handled using the SQL:1999 (ANSI/ISO/IEC 9075-2-1999) query shown in Listing 1. This query starts by creating a table expression (TOPIC_PATHS) populated with all main topic records as parents of themselves and appends (UNION) records for all paths of length one from these nodes to the topics directly below them. The RECURSIVE option continues the process to build all indirect paths from each topic to all its descendants.

The query then joins the end points (Topic_Below) of all paths in the TOPIC_PATHS result set to the documents assigned to these topics. By limiting the start points of these paths to main topics (topics at level 1) and

Figure 1. A topic hierarchy

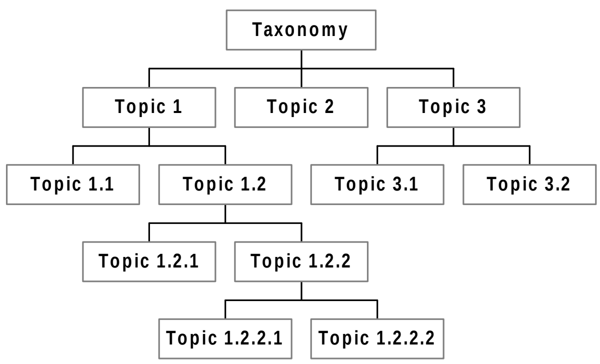
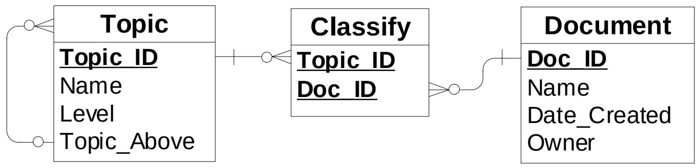


Figure 2. A normalized data model with a topic hierarchy



Listing 1. Recursive hierarchy retrieval using SQL:1999

```
WITH RECURSIVE TOPIC_PATHS (topic_above, topic_below, level) AS
(SELECT topic_id, topic_id, level FROM TOPIC
UNION ALL
SELECT TOPIC_PATHS.topic_above, TOPIC.topic_id, TOPIC_PATHS.level
FROM TOPIC_PATHS, TOPIC
WHERE TOPIC_PATHS.topic_below = TOPIC.topic_above)

SELECT TOPIC_PATHS.topic_above, DistinctCount (CLASSIFY.Doc_ID)
FROM TOPIC_PATHS, CLASSIFY
WHERE TOPIC_PATHS.topic_below = CLASSIFY.Topic_ID AND
TOPIC_PATHS.level = 1
GROUP BY TOPIC_PATHS.topic_above;
```

grouping the end result by those topics, we get the requested information. Avoiding double-counting of documents that were assigned to multiple topics is achieved by using *DistinctCount*(Classify.Doc_ID) rather than *Count*(Classify.Doc_ID).

Relying on such complex SQL is probably beyond the reach of many IT professionals. This can be addressed by implementing the complex portion of these SQL statements as database views. However, someone has to write the SQL for these views and the intensive nature of the required processing may lead to slow performance in reporting applications with large hierarchies and frequent queries.

Celko (2000) reports on a technique leading to significant improvements in query speeds by storing the hierarchy data not as parent-child relationships but as "nested sets" using a somewhat complex numbering scheme. However, alternative approaches can achieve very significant query performance gains while maintaining intuitive data storage and SQL syntax.

THE PATH TABLE APPROACH

The path table approach uses a "navigation bridge table" (Kimball et al., 1998) with records enumerating all paths starting from each node to all nodes in the branch above it, including itself. This approach provides flexibility in the sense that each subtopic node can belong to multiple direct parent topics and there is no limit on the number of levels in the taxonomy.

As demonstrated by Table 1, topic 1.1.1 would require four records in the path table reflecting the paths up to itself, topic 1.1, topic 1, and topic 0 (the top node of the hierarchy). These are just 4 of the 37 records required to capture all paths for the sample hierarchy in Figure 1.

To demonstrate how the path table can simplify data retrieval, consider the same challenge of showing how many documents belong to each main topic, including all subtopics below it. By joining the tables as shown in Figure 3, we can easily select all documents that belong to topics below each main topic. Since the path table includes a zero-length path between each topic and itself, documents that belong directly to a main topic would be included in the result set. Again, the classify table allows us to associate the same document with multiple topics.

Other requests for information can use the same approach or variations such as connecting to the topic table via the Topic_ID column in the path table or adding path length and terminal node information to the path table (Kimbal et al., 1998).

One limitation of the path table approach is that the number of records in the path table can grow quite large for deep hierarchies. The following section describes another approach that avoids that problem.

THE DENORMALIZED TOPIC TABLE APPROACH

The denormalized topic table approach maintains information about all higher-level topics within each topic

Table 1. A path table for the sample hierarchy

Topic_ID	Topic_Above
1.1.1	1.1.1
1.1.1	1.1
1.1.1	1
1.1.1	0

record. The classify table still allows each document to be assigned to multiple topics, but this approach restricts the topic taxonomy to a strict hierarchy whereby each sub-topic can have only one direct parent topic.

As demonstrated by Figure 4, if we assume that our topic hierarchy will not exceed six levels, we need six more columns to maintain this information. Each node will be indicated as its own parent at its own level. For example, topic 1.1.1 in Figure 1 would have 0 (the topic_id for the top node in the hierarchy) as its Topic_Level_1, 1 as its Topic_Level_2, 1.1 as its Topic_Level_2, 1.1.1 (itself) as its Topic_Level_4, and Null values for Topic_Level_5 and Topic_Level_6.

To demonstrate how the denormalized topic table can simplify data retrieval, consider again the challenge of showing how many documents belong to each main topic, including all subtopics below it. By joining the tables as shown in Figure 4, we can easily select and group all topics according to Topic_Level_2. Documents classified directly under that topic would be included in the result set because each topic is its own parent at its own level. Listing 2 shows that using this approach, a simple SQL statement can generate the requested information.

Note that in order to return the topic name we resorted to adding an aliased (Topic2) copy of the topic table to the SQL statement.

The denormalized topic table is less flexible since it limits the topic taxonomy to a strict hierarchy. However, the path table approach requires more complex joins due to the addition of one more physical table. Another disadvantage of the path table approach is that for deep hierarchies, the size of the path table can grow quite large, degrading query performance. A key issue with both approaches is the question of maintaining the redundant data required by these methods.

MAINTENANCE OF DENORMALIZED HIERARCHY DATA

This section suggests methods that can simplify and automate the maintenance of redundant hierarchy data. The proposed approach provides a concrete example of an "incremental evaluation system" for handling recursive SQL queries as advanced by previous literature (Dong et al., 1999; Libkin & Wong, 1997).

Because the denormalized topic table allows for simpler maintenance logic, the remainder of this article uses that approach as the assumed data structure.

A key observation is that the redundant information maintained for each topic is a simple extension of the redundant information already maintained for its parent topic. Consider a situation where Topic 1.2.2.2 in Figure 5 is moved from under Topic 1.2.2 to under Topic 3.2. The

Figure 3. A path table connects each classification with all its parents.

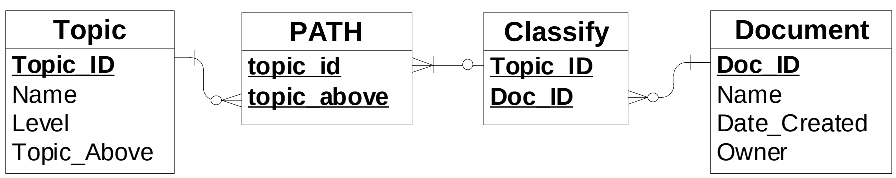
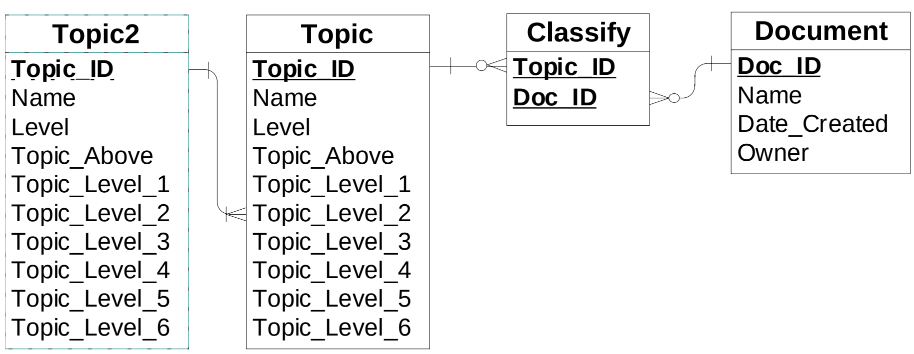


Figure 4. Using a denormalized topic table



Listing 2. Retrieval via a denormalized topic table

```
SELECT TOPIC2.Name, DistinctCount(CLASSIFY.Doc_ID)
FROM (TOPIC2 INNER JOIN TOPIC ON TOPIC2.topic_id = TOPIC.Topic_Level_2)
INNER JOIN CLASSIFY ON PATH.Topic_Level_2 = CLASSIFY.topic_id
GROUP BY topic.Topic_Level_2;
```

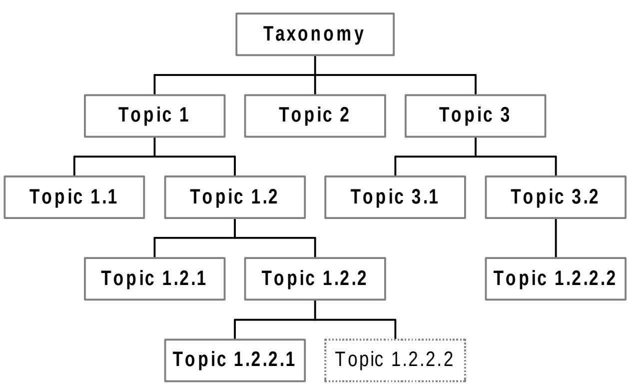
Topic_Level_N columns from the new parent node can simply be copied to the updated topic record (Topic 1.2.2.2). We then simply add the topic as its own parent at its own level and update the level column to the level of the parent node plus one.

Topic taxonomies (such as the Dewey decimal classification system used in libraries) frequently use "intelligent" primary keys reflecting the position of each topic within the taxonomy. In our example, we would neeed to change the primary key from 1.2.2.2 to 3.2.1. Using referential integrity, the database management system can automatically cascade such primary key updates to the foreign keys in documents as well as in Parent_Level_N columns.

The remaining challenge is to handle situations where a whole branch (a topic with subtopics) is moved in the hierarchy. Consider, for example, moving topic 1.2.2 in Figure 5 from under Topic 1.2 to under Topic 1.1. The procedural logic would update the redundant data for Topic 1.2.2, but we now need to update the information for all its decendants. The most elegant solution is to recursively extend the procedural logic by applying it to all descendant nodes, as if their Topic_Above column was updated as well.

Using this update logic, users would update the structure of the hierarchy by specifying only the Topic_Above information for each node. The maintenance of the redundant data can be done either under the control of the application or through database triggers.

Figure 5. Moving Topic 10 from Parent 6 to Parent 4



FRONT-END LOGIC OR DATABASE TRIGGERS

Each time a user changes the Topic_Above column, the application can call a function with the Topic_ID and its new Topic_Above as arguments. After completing the logic for that topic, the function would use embedded SQL to identify the descendants of the topic and call itself recursively against all these descendants.

One limitation of using such front-end logic to maintain the redundant hierarchy data is that it would require multiple communications between the client and the server. A much more important limitation is that we are dependent on uniform and full compliance by all client applications. The integrity of the hierarchy data can be compromised if some screens neglect to call or implement the update function appropriately.

As described in detail by Millet (2003), the update logic can be implemented via recursive Update and Insert triggers declared on the Topic_Above column of the TOPIC table. Moving the procedural logic from front-end functions to back-end triggers removes the threat of multiple points of failure and achieves better performance.

FUTURE TRENDS

The current trend in addressing the challenges of maintaining hierarchical data in relational databases is to provide more complex but more powerful data retrieval syntax. The SQL standard has added recursive queries. Similarly, proprietary SQL extensions such as the START WITH and CONNECT BY clauses provided by the Oracle DBMS are also useful mechanisms for retrieving hierarchy data from normalized data structures.

Clearly, the trend so far has been in the direction of maintaining the simplicity of the data structure at the expense of more complex queries. In the long term, I hope DBMS vendors realize the value of also supporting the opposite approach whereby queries are simplified at the expense of complicating the data structure. Declarative statements are all that is needed today to ask a DBMS to apply referential integrity logic to the data. In a similar way, we can envision a future where declarative statements identify to the DBMS the link between parent and child nodes and the Parent_Level_N columns that should be maintained by the DBMS.

CONCLUSION

Topic hierarchies are used to classify and search for documents, Web sites, and knowledge areas. Beyond the data retrieval and data maintenance challenges posed by any hierarchy domain, the need to support classifying the same document under multiple topic nodes leads to different table designs and requires care in avoiding biased aggregate results due to double counting. Since topic taxonomies frequently utilize intelligent rather than surrogate primary keys, updates to the location of topic nodes may require updates to the primary keys as well.

Fast and simple data retrieval against topic hierarchies can be achieved by maintaining redundant information. The approach of denormalizing the topic table may be preferred over maintaining a separate path table because the path table can grow too large and requires more complex data maintenance logic.

The limitation of denormalized and redundant hierarchy information is that updates to the hierarchy require special processing logic in order to avoid update anomalies. To efficiently refresh the hierarchy data we can exploit the redundant information already maintained for the specified parent topic as topic records are inserted or updated. The process can then be extended recursively for lower level nodes.

If the necessary trigger options are available for the DBMS in use, it is recommended that the processing logic for maintaining the redundant hierarchy information be implemented as triggers. This removes the burden of hierarchy maintenance from client applications. It also ensures that client applications cannot circumvent the hierarchy maintenance logic.

REFERENCES

ANSI/ISO/IEC 9075-2-1999, ANSI's Electronic Standards Store: *http://webstore.ansi.org*

Celko, J. (2000). *Joe Celko's SQL for Smarties: Advanced SQL programming.* San Francisco: Morgan Kaufmann.

Dong, G., Libkin, L., Su, J., & Wong, L. (1999). Maintaining the transitive closure of graphs in SQL. *Int. Journal of Information Technology, 5,* 46-78.

Kimball, R., Reeves, L., Ross, M., & Thornthwaite, W. (1998). *The data warehouse lifecycle toolkit: Expert methods for designing, developing, and deploying data warehouses.* New York: Wiley Computer Publishing.

Libkin, L., & Wong, L. (1997). Incremental recomputation of recursive queries with nested sets and aggregate functions. In *Database programming languages* (pp. 222-238). Springer.

Millet, I. (2001). Accommodating hierarchies in relational databases. In S. Becker (Ed.), *Developing complex database systems: Practices, techniques, and technologies.* Hershey, PA: Idea Group Publishing.

Millet, I. (2003). Managing document taxonomies in relational databases. In S. Becker (Ed.), *Effective databases for text & document management.* Hershey, PA: Idea Group Publishing.

KEY TERMS

Database Trigger: A procedure that gets invoked by the database management system whenever a specific column gets updated or whenever a row gets deleted or inserted.

Denormalized Data Tables: A database design that violates principles of good (normalized) data models. Such database designs may lead to various problems such as data redundancy, reduced flexibility, and update anomalies.

Intelligent Key: An intelligent key contains data that has a meaning beyond the unique identification of a database record. For example, a vehicle identification number (VIN) contains information about the manufacturer, model, and other attributes of the vehicle.

Recursive: A recursive procedure is one that has the ability to call itself.

Relational Database: A method of storing data in tables and columns and operating on that data through a set-oriented query language (SQL) and specialized software the takes care of the procedural and physical aspects of data storage and manipulations. Invented by E.F. Codd at IBM in 1970.

SQL (Structured Query Language): A widely accepted standard for retrieving information from and updating relational databases.

Taxonomy: A classification system used for analysis or information retrieval.

UNION Query: By inserting the key word UNION between multiple SQL SELECT statements, the resulting rows from these statements are appended into a single result set (provided the number and data types of the columns match).

Managing Strategic IT Investment Decisions

Tzu-Chuan Chou
Warwick Business School, UK

Robert G. Dyson
Warwick Business School, UK

Philip L. Powell
University of Bath, UK

INTRODUCTION

IT can have a significant impact on organizational performance, but it can also be a major inhibitor of change and can be a resource-hungry investment that often disappoints. Organizations can best influence the success of IT projects at the decision stage by rejecting poor ones and accepting beneficial ones. However, little is known about IT decision processes. Research demonstrates the importance of managing strategic IT investment decisions (SITIDs) effectively. SITIDs form part of the wider range of corporate strategic investment decisions (SIDs) that cover all aspects that the organization might wish to invest in. SIDs will then have different degrees of IT intensity that may impact on outcome. IT investment intensity is the degree to which IT is present in an investment decision. Here, IT investment intensity is defined as the ratio of IT spending to total investment. The higher IT investment intensity, the more important IT is to the whole investment. For example, Chou et al. (1997) find IT investment intensity to be negatively associated with SID effectiveness. The concept of IT intensity is similar to, but also somewhat different from, the concept of information intensity. Information intensity may be defined as the degree to which information is present in the product or service of a business (Porter & Millar, 1985).

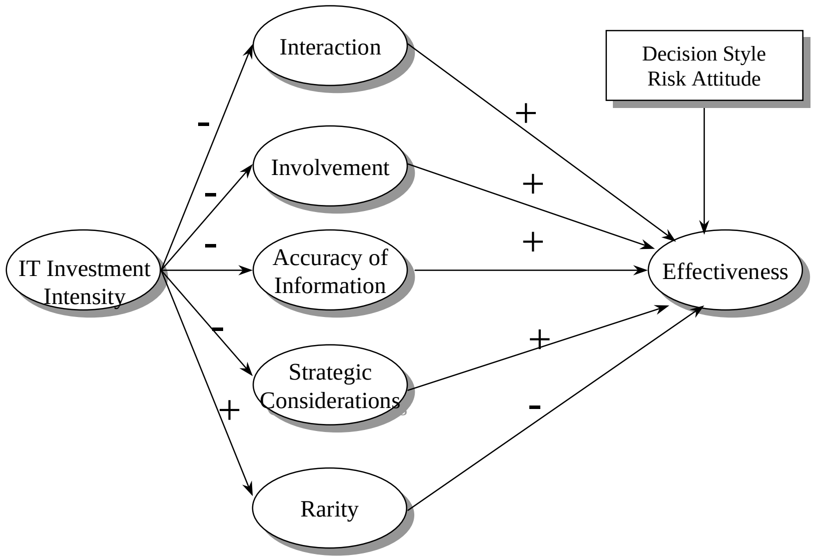
Management may use different processes in order to make different types of decisions (Dean & Sharfman, 1996). The link between decision process and outcome is so intimate that "the process is itself an outcome" (Mohr, 1982). This may imply that the link between IT investment intensity and SID effectiveness is not direct but that the impact of IT investment intensity may be through the decision process. If different IT intensity in projects leads to different decision processes, leading to different outcomes, then it is important to know what factors act in this, in evaluating and managing SITIDs. This chapter presents an integrative framework for exploring the IT investment intensity-SID effectiveness relationship.

BACKGROUND

Studying decisions involves "contextualism" (Pettigrew et al., 1988), which integrates process, content and context, as all decisions need to be studied in context. Content refers to the decision itself, exploring the nature and scope of SIDs. Process refers to actions, reactions and interactions as managers allocate resources for the decision. The context includes the outer context of economic, political and social actions, while the inner context involves on-going strategy, structure, culture, management and political processes.

In the IT investment intensity-SID effectiveness link, the roles of process, content and context are unclear. While unclear, it is likely that the links between variables are not direct; rather they are mediated or moderated by other variables or processes. Moderators and mediators are functions of third variables. A moderator "partitions a focal independent variable into subgroups that establish its domains of maximal effectiveness in regard to given dependent variables," while a mediator function "represents the generative mechanism through which the focal independent variable is able to influence the dependent variable of interest" (Baron & Kenny, 1986). Sambamurthy et al. (2003) employ the moderator concept in their work on reshaping agility through digital options. They, for example, see IT competence as an antecedent of firms' competitive actions but the relationship is mediated by dynamic capabilities. Sambamurthy et al. demonstrate theoretically that IT investments and capabilities influence firm performance through organizational capabilities – agility, digital options and entrepreneurial alertness, and through strategic processes involving capability building, entrepreneurial action and co-evolutionary adaptation. This, they claim, is valid at the enterprise level, for business units and for processes. They finally call for empirical research that might validate their theoretical developments. This chapter takes a slightly different route, as it focuses on IT at the project or at the decision

Figure 1. The theoretical model (adapted from Chou et al., 1997)



level and it is based on data that back up the model development.

Here, the proposal is that the impact of IT investment intensity on SID effectiveness is through decision processes. Accordingly, decision process constructs should have a mediating effect. Greater IT intensity in projects leads, inter alia, to a more technically orientated project, which impacts on SID effectiveness. The decision content has a mediating effect on the IT involvement-SID effectiveness link. The investment context impacts the outcome. Therefore, context constructs should act as covariances that impact SID effectiveness. Decision context, content and process will involve many individual constructs, some unrelated to IT investment intensity. Two criteria can be employed in order to select the constructs of interest here. First, the decision construct is expected to vary with IT investment intensity. Second, it must impact at the decision level. Figure 1 outlines the basic model.

A hypothesized negative impact of IT investment intensity on several constructs suggests projects with high IT investment intensity are more challenging than those with low IT content. Effectiveness compares actual performance against planned target, outcomes and policy objectives, measured by project success, correct choice, unexpected negative outcomes, learning, and satisfactory process (Butler et al., 1993).

Decision Context

The context of any investment is affected by the firm's financial health, its market position, industry pressures, culture, and business strategy. SIDs often involves major change to the organization and environment. This suits managers with an innovative risk attitude. From a style perspective, decision quality is dependent on resources the leader is able to utilize. Consensus-driven management seems able to acquire more information than directive management, and leads to more effective decisions. Management's *attitude to risk* and *decision style* are predicted to relate to SID effectiveness, since other factors impact at an organizational level.

Decision Process

Strategic decision processes involve comprehensiveness, rational activity, participation, duration and conflicts (Rajagopalan et al., 1993). Comprehensiveness measures rationality and is the extent to which the organization attempts to be exhaustive in making and integrating strategic decisions. This includes formal meetings, assignment of primary responsibility, information-seeking and analytical activities, the systematic use of external sources, stakeholder involvement, use of consultants, reviews of historical data, functional expertise (Papadakis, 1995), and informal interaction. Hickson et al. (1986) define "politicality" as the degree to which influence is exerted on the outcome through a decision process. Strategic decision-making is not a matter of explicating alternatives and choosing on the basis of readily available criteria all participants perceive as appropriate (Fahey, 1981). It might be expected that interaction and involvement are related to IT investment intensity.

Interactions are contacts between people. Higher IT intensity reduces interaction and SID effectiveness. Decision-makers' IT knowledge, experience, and educational are associated with alienated attitudes towards IT. Higher IT investment intensity leads to more technically oriented projects. Without IT knowledge, managers cannot discuss the project knowledgeably. It, therefore, reduces interaction and impacts upon decision quality. This article suggests that higher IT intensity reduces *involvement*, reducing SID effectiveness. Less involvement leads to less collective information and reduced decision effectiveness. This suggests that *IT investment intensity reduces interaction and adversely impacts decision effectiveness* and that *IT investment intensity reduces involvement and adversely impacts decision effectiveness.*

The evaluation process is important for investment decisions. An IT investment decision is problematic because the cost and benefits are hard to identify and quantify. Therefore, uncertainty of information used in evaluating IT investment is greater. The higher the information uncertainty, the lower the *information accuracy*. This article expects that lower accuracy of information reduces decision effectiveness; that is, *IT investment intensity reduces information accuracy and adversely impacts decision effectiveness.*

It can be argued that the IT evaluation problem is one of alignment. This article expects that management may fail to link IT's strategic purpose with organizational strategy, reducing decision effectiveness. This suggests that *IT investment intensity reduces strategic considerations and adversely impacts decision effectiveness.*

Decision Content

A strategic decision is characterized by its novelty and complexity. Complexity relates to the number and variety of factors in the environment that impinge on decision-making behavior. SIDs evolve from the organizational context and have their own characteristics. Constructs that contribute to decision complexity include rarity and importance. Uncertainty is due to *rarity*. Rarity assesses the novelty of the decision to the firm. If a firm repeatedly makes similar decisions then it will gain experience; conversely, a rare decision is likely to be more problematic. Importance is common to all SIDs irrespective of IT investment intensity, as they are all strategic. New technologies often require investments of a different nature because of high uncertainty, widespread organizational impact, and greater strategic importance. Even compared with other new technologies, the IT life cycle is short, so that the IT component of projects is constantly changing, increasing rarity. Rarity inhibits effective feedback and learning. This article expects that the higher IT investment intensity, the higher decision rarity, reducing decision effectiveness; *IT investment intensity heightens decision rarity and adversely impacts decision effectiveness.*

In order to investigate these proposals, empirical work investigated a single strategic investment project in each of 80 Taiwanese manufacturers. IT investment intensity is measured by the ratio of IT spending to total investment in the project, while the measure of decision effectiveness is subjective. In order to capture the whole process from decision to outcome, only projects that were complete were researched. Thus each project had a decision process and an outcome to allow assessment of success.

DISCUSSION

A principal components factor analysis highlights five important factors in these decisions - information accuracy, strategic consideration, interaction, involvement, and rarity. However, only three proposed mediators, interaction, information accuracy and strategic consideration are significant in predicting mediators.

The model as a whole is significant in predicting SID effectiveness. When contextual variables are added, IT investment intensity is still significant in predicting SID effectiveness. Interaction, information accuracy and strategic consideration have a negative correlation with IT investment intensity, but a positive correlation with SID effectiveness. Hence, the impact of IT investment intensity is transmitted to interaction, information accuracy, and strategic considerations and, through that, adversely impacts decision effectiveness. This suggests that *IT investment intensity does reduce interaction and adversely impacts decision effectiveness.* It is also the case that IT investment intensity reduces information accuracy and adversely impacts decision effectiveness, and that IT investment intensity reduces strategic considerations and adversely impacts decision effectiveness. However, IT investment intensity is not found to reduce involvement and adversely impact decision effectiveness, nor does IT investment intensity heighten decision rarity and adversely impact decision effectiveness.

Three process-related constructs, interaction, strategic considerations, and information accuracy act as mediating factors in linking IT investment intensity and SID effectiveness, as they reduce the effects of IT investment intensity. Content-related constructs do not act as mediators. Although decision rarity is negatively associated with SID effectiveness, it is unrelated to IT intensity.

Interaction in the formulating process has a mediating effect on the linkage. Interaction is important in developing group behavior. IT investment intensity lowers interaction, reducing SID effectiveness.

Strategic considerations act as a mediating variable. The higher IT intensity, the lower the strategic considerations, leading to reduced SID effectiveness. This demonstrates that the IT evaluation problem is really one of alignment.

Information accuracy acts as a mediating variable. The higher IT investment intensity, the lower information accuracy, reducing SID effectiveness. This supports Freeman and Hobbs (1991), who find managers ignoring reject signals given by capital budgeting techniques, and identify senior management's preference for qualitative information and IT investment as an "act of faith" (Powell, 1993). This suggests that high information uncertainty leads to a limited use of capital budgeting techniques.

IT investment intensity is still significant when interaction is tested as a mediator. This indirect transmission of influence from IT investment intensity to SID effectiveness via interaction shows that the effect of IT investment intensity on effectiveness is only partially mediated by interaction. The effect of IT investment intensity on SID effectiveness is completely mediated by strategic consideration and information accuracy - two evaluation-related constructs. This implies that, in seeking a better outcome of SITIDs, research that focuses on evaluation factors may be insufficient to capture the complexity of SITIDs.

FUTURE TRENDS

As demonstrated, the two evaluation-related constructs are highly correlated. From an IT investment perspective, alignment of IT and business strategy is problematic if there is a lack of accurate information for evaluation. However, evaluation of IT investments is problematic if there is a lack of alignment of IT and business strategy. To improve the effectiveness of IT investment, management needs to increase alignment of IT and business strategy and information accuracy for evaluation techniques simultaneously. This points to the issues to which management needs to attend in the future.

CONCLUSION

Much work on SITIDs ignores the continuous nature of decisions and the relationships between SITIDS and non-IT SIDs. This article proposed a model that explores mediators in the link between IT investment intensity and SID effectiveness. Survey data show that interaction, information accuracy and strategic considerations are important factors that mediate the impact of IT investment intensity. Willcocks (1992) emphasizes that management faces a Catch-22 situation with IT investment. They know how important IT is, but they do not know how to evaluate IT projects. The implication is that managers need to pay special attention to the problematic nature of IT investment intensity in SIDs. They should focus on facilitating interaction, ensuring integration of IT strategy with corporate strategy and improving information accuracy.

REFERENCES

Baron, R., & Kenny, D. (1986). Moderator-mediator variable distinction. *Journal of Personality and Social Psychology, 51*(6), 1173-1182.

Butler, R., Davies, L., Pike, R., & Sharp, J. (1991). Strategic investment decision-making. *Journal of Management Studies, 28*(4), 395-415.

Chou, T., Dyson, R., & Powell, P. (1997). *Does information technology matter?* 4th European Conference on IT Investment Evaluation, Delft, Netherlands.

Dean, J., & Sharfman, M. (1996). Does decision process matter? A study of strategic decision-making effectiveness. *Academy of Management Journal, 39*(2), 368-396.

Fahey, L. (1981). On strategic management decision processes. *Strategic Management Journal, 2*, 43-60.

Freeman, M., & Hobbs, G. (1991). Costly information, informed investors. *AAANZ Conference* (pp. 68-74).

Hickson, D., Butler, R., Cray, D., Mallory, G., & Wilson, D. (1986). *Top decisions: Strategic decision-making in organizations.* San Francisco, CA: Jossey-Bass.

Mohr, L. (1982). *Explaining organizational behavior*. San Francisco: Jossey-Bass.

Papadakis, V. (1995). Contribution of formal planning systems to strategic investment decisions. *British Journal of Management, 16*, 15-28.

Pettigrew, A., McKee, L., & Ferlie, E. (1988). Understanding change in the NHS. *Public Administration, 66*, 297-317.

Porter, M., & Millar, V. (1985, July-August). How information gives you competitive advantage. *Harvard Business Review*, 149-160.

Powell, P. (1993). Causality in the alignment of information technology and business strategy. *Journal of Strategic Information Systems, 2*(4), 320-334.

Rajagopalan, N., Rasheed, A., & Datta, D. (1993). Strategic decision processes. *Journal of Management, 19*(2), 349-384.

Sambamurthy, V., Bharadwaj, A., & Grover V. (2003). Shaping agility through digital options: Reconceptualizing the role of IT in contemporary firms. *MIS Quarterly, 27*(2), 237-263.

Willcocks, L. (1992). IT evaluation: Managing the Catch 22. *European Management Journal*, *10*(2), 220-229.

KEY TERMS

Contextualism: Integrates process, content and context to study organizational decision-making.

Decision Content: Content refers to the particular decision under study. Content explores the basic nature and scope of decisions.

Decision Context: The context includes the outer context, which refers to the national economic, political and social context for an organization, and the inner context that is the on-going strategy, structure, culture, management and political process of the organization. Context helps to shape the process of decision-making.

Decision Process: Refers to the actions, reactions and interactions of the various interested parties as they seek to make a commitment to allocate corporate resources. Process incorporates both the formulation and evaluation processes.

IT Investment Intensity: A concept similar to, but also somewhat different from, the concept of information intensity. Information intensity is the degree to which information is present in the product/service of a business. The degree to which IT is present in an investment decision reflects the IT level of intensity of that decision. Here, IT investment intensity is defined as the ratio of spending on IT to total investment.

Moderators and Mediators: In the social sciences, moderators and mediators have been identified as two functions of third variables. These are sub-groups of independent variables that affect given dependent variables via a mediator function.

Strategic IT Investment Decisions: Significant, long-term decisions to invest in projects that have substantial information systems or information technology components. They form part of corporate strategic investment decisions.

Managing the Organizational Impacts of Information Systems

Neil F. Doherty
Loughborough University, UK

Malcolm King
Loughborough University, UK

INTRODUCTION

Information technology is now a ubiquitous and increasingly critical part of the fabric of the modern organization, supporting its day-to-day operations and all aspects of the decision-making process, as well as its strategic positioning. It is, therefore, not perhaps surprising that the implementation of a new technology or information system is likely to result in wide array of impacts to the organization as well as the working lives of individual employees. There is a growing consensus within the literature that many such impacts are not deterministic and cannot therefore be easily predicted prior to a system's implementation (e.g., DeSanctis & Poole, 1994). The corollary of this is that many of the consequences of an information system's implementation will be unanticipated (Robey & Boudreau, 1999). While some of these unanticipated consequences, or incidental side effects, may be of a positive nature, negative impacts are also quite common, as IT-induced organizational change often results in user resistance and, in extreme cases, possibly even system rejection (Martinsons & Chong, 1999).

While there is strong evidence that information systems projects are not totally predictable, it can be argued that many of their organizational impacts only remain unanticipated, because systems developers are reluctant to tackle the human and organizational aspects of IT. Systems development projects have typically been viewed as exercises in technical change, rather than socio-technical change; *"most investments in IT are technology-led, reflecting too technical an emphasis"* (Clegg, 2000). This is a dangerous strategy, because unforeseen and unresolved negative impacts may increase the likelihood of systems failure. Moreover, beneficial impacts, of both a planned and incidental nature, may not be fully realised without an appropriate programme of organizational change. We would argue that if systems development projects are viewed as an exercise in organizational change, in which all potential organizational impacts are proactively and systematically analysed, then many previously unanticipated impacts could be predicted (Doherty & King, 2002). While the importance of treating organizational issues is now widely acknowledged (e.g., Clegg et al., 1997; Eason, 2001), little progress has been made in the development of practical treatment approaches that have succeeded in making the transition from research laboratory to widespread commercial usage. The primary aim of this article is to present an innovative new approach for their proactive treatment. However, in advance of this, it is important to establish the importance of treating organizational issues.

BACKGROUND: THE NEED TO TREAT ORGANIZATIONAL ISSUES

The information systems' literature is very clear on two points; general levels of failure are far too high and the primary cause of this problem is the failure to adequately treat organizational issues (Clegg et al., 1997; Ewusi-Mensah & Przasnyski, 1994). In this context, the term *"organizational issue"* relates to those organizationally-oriented facets of systems development projects that need to be treated to ensure that the resultant impacts of an information system are likely to be desirable. A comprehensive checklist of important organizational issues, that was originally drawn from the literature, but then validated over a series of studies (e.g., Doherty & King, 1998, 2001, 2003), is presented in Table 1.

To treat a specific organizational issue, it is necessary to firstly evaluate the likely organizational impact associated with it, and then if necessary take steps to ensure that the resultant impact is likely to be desirable. For example, if it is found that a proposed system is likely to be poorly suited to an organization's working practices, then it will be necessary to either modify the system's technical specification, so that the mismatch is avoided, or redesign the working practices so that they are well aligned with the system. In essence, the treatment of organizational issues is the mechanism by which the project team should align

Table 1. Checklist of organizational issues to address

Issue	Description
Current business needs	The system's ability to satisfy the organization's current business needs.
Information systems strategy	The system's alignment with the current information system strategy
Prioritisation of needs	The prioritising of development effort on those aspects which address the most important business needs.
Future needs of organization	The system's ability to satisfy the organization's likely future business needs.
Process design	The system's impact on the design of key business processes.
Health & safety / ergonomic factors	The likely ergonomic and health & safety implications of the system, such as RSI and eye strain.
User motivation / needs	The system's ability to satisfy user needs and support user motivations.
User working styles and personal skills	The implications of user working styles and personal skills for the system's design and ongoing use.
Job redesign	The proposed system's impact on the design of working practices.
Timing of Implementation	The interaction of the system's implementation with other planned concurrent changes.
Organizational disruption	The temporary organizational disruption that may be caused by the implementation of the proposed system.
Organizational structure	The system's effect on the organizational structure, and the lines of authority.
Organizational culture	The proposed system's impact on the culture in the organization (*i.e., the set of important assumptions (often unstated) which members of an organization share in common).*
Organizational power	The proposed system's political implications for the distribution of power in the organization.

the capabilities afforded, and the constraints imposed, by the technical system with the requirements and characteristics of an organization and its individual employees.

System developers typically view the system development process as a science, rather than art, which requires the use of structured methods that focus upon the delivery of technically effective systems, on time and within budget. They are extremely reluctant to tackle intangible, ill-defined and politically-sensitive organizational issues (Doherty & King, 2001), for which they're ill-equipped, in terms of training, competencies and motivation (Clegg, 2000). Consequently, approaches to the treatment of organizational issues have typically been reactive rather than proactive (Clegg et al., 1996) – get the system implemented and then worry about its organizational impacts. There is, therefore, a pressing need to find ways to encourage the systems development community to become more actively engaged in the treatment of organizational issues. One obvious strategy is through the creation of methods, tools and techniques, which are specifically designed to facilitate the treatment of organizational issues. A wide variety of organizationally-oriented approaches have now been proposed, which can be categorised as follows:

1. **Socio-technical methods:** Socio-technical methods attempt to produce information systems that are technically efficient and coherent, while also being sensitive to organizational and human needs, for example, Ethics (Mumford, 1996) or Multi-view (Avison, Wood-Harper, Vidgen, & Wood, 1998).

2. **Tools and techniques for the treatment of specific issues:** Many researchers (e.g., Clegg et al., 1996) have attempted to develop tools and techniques to aid in the treatment of specific organizational issues.
3. **An organizational impacts analysis:** The *"organizational impact analysis"* (e.g., Sauer, 1993) is typically a one-off study to determine the ways in which a proposed system will affect the organization's decision-making, power, structure, culture, working practices, and so forth.

While each of the above contributions has been very useful in increasing our understanding of the nature and treatment of organizational issues, there is little evidence that these contributions have made much of an impact on the practice of systems development (Clegg, 2000). This is probably, at least in part, due to technical specialists' continuing preference for the more technically oriented tools and techniques. However, if a comprehensive, coherent and easy-to-use approach could be found, that complemented their existing methods, then it might have a greater chance of adoption. The remainder of this section describes one such approach, that can best be described as an example of organizational impact analysis.

APPROACH FOR THE TREATMENT OF ORGANIZATIONAL ISSUES

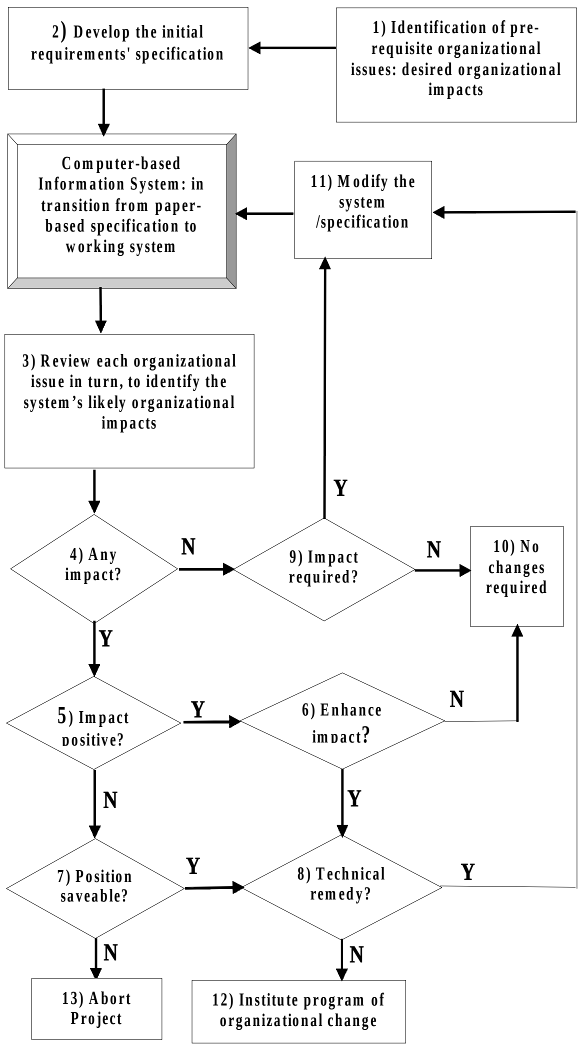
The proposed approach has been formulated from an extensive review of the literature, and the authors' experience working in this domain for the past few years (e.g., Doherty & King, 1998, 2001, 2004). A schematic representation of the approach, which has been conceptualised as a flow diagram, is presented in Figure 1. Each of the major processes and decision points on this diagram, all of which have been numbered, is reviewed in the following:

1. **Identification of planned organizational impacts:** While many organizational issues will not be treated until a systems development project is well under way, others, such as the system's ability system to satisfy *"current organizational needs"*, will need to be considered right at the very outset, if they are to be the planned outputs of the project.
2. **Development of initial requirements specification:** The development of the initial requirement's specification is a fundamental component of all development methods, and can be conducted using the proprietary or in-house method of the developers choosing.

While the previous two stages occur only once, at the project's outset, it is envisaged that the following stages will be repeated at key stages in the systems development process, for each of the organizational issues in the checklist (see Table 1), in turn.

3. **Review of organizational issues:** Assess the system's likely impacts, with regard to each organizational issue. The process is designed to ensure that the planned impacts will ultimately come to fruition, while any incidental impacts will be effectively identified and managed.
4. **Determine existence of organizational impacts:** The output of the review procedure, described previously, will be an assessment of whether there is a significant organizational impact, associated with each organizational issue.
5. **Evaluation of desirability of impacts:** The desirability of each identified impact must be assessed, to determine whether it is likely to be of a positive or negative nature. For example, an assessment of the system's impact on the motivation of users might identify a negative impact, such as user resistance or resentment, due to changes in their work practices. Potential solutions might be of a technical nature, such as changing the system's design to give the user more control over their work, or an organisational orientation, for example, improving the users' terms and conditions, by way of compensation.
6. **Assessment of potential for increasing desirability of impacts:** If the previous stage has identified a desirable impact, associated with an organizational issue, it is important to consider whether the impact could be made even more positive if the information system design were to be modified.
7. **Is the situation retrievable?:** In situations where a potentially undesirable impact of the system's operation has been identified then it is necessary to consider whether the situation is retrievable or not.
8. **Is the remedy technical?:** Having identified potentially negative, yet retrievable, impacts, associated with the system implementation, a decision must be made as to whether the remedy is of a technical or organizational nature.
9. **Evaluation of potential for impacts:** If it has been discovered that there is no impact, associated with a particular organizational issue, then it is important to question whether there should be an impact. If, for example, it has been determined that the system is unlikely to change working practices, questions might be raised as to whether the system could be used to stream-line or enrich the design of jobs.

Figure 1. Approach to treatment of organizational issues



10. **No changes required:** In the cases where there is no actual impact or potential for any specific organizational issue then there is no requirement to either change the system's specification or institute a programme of organizational change.
11. **Modification of specification:** In many situations, the organizational issue's review process will necessitate changes to the system's specification, in order to correct a negative impact, or evoke a more positive impact.
12. **Development of program of organizational change:** In situations where organizational impacts have been identified that have judged to be desirable, it is important that a programme of organizational change is planned and implemented to ensure that the impact is realised.
13. **Abort project:** In situations where it has been found that the introduction of an information system is likely to result in significant organizational impacts, of a negative nature, then the project should be aborted.

At a minimum, it is envisaged that the organizational impacts will be assessed, using this approach, at the following key stages of the development process (Figure 2): on completion of the requirements specification, then again at the end of the development phase and then very soon after implementation. However, for very complex projects, or those that are likely to be the catalyst for significant organizational change, it is recommended that the analysis should be repeated more frequently, particularly during the design phase. Moreover, it is recommended that the organizational impact analysis be repeated a number of times over the systems working life. This would be one way of operationalising Orlikowski, Yates, Okamura, and Fujimoto's (1995, p.424) concept of *"technology-use mediation"*, which they define as:

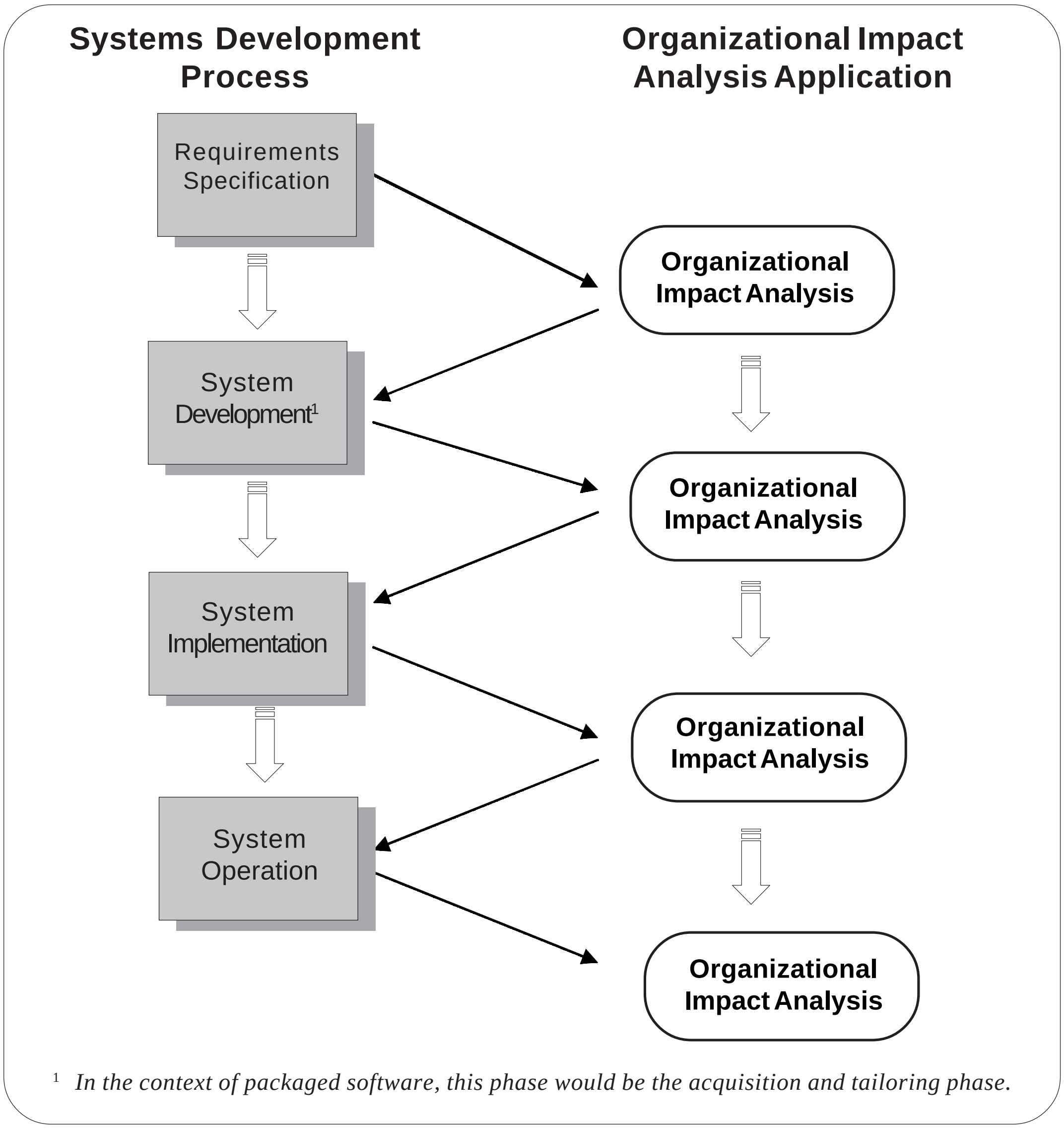
"deliberate, ongoing and organizationally sanctioned intervention within the context of use that helps to adapt new technology to its context, modifies the context as appropriate to accommodate the use of the technology, and facilitates the ongoing effectiveness of the technology over time."

It should be noted that once the system goes live, the organizational impact analysis procedure will not need to address the full range of organizational issues, as some – such as prioritization, timing of implementation, and organizational disruption – will no longer be relevant in an operational context.

While there will be a number of separate applications of the organizational impact analysis, they are not totally independent, as it is envisaged that the outcomes and findings of any given iteration of the of the approach might prompt or flag-up issues that need to be addressed in a subsequent iteration. It should also be remembered that many organizational impacts are interdependent and consequently changes made with respect to one specific issue might engender impacts in other areas. One of the obvious benefits of adopting an iterative approach is that it allows changing requirements to be monitored and accommodated, on an on-going basis.

In terms of who is involved in the analysis of the impacts and the identification of appropriate changes to the system or the organization, to ensure the impacts will ultimately be desirable, it is recommended that the exercise be very inclusive. As a key objective of socio-technical approaches is to achieve consensus among all the system's stakeholders, it is envisaged that the pro-

Figure 2. Relationship between systems development process and application of the organziational impact analysis



posed approach will act as a mechanism for channeling a debate about organizational change. As Markus and Benjamin (1997, p.55) put it:

"the hard reality of IT-enabled transformation is that change is everyone's job."

FUTURE TRENDS

As the scope and strategic importance of information systems continues to grow, so does the need to find better ways of matching them to their organizational context, to ensure that they deliver a significant contribution to organizational performance. Consequently, there is an urgent need for practical and easy-to-use methods to aid systems development professionals in managing the organizational impacts of the systems for which they are responsible. To this end, our immediate priority is to test the provisional framework on a variety of systems development projects, and use the feedback from these exercises to further define exactly how different organizational impacts can best be analyzed and managed.

CONCLUSION

Many information systems researchers have recognised the need for more effective socio-technical tools and methods to be developed. While the work described in this article is not trying to develop a highly specific tool or technique, it does propose a more general framework to promote the systematic and coherent treatment of organizational issues. The chief benefits of the proposed approach are that it presents systems developers with a systematic framework which obliges them to confront organizational issues and provides them with the means to effectively navigate their way through a very complex decision-making process. Moreover, a comparison of this approach with some of its predecessors allows the following distinctions to be made:

- **Complementary:** The proposed approach complements, rather than replaces, existing development tools and methods. There is no requirement for systems developers to abandon there tried and tested practices.
- **Straightforward:** The approach adopts a common-sense perspective, and it should, therefore, be relatively easy to learn and apply.
- **Proactive:** By using this approach, organizations will ensure that potential problems are recognised and opportunities are identified and exploited in a timely and effective manner.
- **Comprehensive:** The approach is comprehensive and can cope with a wide range of potential impacts.
- **Flexible:** The approach is highly flexible and can be adapted to suit the requirements of a wide variety of information systems projects

In essence this approach is inviting systems developers to periodically stand back from the systems development process and, in conjunction with a wide variety of stakeholders, assess the likely impacts of their work on the design and operation of the organization. While there are a number of potential benefits to the proposed approach, it is also important to issue a health warning: These ideas are provisional and exploratory, and there is much further work required to translate them into a robust and reliable tool.

REFERENCES

Avison, D., Wood-Harper, A.T., Vidgen, R.T., & Wood, J.R.G. (1998). A further exploration into information systems development: The evolution of Multiview 2. *Information Technology and People*, *11*(2), 124-139.

Clegg, C.W. (2000). Socio-technical principles for system design. *Applied Ergonomics*, *31*, 463-477.

Clegg, C.W., Axtell, C., Damadoran, L., Farbey, B., Hull, R., Lloyd-Jones, R., Nicholls, J., Sell, R., & Tomlinson, C. (1997). Information technology: A study of performance and the role of human and organizational factors. *Ergonomics*, *40*(9), 851-871.

Clegg, C.W., Coleman, P., Hornby, P., McClaren, R., Robson, J., Carey, N., & Symon, G. (1996). Tools to incorporate some psychological and organizational issues during the development of computer-based systems. *Ergonomics*, *39*(3), 482-511.

DeSanctis, G., & Poole, M.S. (1994). Capturing the complexity in advanced technology use: Adaptive structuration theory. *Organization*, *5*(2), 121-147.

Doherty, N.F., & King, M. (1998). The consideration of organizational issues during the systems development process: An empirical analysis. *Behaviour & Information Technology*, *17*(1), 41-51.

Doherty, N.F., & King, M. (2001). An investigation of the factors affecting the successful treatment of organizational issues in systems development projects. *European Journal of Information Systems*, *10*, 147-160.

Doherty, N.F., & King, M. (2002). From technical change to socio-technical change: Towards a proactive approach to the treatment of organizational issues. In S. Clarke, E. Coakes, M.G. Hunter, & A. Wenn (Eds.), *Socio-Technical and Human Cognition Elements of Information Systems* (pp.22-40). Hershey, PA: Information Science Publishing.

Doherty, N.F., & King, M. (2003). The impact of inadequacies in the treatment of organizational issues on information systems development projects. *Information & Management*, *10*, 147-160.

Eason, K. (2001). Changing perspectives on the organizational consequences of information technology. *Behaviour & Information Technology*, *20*(5), 323-328.

Ewusi-Mensah, K., & Przasnyski, Z. (1994). Factors contributing to the abandonment of information systems development projects. *Journal of Information Technology*, *9*, 185-201.

Markus, M.L., & Benjamin, R.I. (1997). The magi bullet theory in IT-enabled transformation. *Sloan Management Review*, *38*(2), 55-68.

Martinsons, M., & Chong, P. (1999). The influence of human factors and specialist involvement on informations systems success. *Human Relations*, *52*(1), 123-152.

Mumford, E. (1996). *Systems design: Ethical tools for ethical change*. Basingstoke: MacMillan.

Orlikowski, W.J., Yates, J., Okamura, K., & Fujimoto, M. (1995). Shaping electronic communication – the metastructuring of technology, in the context of use. *Organization Science*, *6*(4), 423-444.

Robey, D., & Boudreau, M.-C. (1999). Accounting for the contradictory organizational consequences of information technology: Theoretical directions and methodological implications. *Information Systems Research*, *10*(2), 176-185.

Sauer, C. (1993). *Why information systems fail: A case study approach*. Alfred Waller, Henley.

KEY TERMS

Incidental Impacts: Impacts that are un-planned, by-products of the system's development process that had not, or could not, have been envisaged at the project's outset.

Organizational Impact Analysis: A one-off study to determine the ways in which a proposed system will affect the organization, in areas such as power, structure, culture, working practices, and so forth.

Organizational Issues: Those issues which need to be treated during the systems development process to ensure that the individual human, wider social and economic impacts of the resultant computer-based information system are likely to be desirable.

Planned Impacts: The anticipated outcomes of a systems development project that were identified at the project's outset, and are typically critical to its ultimate success.

Socio-technical Methods: Development methods that attempt to produce systems that are both technically efficient and organizationally sensitive.

Systems Failure: Systems abandoned before completion, systems completed but never used, under used or failing to deliver key aspects of functionality, and projects that are significantly over budget or schedule.

Managing Value-Creation in the Digital Economy

John C. Mc Intosh
Bentley College, USA

Keng L. Siau
University of Nebraska – Lincoln, USA

INTRODUCTION

Value-creation and maintenance of a sustained revenue stream in the digital economy continue to elude most businesses. With the exception of advanced firms such as Dell™, Virgin®, Yahoo!®, and eBay®, relatively few firms have been able to leverage Internet technologies to create sustainable business models. From a strategic perspective, although firms are freed of the strictures of vertical integration, they face the daunting task of orchestrating a constantly changing web of suppliers and partners to create appealing products for increasingly sophisticated, fickle customers.

Contemporary research, often based on frameworks such as Porter's Five Forces (Porter, 1984), has proliferated myriad, largely static business models. For example, Tapscott et al. (2000) proposed five distinct models—agoras, aggregations, alliances, distribution networks, and value chains, while Weill and Vitale (2001) proposed six—direct to customer, content provider, full-service provider, portals, shared infrastructure, and whole of enterprise. By not explicitly accounting for today's dynamic business environment, these models offer limited prescriptions for sustained value-creation. Fundamental to sustainability is business model evolution, which often entails the transfer of knowledge, processes, partnerships, and relational capital from one group of customers to another. For the IT-enabled firm, sustainable value-creation increasingly depends on its ability to combine intangible assets (i.e., brand, information resources, relational capital) with those of its network partners (Brandenberger & Nalebluff, 1996; Gulati et al., 2000; Tapscott et al., 2000).

This entry synthesizes research from strategic management and IT to develop a framework for understanding sustained value-creation. In contrast to much of the literature on IT-enabled competition (Afuah & Tucci, 2001; Osterwalder & Pigneur, 2002; Weill & Vitale, 2001), it does not focus on e-business models per se, but instead reinterprets the traditional three stages of strategy making (conceptualization, planning, and implementation) to describe how distinct types of information may be used for sustained value-creation. We chose to focus on the strategy aspect of value-creation, because business models are tangible representations of a firm's strategic intent. In other words, they are the outcome of a firm's strategy. Hereinafter, we use the term "product" to describe both tangible products and intangible services and the term "strategy" to describe the protocol that firms use to create valuable products and services.

This article is organized into two sections, first, drawing on two features of information: velocity (rate of change) and interoperability (the extent to which information can be combined with other information to create value). It develops a 2×2 matrix depicting four fundamental types of information necessary to understand information-driven competition (Figure 1). Next, it develops a conceptualization of sustained value-creation by using the information typology and IT's considerable information-gathering and coordination capabilities as a lens to reinterpret the three stages of strategy.

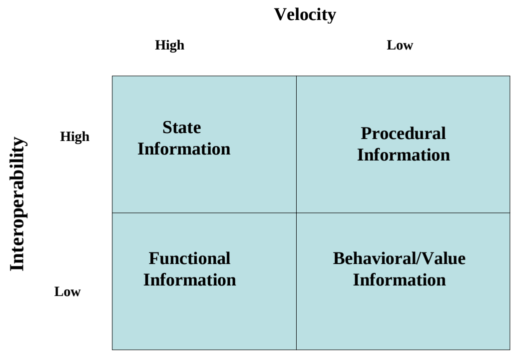
FOUR FUNDAMENTAL INFORMATION TYPES

This section draws on two characteristics of information—velocity and interoperability to develop a typology (Mc Intosh & Siau, 2001). Each type is discussed.

State Information

Typically descriptive in nature, state information changes rapidly (high velocity) and can be easily combined with other information to create value (high interoperability). Examples of state information include the price of a commodity at a specific day and time; the quantity of a product in inventory; and the chronology of a product's journey through a firm's value chain. General Motor's information system, for example, provides state information concerning the number of engine assemblies in inventory and the mix of automobiles on a car carrier in the distribution system (Mc Intosh & Siau, 2001).

Figure 1. Typology of information based on velocity and interoperability



Procedural Information

Procedural information, characterized by low velocity and high interoperability, refers to the steps or protocols required to conduct a process or perform a service. It can be combined with other information to create efficient business processes. For example, best practices for a particular industry may be combined with knowledge of company-specific practices to improve business processes. Procedural information therefore tends to be more explicit and relatively easily codifiable (Mc Intosh & Siau, 2001).

Functional Information

Functional information describes how components and subsystems of a tangible product interact and how those interactions give rise to the performance features that characterize a product. The most common, engineering knowledge, is high velocity (applied technical information changes rapidly), of low interoperability (applicable to a narrow domain), and is grounded in an understanding of the technologies that constitute a product and how variations in those technologies can affect overall system performance. In PC design, functional information regarding microprocessor speed, hard drive access time, and the amount of RAM permits engineers to design systems with differing performance characteristics. Unlike procedural information, functional information tends to be more complex, tacit, and cannot be as easily captured and represented in an organization's databases as can state and procedural information (Mc Intosh & Siau, 2001).

Behavioral/Value Information

This information type, characterized by low velocity/low interoperability, refers to the way large-scale, complex systems interact under different environmental conditions. It may be used to predict the future actions of complex systems, such as the trajectory of a hurricane, the competitive outcomes from adding a new partner to a value web, or customer response to a new channel. For example, capturing information about individual customers and buying preferences might allow the firm to create a model of consumer preference for particular product offerings and thus enable it to fine-tune its marketing and sales initiatives (Mc Intosh & Siau, 2001).

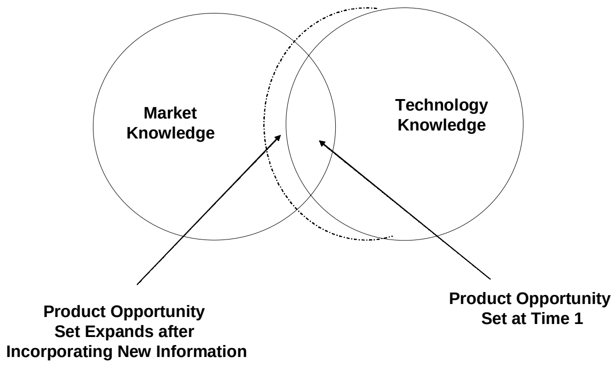
UNDERSTANDING SUSTAINED VALUE-CREATION

The Internet is a double-edged sword for today's managers. Freed from the constraints of vertical integration and costly coordination, managers enjoy unprecedented flexibility to create appealing products and services by mixing and matching resources and capabilities from their value-webs. Nevertheless, the Internet's dynamic nature, hypercompetition (D'Aveni, 1995), and accelerating global competition, force managers to cope with shorter, more frequent design–build cycles, quicker market feedback, and greater opportunities for product evolution. Despite this, most business models, operating under assumptions of linearity, constrain managerial action to the proverbial single-board chess game. Stated differently, the linear, three-stage conception of strategy (conceptualization/planning/implementation) now becomes iterative, with each cycle offering opportunities for learning, adaptation, and sustained value-creation. This arises from IT's information gathering and coordinative capabilities that enable managers to incorporate, at any stage, new information, resources, and capabilities, from the firm's value-web. The chess game, now played on three boards, permits horizontal and vertical moves.

Chess Board 1: Conceptual Stage

Proponents of the resource-based view (Barney, 1991; Grant, 1995; Wernerfelt, 1984) demonstrate that product conceptualization is conditioned by the firm's resources and capabilities. Managers use the firm's market knowledge (i.e., customer tastes and preferences) and technology knowledge (what technological capabilities can be used to meet market demands) to create products. The overlap of these two domains (Figure 2) represents the firm's product opportunity set. By allowing managers to access knowledge resources from value-web partners, IT expands both knowledge domains to offer a broader product opportunity set (Hamel, 2000; Piller et al., 2000; Spencer, 2003). For example, a firm in the automobile performance components industry may combine procedural information (from a value-web partner possessing expertise in carbon-fiber composites) with functional information (concerning how air flows within an intake system) to offer a lightweight, high-performance manifold.

Figure 2. Product opportunity set expansion after new knowledge is added



Chess Board 2: Planning Stage

Management education, steeped in concepts such as specialization (Porter, 1980), business processes (Chowdhury, 2002), and value chains (Porter & Millar, 1985), has fostered a firm-centric view of products. Consequently, managers often create offerings that meet a narrow set of customer needs. Customers, however, focus less on product characteristics and more on services that meet their idiosyncratic needs. Firms tend to focus on specific products, while customers tend to focus on a bundle of services that satisfy particular needs. We advance a customer-centric view of value-creation by proposing the "activity chain" concept. Much like the firm's value chain, a customer's activity chain defines the steps necessary to achieve a particular goal. For example, a consumer desiring to move cross-country would engage in the following: research neighborhoods, find a home, buy that home, hire a moving company, turn on utilities at the new location, and convert licenses and vehicle registrations.

Prior to commercialization of the Internet, specialized firms (such as realtors) tended to create firm-centric, atomistic products that satisfied selected steps in the customer activity chain. IT-enabled companies such as Homefinders.com, taking a customer-centric view, offer a menu of atomistic services from which users may create product bundles that meet their idiosyncratic needs. This is typically achieved by using behavioral/value informationBehavioral /Value Information (to construct a stylized activity chain describing tasks associated with moving) and procedural information (concerning how each web partner delivers its atomistic service). Additional value may be created by offering customers a mortgage application wizard (based on procedural information) or tracking services to monitor the progress of a moving truck transporting their possessions (based on state information).

Chess Board 3: Implementation Level

IT's information-gathering and coordinative capabilities significantly reduce transaction costs associated with finding appropriate value-web partners, certifying their capabilities, mitigating opportunistic behaviors, and reducing switching costs in the event a partnership fails (Mc Intosh & Kim, 1998). Reduced risk and greater coordination allow firms to focus on core activities while hiving off noncore activities to "best-in-breed" partners, thereby conferring unprecedented levels of strategic flexibility (Andersson et al., 2002; Dyer, 1997; Madhok & Tallman, 1998; Tsai & Ghoshal, 1998). These advantages, however, are not without costs, because electronic "arm's length contracting" requires a significant element of trust. This places a premium on managing the firm's social and relational capital (Burt, 1997; Granovetter, 1985). Managing relationships with value-web partners increasingly requires the use of all four types of information. Coordinating a simple buyer–supplier relationship requires state information. Firms relying on just-in-time deliveries need detailed information of inventory stocks and the physical locations of components in the logistic chain. Managers desiring to adopt best practices or transfer learning across organizations require a keen understanding of procedural information to successfully absorb new practices. Additionally, as firms jointly develop new products and processes, procedural information concerning partners' business processes is increasingly important to facilitate synchronized action. Last, arrangements such as strategic alliances or mergers require very high levels of interdependence. In cases where partners are culturally dissimilar, a keen understanding of behavioral/value information is of paramount importance to organizational effectiveness.

CONCLUSION

We proposed a framework for sustained value-creation. The framework consists of three stages: the conceptual stage, the planning stage, and the implementation stage. Four types of information are discussed: state, procedural, functional and behavioral/value. The use of these four types of information in the three stages of the framework to sustain value-creation is discussed in the second section of the chapter. The contributions of the framework are threefold. First, in contrast to traditional business models, the framework is not a static model. Second, the framework integrates strategy management with IT. Third, the framework discusses the use of different types of information to support the three stages of value-creation.

REFERENCES

Afuah, A. C., & Tucci, C. (2001). *Internet business models and strategies*. Boston, MA: McGraw Hill.

Andersson, U., Forsgren, M., & Holm, T. (2002). The strategic impact of external networks: Subsidiary performance and competence development in the multinational corporation. *Strategic Management Journal, 23,* 979–996.

Barney, J. B. (1991). Firm resources and sustained competitive advantage. *Strategic Management J., 17*, 99–120.

Brandenberger, A., & Nalebluff, B. (1996). *Co-Opetition*. New York: Doubleday.

Burt, R. S. (1997). The contingent value of social capital. *Administrative Science Quarterly, 42,* 339–365.

Chowdhury, S. (2002). *Design for six sigma—The revolutionary process for achieving extraordinary profits*. Chicago, IL: Dearborn Trade Publishing.

D'Aveni, R. A. (1995). Coping with hypercompetition: Utilizing the new 7S's framework. *Academy of Management Executive, 9,* 45–57.

Dyer, J. H. (1997). Effective inter-firm collaboration: How firms minimize transaction costs and maximize transaction value. *Strategic Management Journal, 18,* 535–556.

Granovetter, M. (1985). Economic action, social structure, and embeddedness. *Amer. J. of Sociology, 91*, 481–510.

Grant, R. M. (1995). *Contemporary strategy analysis*. Malden, MA: Blackwell.

Gulati, R., Nohria, N., & Zaheer, A. (2000). Strategic networks. *Strategic Management Journal, 21*, 203–215.

Hamel, G. (2000). *Leading the revolution*. Boston, MA: Harvard Business School Press.

Madhok, A., & Tallman, S. B. (1998). Resources, transactions and rents: Managing value through inter-firm collaborative relationships. *Organization Science, 9*, 326–339.

Mc Intosh, J. C., & Kim, E. (1998). Competitive uses of information technology in the global environment: A comparative study of American, Swedish and Korean manufacturing firms. In M. A. Hitt, J. E. Ricart, I. Costa, & R. D. Nixon (Eds.), *New managerial mindsets*. Chichester, UK: John Wiley & Sons.

Mc Intosh, J. C., & Siau, K. (2001). Managing value-creation in the digital economy: Information types and e-business models. In S. Dasgupta (Ed.), *Managing internet and intranet technologies in organizations: Challenges and opportunities*. Hershey, PA: Idea Group Publishing.

Osterwalder, A., & Pigneur, Y. (2000). *An eBusiness model ontology for modeling eBusiness*. Bled Electronic Commerce Conference, Bled, June 17–19.

Piller, F. T., Reichwald, R., & Moselin, K. (2000). *Information as a critical success factor for mass customization or why even a customized shoe not always fits*. ASAC-IFSAM conference, Montreal, July 8–11.

Porter, M. E. (1980). *Competitive strategy*. New York: The Free Press.

Porter, M. E., & Millar, V. (1985). How information gives you competitive advantage. *Harvard Business Review, 63,* 149–160.

Spencer, J. W. (2003). Firms' knowledge-sharing strategies in the global innovation system: Empirical evidence from the flat panel display industry. *Strategic Management Journal, 24*, 217–233.

Tapscott, D., Lowi, A., & Ticoll, D. (2000). *Digital capital: Harnessing the power of business webs*. Boston, MA: Harvard Business School Press.

Tsai, W., & Ghoshal, S. (1998). Social capital and value-creation: The role of intrafirm networks. *Academy of Management Journal, 41,* 464–477.

Weill, P., & Vitale, M. R. (2001). *Place to space: Migrating to eBusiness models*. Boston, MA: Harvard Business School Press.

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal, 5*, 171–180.

KEY TERMS

Behavioral/Value Information: Information that represents how large-scale, complex systems, including social systems, interact under different environmental conditions and in conjunction with different stimuli.

Functional Information: Information, based on engineering knowledge, that describes how components and subsystems of a tangible product interact and how these interactions give rise to different product features and varying levels of product performance.

Procedural Information: Easily codifiable information that depicts the steps or protocols that are required to conduct a business process or perform a service.

State Information: Information that describes the state or condition of a physical product or information good (i.e.,a stock price or the number of SKUs in inventory).

Market of Resources for Agile/Virtual Enterprise Integration

Maria Manuela Cunha
Polytechnic Institute of Cávado and Ave, Portugal

Goran D. Putnik
University of Minho, Portugal

INTRODUCTION

Enterprise Networking and Dynamics

Fast change, uncertainty, and strong competition are challenges of the actual worldwide economic context. Competitiveness is a main requirement of enterprises, whose satisfaction requires the definition of new organisational concepts, with extremely high performances that are strongly time-oriented while highly focused on cost and quality. Several factors appear as supreme factors of competitiveness: (1) the organisations' capability to achieve and explore competitive advantages in synergy, by using or integrating the optimal available resources for the functions that the organisation undertakes; (2) the capability of fast adaptability to the market; together with (3) the capability of managing all business processes independently of distance, to be achieved through the recent virtual enterprise (VE) organisational models (Cunha & Putnik, 2002; Cunha, Putnik, & Ávila, 2000).

The need to keep a close alignment with the market environment in permanent change implies the high dynamics of the organisations' structure reconfigurability, introducing a new concept of dynamically reconfigurable global networked structures, corresponding to the emerging agile/virtual enterprise (A/VE) model (Cunha & Putnik, 2004).

The Market of Resources is the environment proposed by the authors to enable A/VE dynamic integration and business alignment.

BACKGROUND

Critical Factors against Networking and Dynamics

Two critical factors against networking and enterprise dynamics are as follows:

1. The *transaction costs*, i.e., the firm reconfiguration cost, associated with partners search, selection, negotiation, and integration as well as permanent monitoring and evaluation of partnership performance. Resource allocation in the market is normally guided through prices, but within the firm, work/job is done through decisions and commands of management (Coase, 1937). Activities are collected in a firm when transaction costs incurred in using the price mechanism exceed the cost of organising those activities through direct managerial controls, i.e., the decision of keeping a function/activity in-house results from the comparison between expected economic loss that can result from an outsourcing contract (transaction costs and contractual risks), and the expected economic gains.
2. *Preservation of firm's knowledge* of organisational and management processes, as it is the firm's competitive factor. The firm incurs the risk of leakage of private information when opt by performing an activity by an independent market firm.

For the efficient implementation of A/VE, it is necessary to conceive tools to overcome the networking and dynamics disabling factors.

Tools for Managing, Controlling, and Enabling Networking and Dynamics According to BM_VEARM Approach

The main tools conceived by the BM_virtual enterprise reference model (BM_VEARM) (Putnik, 2000) for managing, controlling, and enabling networking and dynamics, overcoming the two critical factors, are:

- The Market of Resources is the environment for enabling and managing efficient configuration, and assuring virtuality, at low transaction costs and reduced risk of knowledge leakage.
- The broker or organisation configuration manager[1] is the main agent of agility and virtuality, acting

either between two operations of the A/VE (off-line reconfigurability, providing agility only) or online with the operation (online reconfigurability, providing virtuality and a higher level of agility).
- Virtuality makes possible the transition from one physical structure (instance) to another in a way so that the enterprise or process owner is not affected by the system reconfiguration and is not aware of the reconfiguration—the underlying service structure and reconfiguration process are hidden.

Additionally, A/VE must satisfy the highest level of integration and (geographic) distribution of the A/VE partners.

MARKET OF RESOURCES CHARACTERIZATION

Market of Resources Definition

Market of Resources is an institutionalised organisational framework and service assuring the accomplishment of the competitiveness requirements for A/VE dynamic integration and business alignment. The operational aspect of the Market of Resources consists of an Internet-based intermediation service, mediating offer and demand of resources to dynamically integrate in an A/VE, assuring low transaction costs (demonstrated in Cunha & Putnik, 2003a, 2003b) and the partners' knowledge preservation. Brokers act within the Market of Resources as intermediation agents for agility and virtuality.

In this "virtual" environment, *offer* corresponds to *resources providers* (individuals, enterprises) that make their *resources* (products, components, operations) available, as potential partners for A/VE integration.[2] *Demand* corresponds to *client*, the A/VE owner, the entity looking for resources to create/integrate/reconfigure an A/VE to satisfy the *Customer*. *Customer* is the entity giving rise to a business opportunity and is considered outside the Market of Resources.

The service provided by the Market of Resources is supported by the following:

1. A knowledge base of resources and history of previous performance results
2. A normalised representation of information
3. Computer-aided tools and algorithms
4. Brokers
5. A regulation, i.e., management of negotiation and integration processes, as well as contract enforcement mechanisms

It is able to offer knowledge for resources search and selection and its integration in an A/VE, specific functions of A/VE operation management, and contracts and formalising procedures to assure the accomplishment of commitments, responsibility, trust, and deontological aspects, envisaging the accomplishment of the A/VE objectives.

Information technology (Internet and WWW technologies, agent technology, e-marketplaces, etc.) supports or automates purchasing activities, helping from procurement processes up to the search for partners for a partnership, including electronic automated negotiation, electronic contracting, and market brokerage services. Although the basic IT infrastructures and tools are necessary as support, the added value comes from the higher-level functions, to support search, selection, and integration of resources under the format of an A/VE, coping with the high reconfigurability dynamics requirements (overcoming the disabling factors) intrinsic to the A/VE (Cunha, Putnik, Carvalho, & Ávila, 2002).

Market of Resources Functionality

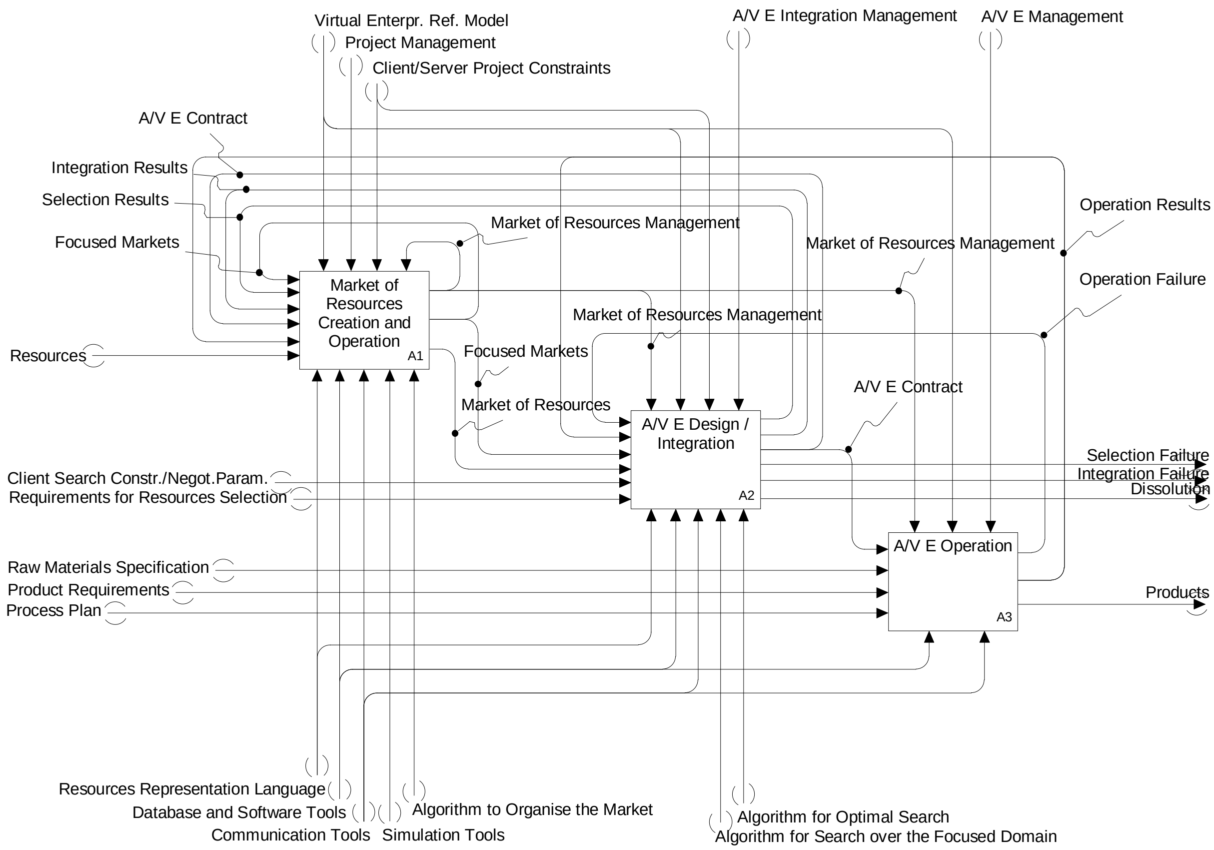
The Market of Resources is designed to offer the following (Cunha, Putnik, & Ávila, 2003; Cunha, Putnik, & Carvalho, 2002):

- Reduction of negotiation time and of time-to-contract, as one instantaneous physical structure (one instance) of an A/VE may last (on a limit) for only a few days or even hours
- The permanent alignment of the A/VE with business requirements, trust of participants, and accomplishment of contracts, requiring a dynamic process of A/VE performance evaluation, monitoring the performance of participating resources, and identifying reconfiguration opportunities
- The ability to find the right potential partners and further efficient negotiation
- Enforcement and risk minimisation by a contractual agreement between the involved parties
- Provision of knowledge/advisory guidance in A/VE design, validation, and reconfiguration, through appropriate algorithms and expert brokerage functions

Market of Resources Organisation

The overall functioning of the Market of Resources (Figure 1) consists of the creation and management of the Market as the environment (Process A.1.) to support the design and integration of the A/VE (Process A.2.) and the A/VE operation (Process A.3.), offering technical and

Figure 1. IDEF0 representation of the global process for the market of resources creation and operation, and for A/VE design, integration, and operation



Source: *Cunha, Putnik, and Gunasekaran, 2002.*

procedural support for the activities of identifying potential partners, qualifying partners, and integrating the A/VE, as well as providing coordination and performance evaluation mechanisms. The model respects the *BM_VEARM*.

The Market of Resources has two components: the organisational or managerial component that integrates the criteria for resources selection, management procedures, control, and environment evaluation; and the infrastructural or informational component (databases). Figure 2 represents how the two components (organisational and informational) are created, for the first time (Process A.1.1.); how the organisational component is maintained (Processes A.1.2. and A.1.3.); and how the Market of Resources operation is managed (Processes A.1.4. and A.1.5.).

Virtual Enterprise Extended Life Cycle

The VE life cycle presented in the literature as comprehending the phases of creation, operation, and dissolution of the partnership (where reconfiguration is a possibility to be considered during peration), traduces the limited ability of the traditional VE model to support dynamics. The extended life cycle introduced by the authors is directly tied with the *BM_VEARM* and the Market of Resources as supporting environment, and includes in it the "traditional" VE life cycle (Figure 3).

It starts with *Identification of the Opportunity* to create an A/VE, followed by the selection (by the A/VE owner) of a Market of Resources where support can be found. After the *contratualisation* with a Market, the process of *A/VE design*, and the search and selection of resources providers toward the *A/VE integration* takes

Figure 2. IDEF0 representation of process A.1—Market of resources creation and operation

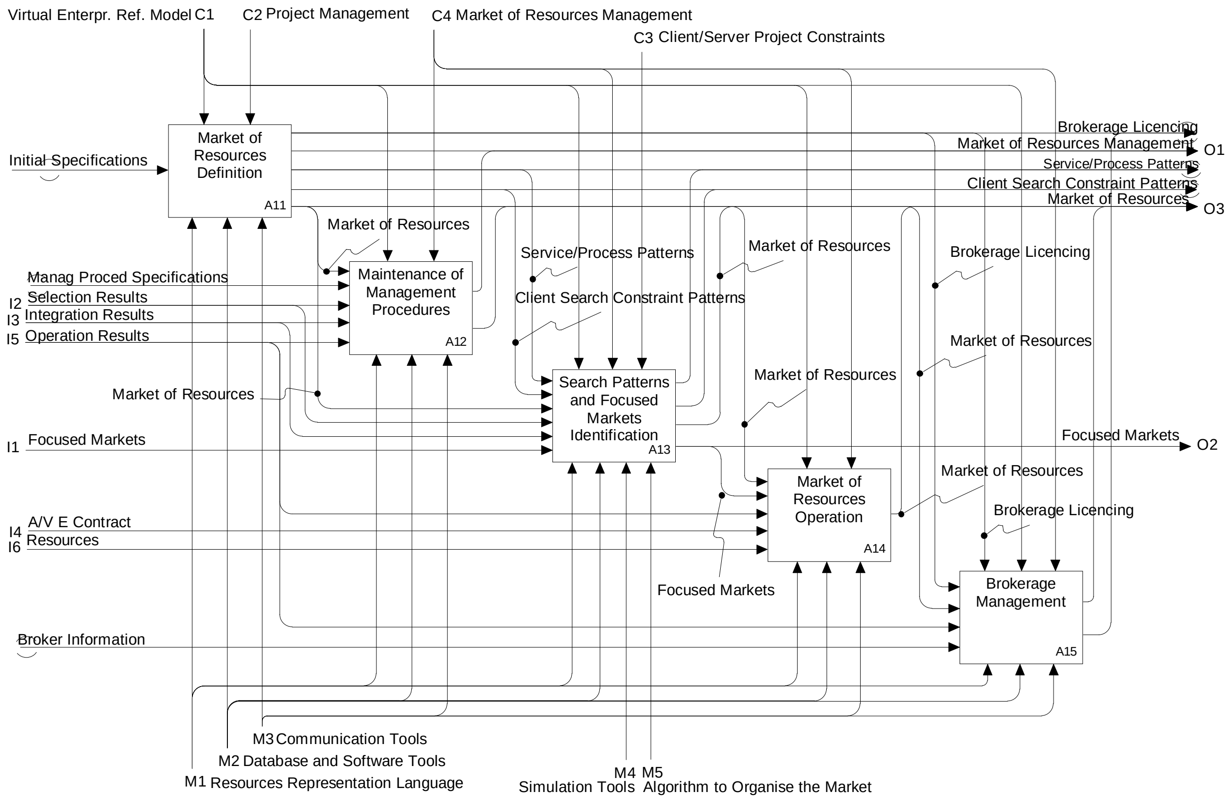
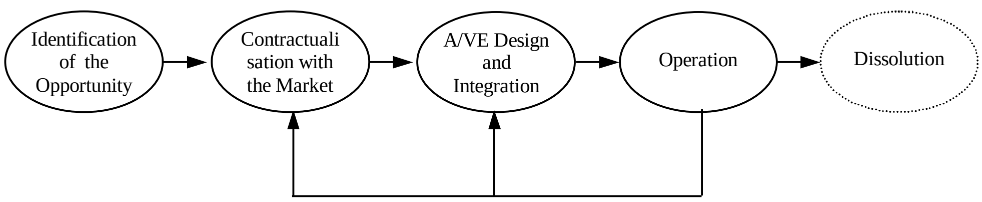


Figure 3. Virtual enterprise extended life cycle (the A/VE life cycle)



Source: Cunha, Putnik, and Ávila (2004)

place. During the *Operation* phase, the A/VE can suffer reconfiguration, or the A/VE owner can even decide to contractualise with another market as an alternative or complementarily. Finally, we have the *Dissolution* phase. Inclusion of the Market of Resources in the A/VE life cycle is necessary to enable the A/VE reconfigurability dynamics.

Supporting Technologies

Table 1 lists some of the Market of Resources functionalities and the corresponding supporting technologies.

Implementation

The Market of Resources is an independent organization, i.e., company, independent both from the offer and the demand sides of A/VE integration, i.e., independent from the Market participants, in order to assure impartiality, fairness, and trust, being able to monitor the activities of all the participants (including brokers) and to enforce the accomplishment of contracts between parties.

In its nonregular implementation, the Market of Resources can be considered as a product and implemented as a function in one company. This implementation loses

Table 1. Technologies to support the main functionalities of the market of resources

Market of Resources functionalities	Supporting technologies and tools
Market contents: user/buyer profile, catalogs, historic, database of resources	- Database management systems - Distributed database management systems - e-Business development platforms - Portals
Negotiation: request for quotes, auctions, optimal selection	- Intelligent agent technology - Electronic negotiation tools - Algorithms or protocols - Regulation of negotiation - Intelligent decision-making systems
Transactions: payment, contractualisation	- Electronic payment - Digital signature - Certification - Other security mechanisms
Management: monitoring, performance evaluation, analysis of operation results, decision making	- Simulation tools - Workflow technology and collaboration techniques - Regulation
Brokerage: expert advice, monitoring, and coordination	- Messaging and conferencing - Algorithms - Management procedures
Integration: file translation, collaboration	- Standards for product/services description - Collaboration tools - Data translation standards and tools - Communication protocols
Resources: final selection	- Algorithms, heuristics, and computer-aided tools - Intelligent decision-making systems - Artificial intelligence

Source: Cunha (2003)

the impartiality as it corresponds to a "buyer-centric" e-marketplace.

Other Contributions Toward A/VE Integration Support

Other similar Market of Resources concepts, services, and products include the new generation of high value-added electronic marketplaces, e-alliances, breeding environments, electronic institutions, and virtual clusters.

FUTURE TRENDS

The Market of Resources is part of an umbrella of research projects under development at the University of Minho addressing virtual enterprise theory, design and control tools and technologies, and the corresponding environment. In particular, the practical objective of the underlying reference model (BM_VEARM) is to serve as a framework for cooperation and coordination of this group of research projects.

Future developments include a more detailed characterisation of the Market of Resources, extended from virtual enterprises to virtual organisations, to cover environments or application domains, such as education (virtual universities), health care, etc., with intrinsic specificities not included in our actual model. The new paradigms and developments in the field of information technologies, such as ubiquitous computing and systems, intelligent software, data mining tools, and intelligent knowledge bases, should also be considered.

CONCLUSION

The A/VE model is of increasing relevance in the organisational panorama, due to its intrinsic agility, dynamic adaptability, and efficiency. The authors introduce the Market of Resources as an environment and enabler that, overtaking the disabling factors of networking and dynamics, is able to cope with the A/VE requirements of reconfigurability dynamics and business alignment.

REFERENCES

Ávila, P., Putnik, G. D., & Cunha, M. M. (2002). Brokerage function in agile/virtual enterprise integration—A literature review. In L. M. Camarinha-Matos (Ed.), *Collaborative business ecosystems and virtual enterprises* (pp. 65–72). Dordrecht: Kluwer Academic Publishers.

Coase, R. (1937). The nature of the firm. *Economica*, *4*, 386–405.

Cunha, M. M. (2003). *Organisation of a market of resources for agile and virtual enterprises integration*. Doctoral thesis, University of Minho, Guimarães, Portugal.

Cunha, M. M., & Putnik, G. D. (2002). Discussion on requirements for agile/virtual enterprises reconfigurability dynamics: The example of the automotive industry. In L. M. Camarinha-Matos (Ed.), *Collaborative business ecosystems and virtual enterprises* (pp. 527–534). Dordrecht: Kluwer Academic Publishers.

Cunha, M. M., & Putnik, G. D. (2003a). Market of Resources versus e-based traditional virtual enterprise integration—A comparative cost analysis. In G. D. Putnik & A. Gunasekaran (Eds.), *Proceedings of the First International Conference on Performance Measures, Benchmarking and Best Practices in New Economy*. Guimarães, Portugal.

Cunha, M. M., & Putnik, G. D. (2003b). Market of Resources versus e-Based Traditional Virtual Enterprise Integration - A cost model definition. In G. D. Putnik & A. Gunasekaran (Eds.), *Proceedings of the First International Conference on Performance Measures, Benchmarking and Best Practices in New Economy*. Guimarães, Portugal.

Cunha, M. M., & Putnik, G. D. (2004). Trends and solutions in virtual enterprise integration. *Tekhné—Review of Politechnical Studies*, *1*(1).

Cunha, M. M., Putnik, G. D., & Ávila, P. (2000). Towards focused markets of resources for agile/virtual enterprise integration. In L. M. Camarinha-Matos, H. Afsarmanesh, & H. Erbe (Eds.), *Advances in networked enterprises: Virtual organisations, balanced automation, and systems integration* (pp. 15–24). Dordrecht: Kluwer Academic Publishers.

Cunha, M. M., Putnik, G. D., & Ávila, P. (2003). Agile/virtual enterprise enablers: A comparative analysis. In *Proceedings of Group Technology/Cellular Manufacturing World Symposium*. Columbus, USA: Ohio University.

Cunha, M. M., Putnik, G. D., & Ávila, P. (2004). Virtual enterprises' extended life cycle. In *Proceedings of SymOrg 2004— IX International Symposium*. Belgrade: University of Belgrade, Faculty of Organisational Sciences.

Cunha, M. M., Putnik, G. D., & Carvalho, J. D. (2002). Infrastructures to support virtual enterprise integration. In R. Hackney (Ed.), *Proceedings of 12th Annual BIT Conference—Business Information Technology Management: Semantic Futures*. Manchester, UK: The Manchester Metropolitan University (CD-ROM).

Cunha, M. M., Putnik, G. D., Carvalho, J. D., & Ávila, P. (2002). A review on environments supporting virtual enterprise integration. In M. Vladimír, L. M. Camarinha-Matos, & H. Afsarmanesh (Eds.), *Balancing knowledge and technology in product and service life cycle* (pp. 133–140). Dordrecht: Kluwer Academic Publishers.

Cunha, M. M., Putnik, G. D., & Gunasekaran, A. (2002). Market of resources as an environment for agile/virtual enterprise dynamic integration and for business alignment. In O. Khalil & A. Gunasekaran (Eds.), *Knowledge and information technology management in the 21st century organisations: Human and social perspectives* (pp. 169–190). Hershey, PA: Idea Group Publishing.

Putnik, G. D. (2000). BM_Virtual enterprise architecture reference model. In A. Gunasekaran (Ed.), *Agile manufacturing: 21st century manufacturing strategy* (pp. 73–93). UK: Elsevier Science.

Silva, J. P., Putnik, G. D., & Cunha, M. M. (2003). Technologies for virtual enterprise integration. In G. Putnik & A. Gunasekaran (Eds.), *BusinessExcellence I—Performance measures, benchmarking and best practices in new economy* (pp. 706–711). University of Minho, Guimarães.

KEY TERMS

Agile/Virtual Enterprise: Dynamically reconfigurable global networked organisation, networked enterprise, or network of enterprises, sharing information, knowledge, skills, core competencies, market and other resources and processes, configured (or constituted) as a temporary alliance (or network) to meet a (fast-changing) market window of opportunity, presenting as its main characteristics agility, virtuality, distributivity, and integrability (see Putnik, 2000). The factors against A/VE are transaction (reconfiguration) costs and preservation of enterprises', or firms' (partners in A/VE) knowledge.

A/VE: Agile/virtual enterprise (see *Agile/virtual enterprise*).

BM_VEARM: BM_Virtual Enterprise Architecture Reference Model: BM_VEARM is a VE reference model conceived to enable the highest organisational/structural/reconfiguration and operational interenterprise dynamics of VE or A/VE (see *Agile/virtual enterprise*), employing three main mechanisms for VE dynamic creation, reconfiguration, and operation: *Market of Resources*, *Broker*, and *Virtuality*. Additionally, BM_VEARM implies the highest level of integration and (geographic) distribution of VE elements (partners in the VE network).

BM_Virtual Enterprise: A virtual enterprise in a total or partial conformance with the BM_Virtual Enterprise Architecture Reference Model (BM_VEARM) (see *BM_VEARM*).

Broker: Also called organisation configuration, structure, organization or architecture manager, it is the main agent of agility and virtuality in A/VE, acting either between two operations of the A/VE (off-line reconfigurability, providing agility only) or online with the operation of the A/VE (online reconfigurability, providing virtuality and a higher level of agility).

Business Alignment: Actions to be undertaken by an organisation, to answer to a market opportunity with the provision of the required product, with the required specifications, at the required time, with the lowest cost, and with the best possible return.

Dynamics (of Organisation): Change in an organisation's structure (see *structural dynamics*) or operation (see *operational dynamics*) along the time, when time as a parameter is indispensable for the organisation, or some aspect of organisation, description and analysis. Otherwise, although the organisation's state changes, if the time as a parameter can be disregarded, the organisation, or some aspect of the organisation, is considered static. The organisation's state changes frequency, state change time, and intensity are examples of an organisation's dynamics features and performance measures.

Enterprise Dynamic Integration: (See *Organisational dynamics*.)

Enterprise Dynamic Reconfiguration: (See *Organisational dynamics*.)

Enterprise Integration: (1) Consists of the establishment of effective and efficient interactions, as well as interaction improvement, among the enterprise's elements, or organisational units and functions, and provides an easy enterprise reconfiguration ["without significant conversion (reconfiguration) costs"]—*intraenterprise integration*. (2) In the case of a virtual enterprise, *enterprise integration* consists of the establishment of effective and efficient interactions among the partners of the VE network and one of the enablers of the VE dynamic reconfiguration—*interenterprise integration*.

Market of Resources: An institutionalised organisational framework and service assuring the A/VE dynamic integration, reconfiguration, and business alignment. The operational aspect of the Market of Resources consists of an Internet-based intermediation service, mediating offer, and demand of resources to dynamically integrate in an A/VE, assuring low transaction costs and partners' knowledge preservation. Brokers act within the Market of Resources as the intermediation agents for agility and virtuality. Its implementation is as an independent organization/company, independent from the offer and the demand sides of A/VE integration, i.e., independent from the Market participants, in order to assure impartiality, fairness, and trust, being able to monitor the activities of all the participants (including brokers) and to enforce the accomplishment of contracts between parties.

Operational Dynamics: Change in an organisation's operation, while the organisation's structure remains unchanged (static), along the time, when time as a parameter is indispensable for the organisation's operations, or some aspect of organisation's operations, description, and analysis [see *Dynamics (of organisation)*]. In the case of A/VE (see *Agile/virtual enterprise*), the operational dynamics refers to *interenterprise operational dynamics*.

Organisational Dynamics: Changing of an organisation's structure, where structure consists of organisation's (company's) elements and relations among them along the time parameter, when time as a parameter is indispensable for the organisation's structure, or some aspect of the organisation's structure, description, and analysis [see *Dynamics (of organisation)*]. In the case of A/VE (see *Agile/virtual enterprise*), the organisational dynamics refers to the changes in the network structure or configuration along the time parameter. Also, *Enterprise dynamic integration*, *enterprise dynamic reconfiguration*.

Resource: This is the entity (individual or organisation) that can contribute or add value to a product, providing either a product (component, assembly) or an operation.

A resource is (a view of) an enterprise object used to perform (or to support the execution of) one or more processes, and it is a subject of control. A resource can be primitive or complex. In the case of A/VE (see *Agile/ virtual enterprise*), a person or enterprise (company), partner in A/VE.

Virtual Enterprise: (See *Agile/virtual enterprise*.)

Virtuality (of Organisation): The specific organisational structure between VE elements/partners and brokers, BM_VEARM (see *BM_VEARM*), providing the feature of virtuality as a further requirement for achievement of the highest VE dynamics. Virtuality makes possible the transition from one A/VE's physical structure (instance) to another in a way that the enterprise or process owner is not affected by the system reconfiguration and is not aware of the reconfiguration—the underlying service structure and reconfiguration process are hidden.

ENDNOTES

[1] See Ávila, Putnik, and Cunha (2002) for a classification of brokerage functions.

[2] A/VE integration, or reconfiguration, includes the processes of search or potential partners, negotiation between them, selection, and integration of the selected in an A/VE.

M

Marketplace Architecture for Enterprise Integration

Hamada H. Ghenniwa
University of Western Ontario, Canada

Michael N. Huhns
University of South Carolina, USA

INTRODUCTION

Businesses today must be fast and flexible, responsive to customers, and cost effective in their operations. They must collaborate more frequently with partners to build virtual organizations and supply-chains that reduce times-to-market and costs. *E-business* is the use of the Internet along with other electronic means and technologies to conduct within-business, business-to-consumer, business-to-business, and business-to-government interactions. A basic model of an e-business is the *e-shop* model, which is based on providing a self-service storefront to a customer by displaying the company catalogs and product offers on a Web site. An *e-procurement* model focuses on the buying aspect of the business. A typical architecture for *e-procurement* consists of a browser-based self-service front-end to the corporate purchasing system or its ERP. The supplier catalogs are presented to end-users through a single unified catalog, thereby facilitating a corporate-wide standard procurement process. Online auction models have also received much attention for automating dynamic trading. Other models are based on creating value-chain businesses, such as service provisioning of specific functions for the value-chain, and electronic payments or logistics.

Although each of the above models attempts to provide an e-business solution, none of them addresses the challenge of how to create and leverage services and supply operations in a way that seamlessly integrates business entities (customers, suppliers, partners, and competitors) in a dynamic trading community. A very important and promising model is the *e-marketplace*. It supports value-chain integration and provisioning in its structure and services. The objective is to develop an e-business solution that relieves business entities of much of the burden of participating effectively in the e-business domain. This model combines the advantages of the sell-side, the buy-side, and the value-chain models.

BACKGROUND AND LITERATURE REVIEW

There have been several recent attempts to promote e-marketplace models by the academic and industrial communities. For example, the Electronic Market-Place (Boll, Gruner, Haaf & Klas, 1999) is an attempt to develop a business-to-business system architecture. It is viewed as a DBMS solution to support many-to-many relationships between customers and suppliers. The Global Electronic Market system (Rachlevsky-Reich, Ben-Shaul, et al., 1999) attempted to develop a logical market framework and infrastructure. In this system, the market provides trading mechanisms that include bids and offers. A more complex architecture for an e-marketplace is MAGMA (Tsvetovatyy, Gini, Mobasher & Wieckowski, 1997), with its special focus on the infrastructure required for conducting commerce on the Internet. OFFER (Bichler, Beam & Segev, 1998) proposed a brokering-based architecture marketplace. A customer can search for a service either directly in the e-catalog of the supplier or use the e-broker to search all the e-catalogs of the suppliers that are registered with this broker. E-brokers employ a simple auction mechanism. MOPPET (Arpinar, Dogac & Tatbul, 2000) proposed an e-marketplace system as agent-oriented workflows. MOPPET viewed the market as a workflow management system carried out by several types of agents: task, scheduling, facilitator, and recovery agents.

Another approach was driven by the bottom-up modeling of market processes with self-organizing capabilities (Arthur, Holland, LeBaron, Palmer & Tayler, 1997). The objective was to develop a computational study of economies modeled as evolving systems of autonomous interacting agents, and known as agent-based computational economics (ACE) (LeBaron, 2000; Timmers, 1999). The ACE researchers relied on computational laboratories (McFadzean, Stewart & Tesfatsion, 2001) to study the evolution of decentralized market economies under controlled experimental conditions.

Several companies have emerged to automate logistics and re-supply within specific industrial segments. For example, Ariba (2000) developed a marketplace based on procurement portals and dynamic exchanges for horizontal marketplaces. The SAP Service Marketplace (SAP AG) is an Internet portal for the SAP community. It provides basic online services, such as catalog browsing, matchmaking, and ordering from SAP and its partners. Other approaches were directed to support vertical marketplaces, such as PaperExchange (PaperExchange Marketplace), which enable customers and suppliers to negotiate pricing and transact directly with one another. VerticalNet (VerticalNet® Marketplaces) also built a set of Web-based marketplaces for specific industrial segments, such as financial services, healthcare, and energy. Each Web site forms a community of vendors and customers in a specific area.

Another direction adopted by major software vendors is to develop Internet-based commerce platforms. Examples are IBM CommercePOINT (IBM Corporation CommercePOINT Payment), Microsoft Site Server Commerce Edition (Microsoft Corporation. Internet Commerce, 1998), Oracle Internet Commerce Server INTERSHOP (Intershop Communications, Inc., 1998), and Sun JavaSoft JECF (Java Electronic Commerce Framework, Sun Microsystems). These proprietary attempts focus on providing infrastructure services, such as security payment directories and catalogs, to be integrated with existing systems and the Web.

In our research work, we view e-marketplace as a cooperative distributed system that integrates participating business entities, including consumers, suppliers, and other intermediaries. This architecture enables and facilitates common economic services and commerce transactions between the buyers and sellers, such as brokering, pricing, and negotiation, as well as cross-enterprise integration and cooperation in an electronic supply-chain. In this architecture, the e-marketplace exists as a collection of economically motivated software agents.

DESIGN ISSUES AND TRENDS

As e-business grows and becomes viable in the real world, its corresponding e-marketplaces must expand to support a broader base of services ranging from baseline interaction and directory services to specialty services, such as dynamic trading, cooperative supply-chain integration, and management. In this new e-marketplace environment, there are significant interactions among the systems deployed by the participating business units of an enterprise, their customers, and other businesses. Therefore, designing e-marketplaces requires embodying greater levels of business knowledge within the e-marketplace transactions, activities, and service definitions. Additionally, it requires a greater degree of communication, coordination, and cooperation within and among the business entities and their systems in the e-marketplace. In other words, the e-marketplace architecture represents an integrated body of people, systems, information, processes, services, and products.

Enterprise Model and Ontologies

At the heart of the integration architecture for an e-marketplace is a model of the enterprise. It is an abstract representation of the structure, activities, processes, information, resources, people, behavior, goals, rules, and constraints of the e-marketplace. From an operational perspective, the enterprise model captures what is planned, what might happen, and what has happened. Therefore, it supplies the information and knowledge necessary to support the operations of an e-marketplace. An appropriate e-marketplace architecture should support enterprise-modeling ontologies. An ontology is a vocabulary along with some specification of the meaning or semantics of the terminology within the vocabulary. The objective is to provide a shared and common understanding of a domain that can be communicated to people, application systems, and businesses. In an e-marketplace model, ontologies are integrated or related to support reasoning among the elements of the model.

Market Structure and Economy Model

An important aspect of the e-marketplace is the economic model of its structure. A market structure governs the trading process and defines the formal rules for market access, traders' interactions, price determination, and trade generations. In classical economic theory there are several market models for specific trading situations and structural behaviors. In the commodity market model, various suppliers and consumers participate to trade goods/services (commodity) of the same type. The market price is publicly agreed upon for each commodity independent of a particular supplier. All consumers and suppliers decide whether and how much to buy or sell at each agreed-upon price. The challenge in this market structure is to deploy a pricing methodology that produces price adjustments that bring about market equilibrium (i.e., equalizes supply and demand). In an auction-based market, each participant (both consumers and suppliers) acts independently and contracts to buy or sell at a price agreed upon privately. An auction-based e-marketplace is a form of centralized facility, or clearinghouse, by which costumers and suppliers execute trades in an open and competitive bidding process.

There are many situations where effective coordination cannot be entirely determined by the market forces. In bargaining, both customers and suppliers have their own objective functions, and they negotiate with each other as long as their objectives are met. The participants can engage in direct negotiations with each other using their respective bargaining strategies to arrive at a "fair" price for a particular item. This market structure does not support a specific negotiation protocol; rather, the participants will use an unrestricted bidding protocol. A major challenge in this structure is how to enable any participant to determine the "fair" price.

Supply-Chain Integration and Management

An e-marketplace can be treated as a physically and logically distributed system of interacting autonomous business entities. Yet, there is a need for well-accepted interoperability standards, which must be meshed for supply-chain integration to meet business demands. Conceptually, a supply-chain manages coordinated information and material flows, production operations, and logistics of the e-marketplace. It provides the e-marketplace with flexibility and agility in responding to customer demand shifts without conflicts in resource utilization. The fundamental objective is to improve coordination within and between various participant business entities in the supply-chain. In an e-marketplace setting, supply-chain management can be viewed as a cooperative distributed problem-solving activity among a society or group formed by autonomous business entities that work together to solve a common problem (Smirnov & Chandra, 2000). The group is responsible for coordination throughout the supply chain, whereas each member provides specialized expert knowledge and product and process technology to the supply chain. The decision-making process is centralized for the group, but decentralized for the local decisions of each member.

Foundation Architecture for Integration

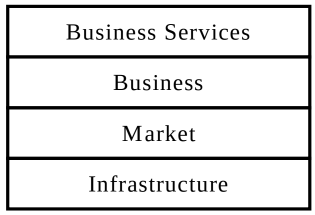
The architecture of the e-marketplace provides the foundation to integrate and leverage the participants' resources, such as applications and databases. Traditionally, the foundation technology that enables enterprises to connect resources together is known as *middleware*. Mainstream middleware solutions focus on integration at the data level. There are several commercial middleware products and standards, such as OMG CORBA™ (Object Management Group, Inc. 1995), J2EE™ (Java™ 2 Platform, Enterprise Edition), and .NET (which has superseded DCOM, the Distributed Component Object Model), that focus on providing infrastructure tools and frameworks of integration. Enterprise application integration (EAI) is a trend that has recently emerged in designing middleware technology with an objective to ease the burden and lower the costs of application integration. However, different EAI solutions are developed to accommodate different levels of integration requirements. Object-level integration provides synchronization of data between different applications or databases. Business process-level integration extends the object level by supporting multiple, distributed, and heterogeneous applications. Finally, cross-enterprise process-level integration involves multiple, distributed, heterogeneous business-process applications across different enterprises. While EAI solutions focus on technology-centered integration, other approaches focus on integration as an architectural aspect. One approach is a mediator-based architecture (Wiederhold, 1992), which comprises a layer of "intelligent" middleware services to link data resources and applications, such as integrating data from multiple sources in a way that is effective for the client application. Another approach is the facilitator (Genesereth, 1992), in which integration is based on the principle that any system (software or hardware) can interoperate with any other system without the intervention of human users or their developers. This level of automation depends on supporting ontologies to describe the resources. Facilitators use meta-level information in converting, translating, or routing of data and information.

FUTURE TRENDS: AGENT-ORIENTED E-MARKETPLACE

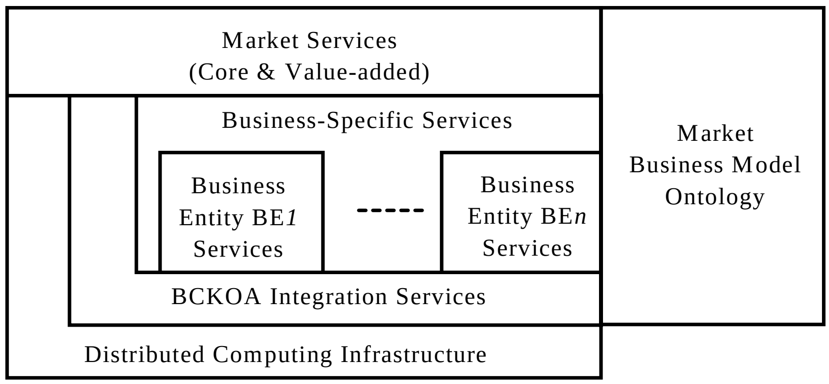
The development of an architecture for an e-marketplace requires a new design paradigm, improved integration architectures, and services. In this architecture, the e-marketplace is a cooperative distributed system composed of economically motivated software agents that interact cooperatively or competitively, find and process information, and disseminate it to humans and to other agents. The architecture also supports common economic services and commerce transactions, such as pricing, negotiation, and automated supply chains, as well as cross-enterprise integration and cooperation.

Fundamentally, integration can be viewed as an abstraction level at which a distributed system environment can be described as a collective coherent universe of cooperative entities. In a cooperative distributed system, integration is captured at the foundation architecture that supports all the entities' individual architectures, and therefore, the complete computing environ-

Figure 1. Use of BCKOA for the architecture of an e-marketplace



(a) Abstraction layers for *e-marketplace*



(b) BCKOA-based *e-marketplace*

ment. Here we describe a business-centric knowledge-oriented architecture (BCKOA) for cooperative distributed systems. The main elements of BCKOA include domain services, integration services, and a domain ontology. Furthermore, BCKOA provides three families of integration services: ontology and semantic integration services, coordination and cooperation services, and wrapping services. Based on our premise that an e-marketplace is a coherent service-oriented universe, the BCKOA-based e-marketplace is shown in Figure 1(b), which builds upon the abstraction architecture of the e-marketplace in Figure 1(a) (Ghenniwa, 2001).

All services (business, market, and integration) in a BCKOA-based e-marketplace usually involve complex and non-deterministic interactions, often producing results that are ambiguous and incomplete. Auctions and *ad hoc* service integrations are some examples. In addition, the dynamic nature of the environment requires that the components of the system be able to change their configuration to participate in different, often simultaneous roles in e-marketplaces. These requirements could not be accomplished easily using traditional ways of manually configuring software. For this domain, we strongly believe that agent-orientation is a very promising design paradigm for integration. In fact, such a paradigm is essential to model an open environment such as an e-marketplace, especially considering the multiple dynamic and simultaneous roles a single business entity may need to participate in, given e-marketplace sessions.

Agent technology provides the next step in the evolution of paradigms for computational modeling, programming methodology, and software engineering (Huhns, 2001, 2003). The first principle of agenthood is that an agent should be able to operate as a part of a community of cooperative distributed systems, including human users. In our view, an agent model such as Coordinated Intelligent and Rational, Agent (CIR-Agent) (Ghenniwa, & Kamel, 2000) can be described as an individual collection of primitive components that provide a focused and cohesive set of capabilities. In the context of a BCKOA-based e-marketplace, an agent has a role that is related to a specific service category and is able to coordinate, cooperatively or competitively, with the other agents, including humans. An agent's role can be categorized as user-interface, business-specific service, business-entity service, market service, or integration service.

User interface agents play an important and interesting role in many applications. The main functionality of user interface agents is to support and collaborate with users in the same work environment to achieve the users' goals.

Business-specific service agents are specialists that provide a collection of business services available in the e-marketplace. Performing the functionality of a business service is typically the cooperative integration of several agents, including business-entity agents and market service agents.

A business-entity service agent may be a representative in the e-marketplace for some functionality that is based on legacy applications or libraries, such as a product catalog Web site. Market service agents are specialists that provide a collection of functions for the generic e-business in an e-marketplace environment in which a single entity (usually an agent) can perform its tasks in the e-marketplace.

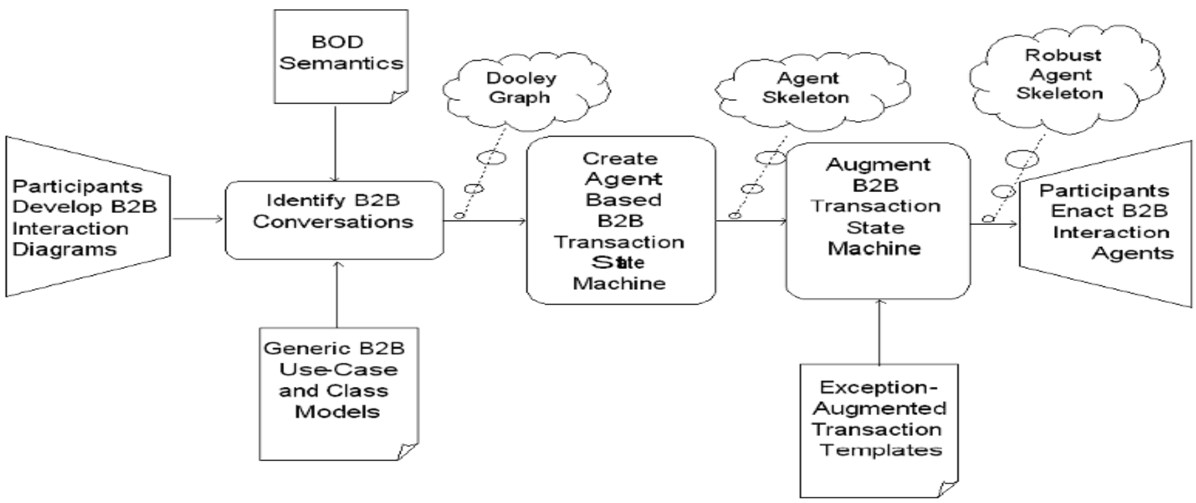
Market service agents are horizontal, in the sense that they are used in several business domains by several business entities. Examples of core services are dynamic trading services, such as commodity market and Vickery auctions (Wellman, 1993; Varian, 1995), and supply-chain integration and management. The commodity market service governs the trading behavior of the participant business entities in the session. This service recognizes three types of agents, namely, market-mediators, consumers, and suppliers. Consumers and suppliers are roles assigned to agents of type business-entity service or user interface. These roles are assigned upon registration with the market session. Each market session is assigned to a mediator to coordinate the actions taken by consumers and suppliers in a way that will eventually clear its respective market. An alternative market session is the auction market. It recognizes three types of agents, representing suppliers, an auctioneer, and buyers. However, the trading process mainly involves the auctioneer and buyers (or bidders). Each bidder agent declares its valuation function to the auctioneer. Under the general Vickery mechanism, it is in the interest (the dominant strategy) of the bidder to report its true valuation function.

Agent-based supply-chain integration and management, in an e-marketplace, is a cooperative distributed problem-solving service. Using this service, business participants need only describe their supply processes using OAG standard business documents and UML interaction diagrams. The methodology, summarized in Figure 2, uses—and begins to formalize—the business object documents (BODs) that OAG and RosettaNet are standardizing. It provides a basis for the convergence of multiple standards for supply-chain management, which could become ready-to-use technology by different participant business entities in the e-marketplace.

Integration agents are specialists that provide a collection of integration functions for a cooperative distributed system in which a single entity (agent, component, object, etc.) can perform its tasks. Integration services are used by several distributed entities.

To validate and experiment with our analyses and foundations described in the previous sections, we have developed a prototype of an agent-oriented BCKOA for an e-marketplace with virtual business entities. They register with the e-marketplace for both purchase and sales services. Both services use a BCKOA-based computation environment. Individual customers or business-entity personnel in the e-marketplace can participate in the market through their user interface agents. Similarly, an agent in the e-marketplace represents each business-entity service. These agents provide thin, intelligent, autonomous interfaces for the business-entity services, which might be based on legacy applications. The implementation utilizes the JADE platform (Bellifemine, Poggi & Rimassa, 1999), which is a software framework to develop agent applications in compliance with the FIPA specifications (The Foundation for Intelligent Physical

Figure 2. Agent-based coordination methodology for B2B automation



Agents, 1998) for multi-agent systems. Although our implementation takes advantage of the JADE platform and FIPA specifications, the architecture of the economic market structure and the application agents is based on the CIR-Agent model and BCKOA.

CONCLUSION AND FUTURE WORK

This article presented state-of-the-art research on developing an agent-oriented architecture for an e-marketplace that provides intelligent enterprise integration. The objective is to establish an engineering foundation for an e-marketplace. To this end, several business and design issues have been discussed and analyzed for e-marketplaces. This article has emphasized an agent-based business-centric and knowledge-oriented approach. An agent-oriented architecture provides an abstraction of the domain entities and applications independent of any specific technology. An e-marketplace supports several types of agents, namely, user-interfaces, business-specific services, market services, and integration services. In continuing the research, our main concern will be investigating the computational effectiveness of agent-oriented e-marketplaces, as well as exploring the appropriate techniques to support secure, reliable, and effective transactions.

REFERENCES

Ariba.(2000, October 19). *B2B marketplaces in the new economy*. Retrieved from *www.commerce.net/other/research/ebusiness-strategies/2000/00_07_r.html*.

Arthur, W.B., Holland, J., LeBaron, B., Palmer, R. & Tayler, P. (1997). Asset pricing under endogenous expectations in an artificial stock market model. In W.B. Arthur, S.N. Durlauf & D.A. Lane (Eds.), *Proceedings on the Economy as an Evolving Complex System*.

Arpinar, S., Dogac, S.A. & Tatbul, N. (2000). An open electronic marketplace through agent-based workflows: MOPPET. *International Journal on Digital Library, 3*(1).

Bellifemine, F., Poggi, A. & Rimassa, G. (1999, April). JADE—a FIPA-compliant agent framework. *Proceedings of PAAM'99* (pp. 97-108), London.

Bichler, M., Beam, C. & Segev, A. (1998). OFFER: A broker-centered object framework for electronic requisitioning. *Proceedings of the IFIP Conference on Trends in Electronic Commerce*.

Boll, S., Gruner, A., Haaf, A. & Klas, W. (1999). EMP—a database-driven electronic marketplace for business-to-business commerce on the Internet. *Journal of Distributed and Parallel Databases, Special Issue on Internet Electronic Commerce*, *7*(2).

Distributed Component Object Model (DCOM). (n.d.). Retrieved from *www.microsoft.com/com/tech/DCOM.asp*.

Genesereth, M. (1992). An agent-based approach to software interoperability. *Proceedings of the DARPA Software Technology Conference*.

Ghenniwa, H. (2001, November). E-marketplace: Cooperative distributed systems architecture. *Proceedings of the 4th International Conference on Electronic Commerce Research*, Dallas, Texas, USA.

Ghenniwa, H. & Kamel, M. (2000). Interaction devices for coordinating cooperative distribution. *Intelligent Automation and Soft Computing*, *6*(2), 173-184.

Huhns, M. (2003). Software agents: The future of Web services. In R. Kowalcyk, J.P. Müller, H. Tianfiled & R. Unland (Eds.), *Agent technologies, infrastructures, tools, and applications for e-services* (pp. 1-18). Berlin: Springer-Verlag (LNAI 2592).

Huhns, M. (2001). Interaction-oriented programming. In P. Ciancarini & M. Wooldridge (Eds.), *Agent-oriented software engineering* (pp. 29-44). Berlin: Springer-Verlag (LNAI 1957).

IBM Corporation CommercePOINT Payment. (n.d.). Retrieved from *www.internet.ibm.com.commerce point.payment*.

Intershop Communications, Inc. (1998). *Intershop 3*. Retrieved from *www.intershop.com*.

LeBaron, B. (2000). Agent-based computational finance: Suggested readings and early research. *Journal of Economic Dynamics and Control, 24*, 679-702.

McFadzean, D., Stewart, D. & Tesfatsion, L. (2001). A computational laboratory for evolutionary trade networks. *IEEE Transactions on Evolutionary Computation, 5*, 546-560.

Microsoft Corporation. Internet Commerce. (1998). Retrieved from www.microsoft.com.

Oracle Corporation. (n.d.). *Oracle Internet Commerce Server*. Retrieved from www.oracle.com/products/asd/ics/ics.html.

Paperexchange Marketplace. (n.d.). Retrieved from www.paperexchange.com.

Rachlevsky-Reich, B., Ben-Shaul, I., et al. (1999). GEM: A global electronic market system. *Information Systems*

Journal, *Special Issue on Electronic Commerce*, *24*(6).

SAP Services Marketplace, SAP AG. (n.d.). Retrieved from www.sap.com.

Smirnov, A. & Chandra, C. (2000, March 20-22). Ontology-based knowledge management for cooperative supply chain configuration. *Proceedings of the AAAI Spring Symposium Bringing Knowledge to Business Processes*. Stanford, CA: AAAI Press.

Sun Microsystems. (n.d.). *Java Electronic Commerce Framework (JECF)*. Retrieved from www.javasoft.com/products/commerce.

Timmers, P. (1999). *Electronic commerce: Strategies and models for business-to-business trading*. New York: John Wiley & Sons.

Tsvetovatyy, M., Gini, M., Mobasher, B. & Wieckowski, Z. (1997). MAGMA: An agent-based virtual market for electronic commerce. *Journal of Applied Artificial Intelligence, Special Issue on Intelligent Agents*, *11*(6), 501-523.

Varian, H.R. (1995). Mechanism design for computerized agents. *The First USENIX Workshop on Electronic Commerce*, *11*(19), 13-21.

VerticalNet® Marketplaces. (n.d.). Retrieved from www.VerticalNet.com.

Wellman, M.P. (1993). A market-oriented programming environment and its application to distributed multicommodity flow problems. *Journal of Artificial Intelligence Research*, *1*, 1-22.

Wiederhold, G. (1992). Mediators in the architecture of future information systems. *IEEE Computer*, *25*(3), 38-49.

KEY TERMS

Agent Orientation: The next step in the evolution of computational modeling, programming methodologies, and software engineering paradigms. Aspects of agent orientation include both cooperative and competitive interactions, knowledge, economic and logical rationality, and learning, all of which are useful for designing distributed computations in open dynamic environments.

Auction-Based Market: A form of centralized facility or clearinghouse by which consumers and suppliers execute trades in an open and competitive bidding process.

Commodity-Based Market: A form of a market in which various suppliers and consumers participate to trade goods and services, i.e., commodities, of the same type. The market price is publicly agreed upon for each commodity independent of any particular supplier. All consumers and suppliers decide whether and how much to buy or sell at each agreed-upon price.

E-Business: The use of the Internet along with other electronic means and technologies to conduct within-business, business-to-consumer, business-to-business, and business-to-government interactions.

E-Marketplace: An electronic marketplace. It is a business model for a particular kind of e-business, which aggregates potentially large numbers of business partners (including buyers, sellers, and intermediaries) and allows them to interact according to a variety of market structures, such as a commodity market, an auction, or an exchange. The result can be significant cost savings.

Enterprise Integration: Refers to the plans, methods, and tools aimed at modernizing, consolidating, and coordinating software applications among a group of businesses or organizations that interact as consumers and suppliers. Enterprise integration might involve developing a total view of the organizations' businesses and applications, seeing how existing applications fit into the new model, and then devising ways to efficiently reuse what already exists while adding new applications and data. Enterprise integration is done for the mutual benefit of all organizations involved.

Market Structure: Governs the trading process and defines the formal rules for market access, traders' interactions, price determination, and trade generations.

Ontology: A representation of knowledge specific to some universe(s) of discourse. It is an agreement about a shared conceptualization, which includes conceptual frameworks for modeling domain knowledge and agreements about the representation of particular domain theories.

Supply-Chain Management: All services that help an enterprise manage its procuring, purchasing, warehousing, distributing, consuming, and monitoring of the supplies it utilizes in its operation. These services enable management to receive the information needed for making decisions along the chain, from acquiring raw materials to manufacturing products to distributing finished goods to retailers. The services consist of consulting, project design, compliance reporting, deployment of resources, and product qualification along the entire chain.

Measurement Issues in Decision Support Systems

William K. Holstein
The College of William and Mary, USA
The Graduate School of Business Administration-Zurich, Switzerland

Jakov Crnkovic
University at Albany, State University of New York, USA

INTRODUCTION

The past decade has seen tremendous progress in systems for information support-flexible and adaptable systems to support decision makers and to accommodate individual needs and preferences. These model- or data-driven or hybrid systems incorporate diverse data drawn from many different internal and external sources. Increasingly, these sources include sophisticated enterprise resource planning systems, data warehouses, and other enterprise-wide systems that contain vast amounts of data and permit relatively easy access to that data by a wide variety of users at many different levels of the organization. Decision support and decision support systems (DSS) have entered our lexicon and are now common topics of discussion and development in large, and even in medium-sized, enterprises. Now that DSS is well established, attention is turning to measurement and the metrics that populate such systems.

BACKGROUND

Decision making as we know it today, supported by computers and vast information systems, is a relatively recent phenomenon. But the concept has been around long enough to permit the methods and theories of decision making to blossom into "a plethora of paradigms, research schools, and competing theories and methods actively argued by thousands of scientists and decision makers worldwide" (Robins, 2003).

Early computer systems focused primarily on accounting and financial data. It is said that information systems are about transforming data. We could say that early systems transformed data into aggregated or summarized data-for example, wage rates, hours worked, benefits and tax data, and so forth, transformed into departmental or corporate payroll reports.

In the mid-1960s, the development of the IBM System 360 and rapidly proliferating competitive systems from other vendors ushered in the era of management information systems (MIS). Applications quickly moved beyond finance and accounting data and into operations. Transaction processing systems began to generate order, usage, and customer data that could be analyzed with (what quickly became quite sophisticated) models. The transformation of data into information became commonplace. For example, data on sales and usage, costs, supplier lead times, and associated uncertainties were transformed into reorder points, safety stocks, and comprehensive inventory management and production scheduling systems.

Despite the broader reach of MIS, such systems are characterized by highly structured, infrequent reports, often with standard formatting. Frequently, because it was 'easier' (for the IT staff), each manager in a given function (e.g., marketing) received the same voluminous report-even though a manager of activities in Japan could not care less about data relating to New Jersey. Despite the tremendous advance of MIS over previous-generation systems, contemporary MIS systems draw most of their data from enterprise resource planning (ERP) systems that contain mostly internal data on transactions, and therefore suffer from many of the same problems as older systems (an internal, historical, and financial focus).

DSS "evolved from the theoretical studies of organizational decision making done at the Carnegie Institute of Technology during the late 1950s and early '60s and the technical work on interactive computer systems, mainly carried out at the Massachusetts Institute of Technology in the 1960s" (Keen & Scott Morton, 1978; Power, 2003). By the end of the 1970s, it was clear that model-based decision support had become a practical, useful tool for managers.

A 1970 article by John Little of MIT clarified the concept of decision support (Little, 1970). In a 1979 paper he provided a definition that is paraphrased here:

"…a coordinated collection of data, systems, tools, and techniques along with requisite software and hardware, by which an organization gathers and interprets relevant

information from the business and environment and turns it into a basis for action." (Little, 1979)

Another useful definition of a DSS is:

"Decision support systems couple the intellectual resources of individuals with the capabilities of the computer to improve the quality of decisions. [They comprise] a computer-based system for management decision makers who deal with semi-structured problems." (Gorry & Scott Morton, 1989)

In these two definitions, we see some important concepts-gathering and interpreting relevant information (related to the decision at hand, not just to transactions), using the intellectual resources of managers, and providing information that can be used as the basis for action. The 'new idea' here was that managers need more than information, they need decision support. If provided with good data, and models and tools to transform the data into useful information, their effectiveness will improve.

METRICS OF BUSINESS AND MANAGEMENT PERFORMANCE

Decision support means supporting managers who are running the business. Increasingly, it refers to supporting middle-level managers who rely on a mix of internal and external data that is steadily tilting towards external data on customers, markets, competitors, and the political, regulatory, and economic environment. If we define the process of *control* as tasks undertaken by middle- and lower-level managers to *ensure that plans come true*, we see clearly the role of data and information in decision support: managers use data and convert it into information to *monitor* the implementation of plans to ensure that strategic goals are met. If the monitoring indicates that plans will not be fulfilled, corrective *action* must be taken *in time* to ensure that the plan is, in fact, met. If the information from a decision support system cannot serve as the basis for action (i.e., cannot first help the decision maker to decide to do something, and then help to decide what to do), the information will not be used and the system will therefore be useless.

The key words in the previous paragraph that lead to action are *monitoring* and *in time*. Monitoring is the management function that is the primary target for DSS implementation. Timeliness is crucial. Advance warning without enough time to steer around the iceberg, or to make the necessary changes to ensure that strategic plans are successful, is not the kind of decision support that managers seek.

As we think about supporting management decision making, we must think of how managers work at decision making. What they do is easy to describe (despite the fact that it is fiendishly difficult to do it): managers abhor irregularities and plans that do not come true, yet they thrive on exceptions. They look for things that don't fit, for things that look funny, for things that are out of line. Then they ask 'why?' Much of their time is spent trying to answer that simple question and searching for actions that will make perceived problems disappear and bring things back to 'normal expectations'. Examples of the 'whys' that plague managers of large companies include: Why can't Cadillac attract younger buyers? Why did the PC manufacturers who dominated the market in the 1990s lose so much share to Dell Computer? Why do practically none of the profits of the newly merged HP/Compaq come from PCs? Which of the newly merged HP/Compaq's 85,000 products should survive? (Abruptly killing off redundant products might scare customers and deflate revenues. But overlap is costly.)

For each of these questions, one can imagine a manager who is conjuring the question as a response to a perceived exception that needs to be 'fixed'. What Cadillac sales manager thinks that his/her product is not attractive to young buyers? What HP manager thinks that Dell's share should be where it is? What former Compaq manager thinks his/her product should be dropped?

We cite this process and these questions to focus clearly on metrics of business performance and management performance. Measurement and metrics are the tools for identifying exceptions. Exceptions, in turn, drive management to seek and find actions that will deal with the exceptions and achieve strategic goals. But think for a moment about traditional metrics. The Cadillac manager knows how to measure the average age of Cadillac buyers. But how does s/he measure the potential attractiveness to younger buyers of a proposed new model? Of a proposed advertising campaign? Of a proposed discount or rebate program? This is where judgment, experience, and intuition come into play-and these are precisely the areas where managers need decision support.

In the current environment, measurement must be related to business matters, business strategies and goals-the stuff that managers deal with in their everyday environment. They are trying to formulate and monitor plans that reflect the strategic mission and goals of the business, that is, accomplish strategic tasks. They need IT that can add value to the business in ways that they can clearly understand.

FUTURE TRENDS

Pulling together what we have stated thus far, we see a situation that precludes significant progress in the development and implementation of decision support systems unless:

- new metrics that focus more clearly on business and management performance are developed and implemented; and
- more attention is given to metrics that focus on monitoring strategic activities, or activities that will have an important effect on the outcomes of strategic initiatives and high-level company goals.

An early initiative in which MIT aimed to provide decision support to higher-level executives (Executive Information Systems or EIS) worked to include 'critical success factors' as the basis for metrics in information systems (Rockart, 1979; Rockart & Treacy, 1982). These systems were broader than DSS and provided a wide variety of information-critical success metrics (metrics for areas of activity in which results, if they are satisfactory, will ensure successful competitive performance for the organization, and which should receive constant and careful attention from management), key performance indicators, reports with the ability to drill down to underlying detail, budget information, plans and objectives, competitive information, news, and more. More recently, the Balanced Scorecard and other initiatives have made it clear that traditional MIS systems and business metrics rely too heavily on financial and accounting measures. Today, the issue of metrics is very active, with performance management systems, digital dashboards and cockpits, and scorecards to drive organizational performance measurement and management down to the individual level in order to support alignment with business strategies (Volonino & Watson, 1990-1991; Power, 2003).

Financial measures tend to be hard, historical, and internal, rather than soft (including judgmental data and estimates that, while 'inaccurate,' are vital to future planning), future oriented (and therefore, by definition, soft), and external (related to the customer, the market, and the environment). We see, therefore, that an important question in the future of metrics in decision support systems is not just *what* to measure, but *how* to measure, and in which areas of the business to seek meaningful metrics.

We turn now to implementation issues surrounding these ideas and suggest some guidelines for DSS development incorporating new metrics.

IMPLEMENTATION GUIDELINES

Breadth in Measurement

A first guideline, as we have discussed, is that breadth in measurement (beyond financial and accounting measures) is important. The ideas behind the Balanced Scorecard (Kaplan & Norton, 1992) should be understood and implemented in decision support systems. Breadth, however, does not imply complexity.

Consider the following quote-it may be true since it does not say "*only* three things"-and it fits the point here:

"The ability of companies to boost profits depends on three things: how high they can lift their prices, how much they can increase output per worker, and how fast wages are rising." (Mandel, 2002)

The ability to raise prices, often described as 'pricing power,' is related to many different metrics and measures. Some of these measures are financial, profitability of other divisions, products and services, and capital structure; but many are not, customer perception of the company's value proposition, features and benefits of products and services, embedded technology, total quality (not just product quality), competitive position in the marketplace, relationship with customers, suppliers, and workers. Increasing worker productivity involves similarly complex metrics.

Simplicity

A second guideline is simplicity. Metrics must be easy to understand and communicate. They must relate to relevant activities and tasks, and be 'drivable,' that is, managers must be able to use the measures to determine (drive) actions that will affect future results.

Selectivity

Our third suggestion relates to simplicity: DSS designers must avoid the impulse to measure everything. The focus must be only on the most important metrics from the user's (the decision maker's) point of view. Metrics should be built in hierarchies, with more details for lower-level managers, and fewer, more summarized measures for higher-level managers.

Research and Learning

Fourth, DSS design should emphasize data and data collection, not just for reporting, evaluation, and auditing, but for research and learning, for finding exceptions, for learning the root causes of exceptions, and for exploring alternative courses of remedial action. 'Research' is perhaps a strange term to use in the context of management practice, but research, in the best sense of the term (investigate, study, explore, delve into, examine) is required to find, first, meaningful measures of exceptions and then their causes.

Benchmarking

As a fifth guideline, we cannot overemphasize the importance of benchmarking against credible external targets. Indeed, without a firm connection to good external benchmarks (best practice, best-of-class indicators), companies can fall victim to *manumation*, simply automating old, outdated, manual processes. A formula for manumation has been around for years:

> OP + NT = EOP
> (**Old Processes** plus **New Technology** equals **Expensive Old Processes)**

Benchmark analysis can identify problems and suggest solutions, and can serve as an excellent idea bank for new metrics. Benchmarks or comparatives are important because metrics need anchor points for comparison. Without a benchmark for 'normal' or 'best in class', how can you gauge results?

Simple benchmarks might include your own past performance, current goals, customer expectations for things like order-to-delivery time, percentage defects, or on-time delivery. In particular, you should know how you compare to others in your industry and leaders in your functional area. Be sure to think carefully about which comparatives will lead to valid conclusions and sensible action.

In some industries and functions, there are a growing number of highly useful benchmarks from trade associations, consulting companies, and other organizations. An example of a widely used set of metrics is the Supply Chain Council's Supply Chain Operations Reference (SCOR) model (Supply-Chain Council, 2004). SCOR allows companies to objectively measure their supply chain practices and compare them against benchmark standards gathered from the more than 700 manufacturing and related companies that are members of the Supply Chain Council. The SCOR model groups supply chain functions into five process categories: Plan, Source, Make, Deliver, and Return. Metrics at each level of the model are supported by progressively more detailed metrics for processes at lower levels.

The following quote highlights the importance of metrics to support process change when new factors, such as Web-enabled processes, are introduced:

"Updating performance metrics for Web-enabled supply chain operations became important the day your company migrated from an information-only Web site to one with interactive customer capabilities. Whether you're tracking customer orders or collaborating with partners on products and processes, your new business dynamics likely don't fit the old criteria, and you may need to update your basic supply chain performance measurements to match an advanced level of attainable objectives." (Schultz, 2001)

Benchmarking is useful beyond performance measurement. It can help to answer the 'why?' when exceptions are identified. For example, is a manufacturer's frequent late delivery an inventory-level issue? Or is it caused by slow order-reaction-time on the part of one or more of the supply chain's participants? Benchmarking can help to target on the exact answer, which can then lead to needed adjustments.

Time

As a sixth and final guideline, we suggest careful consideration of collapsing time. Speeding up of business processes and reaction times has become almost a cliché, but time is the most important element in many new metrics to support decision making. Consider this example from the previously quoted article:

"If I said today that I need 30 percent more of something, how fast can my suppliers, and their component suppliers, deliver that? [Only] part of the answer is in manufacturing lead time. [The other, equally important] part is administrative lead times-getting the information, sharing it, and synchronizing it."

CONCLUSION

The previous quote deftly summarizes much of what we have discussed. We see a clear exception-the need to increase output by 30%, quickly-and the immediate following need to organize the required information internally, and then to communicate that information externally to suppliers and on to their suppliers.

The key element in the processes that are involved here is time, not cost. So, first, metrics to ensure that these process are working properly will be drawn largely from non-financial data.

Next, there is a pressing need to focus only on the most important, but not necessarily the most obvious issues, and to collect relevant data to support understanding of how the task will be accomplished. Much of the learning will be based on intensive communication, with suppliers and their suppliers, probably much of it Web based.

Previous benchmarking outside the organization with suppliers and processes would be highly valuable at this point-indicating who can perform and who cannot.

Summarizing our implementation guidelines for the development of metrics for decision support systems:

- Think beyond ROI and payback metrics for IT investments (Renkema, 2000); measurement for business and management performance is more important.
- Think beyond financial measures.
- Focus only on the most important, but not necessarily obvious issues, and collect relevant data to support exploration, analysis, and understanding.
- Benchmark outside the organization and build relevant knowledge to support change.
- Do it fast, and make it possible for users of the system to work quickly as well.

REFERENCES

Gorry, G.A. & Scott Morton, M.S. (1989). A framework for management information systems. *Sloan Management Review, 13*(1).

Kaplan, R.S. & Norton, D.P. (1992). The Balanced Scorecard-measures that drive performance. *Harvard Business Review*, (January-February). See also the *Balanced Scorecard Collaborative* at www.bscol.com.

Keen, P. & Scott Morton, M.S. (1978). *Decision support systems: An organizational perspective*. Reading, MA: Addison-Wesley.

Little, J.D.C. (1970). Models and managers: The concept of a decision calculus. *Management Science, 16*(8).

Little, J.D.C. (1979). Decision support systems for marketing managers. *Journal of Marketing, 43*(Summer).

Mandel, M.J. (2002). More productivity, more profits? *Business Week,* (June 10).

McCosh, A.M. & Scott Morton, M.S. (1978). *Management decision support systems*. London: Macmillan.

Power, D.J. (2003). A brief history of decision support systems. *DSSResources.com*. Version 2.8 retrieved May 31, 2003, from *DSSResources.com/history/dsshistory.html*.

Renkema, T.J.W. (2000). *The IT value quest: How to capture business value of IT-based infrastructure*. New York: John Wiley & Sons.

Robins, E. (2003). *A brief history of decision-making*. White Paper from the Technology Evaluation Corp. Retrieved from *www.technologyevaluation.com*.

Rockart, J.F. (1979). Chief executives define their own data needs. *Harvard Business Review*, (March-April).

Rockart, J.F. & Treacy, M.E. (1982). The CEO goes on-line. *Harvard Business Review,* (January-February).

Rockart, J.F. & DeLong, D.W. (1988). *Executive support systems: The emergence of top management computer use*. Homewood, IL: Dow Jones-Irwin.

Schultz, G. (2001). Advanced performance metrics for the e-era: Your old gauge for measuring and meeting operational and profitability goals may need recalibration. *Technology Edge*, (June).

Scott Morton, M.S. (1967). *Computer-driven visual display devices-their impact on the management decision-making process*. Doctoral Dissertation, Harvard Business School, USA.

Scott Morton, M.S. & McCosh, A.M. (1968). Terminal costing for better decisions. *Harvard Business Review*, (May-June).

Supply-Chain Council. (2004). *Supply-Chain Operations Reference Model overview, Version 5.0*. Retrieved from *www.supply-chain.org*.

Volonino, L. & Watson, H.J. (1990-1991). The Strategic Business Objectives (SBO) method for guiding executive information systems development. *Journal of Management Information Systems,* (Winter).

KEY TERMS

Actionable Information: Information that can be used as the basis for a decision, or for taking action, usually to change something.

Balanced Scorecard: An approach for measuring business and management results that goes well beyond financial metrics. Several ‘perspectives’ are suggested, e.g., financial, customer, internal processes, and learning and innovation (Kaplan & Norton, 1992).

Benchmark: A standard, usually from outside sources and usually representing the best or better-than-average performance against which an activity’s metric is compared. For example, world-class competitors have 35 defects per unit within the first six months; we have 85.

Decision Support System (DSS): Decision support systems couple the intellectual resources of individuals with the capabilities of the computer to improve the quality of decisions. They comprise a computer-based system for management decision makers who deal with semi-structured problems (Gorry & Scott Morton, 1989).

Hard Data: Historical, usually accurate, data, often from transaction processing systems.

Measurement: The process of determining values representing performance, or the results of the process. For example, the measurement process is now started, or the measurement was 35 days.

Metric: A predetermined measure that will be used as the basis for a measurement process. For example, percentage of customer calls answered within one minute.

Soft Data: Judgmental, qualitative data, often from external sources. Any data involving the future.

Measuring Collaboration in Online Communication

Albert L. Ingram
Kent State University, USA

INTRODUCTION

Collaboration has become a key concept in the workplace, in research laboratories, and in educational settings. Companies want members of different departments located far apart to work together. Various government agencies try to establish collaborative relationships with private organizations. Academics and corporate researchers collaborate with far-flung colleagues to produce new knowledge. Students at all levels of our educational system are increasingly being asked to learn collaboratively. In addition, more work is being done online. Businesses communicate over the Internet, and increasing numbers of educational experiences are being delivered at a distance. Virtual high schools, traditional and for-profit distance education institutions, and colleges and universities are all among the current users of the Internet in education.

In all of these situations—educational and non-educational, face-to-face, and online—several questions need to be addressed. First, what is collaboration? The word is sometimes used as if everyone already understands what it means, but we can find a variety of different definitions in the literature. Second, when we form groups to collaborate, how do we know when they have done so? Is it possible to measure the extent to which collaboration has occurred in a given group and setting? Third, what actions and conditions enhance the collaboration that does take place? And finally, does collaboration work? That is, do groups that are more collaborative produce better results or learning than groups that are less collaborative?

This brief article will not attempt to answer all these questions, but it will concentrate on a specific issue: What methods can be used to determine whether, and how much, collaboration has occurred in online groups in various settings? We will explain our preferred definition of collaboration, based on previous research, and then discuss some of the implications of these ideas for online collaboration and for research into that issue.

BACKGROUND

Collaboration can be generally described in a variety of ways, but perhaps a typical definition is "working in a group of two or more to achieve a common goal" (McInnerney & Roberts, 2004, p. 205). Such a general definition, however, does not tell us how reliably to identify when collaboration has taken place or, assuming that there can be degrees of collaboration, how much of it is going on. To make such measurements, we need an operational definition of collaboration. Recently, Hathorn and Ingram (2002) proposed such a definition. They maintained that collaboration consists of at least three key ingredients: interdependence (Johnson, Johnson & Smith, 1998), a product that is achieved through genuine synthesis of information and contributions from all members (Kaye, 1992), and independence from a single leader (Laffey, Tupper, Musser & Wedman, 1998). In education, this would likely be independence from the class instructor. In other settings, it would mean relative independence from supervisors or others who might otherwise control the process too tightly.

Under this definition, collaboration contrasts sharply with what can be called a *cooperative* way of working. In this characterization, cooperation occurs when a group agrees to divide the work among them, with each taking part of the project. The final product, then, is the sum of separate contributions from each member, rather than being a true synthesis as in a good collaborative effort (Hathorn & Ingram, 2002; Ingram & Hathorn, 2004; Dillenbourg, Baker, Blaye & O'Malley, 1996).

Hathorn and Ingram (2002) operationalized their definition by looking at ways of measuring each of the three components of collaboration. Positive interdependence occurs when group members share information and test their ideas on one another. When individuals in a collaborative group work toward their common goal, they often achieve things that would not have been possible individually (Henri, 1992; Kaye, 1992). Synthesis occurs as the group attains new insights as a result of working together (Henri, 1992; Kaye, 1992). Finally, independence requires that the group function on its own without too much centralized direction (Laffey et al., 1998). Otherwise, it is a directed project, not a collaboration among equals.

MEASURING COLLABORATION

In the literature we can find a variety of ways to measure collaboration that have been used by teachers and researchers. In general, these break down into a few major categories: teacher or leader observations, student and participant self-ratings and self-reports, and quantitative analysis of discussion transcripts. Here we look briefly at each of these in turn.

Teacher/Leader Observations

Sometimes an instructor or a team leader can have a very good "feel" for how well a group is collaborating. By scrutinizing the team in action and examining the products that result from the group work, these observers can often tell who is participating fully and contributing to the results, and who is not. Frequently, however, teachers and others may assume that simply putting people into groups automatically results in high-quality collaborative work. This assumption is false: good collaboration requires many factors, and casual observations may not reveal what is really going on. In many cases, online collaborative groups can be easier to observe than face-to-face groups, because all the conversations may be recorded automatically, depending on the software and systems used.

Student and Participant Self-Ratings and Self-Reports

In many instances, members of groups may know how well they are working together. For instance, a frequent complaint of students doing group work for classes is the uneven distribution of the workload. Finding ways to get clear and reliable self-reports from students and other participants in collaborative groups can lead to better understanding of how the groups operate. There is a danger in this, however, because group members may not have a clear understanding of what it means to collaborate effectively. This is especially true if they have never experienced high-quality collaboration themselves. Many groups, especially in education, seem to prefer a "divide-and-conquer" cooperative strategy that appears to them to be collaborative. In fact, it lacks both the interactions and synthesis necessary for good collaboration, because each member of the group works on just part of the whole project. Therefore, in order both to increase the actual collaboration among group members and to improve the reliability and validity of the self-reporting, it is necessary to teach people the characteristics of good collaboration, how to recognize those characteristics, and how to produce them.

Quantitative Measures

Finally, we look at quantitative measurements of whether collaboration has occurred and of its extent. One approach was taken by Wilczenski, Bontrager, Ventrone, and Correra (2001). They measured the behaviors in a group that facilitated and detracted from the collaboration, under the assumption that groups with more facilitative behaviors would be more collaborative. The study showed that groups exhibiting more facilitative behaviors did better on several measures.

Hathorn and Ingram (2002; Ingram & Hathorn, 2004) also took a quantitative approach. Based on the definition of collaboration cited above, they developed measures of its three main components: interdependence, synthesis of contributions, and independence. Specifically, they applied these concepts to asynchronous threaded discussions, although the same ideas could be useful in other contexts as well (e.g., synchronous online chats). They relied on close and detailed content analysis of the discussions themselves (Silverman, 1993). Rourke, Anderson, Garrison, and Archer (2001) noted that a key step is to develop a way of coding the discussions to illuminate the questions one wants to answer. Hathorn and Ingram (2002; Ingram & Hathorn, 2004) developed such a system for the construct of collaboration, noting the inadequacy of many previous schemes for analyzing online collaboration specifically.

In order to use these measures, one needs complete transcripts of the discussions. Online textual discussions are especially useful in this regard since the transcripts are usually kept automatically in both synchronous and asynchronous discussions. Conceivably, the actual medium of communication could be instant messaging/chat, e-mail (including listservs), threaded discussion boards, or other text-based systems, as long as the technology can keep complete logs of the discussions. In Hathorn and Ingram's (2002) system, coding is based on "statements" made in the discussions. Statements are sentences or complete ideas within sentences that represent individual idea units. A single message can contain just one statement or numerous statements on a variety of topics. Indeed, a single sentence can contain multiple statements.

Interdependence is identified using several criteria. First, it requires roughly equal participation among all members. Without that, it is difficult to see how the members can be meaningfully interdependent. Participation is measured primarily by the number of messages and/or statements contributed by each group member. The count of statements is probably a more accurate measure of actual participation than number of messages, sentences, or words would be. It is unlikely in any group that the members participate exactly equally by any measure, so the requirement for good collaboration is that there be at least roughly equal participation. A simple test for this is a chi-square analysis on the participation of the group members. If the test shows significance, it is likely that the members were not participating equally enough for the group to be considered collaborative.

Beyond simple participation, we can measure interdependence by looking at interactions among the group members. In particular, Hathorn and Ingram (2002) focused on direct participation in the substantive discussion (as opposed to off-task comments and other such contributions). The patterns of discussion that indicate actual interaction are threads where different participants refer, explicitly or implicitly, to one another's comments. The minimum length of such a thread that would indicate true interaction is three statements: an initial comment, a response to that comment, and a synthesizing response. The more such threads (and longer ones) appear in a discussion, the more interaction there is and, all other things being equal, the more collaboration is taking place. A useful technique at this point is to diagram the discussion, using the statement as the unit. Each statement appears in a separate box, and lines connecting them show the patterns. It is important not to diagram the messages, because in our experience these do not reflect the substantive patterns that are revealed by the statement-level analysis.

Next, the question of synthesis arises. True collaborations result in products that cannot be identified as resulting from individual efforts. This can be measured in two ways. First, the patterns of interactions noted above include the necessity for synthesizing responses, so that was the first and more detailed measure of synthesis. The other measure focused on the final group product and whether it was written by an individual or the entire group.

Finally, independence from a central authority can be measured by examining both the basic participation of the group members in relation to the instructor and the number of interactions that take place without the instructor's or leader's participation.

Thus, in this quantitative view, the extent of collaboration is revealed by conducting a close content analysis based on the transcripts of the online discussions. Measuring the three elements of collaboration depends on several key pieces of data: overall participation based on the number of statements made by each individual, the types of statements they make (on-task, off-task, and others), and the patterns of interaction, represented especially by the basic unit of collaboration consisting of comment-response-synthesis. The nature of the actual product developed by the group is also important.

Measuring Collaboration in the Real World-An Example

Jim is project manager for a medium-sized information technology firm. His team is scattered around in several locations and needs to work together using a variety of different technologies. Specifically, his supervisor expects him to ensure that they are collaborating well in spite of the distance, the different time zones, and other obstacles. Naturally, the most important concern for Jim and his supervisor is the quality of the work. So far, this has been acceptable, but not outstanding. Jim thinks his team can do better if they learn to collaborate more effectively.

Being a systematic sort of person, Jim decides to approach the problem carefully. So far, simply exhorting the team to do better has not worked, so he wants to find out where they might be failing to collaborate well. To do that, he needs some data about their current efforts. One advantage he has is the fact that much of the work done by the team is online and text based; frequently there is an automatic record kept of the conversations that people have. For example, the team's private collaborative Web site includes a threaded discussion board where everyone can read all the messages and contribute to the conversation. In addition, all documents produced by the team are stored in the online repository in their space. Finally, there is the opportunity for synchronous chats among members while they are working on specific pieces of the project. Jim has decided to record and archive those chats for future use.

Jim's first step in deciding whether his team is collaborating productively is to scan the communications that they have made. At first glance, it looks as if they are all participating and making contributions, although he did not read everything in detail. He decides to try something else before he goes through that lengthy chore.

First, Jim sends an e-mail to everyone in the team asking them to assess the level of collaboration in the team and to e-mail him back by Friday. Very quickly he receives numerous questions about how he defines collaboration, what the request means, and so forth, so he decides to try another approach. This time, Jim does a little research about what is meant by collaboration and sends a more specific e-mail message to the group. Now, instead of asking a global question, he asks people to rate whether the members of the team participate roughly equally and respond well to one another, whether they synthesize each other's contributions, and whether they work independently of the leader. (That last one is tricky for team members, since Jim *is* the leader.)

This time, people are able to answer, and the responses are interesting. In general the team thought they were participating equally and responding to each other. They had mixed views on whether their resulting products were really the result of synthesizing ideas from many sources within the group. They all agreed (tongue in cheek?) that they could get along without Jim for most of their work.

Now Jim has a choice. Is the information he received from the group enough to make decisions about the direction he needs to take? One possibility is that he could arrange to train his team on collaborative techniques, especially using online technologies. Training can be expensive, though, and Jim does not relish taking people away from their jobs for the time it would take. If the team would benefit, then he believes it would be worth it. He is still not sure that they will (or will not) gain enough from such training.

Perhaps the self-reports of his team members are good enough for Jim to make the decision. If the stakes are high enough, however, Jim might want to look further. In that case a more detailed content analysis might pay off. His team members might overestimate the extent to which they respond to each other's message. Often an online discussion proceeds more as a series of monologues than as a true collaborative conversation. Sometimes what seems to be a synthesis of ideas is really one person dominating the group. Or perhaps the team is not as independent of Jim as they think. If Jim can pinpoint more specific problems with their collaborative skills, then it is more likely that he can arrange to have them trained in the relevant skills. In turn, that could lead to more tangible benefits from the collaborative team structure that he has developed than he is seeing now. He will not really know until he finds a way to measure the actual collaboration taking place.

FUTURE TRENDS

Online collaboration is likely to become even more important in the future for education, research, business, and other fields. In many of these areas, it will be important to ensure that groups working online are actually collaborating. Often in job settings working groups are composed of people with different knowledge and skill sets. If they do not truly collaborate, then their products (reports, designs, and other substantive materials) may be incomplete and substandard. In educational settings group work may produce high-quality learning outcomes primarily in groups that collaborate, rather than those using a less effective "divide-and-conquer" strategy in which group members learn only the pieces that they work on.

One indication of the trend toward more online collaboration is the increasing number of software packages aimed at improving the efficiency and effectiveness of online collaboration. For example, Groove *(www.groove.net)* is inexpensive peer-to-peer software designed specifically for small groups and containing a variety of tools beyond basic synchronous and asynchronous discussion capabilities. Convea *(www.convea.com)* is similarly inexpensive Web-based software designed for much the same purpose. Such packages are increasingly under scrutiny as collaborative tools (Ingram & Parker, 2003; Ingram, Pretti-Frontczak & Parker, 2003).

These trends mean that we need more and better ways to assess the presence, amount, and nature of online collaborative processes. Such measures will be useful in research, in education, and in practical application. Which measures are used depends on the questions being asked and on the purposes to which the assessment will be put. Some measures are extremely subjective and not firmly grounded in theory and previous research. Others may rely too heavily on the untutored impressions of participants. On the other hand, one drawback of a detailed quantitative approach such as the one we outlined here is the time it takes to complete. An approach being pursued by the present author and his colleagues is using the detailed research approach to develop better surveys of team members that will allow the identification of good and poor collaboration more quickly and efficiently. Such tools would have wide applicability.

CONCLUSION

This article has discussed several ways that collaboration in small online working groups might be measured. The need for such measurements seems clear: high-quality work in a variety of fields demands good collaboration. In turn, managers, researchers, educators, and others need to be able to identify the presences, amount, and nature of the collaboration that does or does not take place in different circumstances.

All of the measurement methods discussed here have strengths and drawbacks. Some are quick and easy to use but may give misleading results. Others are precise and based on some of the best research available, but demand significant effort and time to implement. One direction for future research and practice may be to develop measurement tools that are both accurate and quick.

REFERENCES

Dillenbourg, P., Baker, M., Blaye, A. & O'Malley, C. (1996). The evolution of research on collaborative learning. In P. Reinman & H. Spada (Eds.), *Learning in humans and machines: Towards an interdisciplinary learning science* (pp. 189-211). New York: Pergamon.

Hathorn, L.G. & Ingram, A.L. (2002). Cooperation and collaboration using computer-mediated communication. *Journal of Educational Computing Research, 26*(3), 325-247.

Henri, F. (1992). Computer conferencing and content analysis. In A.R. Kaye (Ed.), *Collaborative learning through computer conferencing* (pp. 117-136). Berlin: Springer-Verlag.

Ingram, A.L. & Hathorn, L.G. (2004). Methods for analyzing collaboration in online communications. In T.S. Roberts (Ed.), *Online collaborative learning : Theory and practice* (Chapter 10, pp. 215-241). Hershey, PA: Idea Group, Inc.

Ingram, A.L. & Parker, R.E. (2003). Collaboration and technology for teaching and learning. *Proceedings of the Annual Conference of the Ohio Learning Network*, Columbus, OH. Retrieved from *www.oln.org/conferences/papers/Collaboration_and_Technology.pdf*.

Ingram, A.L., Pretti-Frontczak, K. & Parker, R. (2003). Comparisons of student and faculty use of online collaboration tools. *Proceedings of the Teaching Online in Higher Education Online Conference.* Retrieved from *www.ipfw.edu/as/2003tohe/*.

Johnson, D.W., Johnson, R.T. & Smith, K.A. (1998). Cooperative learning returns to college. *Change*, *30*(4), 26-35.

Kaye, A. (1992). Learning together apart. In A.R. Kaye (Ed.), *Collaborative learning through computer conferencing* (pp. 117-136). Berlin: Springer-Verlag.

Laffey, J., Tupper, T., Musser, D. & Wedman, J. (1998). A computer-mediated support system for project-based learning. *Educational Technology Research and Development*, *46*(1), 73-86.

McInnerney, J.M. & Roberts, T.S. (2004). Collaborative or cooperative learning? In T.S. Roberts (Ed.), *Online collaborative learning : Theory and practice* (Chapter 9, pp. 203-214). Hershey, PA: Idea Group, Inc.

Rourke, L., Anderson, T., Garrison, D.R. & Archer, W. (2001). Methodological issues in the content analysis of computer conference transcripts. *International Journal of Artificial Intelligence in Education, 12*(1), 8-22. Retrieved from *www.atl.ualberta.ca/cmc/publications.html*.

Silverman, D. (1993). *Interpreting qualitative data.* Thousand Oaks, CA: Sage Publications.

Wilczenski, F.L., Bontrager, T., Ventrone, P. & Correra, M. (2001). Observing collaborative problem-solving processes and outcomes. *Psychology in the Schools*, *38*(3), 269-281.

KEY TERMS

Asynchronous Discussion: Online discussions that occur independent of time and space. Participants do not have to be online simultaneously, and can read and contribute to the conversation on their own schedules.

Collaboration: Occurs when small groups of people work together toward a common goal in ways that produce new products and knowledge that are unlikely to be developed by individuals. Three essential elements of collaboration are interdependence, synthesis, and independence.

Cooperation: Cooperative groups work together on group projects in ways that do not necessarily result in high-quality interaction, and new products and knowledge. A typical cooperative strategy is to divide up the work among the members and stitch the various contributions together at the end of the project.

Independence: The independence of a group from a central authority, such as an instructor or manager, ensures that the group can truly collaborate among themselves and produce results that are unique and new.

Interaction: Interaction among members of a group is necessary to produce collaboration. It can be measured by examining the give-and-take nature of the discussion threads.

Interdependence: Interdependence among members of a small group is a necessary element of collaboration. It means that group members could not produce the results they did without one another.

Participation: The most basic requirement of collaboration; it may be measured by the number of postings made and read or by the number of statements contributed to a discussion.

Synchronous Discussion: Occur when all participants are online and actively involved in the discussion at the same time.

Synthesis: Occurs when the final product of a group effort includes information and other elements from all members in such a way that individual contributions are difficult or impossible to identify.

Media and Personal Involvement in the Perceptions of Data Quality

Andrew Borchers
Kettering University, USA

INTRODUCTION

This article introduces the concepts of data quality as described in the literature and discusses research results on how individual perceptions of data quality are influenced by media (World Wide Web versus print) and personal involvement with the topic. A search of literature on "data quality" and "media creditability" reveals that researchers in both the information systems and journalism fields have explored this topic. While these threads have developed separately, these streams of research approach similar issues of how people view the quality of information they receive from different sources.

This topic is important in that the Internet revolution has fundamentally changed how people share information. But with such access comes a challenge as stated by Gilster (as cited in Flanagin & Metzger, 2000): "When is a globe spanning information network dangerous? When people make too many assumptions about what they find on it. For while the Internet offers myriad opportunities for learning, an unconsidered view of its contents can be misleading and deceptive."

The importance of this topic is underscored by recent research and seminars on the topic of data quality on the Web. A 2003 conference in Wadern, Germany, for example, brought researchers together to discuss topics such as "criteria and measurement for quality of Web data, representation and exchange of quality information and usage and maintenance of data quality in Web querying and data integration". The conference included working groups that focused on quality assessment, trust, data integration, and metadata (*http://www.dagstuhl.de/03362/*).

BACKGROUND

The importance of data quality has been echoed among information systems and journalism practitioners for many years. Research by Redman (1998) summarizes the practical implications of poor data quality. He points out the consequences of poor data quality in areas such as decision-making, organizational trust, strategic planning and implementation, and customer satisfaction. Redman conducted detailed studies and found increased cost of 8-12% due to poor data quality. Service organizations can find increased expenses of 40-60% (Redman, 1998). Strong, Lee, Yang, and Wang (1997) support the seriousness of this issue in their study of 42 data quality projects in three organizations. Research by other authors note data quality issues in a number of settings including accounting (Xu, 2000; Kaplan, Krishnan, Padman, & Peters, 1998), airlines, health care (Strong et al., 1997), criminal justice (Laudon, 1986) and data warehousing (Ballou, 1999).

As for a formal definition of data quality, Umar, Karabatis, Ness, Horowitz, and Elmagardmid (1999) quote Redman (1998):

A product, service, or datum X is of higher quality than product, service, or datum Y if X meets customer needs better than Y.

Umar et al. (1999) go on to point out that this definition has been generally accepted and is consistent with author's work. The definition is somewhat incomplete, however, as it does not delve into the various dimensions of data quality.

A number of authors in the information systems field have gone further than Redman and written conceptual articles on "data quality" (Wand & Wang, 1996; Wang, Reddy, & Kon, 1995; Wang & Strong, 1996; Strong et al., 1997). This work suggests that data quality is a multidimensional concept (Wand et al., 1996) that can be viewed from a number of different perspectives. A panel discussion in 2000 (Lee, Bowen, Funk, Jarke, Madnick and Wand) found five different perspectives to discuss data quality. These included an ontological perspective (specification of a conceptualization) that include different views of reality based on actual observation versus computer influenced observations, architectural perspective (a view that focuses on system infrastructure and its influence on data quality), context mediation perspective (focusing on communication across space and time), time-based e-commerce perspective (focusing on the real-time nature of e-commerce) and an information product perspective (focused on data as a product of an organization).

In talking about "data quality", a key beginning is to determine from the literature just what is meant by the term. In a definitive work on the topic, Wang et al. (1996) provides a conceptual framework for data quality. In a way consistent with Redman's customer perspective, they start by defining "high-quality data as data that is fit for use by data consumers". Using a two-stage survey and sorting process, Wang develops a hierarchical framework for data quality that includes four major areas: intrinsic, contextual, representational, and accessibility.

Intrinsic data quality refers to the concept that "data have quality in their own right" (Wang et al., 1996). Intrinsic dimensions include accuracy, objectivity, believability, and reputation. Contextual data quality is based on the idea that data does not exist in a vacuum – it is driven by context. Contextual dimensions include relevancy, timeliness and appropriate amount of data. Representational data quality relates to the "format of the data (concise and consistent representation) and meaning of data (interpretability and ease of understanding)". Accessibility refers to the ease with which one can get to data (Wang et al., 1996).

Beyond the information systems literature, there is a second relevant body of literature that comes from the journalism field. Their focus is on perceptions of Internet credibility (Flanagin & Metzger, 2000; Johnson & Kaye, 1998). The major thrust of this literature is in comparing the Internet to traditional sources with respect to credibility. Note that when referring to "credibility", these authors say "the most consistent dimension of media credibility is believability, but accuracy, trustworthiness, bias and completeness of information are other dimensions commonly used by researchers." (Flanagin et al., 2000, p. 521). Hence, there is a rough correspondence of thinking about "credibility" in the journalism literature to the concept of "intrinsic" and "contextual" data quality in the information systems literature.

COMPARISON OF RESEARCH FINDINGS

Research on data quality and media differences has been undertaken by a number of scholars over the years. In this section, the author will focus on work by Klein (1999, 2001), Flanagin et al. (2000) and Borchers (2003). All of these authors have examined data quality in a similar way, focusing on perceived differences based on media (print versus Internet). In addition, Flanagin et al. (2000) examine whether Internet users verify what they find. Borchers (2003) extends the discussion by examining the effect of personal involvement in the topic.

Klein (1999, 2001) has studied perceptions of data quality by surveying a sample of approximately 70 graduate business students conducting class projects. In one early study, Klein (1999) found Web-based material to be more timely, but less believable and of lower reputation, accuracy and objectivity than printed material. In a more formal result, Klein (2001) found traditional text sources to be perceived as more accurate, objective and to have higher reputation and representational consistency. Internet sources were found to be stronger in timeliness and appropriate amount.

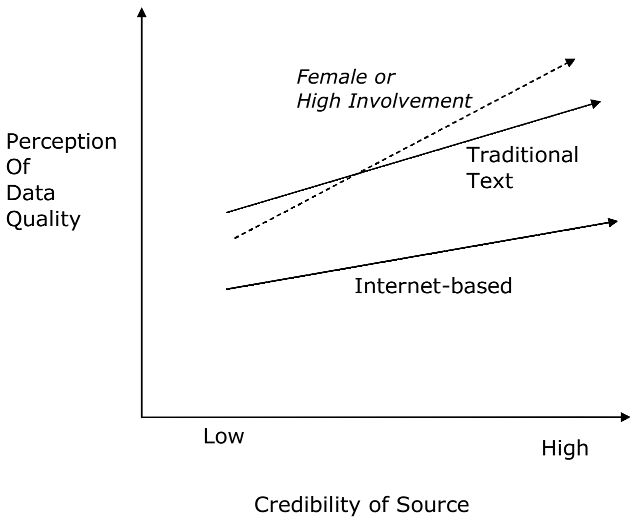
Flanagin et al.'s work (2000) focuses in three areas. First, they look at the perceived credibility of television, newspapers, radio and magazines compared to the Internet. The major finding, unlike Klein, is that there is little difference in credibility between media. Second, Flanagin et al. look at the extent to which Internet users verify what they receive. Here, they find that few Web users verify the information they receive. Those with limited Internet experience verify less than those with more experience. Third, and most important to this discussion, Flanagin et al. look at whether perceived credibility varies depending on the type of information being sought. Flanagin et al. cite Gunther in suggesting that "greater involvement with the message results in, first, a wider latitude of rejection."

Borchers (2003) considered the literature cited previously and examined a number of interesting questions. In keeping with Klein and Flanagin et al., he examined how people perceive Web-based material compared to printed material considering dimensions such as "timely", "believable", "reputation", "accuracy", or "objectivity". Second, Borchers studied whether individuals with personal involvement in a topic (e.g., cancer) are better discriminators of data quality than those who are not involved with a topic. Finally, given their role as health care acquirers (Bates & Gawande, 2000; Looker & Stichler, 2001), Borchers explores whether women are better discriminators of data quality than men on health related topics such as cancer.

Figure 1 demonstrates what Borchers hope to find. H0, his initial hypothesis, is that the perception of low creditable sources is significantly less than high creditable sources. Hence, the two lines for Internet-based and print-based text should have a positive slope. H1 suggests a significant gap between the lines for Internet-based sources and text-based sources on the timeliness, believability, reputation, accuracy, and objectivity dimensions. This assertion was based on prior literature by Klein (1999). H2 suggests that the slope of the lines should vary based on one's personal involvement in cancer. This is to say, that persons with high personal involvement in cancer should be better discriminators of data quality. Finally, H3 suggests that women are better able to differentiate credible from non-credible sources. Hence, the slope of the lines should vary based on gender.

Borchers studied 127 subjects on their perception of information on cancer based on exposure to Internet and print media. Subjects were drawn from mid-career students in MBA and MSIS classes at a Midwestern univer-

Figure 1. Research design (Borchers, 2003)



sity. His sample was strongly multi-cultural with significant U.S., Indian and Chinese representation. Subjects were randomly assigned to one of four groups.

These four groups were shown cancer information based on two sources of information presented in two different formats. One source was a Website of a highly credible national cancer organization. The second source of cancer information was a Website of low creditability, a site that touted alternative medical treatments. The third and fourth sources were identical to the first two with the exception that they were presented in printed form by way of a color document. Subjects were then asked about their perceptions of the data that they have viewed using Wang's intrinsic data quality dimensions (accuracy, objectivity, believability, and reputation) as well as contextual dimensions (timeliness, relevancy and appropriate amount of information) and ease of use. Further, subjects were asked about their personal and family experience with cancer as well as demographic questions (gender, age and country of birth).

After the data was collected, Borchers tested the dimensions using Cronbach alpha to be sure they were reliable. Each dimension had a value of .8 or higher. Second, Borchers generated a second set of statistics using a univariate ANOVA procedure to test each of the research hypotheses. H0 was tested for all eight measured data quality dimensions using the source reputation (high or low) and media (print or WWW) as fixed factors. In testing H2 and H3 cancer involvement or gender were added as random factors. The hypotheses were tested by looking at the product term for source reputation and cancer involvement (H2) or Gender (H3). Table 1 below summarizes Borchers' findings:

With respect to H0, Borchers found that his experimental design worked reasonably well. Subjects could easily discriminate between the low and high credible sources. In comparing Internet to print sources (H1), Borchers observed no difference. In examining personal involvement (H2), Borchers found that believability and reputation were impacted by personal involvement, but only for a subset of respondents. Borchers did not support the gender hypothesis (H3).

FUTURE TRENDS

Future work should be done to extend the work of Klein, Borchers and Flanagin et al. First, methodological issues need to be addressed. Instrumentation in particular is a major concern. How can researchers more accurately measure people's perception of data quality and personal involvement in a topic? Studies should also be conducted among participants other than college students. Experimental designs such as Borchers' can be extended to more precisely test hypotheses. Second, existing work focuses on print and the Internet (especially the WWW) as media. Further research could follow Flanagin et al.'s work in other media, such as TV and radio, or emerging media. Third, tests of personal involvement could be extended to topics other than cancer. The concept of "personal involvement" needs to be developed, perhaps turning to the marketing literature for a base. Finally, as the general population becomes more computer literate and Internet savvy, researchers may find that perceptions of data quality between different media converge (or perhaps diverge).

It should be noted that there are differences in research approach between these authors' work and that each work has methodological weaknesses that can be addressed in future research. For example, in Flanagin et al.'s and Klein's work subjects were asked to complete a one shot survey that asked about their perceptions in general of credibility of Internet and print sources. In Borchers' work subjects were randomly placed in groups that saw exactly the same material in both Internet and text formats. In his study, however, cultural differences among subjects confounded the analysis. All three researchers use college students as a primary source of respondents, an obvious weakness. Each of these authors also uses surveys as a data collection technique, a potential weakness.

Having noted these limitations and future areas of work, this line of research is important for several reasons. First, the Internet has become a de facto standard source of information for younger generations. Their perceptions of data quality, particularly compared to print, are a key factor in understanding how people will interpret what they see. Second, the question of personal involvement raises important concerns – do people become more

Table 1. Hypothesis testing results (Borchers, 2003)

Hypothesis	Dimension	F ratio	Significance
H0 – Initial difference due to reputation	Believable, accuracy, reputation, objectivity, and appropriate amount	10.526 to 24.489	.000 to .002
H0 – Initial difference due to reputation	Timeliness, relevance, ease of use	< 1.4	> .35
H1 – WWW compared to print	Timeliness Believable Reputation Accuracy Objectivity Appropriate amount Relevance Ease of use	2.587 1.036 .340 .483 1.132 .617 .030 .620	.110 .311 .561 .489 .290 .484 .865 .484
H2 – Personal involvement with cancer	Timeliness Believable Reputation Accuracy Objectivity Appropriate amount Relevance Ease of use	With all respondents F ratio < 4 US only respondents had significant interaction on believability and reputation	With all respondents F > .05 US only respondents were significant on believability and reputation
H3 – Gender	Believable Accuracy Reputation Objectivity Appropriate amount Relevance Ease of use	All < 4	All > .05

discriminating in evaluating data quality on topics that have significant influence to their lives? Finally, verification (as described by Flanagin et al.) is yet another interesting topic. To what extent do people verify what they read on the Internet or in print? Future research should address these and other areas.

CONCLUSION

The work of these three researches (Klein, 1999, 2001; Flanagin et al., 2000; Borchers 2003) seeks to understand how perceived data quality is affected by different media. Flanagin et al.'s work extends the discussion to verification of information. Borchers extends the case to include personal involvement.

Do people perceive data quality differently depending on the media that data is presented in? Flanagin et al. (2000) and Borchers (2003) suggest that media is not a significant factor. This comes in contrast to Klein's work (1999, 2001) which suggested a difference in perceived data quality on five dimensions: accuracy, objectivity, reputation, timeliness, and appropriate amount.

Do Internet users verify what they see on the Internet? Flanagin et al.'s work suggests that relatively few do so, and among inexperienced users even fewer verify what they find. This finding should be particularly distressing to academics and practitioners concerned with delivering high quality data. Will students be able to discern data quality in the myriad of Web resources available to them?

Do people become more discriminating of data quality for topics that they are personally involved in? Borchers' study provides a first look, at least for cancer information, and finds only limited support for this notion. Finally, does gender play a role in one's ability to discriminate between reputable and non-reputable sources of cancer information? Borchers' work would suggest that this is not so.

Data quality is an important topic that deserves continued research focus. The Internet revolution has fundamentally changed how people share information. How people perceive the quality of the information they view is an essential research topic for the information systems field.

REFERENCES

Ballou, D.P., & Tayi, G.K. (1999 January). Enhancing data quality in data warehouse environments. *Source Communications of the ACM*, *42*(1), 73-78.

Bates, D.W., & Gawande, A.A. (2000). The impact of the Internet on quality measurement. *Health Affairs, 19*(6), 104-114.

Borchers, A.S. (2003). Intrinsic and contextual data quality: The effect of media and personal involvement. In *ERP & data warehousing in organizations: Issues and challenges*. Hershey, PA: IRM Press.

Bowen, P.L., Funk, J.D., Johnson, S.C., Jarke, M., Madnick, S.E., Wand, Y., & Lee, Y.W. (2000). Data quality in internet time, space, and communities (panel session). *International Conference on Information Systems Proceedings of the twenty first international conference on Information systems table of contents*, Brisbane, Queensland, Australia, 713-716.

Flanagin, A.J., & Metzger, M.J. (2000). Perceptions of Internet information credibility. *Journalism and Mass Communications Quarterly*, *77*(3), 515-540.

Johnson, T.J., & Kaye, B.K. (1998). Cruising is believing? Comparing Internet and traditional sources on media credibility measures. *Journalism and Mass Communications Quarterly*, *75*(2), 325-340.

Kaplan, D., Krishnan, R., Padman, R., & Peters, J. (1998). Assessing data quality in accounting information systems. *Communications of the ACM*, *41*(2).

Klein, B.D. (1999, October). Information quality and the WWW. *Applied Business in Technology Conference.* Rochester, MI: Oakland University.

Klein, B.D. (2001, Summer). User perceptions of data quality: Internet and traditional text sources. *Journal of Computer Information Systems*, *41*(4).

Laudon, K.C. (1986, January). Data quality and due process in large inter-organizational record systems. *Communications of the ACM*, *29*(1).

Looker, P.A., & Stichler, J.F. (2001). Getting to know the women's health care segment. *Marketing Health Services*, *21*(3), 33-34.

Redman, T. (1998). The impact of poor data quality on the typical enterprise. *Communications of the ACM*, *41*(2), 79-83.

Strong, D., Lee, M., Yang, W., & Wang, R.Y. (1997). Data quality in context. *Communications of the ACM*, *40*(5), 103-110.

Umar, A., Karabatis, G., Ness, L. , Horowitz, B., & Elmagardmid, A. (1999). Enterprise data quality: A pragmatic approach. *Information Systems Frontiers, 1*(3), 279.

Wand, Y., & Wang, R.Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, *39*(11), 86-95.

Wang, R., Reddy, M.P., & Kon, H.B. (1995). Towards quality data: An attribute-based approach. *Decision Support Systems*, *13*(3, 4), 349-372.

Wang, R.Y., & Strong, D.M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, *12*(4), 5-33.

Xu, H. (2000, December). Managing accounting information quality. *Proceedings of the Twenty First International Conference on Information Systems.*

KEY TERMS

Accessibility Data Quality: An aspect of data quality that refers to the ease with which one can get to data.

Contextual Data Quality: A concept that data does not exist in a vacuum, it is driven by the circumstance in which data is used. Contextual dimensions include relevancy, timeliness and appropriate amount of data.

Data Quality: A multi-faceted concept in information systems research that focuses on the fitness for use of data by consumers. Data quality can be viewed in four categories: intrinsic (accuracy, objectivity, believability and reputation), contextual (relevancy, timeliness and appropriate amount of data), representational (format of the data) and accessibility (ease of access).

Internet Credibility: A multi-faceted concept in journalism research that consists of believability, accuracy, trustworthiness and bias.

Intrinsic Data Quality: A concept that "data have quality in their own right" (Wang & Strong, 1996) including accuracy, objectivity, believability and reputation dimensions.

Representational Data Quality: A concept that data quality is related to the "format of the data (concise and consistent representation) and meaning of data (interpretability and ease of understanding)" (Wang & Strong, 1996).

MESH Object-Oriented Hypermedia Framework

Wilfried Lemahieu
Katholieke Universiteit Leuven, Belgium

INTRODUCTION

Hypermedia systems represent data as a network of nodes, interconnected by links. The information embodied within the nodes can be accessed by means of navigation along the links, whereby a user's current position in the information space determines which information can be accessed in the next navigation step. This property of navigational data access raises hypermedia systems as utterly suitable to support user-driven exploration and learning. The user autonomously determines the way in which he or she will delve into the information, instead of being confined to the rigid "linear" structure of, for example, pages in a book.

Although the World Wide Web contributed tremendously to the popularity of the hypermedia paradigm, it also amply illustrated its two primary weaknesses: the problem of user disorientation and the difficulty of maintaining the hyperbase. Indeed, whereas non-linear navigation is a very powerful concept in allowing the end user to choose his or her own strategy in discovering an information space, the resulting navigational freedom may easily lead to cognitive overhead and disorientation. The latter is called the "lost in hyperspace" phenomenon. The user becomes unable to assess his or her current position and sort out his/her navigational options. Equally stringent is the maintenance problem. Due to the absence of workable abstractions, many hypermedia systems implement links as direct references to the target node's physical location (e.g., the URL in a Web environment). To make things worse, these references are embedded within the content of a link's source node. As a result, hyperbase maintenance becomes a synonym for manually editing the nodes' content. Moreover, the inability to enforce integrity constraints and submit the network structure to consistency and completeness checks results in a hyperbase with plenty of dangling links. Needless to say that the consequences of inferior maintenance will also frustrate the end user and effect additional orientation problems.

BACKGROUND: A STRUCTURED APPROACH TO HYPERMEDIA AUTHORING AND NAVIGATION

This article introduces the *MESH* hypermedia framework, an acronym for ***M**aintainable*, ***E**nd user friendly*, ***S**tructured*, ***H**ypermedia*. So as to improve maintainability and facilitate end-user orientation, a structured approach is taken towards both authoring and navigation. MESH's fundaments are a solid underlying data model and a context-based navigation paradigm.

The benefits of data modeling abstractions to both orientation and maintainability were already acknowledged in Halasz (1988). They yield richer domain knowledge specifications and more expressive querying. Typed nodes and links offer increased consistency in both node layout and link structure (Knopik & Bapat, 1994). Higher-order information units and perceivable equivalencies (both on a conceptual and a layout level) greatly improve orientation (Thüring et al., 1995). Moreover, the separation of node content from data structure, along with the introduction of a dedicated link storage facility, allows for improved maintainability (Garzotto et al., 1993). Semantic constraints and consistency can be enforced (Ashman et al., 1997), tool-based development is facilitated and reuse is encouraged (Nanard & Nanard, 1995).

Consequently, hypermedia design is to be based on a firm conceptual data model. The pioneering conceptual hypermedia modeling approaches such as HDM (Garzotto et al., 1993) and RMM (Isakowitz et al., 1998) were based on the entity-relationship paradigm. Object-oriented techniques were mainly applied in hypermedia engines to model functional behavior of an application's components, for example Microcosm (Davis et al., 1992) and Hyperform (Wiil & Leggett, 1997). Along with OOHDM (Rossi et al., 2000) and WebML (Ceri et al., 2000), MESH is the first approach where modeling of the application domain is fully accomplished through the object-oriented paradigm.

MESH couples general-purpose object-oriented modeling abstractions to proprietary hypermedia concepts to provide for a formal hypermedia data model. Moreover, its context-based navigation paradigm builds upon the data model to reconcile navigational freedom with nested, dynamically created guided tours, which are to provide a disoriented end user with sequential paths as guidance. This article presents the two elementary concepts behind MESH, its data model and the context-based navigation paradigm, and proposes an implementation framework. As this is only an overview, the discussion is kept very brief. A more elaborate description of the data model and navigation paradigm can be found in Lemahieu (2003) and Lemahieu (2002) respectively.

THE MESH HYPERMEDIA FRAMEWORK

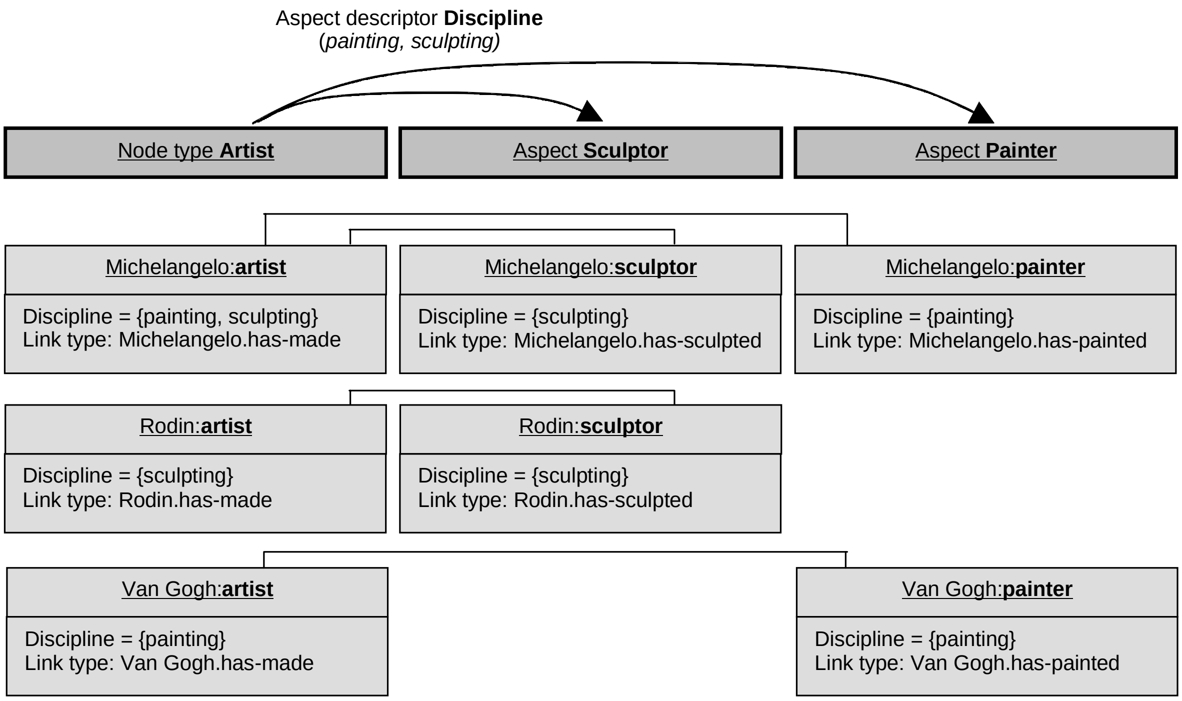
MESH's data model is looked upon as a hierarchy of node types. Each node is defined as an instance of exactly one "most specific type," which remains unchanged during the whole of the node's lifetime. All node types are assorted in an inheritance tree, each node type being compliant with the definition of its parent. Visual properties for a certain node type are abstracted in layout templates. These can be inherited and refined throughout the node type hierarchy. In this way, node instances representing similar real-world objects bear a similar, recognizable layout. In addition to the primary node classification criterion provided by the static node typing hierarchy, the aspect construct offers a means to attribute specific combinations of link and layout properties to certain sets of nodes. At run-time, a node's features can be altered by adding or removing aspects. As such, aspects provide secondary node classification criteria, which are independent of the node typing hierarchy and which, unlike a node's type, are allowed to be non-disjoint and variable over time.

So as to characterize node interaction, a node type is equipped with a set of link types. These classify the links according to the semantic interpretation of the relations they embody. Link type properties include domain, minimum cardinality, maximum cardinality and inverse link type. These can be seen as constraints imposed upon the respective links, in order to be feasible instances of their type. If a link type is inherited by a child node type, these constraints can be overridden, but may never be weaker than the ones imposed at the parent level. A final notable feature of MESH's data model is the concept of link subtyping. In common data modeling literature, subtyping is invariably applied to objects, never to object interrelations. However, in MESH, link types are full-fledged abstract types, which can be equally subject to subtyping. A child link type's instances are a subset of its parent's instances. It models a relation that is more specific than the one modeled by the parent, possibly reflected in more stringent constraints. Link types are deemed extremely important, as they not only enforce semantic constraints but also provide the key to context-based navigation.

An example is show in Figure 1: the node type artist has two aspects: sculptor and painter, denoted by a special attribute (a so-called aspect descriptor) discipline. The node Michelangelo is an instance of artist. Moreover, it has both the values "sculpting" and "painting" for its discipline aspect descriptor; hence Michelangelo has two aspects: a painter aspect and a sculptor aspect. Each such aspect contains discipline-specific properties of the artist Michelangelo. The artist Rodin only has the value "sculpting" for discipline; hence he only has a sculptor aspect. Van Gogh only has a painter aspect. An artist is linked to his artwork in general through a has-made link type. A sculptor is linked to his sculptures through a has-sculpted link type. A painter is linked to his paintings through a has-painted link type. This is an example of link subtyping: both has-sculpted and has-painted are subtypes of has-made. Rodin and Van Gogh only have has-sculpted and respectively has-painted links. Michelangelo has both: the paintings made by Michelangelo, denoted as Michelangelo.has-painted, will be a subset of all of Michelangelo's artwork, denoted as Michelangelo.has-made. We can say that Michelangelo.has made is the union of Michelangelo.has-sculpted and Michelangelo.has-painted. Please refer to Lemahieu (2003) for more details.

The navigation paradigm as presented in MESH combines set-based navigation principles (Kappe, 1999) with the advantages of typed links and a structured data model. The typed links allow for a generalization of the guided tour construct. The latter is defined as a linear structure that eases the burden placed on the reader, hence reducing cognitive overhead and, consequently, disorientation. As opposed to conventional static guided tour implementations, MESH allows for guided tours to be generated at run-time, depending on the context of a user's navigation. Such context is derived from the type of previously followed links. The context-based navigation paradigm defines navigation in two orthogonal dimensions: on the one hand, navigation within the current tour yields linear access to nodes related to the user's current focus of interest. On the other hand, navigation orthogonal to a current tour, changing the context of the user's information requirements, offers the navigational freedom that is the trademark of hypertext systems. In addition, the typed links sustain the generation of very compact overviews and maps of complete navigation sessions. An example of such overview is provided in Figure 2. The latter represents multiple guided tours, nested in one another. The "outer" tour Paris.Museums consists of all museums in Paris. The "current node" in that tour, that is, the museum currently visited, is the

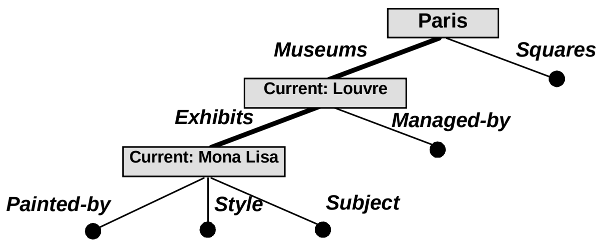
Figure 1. Node types, aspects and link types in MESH



Louvre. At this level, pressing the "next" button would lead to the next node in the tour, that is, another museum in Paris, becoming current. However, within the Paris.Museums tour, there is also a nested tour ongoing: Louvre.Exhibits, showing all artwork in the Louvre. Within that tour, the current node is *Mona Lisa*. For all artwork in the Louvre.Exhibits tour, the following information is visited consecutively: information about the artist (the painted-by link type), about its style (the style link type) and about the subject (the subject link type). If the "next" button is pressed at this level, another artwork in the Louvre is accessed. Upon completion of the Louvre.Exhibits tour (i.e., all artwork in the Louvre is visited), information about the museum's management (the Managed-by link type) becomes accessible. Thereafter, the next node in the Paris.Museums tour can be accessed. Upon completion of the Paris.Museums tour, a Paris.Squares tour will be started. This information can also be bookmarked; that is, bookmarks do not just refer to a single node but to the complete navigational situation, which can be resumed at a later date. More details are provided in Lemahieu (2002).

Although in theory MESH's data model and navigation paradigm can be implemented above any physical hypermedia architecture, this article presents one possible implementation framework that fully supports all of MESH's facets (cf. Figure 3). In broad outlines, this implementation architecture corresponds to the OHP framework proposed by the Open Hypermedia Community (Millard et al., 1998). The information content and navigation structure of the nodes are separated and stored independently. The resulting system consists of three types of components: the nodes, the linkbase/repository and the hyperbase engine. The nodes are considered as a potentially heterogeneous collection of entities, ranging from flat files (e.g., HTML or XML fragments) to objects in an object-oriented database, each containing one or more embedded multimedia objects. Optionally, a node can be endowed with the intelligence to tune its visualization to the context in which it is accessed; hence the so-called context-sensitive visualization principle. Links as well as metadata about node types and link types are captured in a searchable linkbase/repository, implemented in a relational database environment. Only here, references to physical node addresses are stored; these are never to be embedded in a node's body. According to the object-oriented information hiding principle, a node's internal content is shielded from the outside world by the indirection of link types playing the role of a node's public interface. The interaction mechanism can be compared to object-oriented method calls and return values, mediated by the hyperbase engine. This hyperbase engine is con-

Figure 2. Example of guided tours in MESH



ceived as a server-side application that accepts link selections from the current node, invokes a method on the correct destination node, keeps track of session information and provides facilities for generating maps and overviews. Since all relevant linkage and meta information is stored in the relational DBMS, the hyperbase engine can access this information by means of simple, pre-defined database queries, that is, without the need for searching through node content.

FUTURE TRENDS

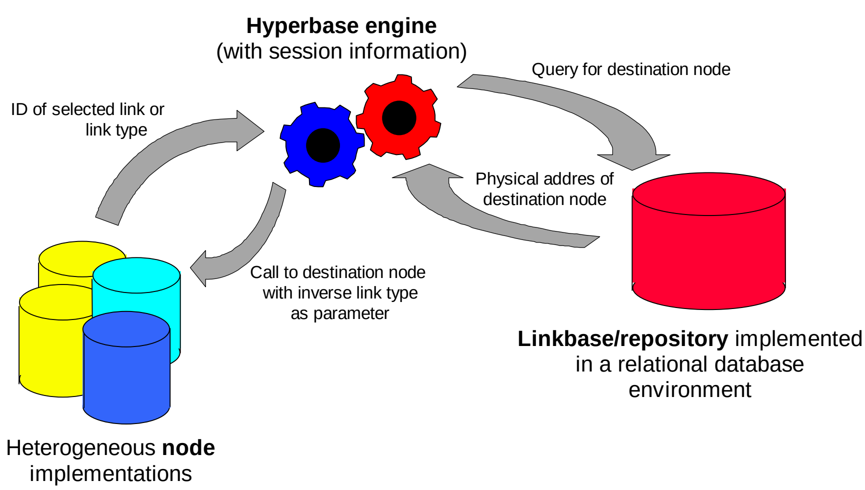
Future research is aimed at translating the platform-independent implementation framework into a Web based system, the hyperbase engine being conceived as a set of Web services. Currently, a standalone, read-only prototype already exists as a Java servlet. A future version is to offer support for authoring and distributed storage of node content and links. Another line of research is the application of Web usage mining techniques to support the personalized "advertising" of relevant guided tours.

CONCLUSION

MESH's structured approach to data modeling enables nodes and links to be treated as real objects, equipped with a unique OID that is independent of physical node location. The links, stored in a separate (and searchable) link facility, can map these OIDs to physical locations, which can be manipulated without the need for nodes to be re-edited upon reallocations. This guarantees a clean separation between authoring content and physical hyperbase maintenance. Authoring is greatly facilitated by object-oriented features such as inheritance and overriding, class properties and layout templates that allow for high-level specification and lower-level refinement of node properties. Authoring is also facilitated by suggestion of mandatory links and valid link destinations, based on metadata present in the linkbase/repository. Moreover, the constraints enforced by the data model permit the automated integrity checking that also exists in database systems. Dangling links and inconsistent link attributions can already be detected during the design phase. As to navigation, apart from the obvious benefit of a well-maintained hyperbase, typed nodes and links reduce cognitive overhead and permit a better comprehension of the semantic relations between information objects. The context-based navigation mechanism further reduces cognitive overhead by dynamically generating guided tours. The linear guided tours assist the user as well as improve his or her sense of position and ability to ascertain his or her navigational options. This effect is even increased by consistent node layout and context sensitive node visualization.

Other hypermedia approaches such as *RMM*, *HDM*, *OOHDM* and *WebML* are also based on conceptual modeling abstractions, either through entity-relationship or object-oriented techniques. RMM, HDM and OOHDM

Figure 3. Implementation framework for MESH



restrict themselves to a "standard" entity-relationship or object-oriented approach. WebML provides a recent, model-driven approach to Web site development, borrowing from entity-relationship modeling and UML. Also worth noting in the context of UML are W2000 (Baresi et al., 2001) and UHDM (Baumeister et al., 1999). These are positioned as frameworks for designing Web applications, that is, apart from hypermedia functionality, they also deal with operations and transactions that affect node content. For that purpose, they extend UML to accommodate for structural and navigational hypermedia abstractions. The underlying hypermedia constructs are borrowed from HDM and OOHDM, respectively. UWE (Koch et al., 2001), combining hypermedia constructs from RMM and OOHDM, stands for a UML-based approach to Web application design as well. However, none of these UML-based approaches allows for link subtyping and multiple specializations over both node and link types in the way *MESH* does.

MESH has the guided tour as default construct for navigating through a set of nodes. All kinds of indexes can be generated just as easily from the information in the linkbase/repository. RMM, HDM, OOHDM and WebML also feature specific topologies such as guided tours, indexes and so forth. A fundamental difference is that these are conceived as explicit design components, requiring extensive authoring efforts for query definitions, node collections and forward/backward links. In MESH, these are generated on-the-fly by the hyperbase engine, based on the current context. The author is not even engaged in their realization. RMM, HDM and OOHDM also in one way or another incorporate node visualization mechanisms that are sensitive to where a node is accessed. MESH, however, is the only approach to use a link's type to determine the destination node's visualization. It is also the only approach that stays true to the object-oriented paradigm by looking upon the associated link types as the public interface to a node, hiding its actual implementation. Finally, in MESH, a user has the unique ability to bookmark a complete navigational situation (instead of single nodes) in a very compact manner, with the possibility to resume navigation later on, from the exact point where it was left.

REFERENCES

Ashman, H., Garrido, A., & Oinas-Kukkonen, H. (1997). Hand-made and computed links, precomputed and dynamic links. *Proceedings of Hypertext - Information Retrieval - Multimedia (HIM '97).*

Baresi, L., Garzotto, F., & Paolini, P. (2001). Extending UML for modeling Web applications. *Proceedings of the 34th Annual Hawaii International Conference on System Sciences (HICSS-34)* (pp. 1285-1294).

Baumeister, H., Koch, N., & Mandel, L. (1999). Towards a UML extension for hypermedia design. *Proceedings of UML '99 The Unified Modeling Language – Beyond the Standard, LNCS 1723*, 614-629.

Ceri, S., Fraternali, P., & Paraboschi, S. (2000). Web modeling language (WebML): A modeling language for designing Web sites. *Proceedings of the 9th International World Wide Web Conference* (pp. 137-159).

Davis, H., Hall, W., Heath, I., Hill, G., & Wilkins, R. (1992). *MICROCOSM: An open hypermedia environment for information integration.* Computer Science Technical Report CSTR 92-15.

Garzotto, F., Paolini, P., & Schwabe, D. (1993). HDM - A model-based approach to hypertext application design. *ACM Transactions on Information Systems, 11*(1).

Halasz, F. (1988). Reflections on NoteCards: Seven issues for next generation hypermedia systems. *Communications of the ACM, 31*(7).

Isakowitz, T., Kamis, A., & Koufaris, M. (1998). *The extended RMM methodology for Web publishing.* Working Paper IS-98-18.

Kappe, F. (1999). *Managing knowledge with Hyperwave Information Server.* Hyperwave White Paper Version 1.2.

Knopik, T., & Bapat, A. (1994). The role of node and link types in open hypermedia systems. *Proceedings of the 6th ACM European Conference on Hypermedia Technology.*

Koch, N., Kraus, A., & Hennicker, R. (2001). The authoring process of the UML-based Web engineering approach. *Proceedings of the* 1st International Workshop on Web-oriented Software Technology (IWWOST´01).

Lemahieu, W. (2002). Context-based navigation in the Web by means of dynamically generated guided tours. *Computer Networks, 39*(3), 311-328.

Lemahieu, W. (2003). Encapsulation and information hiding as the keys to enhanced hypermedia development and maintenance. *Data & Knowledge Engineering, 46*(1), 65-96.

Millard, D., Reich, S., & Davis, H. (1998). Reworking OHP: The road to OHP-Nav. *Proceedings of the 4th Workshop on Open Hypermedia Systems at Hypertext '98* (pp. 48-53).

Nanard, J., & Nanard, M. (1995). Hypertext design environments and the hypertext design process. *Communications of the ACM, 38*(8).

Rossi, G., Schwabe, D., & Lyardet, F. (2000). Abstraction and reuse mechanisms in Web application models. *Proceedings of the World Wide Web and Conceptual Modeling'00 Workshop, ER'00* (pp. 19-21).

Thüring, M., Hannemann, J., & Haake, J. (1995). Hypermedia and cognition: Designing for comprehension. *Communications of the ACM, 38*(8).

Wiil, U., & Leggett, J. (1997). Hyperform: A hypermedia system development environment. *ACM Transactions of Information Systems, 15*(1).

KEY TERMS

Aspect: Defines a set of related node properties, attributable to nodes with the same value for an aspect descriptor. For a given node, aspects can be added and removed at runtime by manipulating the values for its aspect descriptors.

Aspect Descriptor: Attribute for which each possible value defines an aspect of a node.

Authoring: The creation and organization of multimedia or hypermedia content.

Guided Tour: Combination of links that "guides" a user along a collection of hypermedia nodes that have some topic or concept in common. In MESH, guided tours are generated dynamically at runtime, based on the type of previously followed links by a particular user.

Hypermedia: Way of organizing data as a network of nodes and links.

Link: Navigation path between two nodes in a hypermedia environment.

Link Type: Type level definition of properties (domain, inverse link type, cardinalities) for a group of similar links.

Navigation: Means of accessing data in a hypermedia environment, by moving from node to node along the links.

Node: Unit of data (and navigation) in a hypermedia environment.

Node Type: Type-level definition of properties (layout template, link types, aspect descriptors) for a group of similar nodes.

Metadata for Electronic Documents Using the Dublin Core

Till Hänisch
University of Cooperative Education, Heidenheim, Germany

INTRODUCTION

The Dublin Core Element Set was developed at the OCLC/NCSA Metadata Workshop in Dublin (Ohio), 1995 (hence the name). It is maintained by the Dublin Core Metadata Initiative (DCMI).

The Dublin Core Element Set defines attributes, so called elements, which can be used for the description of electronic documents (Table 1).

Some of these elements can be qualified further. For example the "Date" element can be refined to "Created", "Issued", "Valid", "Available" or "Modified". Some elements can be refined by specifying the format or vocabulary used, for example, a qualifier can specify the scheme used to encode a date.

The Dublin Core standard itself does not specify a format (for example a XML schema) for the elements, usually the metadata are expressed either in HTML-documents (Figure 1) or in XML.

BACKGROUND

The definition of the semantics of Dublin Core elements is very flexible. Although, the DCMI recommends best practices for some elements, there is no definitive standard for their content. For example, the "Language" element should be specified according to RFC 3066 and ISO 639, which is a quite formal definition; but for the "Source" element it is recommended to "identify the referenced resource by means of a string or number conforming to a formal identification system" (DCMI, 2003), which gives each implementation a considerable amount of freedom regarding the system to be used. The advantage of these imprecise definitions is the resulting flexibility: Each implementation can use its own vocabulary specific to the actual application. The disadvantage is the lack of interoperability between applications which reduces the benefit of using a common standard.

The Dublin Core standard itself does not specify which elements are required and which are optional (all elements are optional, so an empty set of elements conforms to the Dublin Core). The selection and the concrete semantics of the elements to be used for a given application may be clarified by a so-called profile, for example, the DC-Library Application profile (DCMI, 2002) clarifies the use of the Dublin Core in library related applications. But profiles exist only for a small number of possible applications, so typically each implementation has to define its own semantics.

The Dublin Core element set mainly provides information about the content (for example "Title", "Description", etc.) and "Rights" related topics (for example "Cre-

Table 1. Dublin Core element set (DCMI, 2003)

Title	A name given to the resource
Creator	An entity primarily responsible for making the content of the resource
Subject	A topic of the content of the resource
Description	An account of the content of the resource
Publisher	An entity responsible for making the resource available
Contributor	An entity responsible for making contributions to the content of the resource
Date	A date of an event in the lifecycle of the resource
Type	The nature or genre of the content of the resource
Format	The physical or digital manifestation of the resource
Identifier	An unambiguous reference to the resource within a given context
Source	A Reference to a resource from which the present resource is derived
Language	A language of the intellectual content of the resource
Relation	A reference to a related resource
Coverage	The extent or scope of the content of the resource
Rights	Information about rights held in and over the resource

Figure 1. Example of HTML encoded Dublin Core elements

```
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<link rel="schema.DCTERMS" href="http://purl.org/dc/terms/" />
<meta name="DC.title" lang="german" content="Meine homepage" />
<meta name="DC.creator" content="Till Haenisch" />
<meta name="DC.subject" lang="german" content="Vorlesungen; dlmeta />
<meta name="DC.date" scheme="DCTERMS.W3CDTF" content="2004-02-23" />
<meta name="DC.type" scheme="DCTERMS.DCMIType" content="Text" />
<meta name="DC.format" content="text/html" />
```

ator", "Publisher", etc.) of a resource. There is no (standard) way to include application specific information, for example, the timecodes of the different scenes in a video or the encoding quality of a MP3 file. This leads to a great variety of extensions, either specific for a single application or a class of applications. For example, Einhorn (2003) describes an extension for capturing metadata about presentations. DLmeta (see Abele, 2002) describes an extensible model for multimedia applications. These extensions enable applications to add necessary information to the metadata set, but since there is no common standard, this may lead to incompatibility.

Extending the Dublin Core results in a more complex set of metadata. A recent study (Ward, 2003) shows that even the unqualified elements are often used not completely: "two elements out of fifteen […] make up half the element usage in over half of the DPs [Data Providers]" (Ward, 2003). Nearly everybody (who uses Dublin Core) uses the "Title" and "Creator" elements. Some other elements (for example "Relation" and "Coverage") are rarely used. Maybe the original Dublin Core has too many elements (is too complex) for a lot of applications.

FUTURE TRENDS

There are two ways to solve the problems described above: Extend the element set or reduce it. Models like DLmeta (DLmeta, 2000) try to extend the Dublin Core to allow a more detailed description of electronic documents and their media specific properties and provide a certain degree of flexibility for specialized applications. Other models like the ABC-Model (Legoze, 2001) try to use a more abstract approach to allow the description of arbitrary objects and complex relationships, for example, in museum catalogs. The higher precision of the description when using these models may result in higher cost for the creation of the metadata and application development.

A different approach is to use a minimal common standard which is extensible according to the applications needs. One of the main reasons for using metadata is the ability to locate and retrieve resources. A minimal usable description should at least support this task. Kunze (2001) suggests the use of a set of only 4 elements (Table 2) to describe the basic properties of a resource. If necessary, these elements could be furthermore qualified and extended.

One important reason for a common metadata standard is the interchange of metadata. A local application like a digital library may use a propriety metadata model unless this information should be shared with others, for example, to enable cross domain resource discovery. One solution to this problem is that every application uses the same metadata, another solution is a common minimal subset used (only) for the exchange of information. The Open Archives Initiative (Lagoze, 2001) uses this approach. Metadata is exchanged between data providers using Dublin Core but every provider may deliver additional XML-formatted metadata.

CONCLUSION

The trend to use a more complex metadata set resembles the way, library cataloguing techniques have developed from simple schemes to today's complex standards like MARC. The large number of electronic documents to be described requires efficient techniques for their description. The currently used Dublin Core standard has several deficiencies, but it is not clear if a more complex or a simpler standard should be used. Combinations of a small metadata kernel used for data interchange with optional application specific elements may result in the advantages of both directions.

REFERENCES

Abele, H., Frings, T., Hänisch, T., Aschoff, C. & Winter, C. (2002). DLmeta – noch nie waren Standards so wertvoll wie heute. *Praxis der Informationsverarbeitung und Kommunikation*, *25*(1), 39-43

Byrne, D. J. (1998). MARC Manual: understanding and using MARC (2nd ed.). Englewood, Colo: Libraries Unlimited.

Table 2. Electronic resource citation (ERC) record and corresponding Dublin Core elements

Element	Description	Corresponding DC element	Example
who	A responsible person or party	Creator	Dlmeta initiative
what	A name or other human-oriented identifier	Title	Dlmeta DTD
when	A date important in the object's lifecycle	Date	31.1.2002
where	A location or machine-oriented identifier	Identifier	http://www.dlmeta.de

DLmeta Initiative. (2000). Document type definition (DTD) des DLmeta-Datenmodells. Retrieved from *http://www.dlmeta.de/jdlmeta/dtd/index.jsp*

Dublin Core Metadata Initiative. (2003). Dublin Core metadata element set (Version 1.1) [Reference description]. Retrieved from *http://www.dublincore.org/documents/dces*

Dublin Core Metadata Initiative. (2002). Library application profile. Retrieved from *http://dublincore.org/documents/2002/09/24/library-application-profile/*

Einhorn, R., Olbrich, S. & Nejdl, W. (2003). A metadata model for capturing presentations. *Proceedings of the 3rd IEEE International Conference on Advanced Learning Technologies*, 110-114.

Kunze, J. A., (2001). A metadata kernel for electronic permanence. *Journal of Digital Information*, *2*(2), 84.

Lagoze, C. & Hunter, J. (2001). The ABC ontology and model. *Journal of Digital Information, 2*(2), 77.

Lagoze, C. & Van de Sompel, H. (2001). The open archives initiative: Building a low-barrier interoperability framework. *Proceedings of the first ACM/IEEE-CS joint conference on Digital libraries,* 54-62

Ward, J. (2003). A Quantitative Analysis of Unqualified Dublin Core Metadata Element Set Usage within Data Providers Registered with the Open Archives Initiative. *Proceedings of the 2003 Joint Conference on Digital Libraries*. IEEE Computer Society, 315-317.

KEY TERMS

Catalog: The collection of records used to describe and locate the items contained in a library.

Dublin Core: A widely accepted standard for metadata about electronic documents. Maintained by the Dublin Core Metadata Initiative.

MARC: Machine Readable Cataloging Record. A Standard for the representation and communication of bibliographic and related information in machine readable form, developed by the Library of Congress.

Metadata: Data about other data. This is needed to find, catalogue and organize resources.

Resource: Any information resource like HTML-document, book, video etc.

RFC: Request For Comments. Internet standards are proposed and described in the form of RFCs. They are maintained by the Internet Engineering Task Force.

XML: Extensible Markup Language, an easy to use dialect of SGML, is a flexible technique for storage and interchange of data. One important aspect of XML is the combination of data and metadata in a single document

Methods for Understanding IS Failures

Darren Dalcher
Middlesex University, UK

INTRODUCTION

Researchers with a keen interest in information systems failures are faced with a double challenge. Not only is it difficult to obtain intimate information about the circumstances surrounding such failures, but there is also a dearth of information about the type of methods and approaches that can be utilised in this context to support such information collection and dissemination. The purpose of this chapter is to highlight some of the available approaches and to clarify and enhance the methodological underpinning that is available to researchers interested in investigating and documenting phenomena in context-rich and dynamic environments. The chapter concludes by introducing a new range of antenarrative approaches that represent future developments in the study of IS failures.

BACKGROUND

Contemporary software development practice is regularly characterised by runaway projects, late delivery, exceeded budgets, reduced functionality, and questionable quality that often translate into cancellations, reduced scope, and significant rework cycles (Dalcher, 1994). Failures, in particular, tell a potentially grim tale. In 1995, 31.1% of U.S. software projects were cancelled, while 52.7% were completed late, over budget (cost 189% of their original budget), and lacked essential functionality. Only 16.2% of projects were completed on time and within budget; only 9% in larger companies, where completed projects had an average of 42% of desired functionality (Standish Group, 2000). The 1996 cancellation figure rose to 40% (Standish Group).

The cost of failed U.S. projects in 1995 was $81 billion; in addition, cost overruns added an additional $59 billion ($250 billion was spent on 175,000 U.S. software projects; however, $140 billion out of this was spent on cancelled or over budget activities; Standish Group, 2000). In fact, Jones (1994) contended that the average U.S. cancelled project was a year late, having consumed 200% of its expected budget at the point of cancellation. In 1996, failed projects alone totalled an estimated $100 billion (Luqi & Goguen, 1997). In 1998, 28% of projects were still failing at a cost of $75 billion; while in 2000, 65,000 of U.S. projects were reported to be failing (Standish Group).

The Standish Group makes a distinction between failed projects and challenged projects. Failed projects are cancelled before completion, never implemented, or scrapped following installation. Challenged projects are completed and approved projects which are over budget, late, and with fewer features and functions than initially specified. Lyytinen and Hirschheim (1987) identify: correspondence failures (where the system fails to correspond to what was required), process failures (failure to produce a system or failure to produce it within reasonable budgetary and timescale constraints), interaction failures (where the system cannot be used or is not satisfactory in terms of the interaction), and expectation failures (where the system is unable to meet a specific stakeholder group's expectations). Many situations contain behavioural, social, organisational, or even societal factors that are ignored, and therefore the definition of failure needs to encompass a wider perspective. The general label *system failures* is often utilised in order to embrace a wider grouping of failures, including ones with undesirable side effects which may impact other domains and the organisational context (see, for example, Fortune & Peters, 1995). As information becomes more embedded in other domains, the scope and impact of failure becomes more wide-reaching. This was clearly evident from the extensive effort to minimise the impact of the "year 2000 bug" from any system containing computers and underscores our interest in utilising the term *IS failure* to describe a wider class of systems failures that impact on individuals, organisations, and societal infrastructure.

IS failure investigations start with extensive attempts to collate relevant evidence. However, in most cases the researcher is exposed to specific information post hoc, i.e., once the failure is well established and well publicised and the participants have had a chance to rationalise their version of the story. Most of the available sources are therefore already in place and will have been set up by agencies other than the researcher.

The purpose of a forensic investigation is to explain a given failure by using available information and evidence. The term *forensic* is derived from the Latin *forensis*, which is to do with making public. Forensic science is the applied use of a body of knowledge or practice in determining the cause of death, nowadays extended to include any skilled investigation into how a crime was perpetrated. Forensic systems engineering is the postmortem analysis and study of project disasters (Dalcher, 1994).

The work involves a detailed investigation of a project, its environment, decisions taken, politics, human errors, and the relationship between subsystems. The work draws upon a multidisciplinary body of knowledge and assesses the project from several directions and viewpoints. The aim of forensic analysis is to improve the understanding of failures, their background, and how they come about (Dalcher, 1997). The concept of systems is a central tool for understanding the delicate relationships and their implications in the overall project environment.

Forensic systems engineering is primarily concerned with documentary analysis and (post-event) interviews in an effort to ascertain responsibility lines, causal links, and background information. The primary mode of dissemination of findings, conclusions, and lessons is through the publication of case study reports focusing on specific failures. However, there are limited research methods to explore the dynamic and fragmented nature of complex failure situations. Lyytinen and Hirschheim (1987) noted that more qualitative research methods were needed for IS failure research as well as more extensive case studies that explored problems in more detail and viewed solution arrangements in light of what transpired. The same methods also need to account for group issues and cultural implications. Sadly, 16 years on, the same constraints in terms of methods are still in evidence.

DESCRIBING FAILURE

Making sense of IS failures retrospectively is difficult. In general, there is very little objective quantitative failure information that can be relied upon. This makes the utilisation of quantitative methods less likely until all relevant information is understood. Indeed, a specific feature of failure is the unique interaction between the system, the participants, their perspectives, complexity, and technology (Perrow, 1984). Lyytinen and Hirschheim (1987) pointed out that failure is a multifaceted phenomenon of immense complexity with multiple causes and perspectives. Research into failures often ignores the complex and important role of social arrangement embedded in the actual context. This is often due to the quantitative nature of such research. More recently, Checkland and Holwell (1998) argued that the IS field requires sensemaking to enable a richer concept of information systems.

Understanding the interactions that lead to failures likewise requires a humanistic stance that is outside the conventional positivist norm to capture the real diversity, contention, and complexity embedded in real life. Forensic analysis thus relies on utilising qualitative approaches to obtain a richer understanding of failure phenomena in terms of action and interaction.

The fact that a failure phenomenon is being investigated suggests that attention has already been drawn to the complexities, breakdowns, and messy interactions that such a situation entails (i.e., the investigation is problem-driven). Many such inquiries deal with subjective accounts, including impressions, perceptions, and memories. The aim of the researcher is to increase in a systemic way the understanding of a situation, yet do so from a position that takes in the complexity of the entire situation and incorporates the different perspectives and perceptions of the stakeholders involved.

Overall, the purpose of a failure research method is to enable the researcher to make sense of the complexity of detail and the complexity of interaction and chart the contributory role of different causes and issues in the buildup to failure. However, the armoury of research methods in this domain is often limited to case studies.

The term *case study* is an umbrella term used in different contexts to mean different things that include a wide range of evidence capture and analysis procedures. Yin (1994, p.13) defines the scope of a case study as follows:

“A case study is an empirical inquiry that:

- *investigates a contemporary phenomenon within its real-life context, especially when*
- *the boundaries between phenomenon and context are not clearly identified.”*

A case study can be viewed as a way of establishing valid and reliable evidence for the research process as well as presenting findings which result from research (Remenyi et al., 1998). According to Schramm (1971), the case study tries to illuminate a decision or a set of decisions and in particular emphasise why they were taken, how they were implemented, and with what results. A case study is likely to contain a detailed and in-depth analysis of a phenomenon of interest in context; in our case, the failure scenario. Table 1 summarises some of the main advantages of using case studies.

The general aim of the case study approach is to understand phenomena in terms of issues in the original problem context by providing the mechanism for conducting an in-depth exploration. They often result from the decision to focus an enquiry around an instance or an incident (Adelman, Jenkins, Kemmis, 1977), as they are principally concerned with the interaction of factors and events (Bell, 1999). The combination of a variety of sources offers a richer perspective, which also benefits from the availability of a variety and multiplicity of methods that can be used to obtain new insights about this single instance. A case study allows the researcher to concentrate on specific instances in their natural setting and thereby attempt to identify the interacting perceptions,

Table 1. Main advantages of using case studies

✓	ability to identify and focus on issues
✓	richness of detail
✓	multiple perspectives
✓	multiple explanations (no absolute truth)
✓	cross-disciplinary remit
✓	ability to recognise and minimise inherent complexity
✓	ability to handle conflict, disparity, and disagreement
✓	ability to show interactions
✓	ability to observe emerging patterns
✓	conducted in real-life setting
✓	encompasses original problem context
✓	ability to deal with interpretations
✓	can extend the boundaries to include aspects of wider system environment
✓	can be accumulated to form an archive of cases

Table 2: Main objections to the use of case studies

❖	sometimes viewed as soft data (but some argue it is hard research)
❖	biases inherent in accepting views and perceptions
❖	questions about generalisability of findings (especially from a single case), but it is possible to build a library of such cases
❖	issues regarding objectivity of approach and perceived lack of rigour
❖	negotiating access to settings
❖	boundaries are difficult to define, but this could also be a strength!
❖	mainly retrospective
❖	sometimes viewed as likely to take too long and result in massive documentation
❖	the observer effect
❖	reliability of conclusions
❖	there is little control over events, but this may also be a strength

issues, and processes at work, ultimately resulting in in-depth understanding. Crucially, the focus on a single incident thus enables the study of the particularity and complexity of a case, thereby coming to understand the activity within important circumstances (Stake, 1995).

There are a number of general objections that are associated with the use of case studies (see Table 2). However, one must recognise that case studies are more likely to be used retrospectively rather than as an ongoing perspective (especially from a failure point-of-view), as researchers are unlikely to know the potential for useful results and interest from the outset.

The richness of detail can be controlled through the careful placement of systems boundaries and consideration of the wider system environment that is relevant to the phenomenon under study. Case studies can be utilised as a source of understanding which is tolerant of ambiguity, paradox, and contradiction. A case study is viewed as interpretative when events in the real world are observed and then an effort takes place to make sense of what was observed, i.e., when one tries to make sense of a failure from the perspectives of participants. They also offer the potential for generating alternative explanations from the different stakeholder perspectives, thereby allowing the researcher to highlight contradictions, conflicts, and misunderstandings.

FROM CASE STUDIES TO CASE HISTORIES

The generally liberal use of the term *case study* requires a tighter definition of its meaning in failure research. While there may be a tradition of using case studies within the IS community, this is perhaps more often borrowed from the MBA culture than as a result of self-conscious effort to adopt them as a research approach (Cornford & Smithson, 1996; Walsham, 1995). Indeed, the case study is typically used more in its capacity as a teaching tool than as a *research tool*. The shift to studying the impact of issues within the organisational context renders case studies particularly useful for investigating failure scenarios. However, the use of the term often leads to some confusion.

Moreover, one of the major complications in failure investigations is in relating causes to effects (and possibly events) through extended time horizons (Dalcher, 2000). The implications of actions may not be witnessed for years or even generations. Delays between making a decision and observing the result distort the causal link between the two. As a result, people tend to associate a different level of severity to events occurring following a delay. The perceived severity is thus diminished, with the length of the delay further complicating the task of identifying patterns and interactions that contributed to a given failure. Failure researchers are thus required to provide adequate historical accounts of the interaction between actions, perceptions, and the passage of time.

Case studies have typically been used to explore issues in the present and the past and comprise ethnographic studies, single case studies, and comparative case studies (Jankowicz, 2000), as well as action research, evaluative, exploratory, explanatory, and descriptive case studies (Bassey, 1999). In our experience there is a need to add the failure case study as a special example of a case study focusing primarily on the background, context, perception, interactions, and patterns, especially as the failure investigation is likely to take place *after* the (failure) event. We thus propose the use of the label *case histories* to refer to the specialised historical research studies focusing on failure incidents.

The time dimension (sequencing) is critical to understanding interactions and identifying their impacts. Case histories are concerned with providing the background and context that are required to endow words and events with additional meaning. Background refers to previous history of the system itself, while context refers to interactions with the environment. As failures are time- and place-dependent, the case history framework enables readers to obtain an understanding of the intimate context surrounding the main event. The primary tool available to the community is the case histories of failures (derived from the use of the case study method). These represent a detailed historical description and analysis of actual processes. Their value is in tracing decisions (and recorded rationale) to their eventual outcomes by utilising techniques borrowed from decision analysis and systems engineering. Indeed, the historical description and presentation of a chronology of events are based on the recognition that real life is ambiguous, conflicting, and complex.

Case histories thus contain observations, feelings, and descriptions. They can be used to construct, share, dispute, and confirm meanings, interpretations, and scenarios in the context of real events. Rather than simply highlight a chronicled sequence of happenings, they convey a story encompassing a specific perspective, focus, and possibly some inevitable biases. The interpretation plays a key part in transmuting the chronicle into a meaningful story with plot, coherence, and purpose. However, constructing a convincing narrative of a complex story with competing meanings, alternative perspectives, and inherent prejudices is a challenge in itself.

FUTURE TRENDS

Failures, in common with other activities that take place in organisations, are based on stories. The verbal medium is crucial to understanding behaviour within organisations and systems, and researchers are thus required to collect *stories*, grounded in practice, about what takes place (Easterby-Smith, Thorpe, & Lowe, 2002; Gabriel, 2000). Understanding failures often entails the retrospective untangling of complicated webs of actions and events and emergent interaction patterns. Failure storytelling can thus be understood as a combination of narrative recounting of empirical events with the purposeful unlocking of meaningful patterns, or a plot.

Historically, storytelling has been an acceptable form of conveying ideas, experience, and knowledge of context. It plays a key role in communicating the cultural, moral, or historical context to the listener. Indeed, Arendt (1958) argued that the chief characteristic of human life is that it is always full of events, which ultimately can be told as a story. There are even strong claims that the narrative is the main mode of human knowledge (Bruner 1986, 1990, Schank 1990), as well as the main mode of communication (Denning, 2001; Fisher, 1987; Schank). Moreover, children are often initiated into culture (and its boundaries) through the medium of storytelling, offering models for emulation or avoidance.

In practice, the essence of any good case study revolves around the ability to generate an effective story line, normally with a unique style, plot, or perspective. In a large case, a general theme can be obtained from selected excerpts weaved together to illustrate a particular story. Personal stories that form part of a case study can thus be viewed as a valid source of data organised to make sense of a theme or problem. This is particularly useful when the researcher is trying to portray a personal account of a participant, a stakeholder, or an observer in an incident, accident, or failure. The implication is that the need to address personal aspects of interaction and story is fulfilled by the development of a research-valid narrative. Indeed, Remenyi et al. (1998) contend that a story, or a narrative description, is valid if the resulting narrative adds some knowledge. Furthermore, White (1973, p. 24) describes a story as "the process of selection and arrangement of data from the unprocessed historical record in the interest of rendering the record more comprehensible to an audience of a particular kind" by inserting a sense of perspective and purpose.

Narratives are neither discovered nor found: they are constructed. Understanding IS failures is therefore more complicated than the discovery of a simplistic chronology of events. Narrative inquiry is evolving into an acceptable research approach in its own right in the social sciences and in management research circles (Bell, 1999; Boje, 2001; Czarniawska, 1998; Easterby-Smith et al., 2002; Gabriel, 2000), as the story format provides a powerful way of knowing and linking disparate accounts and perspectives. When different accounts are combined, the story line benefits from the richness of multifaceted insights.

Developing a narrative requires plot as well as coherence, as a story is made out of events and the plot mediates between the events and the story (Boje, 2001; Carr, 2001; Kearney, 2002). In failure stories, the plot often emanates from the actions and perceptions of participants emerging out of the flux of events in (direct) contradiction with expectations. The storyteller is concerned with the perspective and purpose of participants as well as with the plausibility of the emerging plot. The combination of plot, purpose, and perspective dictates the selection of elements, the filling in of links, and the removal of "irrelevant" noise.

Postmodern interpretation contends that most real-life stories are fragmented, nonlinear, multivariate, and incoherent. This has already been highlighted as a feature of failure stories. Such stories also tend to be dynamic, polyphonic (multi-voiced), and collectively produced, as they occur in asymmetrical, random, and turbulent environments. The stories are not plotted as such and they appear to flow, emerge, and network, offering complex clustering of events, emergent phenomena, causes, and effects. Moreover, the accounts are often subjective, counterintuitive, and contradictory. This leads to interacting and conflicting webs of narratives, characterised by coincidences, predicaments, and crises.

Generally, stories appear to be improperly told, as a story is an "ante" state of affairs, existing previously to a carefully constructed narrative (Boje, 2001). The antenarrative, or the "real" story, is the fragmented, messy and dynamic, multi-vocal, multi-plotted, multi-version, and complex tale. Indeed, modern storytellers look for new ways and mediums for weaving and depicting a multi-vocal reality, as exemplified by Mike Finggis' digitally shot film *Time's Arrow*, where the screen is split in four to allow for four separate perspectives and substories that occasionally intersect or overlap. In the tradition of postmodern inquiry, a real-life researcher is often faced with fragments rather than a whole story to tell, and many of the fragments may reflect contrary versions of reality. This is potentially more acute when the accounts attempt to justify roles of participants in the lead-up to disaster. It would also appear from past analysis that there are hierarchies of stories and stories that exist within or interact with other stories. Using the terminology provided by Boje, the purpose of narrative methods is to take a complex situation characterised by collective (yet often conflicting) memory and an antenarrative and construct the plot and coherence that can be used to narrate the story of interest.

The reality in failure stories is of multistranded stories of experiences and reactions that lack collective consensus. Indeed the discipline of decision making has also recognised that making choices is about forming and selecting interpretations from a mosaic of possibilities (March, 1994, 1997; Weick, 1995). Not surprisingly, disasters or traumatic stories are hard to narrate, understand, and justify. Stories have three basic properties: time, place, and mind (Boje, 2001), which interact and build up as the story evolves. In forensic case histories, these are further clarified through the identification of the background and context, which clarify and justify the interpretation in the context of the emerging phenomena.

Boje (1991, 2001) and Kearney (2002) contend that the current view is of sequential single-voice stories and implies excessive reliance on the hypothetical-deductive approach (akin to simplistic causal pairings). The answer is not to develop Harvard case studies but to rewrite stories as polyvocal tapestries enabling different perceptions and interpretations to exist, thereby explaining webs of actions and interactions. What is new in this approach is the antenarrative reading, which enables narrative analysis methods to be supplemented by antenarrative methods, allowing previously fragmented and personal storytelling to be interpreted as a unified whole. This focus offers alternative discourse analysis strategies that can be applied where qualitative story analyses can help to assess subjective, yet "insightful" knowledge in order to obtain "true" understanding of complex interactions.

As for the long-term future, good stories can also benefit from pictures. Once we have mastered the techniques of telling complex, modern stories, we need to focus on composing that information. Even the most gripping story needs to be made attractive and believable. Textual information needs additional support not only in "emplotting" and in maintaining coherence and perspective, but also in ascertaining the plausibility of constructed stories and in differentiating between noise and narrative. Developing improved techniques for organising or visualising knowledge (such as Net maps) can therefore help in untangling some of the fragmented strands as well as in making the stories more readable and understandable, as well as ultimately more appealing.

CONCLUSION

With the benefit of hindsight it is possible to reconstruct a systematic retelling of events that have led to a failure. The narrated structure provides an explanation as to how and why failures occur. The purpose of the structure is to make sense of a rich tapestry of interactions and connections by following an identified story line that chronicles and links the relevant issues within the environment. Indeed, recounted life may prise open perspectives that would have been inaccessible using ordinary methods and thinking arrangements. Moreover, failure tends to highlight missing and incorrect assumptions and faulty defensive mechanisms and can therefore serve as a pretext to updating the frame of reference or the context for understanding.

REFERENCES

Adelman, C., Jenkins, D., & Kemmis, S. (1977). Rethinking case study: Notes from the Second Cambridge Conference. *Cambridge Journal of Education*, *6*, 139-150.

Arendt, H. (1958). *The human condition*. Chicago: University of Chicago Press.

Bassey, M. (1999). *Case study research in educational settings*. Buckingham, England: Open University Press.

Bell, J. (1999). *Doing your research project: A Guide for first-time researchers in education and social science* (3rd ed.). Buckingham, England: Open University Press.

Boje D. M. (1991). The Storytelling Organization: A Study of Story Performance in an Office-Supply Firm, *Administrative Science Quarterly*, 36, March, 106-126.

Boje, D. M. (2001). *Narrative methods for organisational & communication research*. London: Sage.

Bruner, J. (1986). *Actual minds, Possible worlds*. Cambridge, MA: Harvard University Press.

Bruner, J. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.

Carr, D. (2001). Narrative and the real world: An argument for continuity. In G. Roberts (Ed.), *The history and narrative reader* (pp. 143-156). London: Routledge.

Checkland, P., & Holwell, S. (1998). *Information, systems and information systems—Making sense of the field*. Chichester, England: Wiley.

Cornford, T., & Smithson, S. (1996). *Project research in information systems: A student's guide*. Basingstoke, England: Macmillan.

Czarniawska, B. (1998). *A narrative approach to organization studies*. London: Sage.

Dalcher, D. (1994). Falling down is part of growing up; The study of failure and the software engineering community. In *Proceedings of Seventh SEI Education in Software Engineering Conference* (pp. 489-496). New York: Springer-Verlag.

Dalcher, D. (1997). The study of failure and software engineering research. In *Proceeding of the UK Software Engineering Association Easter Workshop* (pp. 14-19). London: Imperial College.

Dalcher, D. (2000). Feedback, planning and control—A dynamic relationship. In *FEAST 2000* (pp. 34-38). London: Imperial College.

Denning, S. (2001). *The springboard: How storytelling ignites action in knowledge-era organizations*. Boston: Butterworth-Heinemann.

Easterby-Smith, M., Thorpe, M., & Lowe, A. (2002). *Management research* (2nd ed.). London: Sage.

Fisher, W. R. (1987). *Human communication as narration: Towards a philosophy of reason, value and action*. Columbia, SC: University of South Carolina Press.

Fortune, J., & Peters, G. (1995). *Learning from failure: The systems approach*. Chichester, England: Wiley.

Gabriel, Y. (2000). *Storytelling in organizations: Facts, fictions and fantasies*. Oxford, England: Oxford University Press.

Jankowicz, A. D. (2000). *Business research projects* (3rd ed.). London: Business Press.

Jones, C. (1994). *Assessment and control of software risks*. Englewood Cliffs, NJ: Prentice Hall.

Kearney, R. (2002). *On stories*. London: Routledge.

Luqi, & Goguen, J. A. (1997). Formal methods: Promises and problems. *IEEE Software*, *14*(1), 73-85.

Lyytinen, K., & Hirschheim, R. (1987). Information systems failures: A survey and classification of the empirical literature. *Oxford Surveys in Information Technology*, *4*, 257-309.

March, J. G. (1994). *A primer on decision making*. New York: Free Press.

March, J. G. (1997). Understanding how decisions happen in organisations. In Z. Shapira (Ed.), *Organisational decision making* (pp. 9-34). Cambridge, England: Cambridge University Press.

Perrow, C. (1984). *Normal accidents: Living with high-risk technologies*. New York: Basic Books.

Remenyi, et al. (1998). *Doing research in business and management: An introduction to process and method*. London: Sage.

Schank, R. C. (1990). *Tell me a story: Narrative and intelligence*. Evanston, IL: Northwestern University Press.

Schramm, W. (1971). *Notes on case studies of instructional media projects* (Working paper). Washington, DC: Academy for Educational Development..

Stake, R. E. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.

Standish Group. (2000). *Chaos 2000*.Dennis, MA: Standish.

Walsham, G. (1993). *Interpreting information systems in organizations*. Chichester, England: Wiley.

Weick, K. E. (1995). *Sensemaking in organisations*. Thousand Oaks, CA: Sage.

White, H. (1973). *Metahistory*. Baltimore: John Hopkins University Press.

Yin, R. K. (1994). *Case study research: Design and methods* (2nd ed.). Newbury Park, CA: Sage.

KEY TERMS

Antenarrative: the fragmented and messy and dynamic stories of real life in their original context before a clear narrative is developed to explain away a certain aspect.

Case History: specialised historical research focusing on failure incidents. Case histories emphasise the background and context that can help in untangling relationships and causes.

Case Study: investigation of phenomena in naturalistic setting, conducted in order to enable in-depth analysis of that phenomena.

Challenged Projects: completed and approved projects that are over budget, late, and have fewer features and functions than originally specified.

Failed Projects: projects that are: cancelled before completion, are never implemented, or are scrapped following installation.

Forensic Systems Engineering: postmortem analysis and study of project failures and disasters aimed at uncovering causes and relationships.

Storytelling: a method of communicating and sharing ideas, experiences, and knowledge in a specific context.

Metrics for Data Warehouse Quality

Manuel Serrano
University of Castilla-La Mancha, Spain

Coral Calero
University of Castilla-La Mancha, Spain

Mario Piattini
University of Castilla-La Mancha, Spain

INTRODUCTION AND BACKGROUND

It is known that organizations are very rich in data but poor in information. Today, technology has made it possible for organizations to store vast amounts of data obtained at a relatively low cost, however these data fail to provide information (Gardner, 1998). Data warehouses have appeared as a solution to this problem supporting decision-making processes and new kinds of applications as marketing.

A data warehouse is defined as a "collection of subject-oriented, integrated, non-volatile data that supports the management decision process" (Inmon, 1997). Data warehouses have become the key trend in corporate computing in the last years, since they provide managers with the most accurate and relevant information to improve strategic decisions. Also, the future for data warehouse is promising. Jarke, Lenzerin, Vassilou, and Vassiliadis (2000) forecast a market of 12 million U.S. dollars for data warehouse markets for the next few years. However, the development of a data warehouse is a difficult and very risky task. It is essential that we can assure the information quality of the data warehouse as it becomes the main tool for strategic decisions (English, 1999).

Information quality of a data warehouse comprises data warehouse system quality and presentation quality (see Figure 1). In fact, it is very important that data in a data warehouse reflect correctly the real world, but it is also very important that data can be easily understood. In data warehouse system quality, as in an operational database (Piattini, Genero, Calero, Polo, & Ruiz, 2000), three different aspects could be considered: DBMSs quality, data model quality, and data quality.

In order to assess DBMS quality, we can use an international standard like ISO 9126 (ISO, 1999), or some of the existing product comparative studies. This type of quality should be addressed in the product selection stage of the data warehouse life cycle.

Data quality must address mostly in the extraction, filtering, cleaning and cleansing, synchronization, aggregation, loading, and so forth, activities of the life cycle. In the last few years, very interesting techniques have been proposed to assure data quality (Bouzeghoub & Kedad, 2002).

Last but not least, data warehouse model quality has a great influence in the overall information quality. The designer must choose the tables, processes, indexes and data partitions representing the logical data warehouse and facilitating its functionality (Jarke et al., 2000).

Figure 1. Information and data warehouse quality

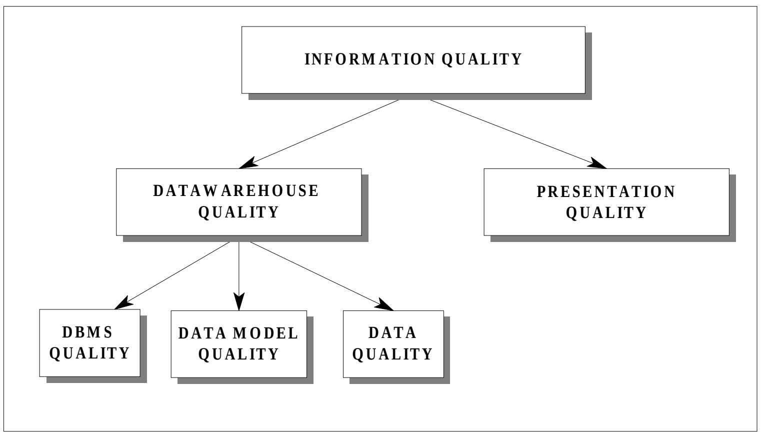
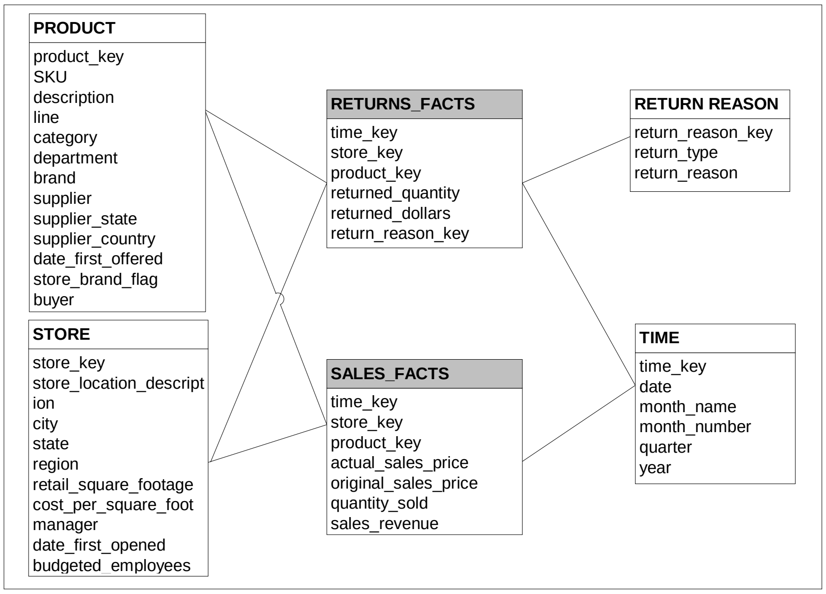


Figure 2. Example of a multidimensional data model design



Multidimensional data models are used to design data warehouses (Petersen & Jensen, 2001). A multidimensional data model is a direct reflection of the manner in which a business process is viewed. It captures the measurements of importance to a business, and the parameters by which the measurements are broken out. The measurements are referred to as *fact* or *measures*. The parameters by which a fact can be viewed are referred to as *dimensions* (Adamson & Venerable, 1998).

Usually multidimensional data models are represented as star schemas, which consist of one central table and several dimensional tables. The measures of interest are stored in the fact table (e.g., sales, inventory). For each dimension of the multidimensional model, there exists a dimensional table (e.g., product, time) that stores the information about the dimension (Jarke et al., 2000).

In Figure 2, we present an example of multidimensional data model design, with two fact tables (Returns_Facts and Sales Facts) and four dimensional tables (Product, Store, Return_Reason and Time).

In recent years, different authors have proposed some useful guidelines for designing multidimensional data models (Bouzeghoub & Kedad, 2002; Jarke et al., 2000; Vassiliadis, 2000). However, more objective indicators are needed to help designers and managers to develop quality multidimensional data models (Hammergren, 1996; Kelly, 1997; Kimball, Reeves, Ross, & Thornthwaite, 1998). Also, interesting recommendations for achieving a "good" multidimensional data model have been suggested (Adamson & Venerable, 1998; Inmon, 1997; Kimball et al., 1998), but quality criteria are not enough on their own to ensure quality in practice, as different people will generally have different interpretations of the same criteria. The goal should be to replace intuitive notions of design "quality" with formal, quantitative, objective metrics in order to reduce subjectivity and bias in the evaluation process.

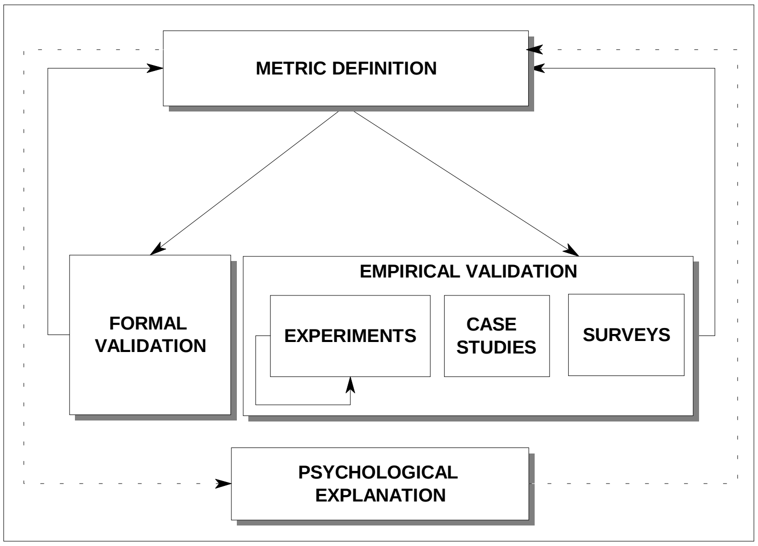
The definition of a set of objective metrics for assuring multidimensional data model quality is the final aim of our work. As we know, quality depends on several factors and characteristics such as functionality, reliability, usability, understandability... (external attributes) (ISO, 1999). Several of these characteristics are influenced by the complexity (internal attribute) of the multidimensional data model. We tried to obtain a set of metrics for measuring the complexity of datawarehouse models that help designers to improve the quality of their datawarehouses.

However, it is not enough with proposing metrics, and it is fundamental to be sure that these metrics are really useful for the goal they were conceived through different kinds of validations.

In this article, we will propose metrics for multidimensional data models quality, which can characterize their complexity and the different validations we have made with them.

In the next section, we will present the framework followed to define and validate metrics. The third section summarizes the proposed metrics, and in the fourth section, the formal validation of these metrics is described. Part of the empirical validations are presented in the fifth section and conclusions and future work will be presented in the final section.

Figure 3. Steps followed in the definition and validation of the metric



A FRAMEWORK FOR DEVELOPING AND VALIDATING DATA WAREHOUSE METRICS

As we have said previously, our goal is to define metrics for controlling data warehouse complexity. Metrics definition must be done in a methodological way, it is necessary to follow a number of steps to ensure the reliability of the proposed metrics. Figure 3 presents the method we apply for the metrics proposal.

In Figure 3, we have four main activities:

- **Metrics definition.** The first step is the definition of metrics. Although it looks simple, it is an important one in ensuring metrics are correctly defined. This definition has made taking into account the specific characteristics of the multidimensional data model we want to measure and the experience of database designers and administrators of these information systems.
- **Formal validation.** The second step is the formal validation of the metrics. The formal validation helps us to know when and how to apply the metrics. There are two main tendencies in metrics formal validation: the frameworks based on axiomatic approaches and the ones based on the measurement theory. The goal of the first ones is merely definitional. On this kind of formal framework, a set of formal properties is defined for a given software attribute, and it is possible to use this property set for classifying the proposed metrics. The most well-known frameworks of this type are those proposed by Weyuker (1988), Briand, Morasca, and Basili (1996), and Morasca and Briand (1997).

  The measurement theory-based frameworks (such as Zuse, 1998; Whitmire, 1998; Poels & Dedene, 2000) specify a general framework in which metrics should be defined. The strength of measurement theory is the formulation of empirical conditions from which we can derive hypothesis of reality.
- **Empirical validation.** The goal of this step is to prove the practical utility of the proposed metrics. Although there are various ways of performing this step, basically we can divide the empirical validation into experimentation, case studies and surveys.
- **Psychological explanation**. Ideally, we will be able to explain the influence of the values of the metrics from a psychological point of view. Some authors, as Siau (1999), propose the use of cognitive psychology as a reference discipline in the engineering of methods and the studying of information modeling.

As shown in Figure 3, the process of defining and validating metrics is evolutionary and iterative. As a result of the feedback, metrics could be redefined based on discarded formal, empirical validation or psychological explanation.

PROPOSED METRICS

In this section, we present the metrics we have proposed for multidimensional data models. As some metrics can be applied at table, star and schema level, we present them separately.

M

Table Level Metrics

In the last few years, we have researched different metrics for assuring relational database quality (Calero, Piattini, & Genero, 2001). Two of these metrics could be useful for data warehouses:

- **NA(T).** Number of attributes of a table.
- **NFK(T).** Number of foreign keys of a table.

In Table 1, we can find the values of the table metrics fort the star schema shown in Figure 2.

Star Level Metrics

- **NDT(S).** Number of dimensional tables of a star.
- **NT(S).** Number of tables of a star, which corresponds to the number of dimensional tables added the fact table.

  NT(S) = NDT(S) + 1

- **NADT(S).** Number of attributes of dimensional tables of a star.
- **NAFT(S).** Number of attributes of the fact table of a star.
- **NA(S).** Number of attributes of a star.

  NA(S) = NAFT(FT) + NADT(S)
  where FT is the fact table of the star S.
- **NFK(S).** Number of foreign keys of a star.

$$NFK(S) = NFK(FT) + \sum_{i=1}^{NDT} NFK(DT_i)$$

  where FT is the fact table of the star S
  and Dti is the dimensional table number i of the star S
- **RSA(S).** Ratio of star attributes. Quantity of attributes of dimensional tables per number of attributes of the fact table of the star.

$$RSA(S) = \frac{NADT(S)}{NAFT(FT)}$$

  where FT is the fact table of the star S.

Table 1. Values for table metrics

	NA	NFK
PRODUCT	13	0
STORE	10	0
RETURN REASON	3	0
TIME	6	0
RETURNS_FACTS	6	4
SALES-FACTS	7	3

- **RFK(S).** Ratio of foreign keys. Quantity of the fact table attributes which are foreign keys.

$$RFK(S) = \frac{NFK(FT)}{NAFT(FT)}$$

  where FT is the fact table of the star S

In Table 2, we can find the values of the star metrics for the star schema shown in Figure 2.

Schema Level Metrics

- **NFT(Sc).** Defined as a number of fact tables of the schema.
- **NDT(Sc).** Number of dimensional tables of the schema.
- **NSDT(Sc).** Number of shared dimensional tables. Number of dimensional tables shared for more than one star of the schema.
- **NT(Sc).** Number of tables. Number of the fact tables plus the number of dimensional tables of the schema.
  NT(Sc) = NFT(Sc) + NDT(Sc)
- **NAFT(Sc).** Number of attributes of fact tables of the schema.

$$NAFT(Sc) = \sum_{i=1}^{NFT} NA(FT_i)$$

  where FTi is the fact table i of the schema Sc
- **NADT(Sc).** Number of attributes of dimensional tables of the schema.

$$NADT(Sc) = \sum_{i=1}^{NDT} NA(DT_i)$$

  where DTi is the dimensional table i of the schema Sc
- **NASDT(Sc).** Number of attributes of shared dimensional tables of the schema.

$$NASDT(Sc) = \sum_{i=1}^{NSDT} NA(DTi)$$

Table 2. Values for star metrics

	Returns_Facts	Sales_Facts
NA	38	36
NFK	4	3
NDT	4	3
NT	5	4
NADT	32	29
NAFT	6	7
RSA	32/6	29/7
RFK	4/6	3/7

where DTi is the dimensional table i of the schema Sc

- **NA(Sc).** Number of attributes of the schema.

  NA(Sc) = NAFT(Sc) + NADT(Sc)
- **NFK(Sc).** Number of foreign keys in all the fact tables of the schema.

$$NFK(Sc) = \sum_{i=1}^{NFT} NFK(FT_i)$$

where FTi is the fact table i of the schema Sc

- **RSDT(Sc).** Ratio of Shared Dimensional Tables. Quantity of dimensional tables, which belong to more than one star.

$$RSDT(Sc) = \frac{NSDT(Sc)}{NDT(Sc)}$$

- **RT(Sc).** Ratio of tables. Quantity of dimension tables per fact table.

$$RT(Sc) = \frac{NDT(Sc)}{NFT(Sc)}$$

- **RScA(Sc).** Ratio of Schema Attributes. Number of attributes in dimension tables per attributes in fact tables.

$$RScA(Sc) = \frac{NADT(Sc)}{NAFT(Sc)}$$

- **RFK(Sc).** Ratio of foreign keys. Quantity of attributes that are foreign keys.

$$RFK(Sc) = \frac{NFK(Sc)}{NA(Sc)}$$

- **RSDTA(Sc).** Ratio of Shared Dimensional Tables Attributes. Number of attributes of the schema that are shared.

$$RSDTA(Sc) = \frac{NASDT(Sc)}{NA(Sc)}$$

In Table 3, we can find the values of the schema metrics for the star schema shown in Figure 2.

Table 3. Values for data warehouse schema metrics

Metric	Value
NA	45
NFK	7
NDT	4
NT	6
NADT	32
NAFT	13
RFK	7/45
NFT	2
NSDT	3
NASDT	29
RSDT	3/4
RT	4/2
RScA	32/13
RSDTA	29/45

METRICS FORMAL VALIDATION

As we have said, there are two basic tendencies in formal metrics validation: axiomatic approaches and measurement theory. In this section, we will present both validation techniques. The formal framework proposed by Briand et al. (1996) is an example of axiomatic approach and the Zuse's formal framework is based on measurement theory.

In Table 4, we present the results obtained for all the presented metrics on both formal frameworks.

METRICS EMPIRICAL VALIDATION

In the past, empirical validation has been an informal process relying on the credibility of the proposer. Oftentimes, when a metric was identified theoretically as an effective metric of complexity, then practitioners and researchers began to use the metric without questioning its validity. Today, many researchers and practitioners assume that validation of a metric (from a theoretical point of view) is not sufficient for widespread acceptance. They expect the empirical validation to demonstrate that the metric itself can be validated.

We have made several experiments (Serrano, Calero, & Piattini, 2002, 2003), and we have found that several metrics (like NT, NFT, NDT and NFK) seem to be a good quality indicator for logical data warehouse schemas.

CONCLUSIONS AND FUTURE TRENDS

If we really consider that information is "the" main organizational asset, one of our primary duties should be assuring its quality. Although some interesting guidelines have been proposed for designing "good" multidimensional models for data warehouses, more objective indicators are needed. Metrics are useful and objective mechanisms for improving the quality of software products and also for determining the best ways to help professionals and researchers.

In this way, our goal is to elaborate a set of metrics for measuring data warehouse quality which can help designers in choosing the best option among more than one alternative design.

M

Table 4. Summary of metrics formal validation

	BRIAND ET AL. (1996)	ZUSE (1998)
NA	SIZE	ABOVE THE ORDINAL
NFK	COMPLEXITY	ABOVE THE ORDINAL
NDT	SIZE	ABOVE THE ORDINAL
NT	SIZE	RATIO
NADT	SIZE	ABOVE THE ORDINAL
NAFT	SIZE	ABOVE THE ORDINAL
NFT	SIZE	RATIO
NSDT	SIZE	ABOVE THE ORDINAL
NASDT	SIZE	RATIO
RSA	NOT CLASSIFIABLE	ABSOLUTE
RFK	NOT CLASSIFIABLE	ABSOLUTE
RSDT	NOT CLASSIFIABLE	ABSOLUTE
RT	NOT CLASSIFIABLE	ABSOLUTE
RSDTA	NOT CLASSIFIABLE	ABSOLUTE

As a start of this work, we presented in this article some metrics that have been defined for measuring the data warehouse star design complexity. We presented the formal validation of these metrics, and we explained the first experiment we developed in order to validate them.

As a conclusion of the formal validation, we obtained that all the metrics are useful from the Zuse (1998) measurement theory framework point of view, and most of them are classifiable by the Briand et al. (1996) axiomatic formal framework. As a result of the experiments we performed, we found that several of the proposed metrics (NFT, NT, NDT and NFK) seem to be good quality indicators for logical data warehouse schemas.

Following the MMLC (Measure Model Life Cycle) of Cantone and Donzelli (2000) with this and other experiments, our research can be classified into the creation phase. At this moment, we are starting different collaborations with companies in order to go to the acceptance phase through the systematic experimentation in a context suitable to reproduce the characteristics of the application environment, with real business cases and real users.

ACKNOWLEDGEMENT

This research is part of the CALIPO project, supported by Dirección General de Investigación of the Ministerio de Ciencia y Tecnologia (TIC2003-07804-C05-03).

REFERENCES

Adamson, C., & Venerable, M. (1998). *Data warehouse design solutions.* John Wiley and Sons.

Bouzeghoub, M., & Kedad, Z. (2002). *Quality in data warehousing in information and database quality* (pp.163-198). Kluwer Academic Publisher.

Briand, L.C., Morasca, S., & Basili, V. (1996). Property-based software engineering measurement. *IEEE Transactions on Software Engineering*, *22*(1), 68-85.

Calero, C., Piattini, M., & Genero, M. (2001). Empirical validation of referential integrity metrics. *Information and Software Technology*, *43*(15), 949-957.

Cantone G., & Donzelli, P. (2000). Production and maintenance of software measurement models. *Journal of Software Engineering and Knowledge Engineering*, *5*, 605-626.

English, L.P. (1999). *Improving data warehouse and business information quality: Methods for reducing costs and increasing profits*. New York: John Wiley & Sons.

Gardner, S.R. (1998). Building the data warehouse. *Communications of the ACM*, *41*(9), 52-60.

Hammergren, T. (1996). *Data warehousing building the corporate knowledge base*. Milford: International Thomson Computer Press.

Inmon, W.H. (1997). *Building the data warehouse* (2nd ed.). John Wiley and Sons.

ISO. (1999). Software product evaluation-quality characteristics and guidelines for their use. *ISO/IEC Standard 9126*, Geneva.

Jarke, M., Lenzerin, I.M., Vassilou, Y., & Vassiliadis, P. (2000). *Fundamentals of data warehouses*. Springer.

Kelly, S. (1997). *Data warehousing in action.* John Wiley & Sons.

Kimball, R., Reeves, L., Ross, M., & Thornthwaite, W. (1998). *The data warehouse lifecycle toolkit*. John Wiley and Sons.

Morasca, S., & Briand, L.C. (1997). Towards a theoretical framework for

measuring software attributes. *Proceeding of the Fourth International,*

Software Metrics Symposium (pp.119-126).

Petersen, T.B., & Jensen, C.S. (2001). Multidimensional database technology. *Computer, 34*(12), 40-46

Piattini, M., Genero, M., Calero, C., Polo, M., & Ruiz, F. (2000). Database quality. In O. Diaz & M. Piattini (Eds.), *Advanced database technology and design*. London: Artech House.

Poels, G., & Dedene, G. (2000). Distance-based software measurement: Necessary and sufficient properties for software measures. *Information and Software Technology*, *42*(1), 35-46.

Serrano, M., Calero, C., & Piattini, M. (2002). Validating metrics for data warehouses. *Proceedings of the Conference on Empirical Assessment in Software Engineering (EASE 2002),* Keele, UK, 8-10 April.

Serrano, M., Calero, C., & Piattini, M. (2003). Experimental validation of multidimensional data models metrics. *Proc of the Hawaii International Conference on System Sciences (HICSS'36),* January 6-9.

Siau, K. (1999). Information modeling and method engineering: A psychological perspective. *Journal of Database Management, 10*(4), 44-50.

Vassiliadis, P. (2000). Data warehouse modeling and quality issues. PhD thesis, Department of Electrical and Computer Engineering.

Weyuker, E.J. (1988). Evaluating software complexity measures. *IEEE Transactions on Software Engineering, 14*(9), 1357-1365.

Whitmire, S.A. (1997). *Object oriented design measurement*, Ed. Wiley.

Zuse, H. (1998). *A framework of software measurement.* Berlin: Walter de Gruyter.

KEY TERMS

Data Mart: A data warehouse that is designed for a particular line of business, such as sales, marketing, or finance.

Data Warehouse: Collection of subject-oriented, integrated, non-volatile data that supports the management decision process.

Multidimensional Data Model: A direct reflection of the manner in which a business process is viewed. It captures the measurements of importance to a business, and the parameters by which the measurements are broken out. The measurements are referred to as *fact* or *measures*. The parameters by which a fact can be viewed are referred to as *dimensions.* Multidimensional data models are used to design datawarehouses.

On-Line Analytical Processing (OLAP): Systems oriented to analyse huge amounts of data. These systems are optimized for querying and their functionality is characterised by dynamic, multidimensional analysis of historical data.

On-Line Transaction Processing (OLTP): Systems that are optimized for transaction handling. These kinds of information systems are used in day-to-day work.

Software Metric: The term metric or software metric has, for many years, been accepted as the standard term used to refer to the instrument of measurement in the field of software. The term software measure was introduced in the 80's. This new term was introduced because, in general, a metric captures two arguments and brings the distance between them, so from that point of view, a software metric cannot really be defined as a metric. This last term emphasises the fact that software metrics should be defined in accordance with the measurement theory. However, today both terms continue to be used.

Star Schema: A relational schema whose design represents a multidimensional data model. A star schema consists of one central table and several dimensional The measures of interest are stored in the fact table (e.g., sales, inventory). For each dimension of the multidimensional model, there exists a dimensional table (e.g., product, time) that stores the information about the dimension.

Metrics for the Evaluation of Test–Delivery Systems

Salvatore Valenti
Università Politecnica delle Marche - Ancona - Italy

INTRODUCTION

Most solutions to the problem of delivering course content supporting both student learning and assessment nowadays imply the use of computers, thanks to the continuous advances of information technology. According to Bull (1999), using computers to perform assessment is more contentious than using them to deliver content and to support student learning. In many papers, the terms computer-assisted assessment (CAA) and computer-based assessment (CBA) are often used interchangeably and somewhat inconsistently. The former refers to the use of computers in assessment. The term encompasses the uses of computers to deliver, mark, and analyze assignments or examinations. It also includes the collation and analysis of data gathered from optical mark readers. The latter (that will be used in this paper) addresses the use of computers for the entire process, including assessment delivery and feedback provision (Charman & Elmes, 1998).

A typical CBA system is composed of the following.

- Test-Management System (TMS) - that is, a tool providing the instructor with an easy-to-use interface, the ability to create questions and to assemble them into tests, and the possibility of grading the tests and making some statistical evaluations of the results
- Test-Delivery System (TDS) - that is, a tool for the delivery of tests to the students. The tool may be used to deliver tests using paper and pencil, or a stand-alone computer on a LAN (local area network) or over the Web. The TDS may be augmented with a Web enabler used to deliver the tests over the Internet. In many cases, producers distribute two different versions of the same TDS: one to deliver tests either on single computers or on a LAN and the other to deliver tests over the WWW (World Wide Web). This is the policy adopted, for instance, by Cogent Computing Co. (2004) with CQuest LAN and CQuest Net.

The TMS and TDS modules may be integrated in a single application as, for instance, Perception developed by Question Mark Computing (2004), or may be delivered as separate applications as it occurs for MicroTest and MicroGrade developed by Chariot Software Group (2004).

BACKGROUND

The interest in developing CBA tools has increased in recent years thanks to the potential market of their application. Many commercial products, as well as freeware and shareware tools, are the result of studies and research in this field made by companies and public institutions.

Thus, for instance, 42 quiz software products are referenced by the Soft411 (2004) directory, 23 by the Educational Software (2004) directory, and 8 by Assessment System Co. (2004). Moreover, it must be noted that almost all course management systems (Edutools, 2004) provide facilities for CBA. This noteworthy growth in the market raises the problem of identifying a set of criteria that may be useful to an educational team wishing to select the most appropriate tool for their assessment needs. The literature on guidelines to support the selection of CBA systems seems to be very poor since no other up-to-date papers are available on the Internet apart from the works by the author and his colleagues (Valenti, Cucchiarelli, & Panti, 2002a, 2002b).

The purpose of this paper is to provide a framework for the evaluation of a test-delivery system.

METRICS FOR THE EVALUATION OF A TDS

Three main functional modules roughly compose a TDS: a student interface, a question-management unit, and a test-delivery unit. Therefore, our framework for the evaluation of a TDS is defined in terms of criteria that may support the evaluation of each functional module and other criteria for the evaluation of the whole system, as shown in Table 1.

The evaluation of the interface is a qualifying aspect for the evaluation of a CBA system and obviously for a TDS. This becomes dramatically true if we take into

Table 1. Metrics for the evaluation of a TDS

Issue		Metrics
Component Level	Interface	❄ Friendly GUI (graphical user interface)
	Question Management	❄ Types of Questions ❄ Question Structure (retries, tutorial building)
	Test Management	❄ Help and Hints ❄ Restricted Availability ❄ Grading
System Level		❄ Security ❄ Survivability ❄ Communication

account the fact that neither the teacher nor the students involved in the use of a TDS necessarily have a degree in computer science, nor may be interested in acquiring skills in this field. According to Nielsen and Molich (1990), the interface must be easy to learn, efficient to use, easy to remember, error free, and subjectively pleasing. Some further criteria that may be adopted to evaluate the usability of the interface are summarized in the following list.

- speak the users' language (multilinguality and multiculturality)
- be accessible
- provide feedback
- provide clearly marked exit points

The question-management unit of a TDS can be evaluated with respect to its ability to provide

- multiple attempts at solving a question (retries),
- feedback and tutorials on the topic covered by the questions, and
- capabilities for the inclusion of multimedia in questions.

The ability of providing retries may be of great importance for self-assessment since it is useful to improve the knowledge of the student whilst reducing the need for providing feedback and/or tutoring. On the other hand, the impossibility to change the answer to a question during an examination is often perceived as unfair by the students (Valenti et al., 2002b). It is worth outlining that allowing multiple attempts at question answering may affect the use of adaptive systems whenever item presentation depends on previous responses.

The feedback may be provided after each question (this solution being preferable for self-assessment), after a set of questions covering a given topic, or at the end of the test, and can be based on the overall performance. Furthermore, the feedback may be used to indicate the correctness of the answer, to correct misconceptions, or to deliver additional material for deepening and/or broadening the coverage of the topic assessed by the question. Tutorials represent an extended approach to provide additional information to the students. The existence of some facility to ease inclusion of tutorials in the TDS represents an important feedback aid. As an example, Perception provides explanation-type questions that may be used for "information screens, title pages, or to display large bodies of text" (Question Mark Computing Ltd., 2004).

The use of questions incorporating multimedia, such as sound and video clips or images, may improve the level of knowledge evaluation. This aspect may be of great importance, for example, in language assessment, where the comprehension of a talk or a movie can be assessed by recurring to multimedia only. The use of multimedia can raise issues related to portability and interoperability since it may require special hardware and software, both for the server delivering the questions and for the client used by the students. Furthermore, it may raise the costs for the adopted solution. These issues may not represent a problem whenever a Web-enabled TDS is selected since the nature of the World Wide Web is inherently multimedial. In this case, the choice of standard plug-ins for the most common browsers may reduce risks of portability and of interoperability. Since most plug-ins used to grant access to multimedia sources are usually free of charge, their use may not interfere with cost problems.

Among the issues taken into account to evaluate the test-management unit of a TDS, we have identified the ability to

- provide help and hints,
- make tests available at a given time, and
- allow scoring procedures.

The capability of a TDS to provide directions about the completion of the test and hints that usually are related to the contents of the questions represents a further measure of the ease of use of the application from the student's point of view.

Tests can be made either available or unavailable at a specified date and time. This allows test designers to specify exactly when people can access a test. It should be possible to leave out either or both restrictions to provide maximum flexibility. This lends itself nicely to the computer-lab setting where students are required to complete an online test during a specified time frame on a specified day.

Obviously, any software for assessment should be able to compute student grades. Furthermore, grades must be delivered as feedback to the course coordinator, to the instructor, and to the students. Each of these categories of stakeholders may require a different kind of feedback on the grades associated with a test. For instance, a student needs to know where he or she stands with respect to other students and to the class average besides his or her own individual and cumulative grades. This need raises obvious concerns about privacy that may be faced through the security facilities provided with the assessment tool.

Among the issues taken into account to evaluate a TDS from a systemic point of view, we have selected

- security,
- survivability, and
- communication with other software.

There is a wide range of security issues related to the use of TDSs. In more detail, many concerns on the security of the availability of the test material, on the HTML (hypertext markup language) code that implements testing, and on the access-control privileges do exist. With respect to security concerns about the test material and its HTML code, it must be outlined that while commercial programs usually implement encrypting approaches, a lot of issues are raised by freeware. In fact, most freeware applications rely either on Perl/CGI (common gateway interface) or on JavaScript. In particular, since a CGI program contains an executable code, the use of CGI-based applications is the equivalent of letting the world run a program on the server side, which is not the safest thing to do. Therefore, there are some security precautions that need to be implemented when it comes to using CGI-based applications. The one that will probably affect the typical Web user is the fact that CGI programs need to reside in a special directory so that the server knows to execute the program rather than just display it to the browser. This directory is usually under direct control of the webmaster, prohibiting the average user from creating CGI programs. On the other hand, the assessment program cannot be completely hidden whenever using a JavaScript code that runs on the client side of the application, so a "smart" student can access the source, discovering the right answer associated to each question. Some sophisticated techniques can be used to partially overcome this problem (Cucchiarelli, Panti, & Valenti, 2000).

The ability of a TDS to perform under adverse conditions (i.e., survivability as discussed in Valenti et al., 2002a) is of great importance. In particular, no termination procedures should result in any loss of data. To ensure this, both student and system files should be updated after each transaction so that no data is lost if the test is terminated because of machine or power failure. The possibility of providing examination printouts may further enforce the survivability of the system. Furthermore, after a crash, the system should be able to restart from the point of termination with all aspects of the original status unchanged, including the answers already given and the clock still displaying the time remaining.

Communication with other existing software may be very useful both for exporting answers and for calling external applications. Furthermore, this feature is required to allow the integration of a TDS with a TMS distributed by different producers. Exporting answers is usually performed through test files and data-conversion utilities. This may be useful to customize the reports generated by the application or whenever an in-depth analysis is needed to evaluate the results obtained. Moreover, many available TDSs enable the calling of a program as a block within a question. The called program returns a score in points that may be added to the test score. This may be useful for assessing abilities that cannot be evaluated through the basic question-answer paradigm of most assessment tools.

Finally, some tools allow external applications to be called at the very end of the test phase for printing certificates for all users who pass the test, for the electronic submission of the answer file to a central location for analysis and evaluation, and for the storage of the results in a file to be accessed by a user program (Question Mark Computing Ltd., 2004).

FUTURE TRENDS

The software-evaluation process is a very complicated task because many, often contradictory, attributes have to be taken into account. Issues such as selection of appropriate attributes, creation of their hierarchy, assignment of relative weights to them, and application of a sound decision-aid methodology arise frequently both for nonexperienced and skilled users. Therefore, a first step toward providing support for the evaluation of test-delivery systems could be the construction of a Web site listing all the available software on the market, providing an independently reviewed, objective source of information to help educators in making the best decision for their institution. A further step in this direction could be the implementation of a decision support system to be used as an (semi) automated aid for helping the educator in selecting the most appropriate criteria and constructing the evaluation model.

CONCLUSION

In this article we have discussed a framework that may be useful in assisting an educational team in the selection of a TDS. Three main functional modules roughly compose a TDS: a student interface, a question-management unit, and a test-delivery unit. Therefore, we decided to organize our framework by identifying some metrics to support the evaluation of the functional modules and other metrics to support the evaluation of the system as a whole.

REFERENCES

Assessment System Co. (2004). Retrieved from http://www.assess.com

Bull, J. (1999). Computer-assisted assessment: Impact on higher education institutions. *Educational Technology & Society*, *2*(3).

Chariot Software Group. (2004). Retrieved from http://www.chariot.com/home/index.asp

Charman, D., & Elmes, A. (1998). Computer based assessment: A guide to good practice (Vol. I). SEED Publications, University of Plymouth.

Cucchiarelli, A., Panti, M., & Valenti, S. (2000). Web-based assessment of student learning. In A. K. Aggarwal (Ed.), *Web-based learning and teaching technologies: Opportunities and challenges* (pp. 175-197). Hershey, PA: Idea Group Publishing.

Educational Software. (2004). Retrieved from http://www.educational-software-directory.net

Edutools. (2004). Retrieved from http://www.edutools.info

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. *Proceedings of CHI 90*, 249-256.

Question Mark Computing Ltd. (2004). *Perception*. Retrieved from http://www.questionmark.com/home.htm

Soft411. (2004). Retrieved from http://www.soft411.com

Valenti, S., Cucchiarelli, A., & Panti, M. (2002a). Computer based assessment systems evaluation via the ISO9126 quality model. *Journal of Information Technology Education* , *1*(3), 157-175.

Valenti, S., Cucchiarelli, A., & Panti, M. (2002b). Relevant aspects for test delivery systems evaluation. In M. Khosrow-Pour (Ed.), *Web-based instructional learning* (pp. 203-216). Hershey, PA: IRM Press.

KEY TERMS

Computer-based Assessment: addresses the use of computers for the entire process of assessment including production, delivery, grading, and provision of feedback

CGI: CGI is the acronym for common gateway interface. A CGI program is any program designed to accept and return data that conforms to the CGI specification. CGI programs are the most common way for Web servers to interact dynamically with users. The program could be written in any programming language including C, Perl, Java, or Visual Basic.

HTML: HTML is the acronym for hypertext markup language, the authoring language used to create documents on the World Wide Web. HTML defines the structure and layout of a Web document by using a variety of tags and attributes. HTML is derived from SGML, although it is not a strict subset.

JavaScript: a scripting language developed to enable Web authors to add dynamic content to sites. Although it shares many of the features and structures of the Java language, it was developed independently. It is supported by recent browsers from Netscape and Microsoft.

Multiculturality: the term used to address the measures now being taken to provide graphical user interfaces with ad hoc icons and texts according to the cultural heritage of the user

Multilinguality: the term used to address the measures now being taken to provide graphical user interfaces with features for internationalization, that is, support of the character sets and encodings used to represent the information being manipulated, and presentation of the data meaningfully

Test-Delivery System: a tool for the delivery of tests to students. The tool may be used to deliver tests using a stand-alone computer on a local area network or over the World Wide Web.

Migrating Legacy Systems to the Web

Lerina Aversano
University of Sannio, Italy

Gerardo Canfora
University of Sannio, Italy

Andrea De Lucia
University of Salerno, Italy

INTRODUCTION

Business Process Reengineering (BPR) is defined as "the fundamental rethinking and radical redesign of business processes to achieve significant improvements of the performances, such as cost, quality, service, and speed" (Hammer & Champy, 1993). Most BPR projects aim at converting business organisations from hierarchical centralised structures to networked decentralised business units cooperating with one another. This conversion is assuming a strategic relevance as the Internet is changing radically business processes, not only because they are purposely reengineered, but also because the Internet and the information and communication technology, offer more convenient means of fulfilling their requirement.

Current business processes have been profoundly fitted to the available software. The technologies involved in process execution impact the way businesses are conceived and conducted. Abstractly, BPR should entail discarding the existing and legacy systems to develop new software systems that meet the new business needs. This is superficially attractive and human appealing. However, legacy systems cannot be simply discarded because they are crucial systems to the business they support and encapsulate a great deal of knowledge and expertise of the application domain. This entails that even the development of a new replacement system may have to rely on knowledge which is encapsulated in the old system. In summary, existing systems are the result of large investments and represent a patrimony to be salvaged. Therefore, to satisfy the goals of a BPR project, it is necessary to work intensively to search a trade-off between the constraints of existing legacy systems and the opportune BPR strategy.

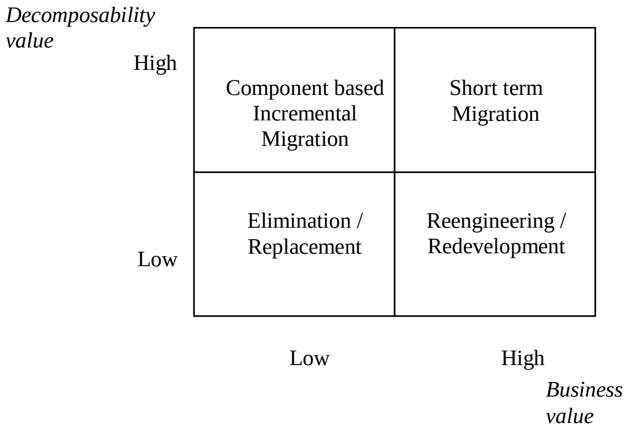
In this article, we discuss a strategy for migrating business processes and the supporting legacy systems to Web-centric architecture. The overall strategy comprises modelling the existing business processes and assessing the business and quality value of the supporting software systems. This is followed by the migration of the legacy systems, which can be in turn enacted with different strategies. The initial step consists of understanding and modelling the business processes together with the involved documents and software systems. The analysis of the existing processes is required to get an inventory of the activity performed, compare them with best practices, and redesign and/or reengineer them. Our overall approach is discussed in details in references (Aversano, Canfora, De Lucia, & Stefanucci, 2002a; Canfora, De Lucia, & Gallucci, 2002b), together with experience concerned with their applications. The final phase related to legacy system analysis and assessment is discussed in details in Aversano, Canfora, and De Lucia (2003) and briefly presented here.

BACKGROUND

Legacy systems are "large software systems that we don't know how to cope with but that are vital to our organization" (Bennett, 1995). There are a number of options available in managing legacy systems. Typical solutions include: discarding the legacy system and building a replacement system; freezing the system and using it as a component of a new larger system; modifying the system to give it another lease of life. Modifications may range from a simplification of the system (reduction of size and complexity) to preventive maintenance (redocumentation, restructuring, and reengineering) or even to extraordinary processes of adaptive maintenance (interface modification, wrapping, and migration) (Pigoski, 1997; De Lucia, Fasolino, & Pompella, 2001).

Several authors have identified possible alternatives for dealing with legacy systems and have proposed decision frameworks to select among the alternatives. In general, decision frameworks require that a legacy system be assessed from two points of views: a business dimension and a technical dimension (Bennett, Ramage, & Munro, 1999; De Lucia et al., 2001; Sneed, 1995). This information measures the complexity of the business processes and administrative rules that a system, or system's component, implements and their relevance to

Figure 1: A decisional framework



achieve business competitiveness. The technical value of a legacy system can be assessed through different quality attributes, such as the obsolescence of the hardware/software platforms, the level of decomposability, the maintainability, and the deterioration (De Lucia et al., 2001).

We assess the technical quality of a legacy system by considering the obsolescence and the decomposability level. In particular, we focus on making decisions on the actions to perform as a consequence of a BPR project aimed at taking advantage of the Internet. We assume that the obsolescence of the system is high, and therefore, extraordinary maintenance is required. Accordingly, the decision about the particular type of intervention to take will be made based on the decomposability and business value of the system.

Two different kinds of decomposability can be considered:

- *vertical decomposability*, which refers to the possibility of decomposing a system into major architectural layers;
- *horizontal decomposability*, which accounts for the possibility of decomposing a legacy system into independent and cooperating business components.

In particular, concerning the vertical decomposability, Brodie and Stonebaker (1995) refer that a software system can be considered as having three types of components: *interface components*, *application logic components*, and *database components*. Depending on how separated and well identified are these components, the architecture of a legacy system can be *decomposable*, *semidecomposable*, or *nondecomposable*. In a decomposable system, the application logic components are independent of each other and interact with the database components and system interfaces. In a semidecomposable system, only interfaces are separate modules, while application logic components and database services are not separated. A nondecomposable system is a black box with no separated components.

Figure 1 shows a decisional framework that takes into account the considerations described previously. The decision about the intervention to take on the legacy system with a high business value is mainly driven by the vertical decomposability of the system. If the vertical decomposability value is sufficiently high, that is, the system is decomposable or semidecomposable in the Brodie and Stonebaker terminology, the best strategy is a short term migration of the system, achieved through wrapping the application logic and/or data management functions (that define the server tier) and reengineering/redeveloping the user interface to a Web-centric style. This strategy represents a good alternative also in the case of a nondecomposable system, provided that the costs and risks of its decomposition are affordable (Canfora, De Lucia, & Di Lucca, 1999; Sneed, 1995). If the decomposability level of a system with high business value is too low, the complete reengineering / redevelopment alternative has to be preferred, as the legacy system can still be used.

For legacy systems with a low business value, the decision about the intervention to take is mainly driven by the horizontal decomposability. In particular, if the horizontal decomposability value is high, it is possible to use again the framework in Figure 1 to make different decisions for the different business components. This provides the basis for a component based incremental migration of the legacy system. Whenever both the business value and the decomposability of a legacy system are low, the only possible option is elimination/replacement. Indeed, in this case, there are no adequate business preconditions to evolve the existing system.

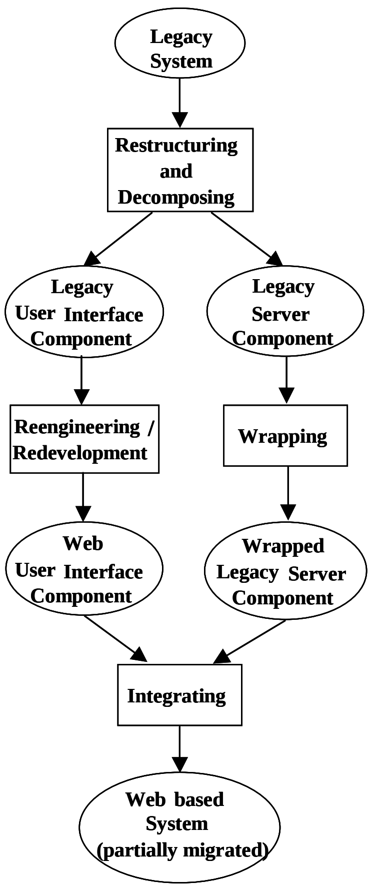
The migration of a legacy system entails the reuse of the system components while moving the system toward newer and more modern technology infrastructure. Brodie and Stonebraker (1995) propose an incremental approach, named Chicken Little, to migrate a legacy information system using gateways. A different approach proposed by Wu et al. (1997) is the Butterfly methodology, which eliminates the needs of accessing both the legacy and new database during the migration process, thus avoiding the complexity induced by the introduction of gateways to maintain the consistency of the data. Both the methodologies aim at migrating a legacy system mainly based on its vertical decomposability.

Migration strategies have also been proposed that take into account the horizontal decomposability. Canfora et al. (1999) and Serrano, Montes de Oca, and Carter (1999) present two strategies for incrementally migrating legacy systems to object-oriented platforms. The main difference is in the method adopted for identifying objects in procedural code. Both strategies use wrapping technologies to enable the coexistence of legacy and newly developed components. According to Mowbray and Zahari (1994), wrapping is the practice for implementing a software architecture using pre-existing components. Using object wrappers, the legacy code is encapsulated in a controllable way. The wrapper implements the interface through which newly developed objects access legacy resources; it is in charge of mapping incoming messages onto calls to the programs that implement the object methods. In the literature different approaches for wrapping have been presented (Arranga & Coyle, 1996; Canfora et al., 1999). The Web offers several other powerful mechanisms for wrapping legacy systems in a distributed environment. For example, XML is widely used by developers as a standard for data representation of the component's interfaces, message exchanging, and for defining scripting languages (Sneed, 2000).

MIGRATION STRATEGY

In this section, we present a Web-centric short term migration strategy based on vertical decomposability and on the use of wrappers. Full details of this migration strategy can be found in Aversano et al. (2003). This strategy applies to legacy systems with high business value and a high vertical decomposability level. Figure 2 shows the main phases of the migration process.

Figure 2: Short-term migration strategy



The first phase is the decomposition and restructuring of the legacy system to a client server style, where the user interface (client side) controls the execution of the application logic and database components (server side). The complexity of this phase depends on the decomposability of the system. Nondecomposable systems are the most difficult to be restructured, because decoupling the user interface from the application logic and database components can be very challenging and risky. Canfora, Cimitile, De Lucia, and Di Lucca (2000) present a method to identify the code implementing the user interface of these types of legacy systems. However, even migrating semidecomposable or decomposable systems might not be trivial. For semidecomposable or decomposable systems, the restructuring actions to be taken depend on the coupling between the user interface components and the other components. If all the data are exchanged through the calls between the client side and the application logic components, there is no need for restructuring. However, if the client and server components exchange some data through global data areas, data flow analysis and restructuring techniques are required to transform the global data into parameters to be exchanged across calls.

Once the legacy system has been restructured to a client server style, if needed, two phases can be conducted in parallel, namely the user interface reengineering and the wrapping of the server tier. The first phase aims at reengineering/redeveloping the user interface of the system by exploiting Web technologies. User interface reengineering entails using reverse engineering (Chikofsky & Cross, 1990) to abstract a user interface conceptual model and forward engineering to re-implement the user interface using Web technologies (Merlo et al., 1995; Moore & Rugaber, 1997). In particular MORPH (Moore & Rugaber, 1997) is a knowledge based approach that exploits transformation rules to restructure and transform the abstract model of the user interface into a new abstract model used to generate the target graphical user interface. We have used the MORPH guidelines to reengineer graphical user interfaces to Web interfaces (Aversano et al., 2003).

The transformation of the legacy user interface into the new Web interface should also be driven by the business needs. For example, if the new interface has to be used by the old users, a goal could be minimising the needs for re-training: in this case, in addition to establishing a one-to-one correspondence between the legacy and the new interaction objects, the mapping should be built maintaining a correspondence between the legacy panels and the new Web pages and forms (Aversano et al., 2003). However, in many cases the BPR process introduces new roles (for example, a legacy system in the reengineered process might be required to be accessed directly by a customer); in this case, the new interface must be radically redesigned to meet the new user needs (Sneed, 2000).

Wrapping the server tier is the key to reuse a legacy system with a high business value. In the last few years, the diffusion of the wrapping technology has had a significant impulse as a consequence of the use of Internet as a means of communication between remote programs. The final step of the migration strategy is the integration and testing of the new Web-based application. It is worth noting that once the Web-based migration of the legacy system is completed, the management of the new system will be based on a more general decision framework than the one in Figure 1 (see, e.g., De Lucia et al., 2001). In particular, the main factor to consider will be the business value, at least until the technical value of the system will be adequate to the evolving business needs. In other words, the migrated system will undergo periodical phases of evolution, to keep its functionality in line with the evolving business goals, interleaved with the ordinary maintenance of the system. Of course, a major shift in the technological business needs will entail a rapid increase of the obsolescence of the system, thus moving the management of the system to the decision framework of Figure 1.

The migration strategy has been applied in an industrial pilot project (Aversano et al., 2003). The project aimed at integrating into a Web-enabled architecture an existing online analytical processing COBOL system used by many banks for querying large archives of data for decision support through statistical and explorative analyses. The system has evolved over the past 15 years from a centralised mainframe version with a character-based user interface to a semi-decomposable client-server multiversion system composed of a graphical user interface running on a PC and several server components running on different platforms and accessed from the PC through a PC-host connection layer. In addition, a stand-alone version of the system working on the PC was available.

The need for migrating the system was imposed by the changes in the underlying business processes. The first and most important reason was the fact that the system was used by several banks to make decisions both at central and peripheral levels. In the old architecture, the PC-host connection layer had to be replicated on each client installation and this increased the application ownership costs. Most banks nowadays use Internet/intranet technologies for their information system infrastructure and this caused a push in the direction of Web migration. In addition, the new architecture adds flexibility; in particular, it opens the way to the migration of the stand alone PC version towards a distributed version which federates several archives/database host sources, thus eliminating the need for maintaining the different versions of the querying and analysis software.

After assessing the system, we opted for a short-term migration strategy with the aim of reducing, at a minimum, the time needed to have a working version of the Web-enabled system. Besides the graphical user interface component, the migrated part of the system consisted of 218 COBOL programs and more than 130 KLOC. The migration was achieved essentially through an extensive use of reverse engineering, to design the new user-interface, and wrapping, to implement the server side. In particular, the presentation layer of the original system was separated from the application logic and database components and re-implemented.

FUTURE TRENDS

Currently, highly dynamic and agile organisations models are emerging in order to compete in global marketplace. This poses significantly new problems for software development and maintenance, and software services are being promoted as the next big step forward in software engineering. A key issue is that services are something that the organizations use, not own. As a consequence, maintainability is not primarily a technical problem. Non-functional properties need to be incorporated into service models so that business drivers, as well as technical issues, can be addressed in maintaining and evolving software systems. Two key issues must be considered. The first is the nature of the service supply chain, enterprises tend to express their needs in models in which consumers and suppliers define requirements together. Real situations are more complex: consumers influence the type and method of supply and the supplier influences a consumer's business processes in order to adapt to their product.

The second issue is that the use of Web services requires a service-level agreement defining terms and conditions for use and maintenance. A critical success factor is the need for a negotiated and agreed architecture by which services can be used and maintained with minimum change.

CONCLUSION

This article has discussed a strategy for migrating business processes and the supporting legacy systems toward Web-centric architecture. The initial step consists of modelling the existing business processes and assessing the business and technical value of the supporting software systems. The analysis of the processes is required to get an inventory of the activity performed, and redesign and/or reengineer them. Based of the assessment, a decisional framework assists software managers to make informed decisions. This is followed by the migration of the legacy system, which can be in turn enacted with different approaches. In particular, the short-term system migration strategy decomposes and reengineers the system into its client and server components and uses wrappers to encapsulate the server components. Reverse engineering is used to abstract a model of the client components, and in particular the user interface, and redevelop them.

The short-term strategy was successful in reducing the time and the costs of migrating the system to adapt it to the new business scenario, however it did not change the quality of the server programs, which remain essentially unchanged, and therefore no benefits are expected in the future maintenance and evolution of the system. This is in contrast with other approaches that reverse engineer meaningful objects from the legacy code and wrap each of them separately. While more costly, such approaches offer more benefits on the long term as they enable incremental replacement strategies.

REFERENCES

Arranga, E., & Coyle, F. (1996). *Object oriented COBOL*. New York: SIGS Books.

Aversano, L., Canfora, G., & De Lucia, A. (2003). Migrating legacy system to the Web: A business process reengineering oriented approach. In M. Polo (Ed.), *Advances in software maintenance management: Technologies and solutions* (pp.151-181). Hershey, PA: Idea Group Publishing.

Aversano, L., Canfora, G., De Lucia, A., & Stefanucci, S. (2002a). Automating the management of software maintenance workflows in a large software enterprise: A case study. *Journal of Software Maintenance and Evolution: Research and Practice*, *14*(4), 229-255.

Aversano, L., Canfora, G., De Lucia, A., & Gallucci, P. (2002b). Business process reengineering and workflow automation: a technology transfer experience. *The Journal of Systems and Software*, *63*(1), 29-44.

Bennett, K.H. (1995). Legacy systems: Coping with success. *IEEE Software*, *12*(1), 19-23.

Bennett, K.H., Ramage, M., & Munro, M. (1999). Decision model for legacy systems. *IEE Proceedings Software*, *146*(3), 153-159.

Brodie, M.L., & Stonebaker, M. (1995). *Migrating legacy systems - Gateways, interfaces & incremental approach*. San Francisco, CA: Morgan Kaufmann Publishers, Inc.

Canfora, G., De Lucia, A., & Di Lucca, G.A. (1999). An incremental object-oriented migration strategy for RPG legacy systems. *International Journal of Software Engineering and Knowledge Engineering*, *9*(1), 5-25.

Canfora, G., Cimitile, A., De Lucia, A., & Di Lucca, G.A. (2000). Decomposing legacy programs: a first step towards migrating to client-server platforms. *The Journal of Systems and Software*, *54*(2), 99-110.

Chikofsky, E.J., & Cross, II, J.H. (1990). Reverse engineering and design recovery: A taxonomy. *IEEE Software*, *7*(1), 13-17.

De Lucia, A., Fasolino, A.R., & Pompella, E. (2001). A decisional framework for legacy system management. *Proceedings of the International Conference on Software Maintenance*, Florence, Italy: IEEE Computer Society (pp.642-651).

Hammer, M., & Champy, J. (1993). *Reengineering the corporation: A manifesto for business revolution*, NY: HarperCollins.

Merlo, E., Gagne, P.Y., Girard, J.F., Kontogiannis, K., Hendren, L., Panangaden, P., & De Mori, R. (1995). Reengineering user interfaces. *IEEE Software*, *12*(1), 64-73.

Moore, M., & Rugaber, S. (1997). Using knowledge representation to understand interactive systems. *Proceedings of 5th International Workshop on Program Comprehension*, Dearborn, MI, 60-67.

Mowbray, T., & Zahari, R. (1994). *The essential CORBA*, John Wiley & Sons, New York.

Pigoski, T.M. (1997). *Practical software maintenance - best practices for managing your software investment*. John Wiley & Sons, New York, NY.

Serrano, M.A., Montes de Oca, C., & Carter, D.L. (1999). Evolutionary migration of legacy systems to an object-based distributed environment. *Proceedings of International Conference on Software Maintenance*, Oxford, UK, IEEE Computer Society (pp.86-95).

Sneed, H.M. (1995). Planning the reengineering of legacy systems. *IEEE Software*, *12*(1), 24-34.

Sneed, H.M. (2000). Business reengineering in the age of the Internet. Technical Report, Case Consult, Wiesbaden, Germany.

Wu, B., Lawless, D., Bisbal, J., Wade, V., Grimson, J., Richardson, R., & O'Sullivan, D. (1997). The butterfly methodology: a gateway-free approach for migrating legacy information systems. *Proceedings of 3rd IEEE International Conference on Engineering of Complex Computer Systems*, Como, Italy, IEEE Computer Society (pp.200-205).

KEY TERMS

Business Process Reengineering (BPR): The fundamental rethinking and radical redesign of business processes to achieve significant improvements of the performances, such as cost, quality, service, and speed.

Legacy System: A software system which continues to be used because of the cost of replacing or redesigning it, despite its poor competitiveness and compatibility with modern equivalents.

Reengineering: The examination and alteration of a subject system to reconstitute it in a new form and the subsequent implementation of the new form.

Restructuring: The transformation from one representation form to another at the same relative abstraction level, while preserving the subject system's external behaviour.

Reverse Engineering: The process of analysing a subject system to (i) identify system components and their interrelationships and (ii) create representations of the system in another form or at a higher level of abstraction.

System Decomposability: The capability to decompose the system into a set of loosely coupled subsystems.

System Encapsulation: The definition and implementation of a software interface, called wrapper, that allows the access to the system, and its subsystems, from other applications.

System Migration: The incremental transformation of a legacy system toward newer and more modern technology. It entails the reuse of system components through their encapsulation.

Minorities and the Digital Divide

Lynette Kvasny
The Pennsylvania State University, USA

Fay Cobb Payton
North Carolina State University, USA

INTRODUCTION

Information and communication technologies (ICT) such as the World Wide Web, e-mail, and computers have become an integral part of America's entertainment, communication, and information culture. Since the mid-1990s, ICT has become prevalent in middle- and upper-class American households. Companies and government agencies are increasingly offering products, services, and information online. Educational institutions are integrating ICT in their curriculum and are offering courses from a distance.

However, while some are advantaged by the efficiencies and convenience that result from these innovations, others may unwittingly become further marginalized by these same innovations since ICT access is not spreading to them as quickly. The "digital divide" is the term used to describe this emerging disparity. Government analysts argue that historically underserved groups, such as racial and ethnic minorities, rural and low-income communities, and older Americans, are at a distinct disadvantage if this divide is not closed because American economic and social life is increasingly becoming networked through the Internet (National Telecommunications and Information Administration, 1995). The digital divide is not only an American social problem. Digital divide issues are of concern in developing countries as well as in information technology marginalized communities within developed nations.

Over the last decade, access to ICT has increased for most Americans, but does this mean that the problem of the digital divide has been solved? Is further research in this area warranted or has the digital divide become passé? In this article, we take on these questions by first reviewing major issues and trends in digital divide research. We do so by reviewing the digital divide literature as it relates to one historically underserved group, namely African-Americans. Next, we present a conceptual framework that contrasts 1) social and technological access perspectives, and 2) asset-based/resource and behavioral/use perspectives. The paper concludes with our recommendations for future research opportunities for examining digital divide issues.

BACKGROUND

There have been numerous definitions for the digital divide, government and industry reports about the digital divide, and competing interpretations of the statistics contained in these reports. For instance, the digital divide has been defined at the Whatis Web site as "the fact that the world can be divided into people who do and people who don't have access to — and the capability to use — modern information technology, such as the telephone, television, or the Internet." Others (PRNewswire, 2000) offer another definition: "arguably the single, largest, and segregating force in today's world. If it is not made a national priority, a generation of children and families will mature without these tools that are proving to be the key to the future."

Most of our knowledge about the digital divide in the U.S. is based on survey research on computer and Internet access in the home, at work, and in public places. The most cited statistics are found in the digital divide series produced by the U.S. Department of Commerce (National Telecommunications and Information Association, 1998, 1999, 2000, 2002). These studies have found that the divide cuts along the lines of ethnicity and race, geographic location, household composition, age, education, and income level. These and other studies have also documented that these gaps are persistent but closing (Goslee & Conte, 1998; Hoffman & Novak, 1998; Spooner & Rainie, 2001).

When we look at any particular demographic group, however, the analysis is much more complex and contradictory. For example, most research on the digital divide for African-Americans has centered on physical access to computers and Internet in the home, as well at technical skills to operate computers and information literacy skills to engage with content. These researchers have found that African-Americans are less likely to have ICT access and skills, even when controlling for other factors, such as income and education (Mossberger & Tolbert, 2003). The Pew Internet and American Life Project suggests that these gaps are closing, but African-Americans with access to the Internet do not go online as often on a typical day as whites do (Spooner & Rainie, 2000).

Blacks also tended to use ICT differently than other racial and ethnic groups. Novak, Hoffman, & Venkatesh (1997) summarize previous research on African-Americans with regard to different media as follows:

African-Americans have the highest participation in radio and TV and the lowest participation in newspapers. In terms of our classification, it means that historically, they have participated in greater measure in entertainment-oriented technologies rather than in information oriented technologies. Previous studies have also shown that African-American ownership of telephones is lower than white ownership, which may be due in part to differences in income.

They go on to theorize that culture helps to explain these results. African-Americans have found their social expression historically through the arts, and have been less successful in gaining entry to other dominant domains such as business, education, technical employment, and professional occupations. Culture may also help to explain Spooner & Rainie's (2000) observation that online African-Americans are 69% more likely than online whites to have listened to music on the Web, and are 65% more likely than online whites to have sought religious information on the Web. Music and spirituality has traditionally been integral components of African-American culture.

Although African-Americans may use ICT relatively less than other ethnic groups, they have more positive attitudes toward ICT than do similarly situated whites (Mossberger & Tolbert, 2003). Kvasny (2003) found that working class African-American women believed that ICT skills would prepare them for higher paying jobs, and improved their parenting abilities. In a study of ICT adoption in community technology project, Youtie et al. (2002) found that African-American women were among the highest adopters of cable TV-based Internet devices.

Although African-Americans harbored favorable attitudes towards ICT, these same technologies may have little impact on social inclusion. In a more recent study, Sipior, Ward, Volonino, & Marzec (2004) examined the digital divide in a public housing community in Delaware County, Pennsylvania. With 31 African-American participants with ages ranging from 13-65, these researchers concluded that effective community-based programs could help reduce the divide. While these interventions notably have improved computing skills about underserved groups, a one-time shot fails to eliminate or even reduce broader feelings of cultural isolation and economic deprivation among minority groups.

MAJOR ISSUES

When analyzing the digital divide literature, one of the foremost issues is whether a gap still exists. For example, in an article titled "True Nature of the 'Digital Divide' Divides Experts" (Jerding, 2000b), four technology watchers provided radically different standpoints. Mark Lloyd, an executive of the Civil Rights Forum on Communications Policy, states that technology inequality is the latest in a history of economic gaps. In his view, although private enterprise has put forth altruistic efforts, "real government action" was needed to bridge this void. Rick Weingarten, the director of the Office for Information Technology Policy, states that being connected wouldn't solve the problem. What is really at stake is the quality of access such as high-speed access and complex information literacy skills. Professor Jorge Schement believes that the digital divide will persist until Americans can put a face on the problem. So long as this is seen as a problem of the "Other," it can be more easily ignored and rationalized. The final panelist, Adam Clayton Powell II, denies the existence of a digital divide. Using the National Telecommunications and Information Administration studies, he argued that the gap between ethnic groups has dissipated. For him, the digital divide is largely a myth, and education rather than race or ethnicity was the highest barrier to technological access and effective use.

We contend that these debates about the existence of the digital divide result from a rather narrow treatment of a complex social phenomenon. In fact, many of the newer studies in this genre call for a rethinking of the digital divide (Warschauer, 2002; Gurstein, 2003; Hacker & Mason, 2003; Kvasny, 2003; Payton 2003). In what follows, we organize a discussion of the current trends in the digital divide discourse. We do so through a framework (*Table 1*) that contrasts two perspectives of access (technological and social) and two perspectives of use (asset-based and behavioral). Technological access focuses on the computing artifact, while social access focuses on know-how and competence. Asset-based perspectives view the divide as a deficiency in requisite resources, such as income or education, that enable ICT use, while behavioral perspectives tend to focus on the effectiveness of ICT use. Although these perspectives are presented as separate categories, authors tend to draw from both categories. For instance, the argument that the digital divide is based upon a lack of access to computing artifacts and computer skills suggests a technological access/asset-based perspective. An argument that the digital divide emerges from a lack of understanding about how to use ICT to further life chances adopts a social/behavioral perspective.

Table 1. Competing perceptions for examining the digital divide

Access Factors	Use Factors
Technological	Asset-based
Social	Behavioral

Technological and Social Access

The technological access view, with its focus on broad statistics on ICT diffusion and use rates, has led some policy analysts to assume that the answer lies in certain characteristics of the technology. Hence, policy solutions tend to employ technological fixes such as wiring public schools and libraries and providing computing resources with Internet access in poorer communities (Norris, 2001). We contend that an over reliance on descriptive statistics largely contributes to this technology-centric understanding of the digital divide. The more important question for studying as technological access increases is "what are people able to do with this access?"

We further argue that emphasis on quantitative descriptions of who has and who lacks access fuels debates about the degree to which the divide is temporary or permanent, whether the divide is widening or narrowing, or whether a divide exists at all. We have already seen the initial have/have not thesis superseded with the more complacent have now/have later prediction. Proponents of the have now/have later position argue that given enough time, competition will eventually alleviate any natural disparities in the marketplace.

Digital divide interventions informed by a technological access perspective are likely to subside once the technology gap has been narrowed through various programs and policies designed to distribute these resources more evenly. From this perspective, the digital divide would not warrant long-term policy remedies. High profile, short-term injections of government, foundation, or corporate assistance will occur until such time as the technology diffusion problem is lessened. Then, further critical attention to social inequities that are deeper than descriptions of technology access and use may be stifled. The digital divide will be simply defined away (Kvasny & Truex, 2001). For instance, in 2002 the Bush Administration declared the end of the digital divide and proposed deep cuts to federal digital divide programs. The biggest proposed cut was levied against the Technology Opportunities Program (TOP), a federal grant program designed to bring aid to communities that are lagging in access to digital technologies. Under the Clinton administration's 2001 budget, the program distributed $42 million in grants to 74 different non-profit organizations. In 2002, that number fell to just over $12 million (Benner, 2002) .

Analysts and researchers who adopt the social access perspective critique this shortsightedness, and assert that the technological access paradigm ignores social constraints, such as workforce literacy, income differentials, and the inevitable household tradeoffs required in making a PC purchase. Simply put, Maslow's Needs Hierarchy must be addressed from the most fundamental level if one is to progress to higher-order affiliation. The digital divide reflects not only differences in the structure of access, but also the ways in which historical, economic, social, cultural, political and other non-technological factors make such differences matter. Technology-centric solutions alone will do little to redress these aspects of the digital divide.

Asset-Based and Behavioral Perspectives

As access diffuses to historically underserved groups, use also becomes an important basis for studying the digital divide (DiMaggio & Hargittai, 2001; Patterson & Wilson, 2000; Warschauer, 2002; Gurstein, 2003). From an asset-based perspective, computer and Internet access are insufficient without the requisite skills and competencies to use the technology effectively (Mossberger, Tolbert, & Stansbury, 2003). These authors take historically underserved groups as their subjects, and point out the ways in which their use of ICT may be hampered. For instance, in a study of African-American youths, Payton (2003) found that these youths were all too aware that the digital divide is not merely about Internet access. Rather, it involves access to the social networks that ease the path to success in high-tech careers. Hargittai (2001) introduces the concept of "second level divides" to signify the disparities in computer skills and how these disparities are patterned along age, racial, gender and income categories. She found, for example, that search time was positively correlated with age, and negatively correlated with education and prior experience with technology.

In contrast, the behavioral perspective sees the digital divide in terms of disparities in benefits like social inclusion, economic opportunity and political participation that one derives from ICT use. These disparities provide primary justification for realizing that the digital divide is a public problem and not simply a matter of private misfortune (Warschauer, 2003). Groups that are historically underserved in their quality of employment, level of qualifications, level of income, quality of education, and consumption opportunities tend to also be marginalized in their access to and use of IT. The digital divide, therefore, is a political outcome rooted in these historical systems of power and privilege, and not simply a gap in access to technological artifacts. Promoting access and basic training to improve the computer skills of individuals is warranted, but may do little to redress the social forces that may limit these actions in the first place. From the behavioral perspective, the divide is about disparities in what individuals and groups are able to do with their ICT access. Gurstein (2003) contends that effective use of ICT occurs when people are able to use ICT purposively and independently to improve their life chances in culturally relevant domains such as economics, employment, health, education, housing, recreation, culture, and civic engagement.

FUTURE TRENDS AND CONCLUSION

Despite one's alignment regarding the nature and context (technological or social access; asset-based or behavioral use) of the digital divide, the topic warrants reconceptualization. Holistic approaches and frameworks will assist academic, government and industry leaders to first understand the "Other" and the social conditions under which these groups function. Technology access and skills can equip; however, taken in isolation, they cannot sustain, maintain, or offer access to the social, financial, or educational networks needed for empowerment of the "total person."

Thus, as digital divide research matures, there is a continued need to better understand the social dimensions of access. Kling (2000) contends that we do not have a good understanding of the ways that social access to the Internet is effectively supported for ordinary people at home and in public service agencies, such as schools and libraries. This is a topic that merits significant inquiry, since a large body of research points to its importance. He also argues that it is important to examine the specific kinds of networked services that will actually be of value to ordinary people. This goes beyond the current survey research and laboratory studies that contrast the use of ICT by various demographic groups. We need to understand why groups use ICT in the manner that they do, be appreciative of the culture relevance, and study how people use ICT in various contexts such as work, home, churches and public access facilities.

REFERENCES

Benner, J. (2002, February 2). Bush plan "Digital Distortion." Wired News. Retrieved from www.wired.com/news/politics/0,1283,50279,00.html.

DiMaggio, P.J., & Hargittai, E. (2001). From "digital divide" to "digital inequality:" Studying Internet use as penetration increases. Sociology Department, Princeton University. Retrieved from www.princeton.edu/~eszter/research.

Goslee, S., & Conte, C. (1998). Losing ground bit by bit: Low-income communities in the information age. Washington D.C.: Benton Foundation. Retrieved from www2.ctcnet.org/ctc/benton/low-income.pdf.

Gurstein, M. (2003). Effective use: A community informatics strategy beyond the digital divide. *First Monday*, *8* (12). Retrieved February 2004, from www.firstmonday.org/issues/issue8_12/gurstein/.

Hacker, K., & Mason, S. (2003). Ethical gaps in studies of the digital divide. *Ethics and Information and Technology*, *5* (2), 99-115.

Hargittai, E. (2002) Second-level digital divide: Differences in people's online skills. *First Monday*, *7* (4). Retrieved from www.firstmonday.dk/issues7_4/hargittai/index.html.

Hoffman, D.L. & Novak, T.P. (1998). Bridging the racial divide on the Internet. *Science*, *280* (5362), 390-391.

Jerding, C. M. (2000, March 17). True nature of "digital divide" divides experts. *The Freedom Forum Online*. Retrieved from www.freedomforum.org/news/2000/03/2000-02-17-06.asp.

Kling, R. (1998). *Technological and social access to computing, information and communication technologies*. White Paper for Presidential Advisory Committee on High-Performance Computing and Communications, Information Technology, and the Next Generation Internet. Retrieved from www.ccic.gov/ac/whitepapers.html.

Kvasny, L. (2003). Liberation or domination: Understanding the digital divide from the standpoint of the "other." In K. Mehdi (ed.), *Proceedings of the Information Resources Management Association (IRMA) Conference*.Hershey, PA: Idea Publishing Group.

Kvasny, L., & Truex, D. (2001). Defining away the digital divide: A content analysis of institutional influences on popular representations of technology. In N. Russo, & J. DeGross (eds.), *Realigning Research and Practice in Information Systems Development: The Social and Organizational Perspective*. Boston: Kluwer Academic Publishers.

Mossenberg, K., & Tolbert, C. (2003). Race, place, and information technology. In *Proceedings of the Telecommunications Policy Research Conference*. Arlington, VA. Retrieved from intel.si.umich.edu/tprc/papers/2003/184/raceplace4.pdf.

Mossberger, K., Tolbert, C., & Stansbury, K. (2003). *Virtual inequality: Beyond the digital divide*. Washington, DC: Georgetown University Press.

National Telecommunications and Information Administration. (1995). Falling through the Net: A survey of the "have nots" in rural and urban America. Washington, DC: U.S. Department of Commerce. Retrieved from www.ntia.doc.gov/ntiahome/fallingthru.html.

National Telecommunications and Information Administration. (1998). Falling through the Net II: New data on the digital divide. Washington, DC: U.S. Department of Commerce. Retrieved from www.ntia.doc.gov/ntiahome/net2/falling.html.

National Telecommunications and Information Administration. (1999). Falling through the Net: Defining the digital divide. Washington, DC: U.S. Department of Commerce.Retrieved from www.ntia.doc.gov./ntiahome/fttn99/contents.html.

National Telecommunications and Information Administration. (2000). Falling through the Net: Toward digital inclusion. Washington, DC: U.S. Department of Commerce. Retrieved from www.ntia.doc.gov/ntiahome/fttn00/contents00.html.

Norris, P. (2001). *Digital divide: Civic engagement, information poverty and the Internet worldwide*. Cambridge: Cambridge University Press.

Novak, T.P., Hoffman, D.L., & Venkatesh, A. (1997, October). *Diversity on the Internet: The relationship of race to access and usage*. Paper presented at the Forum on Diversity and the Media, Aspen Institute. Retrieved from http://elab.vanderbilt.edu/research/papers/html/manuscripts/aspen/diversity.on.the.internet.oct24.1997.html.

Payton, F.C. (2003). Rethinking the digital divide. *Communications of the ACM*, *46* (6), 89-91.

Sipior, J.C., Ward, B.T., Volonino, L., & Marzec, J.Z. (2004). A community initiative that diminished the digital divide. *Communications of the AIS*, *13*, 29-56.

Spooner, T., & Rainie, L. (2000). *African-Americans and the Internet*. Washington, DC: Pew Internet & American Life Project. Retrieved from www.pewinternet.org/report/pdfs/PIP_African_Americans_Report.pdf.

Warschauer, M. (2002). Reconceptualizing the digital divide. *First Monday*, *7* (4). Retrieved from www.firstmonday.org/issues/issue7_7/warschauer/index.html.

Youtie, J., Shapira, P., Brice, K., Laudeman, G., Young, C., Oh, E., & DiMinin, A. (2002). Who uses LaGrange's public Internet system? Results of the LaGrange Internet access survey. Atlanta: Georgia Institute of Technology.

KEY TERMS

Content: The various genres of information available on the Internet. For instance, local content is information that is specific to a community, neighborhood, or area, such as businesses, housing, neighborhood services, and recreation activities. Community content is information about the neighborhood that promotes community development and facilitates community building. Examples include a listing of places where GED courses are offered, or a newsletter. Culturally relevant content is information that is significant to people with different cultural backgrounds.

Digital Divide: Refers to the gap that exists between these who have and those who do not have access to technology (telephones, computers, Internet access) and related services.

Effective Use: The capacity and opportunity to integrate information and communication technology into the accomplishment of self or collaboratively identified goals. What is most important is not so much the physical availability of computers and the Internet but rather people's ability to make use of those technologies to engage in meaningful social practices.

Historically Underserved Groups: Refers to those who lack access to computers and the Internet. Historically this has included Americans who have low incomes, live in rural communities, have limited education, and are members of racial or ethnic minorities.

Social Access: Refers to a mix of professional knowledge, economic resources, and technical skills to use technologies in ways that enhance professional practices and social life.

Social Inclusion: Refers to the extent that individuals, families, and communities are able to fully participate in society and control their own destinies, taking into account a variety of factors related to economic resources, employment, health, education, housing, recreation, culture, and civic engagement.

Technological Access: Refers to the physical availability of suitable information and communication technologies, including computers of adequate speed and equipped with appropriate software for a given activity.

Mobile Agent Authentication and Authorization in E-Commerce

Sheng-Uei Guan
National University of Singapore, Singapore

INTRODUCTION

With the increasing worldwide usage of the Internet, electronic commerce (e-commerce) has been catching on fast in a lot of businesses. As e-commerce booms, there comes a demand for a better system to manage and carry out transactions. This has led to the development of agent-based e-commerce. In this new approach, agents are employed on behalf of users to carry out various e-commerce activities.

Although the tradeoff of employing mobile agents is still a contentious topic (Milojicic, 1999), using mobile agents in e-commerce attracts much research effort, as it may improve the potential of their applications in e-commerce. One advantage of using agents is that communication cost can be reduced. Agents traveling and transferring only the necessary information save the bandwidth and reduce the chances of network clogging. Also, users can let their agents travel asynchronously to their destinations and collect information or execute other applications while they can disconnect from the network (Wong, 1999).

Although agent-based technology offers such advantages, the major factor that is holding people back from employing agents is still the security issues involved. On the one hand, hosts cannot trust incoming agents belonging to unknown owners, because malicious agents may launch attacks on the hosts and other agents. On the other hand, agents may also have concerns on the reliability of hosts and will be reluctant to expose their secrets to distrustful hosts.

To build bilateral trust in an e-commerce environment, the authorization and authentication schemes for mobile agents should be well designed. Authentication checks the credentials of an agent before processing the agent's requests. If the agent is found to be suspicious, the host may decide to deny its service requests. Authorization refers to the permissions granted for the agent to access whichever resource it requested.

In our previous work, we have proposed a SAFER (Secure Agent Fabrication, Evolution & Roaming) architecture (Zhu, 2000), which aims to construct an open, dynamic and evolutionary agent system for e-commerce. We have already elaborated agent fabrication, evolution, and roaming in Guan (1999, 2001, 2002), Wang (2001), and Zhu (2001). This article gives an overview of the authentication and authorization issues on the basis of the SAFER architecture.

BACKGROUND

Many intelligent agent-based systems have been designed to support various aspects of e-commerce applications in recent years, for example: Kasbah (Chavez, 1998), Minnesota AGent Marketplace Architecture (MAGMA) (Tsvetovatyy, 1997), and MAgNet (Dasgupta, 1999). Unfortunately, most current agent-based systems such as Kasbah and MAGMA are serving only stationary agents. Although MAgNet employs mobile agents, it does not consider security issues in its architecture.

D'Agents (Gray, 1998) is a mobile agent system, which employs the PKI for authentication purposes, and uses the RSA (Rivest, Shamir, & Adleman, 1978) public key cryptography (Rivest et al., 1978) to generate the public-private key pair. After the identity of an agent is determined, the system decides what access rights to assign to the agent and sets up the appropriate execution environment for the agent.

IBM Aglets (Lange, 1998; Ono, 2002) are Java-based mobile agents. Each aglet has a globally unique name and a travel itinerary (wherein various places are defined as context in IBM Aglets). The context owner is responsible for keeping the underlying operating system secure, mainly protecting it from malicious aglets. Therefore, he or she will authenticate the aglet and restrict the aglet under the context's security policy.

Ajanta is also a Java-based mobile agent system (Karnik, 1999, 2001, 2002) employing a challenge-response based authentication protocol. Each entity in Ajanta registers its public key with Ajanta's name service. A client has to be authenticated by obtaining a ticket from the server. The Ajanta Security Manager grants agents permissions to resources based on an access control list, which is created using users' Uniform Resource Names (URNs).

iJADE (intelligent Java Agent Development Environment) (Lee, 2002) provides an intelligent agent-based platform in the e-commerce environment. This system can provide fully automatic, mobile and reliable user authentication.

Under the public key infrastructure (PKI), each entity may possess a public-private key pair. The public key is known to all, while the private key is only known to the key owner. Information encrypted with the public key can only be decrypted with the corresponding private key. In the same note, information signed by the private key can only be verified with the corresponding public key (Rivest, 1978; Simonds, 1996). The default algorithm that generates the key pairs is the digital signature algorithm (DSA), working in the same way as a signature on a contract. The signature is unique, so that the other party can be sure that you are the only person who can produce it.

MAIN THRUST OF THE ARTICLE

This article presents an overview of the architecture based on SAFER (Secure Agent Fabrication, Evolution & Roaming) (Zhu, 2000) to ensure a proper authentication and authorization of agent. Here, the public key infrastructure (PKI) is used as the underlying cryptographic scheme. Also, agents can authenticate the hosts to make sure that they are not heading to a wrong place. According to the level of authentication that the incoming agent has passed, the agent will be categorized and associated with a relevant security policy during the authorization phase. The corresponding security policy will be enforced on the agent to restrict its operations at the host. The prototype has been implemented with Java.

Design of Agent Authentication and Authorization

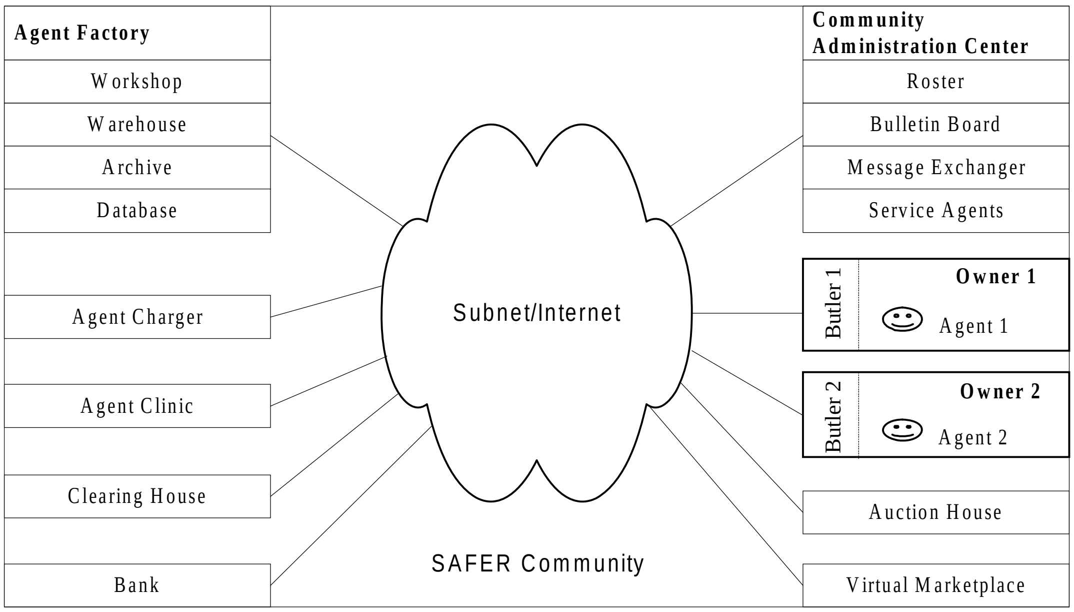
Overview of the SAFER Architecture

The SAFER architecture comprises various communities and each community consists of the following components (see Figure 1): Agent Owner, Agent Factory, Agent Butler, Community Administration Center, and so forth. The Agent Owner is the initiator in the SAFER environment, and requests the Agent Factory to fabricate the agents it requires. The Agent Butler is a representative of the Agent Owner authorized by the owner to coordinate the agents that are dispatched. Owner can go offline after dispatching his or her agents, and thereafter the butler can take over the coordination of the agents. The Agent Factory fabricates all the agents. This is the birthplace of agents and is thus considered a good source to check malicious agents. The Community Administration Center (CAC) is the administrative body, which has a roster that keeps the data of the agents that are in the community. It also collects information, such as addresses of new sites that agents can roam to.

Agent Structure and Cryptographic Schemes

In SAFER, mobile agents have a uniform structure. The agent credentials (hard-coded into the agent (Guan, 2000, 2001)) are the most important part of the agent body and are immutable. This part includes FactoryID, AgentID,

Figure 1. SAFER architecture



Expiry Date, and so forth. The Agent Factory then signs this immutable part. When the receiving host accepts an agent, it can verify with the Agent Factory's public key whether the agent's credentials have been modified. The mutable part of the agent includes the Host Trace, which stores a list of names of the hosts that the agent has visited so far. Upon checking, if any distrusted host is found, a host may decide not to trust this agent and impose a stricter security policy on it.

In SAFER, the main cryptographic technology used is the PKI. The public keys are stored in a common database located in CAC, where the public has read access, but no access to modify existing records.

Authentication Process

Authenticating Host

Before roaming to the next host, it is the duty of the agent to authenticate the next host to make sure that the host it is visiting is a genuine one. A hand-shaking method is devised to authenticate hosts. The agent sends its request to a host, asking for permission to visit. The host will sign on the agent's request message with its private key and send it back to the agent. The agent can then verify the host's identity by extracting the host's public key from the common database and authenticating the signature. If the authentication is successful, then the agent is communicating with the genuine host and starts to ship itself over to the host.

Authenticating Agent

Authentication of an agent involves two major steps: 1) to verify the agents' credentials, and 2) to verify the mutable part of the agent, checking whether it has been tampered with by anyone in its roaming process.
The authentication procedure is shown in Figure 2. Firstly, the agent will be checked for its expiry date. If it has not expired, its host trace will be examined to see if it has been to any distrusted host. If the agent passes these two tests, the final test is to check the trustworthiness of the factory that has manufactured it.

Authorization Process

After the host accepts an agent, it has to determine what resources the agent is allowed to access based on the level of authentication that the agent has passed. Four levels of authorization have been designed, with level 1 being the strictest and level 4 the most lenient. Level 1 authority is given to agents that the host does not have

Figure 2. Authentication & authorization procedure

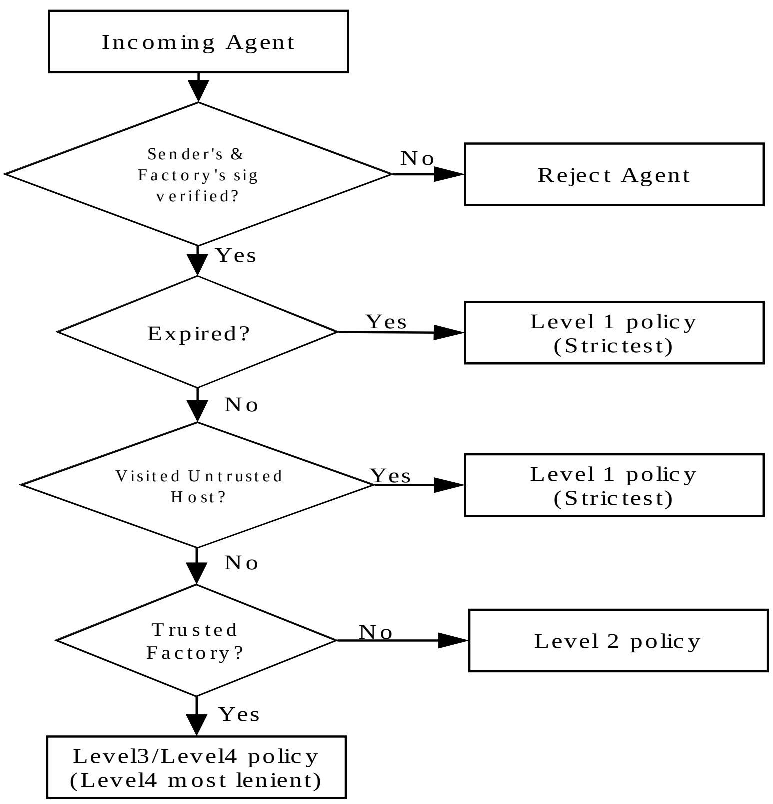


Table 1. Definition of the various security policies

Level of leniency	Policy name	Permissions
Level 4 (Most lenient)	Polfile.policy	FilePermission (Read, write)
		AWT Permission
		Socket Permission (Accept, Listen, Connect)
		Runtime Permission (create/set SecurityManager, queuePrintJob)
Level 3	Pol1.policy	FilePermission (Read, write)
		AWT Permission
		Socket Permission (Accept, Listen, Connect)
		Runtime Permission (create SecurityManager, queuePrintJob)
Level 2	Pol2.policy	FilePermission (Read only)
		AWT Permission
		Socket Permission (Accept, Connect)
		Runtime Permission (create SecurityManager)
Level 1 (Most Strict)	Pol3.policy	FilePermission (Read only)
		No AWT Permission
		No Socket Permission
		No Runtime Permission

much trust in. An agent that passes all levels of authentication and is deemed to be trusted may be awarded the level 4 authority. Table 1 shows the four policies and the restrictions imposed on each policy. The permissions can be customized to meet the requirements of different hosts. Here, AWT stands for Abstract Window Toolkit (AWT), which allows our programs to create a Graphical User Interface (GUI) to interact with the users.

Implementation

Implementation of agent authentication and authorization was done using the Java programming language. The Java Security API and the Java Cryptography Extension were widely used in the implementation. The graphical user interfaces were designed using the Java Swing components.

Discussions

We do not intend to compare our prototype with mobile agent systems such as D'agents and Aglets on detail benchmarks, as the focus of our approach is on the security issues in the context of e-commerce applications. Here, we present our features in comparison to related work, and discuss the advantages and limitations of our system in the following subsections.

Our approach has some features that are similar to the related systems discussed in the second section. For example, the authentication mechanism is based on PKI. The authorization mechanism is implemented using the Java security manager and some user-defined security policies. The major features of our approach lie in the method of authorization. Some agent systems authorize agents based on a role-based access control list and some are based on the identity of the agent. The SAFER system is different in that it allocates a different security policy based on the identity of the agent and the level of authentication it has passed.

Advantages of Our Infrastructure

Storage of Keys

One of the principal advantages of the prototype implemented is that there is no sending of keys over the network. This enhances the security of the system since it is impossible that keys can get intercepted and replaced.

The database storage of the public keys also allows an efficient method of retrieval of keys. This facilitates the verification of all previous signatures in the agent by the current host. For example, the owner may want to verify the signatures of all the previous hosts that the agent has visited. Instead of having all these hosts append their public keys to the agent (which may be compromised later), the owner can simply retrieve the keys from the database according to the hosts' ID.

Examining the Agent's Host Trace

Every host is required to sign on the agent's host trace before it is dispatched to the next destination. The IDs of the hosts visited are compared with the distrusted host list that each host would keep. If a distrusted host were found, the host would then take special precautions against these agents by imposing a stricter security policy on its operations.

Automation of the Authentication and Authorization Process

The beauty of this system is that it can be automated or run manually when the need arises. In the automatic configuration, when an agent is sent over, the host will do the authentication and assign an appropriate security policy to the agent. If the execution is successful, the host signs the agent and adds its identity on the host trace, before sending it out to the next host. In the manual configuration, all the authentication and authorization procedures need prompting from the host owner. The advantage is that the host has more control on what methods to authenticate and what authorization level and policy to enforce on the agent.

Limitations of Our Infrastructure

Pre-Determined Security Policies

In the current design, the agent is assigned to the security policy based on the authentication process. Having predetermined security policies may be stifling to the operations of an agent. It would be useless if the agent is denied the read access but instead granted other permissions that it does not need. The limitation here is an implementation choice because the mechanism to customize the permission for each agent has not been developed. Predetermined security policies are simpler to implement for large-scale systems.

Difficulty in Identifying a Malicious Host

The current implementation does not have a way of identifying the host that is causing the attacks on the agent. The agent owner can only detect that certain information has been tampered with, but does not know which host exactly caused the disparity.

FUTURE TRENDS

The implementation of the prototype has provided a basic infrastructure to authenticate and authorize agents. We are improving our approaches and implementation in two aspects. Firstly, to make the system more flexible in enforcing restrictions on agents, a possible improvement is to let the agent specify the security policy that it requires for its operation at the particular host. It is desirable to have a personalized system with the agent stating what it needs and the host deciding on whether to grant the permission or not. Secondly, the protection of agents against other agents can be another important issue. The authentication and authorization aspects between communicating agents are similar to that of host-to-agent and agent-to-host processes. We are designing certain mechanisms for this type of protection.

M

CONCLUSION

The advantages of employing mobile agents can only be manifested if there is a secure and robust system in place.

In this article, the design and implementation of agent authentication and authorization are elaborated. By combining the features of the Java security environment and the Java Cryptographic Extensions, a secure and robust infrastructure is built. PKI is the main technology used in the authentication module. To verify the integrity of the agent, digital signature is used. The receiving party would use the public keys of the relevant parties to verify that all the information on the agent is intact. In the authorization module, the agent is checked regarding its trustworthiness and a suitable user-defined security policy will be recommended based on the level of authentication the agent has passed. The agent will be run under the security manager and the prescribed security policy.

REFERENCES

Chavez, A., & Maes, P. (1998). Kasbah: An agent marketplace for buying and selling goods. *Proceedings of First International Conference on Practical Application of Intelligent Agents and Multi-Agent Technology*, London (pp. 75-90).

Corradi, A., Montanari, R., & Stefanelli, C. (1999). Mobile agents integrity in e-commerce applications. *Proceedings of 19th IEEE International Conference on Distributed Computing Systems* (pp. 59-64).

Dasgupta, P., Narasimhan, N., Moser, L.E., & Melliar-Smith, P.M. (1999). MAgNET: Mobile agents for networked electronic trading. *IEEE Transactions on Knowledge and Data Engineering*, *11*(4), 509-525.

Gray, R.S., Kotz, D., Cybenko, G., & Rus, D. (1998). D'Agents: Security in a multiple-language, mobile-agent system. In G. Vigna (Ed.), *Mobile agents and security. Lecture notes in computer science.* Springer-Verlag.

Greenberg, M.S., Byington, J.C., & Harper, D.G. (1998). Mobile agents and security. *IEEE Communications Magazine*, *36*(7), 76-85.

Guan, S.U., & Yang, Y. (1999). SAFE: Secure-roaming agent for e-commerce. *Proceedings the 26th International Conference on Computers and Industrial Engineering*, Melbourne, Australia (pp. 33-37).

Guan, S.U., & Zhu, F.M. (2001). Agent fabrication and IS implementation for agent-based electronic commerce. To appear in *Journal of Applied Systems Studies*.

Guan, S.U., Zhu, F.M., & Ko, C.C. (2000). Agent fabrication and authorization in agent-based electronic commerce. *Proceedings of International ICSC Symposium on Multi-Agents and Mobile Agents in Virtual Organizations and E-Commerce*, Wollongong, Australia (pp. 528-534).

Hua, F., & Guan, S.U. (2000). Agent and payment systems in e-commerce. In S.M. Rahman & R.J. Bignall (Eds.), *Internet commerce and software agents: Cases, technologies and opportunities* (pp. 317-330). Hershey, PA: Idea Group Inc.

Jardin, C.A. (1997). *Java electronic commerce sourcebook*. New York: Wiley Computer Publishing.

Karnik, N., & Tripathi, A. (1999). *Security in the Ajanta mobile agent system.* Technical report. Department of Computer Science, University of Minnesota.

Karnik, N.M., & Tripathi A.R. (2001). Security in the Ajanta mobile agent system. *Software Practice and Experience*, *31*(4), 301-329.

Lange, D.B., & Oshima, M. (1998). *Programming and deploying JAVA mobile agents with aglets.* Addison-Wesley.

Lee, R.S.T. (2002) iJADE authenticator - An intelligent multiagent based facial authentication system. *International Journal of Pattern Recognition and Artificial Intelligence*, *16*(4), 481-500.

Marques, P.J., Silva, L.M., & Silva, J.G. (1999). Security mechanisms for using mobile agents in electronic commerce. *Proceedings of the 18th IEEE Symposium on Reliable Distributed Systems* (pp. 378-383).

Milojicic, D. (1999). Mobile agent applications. *IEEE Concurrency*, *7*(3), 80-90.

Ono, K., & Tai, H. (2002). A security scheme for Aglets. *Software Practice and Experience*, *32*(6), 497-514.

Oppliger, R. (1999). Security issues related to mobile code and agent-based systems. *Computer Communications*, *22*(12), 1165-1170.

Pistoia, M., Reller, D.F., Gupta, D., Nagnur, M., & Ramani, A.K. (1999). *Java 2 network security*. Prentice Hall.

Poh, T.K., & Guan, S.U. (2000). Internet-enabled smart card agent environment and applications. In S.M. Rahman & M. Raisinghani (Eds.), *Electronic commerce: Opportunities and challenges* (pp. 246-260). Hershey, PA: Idea Group Inc.

Rivest, R.L., Shamir, A., & Adleman, L.M. (1978). A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM.*

Simonds, F. (1996). *Network security: Data and voice communications.* McGraw-Hill.

Tripathi, A., Karnik, N., Ahmed, T. et al. (2002). Design of the Ajanta system for mobile agent programming. *Journal of Systems and Software.*

Tsvetovatyy, M., Mobasher, B., Gini, M., & Wieckowski, Z. (1997). MAGMA: An agent based virtual market for electronic commerce. *Applied Artificial Intelligence, 11*(6), 501-524.

Wang, T., Guan, S.U., & Chan, T.K. (2001). Integrity protection for code-on-demand mobile agents in e-commerce. To appear in *Journal of Systems and Software.*

Wayner, P. (1995). *Agent unleashed: A public domain look at agent technology.* London: Academic Press.

Wong, D., Paciorek, N., & Moore, D. (1999). Java-based mobile agents. *Communications of the ACM, 42*(3), 92-102.

Zhu, F.M., & Guan, S.U. (2001). Towards evolution of software agents in electronic commerce. *Proceedings of the IEEE Congress on Evolutionary Computation 2001,* Seoul, Korea (pp. 1303-1308).

Zhu, F.M., Guan, S.U., & Yang, Y. (2000). SAFER e-commerce: Secure agent fabrication, evolution & roaming for eE-commerce. In S.M. Rahman & R.J. Bignall (Eds.), *Internet commerce and software agents: Cases, technologies and opportunities* (pp. 190-206). Hershey, PA: Idea Group Inc.

KEY TERMS

Agents: A piece of software, which acts to accomplish tasks on behalf of its user.

Authentication: The process of ensuring that an individual is who he or she claims to be.

Authorization: The process of giving access rights to an individual or entity.

Cryptography: The act of protecting data by encoding them, so that they can only be decoded by individuals who possess the key.

Digital Signature: Extra data appended to the message in order to authenticate the identity of the sender, and to ensure that the original content of the message or document that has been sent is unchanged.

Java: A high-level programming language similar to C++ developed by SUN Microsystems.

Private Key: That key (of a user's public-private key pair) known only to the user.

Public Key: The publicly distributed key that if combined with a private key (derived mathematically from the public key), can be used to effectively encrypt messages and digital signatures.

Mobile Commerce Technology

Chung-wei Lee
Auburn University, USA

Wen-Chen Hu
University of North Dakota, USA

Jyh-haw Yeh
Boise State University, USA

INTRODUCTION

With the introduction of the World Wide Web, electronic commerce has revolutionized traditional commerce and boosted sales and exchanges of merchandise and information. Recently, the emergence of wireless and mobile networks has made possible the admission of electronic commerce to a new application and research subject—mobile commerce, which is defined as the exchange or buying and selling of commodities, services, or information on the Internet through the use of mobile handheld devices. With services provided by mobile commerce, consumers may use the microbrowsers on their cellular phones or PDAs to buy tickets, order meals, locate and book local hotel rooms, even write contracts on the move.

In just a few years, mobile commerce has emerged from nowhere to become the hottest new trend in business transactions. NTT DoCoMo's i-mode (2003) is by far the most successful example of mobile commerce. Introduced in February 1999, i-mode has attracted over 36 million subscribers worldwide. With i-mode, cellular phone users can easily access more than 62,000 Internet sites, as well as specialized services such as e-mail, online shopping and banking, ticket reservations, and personalized ringing melodies that can be downloaded for their phones. The i-mode network structure not only provides access to i-mode and i-mode-compatible contents through the Internet, but also provides access through a dedicated leased-line circuit for added security. i-mode users are charged based on the volume of data transmitted, rather than the amount of time spent connected. In Spring 2001, NTT DoCoMo introduced its next-generation mobile system, based on wideband CDMA (W-CDMA), which can support speeds of 384Kbps or faster, allowing users to download videos and other bandwidth-intensive content with its high-speed packet data communications.

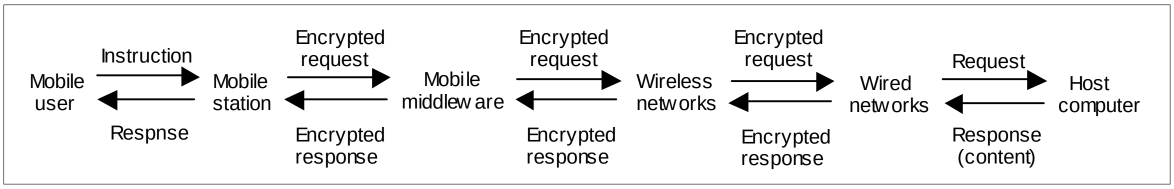
BACKGROUND

A mobile commerce system is very complex because it involves such a wide range of disciplines and technologies. In general, a mobile commerce system can be divided into six components: (1) mobile commerce applications, (2) mobile stations, (3 mobile middleware, (4) wireless networks, (5) wired networks, and (6) host computers.

To explain how these components work together, the following outline gives a brief description of a typical procedure that is initiated by a request submitted by a mobile user:

1. *Mobile commerce applications:* A content provider implements an application by providing two sets of programs: client-side programs, such as a user interface on a microbrowser, and server-side programs, such as database accesses and updating.
2. *Mobile stations:* Mobile stations present user interfaces to the end users, who specify their requests on the interfaces. The mobile stations then relay user requests to the other components and display the processing results later using the interfaces.
3. *Mobile middleware:* The major purpose of mobile middleware is to seamlessly and transparently map Internet contents to mobile stations that support a wide variety of operating systems, markup languages, microbrowsers, and protocols. Most mobile middleware also encrypts the communication in order to provide some level of security for transactions.
4. *Wireless networks:* Mobile commerce is possible mainly because of the availability of wireless networks. User requests are delivered to either the closest wireless access point (in a wireless local area network environment) or a base station (in a cellular network environment).

Figure 1. Flowchart of a user request processed in a mobile commerce system



5. *Wired networks:* This component is optional for a mobile commerce system. However, most computers (servers) usually reside on wired networks such as the Internet, so user requests are routed to these servers using transport and/or security mechanisms provided by wired networks.
6. *Host computers:* This component is similar to the one used in electronic commerce, which includes three kinds of software. User requests are generally acted upon in this component.

To better illustrate the above procedure, Figure 1 depicts a flowchart showing how a user request is processed by the components in a mobile commerce system (Leavitt, 2000).

MOBILE COMMERCE SYSTEMS

Since each component in a mobile commerce system is large enough to be a research area by itself, only elements in components that are specifically related to mobile commerce are explained in this article. Related research on mobile commerce systems can be found in the article by Varshney, Vetter, and Kalakota (2000).

Mobile Commerce Applications

The applications of electronic commerce are already widespread; mobile commerce applications not only cover these but also include new ones. For example, some tasks that are not feasible for electronic commerce, such as mobile inventory tracking and dispatching, are possible for mobile commerce. Table 1 lists some of the major mobile commerce applications (Gordon & Gebauer, 2001; Sadeh, 2002), along with details of each.

Mobile Stations

A mobile station or a mobile handheld device, such as a personal digital assistant (PDA) or Web-enabled cellular phone, may embrace many of the features of computers, telephone/fax, e-mails, and personal information managers (PIMs), such as calendars and address books, and networking features. A mobile station differs from a PC or notebook due to its limited network bandwidth, limited screen/body size, and mobility features. The limited network bandwidth prevents the display of most multimedia on a microbrowser, while the limited screen/body size restricts the mobile stations of today to either a stylus or keyboard version. Table 2 lists some major mobile station

Table 1. Major mobile commerce applications

Mobile Category	**Major Applications**	**Clients**
Commerce	Mobile transactions and payments	Businesses
Education	Mobile classrooms and labs	Schools and training centers
Enterprise resource planning	Resource management	All
Entertainment	Games/images/music/video downloads and online gaming	Entertainment industry
Health care	Accessing and updating patient records	Hospitals and nursing homes
Inventory tracking and dispatching	Product tracking and dispatching	Delivery services and transportation
Traffic	Global positioning, directions, and traffic advisories	Transportation and auto industries
Travel and ticketing	Travel management	Travel industry and ticket sales

Table 2. Specifications of some major mobile stations

Vendor & Device	Operating System	Processor	Installed RAM/ROM	Input Methods	Key Features
Compaq iPAQ H3870	MS Pocket PC 2002	206 MHz Intel StrongARM 32-bit RISC	64 MB/32 MB	Touchscreen	Wireless email/Internet
Handspring Treo 300	Palm OS 3.5.2H	33 MHz Motorola Dragonball VZ	16 MB/8 MB	Keyboard/ Stylus	CDMA network
Motorola Accompli 009	Wisdom OS 5.0	33 MHz Motorola Dragonball VZ	8 MB/4 MB	Keyboard	GPRS network
Nokia 9290 Communicator	Symbian OS	32-bit ARM9 RISC	16 MB/8 MB	Keyboard	WAP
Nokia 6800	Series 40			Keyboard	Innovative keyboard integration
Palm i705	Palm OS 4.1	33 MHz Motorola Dragonball VZ	8 MB/4 MB	Stylus	Wireless Email/Internet
Samsung SPH-i330	Palm OS 4.1	66MHz Motorola Dragonball Super VZ	16 MB/8 MB	Touchscreen/ Stylus	Color screen
Sony Clie PEG-NR70V	Palm OS 4.1	66 MHz Motorola Dragonball Super VZ	16 MB/8 MB	Keyboard/ Stylus/ Touchscreen	Multimedia
Sony Ericsson T68i			800KB	Keyboard	Multimedia Messaging Service
Toshiba E740	MS Pocket PC 2002	400 MHz Intel PXA250	64 MB/32 MB	Stylus/ Touchscreen	Wireless Internet
Sony Ericsson Z1010			32MB	Keyboard	MP3, MMS, WAP2.0

specifications, although several table entries may be incomplete as some of the information is classified as confidential due to business considerations.

Mobile Middleware

The term middleware refers to the software layer between the operating system and the distributed applications that interact via the networks. The primary mission of a middleware layer is to hide the underlying networked environment's complexity by insulating applications from explicit protocol handling disjoint memories, data replication, network faults, and parallelism (Geihs, 2001). Mobile middleware translates requests from mobile stations to a host computer and adapts content from the host to the mobile station (Saha, Jamtgaard, & Villasenor, 2001). According to an article in *Eurotechnology* entitled Frequently asked questions about NTT-DoCoMo's i-mode (2000), 60%of the world's wireless Internet users use i-mode, 39% use WAP, and 1% use Palm middleware. Table 3 compares i-mode and WAP, the two major kinds of mobile middleware.

Wireless Networks

Network infrastructure provides essential voice and data communication capability for consumers and vendors in cyberspace. Evolving from electronic commerce (EC) to

Table 3. Comparisons of WAP and i-mode

	WAP	i-mode
Developer	WAP Forum	NTT DoCoMo
Function	A protocol	A complete mobile Internet service
Host Language	WML (Wireless Markup Language)	CHTML (Compact HTML)
Major Technology	WAP Gateway	TCP/IP modifications
Key Features	Widely adopted and flexible	Highest number of users and easy to use

mobile commerce (MC), it is necessary for a wired network infrastructure, such as the Internet, to be augmented by wireless networks that support mobility for end users. From the perspective of mobile commerce, wireless networks can be categorized into wireless local area networks (WLANs) and wireless cellular networks.

WLAN technologies are suitable for office networks, home networks, personal area networks (PANs), and ad hoc networks. In a one-hop WLAN environment, where an access point (AP) acting as a router or switch is a part of a wired network, mobile devices connect directly to the AP through radio channels. Data packets are relayed by the AP to the other end of a network connection. If no APs are available, mobile devices can form a wireless ad hoc network among themselves and exchange data packets or perform business transactions as necessary. Many WLAN products are available on the market. In general, Bluetooth technology supports very limited coverage range and throughput. Thus it is only suitable for applications in personal area networks. In many parts of the world, the IEEE 802.11b (Wi-Fi) system is now the most popular wireless network and is used in offices, homes, and public spaces such as airports, shopping malls, and restaurants. However, many experts predict that with much higher transmission speeds, 802.11g will replace 802.11b in the near future.

Cellular system users can conduct mobile commerce operations through their cellular phones. Under this scenario, a cellular phone connects directly to the closest base station, where communication is relayed to the service site through a radio access network (RAN) and other fixed networks. Originally designed for voice-only communication, cellular systems are evolving from analog to digital, and from circuit-switched to packet-switched networks, in order to accommodate mobile commerce (data) applications. Currently, most of the cellular wireless networks in the world follow 2G or 2.5G standards. However, there is no doubt that, in the near future, 3G systems with quality-of-service (QoS) capability will dominate wireless cellular services. The two main standards for 3G are Wideband CDMA (WCDMA), proposed by Ericsson, and CDMA2000, proposed by Qualcomm.

Host Computers

A host computer processes, produces, and stores all the information for mobile commerce applications. This component is similar to that used in an electronic commerce system because the host computers are usually not aware of differences among the targets, browsers or microbrowsers they serve. It is the application programs that are responsible for apprehending their clients and responding to them accordingly. Most of the mobile commerce application programs reside in this component, except for some client-side programs such as cookies. Usually this component contains three major elements: a Web server, a database server, and application programs and support software.

FUTURE TRENDS

It is estimated that 50 million wireless phone users in the United States will use their handheld devices to authorize payment for premium content and physical goods at some point during the year of 2006. This represents 17% of the projected total population and 26% of all wireless users (The Yankee Group, 2001). Mobile commerce is an effective and convenient way to deliver electronic commerce to consumers from anywhere and at anytime. Realizing the advantages to be gained from mobile commerce, many major companies have begun to offer mobile commerce options for their customers in addition to the electronic commerce they already provide (Over 50% of large U.S. enterprises plan to implement a wireless/mobile solution by 2003, 2001).

However, without secure commercial information exchange and safe electronic financial transactions over mobile networks, neither service providers nor potential customers will trust mobile commerce systems. Mobile security and payment are hence crucial issues for mobile commerce. Security issues span the whole mobile commerce system, from one end to the other, from the top to the bottom network protocol stack, from machines to humans. For example, in WAP, security is provided through the Wireless Transport Layer Security (WTLS) protocol (in WAP 1.0) and IETF standard Transport Layer Security (TLS) protocol (in WAP 2.0). They provide data integrity, privacy, and authentication. One security problem, known as the “WAP Gap” is caused by the inclusion of the WAP gateway in a security session. That is, encrypted messages sent by end systems might temporarily become clear text on the WAP gateway when messages are processed. One solution is to make the WAP gateway resident within the enterprise (server) network (Ashley, Hinton, & Vandenwauver, 2001), where heavyweight security mechanisms can be enforced.

In an IEEE 802.11 WLAN, security is provided by a data link level protocol called Wired Equivalent Privacy (WEP). When it is enabled, each mobile host has a secret key that is shared with the base station. The encryption algorithm used in WEP is a synchronous stream cipher based on RC4. The ciphertext is generated by XORing the plaintext with a RC4 generated keystream. However, recently published literature has discovered methods for breaking this approach (Borisov, Goldberg, & Wagner, 2001; Fluhrer, Martin, & Shamir, 2001; Stubblefield, Ioannidis, & Rubin, 2002). The next version, 802.11i, is expected to have better security.

Payment on mobile commerce systems is another issue. Although the Secure Electronic Transaction (SET) protocol (SET Secure Electronic Transaction Specification, Version 1.0, 1997) is likely to become the global standard in the domain of electronic commerce over the Internet, a WAP client device normally does not have sufficient processing and memory capability to utilize SET software. A "thin" SET wallet approach (Jin, Ren, Feng, & Hua, 2002) has thus been proposed to adapt the SET protocol for WAP clients. Under the "thin" SET wallet model, most of the functionality of current "fat" SET wallets is moved to the wallet server. To support a SET payment, a WAP client installed with only a "thin" wallet securely connects with a wallet server, which communicates with other SET entities. When SET purchase requests arrive from the "thin" wallet, the wallet server takes over the responsibility of routing requests and managing digital keys and certificates.

CONCLUSION

The emerging wireless and mobile networks have extended electronic commerce to another research and application subject: mobile commerce. A mobile commerce system involves a range of disciplines and technologies. This level of complexity makes understanding and constructing a mobile commerce system an arduous task. To facilitate this process, this article divided a mobile commerce system into six components, which can be summarized as follows:

- Mobile commerce applications: Electronic commerce applications are already broad. Mobile commerce applications not only cover the existing applications, but also include new applications, which can be performed at any time and from anywhere by using mobile computing technology.
- Mobile stations: Mobile stations are limited by their small screens, limited memory, limited processing power, and low battery power, and suffer from wireless network transmission problems. Numerous mobile stations, such as PDAs or Web-enabled cellular phones, are available on the market, but most use one of three major operating systems: Palm OS, Microsoft Pocket PC, and Symbian OS. At this moment, Palm OS leads the market, although it faces a serious challenge from Pocket PC.
- Mobile middleware: WAP and i-mode are the two major kinds of mobile middleware. WAP is widely adopted and flexible, while i-mode has the highest number of users and is easy to use. It is difficult to predict which middleware will be the eventual winner in the end; it is more likely that the two will be blended somehow at some point in the future.
- Wireless and wired networks: Wireless communication capability supports mobility for end users in mobile commerce systems. Wireless LANs and cellular networks are major components used to provide radio communication channels so that mobile service is possible. In the WLAN category, the Wi-Fi standard with 11 Mbps throughput dominates the current market. It is expected that standards with much higher transmission speeds, such as 802.11g, will replace Wi-Fi in the near future. Compared to WLANs, cellular systems can provide longer transmission distances and greater radio coverage, but suffer from the drawback of much lower bandwidth (less than 1 Mbps). In the latest trend for cellular systems, 3G standards supporting wireless multimedia and high-bandwidth services are beginning to be deployed. WCDMA and CDMA2000 are likely to dominate the market in the future.
- Host computers: Host computers process and store all the information needed for mobile commerce applications, and most application programs can be found here. They include three major components: Web servers, database servers, and application programs and support software.

An important trend for mobile commerce is enhancing mobile security mechanisms and payment methods. Mobile commerce systems can prosper only if information can be securely exchanged among end systems (consumers and vendors). Security issues (including payment) include data reliability, integrity, confidentiality, and authentication and are usually a crucial part of implementation in wireless protocols/systems. Solutions are updated frequently, due to the lack of a comprehensive wireless security infrastructure and standard. A unified approach has not yet emerged.

REFERENCES

Ashley, P., Hinton, H., & Vandenwauver, M. (2001). Wired versus wireless security: The Internet, WAP and iMode for E-Commerce. In *Proceedings of Annual Computer Security Applications Conferences (ACSAC)*, New Orleans, LA, December 10-14, 2001 (p. 296).

Borisov, N., Goldberg, I., & Wagner, D. (2001). Intercepting mobile communications: The insecurity of 802.11. In

Proceedings of the 7th International Conference on Mobile Computing and Networking, Rome, Italy, July 16-21, 2001 (pp. 180-189).

Fluhrer, S., Martin, I., & Shamir, A. (2001). Weakness in the key scheduling algorithm of RC4. In *Proceedings of the 8th Annual Workshop on Selected Areas in Cryptography*, Toronto, Ontario, Canada, August 16-17, 2001.

Frequently asked questions about NTT-DoCoMo's i-mode (2000). *Eurotechnology*. Retrieved December 16, 2002, from http://www.eurotechnology.com/imode/faq.html.

Geihs, K. (2001). Middleware challenges ahead. *IEEE computer*, 34(6), 24-31.

Gordon, P. & Gebauer, J. (2001). M-commerce: Revolution + inertia = evolution. *Working Paper 01-WP-1038*, University of California, Berkeley, CA.

i-mode (2003). *NTT-DoCoMo*. Retrieved November 28, 2002 from http://www.nttdocomo.com/.

Jin, L., Ren, S., Feng, L., & Hua, G. Z. (2002). Research on WAP clients supports SET payment protocol. *IEEE Wireless Communications*, 9(1), 90-95.

Leavitt, N. (2000). Will WAP deliver the wireless Internet? *IEEE Computer*, 34(5), 16-20.

Sadeh, N. (2002). *M-commerce: Technologies, services, and business models,* pp. 177-179. New York: John Wiley & Sons.

Saha, S., Jamtgaard, M., & Villasenor, J. (2001). Bringing the wireless Internet to mobile devices. *IEEE Computer*, 34(6), 54-58.

SET Secure Electronic Transaction Specification, Version 1.0 (1997). Retrieved October 11, 2002 from http://www.setco.org/.

Stubblefield, A., Ioannidis, J., & Rubin, A.D. (2002). Using the Fluhrer, Martin, and Shamir attack to break WEP. In *Proceedings of the Network and Distributed Systems Security Symposium*, San Diego, CA, February 6-8, 2002 (pp. 17-22).

Varshney, U., Vetter, R. J., & Kalakota, R. (2000). Mobile commerce: A new frontier. *IEEE Computer*, 33(10), 32-38.

WAP (Wireless Application Protocol) (2003). *Open Mobile Alliance Ltd*. Retrieved November 21, 2002 from http://www.wapforum.org/.

The Yankee Group (2001) Over 50% of large U.S. enterprises plan to implement a wireless/mobile solution by 2003. Retrieved December 10, 2002 from http://www.yankeegroup.com/public/news_releases/news_release_detail.jsp?ID=PressReleases/news_09102002_wmec.htm.

The Yankee Group publishes U.S. mobile commerce forecast (2001). *Reuters*. Retrieved December 16, 2002. from http://about.reuters.com/newsreleases/art_31-10-2001_id765.asp

KEY TERMS

i-mode: the full-color, always-on, and packet-switched Internet service for cellular phone users offered by NTT DoCoMo.

Mobile Commerce: the exchange or buying and selling of commodities, services, or information on the Internet (wired or wireless) through the use of mobile handheld devices.

SET: the Secure Electronic Transaction (SET) protocol is a technical standard designed to provide security for payment transactions among cardholders, merchants, payment gateways, and certification authorities in Internet.

Third Generation (3G): wireless system that can provide fairly high-speed (384 Kbps) packet-switched wide-area wireless Internet access to support multimedia applications.

Wi-Fi: IEEE 802.11b (Wi-Fi) is a wireless local area network standard. It operates in an unlicensed radio frequency band at 2.4 GHz and provides data access at 11 Mbps.

Wired Equivalent Privacy (WEP): a data link-level protocol that provides security for the IEEE 802.11 WLAN standards. The encryption algorithm used in WEP is a stream cipher based on

Wireless Application Protocol (WAP): an open, global specification that allows users with mobile devices to easily access and interact with information and services instantly

Mobile Location Services

George M. Giaglis
Athens University of Economics and Business, Greece

INTRODUCTION

The term "mobile era" as a characterization of the 21st century can hardly be considered an exaggeration (Kalakota & Robinson, 2001). Mobile phones are the fastest penetrating technology in the history of mankind, and global mobile phone ownership has surpassed even the ownership of fixed phones. Mobile applications, despite potentially being very different in nature from each other, all share a common characteristic that distinguishes them from their wire-line counterparts: they allow their users to move around while remaining capable of accessing the network and its services. In the mobility era, *location identification* has naturally become a critical attribute, as it opens the door to a world of applications and services that were unthinkable only a few years ago (May, 2001).

The term "mobile location services" (MLS) [or "location-based services (LBS), as they are sometimes also referred to] has been coined to group together applications and services that utilize information related to the geographical position of their users to provide value-adding services to them (Rao & Minakakis, 2003). This article provides a concise introduction to the major types of MLS and also introduces the reader to the most important positioning technologies that render the provision of MLS possible. Finally, the article also introduces a number of issues that are critical for the future of MLS, including privacy protection, regulation, and standardization.

CATEGORIES OF MOBILE LOCATION SERVICES

Mobile networks are quickly becoming ubiquitous. The ability to reach mobile phone users regardless of their location and, even more importantly, the ability to reach mobile phone users *based* on their location has created a new world of exciting and promising applications. While the possibilities for providing innovative MLS are limited only by one's imagination, we will outline the most important categories of such services in this section.

Emergency Management

Perhaps the clearest market application of MLS is the ability to locate an individual who is either unaware of his or her exact location or is not able to reveal it because of an emergency situation (injury, criminal attack, and so on). MLS are even applicable as a means of overcoming one of the most common problems of motorists, namely, the fact that, most often than not, they are unaware of their exact location when their vehicle breaks down. The ability of a mobile user to call for assistance and at the same time automatically reveal his or her exact location to the automotive assistance agency is considered one of the prime motivators for signing up subscribers to MLS (Hargrave, 2000).

Navigation Services

Navigation services are based on mobile users' needs for directions within their current geographical locations. The ability of a mobile network to locate the exact position of a mobile user can be manifested in a series of navigation-based services:

1. By positioning a mobile phone, an operator can let users know exactly where they are as well as give them detailed *directions* about how to get to a desirable destination.
2. Coupled with the ability to monitor traffic conditions, navigation services can be extended to include destination directions that take account of current *traffic conditions* (for example, traffic congestion or a road-blocking accident) and suggest alternative routes to mobile users.
3. The possibility to provide detailed directions to mobile users can be extended to support *indoor routing* as well. For example, users can be assisted in their navigation in hypermarkets, warehouses, exhibitions, and other information-rich environments to locate products, exhibition stands, and other points of interest.
4. Similarly, *group management* applications can be provided to allow mobile users to locate friends, family, coworkers, or other members of a particular

group that are within close range and, thus, create *virtual communities* of people with similar interests.

Information Provision

Location-sensitive information services mostly refer to the digital distribution of content to mobile terminal devices based on their location, time specificity, and user behavior. The following types of services can be identified within this category:

1. *Travel services*, such as guided tours (either automated or operator-assisted), notification about nearby places of interest (for example, monuments), transportation assistance, and other services that can be provided to tourists moving around in unfamiliar surroundings.
2. *Mobile yellow pages* that provide a mobile user, upon request, with knowledge regarding nearby facilities.
3. *Infotainment services*, such as information about local events, location-specific multimedia content, and so on.

Advertising and Marketing

Mobile advertising is among the first trial applications of MLS, due to its promising revenue potential and its direct links to mobile-commerce activities. Furthermore, mobile advertising has gained significant attention because of the unique attributes, such as *personalization* (Kalakota & Robinson, 2001), that offer new opportunities to advertisers to place effective and efficient promotions on mobile environments. There are various mechanisms for implementing mobile advertising coupled with MLS. Examples of mobile advertising forms include *mobile banners*, *alerts* (usually dispatched as SMS messages), and *proximity-triggered advertisements*.

Tracking

Tracking services can be equally applicable to the consumer and the corporate markets. As far as consumers are concerned, tracking services can be utilized to monitor the exact whereabouts of, for example, children and elderly people. Similarly, tracking services can be effectively applied in corporate situations as well. One popular example refers to tracking vehicles so that companies know where their fleet and goods are at any time. A similar application allows companies to locate their field personnel (for example, salespeople and repair engineers) so that they are able, for example, to dispatch the nearest engineer and provide their customers with accurate personnel arrival times. Finally, the newfound opportunity to provide accurate product tracking within the supply chain offers new possibilities to mobile supply chain management (m-SCM) applications (Kalakota & Robinson, 2001).

Billing

Location-sensitive billing refers to the ability of a mobile service provider to dynamically charge users of a particular service depending on their location when using or accessing the service. For example, mobile network operators may price calls based on the knowledge of the location of the mobile phone when a call is made. Location-sensitive billing includes the ability to offer reduced call rates to subscribers who use their mobile phone when at their home, thereby allowing mobile operators to compete more effectively with their fixed telephony counterparts.

POSITIONING TECHNOLOGIES

The applications and services that were discussed in the previous section are based on underlying technological capabilities that enable the identification of the location of a mobile device, thereby making the provision of MLS possible. Positioning techniques can be implemented in two ways: *self-positioning* and *remote positioning* (Zeimpekis et al., 2003).

In the first approach (self-positioning), the mobile terminal uses signals, transmitted by the gateways/antennas (which can be either terrestrial or satellite) to calculate its own position. More specifically, the positioning receiver makes the appropriate signal measurements from geographically distributed transmitters and uses these measurements to determine its position. A self-positioning receiver, therefore, "knows" where it is, and applications collocated with the receiver can use this information to make position-based decisions, such as those required for vehicle navigation.

In the case of remote positioning, the mobile terminal can be located by measuring the signals travelling to and from a set of receivers. More specifically, the receivers, which can be installed at one or more locations, measure a signal originating from, or reflecting off, the object to be positioned. These signal measurements are used to determine the length and direction of the individual radio paths, and then the mobile terminal position is computed from geometric relationships.

Self-Positioning Techniques

Global Positioning System (GPS) and Assisted GPS (A-GPS): GPS is the worldwide satellite-based radio navigation system, consisting of 24 satellites, equally spaced in six orbital planes 20,200 kilometres above the Earth, that transmit two specially coded carrier signals: one for civilian use and one for military and government use (Djuknic & Richton, 2001). The system's satellites transmit navigation messages that a GPS receiver uses to determine its position. GPS receivers process the signals to compute position in three dimensions—latitude, longitude, and altitude—with an accuracy of 10 meters or less. The main advantage of this technique is that GPS is already in use for many years. However, in order to operate properly, GPS receivers need a clear view of the skies and signals from at least three or four (depending on the type of information needed) satellites, requirements that exclude operation in indoor environments. As far as the A-GPS method is concerned, the mobile network or a third-party service provider can assist the handset by directing it to look for specific satellites and also by collecting data from the handset to perform location identification calculations that the handset itself may be unable to perform due to limited processing power. The A-GPS method can be extremely accurate, ranging from 1 to 10 meters (Giaglis et al., 2002).

Indoor Global Positioning System (Indoor GPS): This system focuses on exploiting the advantages of GPS for developing a location-sensing system for indoor environments. It should be noted that the GPS signal does not typically work indoors, because the signal strength is too low to penetrate a building (Chen & Kotz, 2000). Indoor GPS solutions can be applicable to wide space areas where no significant barriers exist. Indoor GPS takes into account the low power consumption and small size requirements of wireless access devices, such as mobile phones and handheld computers. The navigation signal is generated by a number of pseudolites (pseudo-satellites). These are devices that generate a GPS-like navigation signal. The signal is designed to be similar to the GPS signal in order to allow pseudolite-compatible receivers to be built with minimal modifications to existing GPS receivers. As in GPS, at least four pseudolites have to be visible for navigation, unless additional means, such as altitude aiding, are used (Giaglis et al., 2002).

Remote Positioning Techniques

Cell Identification (Cell-ID): The Cell-ID (or *Cell of Origin, COO*) method is the most widely used technique to provide location services and applications in second-generation mobile communication networks. The method relies on the fact that mobile networks can identify the approximate position of a mobile handset by knowing which cell site the device is using at a given time. The main benefit of the technology is that it is already in use today and can be supported by all mobile handsets. However, the accuracy of the method is generally low (in the range of 200 meters in densely covered areas and much lower in rural environments) (Giaglis et al., 2002).

Angle of Arrival (AOA): The basic idea is to steer in space a directional antenna beam until the direction of maximum signal strength is detected. In terrestrial mobile systems, the directivity required to achieve accurate measurements is obtained by means of antenna arrays (Sakagami et al., 1994). Basically, a single measurement produces a straight-line locus from the base station to the mobile phone. Another AOA measurement will yield a second straight line, and the intersection of the two lines gives the position fix for this system.

Time of Arrival (TOA): Positioning information is derived from the absolute time for a wave to travel between a transmitter and a receiver or vice versa. This implies that the receiver knows the exact time of transmission. Alternatively, this approach might involve the measurement of the round-trip time of a signal transmitted from a source to a destination and then echoed back to the source, giving a result twice that of the one-way measurement. This does not imply synchronization between the transmitter and the receiver and is the most common means of measuring propagation time.

Differential Time of Arrival (DTOA): The problem of having precisely synchronized clocks at transmitter and receiver is solved by using several transmitters synchronized to a common time base and measuring the time difference of arrival at the receiver. More specifically, each DTOA measurement defines a hyperbolic locus on which the mobile terminal must lie. The intersection of the hyperbolic loci will define the position of the mobile device.

CRITICAL ISSUES RELATED TO MLS

Further to the business and technological aspects of mobile location services discussed above, a number of other critical factors will also ultimately determine their successful application in mobile communication networks. Such issues include the need to protect sensitive personal information of individuals, as well as the role of regulation and standardization initiatives in establishing the right climate for market rollout and healthy competition.

Privacy Protection

According to Nokia (2001), "of all the challenges facing mobile location service providers, privacy is undoubtedly the biggest single potential barrier to market take-up" (p. 9). For example, mobile advertising based on a user's location is a sensitive issue and has to be provided only with the explicit consent of the user.

However, even in such a case, the likely exchange of information between third parties (for example, network operators and advertising agencies) may hamper the privacy of user's personal data. To ensure commercial success of mobile location services, user trust must be ensured. A clear prerequisite of the trust-building mechanism is that the control over the use of location information is always on the hands of the user, not of the network operator or the service provider.

Regulation and Standardization

The role of regulatory and policy-making bodies is substantially enhanced in the case of mobile location services. It is not surprising that the initial boost to the market has come from such bodies (the US FCC mandate for emergency services) and that the European Commission has had a very active role in the development of the market on the other side of the Atlantic.

Standardization can also be a serious success or failure factor for any new technology, and mobile location services are not an exception to this rule. A number of bodies worldwide are working toward defining commonly accepted standards for the mobile industry, but prior experience has shown that standardization efforts may have a regional, rather than a global, scope. For example, the presence of incompatible standards for second-generation mobile telephony in Europe and the Americas has created considerable problems for users and the industry alike. Worldwide efforts to define universal standards for third-generation systems provide a more optimistic view of the future; however, the danger of incompatibility and technological "islands" remains. To this end, a number of standardization initiatives are underway, sometimes initiated by the industry. For example, Ericsson, Motorola, and Nokia, have joined forces to establish the *Location Interoperability Forum (LIF)*, with the purpose of developing and promoting common and ubiquitous solutions for mobile location services.

The importance of standardization becomes even more evident when we think of what can be termed as *the paradox of mobile location services*. Although these services are by definition local, any given user will most probably need them when in a nonlocal environment. We can envisage tourists outside the familiar surroundings of their local residences relying on MLS to obtain assistance and directions, and companies also utilizing MLS to track their goods in distant lands. To be useful to their users, mobile location services must, therefore, be provided in a location-independent and user-transparent fashion. From a standardization and technological point of view, this requirement poses a difficult problem: *service portability* and *roaming* issues have to be resolved in order for MLS to be compelling to users (UMTS, 2001).

REFERENCES

Chen, G., & Kotz, D. (2000). A survey of context-aware mobile computing research. Dartmouth Computer Science Technical Report TR2000-381.

Djuknic, G. M., & Richton, R. E. (2001). Geolocation and assisted GPS. *IEEE Computer*, *34*(2), 123–125.

Giaglis, G. M., Kourouthanasis, P., & Tsamakos, A. (2002). Towards a classification framework for mobile location services. In B. E. Mennecke & T. J. Strader (Eds.), *Mobile commerce: Technology, theory, and applications*. Hershey, PA: Idea Group Publishing.

Giaglis, G.M., Pateli A., Fouskas, K., Kourouthanassis, P., Tsamakos, A. (2002). On the potential use of mobile positioning technologies in indoor environments. In C. Loebbecke, R. T. Wigand, J. Gricar, A. Pucihar, G. Lenart (Eds.) *The Proceedings of the 15th Bled Electronic Commerce Conference—E-Reality: Constructing the E-Economy*, Moderna Organizacija, Kranj, Slovenia, Vol 1, 413-429.

Hargrave, S. (2000). Mobile location services: A report into the state of the market. White paper. Cambridge Positioning Systems.

Kalakota, R., & Robinson, M. (2001). *M-business: The race to mobility*. New York: McGraw-Hill.

May, P. (2001). *Mobile commerce: Opportunities, applications, and technologies of wireless business*. London; New York: Cambridge University Press.

Nokia Corporation (2001) Mobile Location Services, White Paper of Nokia Corporation. Available online at *http://www.nokia.com/pc_files_wb2/mposition_mobile_location_services.pdf]*

Rao, B., & Minakakis, L. (2003). Evolution of mobile location-based services. *Communications of the ACM*, *46*(12), 61–65.

Sakagami, S., (1994), Vehicle position estimates by multi-beam antennas. In *Multi-path environments, IEEE Transactions on Vehicular Technologies, 43*(4), 902–908.

UMTS Forum. (2001). The UMTS third generation market—Phase II. UMTS Forum Report #13, April. Retrieved from *http://www.umts-forum.org*

Zeimpekis, V., Giaglis, G. M., & Lekakos, G. (2003). A taxonomy of indoor and outdoor positioning techniques for mobile location services. *ACM SIGECOM Exchanges, 3*(4), 19–27.

KEY TERMS

Angle of arrival (AOA): A positioning technology in which the mobile network sends directional antenna beams to locate a mobile device at the intersection of the directions of maximum signal strength.

Assisted global positioning system (A-GPS): A variation of the *global positioning system (GPS)* in which the mobile network or a third-party service provider assists the mobile handset in determining its geographical position (either by directing it to look for specific satellites or by collecting data from the handset to perform location identification calculations that the handset itself may be unable to perform due to limited processing power).

Cell Identification (Cell-ID): The Cell-ID method is the basic technique to provide location services and applications in second-generation mobile communication networks. The method relies on the fact that mobile networks can identify the approximate position of a mobile handset by knowing which cell site the device is using at a given time.

Differential time of arrival (DTOA): A positioning technology in which several transmitters (synchronized to a common time base) are used to measure time differences of arrival at the receiver and, hence, determine the receiver's geographical position.

Global positioning system (GPS): GPS is the worldwide satellite-based radio navigation system. The system's satellites transmit messages that a receiver uses to determine its own geographical position.

Indoor global positioning system (Indoor GPS): A variation of the *global positioning system (GPS)* for use in indoor environments, where the normal GPS signal does not typically work, because the signal strength is too low to penetrate a building. Indoor GPS navigation signals are generated by a number of pseudolites (pseudo-satellites) and are sent to pseudolite-compatible receivers that use the information to determine their own geographical positions.

Location-based services (LBS): A synonym for *mobile location services (MLS)* denoting applications that utilize the knowledge of one's geographical position to provide added-value services.

Location identification: The ability of mobile hosts to determine the geographical location of wireless access devices.

Mobile location services (MLS): Applications provided over a mobile network that utilize information related to the geographical position of their users to provide added value to them.

Time of arrival (TOA): A positioning technology where information is derived from the absolute time for a wave to travel between a transmitter and a receiver or vice versa.

Mobile Transaction Models Framework

Andrés Coratella
Universidad de Buenos Aires, Argentina

Miguel Felder
Universidad de Buenos Aires, Argentina
Pragma Consultores

Roberto Hirsch
Universidad de Buenos Aires, Argentina

Eduardo Rodriguez
Universidad de Buenos Aires, Argentina

INTRODUCTION

Currently, mobile technology is undergoing a high growth stage, allowing for an increasing plethora of mobile devices (handheld PCs, handsets, etc.) and daily access to distributed resources and information.

This availability entails the requirement for transactional capabilities adapted to the specific characteristics of the mobile environment without losing the consistency and reliability guarantees of traditional online transactional processing (OLTP) systems.

One of the objectives of this work was the study of transactional models applied to mobile environments. After analyzing the state of the art, we observed that none of the models covers all the necessary characteristics for current requirements, and therefore, we propose a framework that allows us to capture and compare the main features to be taken into account in these models.

BACKGROUND

Transactional Systems' Requirements for Mobile Environments

Below we describe the requirements needed for a transactional system in order to be applied to a mobile environment (Tewari et al., 1995; Dunham et al., 1997; Lee et al., 1997):

- ***Ability to distribute the transaction's processing***: Due to memory, power processing, and battery limitations of the mobile hosts (MHs), it may be necessary to execute certain sections of the transaction in the mobile support station (MSS).
- ***Share the state and the partial results***: Because, as was stated in the previous item, parts of a transaction can be executed in the MH, while others run in the MSS, items must be shared.
- ***Capture the movement of mobile transactions***: Due to the physical movement of the MHs, it is necessary to transfer a transaction's control as it moves from cell to cell.
- ***Support long-lived transactions***: This is required, because some processes can take a considerable amount of time, and besides, the search for a computer that has physically moved from a cell can be a time-costly operation.
- ***Support long disconnection periods***: Recall that disconnections can be caused by physical problems, or simply by the MH's own decision. For an MH to continue operating despite being disconnected from the network, it may be necessary to maintain local copies of the data needed (*caching techniques*).
- ***Support partial failures and provide different recovery strategies***: These failures can be caused by battery problems, static electricity, accidental computer turnoffs, etc.

Mobile Transactional Models

The use of transactions in a mobile environment differs substantially from the use of transactions in centralized or distributed systems. The main differences are the high disconnection frequency and the mobility of the transactions. Therefore, transactional models and commit coordination protocols must be revised to take into account the mentioned differences (Dunham et al., 1997).

There are models that explicitly support mobility, such as Kangaroo Transactions (Dunham et al., 1997), which creates a subtransaction on every MSS that the user passes by, establishing a link between them so as to move the data as the user moves. It can use compensations, if necessary, in case of failures. Another example of this kind of model is Moflex (Ku et al., 1998), which allows for the definition of parameters that specify, in a flexible way, the behavior of the transactions. A transaction is made up of subtransactions that can be either compensatable or not. It is possible to define the behavior these transactions will follow, such as the success or failure dependencies they must maintain with other subtransactions, conditions regarding the geographic location, etc. When a transaction faces a handoff, it can behave as specified by the user through a series of rules. Other models, like reporting/cotransactions (Chrysanthis, 1993) allow explicit sharing of partial results between active transactions, while transactions migrate from MSS to MSS as the MH moves from cell to cell, in order to minimize communication costs. Finally, the prewrite model (Madria, 2001) incorporates a prewrite operation before a write operation and also supports mobility.

Another group of models supports disconnections. An example is isolation-only transactions (IOTs) (Lu et al., 1994, 1995), designed to allow disconnected operations in mobile computers, detecting read/write conflicts based in limitations of the serializability. When a transaction commits, results are published only if all the accessed data maintained a connection with the server; otherwise, results are only locally visible, and the transaction is validated at reconnection time. In the weak/strict transactional model (Pitoura et al., 1995), transactions can also execute in disconnected mode, but there are special versions of read and write operations (called *weak read* and *weak write*) that operate on local copies of data items, which must be checked and eventually published at reconnection time. Data clusters of computers are defined that enforce the locality and validity concepts. There is another model, called planned disconnection (Holliday et al., 2000), with the main innovation of planning the disconnections that the user will perform, so as to minimize disruption to the remaining sites. A *planned disconnection* is defined as a disconnection where the user informs the system of his or her intention to disconnect and reconnect in an orderly manner. Finally, it is important to note that the prewrite model (described in the previous paragraph) also supports disconnection.

For more information on these models, please read Coratella et al. (2003).

COMPARISON FRAMEWORK

The idea behind our comparison framework is to act as a guide that should allow us to capture and compare the main characteristics to be taken into account in mobile transaction models.

Definition

Here, we describe the main characteristics that must be taken into account in order to define a transactional model capable of being applied to a mobile environment, trying to take full advantage of its peculiarities.

Relating Physical Aspects

- ***Mobility support***: Maintain the transaction's execution, even though the computer moves from cell to cell.
- ***Disconnection support***: The ability to execute transactions even when the mobile computer is disconnected, using caching techniques.
- ***Replication support***: Support the replication of information to have a lower communication cost.

Relating Transaction Execution

- ***Place of execution***: Some models execute transactions at the MSS, while others execute them at the MHs.
- ***Compensable transactions***: Compensable transactions allow the system to partially commit changes, because they can be later reverted if the whole transaction aborts.
- ***Conditions of execution***: Conditions to be evaluated before, during, and after the execution of the transaction. These conditions can be based on time, etc.
- ***Efficient concurrency handling***: In mobile environments, there is a higher probability that a transaction will become a long-lived transaction (LLT), so the concurrency must be handled efficiently, in order not to lock an object for a long time (for example, Field Call).

Relating the Model's Adaptability

- ***Take advantage of the geographical localization of the mobile unit***: A transaction could specify location-dependent conditions in order to modify its behavior. Furthermore, a handoff may not be always a significant event for the execution of a transaction.

- ***Final states***: Specify acceptable final states for the transaction.

Relating Implementation Issues

- ***Take advantage of mobile techniques***: Use techniques that take into account the mobile environment's characteristics.
- ***Take into account physical limitations***: The limitations imposed by mobile devices, such as low power processing, small memory, etc., should be taken into account.
- ***Take advantage of the particular characteristics of a transaction***: For example, whether a transaction is read-only or not, interactive or not, etc., to increase the system's flexibility and performance.

Comparison of Transactional Models

The mobile transactional models are compared here, using the previously defined framework. For each section in the framework definition, a comparison table is presented, and then an analysis is carried out.

Because mobility is one of the main characteristics to take into account (see Table 1), it is important that transactional models can continue the execution of a transaction even if the user moves from one cell to another (handoff). Disconnection support is also an important issue, because it allows the user to continue working while the MH is disconnected. Another useful characteristic, which can lower communication costs, is the ability to access replicated data, because in this way, a subtransaction can use data located in the local machine, instead of in a possibly-far-away centralized database.

The Kangaroo, Moflex, reporting/cotransaction, and prewrite models all support mobility, because, during a handoff, the original base station must migrate the transaction's state to the target MSS in order to follow the physical movement of the MH. On the other hand, none of these models, except prewrite, allow the user to continue working while the mobile computer is not connected to the network.

In the following table, a comparison of the models is shown, taking into account characteristics related to the execution of transactions. These transactions can be executed either in the MH, MSS, or both (if the model allows this). The decision as to where a transaction should be executed depends on the processing capacity of the computer, if the transaction is interactive or not, etc.

Although nearly all the models allow the definition of compensable transactions and, therefore, the sharing of information before the commitment of the whole transaction, there are differences in the way this facility is provided. The *reporting/cotransactions* model allows a transaction to share its partial results in an explicit way, while this is not possible in the other models.

As can be seen from this table, the Moflex model is the only one that allows the specification of transaction execution conditions.

All the studied models provide facilities for concurrency management, though each one does it in different ways: IOT, prewrite, and planned disconnections allow for the publishing of partial results within the mobile computer; weak/strict allows this at a cluster level; reporting/cotransactions allows for the explicit publishing of these results; the rest of the models expose their changes globally.

In Table 3, we describe those characteristics related to the adaptability of transactional models.

Table 1. Characteristics related to physical aspects.

Transactional Model	Mobility Support	Disconnection Support	Replication Support
Kangaroo	√	×	×
Moflex	√	×	×
IOT	×	√	×
Weak/strict in data clusters	×	√	√
Reporting/cotransa ons	√	×	×
Prewrite	√	√	×
Planned disconnections	×	√	√

Table 2. Characteristics related to the execution of the transactions

Transactional Model	Place of Execution	Compensating Transactions	Execution Conditions	Concurrency Handling
Kangaroo	MSS	√	×	√
Moflex	MSS	√	√	√
IOT	MH	×	×	√
Weak/strict in data clusters	MH	√	×	√
Reporting/ cotransactions	MSS	√	×	√
Prewrite	MH	×	×	√
Planned disconnections	MH	×	×	√

Because, by definition, in a mobile environment, a MH can roam through different geographical regions, it is important for the model to be able to adapt a transaction's execution to the changing conditions generated by this roaming. The main characteristic to consider as regards adaptability is the one related to geographical location, that is, the ability of a transaction to change its behavior depending on the position of the MH.

Here, Moflex is the only model that supports preconditions related to the geographical localization of the mobile computer and the specification of final acceptable states.

Table 3. Characteristics related to adaptability

Transactional Model	Geographical Localization	Acceptable Final States
Kangaroo	×	×
Moflex	√	√
IOT	×	×
Weak/strict in data clusters	×	×
Reporting/cotransa ons	×	×
Prewrite	×	×
Planned disconnections	×	×

Classification of Mobile Models

From all the issues analyzed so far, a classification of mobile models arises based on the presence or not of specific features. This classification allows us to pinpoint the existing voids and to highlight the aspects that deserve further work to reach a more capable architecture (Figure 1).

According to how deep the mobility support, the models can be categorized as follows:

- ***Total mobility***, reflects the ideal model, with both hand-off and disconnection support.
- ***Hand-off support*** provides cell-switching detection but cannot function if disconnected.
- ***Disconnection support*** is only able to work if disconnected and does not take into account the switching of cells.

A second difference is established between those adaptable models that allow for the establishment of locality conditions or nonadaptable conditions. From a third point of view, models allow transactions to be executed only at MSSs, only at MHs, or at both.

All models should exhibit efficient handling of concurrency and LLT because of the specific nature of mobile environments.

From this classification, we observe the lack of models that support hand-offs and disconnections simultaneously. The prewrite model is a special case near total mobility, because it always supports hand-off, but disconnection is supported when it has pre-read all data. If it disconnects from the network before the pre-read phase, it cannot continue the transaction.

Moreover, no current model allows for adapting the site of execution (i.e., MH, MSS, or combined). This analysis shows that the state-of-the-art is not optimal and points out several promising areas for future research (for example, the extension of current models or the definition of new ones).

CONCLUSIONS

In this work, we defined a framework to be used as a tool for studying transactional models that take maximum advantage of the peculiarities of mobile computing environments. We consider it a valuable tool not only to compare existing models, but also to validate new transactional models to come.

As a practical demonstration of its usefulness, we studied some transactional models applied to mobile environments, comparing them with the help of this tool.

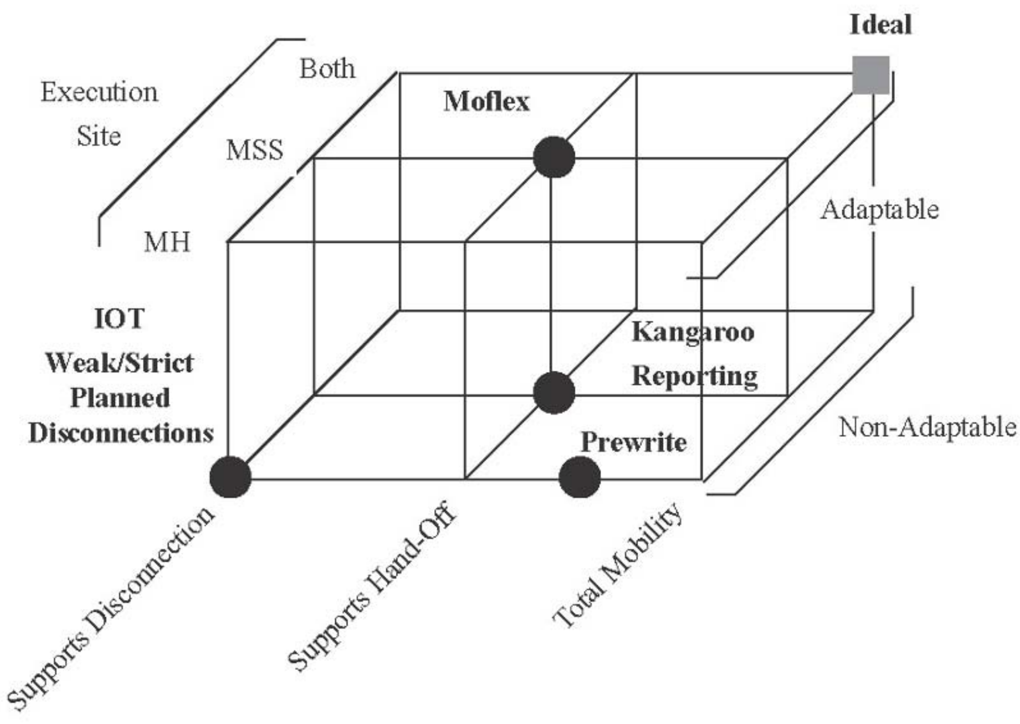
From this comparison and the classification made, we can say that none of the studied models covers all of the characteristics described in the framework. We consider that the essential items every transactional model should meet are *mobility support* and *disconnection support*. It can be seen that none of the models covers both characteristics (prewrite model supports a limited version of total mobility), which leads us to believe in the need to extend some of these models to support those items or to define a completely new mobile transactional model.

FUTURE TRENDS

We consider it necessary to create a flexible and dynamic simulator, so that arbitrary configurations can be easily defined and measured, in order to compare the different approaches of solving problems present in mobile environments. This will allow us, in a further step, to develop concrete implementations to be tested in real environments. To do this, existing implementations must be studied, this being a complex task given the recent definition and development of these systems.

Another possible future work is the definition of mobile models using ObjectZ, as proposed by Graeme (2004). This could lead to a more concrete and clear way of describing a model, rather than just writing a textual specification in a natural language. Besides, having a formal description could greatly enhance the study of these models before the actual simulation/implementation phase, including the possibility of applying model checking to foresee problems.

Figure 1. Classification of mobile models



REFERENCES

Chrysanthis, P. (1993). Transaction processing in a mobile computing environment. In *Proceedings of the IEEE Workshop on Advances in Parallel and Distributed Systems* (pp. 77-82).

Coratella, A., Felder, M., & Hirsch, R. (2003). A framework for analyzing mobile transaction models. In Advanced topics in database research (Vol. II, Chap. XII). Hershey, PA: Idea Group Publishing.

Dunham, M., Helal, A., & Balakrishnan, S. (1997). A mobile transaction model that captures both the data and movement behavior. *ACM Mob. Netw. Appl., 2*, 149-162.

Graeme, S. (2004). A formal framework for modelling and analysing mobile systems. In V. Estivill-Castro (Ed.), *Proceedings of the 27th Conference on Australasian Computer Science* (Vol. 26), (pp.193-202). Australian Computer Society, Dunedin, New Zealand.

Holliday, J., Agrawal, D., & El Abbadi, A. (2000). Planned disconnections for mobile databases. In *Proceedings of the 11th International Workshop on Database and Expert Systems Applications (DEXA '00)*, (pp.165-172), Greenwich, London, UK. IEEE Computer Society.

Ku, K., & Kim, Y. (1998). Moflex transaction model for mobile heterogeneous multidatabase systems. In *Proceedings of the 10th International Workshop on Research Issues in Data Engineering*, (pp.39-46), San Diego, CA. IEEE Computer Society.

Lee, J., & Simpson, K. (1997). A high performance transaction processing algorithm for mobile computing. *IASTED International Conference on Intelligent Information Systems (IIS '97)*, (pp.486-491), Grand Bahama Island, Bahamas. IEEE Computer Society.

Lu, Q., & Satyanarayanan, M. (1994). Isolation-only transactions for mobile computing. *ACM Operating Systems Review, 28*(2), 81-87.

Lu, Q., & Satyanarayanan, M. (1995). Improving data consistency in mobile computing using isolation-only transactions. In *Proceedings of the Fifth workshop on Hot Topics in Operating Systems*, (p.124), Washington, DC. IEEE Computer Society.

Madria, S. K., & Bhargava, B. (2001). A transaction model to improve data availability in mobile computing. *Distributed and Parallel Databases, 10*(2), 127-160.

Pitoura, E., & Bhargava, B. (1995). Maintaining consistency of data in mobile distributed environments. In *Proceedings of the 15th International Conference on Distributing Computing Systems*, (pp.404-413), Vancouver, Canada. IEEE Computer Society.

Tewari, R., & Grillo, P. (1995). Data management for mobile computing on the Internet. Computer & Information Center-Temple University, ACM, Philadelphia, PA, USA.

KEY TERMS

Handoff: Process that occurs after a cell change. The MH identifies itself at the MSS of the cell into which it has moved.

MH (Mobile Host): This is a mobile computer or device.

Mobile Computing: This is an environment made up of a set of devices that interact to give the user the possibility of staying connected while moving from one location to another.

Mobile Technique: This is an implementation technique used to support mobile transactional models, for example, techniques relating to geographic data usage, caching, concurrency handling, dissemination of information, etc.

MSS (Mobile Support Station): This refers to a computer with a wireless interface that offers information and services support to MHs.

Serializability: An execution is said to be serializable if its effect is the same as if the transactions were executed in serial order.

Transactional Model: This defines a way to handle transactions in a database environment, including using subtransactions, scope, limitations, etc.

Mobile Telecommunications and M-Commerce Applications

Clarence N.W. Tan
Bond University, Australia

Tiok-Woo Teo
Bond University, Australia

INTRODUCTION

This article presents an overview of prevailing trends and developments shaping mobile commerce (m-commerce) and the wireless economy. A review of wireless telecommunications infrastructure attempts to demystify the evolving technology landscape. Mobile Internet deployment and adoption demographics are highlighted, as are innovative wireless applications and current m-commerce issues.

BACKGROUND

The World Wide Web (WWW) and Web browser software brought mass market accessibility to the Internet. Riding on this ubiquity and reach is electronic commerce (e-commerce) in its many forms: inter-business dealing, intra-organization transactions and business-to-consumer trade, and so forth. E-commerce has witnessed impressive growth and continues to be a significant propellant of Internet progress. Participants have, however, hitherto been essentially tethered to fixed line connections. The development of novel wireless services and mobile adaptations of familiar applications (Ghini, 2000) is fueled by demand from increasingly nomadic users looking to access familiar online facilities, and the steady convergence of telecommunications and computing technologies (Messerschmitt, 1996).

Wireless telecommunications was conceived in the 1980s to carry voice, but has evolved to become data bearer, including Internet communications. The cellular telephone is now commonplace and more personal digital assistants (PDAs), hand-held computers and the like are sporting cellular radio connectivity. These devices form a sizable platform for deploying m-commerce applications. M-commerce refers to the ability to browse, interact with and make payment for goods and services directly from mobile terminals such as cell phones, PDAs and portable computers (Tan, 2002). Industry forecast statistics point to exponential growth in the sector:

- Worldwide shipment of Web-enabled wireless devices rose 796% in 2000 over 1999 and consumer transactions committed from such terminals will total US $1.8 trillion worldwide by 2005 (Source: Gartner Group).
- International wireless data market was expected to grow from 170 million to more than 1.3 billion subscribers between 2000–2004, equipping themselves with 1.5 billion wireless-capable handsets and other Internet appliances by end of 2004 (Source: Cahners In-Stat Group).
- Wireless Internet users in the Asia-Pacific region alone will rise 10-fold from 20 to 216.3 million between 2000–2007 (Source: Strategis Group).

As Internet and mobile communications converge, e-commerce evolves into m-commerce. The tremendous potential of “anytime” convenience and “anywhere” mobility in carrying out everyday online transactions will spur many unique mobile services yet.

TECHNOLOGY ROAD MAP

Early wireless telecommunications architecture in the late 1940s was modeled after television broadcasting. Tall, centralized transmitter towers provided radio coverage. Limitations like restricted user mobility and capacity, poor voice quality and high cost saw the introduction of new cellular technology in late 1970s—a superior architecture persisting to this day.

A cellular mobile communications system comprises a vast collective of low-power antenna subsystems, dispersed in small overlapping geographical units called cells. Individual cellular base stations provide local coverage and interconnect for a combined footprint that constitutes the wireless network. Modern implementations are typified by larger, sparse cells in rural areas and small, dense ones in metropolitan districts. The technology road map is demarcated by milestones corresponding to transmission bandwidth.

First Generation—1G: Analogue radio transmission characterized 1G cellular systems. The one-time de facto standard throughout the Americas and the Asia-Pacific was the Advanced Mobile Phone Service (AMPS) introduced in the United States in 1983. Despite technical imperfections such as limited growth capacity, poor data transport and deficient transmission security, 1G systems maintained their popularity till the early 1990s. Improved derivatives of AMPS are still deployed in the remaining analogue cellular networks around the world today.

Second Generation—2G: Digital radio transmission heralded the 2G era. Information is digitized into a stream of computer binary coded data packets for transmission and reassembly at the receiving end. Two competing digitization schemes are Time Division Multiple Access (TDMA) and Code Division Multiple Access (CDMA). Better bandwidth utilization boosts network capacity, enhances coverage and improves voice quality. New service features such as data encryption, short text messaging, fax and data transmission can also be offered.

Launched commercially in 1991, the European developed, TDMA-based Global System for Mobile Communications (GSM) is the de facto international 2G standard today, with 863.6 million subscribers in 174 countries at end of May 2003 (Source: GSM Association, *http://www.gsmworld.com*). An American adaptation of GSM called PCS 1900 was launched in late 1995. CDMA-based networks began commercial operation in 1995 and are now the predominant standard in the Americas, Korea and Japan, with 164.1 million subscribers in 60 countries as at June 2003 (Source: CDMA Development Group, *http://www.cdg.org*).

Conventional fixed line Internet connectivity offers varying data rates:

- Up to 56 kbps (kilobits per second): Analog modem
- 64–128 kbps: Integrated Services Digital Network (ISDN)
- 256 kbps–1.5 Mbps (megabits per second): Optical fiber and Digital Subscriber Line (DSL)

In comparison, 2G data service provides 9.6–57.6 kbps throughput, with most network operators supporting only speeds no more than 14.4 kbps. This makes for poor overall user experience for consumers increasingly accustomed to higher-speed, broadband Internet access through fixed connections. Real-time multimedia applications such as live video are also impracticable on 2G architecture. Solutions better optimized for wireless data transmission are clearly needed to meet growing demand for quality mobile Internet access.

Second-and-a-Half Generation—2.5G: 2.5G systems extend 2G infrastructures for upgraded data throughput. New handsets, however, are required to tap the improvements and other added functionalities. The enhancement to GSM is called the General Packet Radio Service (GPRS). Capable of a theoretical maximum 171.2 kbps transmission, the average is 40–80 kbps deployed in practice. Like the Internet, GPRS networks are also based on the Internet Protocol (IP) standard, so GPRS terminals function just like other wired Internet sub-nodes with seamless access to familiar applications such as the WWW, e-mail, Telnet and FTP (File Transfer Protocol), and so forth. The first commercial GPRS service was inaugurated in the United Kingdom in June 2000. 2.5G CDMA technology is known as the IS-95B standard and offers ISDN-like data rates. Its first commercial debut in Japan in early January 2000 beat GPRS to market by some 6 months, and the 64 kbps throughput was superior to prevailing 2G GSM offerings.

While data speed boost is useful in time-sensitive applications like online transactions and credit authorization, there remain technical idiosyncrasies and transmission latencies detrimental to time-critical multimedia functions such as video decoding and playback. Hence, 2.5G is still not the ideal platform for deploying high quality, real-time video conferencing.

"2.75G": When wireless data rates approach those of conventional fixed line broadband connectivity, user experience will improve significantly and truly mobile Internet access would be a reality. Some have termed this stage of the road map loosely as "2.75G".

The Enhanced Data rates for Global Evolution (EDGE) standard is engineered as an extension to GSM technology and leverages past investments in TDMA and GPRS. Throughput of 384–553.6 kbps is theoretically possible and early EDGE-compatible handsets support rates in excess of 100 kbps. The world's first commercial EDGE service was deployed in the United States in July 2003. The equivalent evolution in CDMA technology is the CDMA2000 standards, providing network speeds from 144–614 kbps. From the first commercial service in Korea in October 2000, CDMA2000 subscribers have grown to number 54.1 million by June 2003 (Source: CDMA Development Group).

Third Generation—3G: Wireless broadband connectivity and industry standards harmonization characterize 3G systems. New handsets are again necessary to exploit multimedia capabilities and applications enabled by high-performance wireless networking at unprecedented 384 kbps–2 Mbps speeds.

More importantly, the vision is for 3G devices to roam seamlessly on enabled networks within a country and across continents, creating truly borderless mobile services. To this end, the International Mobile Telecommunications 2000 (IMT-2000) recommendations were promulgated in late 1999 with international participation. IMT-2000 prescribes how mobile service providers should evolve existing cellular networks towards full inter-network compatibility independent of underlying radio technologies. Disparate TDMA, CDMA, GSM systems and their derivatives will be accommodated. For example, under IMT-2000, GSM will evolve into the Universal Mobile Telecommunications System (UMTS) and employ a new transmission technology called Wideband CDMA (W-CDMA). The first commercial UMTS 3G service was introduced in Japan in October 2001.

WIRELESS MESSAGING

SMS—Short Message Service: The Short Message Service is a wireless radio service for bidirectional transfer of alphanumeric messages of up to 160 characters each among mobile terminals in GSM and UMTS networks (European Telecommunications Standards Institute, 2002). Multiple text messages (up to 255) may be concatenated to form longer ones. Messaging is near instantaneous and operates on a store-and-forward scheme that guarantees delivery. Sent messages are routed to their destinations by an electronic intermediary in the cellular network called a Short Message Service Center, or SMSC (European Telecommunications Standards Institute, 2002). If the intended recipient is not immediately contactable, the message is stored for a stipulated time period to facilitate reattempted delivery. The SMSC can notify a message sender of the ultimate delivery success or failure via an SMS notice called a delivery report.

Introduced in 1991, SMS began as a supplement to voice calls, alerting subscribers to awaiting recorded messages. Its use has since grown to encompass notification services like paging alerts, new e-mail notice and calendar reminders; and information services like weather forecast, traffic watch, stock quotes, gaming results and horoscope readings, and so forth. Interpersonal communications still dominate SMS messaging volume by far today but SMS is also finding increasingly sophisticated uses in corporate applications such as job dispatch (Tan, 2002) and secure m-commerce transactions systems (Tan, 2003).

In 2002, an estimated 360 billion SMS messages were sent worldwide with some 30 billion sent monthly in December—double the rate 24 months ago (Source: GSM Association). The growth trend is expected to continue.

EMS—Enhanced Messaging Service: EMS is a minor extension to SMS whereby a mix of formatted text (e.g., bold, italics, etc.), simple animations, tiny pictures and short melodies can be included as message content. The specification provides for amalgamating multiple classic SMS messages to convey the increased EMS message complexity. Compatible handsets have been commercially available since 2001, but there has been a lack of full interoperability among different makes and models due to proprietary implementations of the EMS standard by manufacturers.

MMS—Multimedia Messaging Service: MMS is the latest generation wireless messaging standard that inherits the SMS store-and-forward schema and features improved security mechanisms for managing message encryption, authentication and privacy. MMS is devised to carry a full range of multimedia elements including text, audio clips, still pictures, animated images and full-motion video, all encoded using industry standard file formats (e.g., JPEG for photographs, MP3 for audio, MPEG for video, etc.). Such rich content increases message size and considerably more bandwidth is needed to deliver MMS messages expeditiously. Hence, the takeoff of MMS is only expected to dovetail on the rollout of 3G networks internationally. The world's first commercial MMS service was introduced in Hungary in April 2002.

INNOVATIVE APPLICATIONS

The sheer success of the mobile telephone as a consumer device is indisputable. Mobile phones have already outstripped conventional telephone subscriptions in many countries. As cellular telecommunications networks and the Internet converge, the spillover of Internet e-commerce will fuel m-commerce takeoff. Wireless broadband technology can be used to deploy not only mobile variants of existing Web-based applications, but also fundamentally new ones. This creates new value proposition for consumers and revenue streams for vendors. The true power of the mobile Internet lies in these innovative applications that will be spawned, some of which are listed next.

- **Location-based services (LBS):** LBS refers to the provisioning of value-added services to cellular subscribers based on the physical location of their mobile devices within the network. Potential applications: location-based call billing, emergency medical services, courier fleet tracking, proximity-based personalized marketing, locality information services like navigation and directory assistance, and so forth.
- **Mobile multimedia entertainment:** Higher bandwidth afforded by 2.5G and 3G networks creates new distribution channels for paid multimedia entertainment like computer games, music and video. For example, games can be downloaded wirelessly for play on enabled handsets or with other gamers on the network. Music and video vendors may stream

content wirelessly on demand to mobile devices for instantaneous preview playback and subsequent purchase.

- **Wireless telemetry:** The facility for gathering data or remotely controlling devices over the footprint of a cellular network through two-way communication between remote equipment and a central facility. Potential applications: security alarm monitoring, climate control, meter readings and inventory status polling, and so forth.
- **Wireless electronic payment systems:** Cellular phones become secure, self-contained tools for instantly authorizing payment for goods and services wirelessly over the network. One example is the use of SMS messages for making purchases at vending machines.
- **Telematics:** The integration of wireless telecommunications and in-vehicular monitoring and location systems will bring mobile Internet facilities, location-based services and multimedia entertainment to vehicle dashboards. Potential applications: online information services (e.g., traffic news, weather forecast, route assistance, etc.), target advertising based on vehicle locale and streaming media content, and so forth.
- **Wireless telemedicine:** Cellular telecommunications and wireless broadband multimedia technologies can deliver health care expertise and services remotely. Potential applications: remote monitoring of patients in ambulances and long-distant specialist teleconsultation for patients in remote regions.

CURRENT ISSUES

The realm of m-commerce is still in its infancy and to the extent that it bears roots in e-commerce, many technical, business, marketing and legal perspectives in m-commerce may be extrapolated from e-commerce ones. But as a fundamentally new business paradigm, m-commerce will elicit contentious issues of its own. A few have been highlighted as follows.

- **Wireless privacy:** Industry self-regulation and legislation are needed to guard against potential threat of undue surveillance and profiling, invasion to personal privacy and wireless spam resulting from abuse of location data gathered for provisioning LBS.
- **Wireless security:** Increased vigilance is necessary as computer viruses and like malicious attacks begin to target PDAs and mobile phones, as such devices gain more processing power and intelligence, and users become ever more reliant on them for voice and data services.
- **Wireless emission and public health**: Concerns persist on the long-term impact on public health from cellular emission, radiation and electromagnetic interference in the face of rapid consumer adoption of mobile phones. While no conclusive evidence exists to suggest significant risks, a more precautionary approach is prudent pending further research.

CONCLUSION

Mobile telecommunications have seen an evolution from a voice focus to data-centricity. For now, 3G networks present a viable channel for distributing practically any digital content. The future along the technology roadmap continues to emphasize quality multimedia delivery. Cellular devices are expected to improve in form and function, benefiting from advances in display, storage, imaging, processing and battery technologies. Historically, new handsets were required for each successive generation of wireless infrastructure and their timely volume availability at mass market prices remains crucial to the success of future networks.

The continued proliferation of mobile-enabled portable consumer electronics and the rising demand for ubiquitous computing will contribute to the volume and value of m-commerce, and begin to unleash the power of the mobile Internet.

REFERENCES

European Telecommunications Standards Institute. (2002). *ETSI TS 122 105 V5.2.0—UMTS Services and Service Capabilities.* France.

European Telecommunications Standards Institute. (2002). *ETSI TS 123 040 V5.5.1—Technical Realization of Short Message Service (SMS).* France.

Ghini, V., Pau, G., & Salomoni, P. (2000). Integrating notification services in computer network and mobile telephony. *Proceedings of the 2000 ACM Symposium on Applied Computing*.

Messerschmitt, D.G. (1996). The convergence of telecommunications and computing: What are the implications today? *Proceedings of the IEEE, 84*(1), 1167–1186.

Tan, C.N.W., & Teo, T.W. (2002). A short message service (SMS) enabled job dispatch system. In C.-H. Yeh & S. Tekinay (Eds.), *Proceedings of the International*

Conference on Wireless Networks 2002 (pp. 346-352). Pennsylvania: CSREA Press.

Tan, C.N.W., & Teo, T.W. (2002). From e-commerce to m-commerce: The power of the mobile Internet. In J.D. Haynes (Ed.), *Internet management issues: A global perspective* (pp. 27-53). Hershey, PA: Idea Group Publishing.

Tan, C.N.W., & Teo, T.W. (2003). An authenticated short message service (SMS)-based transactions system without SIM modification. In W. Zhuang, C.-H. Yeh, O. Droegehorn, C.K. Toh & H.R. Arabnia (Eds.), *Proceedings of the International Conference on Wireless Networks 2003*. Pennsylvania: CSREA Press.

KEY WORDS

LBS (Location-Based Services): The provisioning of value-added services to cellular subscribers based on the physical location of their mobile devices within the network.

M-Commerce (Mobile Commerce): The ability to browse, interact with and make payment for goods and services directly from mobile terminals such as cell phones, personal digital assistants (PDAs) and portable computers.

SMS (Short Message Service): A wireless service for bidirectional transfer of alphanumeric messages among mobile terminals in digital cellular networks.

WAP (Wireless Application Protocol): An open, extensible and unified global standard for delivering information and providing interactive services over cellular networks to mobile devices.

Modeling ERP Academic Deployment via AST

M

Cynthia LeRouge
Saint Louis University, USA

Harold Webb
The University of Tampa, USA

INTRODUCTION

Many educators and industry leaders believe industry-academic collaboration can change learning processes and improve learning outcomes by providing opportunities and resources not otherwise attainable (Sikkel, Spil, & van de Weg, 1999; Wohlin & Regnell, 1999). In the case of deploying advanced information technologies (AITs; e.g., computer-aided software engineering tools, enterprise resource planning systems, and database tools) via industry-academic collaboration, each institution's culture and unique organizational objectives influence collaborative relationships. The challenge in these relationships is to facilitate mutual understanding and clarify distinctions in addressing each organization's goals. The aim of these relationships is the appropriation of AITs in a manner that enriches educational experiences while meeting industry needs.

There are many quandaries associated with this phenomenon. How does the deployment of AITs within the educational curriculum facilitate the educational process? To what degree should these resources be utilized in education? What tools and methods should be used to deploy these resources? What should the role of the AIT vendor or consultant be? Can academic independence be maintained within a strong industry collaborative relationship?

Without a framework that identifies relevant variables, it is daunting at best to begin to assess the impact of varying degrees of adoption, identify effective processes of deployment, and move toward assessing costs and benefits. Though some frameworks address industry-academic collaboration (Mead, Beckman, Lawrence, O'Mary, Parish, Unpingco, et al., 1999), none have considered the implications of AITs on the evolution of interinstitutional collaborative relationships. This exposition discusses a framework seeking to address this need. The framework proposes a theoretical model for understanding the forces at work when integrating AITs into educational settings (LeRouge & Webb, 2002).

We begin our discussion by introducing adaptive structuration theory as the foundation for the industry-academic AIT collaboration framework (DeSanctis & Poole, 1994). We then introduce constructs and relationships related to industry-academic collaboration within the context of enterprise resource planning (ERP) systems.

BACKGROUND: THEORETICAL FRAMEWORK - ADAPTIVE STRUCTURATION

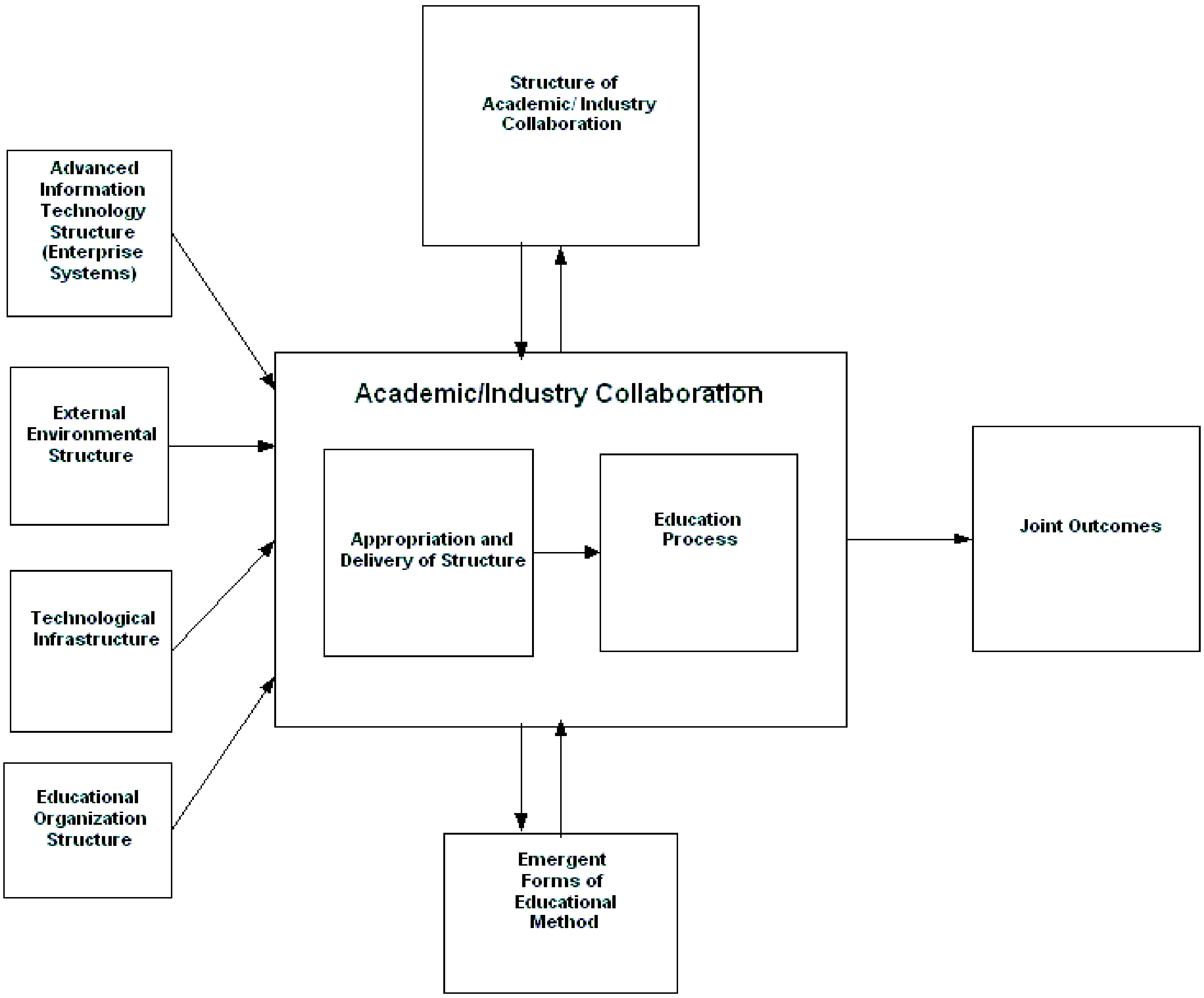
Adaptive structuration theory (AST), an extension of structuration theory, has been used as a framework to study organizational change processes when advanced technologies are adopted. Adaptive structuration is a sociotechnical perspective that introduces human actors and organizational context as moderators of the impacts of technology (Gopal, Bostrom, & Chin, 1993; Poole & DeSanctis, 1992). This perspective posits that the adoption of an advanced technology is a process of organizational change that results from the mutual influence of technology and social processes.

It is our premise that in an academic setting, human actors and organizational context are moderators of the collaborative process in which AITs are appropriated as each entity strives to fulfill its own needs. Such dynamic processes have an effect not only on the outcome of the appropriation, but also on the evolution of the relationship between industry and academia. As the number of academic institutions adopting AITs increase, we feel AITs will serve some groups better than others.

We introduce the framework's components in the context of ERP systems adoption in academe. We exemplify ERP systems within the specific context of colleges of business for a number of reasons: market demand, level of commitment required, interdisciplinary functionality, and level of system sophistication. The level of commitment and potential impact implied by the magnitude of ERP systems suggests that the study of ERP systems in business colleges may be associated with a considerable effect size detectable in future research efforts.

Our model of the constructs and relationships is represented in Figure 1. Model constructs and attributes are summarized in Table 1.

Figure 1. Adaptive structuration theory applied to industry-academic collaborations involving AITs (adapted from DeSanctis and Poole, 1994; LeRouge & Webb, 2002)



ADVANCED INFORMATION TECHNOLOGY STRUCTURE

Two ways have been suggested to describe the contributing social structures offered by an AIT. The first, "structural features," refers to the types of rules and resources offered by the system. The second, "spirit," can be thought of as the intended purpose and utilization of the system (DeSanctis & Poole, 1994).

With respect to technology spirit, there is a natural gap in appropriation between business use and academic use of ERP systems. The spirit of ERP systems can be described as information technology structures designed to integrate operational-level transaction processing, multilevel decision support, and strategic management of major corporations. However, the goals of ERP use in colleges of business are primarily educational and exploratory in nature, often focusing on discipline-related subsystems. This gap in system goals and values may have implications for academic-industry collaboration.

With regard to structural features, ERPs are a comprehensive suite of applications that support diverse organizational functions and processes. The suites developed by major ERP vendors commonly contain more than 15 modules, each of which is geared toward a different industry process and requires a significant training investment. For example, SAP offers some certification programs requiring over 40 days of classroom training. Accordingly, ERP systems challenge colleges with a level of sophistication that affords flexibility in feature set and appropriation, but requires substantial training to achieve acceptable performance.

Table 1. Constructs of interest (LeRouge & Webb, 2002)

Construct	Attributes
Advanced Information Technology Structure (Enterprise Systems)	♦ Structural features (restrictiveness, level of sophistication, comprehensiveness [suite of applications supporting diverse functions], learning curve) ♦ Spirit (enterprise-wide strategy, daily operations, management support system, strategic planning tool, convergence with educational goals)
External Environmental Structure	♦ Accreditation standards and curriculum studies ♦ Technology-vendor market position and status ♦ Industry standards and technology trends ♦ Technology market competition and end-user demands ♦ Technology-enabled labor supply
Technology Infrastructure	♦ Software and hardware ♦ Internal maintenance and software support ♦ Database creation and maintenance ♦ Computer lab facility and student remote access ♦ Industry donation or grants to support technology infrastructure
Educational Organization Structure	♦ Departmental structure ♦ Major program requirements ♦ Course objectives ♦ Instructor preferences
Appropriation and Delivery Structure	♦ Appropriation moves (direct use, relation to other structures, interpretation of structures, or judgment of features) ♦ Faithfulness ♦ Instrumental uses ♦ Attitude
Education Process	♦ Learning models (collaborative learning, hands-on experience, simulations, conceptual presentations, programmed instruction, real-world exposure, case studies) ♦ Supporting technologies (textbooks on technology, presentation tools, asynchronous communication tools, synchronous communication tools, computer-based training modules)
Emergent Forms of Educational Method	♦ Educators enrolling in corporate training programs ♦ Project- or task-specific internships ♦ Industry experts participating in classroom presentation ♦ Students and educators participating in AIT-specific list serves ♦ Credit for and/or increased access to technology training programs for students ♦ Industry development of targeted educational tools, databases, and exercises
Joint Outcomes	♦ Student learning and education in technology arena ♦ Increased work pool and employable students ♦ AIT market exposure ♦ Contribution to industrial research and development effort and academic research ♦ Continued and enhanced program attractiveness
Structure of Academic-Industry Collaboration	♦ Rules (industry participation in curriculum development studies, inclusion of AIT in curriculum development research, academic participation in industry development, educator participation in corporate training programs) ♦ Resources (technology alliance programs, opportunities and invitations for field research)

EXTERNAL SOURCES OF STRUCTURE

A report of the American Accounting Association determined that "the markets for accounting students will drive the 'models' of education, including delivery, content and pedagogy" (Lea, Berlin, Butterwick, Carpenter, Flaherty, Glazerman, et al., 1998). However, ERP systems have found new avenues of opportunity to continue past growth through such channels as e-commerce and service-firm-oriented modules (Stedman, 2000; Wilson, 2000). As such, the ERP market seems strong as is the demand for graduates to work with such systems (Watson & Schneider, 1999). Examples of sources of structure are detailed curriculum guides for computer science (Lidtke, Stokes, Haines, & Mulder, 1999) and information systems (Gorgone & Gray, 2000), as well as documented industry needs for trained graduates and focused ERP research that are driving vendor alliance programs (SAP, Inc., 2001).

TECHNOLOGY INFRASTRUCTURE

A practical component in any study of the appropriation of AITs is a basic cost-benefit analysis. Are the total costs and commitments in deploying these systems rewarded with an enriched educational experience and increased marketability?

Technology infrastructure can be a major cost consideration. ERP systems are large-scale AITs requiring a significant commitment of academic resources (Becerra-Fernandez , Murphy, & Simon, 2000; Watson & Schneider, 1999). These systems typically cannot be deployed or maintained without support from ERP vendors or consultants. Industry may facilitate appropriation through donated services; however, colleges may face additional costs. Hence, some form of collaborative relationship seems a precursor for the existence of these systems in academic settings.

EDUCATIONAL ORGANIZATION STRUCTURE

The philosophies of appropriating ERP systems among educational entities vary widely. The overall philosophic quandary involves balancing conceptual technology education and the development of technology skills. ERP exposure may be used to emphasize the informational aspects of various topics via conceptual illustration rather than dwelling on the technology. Conversely, ERP systems can be used to create the learning environment necessary for experiential learning where students work through business problems via scenarios using a system running training data. The degree of centralization and support in an academic department and the degree of coupling existing between courses included within the curriculum can affect the adoption of ERP systems.

This educational organization structure also impacts the appropriation of technology and the collaborative relationship. If constraints mandate deploying only one ERP system within the curriculum, could the college be transmitting a product endorsement message that constrains the scope of the curriculum? Are there positive or negative connotations, from a student or industry perspective, for a college of business known as an Oracle, SAP, or Sybase "shop"? Accordingly, do companies place any implicit or direct pressure on colleges to exclusively employ their software?

APPROPRIATION AND EDUCATION PROCESS

Structuration has at its core motivated and practical actions. Rules and resources embodied in social institutions are appropriated by participants and enter into the production and reproduction of the social system (Poole & DeSanctis, 1992). Industry-academic collaborative interaction is embodied in the appropriation of the AIT into the educational process.

Educators determine the curriculum strategy and the degree of appropriation for these mammoth ERP systems. The degree of appropriation has been addressed by academic institutions in a variety of ways ranging from inclusion of exemplary material within courses to new course creation, to establishing new degree programs.

The appropriation structure leads to choices regarding strategies for integrating the learning model(s) and supporting technologies within an instructional design. ERP systems are adaptable to many models of learning depending upon the educator's objectives. The model of learning chosen may affect the instructor's approach and utilization of these tools. To illustrate, conceptual presentation and demonstration may adequately support knowledge-level learning objectives, while experiential learning models may better support higher order learning objectives (Leidner & Jarvenpaa, 1995). One example of educational process is the case-based learning approach for ERP system development in undergraduate classes (Stewart & Rosemann, 2001), while another is the integration of vendor training materials into an existing college course (SAP, Inc., 2001).

EMERGENT FORMS OF EDUCATIONAL METHOD

Educational approaches discussed in the research on collaboration among software engineering programs and vendors that potentially influence collaboration with ERP system vendors include the following.

- Accelerated internship programs providing hands-on experiences (Powell, Diaz-Herrera, & Turner, 1997; Wohlin & Regnell, 1999)
- Independent study support by vendors and institutions
- Industrially experienced teachers (Wohlin & Regnell, 1999)
- Educational credit for vendor-sponsored training programs (Beckman, Coulter, & Mead, 1997)
- Jointly developed courses (Wohlin & Regnell, 1999).
- Participation in upgrade and maintenance efforts
- Case studies inspired by industry (Dawson & Newsham, 1997; Sikkel et al., 1999)
- Research within industry settings (Wohlin & Regnell, 1999)

An example of emergent forms of ERP-related education is the use of problem-based learning in which graduate students work with industry partners to design ERP solutions to solve a business problem. The completed solutions are then used as the basis for the development of teaching cases to support learning in undergraduate ERP-related courses (Stewart & Rosemann, 2001).

JOINT OUTCOMES

The purpose of ERP appropriation and industry-academic collaboration is to achieve mutually beneficial joint outcomes for both academia and industry (SAP, Inc., 2001; Stewart & Rosemann, 2001). The desired joint outcomes may include facilitating the educational mission, gaining competitive advantage, accessing educational resources, enhancing reputation, increasing revenue, and providing a staffing source (Mead et al., 1999). The academic institution, industry, or both may desire each of these goals. However, just as success is not guaranteed in implementing these systems in an industry setting, desired joint outcomes from academic appropriation are not guaranteed.

STRUCTURE OF ACADEMIC-INDUSTRY COLLABORATION

The collaborative system is not a recognized organization, but a structured social practice of interdependence that has broad spatial and temporal extension (Giddens, 1982). There are no global, formalized standards for the appropriation of ERP systems in educational settings or specifications for industry's participatory role in this phenomenon. Though some industry alliance programs may require agreements regarding the fees, donations, or support of AITs, educational methods and processes are guided by practice and choice (SAP, Inc., 2001). Certain practices, or implied dynamic rules, are embodied in the collaborative relationship created between industry and academia. The existence of ERP alliance programs may be considered a representation of social practices affecting the structure of academic-industry collaboration. Representations of the implied social practice may also be found in such documents as curriculum guides that encourage industry participation and the study of ERP system concepts in academic programs (Gorgone & Gray, 2000; Lidtke & Stokes, 1999; Lidtke et al. 1999).

Practices suggested for a successful collaboration based upon case study experience (Powell et al., 1997) include the following.

- Central coordination point
- Right mix of knowledge and experience
- Cooperative planning and scheduling
- Flexibility to change curriculum
- Communication to build teamwork
- Balance to maintain objectivity

Mead, Unpingco, Beckman, Walker, Parish, and O'Mary (2000) provide examples of how the academic-industry collaboration process could work. Based on a study of 21 collaborations, this article discusses four alternative collaboration models in use, varying governance structures observed, and the need to establish metrics to monitor and modify the collaborative process.

FUTURE TRENDS

The appropriation of AITs in colleges of business is a modern phenomenon that aspires to bridge the industry-academic gap while fulfilling educational goals. From a practical perspective, industry and consumers of academic education often desire opportune, practical training and education from academic institutions (SAP, Inc., 2001; Stewart & Rosemann, 2001). However, the costs of appropriation may be high and the impact of appropriation on educational processes and collaborative relationships may be either minor or dramatic. Thus, the joint outcomes of AIT appropriations and collaboration are uncertain. Stakeholders should recognize the potential influence of structure and social context on desired outcomes when embarking on the process of industry-academic collaboration. AIT appropriation and associated collaboration decisions may affect the educational foundation and career prospects of the technological workforce. These changes are already underway in both industry and academe (SAP Inc.; Stewart & Rosemann). As such, further research is needed to investigate relationships among the constructs and attributes identified in the framework and their effect on both academe and industry.

CONCLUSION

Adaptive structuration recognizes that technology appropriation may be a key factor in the evolution of affected social structures. Adaptive structuration theory gains predictive and descriptive power by identifying social complexities and relationships associated with the adoption of advanced information technologies. The major contribution of this exposition is the extension of adaptive structuration theory to better understand the dynamics of industry-academic collaboration as they evolve on campus and in the workplace.

REFERENCES

Becerra-Fernandez, I., Murphy, K., & Simon, S. J. (2000). Integrating enterprise in the business school curriculum. *Communications of the ACM, 43*(4), 1-4.

Beckman, K., Coulter, N., Khajenoori, S., & Mead, N. R. (1997). Collaborations: Closing the industry-academia gap. *IEEE Software, 14*(6), 49-57.

Dawson, R., & Newsham, R. (1997). Introducing software engineers to the real world. *IEEE Software, 14*(6), 37-43.

DeSanctis, G., & Poole, M. S. (1994). Capturing the complexity in advanced technology use: Adaptive structuration theory. *Organizational Science, 5*(2), 121-147.

Giddens, A. (1982). *Profiles and critiques in social theory*. Berkeley, CA: University of California Press.

Gopal, A., Bostrom, R., & Chin, R. P. (1993). Applying adaptive structuration theory to investigate the process of group support system use. *Journal of Management Information Systems, 9*(3), 45-63.

Gorgone, J. T., & Gray, P. (2000). MSIS 2000 model curriculum and guidelines for graduate degree programs in information systems. *Communication of the AIS, 3*(1), 1.

Lea, R. B., Berlin, S., Butterwick, B., Carpenter, C. G., Flaherty, R. E., Glazerman, E. J., et al. (1998). *The future viability of AAA members' programs*: *Report of the changing environment committee*. Retrieved April 14, 2004, from American Accounting Association Web site: http://aaahq.org/about/reports/chngenv98.htm

Leidner, D. E., & Jarvenpaa, S. L. (1995). The use of information technology to enhance management school education: A theoretical view. *MIS Quarterly, 19*(3), 265-292.

LeRouge, C., & Webb, H. W. (2002). Theoretical foundations for enterprise systems technology collaborations: An adaptive structuration framework. In J. Lazar (Ed.), *Managing IT/community partnerships in the 21st century* (pp. 178-203). Hershey, PA: Idea Group Publishing.

Lidtke, D. K., & Stokes, G. E. (1999). An information systems-centric curriculum '99. *The Journal of Systems and Software, 49*(2, 3), 171-176.

Lidtke, D. K., Stokes, G. E., Haines, J., & Mulder, M. C. (1999). An information systems-centric curriculum '99 program guidelines for educating the next generations of information systems specialists in collaboration with industry. *National Science Foundation Report*. Retrieved April 14, 2004, from University of Nebraska at Omaha Information Systems-Centric Curriculum Web site: http://www.iscc.unomaha.edu/

Mead, N., Beckman, K., Lawrence, J., O'Mary, G., Parish, C., Unpingco, P., et al. (1999). Industry/university collaborations: Different perspectives heighten mutual opportunities. *The Journal of Systems and Software, 49*(2, 3), 155-162.

Mead, N., Unpingco, P., Beckman, K., Walker, H., Parish, C. L., & O'Mary, G. (2000, May). Industry/university collaborations. *Crosstalk: The Journal of Defense Software Engineering.* Retrieved April 7, 2004, from http://www.stsc.hill.af.mil/crosstalk/2000/03/mead.html

Poole, M. S., & DeSanctis, G. (1992). Micro level structuration in computer-supported group decision making. *Human Communication Research, 19*(1), 5-49.

Powell, G. M. J., Diaz-Herrera, L., & Turner, D. J. (1997). Achieving synergy in collaborative education. *IEEE Software, 14*(6), 58-65.

SAP, Inc. (2001, September). *What's new with the SAP education alliance program in the U.S.* Retrieved April 14, 2004, from http://www.sap.com/usa/company/alliances/pdf/university_program.pdf

Sikkel, K. T., Spil, A. M., & van de Weg, R. L. W. (1999). A real-world case study in information technology for undergraduate students. *The Journal of Systems and Software, 49*(2, 3), 117-123.

Stedman, C. (2000, January 3). Expand ERP beyond the back office. *Computerworld, 34*(1), 13-16.

Stewart, G., & Rosemann, M. (2001). Industry-oriented design of ERP-related curriculum: An Australian initiative. *Business Process Management, 7*(3), 234-242.

Watson, E. E., & Schneider, H. (1999). Using enterprise systems in education. *Communications of the Association for Information Systems, 1*(9), 1-48.

Wilson, T. (2000, January 17). Handing off the burden: Customers gain outsourcing options for J. D. Edwards ERP packages. *Internetweek*, 13.

Wohlin, C., & Regnell, B. (1999). Strategies for industrial relevance in software engineering education. *The Journal of Systems and Software, 49*(2, 3), 124-134.

KEY TERMS

Advanced Information Technology Structure: The rules and resources offered by systems such as computer-aided software engineering tools, enterprise resource planning systems, and database management systems that support the intended purposes and utilization of those systems.

Appropriation and Delivery Structure: The rules and resources that determine the choices made by educators regarding strategies for integrating the learning model(s) and supporting technologies within a selected instructional design.

External Environmental Structure: The rules and resources offered by outside interests including academic standards bodies, technology developers and vendors, industrial organizations, employers, and end users.

Educational Organization Structure: The rules and resources offered by the internal educational institution that are derived from how it is organized as well as program requirements, curriculum, and course objectives.

Education Process: The use of learning models and supporting learning technologies to deliver the learning experience and/or training with students.

Emergent Forms of Educational Method: New methods and techniques that may be employed or used in the education process related to the deployment of AIT in academe.

Joint Outcomes: The direct output and by-products of the education process including student learning, an employable work force, AIT market exposure, and contributions to industrial and academic research.

Structure of Academic-Industry Collaboration: The representation of social practices among the stakeholders affecting academic-industry collaboration that result in the establishment of rules of practice and the provision of resources.

Technology Infrastructure: The required supporting activities and facilities including network, hardware, software, development, and maintenance.

Modeling for E-Learning Systems

Maria Alexandra Rentroia-Bonito
Instituto Superior Técnico/Technical University of Lisbon, Portugal

Joaquim Armando Pires Jorge
Instituto Superior Técnico/Technical University of Lisbon, Portugal

INTRODUCTION

Computer-based instruction is touted as an effective tool to support knowledge dissemination within predefined learning environments. Indeed, many see it as a way to overcome geographical or social barriers to knowledge transmission and educational institutions. However, its domain of application has traditionally been restricted to basic skills and educational contexts. Recently, dynamic and complex business environments shaped by technological changes and the downsizing trend of the '90s placed new constraints on the underlying assumptions (Fuglseth, 2003). Organizations are now pushing for skill flexibility, demanding specialized knowledge and requiring faster learning curves from employees. Many advocate Internet-based education materials as one way to meet those challenges (Bernardes & O'Donoghue, 2003; Karoulis et al., 2004; Storey et al., 2002; Strazzo & Wentling, 2001). However, this raises important questions concerning both effectiveness and efficiency of such tools and materials. Indeed, developing interactive multimedia-based courseware remains pretty much a black art, consuming enormous resources. So far, there is a lack of established models to predict the performance and evaluate how adequately courseware can meet user needs. In fact, developing courseware should take into account the target constituency requirements, organizational context, and the stated educational or training goals. Developing the wrong training materials can lead to costly investments in creating and maintaining content to match the increasing expectations on e-learning. Perhaps this can explain the recent rash of failed e-learning projects—current results do not measure up to business and individual expectations yet.

A better understanding of the many factors affecting e-learning performance would allow individuals and organizations to achieve the expected benefits. In so doing, development teams need methods, techniques, and tools to evaluate in advance which features are needed to achieve higher outcomes, namely, performance and satisfaction. Thus, the need to develop predictive models to improve learning effectiveness is in order.

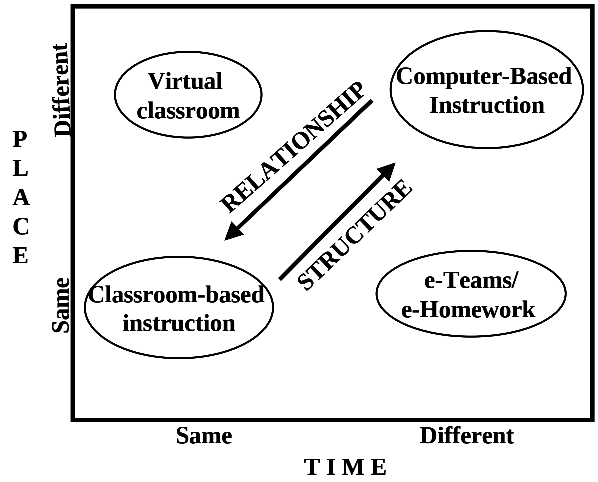
This overview includes four sections. "Background" presents a proposed e-learning theoretical framework to guide our analysis based upon the reviewed literature. "Key Issues" section describes main issues arising from the proposed e-learning conceptual framework. "Future Trends" describes our vision on how to approach e-learning initiatives and future trends. Finally, we present a general conclusion.

BACKGROUND

Organizational investment in e-learning strategies reflects strategic choices regarding skill development through e-learning. According to Wentling, Waight et al. (2000), e-learning involves acquiring and using distributed knowledge facilitated by electronic means in synchronous or asynchronous modes. As shown in Figure 1, knowledge could be distributed geographically within varying time frames.

Thus, the effective use of technology-based instruction would provide to organizations the ability to succeed at operational levels. This justifies the adoption of a holistic approach to courseware evaluation as a diagnos-

Figure 1. Proposed types of e-learning in terms of time and place



tic and managerial tool. We propose a framework, shown in Figure 2, which comprises three basic entities, business processes, people, and information systems, and three main relationships: (a) interaction between people and systems, (b) process-based roles played by people during this interaction, and (c) having the learning task be executed, as part of the e-learning experience, by people performing their process-based roles. This framework could lead to working techniques and approaches that assist development team members in designing work-related e-learning experiences within organizational contexts. To motivate a workable approach, we will now discuss each of these entities and relationships.

Reviewed literature strongly suggests that the external and internal fit among business strategies, culture, human resource practices, and leadership styles is critical to worker performance. Moreover, work contexts, for example, physical and technological conditions surrounding individual tasks, affect people's perceptions and, in turn, influence their motivation to engage into and perform learning tasks (Astleitner, 2001; Bandura, 2000; Chen, 2002; Dix et al., 1998; Kim, 2000; Liu & Dean, 1999; Reeves & Nass, 1996; Strazzo & Wentling, 2001; Vouk et al., 1999; Welbourne et al., 2000; Wentling et al., 2000).

Within the e-learning experience, business processes provide yardsticks to define educational or training goals and monitor outcomes. However, we need also to consider the roles people perform when interacting with courseware. Such process-based roles could be as diverse as e-learners, e-instructors, e-speakers, systems and courseware designers, supervisors, reviewers, human resource managers, and information technology officers among many others.

Human-computer interaction can model parts of the e-learning experience in accordance with Norman's extended model (Dix et al., 1998). Furthermore, the experience is also shaped by the way people relate to systems. This is supported by Reeves' and Nass' (1996) work, which suggests that people relate to media as they would relate to real people, treating them with affection and courtesy. Building on these findings, we argue that the more e-learning systems themselves are easy to use and learn and are "nicely behaved," the likelier e-learners will engage in the experience and profit from their outcomes.

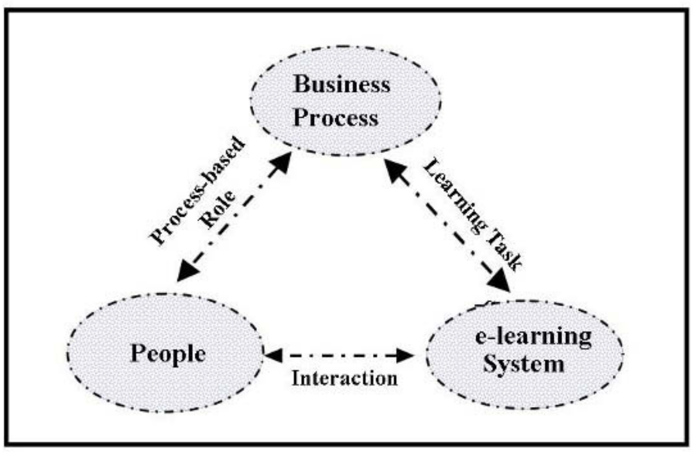
The interplay among these three relationships (process-based role, learning task, and interaction) relates to a just-in-time learning concept. Strategic knowledge acquisition should be enmeshed in current activities to support employees in learning new skills when performing day-to-day business tasks. We believe this concept can foster gradual alignment between learning outcomes, and technology with strategic aspects of business.

KEY ISSUES

We identify structure and relationship as the main issues within our framework as presented in the previous section. Figure 1 shows different modes of e-learning regarding the use of technology in education, both in terms of distance and time. As technology gets more extensively used for delivery, the need for course structure becomes higher and the relationship between instructor and e-learner turns increasingly weaker. Figure 1 also shows this relationship as defining three types of e-learning, which are set apart from conventional classroom instruction.

This shows that using technology to support learning requires higher course structure than traditional classroom-based instruction to be effective (Karoulis et al., 2004; Liu & Dean, 1999). However, current approaches take a one-size-fits-all method to provide courseware

Figure 2. Proposed e-learning framework



delivery regardless of differences in place and time. We cannot argue strongly enough that delivery needs to be tailored to context (space and time) to overcome the barriers imposed by structure and to improve the e-learning experience. This should be done differently for different students with diverse cognitive styles, roles, and tasks within organizational contexts. We will now discuss factors affecting structure and relationship.

As for structure, organizations identify training needs taking into account work context, business process dynamics, individual tasks, objectives, and areas for performance improvement. A business-process approach driving the design and the development of interactive course contents should focus on skill gaps to define instructional objectives in order to meet performance standards. In this way, setting up appropriate goals for training evaluation poses the same requirements both for electronic and traditional media. However, as courseware becomes available and distributed through the Internet, quality of service (QoS) becomes an increasingly important factor to e-learner satisfaction. Thus, technology becomes another structural issue.

From the technology standpoint, three aspects are critical to e-learner satisfaction. The first is courseware evaluation (Chen, 2002; Karoulis et al., 2004; Kim, 2000; Liu & Dean, 1999; Storey et al., 2002; Strazzo & Wentling, 2001; Wentling et al., 2000). Indeed, users' perceptions of mismatch between content and structure reduce their motivation to learn and perform (Astleitner, 2001). Usability is a second aspect affecting both engagement and acceptance. It measures the extent to which a computer system can be used to complete well-defined tasks or achieve specified goals productively and satisfactorily for the intended users in a given context (Dix et al., 1998). Last, but not the least, user modeling completes the set of key technological aspects for e-learners' satisfaction. User modeling is the knowledge a system has about the user's level of knowledge and intentions, processed as users interact with systems. Knowledge of both user and task domains should allow intelligent and adaptable systems to properly respond to the competence levels and needs of the tasks within contexts of use (Dix et al., 1998). This holistic understanding would help developers take into consideration users' expectations at the early stages of system design. In this way, expected learner performance would supplement the metrics suggested by the literature (Dix et al., 1998; Wentling et al., 2000) concerning the implementation of strategies and quality-of-service goals.

Regarding relationship issues, two factors are relevant for this overview: cognitive styles and motivation. Cognitive styles are individual characteristics that serve as stable indicators of how learners perceive, think of, remember, and solve problems (Kim, 2000; Liu & Dean, 1999). This characteristic is consistent with major dimensions of individual differences and explains stable individual performance across tasks over time. Thus, it is a key variable in designing effective systems for a particular user group, especially at early stages of the interaction. Indeed, results show that cognitive styles can help in explaining usability problems when browsing hypermedia documents (Chen, 2002; Kim, 2000; Liu & Dean, 1999). However, research results are numerous and mixed and give no consistent evidence about the relationship between cognitive styles and learners' performance in computer-based settings (Shih & Gamon, 2001; Wentling et al., 2000).

Motivation can be defined as the internal set of processes (both cognitive and behavioral) through which human energy becomes focused on the direction of effort (Welbourne et al., 2000). This definition is twofold. First, the internal set of processes by which human energy becomes focused describes the "agentic" nature of individuals in interaction with their environment (Bandura, 2000; Liu et al., 2002). A second relevant aspect of motivation is the direction of effort, which implies individual goal orientation assessed using defined internal success criteria. These two motivation-related aspects require that development teams actively involve different user roles at early stages of system design to get high usability, learnability, acceptance, and usage levels in order to likely match specific individual, task, and contextual characteristics. In this way, organizations could be more effective in creating the conditions to foster knowledge acquisition, transfer, and reutilization across different learner groups.

FUTURE TRENDS

Currently, people perceive learning as a product rather than a process. Adopting the latter view would require a holistic approach to analyze e-learner experience. This may lead to a paradigm shift in dealing with process, especially since learning, as an activity, is deeply enmeshed in interdependent and rapidly evolving organizational processes. Our vision is to pursue useful endeavors to effectively support learning by people within organizational contexts during their performance of work-related tasks. In this sense, e-learning systems would become tools for people to use, as often as needed, to acquire new knowledge related to current or future tasks. For this to happen, we need to develop a close people-process-system fit. To achieve that, cost-effective, integrated tools for courseware and competence-building evaluation within specific knowledge domains are in order. This requires coordination of efforts, information, and competencies among the key players, including

Table 1. Identified research topics

Issues	Proposed research topics
Learning task	How do e-learning initiatives relate to organizational knowledge management practices?
E-learner	In what conditions could e-learning be effective for everybody? Are there usability metrics universal across roles, levels of technology experience, cognitive styles, job levels, organizational contexts, and cultures? Does task complexity level mediate users' perception about usability? Would the initial role assigned by people to media hold steadily throughout the learning process, or does it change over time, motivated by learners' habits?
Interaction	To what extent is learners' trust and performance affected by perceived support for privacy, confidentiality, and security at the system level? Could software agents be the new "hidden persuaders" (Packard, 1957/1981) to get learners to go further into the skill-development cycle, overcoming obstacles along the way? Should such concerns be incorporated into a design discipline centered on ethical and deontological roles?

universities, companies, government, and research communities. Such coordination is fundamental to achieve skill development at individual, organizational, and societal levels, articulated with defined strategies. Table 1 summarizes the topics identified in our analysis and describes specific research goals toward the fulfillment of this vision.

CONCLUSION

We have discussed a holistic approach covering business processes, e-learners, and information systems fit, while stressing the need for quantitative models, especially in evaluation. The interplay among these entities defines the organizational space for the just-in-time learning concept. The expected benefits lie in aligning learning outcomes with business strategies.

REFERENCES

Astleitner, H. (2001). Web-based instruction and learning: What do we know from experimental research? Retrieved November 2001 from http://rilw.emp.paed.uni-muenchen.de/2001/papers/astleitner.html

Bandura, A. (2000). Cultivate self-efficacy for personal and organizational effectiveness. In E. A. Locke (Ed.), *Handbook of principles of organizational behavior* (pp. 20-136). Oxford, United Kingdom: Blackwell.

Bernardes, J., & O'Donoghue, J. (2003). Implementing online delivery and learning support systems: Issues, evaluation and lessons. In C. Ghaoui (Ed.), *Usability evaluation of online learning programs* (pp. 19-39). Hershey, PA: Idea Group Publishing.

Chen, S. (2002). A cognitive model for non-linear learning in hypermedia programmes. *British Journal of Educational Technology, 33*(4), 449-460.

Dix, A., Finlay, J., Abowd, G., & Beale, R. (1998). *Human-computer interaction* (2nd ed.). Prentice Hall Europe.

Fuglseth, A. M. (2003). A tool kit for measurement of organizational learning: Methodological requirements and an illustrative example. *Journal of Universal Computer Science, 9*(12), 1487-1499.

Karoulis, A., Tarnanas, I., & Pombortsis, A. (2003). An expert-based evaluation concerning human factors in ODL programs: A preliminary investigation. In C. Ghaoui (Ed.), *Usability evaluation of online learning programs* (pp. 84-97). Hershey, PA: Idea Group Publishing.

Kim, K. (2000). *Individual differences and information retrieval: Implications on Web design.* Retrieved September 2001 from http://citeseer.nj.nec.com/update/409393

Liu, Y., & Dean, G. (1999). Cognitive styles and distance education. *Online Journal of Distance Learning Administration, II*(III). Retrieved September 2001 from http://www.westga.edu/~distance/and23.html

Liu, Y., Lavelle, E., & Andris, J. (2002). Experimental effects of online instruction on locus of control. *USDLA Journal, 16*(6). Retrieved April 2004 from http://www.usdla.org/html/journal/JUN02_issue/article02.html

Packard, V. (1981). *The hidden persuaders.* Penguin Books. (Original work published 1957)

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press.

Shih, C., & Gamon, J. (2001). Web-based learning: Relationships among student motivation, attitude, learning styles, and achievement. *Journal of Agricultural Education*, *42*(4). Retrieved April 2004 from http://pubs.aged.tamu.edu/jae/pdf/Vol42/42-04-12.pdf

Storey, M. A., Phillips, B., Maczewski, M., & Wang, M. (2002). Evaluating the usability of Web-based learning tools. *Educational Technology and Society*, *5*(3). Retrieved April 2004 from http://ifets.ieee.org/periodical/Vol_3_2002/ storey.html

Strazzo, D., & Wentling, T. (2001). *A study of e-learning practices in selected Fortune 100 companies.* Retrieved September 2001 from http://learning. ncsa.uiuc.edu/papers/elearnprac.pdf

Vouk, M., Bilzer, D., & Klevans, R. (1999). *Workflow and end-user quality of service issues in Web-based education.* NC: North Carolina State University, Department of Computer Science. Retrieved September 2001 from http://www.computer.org/tkde/tk1999/k0673abs.htm

Welbourne, T., Andrews, S., & Andrews, A. (2000). *Back to basics: Learning about energy and motivation from running on my treadmill.* Retrieved September 2001 from http://www.eepulse.com/pdfs/treadmill%20adobe%203.1.01.pdf

Wentling, T., Waight, C., Gallager, J., La Fleur, J., Wang, C., & Kanfer, A. (2000). *E-learning: A review of literature.* Retrieved September 2001 from http://learning.ncsa.uiuc.edu/papers/elearnlit.pdf

KEY TERMS

Business Process: A set of organized work-related tasks and resources to pursue a specific organizational objective influencing learning experiences by defining two specific relationships: process-based roles (between business process and people) and learning tasks (between business process and information systems).

E-Learning Experience: A process by which people identify work-related learning needs, formulate related goals and the associated internal level-of-success criteria, search for feasible online options to achieve defined learning goals, select and acquire choices, and engage into and complete them successfully by achieving the related goals in a productive and satisfactory manner.

E-Learning Framework: A formal construct to diagnose and manage learning outcomes in terms of the operational dynamic of three basic entities: business process, information systems, and people.

Just-In-Time Learning: Strategic knowledge acquisition enmeshed in business activities to support employees in learning new skills when performing day-to-day tasks, while fostering the alignment between learning outcomes, technological and strategic business issues.

Learning Task: A set of steps with a defined learning goal addressing specific training needs identified within business processes driving the definition of proper instructional design and e-learning system requirements.

Motivation to E-Learn: An individual variable denoting an internal set of processes (both cognitive and behavioral) by which human energy becomes focused on learning particular work-related content (whether by actively interacting with courseware, participating in a virtual class, self-studying, doing e-homework alone or in group) to achieve specific learning goals.

Process-Based Role: The combination of a set of expected work-related behaviors, responsibilities, and the associated set of required competencies to perform within business settings in order to achieve organizational goals.

People-Process-System Fit: A degree of consistency among learner groups, business processes, and e-learning systems that (a) reflects the target constituency requirements, organizational context, and the stated educational or training goals, (b) applies the principles and approaches of constructivism, user-centeredness, participatory design, quality management, and organizational development to instructional design of courseware, and (c) translates into expected performance levels.

Modeling Information Systems in UML

Peter Rittgen
University College of Borås, Sweden

INTRODUCTION

The first approaches to object-oriented modeling appeared by the second half of the 1970s, but not much happened for more than a decade, so there were still barely more than a handful of modeling languages at the end of the 1980s. It was the early 1990s that witnessed an ever-growing market in competing object-oriented methods so that potential users found it increasingly difficult to identify any single method that suited their needs. This phenomenon came to be known as the "method wars." Towards the end of 1994 two of the "big" players, Grady Booch and Jim Rumbaugh, decided to join forces by integrating their respective approaches, the Booch method and OMT (Object Modeling Technique). In late 1995, Ivar Jacobson became a member of this team merging in his OOSE method (Object-Oriented Software Engineering). The efforts of the "three amigos" aimed at overcoming unnecessary differences between the individual approaches and also improving each of them by creating a common, standardized modeling language that could serve as an industry standard. The result was the release of the Unified Modeling Language (UML), version 0.9, in June 1996. The UML partners, an industry consortium, performed further work on UML. This led to the versions 1.0 and 1.1 being introduced in 1997. The latter was adopted by the OMG (Object Management Group) in the same year. The current version is 1.5 (OMG, 2003) but a major upgrade to 2.0 is in preparation (Björkander & Kobryn, 2003).

BACKGROUND

UML is a language to support the development of software systems. It features a common framework, which provides a set of diagram types covering different aspects of an information system. Here we will focus on the ones we deem to be most important: class diagram, use case diagram, and activity diagram. The first is the principal diagram for the static view on a system. The second is often put forward as the central diagram in communication with the user (see, e.g., Dobing & Parsons, 2000). The latter plays a central role in the information system (or business-oriented) perspective (see following sections). Elementary concepts of the UML are actor, activity (and state), object, and class.

An **actor** is a human being or a (computer) system who/which is able to perform activities on his/her/its own. The actor typically represents a group of similar human beings/systems and corresponds to the role the actor performs in the given context. *Teacher* is an example for such an actor. In UML actors are shown as stick figures.

An **activity** refers to a logically connected set of actions that are carried out as a unit in some order. These actions might be executed sequentially, alternatively, concurrently or in any combination thereof. *Grade exam* is an example for an activity.

*"An **object** is an abstraction of a set of real-world things such that:*

- *all of the real-world things in the set - the instances - have the same characteristics,*
- *all instances are subject to and conform to the same rules" (Shlaer & Mellor, 1988, p. 14).*

The structure (attributes) and behavior (operations) of similar objects are gathered in their common **class**. The values of the attributes at a certain point in time represent the state of the object. Classes are drawn as rectangles with the object identifier printed in bold face. Additional horizontal compartments can be introduced for the attributes and operations of the class. The object is depicted in the same way with the object identifier underlined (optionally followed by a colon and the respective class identifier). *Grade* is an example of a class, *F: Grade* that of an object.

Figure 1 shows how these concepts are related to each other: actors perform activities that involve objects. Each object is an instance of precisely one class. The class diagram shows classes and their relations (called associations). An activity diagram gives a detailed account of the order in which activities are executed. It can also show the relevant states of the involved objects. The use case diagram visualizes use cases (complex activities) and their relation to actors.

Figure 1. Language overview

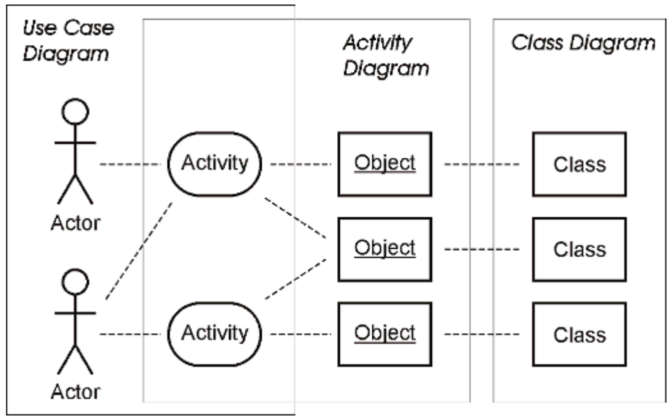
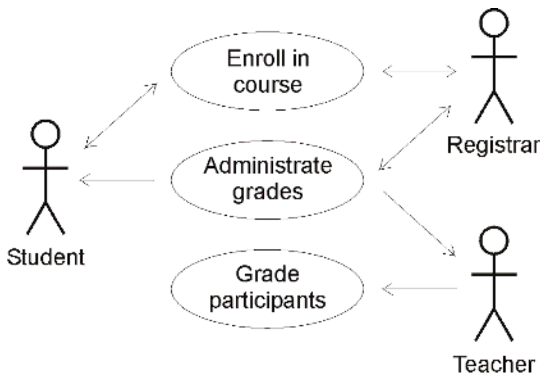


Figure 2. Use case diagram



Use Case Diagram

Use cases were introduced in 1992 by Jacobson and his colleagues. Their book is now available in the 2nd edition (Christerson, Jonsson, & Övergaard, 2004). A use case describes a way in which an actor can interact with your organization. It forms part of a use case scenario, a business situation that is supposed to be supported by the information system in question. *Figure 2* gives an example of a use case diagram involving 3 use cases and 3 actors.

The arrows indicate the direction of the information flow. So the registrar can both read and enter/change grades, whereas teachers and students can only read them. Often the arrowheads are omitted to simplify modeling. Each use case is detailed by a description in structured English of the activities that make up this use case. The use case diagram and the use cases form the basis for the development of class diagrams and activity diagrams. The former specify the static structure of the information that is handled by the use cases, and the latter give a precise, formalized account of the process logic.

Activity Diagram

An activity diagram consists of the detailed activities that make up the use cases. *Figure 3* gives an example of such a diagram in relation to the use cases of *Figure 2*.

The use case *Enroll in course* comprises the activities *Apply for course* (performed by the student) and *Enroll student* (performed by the registrar). *Grade participants* maps to *Grade exam* assuming that the grade for the course is determined by a final exam only. The use case *Administrate grades* involves the activities *Enter grade*, *Check grade* and *Change grade*. Note that the activity diagram also contains activities that are not present in the use case diagram, such as *Deliver course* and *Write exam*. That is because, in our example, they are not supposed to be supported by the information system and hence do not constitute use cases of such a system. But they are still an important part of the overall business process.

The business process in *Figure 3* starts with the initial state (the black dot). The activities are in the boxes with round sides. The rectangles contain objects-in-state where the object identifier is underlined and the object's state is enclosed in square brackets. So after *Enter grade*, for example, the object *Grade* is in state *entered*. Alternative paths leaving an activity must be labeled by so-called guards that determine under which condition each path is taken. These guards are also enclosed in square brackets and should be mutually exclusive. The process terminates when the final state is reached (i.e., the circle containing the black dot). Actors are included in the form of so-called

Figure 3. Activity diagram

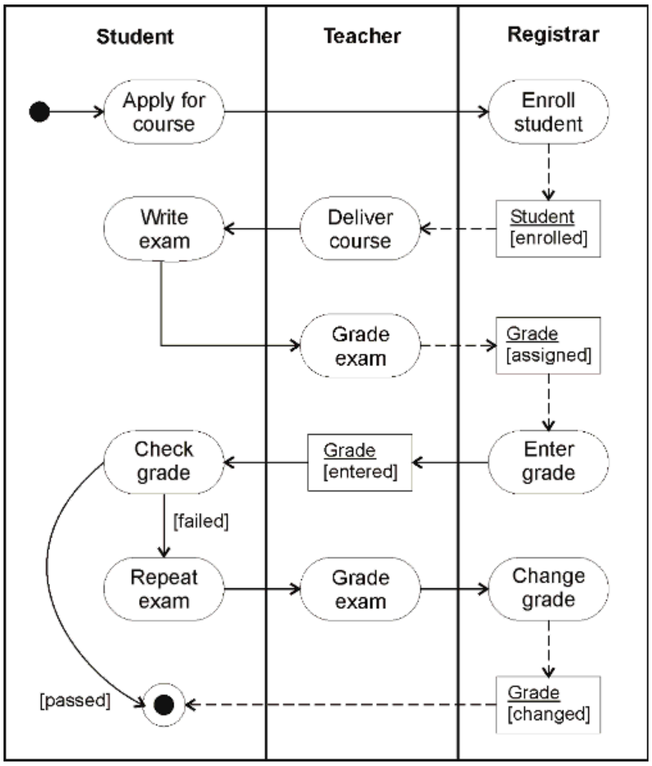
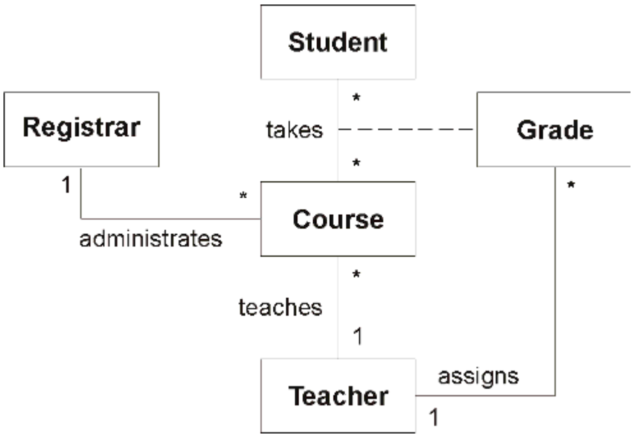


Figure 4. Class diagram



swim lanes. Concurrent execution is achieved by introducing synchronization bars. For more detailed information on activity diagrams we refer the reader to the OMG (2003, p. 3-155 ff).

Class Diagram

A typical approach to identifying classes is that of looking at the nouns contained in a textual specification of the system. Jacobson, Christerson, Jonsson, & Övergaard (2004) suggest that we should limit our attention to the nouns contained in the Use Case Diagram instead to avoid having to eliminate inappropriate class candidates. If we apply this to the use case diagram of *Figure 2* we arrive at the class diagram of *Figure 4* where all the nouns of the former map to a class of the latter. Observe that the noun *participants* maps to the class *Student*.

The class diagram also shows associations between the classes that indicate the way in which they are related to each other. A single line is drawn connecting the respective classes. It can be labeled with a verb that determines the nature of the association. In addition we can also specify the cardinality with the asterisk referring to an unlimited number. In the example a *Teacher teaches* a number of *Course*s but each *Course* is only taught by one *Teacher*. *Grade* is a so-called association class. It qualifies the relation between *Student* and *Course* and is therefore attached to the association path by a dashed line. For further details refer to the OMG (2003, pp. 3-34 ff).

MODELING INFORMATION SYSTEMS IN UML

UML has been designed primarily to support the development of software systems. In such systems the software artifact plays the central role and the role of human beings is restricted to the specification of requirements and, later on, the use of the system. An information system, on the other hand, is a human activity system (sometimes also called a socio-technical system) where some part of the system *might* be *supported* by software (Buckingham, Hirschheim, Land, & Tully, 1987).

The focus being on software systems, it is natural that UML does not emphasize the aspects of an information system that aim at the value and support it can provide to the business, such as strategy (e.g., value chains & strategic goals) and organization (e.g., organizational charts and business processes). These issues are dealt with in "business modeling" (also called enterprise modeling) but UML is beginning to advance into that field, too. The OMG (2003, pp. 1-9), for instance, claims that business processes can be represented in UML. This claim can, however, be challenged in a number of ways. First of all there is very little direct support for business processes in the basic UML diagrams (Bruno, Torchiano, & Agarwal, 2002). Instead many of the required concepts are "hidden" in extensions to UML, such as Enterprise Collaboration Architecture (ECA) (OMG, 2004). Another issue is whether UML models are indeed "understandable" enough to form a solid basis for the communication between "developers" and "users." An empirical study of UML's usability (Agarwal & Sinha, 2003) showed that UML diagrams score between 4.3 and 5.3 on a 7-point Likert scale,that is, the ease-of-use of UML is closer to indifference (4) than to the optimum (7). Understandability of UML is also weakened by construct redundancies, inconsistencies and ambiguities (Shen & Siau, 2003).

Candidate languages for business processes (in native UML) are use cases and activity diagrams. The former have been criticized for their conceptual flaws (Dobing & Parsons, 2000), such as fragmentation of objects, and for construct overload and ambiguity (Shen & Siau, 2003). In addition they restrict the specification of process logic to a textual form, which is neither rigorous nor expressive enough for typical flows of control. Activity diagrams come closer to fulfilling the requirements of a business-process language and are therefore often used in that way (see next section).

FUTURE TRENDS

So far, there is still an abundance of languages used for business modeling, for example, Architecture of Integrated Information Systems ARIS (Scheer, 1999), Open System Architecture for Computer-Integrated Manufacturing CIMOSA (AMICE, 1993), Integrated Definition IDEF* (Ang, Pheng, & Leng, 1999), Integrated Enterprise Modeling IEM (Spur, Mertins, & Jochem, 1995), Dynamic Essential Modeling of Organization DEMO (Reijswoud, Mulder, & Dietz, 1999), and GRAI Integrated Methodology GIM (Doumeingts, 1998). Often modelers also pick out particular model types from integrated methods or use "stand-alone" languages that were developed for specific purposes. In the area of business process modeling these include: Event-Driven Process Chains (EPC) (from ARIS), Business Process Modeling Language (Arkin, 2002), IDEF3 (Mayer et al., 1997), Tropos (Castro, Kolp, & Mylopoulos, 2002), and Petri Nets (Aalst, Desel, & Oberweis, 2000).

But how can we bridge the gap between these enterprise-modeling languages and UML? One approach consists of devising mechanisms that "translate" domain models into UML models. Such a translation is by no means straightforward but typically involves the reconstruction of the domain model as a UML diagram by a modeler who is experienced in both the domain language and UML. It usually involves the loss of information that was present in the domain model and the necessity to add information not yet represented. This implies also the risk of introducing errors, or more precisely inconsistencies between the original and the translated model. Examples of such an approach can be found in many of the languages mentioned due to the popularity of UML, for example, EPC to UML (Nüttgens, Feld & Zimmermann, 1998), IDEF and UML (Noran, 2000), Tropos and UML (Mylopoulos, Kolp, & Castro, 2001), DEMO and UML (Mallens, Dietz, & Hommes, 2001), or Petri Nets and UML (Gou, Huang, & Ren, 2000).

Another – and perhaps even more promising – approach is to equip UML itself with domain support. This can be done in at least two ways: extending existing concepts or introducing new ones (typically in the form of profiles). A major effort towards the latter is being made under the heading "Enterprise Distributed Object Computing (EDOC)" and was adopted by the OMG in February 2004 (version 1.0) but work is still ongoing. Its core is the Enterprise Collaboration Architecture ECA (OMG, 2004). The former approach is followed by several researchers independently. In the area of business processes and workflows they primarily investigate extensions of use cases and activity diagrams (see, for example, Rittgen, 2003; Dumas & Hofstede, 2001).

CONCLUSION

The UML is a widely used modeling language in the area of software systems modeling but it still has to gain ground in domain modeling, especially in business (or enterprise) modeling. Promising steps have already been taken that have so far led to results such as EDOC. But there is still a need for research that provides the core concepts and diagrams of the UML with domain-specific semantics to make them more useful for business modeling. On the other hand, add-on concepts such as EDOC processes require a tighter semantical integration with the existing models. Future research and development will show which of the paths towards integration is more promising and will hence be taken. But all approaches mentioned share a common goal: to improve the communication and understanding between the different people involved in shaping an information system and thereby also improving the system itself.

REFERENCES

AMICE. (1993). *CIMOSA: Open system architecture for CIM* (2nd rev. and ext. ed.). Berlin: Springer.

Agarwal, R., & Sinha, A.P. (2003). Object-oriented modeling with UML: A study of developers' perceptions. *Communications of the ACM*, *46* (9), 248-256.

Ang, C.L., Pheng, K.L., & Leng, G.R.K. (1999). IDEF*: A comprehensive modelling methodology for the development of manufacturing enterprise systems. *International Journal of Production Research*, *37* (17), 3839-3858.

Arkin, A. (2002). *Business process modeling language*. Aurora, CO: BPMI.org.

Björkander, M., & Kobryn, C. (2003). Architecting systems with UML 2.0. *IEEE Software*, *20* (4), 57-61.

Bruno, G., Torchiano, M., & Agarwal, R. (2002). UML enterprise instance models. In S. Iyer & S. Naik (eds.), *CIT 2002*: *Proceedings of the Fifth International Conference on Information Technology*. Bhubaneswar, India, December 21-24, 2002. New Delhi, India: Tata McGraw-Hill.

Buckingham, R.A., Hirschheim, R.A., Land, F.F., & Tully, C.J. (1987). Information systems curriculum: A basis for course design. In R.A. Buckingham, R.A. Hirschheim, F.F. Land, & C.J. Tully (eds.), *Information Systems Education. Recommendations and Implementation* (pp. 14-133). Cambridge: Cambridge University Press,.

Castro, J., Kolp, M., & Mylopoulos, J. (2002). Towards requirements-driven information systems engineering: The Tropos project. *Information Systems*, *27* (6), 365-389.

Dobing, B., & Parsons, J. (2000). Understanding the role of use cases in UML: A review and research agenda. *Journal of Database Management*, *11* (4), 28-36.

Doumeingts, G. (1998). GIM: GRAI integrated methodology. In A. Molina, A. Kusiaka, & J. Sanchez (eds.), *Handbook of Life Cycle Engineering: Concepts, Models and Methodologies* (pp. 227-288). Dordrecht, Netherlands: Kluwer Academic Publishers.

Dumas, M., & Hofstede, A.H.M. (2001). UML activity diagrams as a workflow specification language. In M. Gogolla & C. Kobryn (eds.), *UML 2001: The Unified Modeling Language*. Lecture Notes in Computer Science. (Vol. 2185) (pp. 76-90). 4th International Conference, Toronto, Canada, October 1-5, 2001.,. Berlin: Springer.

Gou, H., Huang, B., & Ren, S. (2000). A UML and Petri nets integrated modeling method for business processes in virtual enterprises. In S. Staab & D. O'Leary (eds.), *Bringing Knowledge to Business Processes*. Papers from 2000 AAAI Spring Symposium (pp. 142-144). Menlo Park, CA: AAAI Press.

Jacobson, I., Christerson, M., Jonsson, P., & Övergaard, G. (2004). *Object-oriented software engineering: A use case driven approach* (2nd ed.). Wokingham, U.K.: Addison-Wesley.

Mallens, P., Dietz, J., & Hommes, B.J. (2001). The value of business process modeling with DEMO prior to information systems modeling with UML. In J. Krogstie, K. Siau, & T. Halpin (eds.), *EMMSAD'01: 6th IFIP 8.1 Workshop on Evaluation of Modeling Methods in Systems Analysis and Design*. Oslo, Norway: SINTEF.

Mayer, R.J., Menzel, C.P., Painter, M.K., deWitte, P.S., Blinn, T., Perakath, B., Sartor, J.M., & McManus, J.C. (1997). *Information integration for concurrent engineering (IICE): IDEF3 process description capture method report*. AL/HR-TP-1996-0030. College Station, TX: Knowledge Based Systems Inc. & Wright-Patterson AFB, OH: Armstrong Laboratory.

McLeod, G. (2000). Beyond use cases. In K. Siau, Y. Wand, & A. Gemino (eds.), *EMMSAD'00: 5th IFIP 8.1 Workshop on Evaluation of Modeling Methods in Systems Analysis and Design*. Stockholm, Sweden, June 4-5, 2000.

Mylopoulos, J., Kolp, M., & Castro, J. (2001). UML for agent-oriented software development: The Tropos proposal. In M. Gogolla & C. Kobryn (eds.), *UML 2001: The Unified Modeling Language*. Lecture Notes in Computer Science (vol. 2185) (pp. 422-441). 4th Int. Conference, Toronto, Canada, October 1-5, 2001., Berlin: Springer.

Noran, O. (2000). *Business modelling: UML vs. IDEF*. Brisbane, Australia: Griffith University. Retrieved from www.cit.gu.edu.au/~noran.

Nüttgens, M., Feld, T., & Zimmermann, V. (1998). Business process modeling with EPC and UML: Transformation or integration? In M. Schader & A. Korthaus (eds.), *The Unified Modeling Language: Technical Aspects and Applications* (pp. 250-261). Berlin: Springer.

OMG. (2003, March). *Unified Modeling Language Specification*. Version 1.5. Needham: OMG. Retrieved from www.uml.org.

OMG. (2004, February). *Enterprise collaboration architecture specification*. Version 1.0. Needham: OMG. Retrieved from www.uml.org.

Palanque, P., & Bastide, R. (2003). UML for interactive systems: What is missing. In G.W.M. Rauterberg, M. Menozzi, & J. Wesson(eds.), *Human-Computer Interaction INTERACT '03*. 9th IFIP TC13 International Conference on Human-Computer Interaction, September 1-5, 2003, Zürich, Switzerland. Amsterdam, Netherlands: IOS Press.

Rittgen, P. (2003). Business processes in UML. In L. Favre (ed.), *UML and the Unified Process* (pp. 315-331). Hershey, PA: IRM Press.

Scheer, A.W. (1999). *ARIS: Business process modeling*. Berlin: Springer.

Shen, Z., & Siau, K. (2003). An empirical evaluation of UML notational elements using a concept mapping approach. In S.T. March, A. Massey, & J.I. DeGross (eds.), *Proceedings of the 24th Int. Conference on Information Systems* (pp 194-206). Seattle, WA, December 14-17, 2003. Atlanta, GA: AIS.

Shlaer, S., & Mellor, S.J. (1988). *Object-oriented systems analysis: Modeling the world in data*. Englewood Cliffs, NJ: Prentice-Hall.

Spur, G., Mertins, K., & Jochem, R. (1995). *Integrated enterprise modeling*. Berlin: Beuth.

Van der Aalst, W., Desel, J., & Oberweis, A. (2000). *Business process management: Models, techniques and empirical studies*. Lecture Notes in Computer Science (Vol. 1806). Berlin: Springer.

Van Reijswoud, V.E., Mulder, J.B.F., & Dietz, J.L.G. (1999). Speech act based business process and information modelling with DEMO. *Information Systems Journal, 9* (2), 117-138.

KEY TERMS

Activity: A logically connected set of actions that are carried out as a unit in some order. It is associated with a state (called action state) in which the system remains while the activity is performed.

Actor: A person or (computer) system that can perform an activity. The actor does not refer to a particular individual but rather to a role (e.g., *Teacher*).

Attribute: A property of an object/class. The class *Car*, for example, can have an attribute *Color*, its value for the object *MyCar: Car* might be *blue*.

Class: A template for similar objects defining attributes and operations.

Object: An abstraction of a set of real-world things that has state, behavior and identity. An instance of its class where the values of the attributes determine the state, and the operations the behavior.

Operation: A function or transformation that may be applied to or by objects in a class. The class *Account*, for example, might have the operations *Open*, *Close*, *PayIn* (*Amount*), and *Withdraw* (*Amount*).

State: A certain combination of attribute values of an object.

Modeling Technological Change in Small Business

Arthur Tatnall
Victoria University, Australia

INTRODUCTION

The introduction of a new information system into a small business, or upgrading an existing system, should be seen as an innovation and considered through the lens of innovation theory. The most widely accepted theory of how technological innovation takes place is provided by innovation diffusion, but most of the research based on this model involves studies of large organisations or societal groups. This article argues that another approach, innovation translation, has more to offer in the case of innovations that take place in smaller organisations (Burgess, Tatnall & Darbyshire, 1999; Tatnall, 2002; Tatnall & Burgess, 2004).

BACKGROUND

There are important differences in the processes by which small and large enterprises choose to adopt or reject computers (Burgess et al., 1999), and this article concerns itself only with issues related to small business. To begin, however, it is important to distinguish between invention and innovation. Whereas invention can be seen in the discovery or creation of new ideas, innovation involves putting these ideas into commercial or organisational practice (Maguire, Kazlauskas & Weir, 1994). Invention does not necessarily invoke innovation and it is fallacious to think that invention is necessary and sufficient for innovation to occur (Tatnall, 2004).

Changing the way things are done is a complex affair (Machiavelli, 1995, p. 19) and one that is difficult to achieve successfully. The dominant paradigm, by far, in innovation research is that of *innovation diffusion* and no discussion would be complete without consideration of this approach. Innovation diffusion has had success in describing how innovations diffuse through large populations (Rogers, 1995). There are occasions, however, when diffusion does not occur and the diffusion model finds these difficult to explain (Latour, 1996). The approach offered in *innovation translation*, informed by actor-network theory (ANT), is also worthy of consideration. In the translation model the key to innovation is creating a powerful enough consortium of actors to carry it through, and when an innovation fails this can be considered to reflect on the inability of those involved to construct the necessary network of alliances amongst the other actors. This article will compare these two models of technological innovation.

INNOVATION DIFFUSION

Rogers (1995), perhaps its most influential advocate, approaches the topic of innovation diffusion by considering a variety of case studies, the prime concern of which is the identification of factors that affect the speed with which an innovation is adopted, or that cause it not to be adopted at all.

In diffusion theory the existence of an innovation is seen to cause uncertainty in the minds of potential adopters, causing a lack of predictability and of information. Rogers (1995) asserts that a technological innovation embodies information, and that this has the potential to reduce uncertainty. Diffusion is thus considered to be an information exchange process amongst members of a communicating social network driven by the need to reduce uncertainty (Lepa & Tatnall, 2002). There are four main elements of the theory of innovation diffusion (Rogers, 1995):

Characteristics of the Innovation Itself

Rogers argues that the attributes and characteristics of the innovation are important in determining the manner of its diffusion and the rate of its adoption, and outlines five important characteristics of an innovation that affect its diffusion: relative advantage, compatibility, complexity, trialability and observability. The attributes of the potential adopter are also seen as an important consideration and Rogers maintains that these include social status, level of education, degree of cosmopolitanism and amount of innovativeness.

Nature of the Communications Channels

Acts of communication are a necessary part of any change process, and to reach a potential adopter the innovation must be diffused through some communications channel. Channels involving mass media are the most rapid and efficient means of spreading awareness of an innovation, but interpersonal channels are generally more effective in persuading someone to accept a new idea.

The Passage of Time

In common with earlier researchers, Rogers found that different individuals in a social system do not necessarily adopt an innovation at the same time. Borrowing from work by Deutschmann and Fals Borda (1962), he proposes that adopters can be classified in their degree of "innovativeness" into five categories: innovators, early adopters, early majority, late majority and laggards, and that if the number of individuals adopting a new idea is plotted over time it usually follows a normal curve.

The Social System

Diffusion occurs within a social system in which the structure constitutes a boundary inside which this diffuses. Rogers argues that the system's social structure affects diffusion through the action of social norms, the roles taken by opinion leaders and change agents, the types of innovation decisions that are taken, and the social consequences of the innovation.

INNOVATION TRANSLATION

An alternative view is that of innovation translation, which draws on the sociology of translations, more commonly known as actor-network theory (ANT). The core of the actor-network approach is translation (Law, 1992), which can be defined as: "... the means by which one entity gives a role to others" (Singleton & Michael, 1993, p. 229).

Essentialism

Diffusion theory asserts that a technological innovation embodies information: some essential capacity or essence instrumental in determining its rate of adoption. A significant problem with an essentialist paradigm like this arises if a researcher tries to reconcile the views of all parties involved in the innovation on what *particular* essences are significant. The difficulty is that people often see *different* essential attributes in any specific technological or human entity, making it hard to identify and settle on the ones that allegedly were responsible for the diffusion.

To illustrate this difficulty, consider the case of a small business deciding whether to purchase their first computer. Researchers using an innovation diffusion model would begin by looking for innate characteristics of the PC that would make a potential adopter more likely to accept it. They would consider the relative advantages of a PC over alternatives like a filing cabinet. An examination of the compatibility, trialability and observability of a PC with this older office technology would show good reasons for acceptance. An examination of the PC's complexity would, however, bring out some reasons for reluctance in its adoption. The researchers would then investigate characteristics of the potential adopters, considering factors like their educational background, innovativeness, and how they heard about the innovation. If, however, you *ask* small business people why they purchased their first PC, the answers often do not match with this view.

Actor-Network Theory: The Sociology of Translations

Rather than recognising in advance the essences of humans and of social organisations and distinguishing their actions from the inanimate behaviour of technological and natural objects, ANT adopts an anti-essentialist position in which it rejects there being some difference in essence between humans and non-humans. ANT considers both social and technical determinism to be flawed and proposes instead a socio-technical account (Latour, 1986) in which neither social nor technical positions are privileged. To address the need to properly consider the contributions of both human and non-human actors, actor-network theory attempts impartiality towards all actors in consideration, whether human or non-human, and makes no distinction in approach between the social, the natural and the technological (Callon 1986).

Mechanisms of Translation

The process of translation has four aspects or "moments" (Callon, 1986), the first of which is known as *problematisation*. In this stage, a group of one or more key actors attempts to define the nature of the problem and the roles of other actors so that these key actors are seen as having the answer, and as being indispensable to the solution of the problem. In other words, the problem is re-defined (translated) in terms of solutions offered by these actors (Bloomfield & Best, 1992). The second moment is *interessement* and is a series of processes that attempt to impose the identities and roles defined in the problematisation on the other actors. It means interesting and attracting an entity by coming between it and some other entity. Here the enrollers attempt to lock the other actors into the roles proposed for them (Callon, 1986) and to gradually dissolve existing networks, replacing them by a network created by the enrollers themselves.

If the interessement is successful then the third moment, *enrolment,* will follow through a process of coercion, seduction, or consent (Grint & Woolgar, 1997), leading to the establishment of a solid, stable network of alliances. Enrolment, however, involves more than just one set of actors imposing their will on others; it also requires these others to yield (Singleton & Michael, 1993). Finally, *mobilisation* occurs as the proposed solution gains wider acceptance and an even larger network of absent entities is created (Grint & Woolgar, 1997) through some actors acting as spokespersons for others (Tatnall, 2001).

FUTURE TRENDS: TRANSLATION VERSUS DIFFUSION

Many small businesses are family-related concerns with several members working in the business. Investigations (Tatnall, 2001, 2002; Tatnall & Burgess, 2002) suggest that a common reason why a small business first acquired a PC is that it was, at least partly, intended for family use. In this regard, one reason that the PC was obtained was in the belief that it would assist with their children's education. Once obtained, other uses almost always also found for the technology, but the question remains: Would the characteristics of the PC and of the people involved in its adoption have been identified by a diffusion model? Table 1 summarises difference in approaches between the diffusion and translation frameworks.

An Example: Adoption of a Slide Scanner by a Small Publishing Company

To illustrate the use of the two innovation models, consider the case of DP Pty Ltd – a small publishing company where four people work on the publication of information systems textbooks. DP is a very small business with a well-established market. Members of the company do much of the writing, and all of the work involved in publication of their books, but send the work off for printing and binding. Most of DP's print runs are small. All those involved in the work of the company are computer literate and make good use of IT. None of them, however, has much knowledge of computer graphics.

Several years ago, the company decided it needed to improve the appearance of the covers on its textbooks.

Table 1. Innovation diffusion versus innovation translation

	Innovation Diffusion	**Innovation Translation**
Innovation	A technology perceived to be new by the potential adopter.	A technology that has yet to be "black-boxed".
Communication	Communication channels can be categorised as cosmopolite or localite, and mass media or interpersonal. Innovations are transferred through these channels.	Translations are made by actors in enrolling the innovation.
Time	Speed of decision to innovate, earliness of adoption and rate of adoption are important.	Network dynamics in enrolment, control, and dissemination are what matter.
The Social System	Homophily versus heterophily. Sharing of interests of human actors.	Interessement between actants, both human and non-human, and goals. Black boxes form when interests move in the same direction.
The Technology	Changes are made to the form and content of the technology as a result of experiences during implementation (re-invention).	The technology is translated through being enrolled, regardless of whether its form or content is modified.
Socio-Technical Stance	The social system and the technology are separate. Diffusion is the adoption of technology by a social system. Technology transfer requires the bringing together of social and technical elements.	The social system and the technology are inseparable. Successful innovation and technology transfer gives the appearance of separation, but this is merely evidence that the actor-network has stabilised.

Adapted from McMaster et al. (1997)

Several options were considered until someone thought of using a photograph. The brother of one of the directors is a landscape photographer who was able to provide a suitable colour photograph. This was supplied in print form, and a problem arose in how to convert it (or its negative) into a suitable format to print on the cover along with the cover text. A digital image seemed to be the answer. The photograph was scanned (using a flat-bed scanner) and the digital image inserted into Microsoft Word so that text could easily be added. The final result was then converted into an Acrobat file and sent off to the printer. This process, however, proved to be quite a lot of bother. Today the company makes use of a Nikon slide and negative scanner, considerably improving the quality of the covers, but also making the process of producing them much simpler. This device is capable of producing digital images of slides and negatives at various resolutions. The question is: why did DP decide to adopt this *particular* item of technology?

Consider first the application of a diffusion model. Here the company would have been mainly influenced by attributes and characteristics of the technology itself. The directors would have considered the relative advantage, compatibility, complexity, trialability and observability of this technology compared with the alternatives. Two of the directors of the company certainly did see some relative advantage in using the slide scanner, particularly as they both had large numbers of colour slides that they had taken over the years. There was, however, one major disadvantage of this technology and that was its high cost: the slide scanner was around four times as expensive as a good flat-bed scanner. The slide scanner did not come out well on compatibility or complexity, as it was quite different and difficult to use. The company arranged trial use of the scanner, which was lucky as it proved difficult to find anyone else using one – its observability was low. On this basis, it is difficult to see why DP would have adopted the slide scanner at all.

When a translation model is applied, the situation is seen quite differently. The socio-technical network consisting of the publishing company personnel, their computers, and their books was destabilised by the need to find a new way of producing book covers. The addition of a new actor, the landscape photographer, introduced new possibilities that worked to further destabilise this network. The slide scanner (also seen as an actor seeking to enter the network) offered new possibilities in which existing (and future) slides and negatives could easily be turned into digital images. This included *any* of the directors' old slides, not just those required for book covers. As well as the main application of producing book covers, both directors quickly saw advantages in a device that could also easily convert the old slides and negatives they had each taken of their children and of their holidays into digital format. It was thus a combination of factors, some business-related and others rather more personal, that the translation modes suggests could be seen as leading to the adoption of this technology.

It should also be pointed out that a significant translation occurred from the slide scanner advertised by Nikon in the magazine article and on their Web page to the device adopted by the publishing company. DP was not interested in using the scanner in all the ways that Nikon offered. They wanted a device to digitise slides and negatives for two reasons: the easy creation of attractive book covers, and the conversion of their own colour slides into a format that could easily be displayed on a video monitor. They did not adopt the scanner advertised by Nikon, but a device for creating book covers and formatting their slides for video display.

CONCLUSION

The translation approach to innovation details the construction of networks and alliances to support and embed changes in order to make them durable. Innovation diffusion and innovation translation are based on quite different philosophies, one stressing the properties of the innovation and the change agent and routed in essentialism, and the other emphasising negotiations and network formation.

Innovation translation is more attuned to the human and political issues involved in small business decision making, and so offers a useful approach to modelling innovation in small business. Most current writings on technology uptake and innovation employ either no theoretical model at all, or else use a diffusion model. The author suggests that future research should further investigate the use of a translation model to explain the adoption of information technology in small business, and encourages future researchers to consider this approach.

REFERENCES

Bloomfield, B.P., & Best, A. (1992). Management consultants: Systems development, power and the translation of problems. *The Sociological Review, 40*(3), 533-560.

Burgess, S., Tatnall, A., & Darbyshire, P. (1999). *Teaching small business entrepreneurs about computers*. EuroPME - Entrepreneurship: Building for the Future, Rennes, France, Groupe ESC, Rennes.

Callon, M. (1986). Some elements of a sociology of translation: Domestication of the scallops and the fishermen of

St Brieuc Bay. In J. Law, R. London & P. Kegan (Eds.), *Power, action & belief. A new sociology of knowledge?* (pp. 196-229).

Deutschmann, P.J., & Fals Borda, O. (1962). *Communication and adoption patterns in an Andean village.* San Jose, Costa Rica, Programa Interamericano de Informacion Popular.

Grint, K., & Woolgar, S. (1997). *The machine at work - Technology, work and organisation.* Cambridge, MA: Polity Press.

Latour, B. (1986). The powers of association. In J. Law, R. London, & P. Kegan (Eds.), *Power, action and belief. A new sociology of knowledge? Sociological Review monograph 32* (pp. 264-280).

Latour, B. (1996). *Aramis or the love of technology*. Cambridge, MA: Harvard University Press.

Law, J. (1992). Notes on the theory of the actor-network: Ordering, strategy and heterogeneity. *Systems Practice, 5*(4), 379-393.

Lepa, J., & Tatnall, A. (2002). *Older people adopting the GreyPath Village Lyceum: An analysis informed by innovation diffusion*. Queensland: AusWeb.

Machiavelli, N. (1995). *The prince.* London: Penguin Classics.

Maguire, C., Kazlauskas, E.J., & Weir, A.D. (1994). *Information services for innovative organizations*. San Diego, CA: Academic Press.

McMaster, T., Vidgen, R.T., & Wastell, D.G. (1997). *Towards an understanding of technology in transition. Two conflicting theories*. Information Systems Research in Scandinavia, IRIS20 Conference, Hanko, Norway, University of Oslo.

Rogers, E.M. (1995). *Diffusion of innovations*. New York: The Free Press.

Singleton, V., & Michael, M. (1993). Actor-networks and ambivalence: General practitioners in the UK cervical screening programme. *Social Studies of Science, 23*, 227-264.

Tatnall, A. (2001). *Adoption of information technology by small business - Two different approaches to modelling innovation*. Managing Information Technology in a Global Economy - (IRMA'2001), Toronto, Canada, Idea Group Publishing.

Tatnall, A. (2002). Modelling technological change in small business: Two approaches to theorising innovation. In S. Burgess (Ed.), *Managing information technology in small business: Challenges and solutions* (pp. 83-97). Hershey, PA: Idea Group Publishing.

Tatnall, A. (2004). To adopt or not to adopt computer-based school management systems? An ITEM research agenda. In A.J. Visscher (Ed.), *Information technology and educational management in the knowledge society*. Assinippi Park, MA: Kluwer Academic Publishers / IFIP.

Tatnall, A., & Burgess, S. (2002). *Using actor-network theory to research the implementation of a B-B portal for regional SMEs in Melbourne, Australia*. 15th Bled Electronic Commerce Conference - eReality: Constructing the eEconomy, Bled, Slovenia, University of Maribor.

Tatnall, A., & Burgess, S. (2004). Using actor-network theory to identify factors affecting the adoption of e-commerce in SMEs. In D. Waddell (Ed.), *E-business: Innovation and change management* (pp. 152-169). Hershey, PA, Idea Group Publishing.

KEY TERMS

Actor-Network Theory: An approach to research in which networks' associations and interactions between actors (both human and non-human) are the basis for investigation.

Innovation: The application, in any organisation, of ideas new to it, whether they are embodied in products, processes or services.

Innovation Diffusion: A theory of innovation in which the main elements are: characteristics of the innovation itself, the nature of the communication channels, the passage of time, and the social system through which the innovation diffuses.

Innovation Translation: A theory of innovation in which, instead of using an innovation in the form it is proposed, potential adopters *translate* into a form that suits their needs.

Invention: The discovery or creation of new ideas.

Sociology of Translations: Another term used to refer to actor-network theory.

Socio-Technical Research: Involving both social and technical interactions, occurring in such a way that it is not easily possible to disentangle them.

Technological Innovation: The introduction or alteration of some form of technology (often information technology) into an organisation.

Model–Supported Alignment of IS Architecture

Andreas L. Opdahl
University of Bergen, Norway

INTRODUCTION

An *information system (IS)* is a system that communicates, transforms, and preserves information for human users. An information system comprises one or more software applications and databases, and their relationships to their human users, operators, and maintainers.

A modern enterprise has many information systems that can be related in various ways. For example, information systems can be *related by exchange* because they exchange data through message passing or shared databases, or because they exchange functions through remote procedure calls or Web services. Information systems can also be *related by overlap* because they maintain the same data or provide the same functions. Information systems can be related in many other ways too, either *directly*, such as when one IS controls another, or *indirectly*, for example, because several ISs depend on the same run-time platforms or because they compete for their users' attention or for computer resources. In addition to being related to one another, information systems can be related to the *surrounding organization* in many ways. For example, organization units such as departments, individuals, or roles may be the *owners*, *users*, *operators*, or *maintainers* of ISs; organizational goals and strategies can be *realized by* ISs; organizational processes can be *supported* or *automated by* ISs; and so on.

The *information systems (IS) architecture* of an enterprise comprises its information systems, the relationships between those information systems, and their relationships to the surrounding organization. In addition to single enterprises, *alliances* of enterprises and *parts* of enterprises, such as divisions and departments, can have IS-architectures too. The above definition implies that *every* enterprise has an IS-architecture, even if that architecture is not explicitly talked about, described, or managed: 'IS-architecture' is a way to look at organizations and their information systems.[1]

IS-architecture alignment is the process of selecting an *IS-architecture vision* towards which the architecture should be incrementally but systematically evolved. This article will present a model-supported framework for aligning an IS-architecture with its surrounding organization (Opdahl, 2003a). The framework shows how an enterprise's *current* IS-architecture can be represented in an enterprise model, from which *candidate architecture visions* can then be generated, before one of them is selected as the enterprise's IS-architecture vision.

BACKGROUND

Zachman (1978) defines 'IS-architecture' as "the sum total of all information-related flows, structures, functions, and so on, both manual and automated, which are in place and/or required to support the relationships between the entities that make up the business." In the last few decades, several IS-architecture methods have been proposed in both industry and academia (Opdahl, 2003a).

A related term is *information architecture (IA)*, used by some authors (e.g., Periasamy & Feeny, 1997) as a synonym to 'IS-architecture', although IA can also be used to emphasize the information sharing and information management aspects of IS-architecture. Another related term is *enterprise architecture (EA)* (McGovern et al., 2004), sometimes called *enterprise information architecture (EIA)* (Cook, 1996), which, according to Chorafas (2002), "is to align the implementation of technology to the company's business strategy" and "to make technology serve innovation economics." 'EA'/'EIA' is sometimes used synonymously with 'IS-architecture', but can also be used to emphasize organizational aspects such as process structure and organizational roles.

IS-architecture alignment can also be understood as an intermediate step (or level) between ICT strategy and detailed IS planning (Brancheau & Wetherbe, 1986).

IS-ARCHITECTURE ALIGNMENT

A *good IS-architecture* should be *strategically* and *operationally fit* to the enterprise, *simple and well structured*, *well managed*, and *clearly and explicitly described*. These characteristics are explained as follows:

- *Strategically fit* means that the IS-architecture should support the enterprise in pursuit of its goals and strategies. This is of course the primary characteristic of a good IS-architecture.
- *Operationally fit* means that the IS-architecture should be integrated with the enterprise's *organi-*

zational structures, such as its *market structure*, *product structure*, *process structure*, *function structure*, *organization structure*, and so on. Although operational fitness may not be a goal in itself, some degree of operational fitness is necessary to achieve strategic fitness.

- *Simple and well structured* means that the IS-architecture should not be unnecessarily complex, because a complex IS-architecture will be difficult to comprehend and understand, and difficult to change without unanticipated consequences. It will therefore be hard to manage.
- *Well managed* means that the principles, activities, roles, and responsibilities for IS-architecture maintenance and evolution should be well-defined and properly taken care of. An IS-architecture that is not explicitly and properly taken care of may start to drift and quickly become unnecessarily complex and/or strategically and operationally unfit.
- *Clearly and explicitly described* means that the enterprise should always document both its *current* IS-architecture and its IS-architecture *vision*. Whereas the current architecture should be represented by a sketch or blueprint,[2] the vision should

Table 1. The core metatypes in the representation framework. For each metatype, a brief description is given, along with examples of possible sub-metatypes (from Opdahl, 2003a).

Metatype (Subtype examples)	*Description*
Goal (mission, vision, business objectives, etc.)	The motives/rationales for the **Activities** carried out by the **Organization Units** and for other **Phenomena**. ***Goals*** *can be either explicit statements or implicit ideas. They can be either shared or individual and either official or private.*
Strategy (business strategies, principles, plans and standards, etc.)	Guidelines for how **Organization Units** carry out **Activities**. *G**uidelines** can be either formal or informal.*
Organization Unit (divisions, business units, departments, work groups, employees, project groups, boards, committees, etc.)	One or more persons. *An **Organization Unit** can be either an individual or a group. It can be either permanent or temporary.* *Note that a **Role** is a subtype of **Organization Unit**, i.e., an individual unit at the type level. The **Role** subtype is so important in enterprise modeling that it should often have its own icon.*
Activity (functions, processes, tasks, some projects, etc.)	Actions or events that occur in the enterprise. ***Activities*** *can either be singular, continuous, or repeated.*
Information	A pattern of information or data that is used and/or produced in the enterprise. ***Information*** *can be on electronic or other formats, e.g., paper.*
Application	A software system that automates or supports an **Activity** in order to let an **Organization Unit** accomplish an **Goal**.
Database (electronic archives, libraries, etc.)	A collection of data or information in the enterprise. *A D**atabase** can be in electronic or other form.*
Basic Software (operating systems, protocols, etc.)	A group of cooperating programs that are used by Applications and Databases.
Computing Equipment (computers, peripherals, etc.)	A piece of hardware.
Network	A communication network that connects computers with other computers, peripherals, and/or networks.
Phenomenon	Any of the above, i.e., either an objective, a strategy, an organization unit, an activity, information, an application, a database, basic software, computing equipment, a network, or an instance of one of the extensional metatypes.

additionally be documented by a set of higher-level evolution principles.[3]

IS-architecture alignment is the process of selecting such a set of higher-level principles—expressed as an IS-architecture vision—towards which the IS-architecture is to be incrementally but systematically evolved.

Henderson and Venkatraman's (1993) *strategic alignment model* distinguishes between the external and internal domains of businesses on the one hand, and between the business domain and the ICT domain on the other hand. In consequence, their framework distinguishes between *strategic integration*, which is "the link between business strategy and I/T strategy," and *functional integration*, which is "the link between organizational infrastructure and processes and I/S infrastructure and processes" (Henderson & Venkatraman, 1993). Relative to Henderson and Venkatraman's model, IS-architecture alignment focuses on 'functional integration' and on the 'internal domain' of enterprises.

MODELING ORGANIZATIONS AND IS-ARCHITECTURES

The framework for model-supported alignment relies on a structured view of an *organization* (Opdahl, 2003a) as a collection of *organizational elements* with *organizational relationships* between them. There are different types of elements, such as *goals*, *organization units*, and *activities*. Table 1 lists the most important ones, with further details given in Opdahl (2003a). The terms used in Table 1 are only suggestions. In real projects, they should be refined into more organization-specific terms. For example, 'organization unit' could be replaced by terms such as 'division', 'department', 'project group', 'role', and 'position', and the term 'activity' could be refined into 'project', 'process group', 'process', 'process step', 'operation', 'business function', and so forth. The framework defines many types of organizational relationships too (not shown in Table 1), such as goals that are *realized-by* activities and organization units that *carry-out* activities and that are *responsible-for* goals.

According to the framework, a part of the organization forms an *IS-architecture*, which is correspondingly viewed as a collection of *IS-architecture elements* and *relationships*. Examples of IS-architectural element types are *applications* and *databases*, and an example of a relationship type between IS-architecture elements is that applications *manipulate* databases. There are also various types of relationships between IS-architecture elements and other organizational elements, such as databases that *store* information, and activities that are *supported-by* applications and that *manipulate* information.

IS-architecture elements form *IS-architecture areas*, which may comprise several different types of elements. A particularly important type of IS-architecture area is the enterprise's *information systems*, which are collections of tightly coupled applications and databases that are related to the rest of the organization. But the framework also allows for other types of IS-architecture areas that group other kinds of IS-architecture elements, for example *responsibility areas* that group information systems into larger clusters. IS-architecture areas are important because they can be used to make the IS-architecture simple and well structured, and thereby more manageable: in a good IS-architecture, the architecture areas support the enterprise's goals and strategies, and the elements in each architecture area are closely related to one another, whereas there are few relationships between elements that belong to distinct areas.

Figure 1 shows the most important organizational and IS-architectural element types in the framework. Goals and strategies are executed by the grey-shaded operational organization, whose three domains are shown as layers in Figure 1. The *organizational domain* comprises organization units, activities, and information; the *information systems domain* comprises IS-architecture areas, applications, and databases, whereas the *ICT-infrastructure domain* comprises basic software, computing equipment, and computer networks.

Importantly, many elements in the framework are *decomposable*. For example, an organization unit can have sub-units, and an information element can have sub-elements. Decomposable elements of the same type form *organizational structures*, similar to organizational *dimensions* (Armour, 2003). Examples of organizational structures are the hierarchy of organization units in an enterprise and its business function hierarchy.

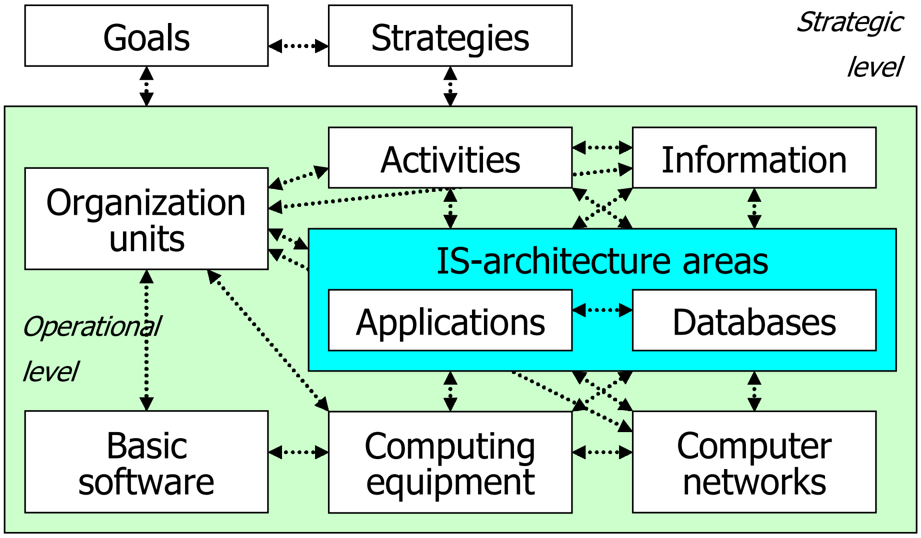
Figure 1 is also a sketch of a more detailed underlying *metamodel* (Opdahl, 2003a). The metamodel can be seen as a model of a modeling language for representing IS-architectures and the organizations that surround them. The metamodel has been implemented in a graphical enterprise-modeling tool, supported by the Computas-Metis modeling and metamodeling tool family. The framework and tool can be used to represent an enterprise's *current* IS-architecture, as well as *candidate* and *selected* IS-architecture *visions*.

ALIGNING ORGANIZATIONS AND IS-ARCHITECTURES

The framework for model-supported alignment centers on *operational* and *strategic fitness*, the two most important

M

Figure 1. The core of the IS-architecture representation framework. The core metamodel can be extended with metamodels that represent other metatypes, such as products and locations and IS architecture areas other than information systems (Opdahl, 2003a).



characteristics of a good IS-architecture. Firstly, the framework systematically investigates alternative ways to achieve operational fitness by generating a candidate architecture vision for each of them. Secondly, the generated candidate visions are evaluated and compared according to their strategic fitness, before one of them is selected as the enterprise's IS-architecture vision. In other words, the framework investigates different ways to achieve operational fitness and selects one of them according to strategic fitness.

In the first step, different ways to align the IS-architecture with the surrounding organization are investigated systematically, inspired by an approach outlined by Kiewiet and Stegwee (1991). Each candidate architecture vision is expressed as a set of *IS-architecture principles* used for grouping IS-architecture elements into IS-architecture areas. Each principle prescribes one way to group one type of IS-architecture element and, together, the set of principles prescribes one way to group all the element types in the architecture.

A principle for grouping IS-architecture elements (of a particular type) comprises (a) an organizational structure, (b) possibly a hierarchical level of decomposition of that structure, and (c) a relationship, possibly indirect through several intermediate relationships, between that structure (at that level of decomposition) and an IS-architecture element (of that type). According to the principle, two IS-architecture elements will belong to the same group if they have the same relationship (c) to an element at the appropriate level of decomposition (b) in the organizational structure (a). For example, applications (a type of element) may be grouped into information systems (a type of architecture area) according to the business function (an organizational structure) that they support (a particular relationship). Furthermore, applications can be grouped into a few large information systems according to high-level business functions or into many smaller information systems according to lower-level, more decomposed functions (decomposition levels in the hierarchy of business functions) (Opdahl, 2003a).

This structured view of IS-architecture alignment, along with the equally structured views of organizations and IS-architectures presented in the previous section, makes it possible to systematically investigate all possible sets of principles, either manually or automatically. If the current IS-architecture and its surrounding organization are represented in an IS-architecture model, it is even possible to automatically generate models of each candidate architecture vision.

The first step of the alignment framework ensures that the generated candidate visions will be both *operationally fit, simple and well structured*, and *clearly and explicitly described*. But it does nothing to satisfy the most important criterion of all, *strategic fitness*. In the second step, the candidate architectures are therefore evaluated and compared according to strategic fitness. The framework does *not* currently support automatic selection of optimal or satisficing architectures leaving this task for manual assessment according to, for example, the enterprise's goals and strategies. However, automatic generation of enterprise models for candidate architecture visions should make selection easier by making assessments of and comparisons between candidates more concrete. Developing heuristics for critiquing, shortlisting, prioritizing, and selecting candidate architectures remains a topic for further work.

FUTURE TRENDS

This article has outlined how enterprise models can be used to help enterprises align their IS-architectures with the surrounding organization. But model-supported alignment of IS-architectures is only one among many important ways in which future enterprises can benefit from enterprise models. For example, new ICT systems will gradually become adaptable and manageable through enterprise models, as when a new customer resource management (CRM) system is tailored by modifying associated models of information processes and markets. Later changes to the enterprise model can then be automatically reflected in the running CRM system, so that both initial adaptation and subsequent management can be done at the model level.

When the enterprise's ICT systems thus become controllable through enterprise models, there is a danger that enterprises become more rigid and uncoordinated as a result, because the models are expressed in many different and unrelated modeling languages. This danger can be avoided by *integrating* the modeling languages and technologies used to control different ICT systems. As a result, the enterprise can become better coordinated because its ICT systems become semantically integrated. The enterprise can also become more flexible because changing ICT systems through enterprise models is easier than changing them at the implementation level. There is therefore a need for theories and tools for tightly integrating a broad variety of enterprise models and modeling languages (Opdahl, 2003b; Opdahl & Sindre, 1997), including information and process models, actor and organization models, and goal and business-rule models. In this light, the alignment framework is only a partial contribution to developing theories and tools for tying together ICT systems in a flexible way through tightly integrated enterprise models.

CONCLUSION

The article has presented a model-supported framework for aligning an IS-architecture with the surrounding organization. One advantage of the framework is that it does not only derive blueprints of future architectures, but also generates *principles* for evolving the current IS-architecture towards the architecture vision. Another advantage is that the framework can in principle be supported by a tool, which can in the long run be developed into an *IS-architecture* (or *enterprise architecture*) *workbench*. The workbench could gradually be extended to supplement the framework with a variety of alternative approaches to ICT strategic alignment.

The framework needs to be developed further. When tool support becomes available, it must be validated by industrial case studies. Further research is also needed on how to compare candidate architectures and how to best represent and visualize IS-architectures. Also, the framework in its present form is *reactive* rather than *proactive*, because it takes the surrounding organization as a given. Although it should not be difficult to modify the framework to better support proactive use, this needs to be investigated.

Behind the alignment framework is a broader view of tomorrow's enterprises, whose ICT systems will be controlled by comprehensive enterprise models. As a result, *enterprise model integration* will become a prerequisite for ICT systems integration and thereby become the key to successful *enterprise integration*. In the enterprises of the future, the cost of establishing and maintaining large models will be shared by many different areas of use. For this to happen, new theories, technologies, and tools are needed to develop, maintain, and operate large models that integrate multiple perspectives on the enterprise and that are used for different purposes (Opdahl, 2003b; Opdahl & Sindre, 1997).

REFERENCES

Armour, P. (2003). The reorg cycle. *Communications of the ACM*, *46*(2), 19-22.

Brancheau, J.C. & Wetherbe, J.C. (1986). Information architectures: Methods and practice. *Information Processing & Management*, *22*(6), 453-463.

Chorafas, D.N. (2002). *Enterprise architecture and new generation information systems*. St. Lucie Press/CRC Press.

Cook, M.A. (1996). *Building enterprise information architectures—Reengineering information systems*. Englewood Cliffs, NJ: Prentice-Hall.

Henderson, J.C. & Venkatraman, N. (1993). Strategic alignment: Leveraging information technology for transforming organizations. *IBM Systems Journal*, *32*(1), 4-15.

Kiewiet, D.J. & Stegwee, R.A. (1991). Conceptual modeling and cluster analysis: Design strategies for information architectures. In J.I. DeGross, I. Benbasat, G. DeSanctis & C.M. Beath (Eds.), *Proceedings of the 12th Annual International Conference on Information Systems* (pp. 315-326).

McGovern, J., Ambler, S.W., Stevens, M.E., Linn, J., Sharan, V. & Jo, E.K. (2004). *A practical guide to enterprise architecture.* Pearson.

Opdahl, A.L. (2003a). Model-supported alignment of information systems architecture. In K. Kangas (Ed.), *Business strategies for information technology management.* Hershey, PA: Idea Group Publishing.

Opdahl, A.L. (2003b). Multi-perspective multi-purpose enterprise knowledge modeling. In R. Jardim-Goncalves, J. Cha & A. Steiger-Garcão (Eds.), *Concurrent engineering: Enhanced interoperable systems—The vision for the future generation in research and applications* (pp. 609-617). A.A. Balkema Publishers.

Opdahl, A.L. & Sindre, G. (1997). Facet modeling: An approach to flexible and integrated conceptual modeling. *Information Systems, 22*(5), 291-323.

Periasamy, K.P. & Feeny, D.F. (1997). Information architecture practice: Research-based recommendations for the practitioner. *Journal of Information Technology, 12*, 197-205.

Zachman, J.A. (1978). The information systems management system: A framework for planning. *Data Base*.

KEY TERMS

Enterprise Model: A diagrammatic representation of an enterprise or part of an enterprise. An enterprise usually focuses on certain aspects of the enterprise, such as its goals and strategies, its business processes, its organization structure, its information and knowledge, etc.

Information System (IS): A system that communicates, transforms, and preserves information for human users. An information system comprises one or more computerized data systems along with their human users, operators, and maintainers.

Information Systems Architecture, IS-Architecture: The set of information systems in an organization, the relationships between those information systems, and the relationships between the information systems and the rest of the organization.

IS-Architecture Alignment: The process of selecting an IS-architecture vision that is strategically and operationally fit for the enterprise, simple and well structured, well managed, and clearly and explicitly described.

IS-Architecture Model: An enterprise model that focuses on the enterprise's IS-architecture and that can be used to represent a current architecture or to illustrate a candidate or selected architecture vision. An IS-architecture sketch is a high-level model, whereas an architecture blueprint is more detailed.

IS-Architecture Principle: A high-level rule that can be used to make decisions about developing and/or evolving individual ICT systems.

IS-Architecture Vision: A coherent set of IS-architecture principles that together guide all the aspects of IS-architecture evolution that are considered important.

ENDNOTES

1. This contrasts authors who define 'IS-architecture' as a 'blueprint' or 'sketch' of how the enterprise's ISs *are* or *should be* organized. In the terminology of this article, although a blueprint or sketch can *represent* or *describe* an IS-architecture, the blueprint or sketch *is not* the architecture.
2. The difference between the two is that a *sketch* is a rough representation that is used to communicate central aspects of an IS-architecture, whereas a *blueprint* is intended as a complete description of the architecture.
3. Principles are preferable to blueprints because a blueprint is only valid for as long as the rest of the organization stays roughly the same, whereas a principle can be useful even after the organization has undergone changes, although the enterprise's principles must of course be re-evaluated from time to time as it evolves.

Moderator in Government-Initiated Online Discussions

Arthur R. Edwards
Erasmus University Rotterdam, The Netherlands

INTRODUCTION

In an older version of a Dutch Internet dictionary the moderator is defined as "a person who exercises censorship on a mailing list or newsgroup."[1] In the libertarian tradition of the Internet, moderation has often been viewed as conflicting with free speech and unrestrained communication (Tsagarousianou, 1998). However, as the history of the famous PEN-experiment (Public Electronic Network) in Santa Monica (1990-96) already showed, the free speech principle has to be weighed against other legitimate concerns, like the need to facilitate a genuine discussion and to counteract possible abuses of the medium (Docter & Dutton, 1998).

This article covers government-initiated online discussions between citizens, government officials, politicians, and social organizations. The importance of moderation of these discussions is now generally recognized (see, for instance, Coleman & Gøtze, 2001). Moderation is important to stimulate and regulate online discussions as purposeful social action. Some characteristics of online interaction, such as its virtuality or anonymity may diminish the psychological thresholds to participate in a discussion, but they may also inhibit the social cooperation that is needed to accomplish complex communicative tasks. From research on discussions in political newsgroups we know that discussions often serve more as a means for reinforcing preexisting views than to persuade others (Hill & Hughes, 1998; Davis, 1999; Wilhelm, 2000). These findings do not imply that the moderation of political newsgroups is imperative. As far as they can be characterized as online "free-for-all-discussions" that satisfy a social need to express opinions and concerns, this is an open question that can be left to the participants. Online discussions, however, that are initiated to involve citizens in dealing with public issues, do require moderation. In these settings, moderation is also necessary to realize some potential advantages of online discussions. Because of their asynchronous nature, there are more possibilities for structuring them. Various discussion lines can be opened and managed. Also, there is more flexibility possible in providing information. To reap these fruits, moderation is necessary.

BACKGROUND

A moderator can be defined as a person (or group of persons) who facilitates a discussion in view of its goals and agenda.

The moderator roles have been discussed since the inception of the Internet community. The Guide for Electronic Citizen Consultation, published by the Dutch Ministry of the Interior (1998), mentions three moderator roles. First, the moderator functions as a "host" so that the participants feel at ease. He shows them the way, so to speak, in the discussion, how it works, where information can be found, etc. Second, the moderator is seen as a "discussion leader." In this role, he furthers the progress of the discussion. Also, he makes sure that all participants actually take part in the discussion. Third, the moderator has a role as an "arbiter." He may designate certain postings as inappropriate and decide to remove them. Coleman and Gøtze (2001) have listed a number of metaphors to designate various moderator roles, based on work by White (2002) and others. These include the roles of "social host," "project manager," "community of practice facilitator," "'cybrarian," "help desk," "referee" and "janitor." These designations are useful, as they give an impression of the variety of moderator roles. White (2002) relates each role to specific types of communities and applications, and also indicates which key skills are required.

In this article, a more theoretical approach will be proposed by outlining a conceptual model of the "management" of Internet discussions. The underlying claim of this model is that it specifies all (main) tasks that have to be performed in the design and management of online policy exercises that should carry at least some weight in the political decision-making. A management approach suggests that certain general "management functions" have to be performed. I distinguish (1) the strategic function, (2) the conditioning function and (3) the process function (see figure 1).

The *strategic* function is to establish the boundaries of the discussion and to embed it in the political and organizational environment. This includes the following tasks:

- Establish the *goals* that the discussion is designed to achieve, both towards the citizenry and the institutional decision-making system;
- Establish and maintain the *substantive domain* of the discussion, i.e., the boundaries of the agenda within which themes and issues may be raised;
- Obtain *political and organizational support* for the discussion;
- Establish the *status* of the discussion in terms of influence on decision-making;
- Ensure that the *results* of the discussion will actually be carried over into the decision-making process and to give feedback on this to the participants.

The *conditioning* function refers to all kinds of conditions and provisions that have to be taken care of to further the discussion. This may include the following:

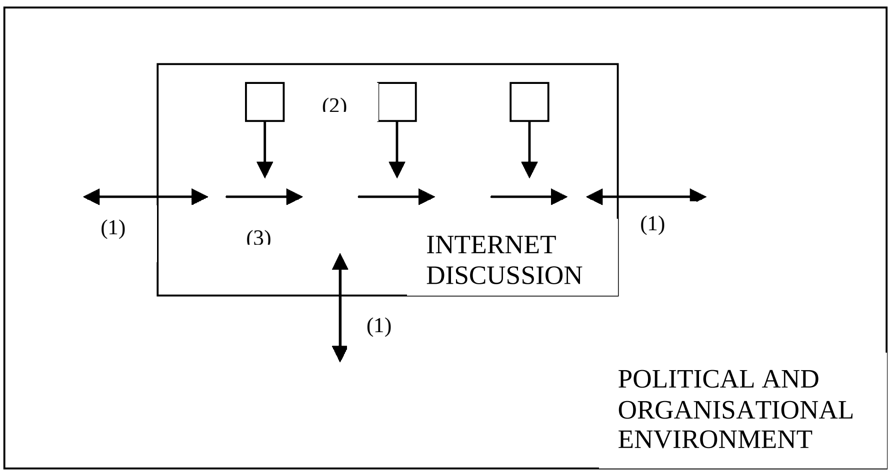
- *Solicit people* to join the discussion as *participants*;
- Provide *information to facilitate informed discussion*;
- Provide *supporting technologies*, such as moderation software, simulation models and visualization.

The *process* function includes all tasks that have to do with the discussion process as a cooperative, purposeful activity:

- Set the *interactional goal* of the discussion, i.e., the kind of results to be reached by the participants within the discussion, for instance, exploration of problem definitions or consensus about a proposal of policy measures;
- Specify the *agenda* of the discussion, within the substantive domain that has been established in the strategic function: the questions, propositions or themes to be discussed;
- Set the *schedule* of the discussion;
- Manage the *discussion process* in view of its interactional goal, its agenda, and its schedule, for example, assign messages to discussion lines or open new discussion lines;
- Facilitate the progress of the discussion by making *summaries* during the discussion;
- Stimulate the *interactivity* of the discussion, for instance, by stirring up participants to take part in the discussion and to give reactions to specific contributions;
- Set and maintain the *rules of the game.*

As an analytical tool, this model can be used in two ways. First, in an actor-oriented way, it can be used as an instrument to discover what moderators do (Edwards, 2002). Second, in a process-oriented way, it can be used to ascertain how the different management functions are performed and which actors are involved. Used in this way, the model allows for contributions to the manage-

Figure 1. The management of online discussions (Edwards, 2002; reprinted)



(1) strategic function: establish the boundaries of the discussion and embedding it in the political and organizational environment.
(2) conditioning function: take care of conditions and provisions.
(3) process function: further the progress of the discussion as a cooperative, purposeful activity.

ment of online discussions by other actors than the moderator (Edwards, 2004).

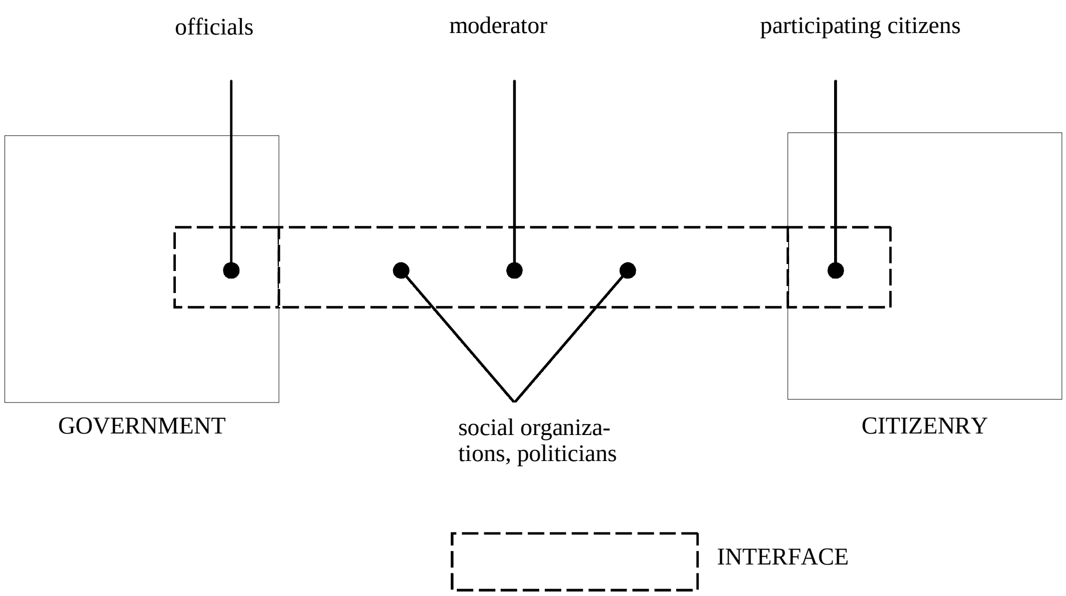
EMPIRICAL FINDINGS ON MODERATION

Case studies indicate that the moderator's primary involvement is in the process function (Edwards, 2002, 2004; Märker et al., 2002; Lührs et al., 2003). First of all, the moderators put a lot of effort into enhancing the interactivity of the discussion, for instance by stirring politicians and other institutional actors to react to contributions of citizens. Moderators are also involved in the formulation of the discussion themes (the agenda*)*. In most cases this takes place in cooperation with the staff of the initiating agency. Furthermore, to keep the discussion on track as a purposeful activity, moderators manage discussion lines and make summaries. Finally, moderators set and maintain some rules of the game, although their actual interventions as a filter are relatively seldom needed. As far as the conditioning function is concerned, moderators are often involved in the information provision. They may do this by inviting others, notably public agencies, interest groups, and political parties, to provide their views on the subject. With regard to the strategic function, moderators are involved in carrying over the results of the discussion to the policy makers. It seems plausible that other strategic tasks, such as establishing the strategic goals of the discussion, its status, as well as obtaining organizational and political support are accomplished by the initiators themselves. We conclude that moderators fulfill important intermediary roles between citizens and public administration.

The significance of this can be clarified further, when we relate our findings to the features of a *deliberative procedure*. In a deliberative democracy, citizens share a commitment to the resolution of problems of collective choice through free public deliberation, and regard their basic institutions as legitimate in so far as these institutions establish a framework for this (Cohen, 1989). We may take an approach to online discussions as providing forums for free public deliberation (Schalken, 1998; Tsagarousianou, 1998). From Benhabib (1994), three principles can be derived that constitute a deliberative procedure. The first principle states that the participation in deliberation is governed by the norms of equality and symmetry. This principle reflects Habermas's "ideal speech situation." It stipulates that all citizens should have equal opportunities to start or to enter a public discussion, as well as to advance stands or proposals and to oppose the stands or proposals of others. Furthermore, the participants are required to defend their proposals by arguments and to refrain from using power (Habermas, 1975). Benhabib's second principle states that all participants have the right to question the assigned topics of the discussion. The third principle is that all have the right to initiate reflexive arguments about the very rules of the discourse procedure and the way in which they are applied or carried out.

With regard to the first principle, the efforts of moderators to enhance the interactivity of the discussion and to create a forum on which politicians and other institutional actors probe the views of citizens and expose their

Figure 2. Actors involved in government-initiated online discussions (Edwards, 2004)



own points of view to critical discussion, can be seen as contributing to an ideal speech situation. Moderators see to it that participants refrain from using power during the discussion, so that only the "force of arguments" can settle the dispute. In addition, moderators can do a lot to ensure that participants defend their proposals by arguments. In particular, a moderator is legitimated to intervene if a participant does not react when a point of view that he advanced earlier in the discussion is attacked by another participant. Benhabib's second principle, the right of participants to question the assigned topics of conversation, implies that the agenda of the discussion (the specific questions or propositions) should be open to amendments, both at the start and during the discussion. Benhabib's last principle suggests that the moderation policy should be transparant and negotiable. Some discussion rules have to be formulated at the start of the discussion. Together with the agenda, they can be consolidated in a commonly agreed charter of the discussion group.

The empirical findings suggest that moderators can do a lot to enhance the interactivity and openness of the discussions as well as their relevance for public policy making. However, the performance of the moderator is dependent on several *institutional and organizational factors*. Moderators do not function in isolation—they have to cooperate with other actors. Figure 2 depicts the interface of actors that are involved in the discussions.

A government that initiates an online discussion with citizens, can recruit a moderator from its own staff, appoint an independent ("third party") moderator, or recruit a moderator from the citizenry. An independent moderator has an autonomous stand towards the various parties in the discussion. While acting as an intermediary, a moderator serves both as an agent of the initiating government, and as an agent of the participating citizens. Here lies a field of tension in which an independent moderator is in a better position to strike a balance. Nevertheless, an independent moderator might also be a source of bias in its own right. For instance, when the moderator is involved in information provision or in making summaries of the results of the discussion, he or she is always vulnerable to allegations of being partisan.

Moderation by civil servants is an acceptable option, if the moderation is transparant and negotiable and citizens have no reason to question the impartiality of the moderation. In many cases, it is an interesting option to involve citizens in the moderation. There are a few cases documented of online discussions, in which citizen-moderation took place. In an online discussion in 2001 in the German city of Bremen, the moderation was performed by a team consisting of four citizens, a civil servant and an expert. It came out that citizen moderators can take over an important feedback function, and can contribute to the success of the discussion by their knowledge of the local situation and relevant people (Westholm, 2003).[2] Citizen moderation and moderation by government staff are less appropriate, however, in cases where the issue is controversial.

In most cases the moderator's position is embedded in *organizational arrangements* in which the initiating government and sometimes also social organizations are represented. In the "project groups," "'supervisory teams," or "editorial groups," the moderator, public officials and representatives of other organizations work together in organizing and managing the discussion. These organizational arrangements are of strategic importance, because they embed the discussion in the organizational and political environments. On the one hand, they impose some limits on the autonomy of the moderator, but on the other hand they may enhance the relevance of the discussion, as to its impact on the regular policy process.

The position of *social organizations* deserves some special attention. The social organizations can figure in five roles: (a) as initiators of a discussion, (b) as partners of the initiating government to give legitimacy to the discussion and to participate in the formulation of its agenda, (c) as information providers, (d) as agencies for soliciting participants, and (e) as participants in the discussion. The involvement of social organizations in online discussions, their interaction with individual citizens, civil servants, and politicians, is an important subject for further research.

The participation of *politicians* is important, as they are the final decision makers in a representative democracy. Their commitment will be beneficial to the discussion's impact on political decision making. Politicians often show a reluctance to involve themselves in "interactive policy exercises" with citizens, whether face-to-face or online (Coleman & Gøtze, 2001). Moderators should encourage their participation, although a too dominant presence of politicians may thwart a free expression of citizens' points of views. Herein lies another example of the "sense of balance" that moderators have to show.

A possibility for strengthening the position of the *participating citizens*, and thereby enhancing the openness of the discussion, is to present the agenda and the moderation policy at the start of the discussion as proposals that are open to amendments. The outline of the discussion would then be consolidated in a commonly agreed upon charter. Also, procedural rules could be established for amendments or additions to the agenda during the discussion. Other provisions can be made to further the openness and negotiability of the moderation.[3] Furthermore, citizens can be invited to make suggestions as to relevant information or to forward information to be placed on the discussion site. Finally, the conclusions of the discussion could be consolidated in a collaboratively drafted summary.[4]

FUTURE TRENDS

A recent trend is the development of specific *discourse support systems*, i.e., systems of information and communication technologies for supporting a discussion. They provide a Web-based platform as well as a methodology and appropriate tools for enabling a fruitful dialogue. Generally, they also provide a set of features for moderating the discussion (Gordon & Richter, 2002; see also Coleman & Gøtze, 2001).5 Moderator tasks will be more and more embedded in such systems of both social and technical components. More and more, the technical components will enable cooperative work and design by the participants.

CONCLUSION

Moderators are emerging new intermediaries in forms of deliberative democracy that are supported by modern information and communication technologies (ICTs). In this article we focused on online discussions as one category of these new forms. Another category includes ICT-facilitated face-to-face discussions.6 Insofar as these forms will enter into the practice of democratic governance, moderators will establish themselves as elements of the information and communication infrastructure between the citizenry and public administration. Moderators enhance the quality of online discussions and their relevance for political decision-making.

Moderator tasks will be more and more embedded in systems of both social and technical components. In the organizational arrangements, procedures, and discussion rules, as well as in the design of the technological components, due attention has to be given to the openness of the discussion, and to the transparency and negotiableness of the moderation.

REFERENCES

Benhabib, S. (1994). Deliberative rationality and models of democratic legitimacy. *Constellations*, 1, 26-52.

Cohen, J. (1989). Deliberation and Democratic Legitimacy. In A. Hamlin & P. Pettit (Eds.), *The Good Polity: Normative Analysis of the State* (pp.1-13). Oxford: Blackwell.

Coleman , S., & Gøtze, J. (2001). *Bowling Together. Online Public Engagement in Policy Deliberation.* Hansard Society. Retrieved March 25, 2004 from the World Wide Web at: http://bowlingtogether.net/

Davis, R. (1999). *The Web of Politics. The Internet's Impact on the American Political System.* New York, Oxford: Oxford University Press.

Docter, S., & Dutton, W.H. (1998). The First Amendment Online. Santa Monica's Public Electronic Network. In R. Tsagarousianou, D. Tambini, & C. Bryan (Eds.), *Cyberdemocracy. Technology, Cities and Civic Networks* (pp.125-151). London: Routledge.

Edwards, A.R. (2002). The moderator as an emerging democratic intermediary. The role of the moderator in Internet discussions about public issues. *Information Polity, 7*, 3-20.

Edwards, A.R. (2004). The Moderator in Government-Initiated Internet Discussions: Facilitator or Source of Bias? In M. Mälkiä, A.V. Anttiroiko, & R. Savolainen (Eds.), *eTransformation in Governance. New Directions in Government and Politics* (pp. 150-167). Hershey, PA: Idea Group.

Gordon, T.F., & Richter, G. (2002). Discourse Support Systems for Deliberative Democracy. In R.Traunmüller, & K. Lenk (Eds.), *Electronic Government. Proceedings EGOV 2002 Aix-en-Provence* (pp. 248-255). Berlin: Springer Verlag.

Habermas, J. (1971). *Toward a Rational Society.* London: Heineman.

Hill, K.A., & Hughes, J.E. (1998). *Cyberpolitics. Citizen Activism in the Age of the Internet.* Lanham: Rowman & Littlefield.

Lührs, R., Albrecht, S., Lübcke, M., & Hohberg, B. (2003). How to Grow? Online Consultation about Growth in the City of Hamburg: Methods, Techniques, Success factors. In R. Traunmüller (Ed.), *Electronic Government. Proceedings EGOV 2003, Prague* (pp. 79-84). Berlin: Springer Verlag.

Ministry of the Interior (1998). *Electronic Civic Consultation; First Experiences* (in Dutch): Den Haag.

Schalken, C.A.T. (1998). Internet as a New Public Sphere for Democracy. In I.Th.M. Snellen, & W.B.H.J. van de Donk (Eds.), *Public Administration in an Information Age: A Handbook* (pp.159-174). Amsterdam: IOS.

Tsagarousianou, R. (1998). Electronic Democracy and the Public Sphere. Opportunities and Challenges. In R. Tsagarousianou, D. Tambini, & C. Bryan (Eds.), *Cyberdemocracy. Technology, Cities and Civic Networks,* (pp. 167-178). London: Routledge.

Westholm, H. (2003). Neue Medien für bessere Bürgerbeteiligung in der “Bürgerkommune”? Ein Praxisbericht. In W. Prigge, & W. Osthorst (Eds.), *Bremen auf dem Weg zur Bürgerkommune?* Universität Bremen: Institut Arbeit und Wirtschaft.

White, N. (2002). *Facilitating and hosting a virtual community*. Retrieved March 25, 2004 from the World Wide Web at: http://www.fullcirc.com/community/communityfacilitation.htm

Wilhelm, A.G. (2000). *Democracy in the Digital Age. Challenges to Political Life in Cyberspace*. New York/London: Routledge.

KEY TERMS

Deliberative democracy: a form of democracy in which citizens share a commitment to the resolution of problems of collective choice through free public deliberation, and in which the basic democratic institutions provide a framework for this.

Deliberative procedure: a discussion that is governed by the norms of equality and symmetry in participation and the right of the participants to question the agenda and the discussion rules, as well as the way in which the agenda and rules are applied.

Discourse support system: a system of information and communication technologies, providing a Web-based platform, a methodology, and appropriate tools for fruitful discussions.

Moderator: a person or group of persons who facilitates a discussion in view of its goals and agenda.

The conditioning function of the management of online discussions: take care of all kinds of conditions and provisions to further the discussion.

The process function of the management of online discussions: all tasks that have to do with the discussion process as a cooperative, purposeful activity.

The strategic function of the management of online discussions: establish the boundaries of the discussion and embed it in the political and organizational environment.

ENDNOTES

1 Het Internet Woordenboek, Furore, 1999.

2 The discussion was in the district Horn-Lehe. One of the citizens in the moderation team was recruited form the local citizen association. Retrieved March, 25, 2004 from the World Wide Web at: http://infosoc2.informatik.uni-bremen.de/~hassan/about.php3.

3 In an online discussion that took place in the German city of Esslingen, there was a separate forum on the site for a reflexive “meta-discussion” on the moderation, the relevance of the discussion, the user-friendliness of the technology, and similar issues (Märker et al., 2002).

4 Some of these suggestions are also formulated in a “Dispute Resolution Flowchart” designed by the Centre for Information Technology and Dispute Resolution (1999), in Wilhelm (2000).

5 An example is the Delphi Mediation Online System (DEMOS). Retrieved March 25, 2004 from the World Wide Web at: www.demos-project.org.

6 An example are the dialogues that took place in New York City (summer 2002) about the rebuilding of Lower Manhattan. Retrieved March 25, 2004 from the World Wide Web at :www.listeningtothecity.org

Monitoring Strategies for Internet Technologies

Andrew Urbaczewski
University of Michigan, Dearborn, USA

INTRODUCTION

Most large organizations that provide Internet access to employees also employ some means to monitor and/or control that usage (Reuters, 2002). This article provides a classification and description of various control mechanisms that an organization can use to curb or control personal Internet usage. Some of these solutions are technical, while others rely instead on interpersonal skills to curb cyberslacking.

After a review of goals for a monitoring program, a list of different activities to monitor and/or control will also be provided. Then a discussion of different techniques for monitoring and associated products will be explored, followed by a discussion of fit between corporate culture and monitoring.

BACKGROUND

The Worker's Perspective

In this age of cell phones, pagers, wireless PDAs, email, and home network links, many employees feel like the employer owns them not just during the workday, but perhaps constantly. Though tiresome, the worker may accept this as an unfortunate circumstance of 21st century knowledge work. However, in the tit-for-tat that this availability demands, the employee may feel that (s)he should be allowed to use the Internet at work to take care of quick business tasks, such as paying bills, sending an email, or checking that evening's movie listings. So long as it isn't excessive, the employee may wonder why the employer even cares. Employers can and do care for many reasons, some more profound than others.

Goals for Monitoring

Why do companies monitor their employees? Organizations monitor for many reasons, including simply "because they can." An electronic monitoring effort is often difficult to establish and maintain, so before an organization begins such an effort, there should be clear monitoring goals.

The popular press is filled with stories of employees frittering away time on the Internet (Swanson, 2002). In the beginning, employees were likely to spend unauthorized time on the Internet at pornography and gambling sites, but now news and online shopping are more likely outlets for cyberslacking (Reuters, 2002). This is quite the opposite of employers' expectations when they implemented Internet connections.

Responding to these challenges, employers created acceptable use policies (AUPs). Some organizations already had AUPs implemented to keep out electronic games, and they simply modified those policies. Other organizations created new AUPs, which directly addressed the Internet's productivity threat. AUPs are useless without enforcement, but in today's litigious society, it behooves accusers to be certain of transgressions before enforcing the policy. Monitoring tools create an irrefutable log of usage which can stand as legal evidence. Some employers hope the mere threat of punishment will keep employees from cyberslacking, often with some success (Urbaczewski & Jessup, 2002). Below are listed some possible goals of a monitoring effort.

Increase Employee Productivity

The Internet was introduced into many organizations as a tool to increase employees' efficiency. While traditional IT packages provided few opportunities for employees seeking to slouch on employer time, the Internet posed an entirely different situation. Computers now had the capability to be an electronic equivalent of a water cooler, break room, or smokers' perch. To curb the potential problem of employees wasting time while appearing to be busy, an organization could implement a monitoring program which completely blocks and/or records the amount of time spent at non-work-related Internet sites. An alternative could be limiting access to frivilous sites to non-production hours only, such as during lunchtime.

Bandwidth Preservation

In some organizations, concerns are not productivity-based but rather that network bandwidth is being dominated by applications and instances not directly work-related. An example might be listening to streaming audio or watching streaming video, both constant drains on bandwidth. People can also engage in excessive file transfers across networks which results in reduced network performance. Two possible solutions to this problem are to purchase more bandwidth or limit the usage of existing bandwidth, with monitoring programs aiding in the latter solution.

Legal Liability Reduction

Along with productivity and bandwidth usage, organizations are also concerned about Internet usage from the potential exposure it brings to legal liability. Consider the following fictitious scenarios:

"Organization X today was sued for negligence, as an employee was running a child pornography server inside the corporate network."

"ABC Corporation today was sued by a former employee who is now in treatment with Gambler's Anonymous. He is charging that ABC, by placing an unrestricted Internet terminal on his desktop, essentially gave him unfettered access to the virtual casinos thriving on the Internet."

"Company B is defending itself today against a privacy lawsuit. It is charged that when an employee downloaded a file-sharing program, that program was equipped with a backdoor which allowed malicious hackers entrance into Company B's networks. These hackers then took thousands of credit card numbers and personal data from the databases..."

Other possibilities like sexual harassment suits and industrial espionage make the legal risks mount. Organizations indeed may wish to monitor Internet connections to prevent any potential legal liabilities from allowing illegal activities to be conducted on their networks.

MAIN THRUST

Different Monitoring Strategies

Once an organization decides it will monitor, it needs to know what to monitor. While Web porn is probably the most reported off-topic use of the Internet in an organization, it is certainly not the only transgression that might come from an Ethernet card. Excessive personal e-mail, filesharing, instant messaging, multimedia streaming, and Usenet browsing and posting are among other ways that employees use the corporate Internet connection for personal enjoyment.

There are several different control mechanisms that an organization might use, generally grouped into one of two categories: managerial and technical. The managerial techniques for monitoring are similar to ways that monitoring of employees has been done for decades: walking around and keeping one's eyes open. When a manager starts to wonder about an employee's performance or collegiality, then the manager starts to pay more attention to that employee's work habits.

Overall, however, the most popular means of monitoring employees is through technology. In many ways, this makes sense – a technical solution to a technological problem. Electronic monitoring operates like "big brother" (Zuboff, 1988), keeping a constant watchful eye on the network and its connected systems (or whatever subset of those systems/hours that a manager may choose to watch). Records can then be kept and offered as later "proof" of an offender's cyberslacking or lack thereof.

Electronic Monitoring Techniques

Logging at the Gateway

Many logging technologies are designed to capture and record packets as they enter and leave the organization, or at least the header information that indicates the sender, recipient, and content of the message. Gateway logging is useful in that it provides a central point of network control. However, it is difficult to accurately gauge how long an employee stares at a particular page, and if all that time (s)he is actually staring at that page or if (s)he has actually gone to lunch and returned later. Moreover, gateway logging can be defeated by the use of encryption tools like PGP (www.pgp.com, see McCullagh, 2001, for a more detailed description of an FBI case with the Philadelphia organized crime ring), or even tools like Anonymizer.com that allows a person to surf the Web anonymously using their gateways and encryption tools. In cases where these technologies are used, a separate technology might also be needed.

Spying at the Client

When gateway logging is insufficient, one can monitor and record connections directly at the source. A keystroke logging program can record everything that a person types on a computer, and many even include technologies to take screenshots or use the Web camera on the desk to prove that it was the person actually sitting at the computer and not someone who just walked up to the terminal.

Client sniffing programs are excellent at recording exactly what the user is doing with the computer at any given time. Many will record all of the user's keystrokes, mouse movements, and active windows, allowing the reconstruction of the entire computing session. Moreover, they can capture other undesirable activity, such as playing games and typing job application letters. However, these programs are not without their own faults. First of all, the manager must install the program on the user's computer, which may not be as easy as it sounds, especially with laptop and other mobile computers. Second, the program must not be detectable (and thus deletable) by the monitored employees. Managers then must sift through mountains of captured data to determine if there is any untoward activity, or enough to warrant further investigation. However, products are available which meet the previously described concerns to varying degrees, and the next section will discuss some of those products.

Software Products for Controlling Internet Usage

As mentioned previously, there are various products available to serve as control mechanisms. They are grouped in the following into five categories. Note that software products come and go, and the availability of certain products and companies are subject to market forces. The categories themselves should remain stable for the time being, although who knows what the future may hold. For example, if this chapter was being written in 1999, there would likely be no section on file-sharing.

Web Monitoring Products

As the popular press articles have largely focused on employee cyberslacking as a problem with personal Web usage, a number of products have been created to help employers manage these problems. The software products are all customizable to some degree, but there are two main classifications of these products: those that *monitor* and record Web usage, and those that actively *block* access to certain Web sites deemed inappropriate. The listing to follow, which is not intended to be exhaustive, details several of these products.

Cybersitter (www.cybersitter.com) and *NetNanny* (www.netnanny.com) are two programs that are geared largely at individuals, as they are installed on the client and maintain logs of Web pages seen by the users. *Websense* (www.websense.com) however, is designed to monitor the Web usage of an entire network. It runs near the firewall and logs all Web requests leaving the network. All of these programs can be configured to block and/or record access to certain Websites. Some of these programs can be tailored to allow different access rules at different times of day. For example, an organization may wish to allow its employees to use the Internet for shopping and other personal entertainment before and after normal business hours and on the lunch hour but not during the work day.

E-Mail Monitoring Products

E-mail can easily be monitored by simply examining accounts for incoming mail or logging the actions of the simple mail transport protocol (SMTP) server for outgoing mail. These logs are often difficult to read, especially with large volumes of mail and users. A series of products can assist in parsing the logs or searching for users or keywords, like *MIMESweeper* (www.mimesweeper.com) and *Message Inspector* (http://www.zixcorp.com/compro/messageinspector.php). Encrypted e-mail can present its own challenges, and additional policies may be necessary regarding the use of strong encryption in an organization.

Monitoring e-mail sent through popular Web-based providers like Yahoo! or Hotmail can be difficult as well, because the message never passes through the SMTP servers for the organization, nor does the organization have direct control over users' mailboxes. Monitoring these types of mail services is usually done through a general monitoring tool, as listed in another section next.

File-Sharing Monitoring Products

File-sharing has a history of waxing and waning between one of the easiest applications to monitor to one of the toughest. Users of file-sharing services often try to devise ways to run their services around and through corporate attempts to halt them. Other problems were created by users demanding that they be allowed to use these programs, especially at high-profile universities like Yale and Indiana. In those cases, something had to be done to limit the amount of bandwidth these services could use, because other legitimate traffic was being delayed. A number of hardware and software solutions cropped up to aid network managers in their quest to reduce or eliminate file-sharing traffic.

On the software side, it was mentioned above that already existing firewalls can be configured to block traffic on certain TCP (OSI layer 4) ports. Other programs, like *DynaComm I:scan* (http://www.dciseries.com/products/iscan/), are designed to examine the packets at the application layer (OSI layer 7) to determine the type of packet and whether or not to block it. Hardware solutions like *Packeteer* (www.packeteer.com) plug into the network to control the amount of bandwidth available to certain applications. Packeteer has been most popular at colleges and universities, which in general do not want to be accused of censorship or limiting access to resources, but still have to deal with bandwidth concerns amongst thousands of users.

Instant Messaging (IM) Monitoring Products

IM's "fire-and forget" nature made it one of the toughest applications to monitor for a long time. The problem was exacerbated because the employers generally did not control the IM servers or the clients. In 2002, applications were created which successfully monitor IM applications and content, implemented largely to comply with certain US Securities and Exchange Commission (SEC) requirements on keeping logs of all transactions between brokerage houses and their customers. IM is usually not monitored to conserve bandwidth, but it can become a productivity drain similar to e-mail.

Facetime (www.facetime.com) is probably the leader in monitoring organizational IM. The Facetime product can record all IM activity, important in SEC-regulated businesses and government agencies with freedom of information requirements. *Vericept* (www.vericept.com) also has all-purpose monitoring capabilities, but focuses largely on monitoring, blocking and recording IM activity. Organizations looking to block and monitor IM and communicate securely using encryption might turn to *Compliancer Hub* (www.communicatorinc.com).

General Monitoring (Spying at the Client) Tools

There are a series of more powerful, less-specific tools available for almost total user monitoring, classified as general monitoring tools. These tools are installed at the client and can create a record of almost everything a user does with the computer. Records can be written to a network database or even e-mailed to another account.

TrueActive (www.winwhatwhere.com) is probably the original instantiation of this type of program, marketed largely to businesses for specific monitoring purposes. There are also many programs targeted at individuals, including *Spy Agent* (http://www.spytech-web.com/spyagent.shtml) and *eBlaster* (http://www.spectorsoft.com/products/eBlaster_Windows/index.html). Managers are often surprised at the sheer amount of data these programs provide, quite often more than a manager really wants to know. One should carefully consider the implications before implementing a general monitoring tool. Employees may react strongly to such total monitoring.

FUTURE TRENDS

Seeking the Recommended Fit between Goals and Monitoring Solutions

If productivity is a major concern, one might begin with a passive but comprehensive logging tool. Cyberslacking can be easily seen and measured, but it is done unobtrusively. When an employee's productivity falls, as observed traditionally, the technical data showing cyberslacking are available. This can be used for implementing positive disciplinary measures, or for supporting a termination. Periodically and when a situation occurs, employees should be reminded of the organization's policy about personal Internet usage and resulting enforcement actions. This is not to embarrass potential offending parties, but rather to keep the policy salient.

If legal liability is the major concern, minimally intrusive means can also be used for the monitoring and recording of transmitted data. In December 2002, five major Wall Street brokerage houses were fined $1.65 million for not keeping e-mails the required two years. An organization can avoid these types of penalties by simply logging and maintaining records of Internet traffic without any review, except on an as required basis. The RIAA and other entertainment industry groups in early 2003 began warning organizations to actively ensure that their employees were not using company networks to access copyrighted music and video files, lest the companies themselves be held liable. The RIAA has been supplying offenders' IP addresses and access times to companies and universities, identifying individuals who may have traded in music and video files. In the end, a company pursuing this strategy would be more concerned with record-keeping than record-reviewing.

An organization pursuing the third goal, bandwidth preservation, can likely use the least passive of all monitoring tools – simply observing bandwidth usage spikes where they occur, and witnessing their relationship to organizational goals. A firm that sees bandwidth constantly full with apparently work-related material may want to investigate adding additional bandwidth resources. At the same time, organizations that suddenly block or "throttle" access to popular Internet destinations known for non-work-related information will likely solve many problems. Employees are more likely to realize that they need to get to work or identify another means of entertainment than they are to complain and cause a ruckus over no longer being able to access these sites at all or at high speeds.

CONCLUSION

This article has detailed several types of control mechnisms. It has listed goals for monitoring, applications to monitor, and means of accomplishing the monitoring. Furthermore, it lists names of actual tools that can be used to accomplish the monitoring. What this article so far has not discussed is the viability and consequences of the monitoring.

In a series of studies, it was found (not surprisingly) that Internet monitoring indeed has a significant positive effect on keeping employees on task (Urbaczewski & Jessup, 2002). However, it was also found that monitored employees were more likely to turnover and less likely to participate in other organizational activities. Could this happen in all organizations? Possibly, but the key to remember when establishing monitoring is:

"Make the monitoring strategy fit the corporate culture"

Some organizations have a culture where monitoring of most (if not all) activities is simply expected. These industries generally include those handling cash or financially-related transactions (e.g., banks, casinos) or deal with physical and/or national security (CIA, FBI, R&D labs, etc.). In these cases, monitoring fits the culture, and if the organization is already monitoring employees in other ways, it also makes sense to monitor computer systems.

Other organizations have cultures which generally detest monitoring. Some industries, like publishing, academia, and other "civil liberties" organizations do not generally monitor their employees, as their foundations are centered on freedom of speech and the unfettered search for the truth. The introduction of monitoring in these industries will likely result in a culture clash between employees and management (Simmers, 2002).

The question then becomes "how does one then reap the benefits of monitoring without angering employees and creating unwanted stress in the organization?" Communication of the control mechanisms to the employees, with a clear understanding of how it supports corporate goals and principles, is key. Explicit statements of who will be monitored, what will be monitored, and when monitoring will occur should also be communicated to the employees, largely through AUPs.

REFERENCES

McCullagh, D. (2001). Scarfo: Feds plead for secrecy. Available online at *http://www.wired.com/news/politics/0,1283,46329,00.html*

Reuters (author unknown). New sites top spots for work surfing, as printed in CNN.com,

http://www.cnn.com/2002/TECH/internet/09/23/workplace.surfing.reut/index.html

Simmers, C.A. (2002). Aligning Internet usage with business priorities. *Communications of the ACM, 45*(1), 71-74.

Swanson, S. (2002). Employers take a closer look. *Information Week, 901*, 40-41.

Urbaczewski, A., & Jessup, L.M. (2002). Does electronic monitoring of employee Internet usage work? *Communications of the ACM, 45*(1), 80-83.

Zuboff, S. (1988). *In the Age of the smart machine: The future of work and power*. New York: Basic Books.

www.communicatorinc.com

www.cybersitter.com

www.dciseries.com/products/iscan/

www.facetime.com

www.mimesweeper.com

www.netnanny.com

www.packeteer.com

www.spectorsoft.com/products/eBlaster_Windows/index.html

www.spytech-web.com/spyagent.shtml

www.winwhatwhere.com

www.zixcorp.com/compro/messageinspector.php

KEY TERMS*

Acceptable Use Policy (AUP): A policy created in an organization to outline the permitted and restricted uses of the company's networks and computer systems.

Bandwidth: Colloquially, the amount of network capacity available for a connection.

Blocking: A means of disallowing access to Internet content and services by restricting access at the corporate gateway.

Cyberslacking: The process of using the Internet to waste time during a workday, similar to how an employee might spend time in a colleague's office or on the telephone.

Logging: Creating a record of all employee Internet usage.

OSI Layer 4: The Transport Layer of a network connection, which provides service differentiation and connection management. From the Open Systems Interconnect model.

OSI Layer 7: The Application Layer of a network connection. It describes how applications work with the network operating system. From the Open Systems Interconnect model.

*Definitions specific to their use in this chapter.

Motivation for Using Microcomputers

Donaldo de Souza Dias
Federal University of Rio de Janeiro, Brazil

INTRODUCTION

Information technology implementation is an intervention we make in order to improve the effectiveness and efficiency of a sociotechnical system. Using microcomputers to help individuals perform their jobs and tasks is one of the most important actions we take when implementing this technology effectively. Information systems effectiveness has been extensively studied using, mainly, user satisfaction and quality of information constructs to evaluate users' acceptability (Iivari & Ervasti, 1994; Ives et al., 1983; Neumann & Segev, 1979). However, sometimes, the result of this intervention is not successful and may even generate difficulties related to people participation in the process. This leaves us with a question: What motivates individuals to use microcomputer technology in their daily activities?

Theorists and empirical researchers have been trying to understand the relevant motivators for the implementation and use of computer technology based on the idea that people make an effort if an activity is enjoyable or offers external rewards (Igbaria et al., 1996; Schwartz, 1983). They have been aiming to find out how individuals feel motivated to work with computers, and what motivates them to use computers in their daily activities.

BACKGROUND

Computer and information technology usage is determined by intrinsic as well as extrinsic motivation (Deci, 1975; Igbaria et al., 1996). The main driving forces considered in the literature as motivators for computer and information technology adoption are perceived usefulness, perceived ease of use, and perceived enjoyment (Davis, 1986, 1989; Igbaria et al., 1996). However, it is known that some individuals create personal obstructions to using technology (Pirsig, 1981), particularly, microcomputer technology (Igbaria & Parasuraman, 1989; Martocchio, 1994). They resist microcomputers usage and experience anxiety when they have to deal with them. We present results found in previous studies for relations and comparisons among the motivational forces above (Dias, 1998a, 1998b, 2002; Dias et al., 2002). The results presented here, all statistically significant at $p < 0.05$, were based on constructs measured using the instrument developed in Dias (1998a) and presented in the Appendix.

MAIN MOTIVATIONAL FORCES

Figure 1 shows the results for the relationships among perceived enjoyment, perceived ease of use, and perceived usefulness found in Dias (1998a). The author focused on the motivators perceived usefulness, perceived ease of use, and perceived enjoyment. The aim was to find out how Brazilian operations managers felt about using computer technology in their workplaces, how the perceived usefulness of computers is affected by ease of use and users' enjoyment in working with them, and how to find opportunities to act according to this acquired knowledge, in order to increase the quality of microcomputer technology usage in organizations. In his study, the author emphasized the relationships among these perceived motivators for using microcomputer technology. The impact of the motivators on systems usage or microcomputer adoption was considered to be beyond the scope of his research.

The path analysis model used was based on the natural precedence of intrinsic motivational factors over extrinsic motivational factors, as proposed by the Freudian theory of psychoanalysis (Freud, 1976).

The data for that study were gathered using a questionnaire administered personally to 79 Executive MBA students at a Brazilian university. Respondents held managerial positions in 55 companies, ranging from small firms to large corporations, located in Rio de Janeiro. The average age of respondents was 36, and they had an average of 11 years working experience. All of the participants were college graduates. Managers said they used microcomputer technology mainly because they perceived it as a useful tool to increase the quality of their work, to accomplish tasks more quickly, and to increase the productivity of their jobs.

Figure 2 shows results of a model in which computer anxiety and enjoyment were considered as antecedent variables to ease of use and usefulness (Dias, 1998b). We found that managers who were more anxious about computer technology tended to find it more difficult to use. On the other hand, enjoyment had a positive direct effect on ease of use and usefulness, as stated before.

M

Figure 1. Relationships among enjoyment, ease of use, and usefulness.

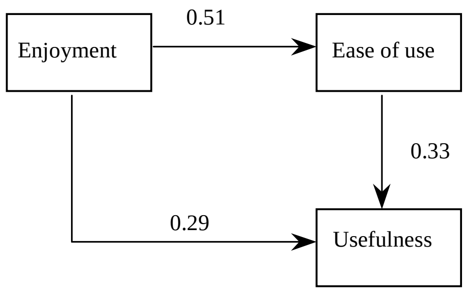


Figure 2. Anxiety and enjoyment as antecedent variables.

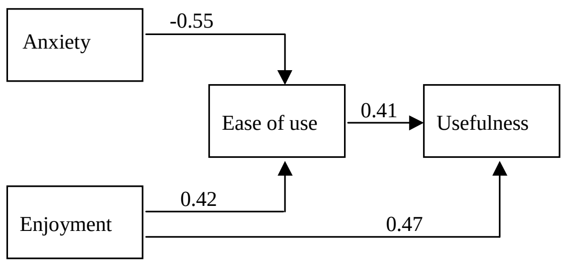
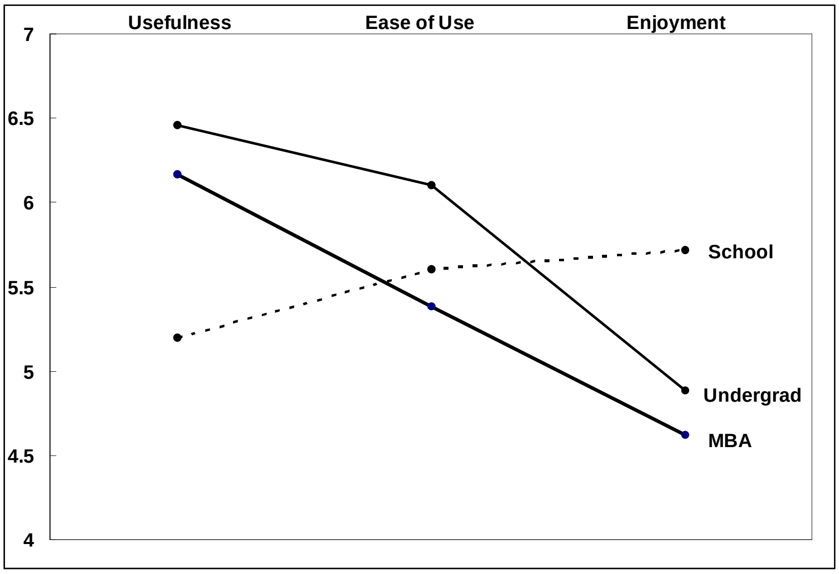


Figure 3. Motivation level for graduate, undergraduate, and elementary school students.



In a study made at a private university located in Rio de Janeiro, with data gathered from 336 undergraduate computer information systems students, Dias et al. (2002) tested the influence of some antecedent variables on enjoyment, ease of use, and usefulness. They found that (a) the fact that a student worked part-time in an area related to information technology positively influenced his or her perception of how easy it was to use microcomputers; (b) enjoyment with microcomputers seemed to decrease as students attained seniority in the university; and (c) older students perceived greater usefulness for microcomputers.

Level of education and age have shown influence on microcomputer attitudes (Igbaria & Parasuraman, 1989). Dias (2002) did a study on the motivation for using microcomputers among different classes of users. He aimed at finding out how graduate, undergraduate, and elementary school students, which represent very specific strata of educational level and age, would differ on the motivational factors examined here. The data for his study were gathered as follows:

- Fifty-three Executive MBA students of a leading Brazilian public university: The average age of respondents was 36, and they had an average of 11 years working experience, all participants were managers and had a college degree.
- Forty-six students aiming for degrees in Business Administration at a private university located in Rio de Janeiro: The average age of respondents was 22.
- Thirty-nine elementary schools students enrolled in the fourth to eighth grades of private (82%) and public schools located in the city of Rio de Janeiro: The students used microcomputers regularly at school, at home, or at relatives' homes.

Factor analysis confirmed that the statements for usefulness, ease of use, and enjoyment constituted three

Table 1. Actions for successful implementation of microcomputers.

Action	Motivational Forces
Develop friendly systems	Ease of use/enjoyment
Encourage user participation in systems development	Usefulness/ease of use
Implement intensive user training	Ease of use
Sell the system to users	Usefulness
Facilitate access to microcomputers	Ease of use
Respond to user needs	Usefulness
Use up-to-date technology	Usefulness/enjoyment
Align business and information technology	Usefulness

distinct perception constructs for the three classes of users. It confirmed the existence of three factors that accounted for 63.5% of the total variance in the 138 interviewees.

Figure 3 shows the motivational profile for graduate, undergraduate, and elementary school students on the usage of microcomputers.

Dias (2002) found that there was a statistically significant difference in the perceived enjoyment, perceived ease of use, and perceived usefulness of using microcomputers among MBA, undergraduate, and elementary school students. MBA and undergraduate students were most motivated by microcomputer technology usefulness. Elementary schools students mostly enjoyed the play aspect of microcomputers.

We did a study with managers in order to better understand these motivational forces and to generate recommendations for increasing the motivation for using microcomputers. Thirty-six managers were split into six working groups and asked to (1) discuss their experience in using microcomputers in their organizations (taking into account personal resistances for using computers in the workplace), situations in which this technology is used compulsively, motivators they thought were important for using computers for task execution, and organizational culture; and (2) propose actions they thought should be taken in the implementation of microcomputers in their workplaces that would lead to a more effective usage of this technology.

The main actions suggested and the corresponding motivational forces are presented in Table 1. We found that in order to attract adult users, information systems should be friendlier, use up-to-date interfaces, and be developed with user participation. In addition, it is important to offer intensive user training, to market the new systems to users heavily, and to fully meet the users' needs in the specification of the systems.

Although managers showed less unanimity on the motivating power of enjoying themselves while using microcomputers, this motivation factor clearly showed its significance. Enjoyment seems to serve more as self-motivation, while usefulness seems to be linked to a feeling of duty—something that the organization expects managers to do or to attain.

FUTURE TRENDS

According to the World Summit on the Information Society (WSIS, 2003):

The global Information Society is evolving at breakneck speed. The accelerating convergence between telecommunications, broadcasting multimedia and information and communication technologies is driving new products and services, as well as ways of conducting business and commerce. At the same time, commercial, social and professional opportunities are exploding as new markets open to competition and foreign investment and participation.

The modern world is undergoing a fundamental transformation as the Industrial Society that marked the 20th century rapidly gives way to the Information Society of the 21st century. This dynamic process promises a fundamental change in all aspects of our lives, including knowledge dissemination, social interaction, economic and business practices, political engagement, media, education, health, leisure and *entertainment. We are indeed in the midst of a revolution, perhaps the greatest that humanity has ever experienced.*

Several new concepts lay the foundation for prospering in the next form of information and communication technology. The future microcomputer usage will be based upon ubiquity, universality, uniqueness, and unison (Watson et al., 2002). The keys to managing network-driven firms will be based on this notion of ubiquitous networks. In addition, we have to consider the new generation of students entering business schools around the world. These students are quite different from the students of the last decades, and we have to consider the technological, social, and cultural changes that are developing. The implications for the next-generation organization and information systems based on the Internet and mobile microcomputer technology have to be considered, and the motivational forces discussed here will probably have to be readdressed.

CONCLUSION

The research described here on motivation for using microcomputers offers several contributions for theory and practice. It confirmed that there are three positively interrelated motivators for computer technology usage: perceived usefulness, perceived ease of use, and perceived enjoyment. It also found that perceiving microcomputers as easy to use is negatively related to people's anxiety toward using them.

The research also compared the perceived reasons for using microcomputers among different classes of users and showed that there are significant differences in the motivation for using them among graduate, undergraduate, and elementary school students. Elementary school students have greater enjoyment in using microcomputers than MBA and undergraduate students in Business Administration. MBA and undergraduate students in Business Administration perceive greater usefulness in the usage of microcomputers than elementary school students, and they said that they used microcomputers mainly because this increased the quality of their work and allowed them to accomplish tasks more easily and quickly. Undergraduate students in Business Administration think it is easier to use microcomputers than do MBA and elementary school students. MBA students think it is most difficult to use microcomputers in their daily tasks.

We should emphasize that the motivational forces studied here could not apply equally to different countries, cultures, or organizational environments. A study made by Straub et al. (1997) compared the technology acceptance model (TAM) (Davis, 1986) across three different countries: Japan, Switzerland, and the United States. The study was conducted by administering the same instrument to employees of three different airlines, all of whom had access to the same computer technology innovation—e-mail. They used perceived ease of use and perceived usefulness as independent variables. The results indicated that the model holds both for the United States and Switzerland, but not for Japan, suggesting that the model may not predict motivation for technology use across all cultures. Harris and Davidson (1999) examined microcomputer anxiety and involvement of groups of microcomputer-using students in six developing countries: China, Hong Kong, Malaysia, New Zealand, Tanzania, and Thailand. Differences in computer anxiety were found to exist between some of the groups, which were probably attributable to demographic factors. Differences

Appendix 1

	Fully disagree Fully agree
1. I do not see time go by when I am using a computer.	1 2 3 4 5 6 7
2. Using computers enables me to accomplish my tasks more quickly.	1 2 3 4 5 6 7
3. I feel motivated to perform activities using computers.	1 2 3 4 5 6 7
4. I find it is easy to use a computer to do my work.	1 2 3 4 5 6 7
5. Using computers is fun.	1 2 3 4 5 6 7
6. Using computers improves my job productivity.	1 2 3 4 5 6 7
7. Using computers makes it easier to perform my tasks.	1 2 3 4 5 6 7
8. Using computers is exciting.	1 2 3 4 5 6 7
9. Using computers increases the quality of my work.	1 2 3 4 5 6 7
10. I feel anxiety when I have to perform a task using microcomputers.	1 2 3 4 5 6 7
11. I think it is easy to use computers.	1 2 3 4 5 6 7
12. Using computers is pleasant.	1 2 3 4 5 6 7
13. I find computers useful for my job.	1 2 3 4 5 6 7
14. I think we should use computers as much as possible.	1 2 3 4 5 6 7

were found to exist between the microcomputer involvements of some of the groups, which could be attributed to cultural factors.

APPENDIX

Please indicate your agreement or disagreement with the following statements, related to the usage of computers in the workplace, using a 7-point scale ranging from *fully disagree* (1) to *fully agree* (7). Please refer to Appendix 1 on the previous page.

REFERENCES

Davis, F. (1986). *A technology acceptance model for empirically testing new end user information systems: Theory and results*. Unpublished Doctoral Dissertation. Boston, MA: MIT.

Davis, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340.

Deci, E. (1975). *Intrinsic motivation*. New York: Plenum Press.

Dias, D. (1998a). Managers' motivation for using information technology. *Industrial Management and Data Systems, 98*(7), 338–342.

Dias, D. (1998b). Intrinsic and extrinsic motivators for microcomputer usage. In M. Khosrow-Pour (Ed.), *Effective utilization and management of emerging information technologies* (pp. 664–667). Hershey, PA: Idea Group Publishing.

Dias, D. (2002). Motivation for using information technology. In E. Szewazak & C. Snodgrass (Eds.), *Human factors in information* systems (pp. 55–60). Hershey, PA: IRM Press.

Dias, D., Mariano, S., & Vasques, R. (2002). Antecedents of Internet use among Brazilian information systems students. *Issues in Information Systems, 3*, 144–150.

Freud, S. (1976). *Obras Completas*, Rio de Janeiro: Imago, 22, 75–102.

Harris, R., & Davidson, R. (1999). Anxiety and involvement: Cultural dimensions of attitudes toward computers in developing societies. *Journal of Global Information Management, 7*(1), 26–38.

Igbaria, M., & Parasuraman, S. (1989). A path analytic study of individual characteristics, computer anxiety and attitudes toward microcomputers. *Journal of Management, 15*(3), 373–388.

Igbaria, M., Parasuraman, S., & Baroudi, J. (1996). A motivational model of microcomputer usage. *Journal of Management Information Systems, 13*(1), 127–143.

Iivari, J., & Ervasti, I. (1994). User information satisfaction: IS implementability and effectiveness. *Information and Management, 27*(4), 205–220.

Ives, B., Olson, M., & Baroudi, J. (1983). The measurement of user information satisfaction. *Communications of ACM, 26*(10), 785–793.

Martocchio, J. (1994). Effects of conceptions of ability on anxiety, self-efficacy, and learning in training. *Journal of Applied Psychology, 79*(6), 819–825.

Neumann, S., & Segev, E. (1979). A case study of user evaluation of information characteristics for systems improvement. *Information and Management, 2*(6), 271–278.

Pirsig, R. (1981). *Zen and the art of motorcycle maintenance*. New York: Bantam Books.

Schwartz, H. (1983). A theory of deontic work motivation. *The Journal of Applied Behavioral Science, 19*(2), 204–214.

Straub, D., Keil, M., & Brenner, W. (1997). Testing the technology acceptance model across cultures: A three country study. *Information and Management, 33*, 1–11.

Watson, R., Pitt, L., Berthon, P., & Zinkhan, G. (2002). U-Commerce: Expanding the universe of marketing. *Journal of the Academy of Marketing Science, 30*(4), 333–347.

WSIS. (2003). World summit on the information society: Newsroom fact sheets: The challenge, Geneva, Switzerland. Retrieved May 6, 2004, from *http://www.itu.int/wsis/newsroom/fact/whynow.html*

KEY TERMS

Computer Anxiety: Degree to which an individual is nervous in his or her interaction with computers; the uneasiness some people feel when they have to use a microcomputer. Anxiety results from a danger or a danger threat. As a feeling, it has a clearly unpleasant character.

Computer Self-efficacy: A judgment of one's capability to use a computer. It incorporates judgments of an individual on his or her skills to perform tasks using a microcomputer.

Extrinsic Motivation: Motivation that derives from what you obtain from engaging in an activity. An example of extrinsic motivation for using microcomputers is using it because you think it is useful for your job.

Intrinsic Motivation: Motivation that derives from the activity itself. An example of intrinsic motivation for using microcomputers is using it because you enjoy it.

Microcomputers Ease of Use: User perception on how simple and easy it is to understand and use microcomputers; degree to which an individual believes that using a particular computer system would be free of physical or mental effort.

Microcomputers Enjoyment: The extent to which the activity of using microcomputers is perceived as being enjoyable in its own right, apart from any performance consequences. It encompasses the feelings of joy, elation, or pleasure associated by an individual to a particular act.

Microcomputers Usefulness: The degree to which an individual believes that using a particular computer system would enhance his or her job performance.

Sociotechnical System: One that focuses on the interaction between the technical and social subsystems that exists in any work situation. As there is always a social system operating along with any technological system, we need to jointly design these two interrelated systems to get the best results from any work system.

Multicast Routing Protocols, Algorithms and its QOS Extensions

D. Chakraborty
Tohoku University, Japan

G. Chakraborty
Iwate Prefectural University, Japan

N. Shiratori
Tohoku University, Japan

INTRODUCTION

The advancement in optical fiber and switching technologies has resulted in a new generation of high-speed networks that can achieve speeds of up to a few gigabits per second. Also, the progress in audio, video and data storage technologies has given rise to new distributed real-time applications. These applications may involve multimedia, which require low end-to-end delay. The applications' requirements, such as the end-to-end delay, delay jitter, and loss rate, are expressed as QoS parameters, which must be guaranteed. In addition, many of these new applications involve multiple users, and hence the importance of multicast communication. Multimedia applications are becoming increasingly important, as networks are now capable of carrying continuous media traffic, such as voice and video, to the end user. When there is a lot of information to transmit to a subset of hosts, then multicast is the best possible way to facilitate it. This article addresses different multicast routing algorithms and protocols. We have also discussed about the QoS multicast routing and conclude this article with mobile multicasting.

BACKGROUND

Multicast consists of concurrently sending the same information to a group of destinations such that exactly one copy of the packet traverses each link in the delivery tree. Interactive multicast applications include video conferencing, computer-supported cooperative work, and virtual whiteboard applications. Other multicast applications such as remote education require a lesser amount of interaction. A third group of multicast applications are noninteractive, for example mailing lists and some real-time control applications.

In a true multicasting, the least-cost path from the source to each network that includes members of the multicast group is determined. This results in a spanning tree of the required configuration. This is not a full spanning tree, but includes at least those networks containing group members. The source transmits a single packet along the spanning tree. The packet is replicated by routers only at branch points of the spanning tree.

When the same data need to be sent to only a subset of the clients on the network, both broadcast and multiple unicast methods waste network bandwidth by sending multiple copies of the data. Broadcast wastes bandwidth by sending the data to the whole network, whether the data are wanted or not. Broadcast also needlessly slows the performance of client machines. Each client must process the broadcast data, whether the client is interested or not. Multicast falls between these two extremes. It is useful for building distributed pseudo-real-time applications such as videoconferencing and audioconferencing. However, its use is not restricted to these kinds of applications. Any application that involves sending copies of data to multiple places can benefit. For instance, one could distribute network routing tables to all routers in an enterprise, while not burdening all of the workstations with processing these messages. If one has to send audio and video, which needs a huge amount of bandwidth compared to Web applications, multicast is the best possible solution. Multicasting has three advantages over broadcasting, particularly when the recipient group size is significantly smaller than all the possible recipients in a network. First, by routing a message only where it needs to go, multicasting conserves network bandwidth, facilitating more efficient use of the network infrastructural resources. It can result in user toll charges that are lower than broadcast delivery. Second, data transmission can be restricted to only the paying subscribers for services such as video on demand. Third, the dissemination of sensitive information can be limited to a select group of recipients.

MULTICAST ROUTING PROTOCOLS

- Distance-Vector Multicast Routing Protocol (DVMRP) designed to deliver multicast datagrams to a group of hosts across the Internet. DVMRP constructs source-based multicast delivery trees using the reverse-path multicasting (RPM) algorithm. In DVMRP a datagram from a multicast source is initially propagated downstream by a designated multicast router to all other multicast routers, regardless of whether they have multicast group members or not. Multicast routers without downstream or local members send explicit prune messages upstream to remove themselves from the distribution tree. The net effect is a source-specific shortest path tree, with the members forming the leaves of the tree. Once the multicast tree is set up, multicast routers keep track of the reverse path to the multicast source. If an arriving datagram does not come through the interface that the router uses to send datagrams to the source of the multicast, then the arriving datagram is dropped.
- Core-Based Trees (CBT): The CBT protocol was first discussed in the research community and then standardized by the IETF (Ballardie, 1997). The CBT uses the basic sparse mode paradigm to create a single shared tree used by all sources. The tree is rooted at a core. All sources send their data to the core, and all receivers send explicit join messages to the core. CBT uses only a shared traffic tree, and is not designed to use shortest path trees. CBT uses bidirectional shared trees, but PIM-SM uses unidirectional shared trees. Bidirectional shared trees involve slightly more complexity, but are more efficient when packets are traveling from a source to the core across branches of the multicast tree. In this case, instead of only sending "up" to the core, packets can also be sent "down" the tree. While CBT has significant technical merits and is on par technically with PIM-SM, few routing vendors provide support for CBT.
- Protocol Independent Multicast – PIM: PIM provides both dense mode (PIM-DM) (Deering, 1998) and sparse mode (PIM-SM) (Estrin, 1998) group membership. As the name implies, the multicast architecture is independent of the protocol employed for unicast routing. PIM can scale to wide-area networks, and is particularly attractive for sparse multicast group. Essentially PIM can use either the shared tree approach of CBT or the shortest-path approach of DVMRP, with appropriate choice made on a per group or per host basis. The PIM architecture relies upon choosing a suitable rendezvous point (RP), similar to a core in CBT, when constructing the multicast delivery tree for a group. The RP provides a place for multicast sources to "meet" multicast recipients.
- Border Gateway Multicast Protocol (BGMP) (Thaler, Estrin & Meyer, 1998**):** is a new inter-domain multicast routing protocol that addresses many of the scaling problems of earlier protocols. BGMP attempts to bring together many of the ideas of previous protocols and adds features that make it more service provider friendly. BGMP is designed to be unified inter-domain multicast protocol in much the same way that Border Gateway Protocol (BGP) is used for unicast routing.

MULTICAST ROUTING ALGORITHMS

Different multicast routing algorithms are summarized in this section, both unconstrained and constrained. A taxonomy of these multicast routing algorithms is given in Table 1.

- Shortest Path Tree: A shortest path algorithm minimizes the sum of the weights on the links along each individual path from the source to a receiver in the multicast group. If unit weight is used per hop, the resulting tree is a least-hop tree. If the weight represents the link delay, the resulting tree is a least-delay tree. The Bellman-Ford and Dijkstra algorithms are the two best-known shortest path algorithms. Both are exact and run in polynomial time.
- Minimum Spanning Tree: A minimum spanning tree is a tree that spans all the group members and minimizes the total weight of the tree. The well-known centralized minimum spanning tree algorithm is Prim's algorithm. In Prim's algorithm the tree construction starts from an arbitrary root node and grows until the tree spans all the nodes in the network. Minimum spanning tree algorithms run in polynomial time and can be used to solve tree optimization problems.
- Steiner Tree: The Steiner tree aims to minimize the total cost of the multicast tree, and is known as NP-complete. If the multicast group includes all nodes in the network, the Steiner tree problem reduces to the minimum spanning tree problem. Unconstrained Steiner tree algorithms can be used to solve tree optimization problems. Tree cost optimization for a whole session duration for unconstrained routing algorithm is proposed in Chakraborty, Chakraborty, Pornavalai and Shiratori (1999). Salama et al. (Salama, Reeves & Viniotis, 1997) gave very good reviews on

Table 1. A taxonomy of multicast routing problems

	No optimization	**Link optimization**	**Tree optimization**
No constraints		Link optimization (polynomial time complexity)	Tree optimization (NP-complete complexity)
Link constraints	Link-constraints (Polynomial complexity) Multiple-link-constrained (Polynomial complexity)	Link-constrained link optimization (Polynomial time complexity)	Link-constrained tree optimization (NP-complete complexity)
Tree constraint	Tree-constrained (Polynomial time complexity) Multiple-tree-constrained (NP-complete complexity)	Tree-constrained link optimization (Polynomial time complexity)	Tree-constrained tree optimization (NP-complete complexity)
Link and tree constraints	Link and tree-constrained (Polynomial time complexity)		Link and tree-constrained tree optimization (NP-complete complexity)

algorithms on the Steiner tree problem. A summary can be found in Wang & Hou (2000).

Other than the these algorithms, it is worth mentioning four of the approaches that have been discussed in the Internet community:

- Addressable Internet Multicast (AIM) by Brian Levine et al. (Levine & Garcia-Lun-Aceves, 1997) attempts to provide explicit addressing of the multicast tree. The routers run an algorithm to label all the branch points uniquely, and then make these labels available to end systems. This allows numerous interesting services or refinement of multicast services to be built.
- Explicitly Requested Single-Source (Express) by Hugh Halbrook et al. (Holbrook & Cheriton, 1999), is aimed at optimizing multicast for a single source. The proposal includes additional features such as authentication and counting of receivers, which could be added to many other multicast protocols usefully. Express makes use of an extended address (channel + group) to provide routing without global agreement on address assignment.
- Connectionless Multicast (CM) (Oooms et al., 2002) is a proposal for small, very sparse groups to be implemented by carrying lists of IP unicast addresses in packets. The scheme may be well suited to IP telephony application where a user starts with a unicast call, but then adds a third or fourth participant.

REAL-TIME MULTICASTING AND QOS

Multicast applications have been consistently increasing on the Internet. Most of these applications such as video conferencing, video-on-demand and Webcasting are multimedia oriented, which usually require large bandwidth, stringent delay and delay jitter, while the multicast protocols running on today's Internet are for best-effort services that have no guarantee for these metrics. There is an urgent need to design and implement multicast protocols with QoS guarantee, and there have been considerable efforts made addressing this issue in recent years.

The basic task of multicast communication is to deliver traffic through a multicast tree. If the objective is to find the minimum cost tree and with QoS constraints added, the problem is usually termed as a Constrained Steiner Minimum Tree (CSMT) problem.

To solve the CSMT problem there are many heuristic solutions available. Although these heuristics give insights into how we can efficiently build multicast trees, they are more theoretical than practical, and may not be suitable for the Internet environment. Hence another group of researchers developed some practical QoS multicast protocols.

HEURISTIC FOR CSMT

The QoS constraints for real-time multicast applications usually include three metrics: delay, delay jitter and delay variation. Most of the heuristic algorithms are for constraints' SMT problem. We summarize them as follows:

- Constrained Shortest Path Tree (CSPT): The basic idea is to find an unconstrained shortest path tree (SPT) first. If the end-to-end delay to any group member violates the delay bound, the path from the source to that group member is replaced by the path with the least delay. A SPT can be constructed by merging the shortest paths from the source to each of the destinations. The worst case complexity of CSPT algorithm is $O(v^2)$, where v is the number of nodes in the network.
- Constrained Minimum Spanning Tree (CMST): Basically, this algorithm constructs the broadcast tree rooted at the source that does not violate the delay constraint using Prim's minimum spanning tree algorithm. This tree may not span all the nodes in the network. However, all the destinations should be in this tree if there exists a QoS satisfying tree. Then this tree is pruned beyond the multicast nodes. The worst-cast complexity is $O(v^2)$, and similarly CMST algorithm can always find a path when there exists one.
- KPP (Kompella, Pasquale & Polyzos, 1993)**:** This algorithm is based on KMB (Kou, Markowsky & Berman, 1981) algorithm. Similar to KMB algorithm, KPP algorithm first constructs a constrained distance network, and then follows the steps in KMB algorithm to find the CSMT. However, the delay constrained shortest path problem is proved to be NP-complete if the link delays and the delay bound are real values. The worst case complexity of KPP is $O(\Delta v^3)$, which is dominated by the calculation of the constrained distance network, where Δ is delay bound.
- Constrained Adaptive Ordering (CAO): This algorithm is based on shortest paths heuristic (SPH), which is another heuristic algorithm for SMT problem. It begins by arbitrarily choosing a group member node, and then adds the member node that is closest to the already built tree at each iteration, until all the group members attach to the tree. The CAO algorithm calculates the constrained shortest path to the tree at each iteration instead of the unconstrained shortest path. The constrained Bellman-Ford algorithm is used in the calculation. The CAO can also always find a QoS satisfying path if there exists one, but there are cases in which the running time of the algorithm grows exponentially.

Practical QoS Multicast Algorithms and Protocols

All these algorithms may not be deployable in practical networks. This is because these algorithms have excessive computation overhead, require knowledge about the global network state, and do not handle dynamic group membership. Hence, some QoS multicast algorithms and protocols with practical use are developed.

For source-based trees, Guerin et al. (Guerin et al., 1998) proposed a QoS extension of OSPF, with the minimum bandwidth requirement as the QoS parameter. Although the extension focuses on unicast flows, it can be extended with moderate modification to multicast flows as well.

Calberg et al. (Carlberg & Crowcroft, 1997) proposed the Yet Another Multicast (YAM) protocol. In this protocol, a new router that intends to join a multicast tree does a *bid-order* broadcast with limited scope using the TTL field. Banerjea et al. (Yan, Faloutsos & Banerjea, 2002) extended YAM and proposed the QoSMIC protocol. In this protocol, both local search with bidding and multicast tree search are used to locate routes. Specifically, a joining router broadcasts *bid-order* messages and at the same time sends a multicast-join message to a manager router. If the manager router has sufficient knowledge of the tree and the network topology, it sends bid-order messages to a set of selected candidates on-tree routers; otherwise, the manager router joins the multicast group and multicasts a bid-order message on the tree. On-tree routers that receive bid-order messages from either the manager router or the requesting router then respond by unicasting bid messages to the joining router. Bid messages collect dynamic QoS information on their way to the requesting router. QMRP (Chen, Nahrstedt & Shavitt, 2000) is new Qos-aware Multicast Routing Protocol. QMRP achieves scalability by significantly reducing the communication overhead of constructing a multicast tree; yet it retains a high chance of success. This is achieved by switching between single-path routing and multiple-path routing according to the current network conditions. MQDMR (Chakraborty, Chakraborty & Shiratori, 2003) is a modified Bellman-Ford algorithm based on Weighted Fair Queuing (WFQ). It utilizes the time duration of the participating nodes to find a path that satisfies multiple QoS constrains, if there exists one.

Having said that, it also should be mentioned that the Internet community is sharply divided into two groups on the issue of the need of QoS-sensitive routing. To some, the need for QoS sensitive routing is self-evident. However, a significant section of the community doubts the importance of QoS-sensitive routing. Some of the objections are technical: a QoS-sensitive network may be unstable; or it may not scale well enough to be applicable to the Internet. But the most fundamental objections are related to the very need for QoS-sensitive routing.

FUTURE TRENDS

A recent development is dynamically reconfigurable wireless ad hoc networks to interconnect mobile users for applications ranging from disaster recovery to distributed collaborative computing. In a mobile environment, the network not only must manage multicast group membership and establish the necessary routes, but also must contend with the fact that the established routes are themselves transient in nature. Regardless of the network environment, multicast communication is a very useful and efficient means of supporting group-oriented applications. This is especially the case in mobile/wireless environments where bandwidth is scarce and hosts have limited power. Another important application of ad hoc network is in critical situations such as disaster recovery or battlefield scenario. The fact that the network must deal not only with the dynamic group membership, but also with the dynamic location characteristics of mobile hosts, makes multicast in a mobile environment a challenging problem.

CONCLUSIONS

The aim of this article is to give an overview of multicasting and its different routing schemes over communication network and possible applications. We introduced the concept of multicasting and how it works. Then we discussed its routing protocols and algorithms. The advantages of multicasting have also been described, which is the main reason for the recent increase in interests in multicasting. It seems that multicast is the answer for the future multimedia data-transfer and various other real-time and nonreal-time applications as well. When multicast trees are considered, it is inclined towards graph-theory problems. To maintain optimal tree-cost in multicasting is an NP-complete problem, known as Steiner Tree problem. With respect to the existing networks infrastructure, there are different interesting solutions and related algorithms. These problems become more complicated when we consider the dynamic situation of the group membership. Most of the existing algorithms discussed are unconstrained routing algorithms. To find an optimal solution considering more than one link metric is an NP-complete problem. We also briefly discussed the mobile and ad hoc multicast, which has recently become an important topic for research.

REFERENCES

Ballardie, A. Core Based Trees (CBTv2). (1997). *Multicast routing, RFC 2189*.

Carlberg, K., & Crowcroft, J. (1997). *Yet another multicast (YAM) routing protocol: Specification, version 1*. Unpublished manuscript.

Chakraborty, D., Chakraborty, G., Pornavalai, C., & Shiratori, N. (1999). Optimal routing for dynamic multipoint connection. *European Transaction of Telecommunication, 10*(2),183-190.

Chakraborty, D., Chakraborty, G., & Shiratori, N. (2003). A dynamic multicast routing satisfying multiple QoS constraints. *International Journal of Network Management,13,* 321-335.

Chen, S., Nahrstedt, K., & Shavitt, Y. (2000). A QoS-aware multicast routing protocol. *IEEE JSAC, 18*(12), 2580-2592.

Deering, S. et al. (1998). *Protocol Independent Multicast version 2 Dense Mode Specification.* Internet draft, draft-ietf-pim-v2-dm-*txt.

Estrin, D. et al. (1998). *Protocol Independent Multicast Sparse Mode (PIM-SM): Protocol specification.* RFC 2362.

Guerin, R. et al. (1998). *QoS routing mechanism and OSPF extensions.* Internet draft.

Holbrook, H., & Cheriton, D. (1999). IP multicast channels: Express support for large-scale single-source applications. *Proceedings of SIGCOMM'99*.

Kompella, V.P., Pasquale, J.C., & Polyzos, G.C. (1993). Multicast routing for multimedia communications. *IEEE/ACM Transactions on Networking, 1*(3), 286-292.

Kou, L., Markowsky, G., & Berman, L. (1981). A fast algorithm for Steiner trees. *Acta Informatica, 15,* 141-145.

Levine, B.N., & Garcia-Luna-Aceves, J. (1997). Improving Internet multicast with routing labels. *IEEE INCP'97,* 241-150.

Ooms, D., Sales, B., Livens, W., Acharya, A., Griffoul, F., & Ansari, F. (2002). *Overview of IP multicast in a multi-protocol label.* RFC 3353.

Salama, H.F., Reeves, D.S., & Viniotis, Y. (1997). Evaluation of multicast routing algorithms for real-time communication on high-speed networks. *IEEE JSAC, 15*(13), 332-345.

Thaler, D., Estrin, D., & Meyer, D. (1998). Border Gateway Multicast Protocol (BGMP). *Proceedings of ACM SIGCOMM '98*.

Wang, B., & Hou, J.C. (2000). Multicast routing and its QoS extension: Problems, algorithms, and protocols. *IEEE Network, 14*(2), 22-36.

Yan, S., Faloutsos, M., & Banerjea, A. (2000). QoS-aware multicast routing for the Internet: The design and implementation of QoSMIC. *IEEE/ACM Transactions on Networking, 10*(1), 54-66.

KEY TERMS

Greedy Algorithm: An algorithm that always takes the best immediate, or local, solution while finding an answer. Greedy algorithms find the overall, or globally, optimal solution for some *optimization problems* but may find less-than-optimal solutions for some instances of other problems.

IGMP: The Internet Group Management Protocol (IGMP) is implemented within the IP module of a host and extends the host's IP implementation to support multicasting.

ODMRP: On-Demand Multicast Routing Protocol for mobile ad hoc networks. ODMRM is a mesh-based, rather than conventional tree-based, multicast scheme and uses a forwarding group concept.

PGM: An extreme example of maintaining state information is the Pragmatic General Multicast (PGM) protocol, formerly known as Pretty Good Multicast, which is implemented by Cisco, where routers maintain state information on a per-packet basis.

RMTP II: Reliable Multicast Transport Protocol II is a transport protocol for IP multicast, designed to reliably and efficiently send data from a few senders to large groups of simultaneous recipients.

RP: PIM-SM constructs a multicast distribution tree around a router called a rendezvous point (RP). This rendezvous point plays the same role as the core in the CBT protocol; receivers "meet" new sources at this rendezvous point.

RPF: In multicast, the router forwards the packet *away* from the source to make progress along the distribution tree and prevent routing loops. The router's multicast forwarding state runs more logically by organizing tables based on the reverse path, from the receiver back to the root of the distribution tree. This process is known as *reverse-path forwarding (RPF)*.

Multilingual Electronic Commerce in a Global Economy

Aryya Gangopadhyay
University of Maryland Baltimore County, USA

Zhensen Huang
University of Maryland Baltimore County, USA

INTRODUCTION

One of the major challenges facing organizations involved in electronic commerce (e-commerce) today is to organize and summarize information in such a way that end users can effectively and efficiently search for and analyze relevant information. Users can look for both structured as well as unstructured information in a system designed for electronic commerce. An example of structured information is the price of a specified product. Unstructured information, on the other hand, is information that is not well specified or that has multiple specifications. For example, the user may be looking for spices for cooking a shrimp dish where they can choose from a number of options. The user may have individual preferences1 for the selection of spices and may not know exactly how the information can be found in the system.

The problem of finding relevant information is exacerbated in global information management, especially in global electronic commerce. While globalization is presenting new opportunities for people and businesses worldwide, several challenges must be addressed in order to realize its full potential. Examples of these challenges include differences in culture and language, which can be an obstacle to unrestricted and free access of information, as well as the disorganization of the potentially precious knowledge asset. While language technology (Nirenburg, 1992; Onyshkevych & Nirenburg, 1995; Sheremetyeva & Nirenburg, 1996) is making rapid progress, much research is needed in managing and accessing multilingual information in order to reach the full potential of global electronic commerce (e.g., Malhotra, 1997, 1998). In Gangopadhyay and Huang (2000), the issue of knowledge requirements for building information systems that operate in multiple languages has been studied. Specifically, the focus was on studying user behavior in performing various tasks in a multilingual system. In order to study user behavior and performance in a multilingual electronic-commerce setting, a bilingual electronic catalog was designed and tested by online retailers selling products and/or services to customers interacting either in English or Chinese.

BACKGROUND

An electronic catalog is a graphical user interface that presents product and/or service information to users, typically using the World Wide Web. An electronic catalog is a key component of electronic commerce that has been used for business-to-consumer commerce as well as business-to-business commerce (Adam, Dogramaci, Gangopadhyay, & Yesha, 1998; Gangopadhyay, 2002). Although the term electronic catalog might sound like an electronic extension of paper catalogs, it offers features that are far beyond those found in paper catalogs. Such features include computational services such as efficient browsing and searching, online order processing such as checking out products using shopping carts and secure payment mechanisms, and back-end processing such as integration with company databases (Segev, Wan, & Beam, 1995). These features have extended the role of electronic catalogs to the point of being used as electronic storefronts.

With the rapid proliferation of electronic commerce both in local and global markets, there is an increasing need to provide support for internationalization such as foreign currencies, different date and time formats, sort order, and multiple languages (Broin, 1999). The need for providing multilingual support is echoed by the rapid increase of non-English-speaking users on the Internet. For example, it is reported that 75% of Internet users may be non-English speaking by 2005 (Thomason, n.d.). A side effect of the increasing number of non-English users of the Internet is the "contamination" of other languages with English terms (Voiskounsky, 1998). Intercultural class differences can also lead to differences in perceptions of abstractions and generalizations. These differences should be taken into account in designing the graphical user interfaces of e-commerce sites (e.g., Stander, 1998). An example of a multilingual electronic catalog is shown in Figure 1a and 1b, taken from Gangopadhyay & Huang (2000).

Empirical investigations in this field (e.g., Gangopadhyay & Huang, 2000) show that users prefer to use their ethnic language when conducting tasks that are

Figure 1a. Three modes of search in English

Figure 1b. Three modes of search in Chinese

unstructured or semistructured. However, language preferences diminish when the tasks become more structured. Also, nonethnic terms, such as computer, are difficult to express in an ethnic language where others, such as kung pao, are difficult to express in a nonnative language such as English. Hence, e-commerce Web sites designed for structured tasks dealing with nonethnic products should use international languages such as English. On the other hand, if the tasks performed by the users are likely to be unstructured and the products or services are ethnic in nature, ethnic languages should be used for designing the Web site.

FUTURE TRENDS

From the research results described above, we can surmise several implications for future research and practice in global electronic commerce. First, it is quite clear that users prefer to use their ethnic language when searching for ethnic products because it is difficult to translate them into another language, such as English. Some researchers (e.g., Doherty, 1999) assert that the *unified content model* is easier to implement in a global electronic commerce system. However, the unified content model may not be good from a user-interface standpoint for all product categories since users may prefer to use their own ethnic language.

Another closely related issue is the level of understanding of the information presented for bilingual users. The previous studies indicate that the language preference is not a major issue when dealing with structured information. However, when dealing with unstructured information, there is a significant preference toward using one's native language. While it takes more research to establish a relationship between language preference and information complexity, such research can render significant insights into the design of multilingual interfaces for global electronic commerce.

In multilingual systems that depend on word-for-word translation from one language to another, a lot of information may be lost during the translation process. For example, it is hard to tell the differences between *table* and *desk* in Chinese. In these cases, images may be helpful to present product information. It is also worthwhile to study the effect of multimedia information on user performance and satisfaction in global electronic commerce.

CONCLUSION

With the rapid proliferation of non-English speakers on the Internet, it is becoming critical to be able to provide interactions in multiple languages. The critical issues discussed in this paper address the importance of multilingual interfaces, as well as language preferences in conducting various types of tasks. As the global economy continues to grow, technological solutions will be sought after to solve existing as well as newer issues that will arise in the future.

REFERENCES

Adam, N. R., Dogramaci, O., Gangopadhyay, A., & Yesha, Y. (1998). *Electronic commerce: Business, technical, and legal Issues.* Prentice-Hall.

Broin, U. Ó. (1999). International aspects of user interface design. MultiLingual Computing & Technology, 10*(3), 50-54.*

Computer Economics. (n.d.) English will dominate Web for only three more years. *Retrieved from http://www.computereconomics.com/new4/pr/pr990610.html*

Doherty, W. (1999). Creating multilingual Web sites. MultiLingual Computing & Technology, 10*(3), 34-37.*

Gangopadhyay, A. (Ed.). (2002). *Managing business with electronic commerce: Issues and trends.* Hershey, PA: Idea Group Publishing.

Gangopadhyay, A., & Huang, Z. (2000, July-September). Online user interaction with electronic catalogs: Language preferences among global users. *Journal of Global Information Management* [Special issue on knowledge management], *8*(3), 16-23.

Malhotra, Y. (1997). Knowledge management in inquiring organizations. *Proceedings of Third Americas Conference on Information Systems*, 293-295.

Malhotra, Y. (1998, July/August). TOOLS@WORK: Deciphering the knowledge management hype. *Journal for Quality & Participation* [Special issue on learning and information management], *21*(4), 58-60.

Nirenburg, S. (1992). Text planning with opportunistic control. Machine Translation *[Special issue on natural language generation]*, ***7*(1-2**), 99-124.

Onyshkevych, B., & Nirenburg, S. (1995). A lexicon for knowledge-based MT. Machine Translation***, 10***(1-2).

Segev, A., Wan, D., & Beam, C. (1995, October). *Designing electronic catalogs for business value: Results of the CommerceNet pilot* (Tech. Rep.). Berkeley: University of California, Haas School of Business, The Fisher Center for Management and Information Technology.

Sheremetyeva, S. O., & Nirenburg, S. (1996). Empirical modeling in computational morphology. Nauchno-Technicheskaja Informacija (Scientific and Technological Information)***, 2*(7).**

Stander, A. (1998). *Bridging the gap: Issues in the design of computer user interfaces for multicultural communities.* Proceedings of Cultural Attitudes Towards Communication and Technology, Sydney, Australia.

Thomason, L. (n.d.). Design tip: Consider your international audience. *NetMechanic.* Retrieved July 2004 from http://www.netmechanic.com/news/vol6/design_no15.htm

Voiskounsky, A. (1998). *Internet: Cultural diversity and unification.* Proceedings of Cultural Attitudes Towards Communication and Technology, Sydney, Australia.

KEY TERMS

Business-to-Business Electronic Commerce: Any business transaction conducted between two business entities. An example is where a manufacturer buys raw material from a supplier over the Internet.

Business-to-Consumer Electronic Commerce: Any business transaction conducted through the Internet between a business and a consumer. An example includes a commercial Web site that sells products to individual customers.

Electronic Catalog: An electronic catalog is a graphical user interface that presents product and/or service information to users, typically using the World Wide Web.

Electronic Commerce: Any business transaction that is performed over the Internet.

Multilingual Interfaces: Interfaces that are written in multiple languages. These interfaces may be identical to each other or vary based on differential needs.

Semistructured Information: The type of information that is a combination of structured and unstructured information. An example of semistructured information is an HTML (hypertext markup language) page. It is structured because it uses tags to describe the content of a Web page, but it is unstructured because there is free text between two tags.

Structured Information: Information that is specific. An example is the price of a product.

Unified Content Model: As used in multilingual Web-based systems, the unified content model refers to word-for-word translation of the content from one language to another.

Unstructured Information: Information that is not well specified. An example is a free-text description of a certain product that cannot be exhaustively characterized by specific features.

ENDNOTE

[1] By individual preference, we mean that the choice of the spices may vary from one user to another.

Multimedia Computing Environment for Telemedical Applications

M

V.K.Murthy
University of New South Wales, Australia

E.V.Krishnamurthy
Australian National University, Australia

INTRODUCTION

Telemedicine (in short, e-medicine) is a means of delivering medical services to any place, no matter how remote, thereby removing the limitations of space and time that exists in today's health-care settings. Computers are indispensable in telemedicine, since they provide for efficient, relevant data gathering for large-scale applications. Besides providing immediate feedback of results to patients and doctors, they also can compare past patient records and evaluate relative improvement or deterioration. Further, they are readily available at any time, fatigue-free and can be more objective.

Five important application areas of telemedicine are:

1. Lifetime health care;
2. Personalized health information;
3. Tele-consultation;
4. Continuing Medical education; and
5. Context-aware Health monitoring.

For example, computers provide for multimedia imaging: ultra sound, digital X-rays, 3D spiral Cat Scanner, magnetic resonance imaging, PET scanning, and so forth, and can fuse them into a single multi-purpose image using fusion software. Adding mobility to computers enhances their role in telemedical applications considerably, especially at times of emergency since the patients, doctors, the data collecting and retrieval machines, as well as their communication links can always be on the move. Very simple, inexpensive mobile communication and computing devices can be of great help in telemedicine, as explained in the following:

- ***Low Cost Radio:*** Even the simplest of mobile devices, such as a low power radio that can transmit messages to a home computer from which medical data can be sent through telephone line and the Internet can be of great value in saving lives (Wilson et al., 2000).
- ***Personal Digital Assistants (PDA):*** The simplest of the computers, such as palmtops and PDA can assist the doctors for instant nomadic information sharing, and look for diagnosis of different diseases and treatment. PDA can help the doctors to figure out drug interactions, storing summaries of sick patients and their drug list. Further, PDA can provide for downloading suitable programs from the Web, and can be programmed for alert, sending and receiving email, jotting down pertinent points, and for storing immediately needed clinical results to carry out ward rounds. Also a hand held system can provide context-awareness to support intensive and distributed information management within a hospital setting (Munoz et al., 2003).
- ***Internet:*** The Internet is an important tool for medical professionals and will completely change the manner in which medical consultations are provided (Coiera, 1997); for further details on telehealth and telemedicine practice and their real life implementation issues, refer to Orlov and Grigoriev (2003), Jennett and Anddruchuk (2001), and Suleiman (2001).

For minor ailments, Internet-based consultations to doctors can provide prescriptions for medical/pathological examinations by laboratories. The results are then posted in the Internet for subsequent reading of the results by the concerned doctors who can prescribe medicines that can be posted on the Internet. This prescription can then be handled by a pharmacy to dispense the medicines to the concerned individual. Kim and Hwang (2001) have proposed a password controlled Internet-based medical system that brings in a variety of services to doctors, patients, pharmacists and health-care professionals. It allows people to receive medical examinations and medical advice.

BACKGROUND: TELEMEDICAL INFORMATION SERVICES

The first step in telemedicine is the telemedical diagnosis (or telediagnosis) based on information obtainable from medical images, blood, urine and other pathological test reports. Usually, for diagnostic purposes, the doctor sends a patient for such examinations. The laboratory assistant takes the required X-ray or ultrasound images or carries out pathological tests and passes these images (or readings) on to a radiologist/pathologist who then makes analysis and sends a report to a doctor. These manual actions are totally sequential and slow. This whole procedure can be made cooperative and faster, if the images and data are stored in a database and these can be simultaneously retrieved by specialists in their offices or homes to make a cooperative diagnosis (Alfano, 1997; Coiera, 1997; Ganapathy, 2001; Gomez et al., 1997; Jameson et al., 1996; Kleinholz et al., 1994; Lauterbach et al., 1997).

Principal Aims

The principal aims of e-medical informatics are to:

(i) provide online services of patient records (medical and pathological databases) to medical practitioners and radiologists;
(ii) provide primary specialist diagnosis, offer second opinion, provide pre- and post treatment advice through email;
(iii) reduce the cost of imaging equipment, delay, and increase the speed and volume of diagnosis;
(iv) aid cooperative diagnosis and provide assistance for remote surgery;
(v) provide student /resident education;
(vi) reduce professional isolation, increase collaboration; and
(vii) provide home-care.

Advantages

E-medicine offers the following advantages:

(i) Provides health care to under-served and isolated areas so that we can make a better allocation and utilisation of health resources.
(ii) Since communication cost is much cheaper than the transportation cost, patients in remote areas can outreach physicians quickly.
(iii) Increases the speed of diagnosis and treatment especially when used for teleradiology, cardiology, psychiatry.
(iv) Allows access to specialty care using time-oriented clinical data.
(v) Real-time monitoring of public health databases to prepare and respond during epidemics, biological and chemical terrorism.
(vi) Internet can provide the following support:
a. Health information;
b. Administrative infrastructure;
c. Online health records;
d. Pharmaceutical information and sales outlets; and
e. Online training for telemedical professionals.

Prerequisites

The pre-requisites for a successful implementation of a telemedical system are:

- *Infrastructure:* A suitable infrastructure of health care providers, doctors, engineers, computing specialists, communication engineers, information technology professionals and medical statisticians to analyse outcomes and suitable outreach clinics with telemedical facilities.
- *Communication Network:* Reliable, inexpensive and readily accessible communication network from outreach clinics to hospitals, doctors and patients and pathological laboratories.
- *Low-cost Computers:* Suitable low-cost hardware-software and a good communication bandwidth to transmit medical data in different modes; for example, radiological images, video images of signals and text. While using wired in or wireless mobile devices and monitors, the effect of electromagnetic interference (EMI) and radio frequency interference (RFI) on data collection and transmission, and the side-effects on patients, both physiological and psychological aspects, have to be taken care of so that improper diagnosis does not result.
- *Training Facility:* Training of personnel to provide proper maintenance of equipment and safety standards to patients.
- *Security, Reliability, Efficiency:* Reliability, Efficiency, Security, Privacy and Confidentiality in handling, storing and communicating patient information.

Economic Necessity

In densely-populated countries (e.g., India), the rate of growth in hospital beds to cope up with the increasing population is economically unsustainable and technically not viable, since the number of medical specialists also cannot grow to meet this demand (Ganapathy, 2001). The use of telemedicine avoids unnecessary strain involved in travel and associated expenses, provide immediate attention and care, can avoid hospitalization, and allow the patients to stay home enjoying family support. Chan et al. (2000) describe a real-time tertiary foetal ultrasound telemedical consultation system using standard integrated system digital network (ISDN) that operates in Queensland, Australia. This consultation system has gained acceptance from the clinicians and patients.

Aging population and rising health costs have created the need to care for more patients in their own homes. Hospitals without walls (e-hospitals or virtual hospitals) provide for continuous monitoring of patients in certain diagnostic categories. Wilson et al. (2000) describe how to build such "hospital without walls". This system uses a miniature, wearable low power radio to transmit vital and activity information to a home computer, from which data is sent by telephone line and the Internet to the concerned doctors. Thus, telemedicine and telediagnosis are economic necessities for both the developing and the developed world.

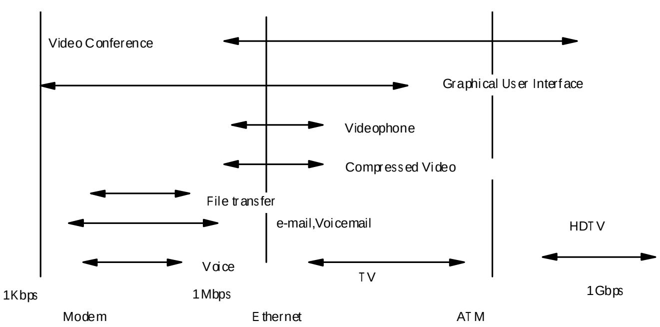
MAIN THRUST OF THE ARTICLE

The main thrust of this chapter is to describe the multimedia telediagnostic environment (MMTE), its features, and some practical MMTE systems that are currently operational.

MULTIMEDIA TELEDIAGNOSTIC ENVIRONMENT

A Mobile Multimedia Telediagnostic Environment (MMTE) permits users to work remotely on common shared resources and applications and simultaneously communicate both visually and through audio. Such an environment becomes more flexible and useful if it can work in a wireless and wired-in (integrated) environment so that the services can be provided for mobile hosts. Here, the users have interfaces of PDA or laptops interacting remotely. In this application, we need cooperation among the participants through special communication tools for conversation (Roth, 2001) and E-pointing (MacKenzie & Jusoh, 2001). The conversation can be of conference type where two or more participants are involved. The e-pointers aid each participant to point out a particular segment of an image or a video image of a signal so that each of the other participants can visualize the movements of the remote pointer. Also in telesurgery (e-surgery) where two or more operating rooms are connected by a network, live video signals may required to be transferred to other remote locations from endo-camera (camera for photographing internal organs) and operating room camera to other remote locations for consultancy. This would allow surgeons not only to see but also visualize surgical instrument movements with 3-D models in real time during the surgery.

Figure 1. Bandwidth (log scale) requirements



Pre-requisites for the Deployment of MMTE

Reliable Communication Network and Equipment

A reliable multimedia communication network that links the remote centres with hospitals is essential. The delivery of multimedia content in a timely fashion is very important. Internet provides a uniform communication medium to link users together to access or deliver multimedia information.

- Communication Bandwidth Requirements: Text requires the lowest bandwidth, while audio and video data and signals require significant increase in bandwidth. Specification of bandwidth to be used and compression techniques used are to be laid down so that the data and images that are transmitted are of diagnostic quality.
- Bandwidth Management: Bandwidth determines the information capacity of a network per unit of time. Wireless networks deliver lower bandwidth than wired network. Hence software techniques based on compression should be used. Also scheduling communication intelligently can save bandwidth. For use in telemedicine the techniques should be extremely reliable. Current cutting edge technologies are yet to develop. Bandwidth requirements along with applications are given in (approximate) logarithmic scale in Figure 1.

Three common technologies used are: (bps = bits per second; K= Kilo; M= Mega; G = Giga)

Dial up mode: Rate 28.8Kbps
T1: 1.544 Mbps; DS3: 45 Mbps

For example, standard X-ray transmission takes 30 minutes in dial-up mode, 30 seconds in T1 mode and 1 second in DS3. It is obvious that the cost goes up, as we want to increase the bandwidth to communicate voice, data and pictures. DS3 is a dedicated, private line service designed for point to point communication. This service uses fibre optic cable. One can have the option of tiered DS3 service from lower to higher bandwidth from 5 Mbps to 45 Mbps depending upon cost considerations and transmission requirements.

- Choice of Multimedia Hardware: We need a proper choice of graphics and audio equipment for quality images and audio for diagnosis.

FEATURES OF A TELEDIAGNOSTIC ENVIRONMENT

A typical cooperative telediagnostic (e-diagnostic) environment is shown in Figure 2. It consists of databases containing medical/pathological image and textual information, cooperative interaction facility, and telecommunication. Typically, its characteristics are:

Remoteness of Users

Users work remotely on common shared resources and applications and simultaneously communicate both visually and through audio. This environment becomes more flexible and useful, if it can work in an integrated (wireless and wired-in) environment so that the services can be provided for mobile hosts.

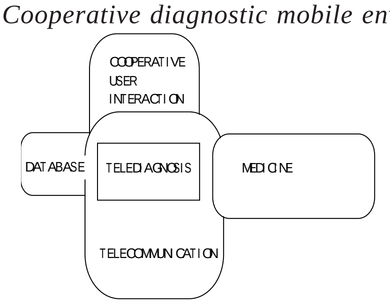
Cooperation

Many e-medicine applications require doctors, specialists and health-record officials to cooperate and discuss for particular medical treatment. Cooperation among the participants requires special communication tools for conversation, e-pointing, e-surgery.

E-pointing Facility

E-pointing is the act of moving an on-screen tracking symbol such as a cursor, by manipulating the input device. The interactive e-pointers (MacKenzie & Jusoh, 2001) aid each participant to point out a particular segment of an image or a video image of a signal so that each of the other participants can visualize the movements of the remote pointer and discuss any abnormality for diagnostic purposes.

Figure 2. Cooperative diagnostic mobile environment



Facility for Internet Link

Also users can remotely access worldwide knowledge bases/databases, download files from the Internet and browse the World Wide Web using portable and mobile devices such as laptops, palmtops and PDA.

Interoperability

Software interoperability is a key requirement in telemedicine, since different modalities of patient records are used on different platforms.

- **Security Requirements**: Security of patient data during storage and transmission are vital to safeguard confidentiality and privacy of patient information. Biometric authentication (Nanavati et al., 2002) could be an answer in dealing with patient information. Health Insurance Portability and Accountability Act (HIPAA) has devised certain security measures to handle telemedical information, see Web document of HIPAA.

FUTURE TRENDS: SOME PRACTICAL MMTE SYSTEMS

Context-aware Mobile Communicaton in Hospitals

Handheld systems can provide for context-awareness in hospitals. This can provide speedy collaboration among doctors to deal with time sensitive cases, for example, delivery, heart attacks, accidents (Munoz et al., 2003).

Presently, portable and mobile devices such as laptops and personal digital assistants (PDA) are useful for remotely accessing resources in an efficient and reliable manner and for reliable telediagnosis (Lauterbach et al., 1997; Roth, 2001). For instance, information sharing among doctors through handheld appliances are useful for transmitting small amounts of data, such as heart rate, blood pressure and other simple monitoring devices. Typical datatypes for current handheld appliances are: text, data entries, numbers, tables. None of the applications described previously can deal with multimedia data, such as audio, video which require considerable bandwidth, sufficient output devices, and a very powerful battery. Presently, multimedia data are not yet suitable for handheld appliances; they need to undergo major developments (Roth, 2001). Currently available tools, such as GyroPoint and RemotePoint (MacKenzie & Jusoh, 2001), lack accuracy and speed. The telepointer technology has yet to mature to provide reliable service for e-medicine and e-surgery. Along with the use of virtual reality, the pointer technology can be useful in teaching surgery, and for trial-error planning of operations.

Image-Guided Telemedical System

Lee (2001) describes a Java-applet based image-guided telemedicine system via the Internet for visual diagnosis of breast cancer. This system automatically indexes objects based on shape and groups them into a set of clusters. Any doctor with Internet access can use this system to analyze an image and can query about its features and obtain a satisfactory performance (http://dollar.biz.uiowa.edu/~kelly/telemedicine.html).

Software Agent based Telemedical Systems

Software agents, which are personalized, continuously running and semi-autonomous objects, can play an important role in cooperative telemedicine. Such agents can be programmed for supporting medical diagnostic intelligence and keep a watchful eye to discover patterns and react to pattern changes that usually occur in epidemics, biological and chemical terrorism.

For example, diabetes is a chronic disease with a sustained elevated glucose level. The diabetics need to be injected insulin to control the glucose level. This is quite a tedious book-keeping procedure. Real-time monitoring using agents and immediate therapy can save a life.

Also agents can help in the appropriate choice of doctors, allocation of hospitals and beds for patients. Ganguly and Ray (2000) describe a methodology for the development of software agent-based interoperable telemedicine system. This system has been illustrated using tele-electrocardiography application. Also recently, a practical agent-based approach to telemedicine has been suggested by Tian and Tianfield (2003). For medical education, this will be a very useful tool. The Foundation for Intelligent Physical Agent Architecture (FIPA, http://www.fipa.org) aims to improve agent interoperability by providing standards for protocols and languages. Also the Java 2 MicroEdition (J2ME) is targeted at PDA. These developments will provide an agent execution environment in PDA.

CONCLUSION

Cooperative telemedical informatics will have a direct impact on the rapidly changing scene of information and communication technology (including advanced communication systems) and will provide for greater interaction among doctors and radiologists and health-care professionals to integrate information quickly, efficiently and make effective decisions at all levels. Virtual e-hospital or hospital without walls will be a reality soon providing great benefit to the society.

REFERENCES

Alfano, M. (1997). User requirements and resource control for cooperative multimedia applications, *Multimedia Applications, Services and Techniques, Lecture Notes in Computer Science*, 1242, 537-553. New York: Springer Verlag.

Chan, F.Y. et al. (2000). Clinical value of real-time tertiary foetal ultrasound consultation by telemedicine. *Telemedicine Journal*, *6*, 237-242.

Coiera, E. (1997). *Guide to medical informatics, the Internet and telemedicine*. London: Chapman & Hall.

Ganapathy, K. (2001).Telemedicine in action •The Apollo experience. *Proceedings of the International Conference on Medical Informatics*, Hyderabad, India, November 2001.

Ganguly, P. & Ray, P. (2000). A methodology for the development of software agent based interoperable systems: A tele-electrocardiography perspective. *Telemedicine Journal, 6*, 283-294.

Gomez, E.J. et al. (1997). The Bonaparte telemedicine. In *Multimedia Applications, Services and Techniques, Lecture Notes in Computer Science*, 1242, 693-708. New York: Springer Verlag.

Jameson, D.G. et al. (1996). Broad band telemedicine teaching on the information superhighway. *Journal of Telemedicine and Telecare*, *1*, 111-116

Jennett, A., & Anddruchuk, K. (2001). Telehealth: Real life implementation issues. *Computer Methods and Programs in Biomedicine*, *64*,169-174.

Kim, S.S. & Hwang, D.-J. (2001). An algorithm for formation and confirmation of password for paid members on the Internet-based telemedicine. *Lecture Notes in Computer Science*, 2105, 333-340. New York: Springer Verlag.

Kleinholz, L. et al. (1994). Supporting cooperative medicine. *IEEE Multimedia*,Winter, 44-53.

Lauterbach, Th. et al. (1997). Using DAB and GSM to provide interactive multimedia services to portable and mobile terminals. In *Multimedia Applications, Services and Techniques, Lecture Notes in Computer Science*,1242, 593-609. New York: Springer Verlag.

Lee, K.-M. (2001). Image-guided telemedicine system via the Internet. *Lecture Notes in Computer Science*, 2105, 323-333. New York: Springer Verlag.

MacKenzie, S. & Jusoh, S. (2001). An evaluation of two input devices for remote sensing. In *Engineering for Human Computer Interaction, Lecture Notes in Computer Science*, 2254, 235-250. New York: Springer Verlag.

Munoz, M.A. et al. (2003). Context-aware mobile communication in hospitals. *IEEE Computer*, *36*(9), 38-47.

Nanavati, S. et al. (2002). *Biometrics*. New York: John Wiley.

Orlov, O., & Grigoriev, A.(2003). Space technologies in routine telemedicine practice: Commercial approach. *Acta Astronautia*, *51*(July), 295-300.

Roth, J. (2001). Information sharing with hand-held appliance. In *Engineering for Human Computer Interaction, Lecture Notes in Computer Science*, 2254, 263-279. New York: Springer Verlag.

Suleiman, A.B. (2001). The untapped potential of telehealth. *International Journal of Medical Informatics*, 61, 103-112.

Tian, J. & Tianfield, H. (2003). A multi-agent approach to the design of an e-medicine system. *Lecture Notes on Artificial Intelligence*, 2831, 85-94. New York: Springer Verlag.

Wilson, L.S. et al. (2000). Building hospitals without walls: A CSIRO home telecare initiative. *Telemedicine Journal*, *6*, 275-281.

WEB-DOCUMENTS

http://telemed.medicine.uiowa.edu: Zollo Susan: Introduction to Telemedicine

http://www.nlm.nih.gov/reserach/telesymp.html: Provides information on Telemedicine Symposium.

http:// telemed.medicine.uiowa.edu: Provides slide shows on various aspects of telemedicine prepared by the National Laboratory for the study of Rural telemedicine, of the University of Iowa.

http://telemed.medicine.uiowa.edu/M.G.Kienzle, Telemedicine for home monitoring

http://www.ntia.doc.gov/reports/telemed/, Telemedicine report to Congress.

http://telemed.medicine.uiowa.edu/ (UI TRC).

http://tie.telemed.org/(TelemedicineInformation Exchange).

http://www.nlm.nih.gov/research/telemedinit.html (NLM National Telemedicine Initiative).

http://www.tmgateway.org/ (Federal Telemedicine Gateway

http: //www. hipaadivsory.comregs/securityoverview.htm

KEY TERMS

Bandwidth Management: Determines the information capacity of a network per unit of time. Wireless networks deliver lower bandwidth than wired network. The choice of appropriate bandwidth for efficient and cost effective transmission of voice, data and pictures is called bandwidth management.

Confidentiality and Security in Telemedicine: Security and confidentiality of patient data during storage and transmission are vital to safeguard confidentiality and privacy of patient information. Biometric authentication, could be an answer in dealing with patient information. HIPAA (Health Insurance Portability and Accountability Act) has devised certain security measures to handle telemedical information (http:// www.hipaadivsory.comregs/securityoverview.htm)

Multimedia Telediagnostic Environment (MMTE): Permits users to work remotely on common shared resources and applications and simultaneously communicate both visually and through audio.

Software Agents: Personalized, continuously running and semi-autonomous objects. They can play important role in cooperative telemedicine. Such agents can be programmed for supporting medical diagnostic intelligence and keep a watchful eye to discover patterns and react to pattern changes that usually occur in epidemics, biological and chemical terrorism

Software Interoperability: Permits different software to run on different platforms with different operating systems.

Telemedicine (E-Medicine): A means of delivering medical services to any place, no matter how remote, thereby removing the limitations of space and time that exists in today's health-care settings.

Telepointers: The act of moving an on-screen tracking symbol, such as a cursor, by manipulating the input device from a remote site.

Virtual E-Hospital (or Hospital Without Walls): A telemedical facility that provides continuous monitoring of patients staying in their homes and enjoying family support and creates a psychological satisfaction to the patients that they are receiving immediate attention and care as if they are in a real hospital.

Multimedia Content Adaptation

David Knight
Brunel University, UK

Marios C Angelides
Brunel University, UK

INTRODUCTION

The previous decade has witnessed a wealth of advancements and trends in the field of communications and subsequently, multimedia access. Four main developments from the last few years have opened up the prospect for ubiquitous multimedia consumption: wireless communications and mobility, standardised multimedia content, interactive versus passive consumption and the Internet and the World Wide Web. While individual and isolated developments have produced modest boosts to this existing state of affairs, their combination and cross-fertilisation have resulted in today's complex but exciting landscape. In particular, we are beginning to see delivery of all types of data for all types of users in all types of conditions (Pereira & Burnett, 2003).

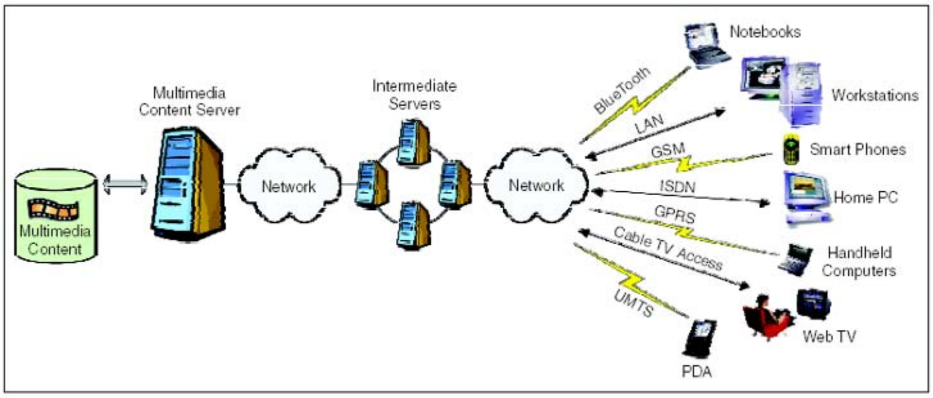
Compression, transport, and multimedia description are examples of individual technologies that are improving all the time. However, the lack of interoperable solutions across these spaces is holding back the deployment of advanced multimedia packaging and distribution applications. To enable transparent access to multimedia content, it is essential to have available not only the description of the content but also a description of its format and of the usage environment in order that content adaptation may be performed to provide the end-user with the best content experience for the content requested with the conditions available (Vetro, 2003).

In the following sections, we will look at the background of multimedia content adaptation, why do we require it and why are present solutions not adequate. We then go onto the main focus of the article, which describes the main themes of modern multimedia content adaptation, such as present day work that defines the area and overviews and descriptions of techniques used. We then look at what this research will lead to in the future and what we can expect in years to come. Finally, we conclude this article by reviewing what has been discussed.

BACKGROUND

More and more digital audio-visual content is now available online. Also more access networks are available for the same network different devices (with different resources) that are being introduced in the marketplace. Structured multimedia content (even if that structure is still limited) increasingly needs to be accessed from a diverse set of networks and terminals. The latter range (with increasing diversity) from gigabit Ethernet-connected workstations and Internet-enabled TV sets to mobile video-enabled terminals (Figure 1) (Pereira & Burnett, 2003).

Figure 1. Different terminals access multimedia content through different networks



Adaptation is becoming an increasingly important tool for resource and media management in distributed multimedia systems. Best-effort scheduling and worst-case reservation of resources are two extreme cases, neither of them well-suited to cope with large-scale, dynamic multimedia systems. The middle course can be met by a system that dynamically adapts its data, resource requirements, and processing components to achieve user satisfaction. Nevertheless, there is no agreement about questions concerning where, when, what and who should adapt (Bormans et al., 2003).

On deploying an adaptation technique, a lot of considerations have to be done with respect to how to realise the mechanism. Principally, it is always useful to make the technique as simple as possible, i.e., not to change too many layers in the application hierarchy. Changes of the system layer or the network layer are usually always quite problematic because deployment is rather difficult. Generally, one cannot say that adaptation technique *X* is the best and *Y* is the worst, as it highly depends on the application area.

The variety of delivery mechanisms to those terminals is also growing and currently these include satellite, radio broadcasting, cable, mobile, and copper using xDSL. At the end of the distribution path are the users, with different devices, preferences, locations, environments, needs, and possibly disabilities.

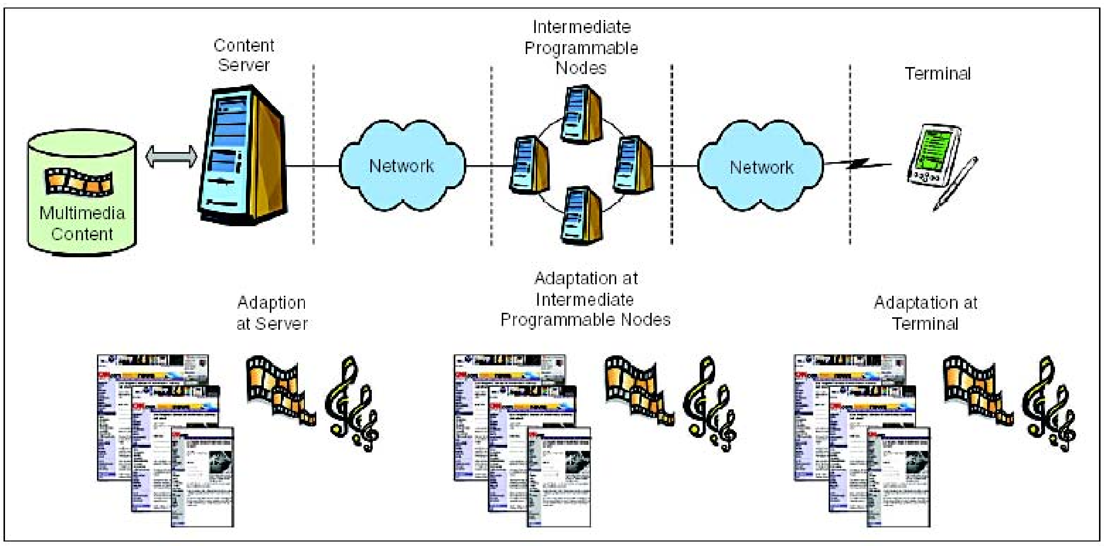
In addition the processing of the content to provide the best user experience may be performed at one location or distributed over various locations. The candidate locations are: the content server(s), any processing server(s) in the network, and the consumption terminal(s). The choice of the processing location(s) may be determined by several factors: transmission bandwidth, storage and computational capacity, acceptable latency, acceptable costs, and privacy and rights issues (see Figure 2).

Present adaptation technologies concerning content adaptation mainly focus on the adaptation of text documents. Therefore, one text document will be adapted on demand to the capabilities of different devices or applications. To fulfill this functionality the structure of the content must be separated from its presentation, i.e., the source document is structured using XML (Extensible Markup Language) and then dynamically processed to generate a presentation tailored to the available resources. One possible use case scenario will be to present the same information either on a standard Web browser or a WAP browser.

Efficient adaptation requires that the participating components know from each other and take advantage of adaptation steps done by other components, which needs standardised media, metadata, and communication. Several standardisation bodies (W3C, MPEG, and WAP) have already been established or are currently under development, which have recognised the need to create a framework that facilitates the efficient adaptation of content to the constraints and preferences of the receiving end.

MPEG-7 (ISO/IEC 15938-5:2002) provides tools for content description, whilst capability description and negotiation is provided for with CC/PP (Composite Capabilities/Preference Profiles, 2003) and UAProf (WAG User Agent Profile, 2001). MPEG-21 (ISO/IEC JTC 1/SC 29/WG 11), the "multimedia framework" includes Digital Item Adaptation (DIA), which enables standard communication of dynamic adaptation of both media resources and

Figure 2. Adaptation may be performed at different places



meta-data, enabling negotiation of device characteristics and QoS parameters. (Böszörményi et al., 2002)

In this section, the reasons for the need for interoperable and efficient multimedia content adaptation have been introduced. A number of standards groups (such as W3C, MPEG and WAP) that facilitate multimedia content adaptation by concentrating on the adaptation of associated XML-type documents have also been mentioned. The next section delves into the different technologies that help make up this exciting field.

MAIN THRUST OF THE CHAPTER

In this section we will look at the main themes found in multimedia content adaptation. We start with a look at a multimedia content adaptation architecture, a discussion on the present state of affairs regarding scalable coding and transcoding, an analysis of the effect the actual location point the adaptation takes place and a brief summary of the relevance of user profiling.

Multimedia Content Adaptation Architecture

The networking access paradigm known as Universal Multimedia Access (UMA) refers to the way in which multimedia data can be accessed by a large number of users/clients to view any desired video stream anytime and from anywhere. In the UMA framework, multimedia information is accessed from the network depending on the following three parameters: channel characteristics, device capabilities, and user preference.

Figure 3 gives an example of different presentations (to suit different capabilities such as formats, devices, networks, and user interests) of the same information.

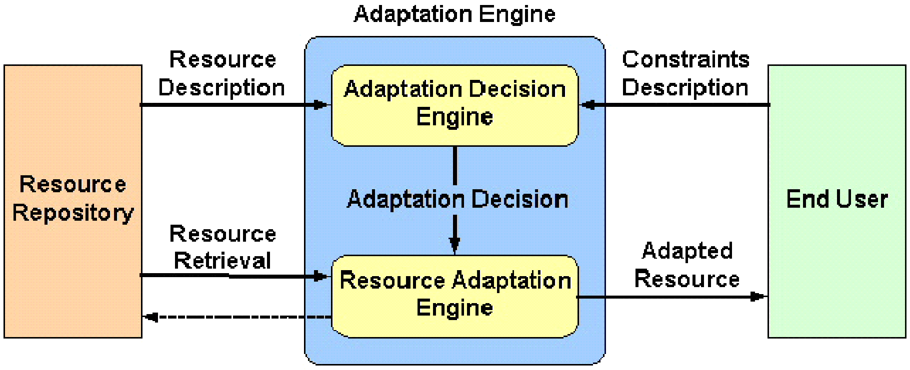
One option for UMA is to provide different variations of the content with different quality, bit rate, media modality (e.g., audio to text), etc. The problem with this option is that it is not too efficient from the viewpoint of variation generations and storage space. On the other hand, real-time transformation of any content implies some delay for the processing and a lot of computing resources at the server (or proxy server) side. Pursuing either of these two options assumes the use of an adaptation engine. Figure 4 gives a bird's eye-view of such an adaptation engine architecture that is applicable to the adaptation of any type of content.

The architecture consists of an adaptation engine that can be located on the server, an intermediate network device such as a gateway, router, or proxy, or even on the client. This engine comprises of two logical engine modules, the adaptation decision engine and the resource adaptation engine. The adaptation decision engine receives the metadata information about the available content (context, format, and adaptation options) from the resource repository and the constraints (terminal and network capabilities, user characteristics, and preferences) from the receiving side. If there are multiple versions of the content pre-stored in the repository and one of these versions matches the constraints, then this version is selected, retrieved, and sent to the end user. However, if the available resource does not match the constraints, but can be adapted, then the adaptation decision engine determines the optimal adaptation for the given constraints and passes this decision to the resource adapta-

Figure 3. Different presentations of the same information

Workstation/LAN	PC/Dialup	TV Browser	Gray PDA	BW PDA	Text Browser
					"bridge"
38 KB	23 KB	8 KB	4 KB	0.6 KB	0.01 KB
24 bit color	24 bit color	256 colors	4 bit gray	B/W	-
256 x 256	192 x 192	128 x 128	96 x 96	64 x 64	-
22 sec	13.5 sec	4.7 sec	2.4 sec	0.35 sec	0.01 sec

Figure 4. Bird's-eye view of an adaptation engine architecture



tion engine. The resource adaptation engine retrieves the resource from the repository, applies the selected adaptation, and sends the adapted resource to the end user.

Constraints can be grouped into four broad categories: user and natural environment characteristics, terminal capabilities, and network characteristics. The terminal and network constraints will set an upper bound on the resources that can be transmitted over the network and rendered by the terminal. Information like the network's maximum bandwidth, delay and jitter, or the terminal's resolution, buffer size, and processing power, will help the adaptation engine determine the optimal version of the resource for the given network and terminal device. As the user is the actual consumer (and judge) of the resource, user-related information, including user preferences, user demographics, usage history and natural environment, is equally important in deciding which resource should be delivered to the terminal.

The adaptation engine needs to have sufficient information about the context and the format of the multimedia resources in order to make a decision whether the resource is suitable for the user or how it should be adapted in order to offer the user the optimal context and quality. The description should therefore include information on the resource type, semantics, available adaptation options, and characteristics of the adapted versions of the resource (Panis et al., 2003).

Scalable Coding

If content is scalable, then adapting content may be done using scalable coding. This removes or alters parts of resources in such a way as to reduce their quality in order to satisfy the receiver's capabilities and needs. Currently available scaling options depend on the coding format to be used. The Holy Grail of scalable video coding is to encode the video once, and then by simply truncating certain layers or bits from the original stream, lower qualities, spatial resolutions, and/or temporal resolutions could be obtained (Vetro, 2003).

Current scalable coding schemes fall short of this goal. MPEG-4 is currently the content representation standard where the widest range of scalability mechanisms is available, notably in terms of data types, granularities, and scalability domains (Pereira & Ebrahimi, 2002).

Transcoding

This is another more complex option that typically refers to transforming the resource from one coding format into another one, i.e., decoding the resource and encoding the resource using another codec (e.g., transcode an MPEG-2 video to an MPEG-1 video). According to Sun et al. (2003), the key design goals of transcoding are to maintain the video quality during the transcoding process and to keep complexity as low as possible. Cavallaro et al. (2003) identify three main approaches to video transcoding: content-blind transcoding, semantic transcoding, and description-based transcoding:

- **Content-blind transcoding** does not perform any semantic analysis of the content prior to conversion. The choice of the output format is determined by network and appliance constraints, independent of the video content (i.e., independent of the way humans perceive visual information). The three main content-blind transcoding categories are spatial conversion, temporal conversion, and colour-depth reduction.
- **Semantic (or intramedia) transcoding** analyses the video content prior to conversion. An example of such analysis is the separation of the video content

into two classes of interest, namely foreground and background. Once this separation has been accomplished, the two classes can be coded differently to better accommodate the way humans perceive visual information, given the available network and device capabilities.

- **Description-based (or intermedia) transcoding** transforms the foreground objects extracted through semantic segmentation into quantitative descriptors. These quantitative descriptors are transmitted instead of the video content itself. In this specific case, video is transformed into descriptors so as to produce a textual output from the input video. Such textual output can be used not only for transcoding, but also for annotating the video content and for translating the visual content into speech for visually impaired users. This transformation is also referred to as cross-media adaptation.

Location of Adaptation

As well as the technologies used, the location used for multimedia content adaptation needs to be addressed. Resource delivery and adaptation can be sender-driven, receiver-driven, or network-driven.

Sender-driven proceeds to adapt resources at the sender/server node depending on the terminal and/or network capabilities received beforehand. After successful adaptation the sender transmits the adapted version of the resource to the receiver. This action requires a serious amount of computational power at the server node and goes at the expense of latency between the receiver's request and the server's delivery.

In contrast, the receiver-driven approach decides what and how to adapt at the terminal side although the real adaptation could take place somewhere else, e.g., on a proxy node. Adaptation directly at the end node could fail due to insufficient capabilities. Additionally network bandwidth will be wasted, too. Nevertheless, adaptation on terminal devices should not be strictly excluded.

The pure network-driven approach is transparent where the network, i.e., the transport system, is responsible for adaptation only. Typical use case scenarios will cover all kind of adaptation approaches described so far, i.e., resource adaptability along the delivery chain, from resource provider to resource consumer. A high-performance server node will provide some kind of pre-processing in order to facilitate easy adaptation along the delivery chain across a wide range of network and terminal devices. Network nodes such as routers or gateways will then perform so-called light-weight adaptations using segment dropping or minor editing techniques whereas proxy nodes could utilise more complex adaptation techniques. Such complex adaptation techniques include not only scaling but also transcoding and cross-media. An adaptive terminal device could perform adjustments due to user and/or usage preferences. The complexity of these adaptations to be done in terminal devices depends on its capabilities, e.g., display resolution, computational power, local storage capacity, and buffer size.

User Profiling

In order for the personalisation and adaptation of multimedia content to take place, the users' preferences, interests, usage, and environment need to be described and modelled. This is a fundamental realisation for the design of any system that aims to aid the users while navigating through large volumes of audio-visual data. The expectation is that by making use of certain aspects of the user model, one can improve the efficacy of the system and further help the user (Kobsa et al., 2001).

This section described the components needed to make up a multimedia content adaptation architecture: transcoding, scalable coding, location of adaptation, and user profiling. The next section discusses the future of multimedia content adaptation by looking at UMA, transcoding, and scalability, specifically.

FUTURE TRENDS

The major problem for multimedia content adaptation is to fix the mismatch between the content formats, the conditions of transmission networks, and the capability of receiving terminals. A mechanism for adaptation needs to be created for this purpose.

Scalable coding and transcoding are both assisting in this. It can be seen that scalable coding and transcoding should not be viewed as opposing or competing technologies. Instead, they are technologies that meet different needs regarding multimedia content adaptation and it is likely that they will coexist.

Looking to the future of video transcoding, there are still quite a number of topics that require further study. One problem is finding an optimal transcoding strategy. Given several transcoding operations that would satisfy given constraints, a means for deciding the best one in a dynamic way has yet to be determined. Another topic is the transcoding of encrypted bit streams. The problems associated with the transcoding of encrypted bit streams include breaches in security by decrypting and re-encrypting within the network, as well as computational issues (Vetro, 2003).

The inherent problem with cross-media (description-based) adaptation is in preserving the intended semantics. What are required are not the blindfolded exchange of media elements and fragments, but their substitution by semantically equivalent alternatives. Unfortunately, current multimedia authoring tools provide little support for producing annotated multimedia presentations. Richly annotated multimedia content, created using document-oriented standards, such as MPEG-7 and MPEG-21 DIA, will help facilitate sophisticated cross-modal adaptation in the future.

For the implementation of UMA, "universal," scalable, video-coding techniques are essential components. Enhancements to existing video-coding schemes, such as MPEG-4 FGS (Fine-Granular-Scalability) and entirely new schemes will help drive the UMA ideal. More efficient FGS-encoders, tests on the visual impact of variability and more improved error resilient techniques are improvements that can be made to scalable coding schemes.

While some technologies such as content scalability and transcoding are fairly well established, there are still vital technologies missing for a complete multimedia content adaptation system vision. Many of these technologies are directly related to particular usage environments. While multimedia adaptation for improved experiences is typically thought of in the context of more constrained environments (e.g., mobile terminals and networks), it is also possible that the content has to be adapted to more sophisticated environments, e.g., with three-dimensional (3-D) capabilities. Whether the adaptation processing is to be performed at the server, at the terminal, or partially at both, is something that may have to be determined case–by-case, depending on such criteria as computational power, bandwidth, interfacing conditions, and privacy issues.

CONCLUSION

The development and use of distributed multimedia applications is growing rapidly. The subsequent desire for multimedia content adaptation is leading to new demands on transcoding, scaling, and, more generally, adaptation technologies. Metadata-based standards, such as MPEG-7 and MPEG-21, which describe the semantics, structure, and the playback environment for multimedia content are breakthroughs in this area because they can assist more intelligent adaptation than has previously been possible.

A prerequisite for efficient adaptation of multimedia information is a careful analysis of the properties of different media types. Video, voice, images, and text require different adaptation algorithms. The complex nature of multimedia makes the adaptation difficult to design and implement. By mixing intelligence that combines the requirements and semantic (content) information with low-level processing, the dream of UMA could be closer than we envision.

REFERENCES

Bormans, J., Gelissen, J., & Perkis, A. (2003). MPEG-21: The 21st century multimedia framework. *IEEE Signal Processing Magazine*, 20(2), 53- 62.

Böszörményi, L., Doller, M., Hellwagner, H., Kosch, H., Libsie, M., & Schojer, P. (2002). Comprehensive Treatment of Adaptation in Distributed Multimedia Systems in the ADMITS Project. ACM International Multimedia Conference, 429-430.

Cavallaro, A., Steiger, O., & Ebrahimi, T. (2003). Semantic segmentation and description for video transcoding. Paper presented at the Proceedings of the 2003 International Conference on Multimedia and Expo, 2003, ICME '03..

Composite Capabilities/Preference Profiles. (2003). Retrieved from the World Wide Web March 2003 at: http://www.w3.org/Mobile/CCPP/

Extensible Markup Language (XML). (2003). 1.0 (3rd Edition). Retrieved from the World Wide Web October 2003 at: www.w3.org/TR/2003/PER-xml-20031030/

Kobsa, A., Koenemann, J., & Pohl, W. (2001). Personalized Hypermedia Presentation Techniques for Improving Online Customer Relationships. *CiteSeer*.

ISO/IEC. (2002). ISO/IEC 15938-5: 2002: Information Technology—Multimedia Content Description Interface—Part 5: Multimedia Description Schemes.

ISO/IEC (2003). ISO/IEC JTC 1/SC 29/WG 11: MPEG-21 Digital Item Adaptation Final Committee Draft. Document N5845, Trondheim, Norway. Retrieved from the World Wide Web July 2003 at: http://www.chiariglione.org/mpeg/working documents.htm#MPEG-21

Panis et al. (2003). Bitstream Syntax Description: A Tool for Multimedia Resource Adaptation within MPEG-21, EURASIP Signal Processing. *Image Communication*, Special Issue on Multimedia Adaptation, 18(8), 721-74

Pereira, F., & Burnett, I. (2003). Universal multimedia experiences for tomorrow. *Signal Processing Magazine*, IEEE, 20(2), 63-73.

Pereira, F., & Ebrahimi, T., (2002). *The MPEG-4 Book*. Englewood Cliffs, NJ: Prentice-Hall.

Sun, H., Vetro, A., & Asai, K. (2003). Resource Adaptation Based on MPEG-21 Usage Environment Descriptions. Proceedings of the IEEE International Conference on Circuits and Systems, 2, 536-539.

Van Beek, P., Smith, J. R., Ebrahimi, T., Suzuki, T., & Askelof, J. (2003). Metadata-driven multimedia access. *Signal Processing Magazine*, IEEE, 20(2), 40-52.

Vetro, A. (2003). Visual Content Processing and Representation, Lecture Notes in Computer Science, (pp. 2849). Heidelberg: Springer-Verlag.

WAG. (2001.) User Agent Profile. Retrieved October 2001 from the World Wide Web at: http://www1.wapforum.org/tech/documents/WAP-248-UAProf-20011020-a.pdf

KEY TERMS

Bit Stream: The actual data stream, which is the transmission of characters at a fixed rate of speed. No stop and start elements are used, and there are no pauses between bits of data in the stream.

Content Scalability: The removal or alteration of certain subsets of the total coded bit stream to satisfy the usage environment, whilst providing a useful representation of the original content.

Cross-Media Adaptation: Conversion of one multimedia format into another one, e.g., video to image or image to text.

Multimedia Content Adaptation: The process of adapting a multimedia resource to the usage environment. The following factors make up this usage environment: users preferences, device, network, natural environment, session mobility, adaptation QoS, and resource adaptability.

Transcoding: The process of changing one multimedia object format into another.

UMA: How users can access the same media resources with different terminal equipment and preferences.

User Modelling: In the context of adaptation, the describing/modelling of the users preferences, interests, usage, and environment.

M

Multimedia Evaluations Based on Cognitive Science Findings

Eshaa M. Alkhalifa
University of Bahrain, Bahrain

INTRODUCTION

Multi-media systems waltzed into the lives of students and educators without allowing for the time required for the development of suitable evaluation techniques. Although everyone in the field is aware that judging this type of teaching software can only come through evaluations, the work done in this regard is scarce and ill organized. Unfortunately, in many of the cases the evaluation forms were just filled in by instructors who pretended they were students when they went through the tutorial systems (Reiser et al., 1994).

BACKGROUND

In the early days, some researchers regarded the evaluation of the program's functional abilities and efficiency to be important, so they defined them as formative evaluation, and they also defined the effectiveness of the system as summative evaluation (Bloom et al., 1971; Scriven, 1967).

Others believe the evaluation of the system is unimportant, so they focused on the latter by comparing student performance in pre- and post-test questionnaires prior to and following the use of the system, learning style questionnaires that targeted their learning preferences and a subjective questionnaire that investigated whether students like the system (Kinshuk et al., 2000). Unfortunately, many of the pre- and post-tests resulted in no significant1 differences in student grades when multimedia is compared to classroom lectures or to carefully organized, well-illustrated textbooks (Pane et al., 1996). These disappointing results caused researchers to question whether or not the correct evaluation questions are being asked; for example should the test be of interactivity versus lack of interactivity, or should one compare animation with textual media (McKenna, 1995)? If Pane et al. (1996) were aware of the work done by Freyd (1987), who studied the cognitive effects of exposing subjects to a series of still images to find that they are equivalent in the reactions they elicit to being exposed to a moving picture, then perhaps they would not have asked whether animation is equivalent to a textbook with carefully set images of all stages.

Since the problem that arose is the evaluation question, researchers continued to alter it in order to recognize what should be emphasized. Tam et al. (1997) proposed a three-part evaluation procedure that includes peer review, student evaluation as well as pre- and post-testing (Tam et. al., 1997). They were not able to get rid of the pre- and post-test evaluation, as it is the primary test for how much learning was achieved, and they still got no significant differences.

At this stage, researchers recognized that evaluations did not target the appropriate level of detail, so Song et al. (2000, 2001) presented empirical support that animation helps reduce the cognitive load on the learner. They also showed that multi-media is more effective in teaching processes than in teaching conceptual definitions, while textual presentations are better at the latter. However, all this was done in very limited test domains that lacked the realistic world of an educational system. Albalooshi and Alkhalifa (2002) implemented some of these ideas in addition to offering both textual representations and animations within the same screen to students. This supports individual learning preferences while offering multi-media systems as a cognitive tool. Such a tool requires an evaluation framework that is well informed of the justification behind its design and the way its main modules interact.

A 3-DIMENTIONAL FRAMEWORK FOR EVALUATION

In the reported cases, most of the evaluated systems failed to reflect their true abilities because some aspects of the design or effects were neglected. Consequently, a complete framework of evaluation is required to take into account all issues concerning the software and the learning process. Evaluation questions can be channeled into three main dimensions of evaluation that could then be subdivided into the various methods that form possible criteria that guide the evaluation process.

Table 1. A three-dimensional framework of evaluation

1st Dimension: System Architecture

This dimension is concerned with the system's main modules, their programming complexity as well as their interactions. Evaluation within this dimension should be performed in any or all of the following methods:

- Full description of system modules and complete check of interaction.
- Expert survey of the system filled by experts or educators.
- Student evaluations to consider their perspective of the system.
- Architectural design must be based on cognitive science findings rather than chance.
- Everything else concerning the system design such as cost analysis and portability.

2nd Dimension: Educational Impact

This dimension is concerned with assessing the benefits that could be gained by students when they use the system. Classically, these are done in pre- and post-tests, and this is carried on in this framework with more attention given to detail.

- Students grouped according to their mean grade in a quiz.
- Post-tests are used to compare one group with system only and another classroom only. A third group attends the classroom lecture with the class group and does a pre-test, and then uses the system before doing a post-test for comparison with the other two.
- Questions in the pre/post-tests must be mapped to each other to test the same types of knowledge, mainly consisting of declarative and procedural knowledge.
- The tests should best be attempted with students who were never exposed to this material previously to assess their learning rate.

3rd Dimension: Affective Measures

This dimension is mainly concerned with student opinions on the user friendliness of the system and allows them to express any shortcomings in the system. This could best be done through a survey where students are allowed to add any comments they wish freely and without restraints.

The framework will be explained through a case study that was performed of a data structure tutorial system (DAST) that was developed and evaluated at the University of Bahrain (AlBalooshi & Alkhalifa, 2003). The process started with a pre-evaluation stage, where students were all given a test and then were divided into groups of equivalent mean grades. This was done to allow each group to have members of all learning levels.

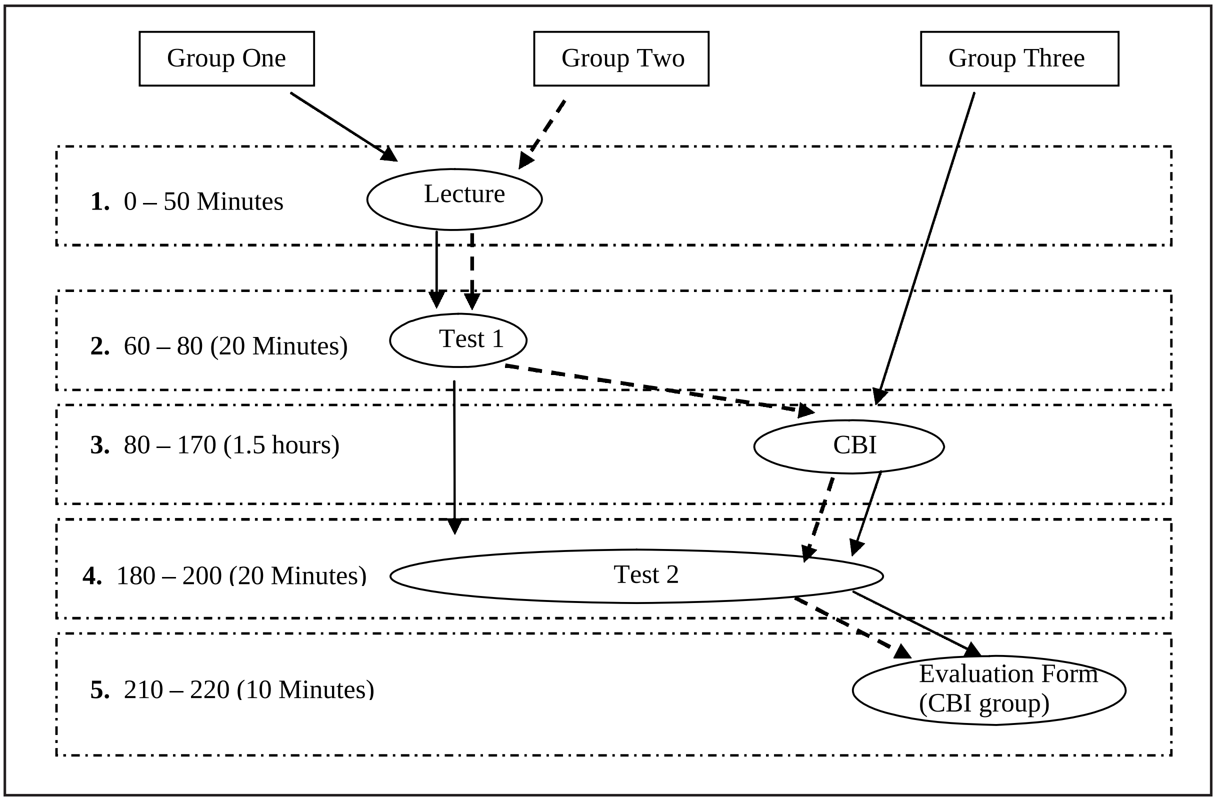
Then the pre- and post-tests were written to ensure that one set of questions mapped onto the next by altering their order while ensuring they include declarative questions that require verbalization of how students understand concepts as well as procedural questions that test if students understand how the concepts can be applied. Last but not least, a questionnaire was prepared to allow students to highlight what they regard as any weak areas or strong areas based upon their interaction with the system. The evaluation procedure for students is shown in Figure 1. Educators were also asked to fill in an evaluation form as experts.

ANALYSIS OF RESULTS

First of all, student grades were analyzed using the Analysis of Variance (ANOVA) test. This test allows the evaluation of the difference between the means by placing all the data into one number, which is F, and returning as a result one p for the null hypothesis. It will also compare the variability that is observed between conditions to the variability observed within each condition.

The statistic F is obtained as a ratio of two estimates of students' variances. If the ratio is sufficiently larger than 1, then the observed differences among the obtained means are described as being statistically significant. The term "null hypothesis" represents an investigation done between samples of the groups with the assumption that additional learning will not occur as a result of the treatment. In order to conduct a significance test, it is necessary to know the sampling distribution of F given that the significance level needed to investigate the null hypoth-

Figure 1. The evaluation procedure



esis. It must also be mentioned that the range of variation of averages is given by the standard deviation of the estimated means.

The ANOVA test did indeed show that there is a significant improvement in Group Two between the first test, which was taken after the lecture, and the second test, which was taken after using the system. However, this was not sufficient to be able to pinpoint the strengths of the system. Therefore, a more detailed analysis was done of student performance in the individual questions of test one and test two. Since the questions were mapped onto each other by design, it was easy to identify significant changes in student grades in a particular question type for students of group two who responded to similar questions before and after the use of the system. For example, a highly significant improvement with F=58 and $p<.000$ was observed in the question: "*Using an example, explain the stack concept and its possible use?*", which is an indication that the use of the system did strongly impact the student understanding of the concept of a "stack" in a functional manner.

Another point of view is to examine the scores by using the total average, which is 10.639, which can be approximated to 10.5, which can be used as a borderline. The rest of the scores can then be divided around this line. It was noticed that the average score of the third group was not high, yet 10 of the scores were above the borderline, while comparatively 6 scores were above it from the second group and only 6 of group one, which took the class only option. This shows the results of the third group, who took the CBI package alone and the second group, which had both the classroom lecture and the CBI package exposure to be close. It also underlines how much the second group improved their test results after taking the CBI and at the same time shows that the first group had not improved much only with the lecture learning. Alkhalifa (2001) tested in abstract logical context the effects of describing a state of a "*moving system*" versus describing it as a static system to find that this affects logical conclusions subjects eventually arrive at. This tutorial implemented those findings through this particular statistic, which revealed the effects of having animation in this multi-media system.

These results indicate that the use of the system may introduce a "limiting" effect that follows the initial introduction to the concepts (Albalooshi & Alkhalifa, 2002). Classroom lectures introduce students to the concepts, allowing them all the freedom to select all types of applications, which is in some ways overwhelming. The use of the system, on the other hand, produces a safe haven to test their ideas and strongly pursue the examples they can imagine, which helps them arrive at a solid procedural understanding of the concepts. It goes without saying that such a conclusion would have been impossible to make if the questions were not purposely set in the shown mapped fashion.

Additionally, students of groups two and three, who were exposed to the system, were asked to fill in an evaluation form composed of a series of questions as proposed by Caffarella (1987). They generally gave ratings of around 4 to 5 on a scale that went 0 to 6 with the highest for "The use of graphics, sound, and color contributes to the student's achievement of the objectives," and "The user can control the sequence of topics within the CBI program." The lowest score was 3.559 for "The level of difficulty is appropriate for you". Therefore, it seems that the students in general enjoyed learning through the system, although they found the level of difficulty of the concepts presented challenging.

In addition to all this, three peer experts filled in evaluation forms to rate the system from an instructor's point of view and they gave the system an average rating of 5.33 on the same scale of 0 to 6.

FINE-GRAINED EVALUATION

The evaluation framework proposed here evaluates multimedia educational software at a finer level of detail than what was previously followed. The system architecture, for example, is evaluated in terms of the cognitive assumptions on which it relies. In many cases this dimension was overlooked, as in the studies that found animation and carefully organized images to be equivalent (Byrne et al., 1999; Lawrence et al., 1994) because they were not informed of the work done to show that a sequence of images are translated as animation (Freyd, 1987). In fact, there is a difference between the two in that animation presents subjects with a cognitive tool that reduces the cognitive load they endure while learning and allows them to concentrate on the transfer of information rather than assimilating the presented material into animation (Albalooshi & Alkhalifa, 2002).

The educational impact question must also be carried out at a finer level of detail than just to report total grades for comparison. Specific questions may be designed to probe key types of learning, as was shown in the presented case study.

FUTURE TRENDS

Only through a fine-grained analysis can the true features of the system be revealed. It is this type of information that allows designers to take advantage of the opportunities offered by an educational medium that offers to transform the learning experience into a joy for both students and educators alike. Consequently, cognitive science concepts represent themselves to future research as a viable means to comprehend the learning process. This would guide the evaluative process to focus on the points of strengths and weaknesses of each system rather than assess it as a whole unit.

CONCLUSION

A three-dimensional framework is presented as a means to evaluating multimedia educational software in order to resolve the shortcomings of the current evaluation techniques. It differs from the other, in that it seeks a more fine-grained analysis while being informed through cognitive science findings. The extra detailed review reveals specific strengths of the system that may have been otherwise concealed if the comparison was done in the traditional manner. In other words, this process focuses on students' cognitive interpretations of what they are presented with rather than assuming to know how they will approach the educational content presented to them.

REFERENCES

Al Balooshi, F., & Alkhalifa, E.M. (2002). Multi-modality as a cognitive tool. In T. Okamoto & R. Hartley (Eds.), *Journal of International Forum of Educational Technology and Society, IEEE, Special Issues: Innovations in Learning Technology*, *5*(4), 49-55.

Alkhalifa, E.M. (2001). Directional thought effect in the selection task. *Proceedings of the Third International Conference on Cognitive Science, ICCS 2001*, Beijing, China (pp. 171-176).

Alkhalifa, E.M., & Al Balooshi, F. (2003). A 3-dimensional framework for evaluating multimedia educational software. In F. Al Balooshi (Ed.), *Virtual education: Cases in learning & teaching technologies*. Hershey, PA: Idea Group Publishing.

Bloom, B.S., Hastings, J.T., & Madaus, G.F. (1971). *Handbook on formative and summative evaluation of learning*. New York: McGraw-Hill.

Byrne, M.D., Catrambone, R., & Stasko, J.T. (1999). Evaluating animation as student aids in learning computer algorithms. *Computers and Education, 33*(4) 253-278.

Caffarella, E.P. (1987). Evaluating the new generation of computer-based instructional software. *Educational Technology, 27*(4), 19-24.

Freyd, J. (1987). Dynamic mental representations. *Psychological Review, 94*(4), 427-438.

Kinshuk, P.A., & Russell, D. (2000). A multi-institutional evaluation of intelligent tutoring tools in numeric disciplines. *Educational Technology & Society, 3*(4). *http://ifets.ieee.org/periodical/vol_4_2000/kinshuk.html*

Lawrence, W., Badre, A.N., & Stasko, J.T. (1994). *Empirically evaluating the use of animation to teach algorithms.* Technical Report GIT-GVU-94-07. Georgia Institute of Technology, Atlanta.

McKenna, S. (1995). Evaluating IMM: Issues for researchers. *Occasional Papers in Open and Distance Learning, 17.* Open Learning Institute, Charles Sturt University.

Pane, J.F., Corbett, A.T., & John, B.E. (1996, April). Assessing dynamics in computer-based instruction. *Proceedings of the 1996 ACM SIGCHI Conference on Human Factors in Computing Systems,* Vancouver, B.C. (pp.197-204).

Reiser, R.A., & Kegelmann, H.W. (1994). Evaluating instructional software: A review and critique of current methods. *Educational Technology, Research and Development, 42*(3), 63-69.

Scriven, M. (1967). The methodology of evaluation. In R.E. Stake (Ed.), *Curriculum evaluation.* Chicago: Rand-McNally.

Song, S.J., Cho, K.J., & Han, K.H. (2000). The effectiveness of cognitive load in multimedia learning. *Proceedings of the Korean Society for Cognitive Science* (pp. 93-98).

Song, S.J., Cho, K.J., & Han, K.H. (2001). Effects of presentation condition and content type on multimedia learning. *Proceedings of the Third International Conference on Cognitive Science, ICCS 2001*, Beijing, China (pp. 654-657).

Tam, M., Wedd, S., & McKerchar, M. (1997). Development and evaluation of a computer-based learning pilot project for teaching of holistic accounting concepts. *Australian Journal of Educational Technology, 13*(1), 54-67.

KEY TERMS

Cognition: The psychological result of perception, learning and reasoning.

Cognitive Load: The degree of cognitive processes required to accomplish a specific task.

Cognitive Science: The field of science concerned with cognition and includes parts of cognitive psychology, linguistics, computer science, cognitive neuroscience and philosophy of mind.

Cognitive Tool: A tool that reduces the cognitive load required by a specific task.

Declarative versus Procedural Knowledge: The verbalized form of knowledge versus the implemented form of knowledge.

Learning Style: This is the manner in which an individual acquires information.

Multimedia System: Any computer delivered electronic system that presents information through different media that may include text, sound, video computer graphics and animation.

ENDNOTE

[1] This term is defined in the Analysis of Results section.

Multimedia Information Filtering

Minaz J. Parmar
Brunel University, UK

Marios C. Angelides
Brunel University, UK

INTRODUCTION

In the film *Minority Report* (20th Century Fox, 2002), which is set in the near future, there is a scene where a man walks into a department store and is confronted by a holographic shop assistant. The holographic shop assistant recognises the potential customer by iris-recognition technology. The holographic assistant then welcomes the man by his name and starts to inform him of offers and items that he would be interested in based on his past purchases and what other shoppers who have similar tastes have purchased. This example of future personalised shopping assistants that can help a customer find shopping goods is not too far away from becoming reality in some form or another.

Malone, Grant, Turbak, Brobst, and Cohen (1987) introduced three paradigms for information selection, *cognitive*, *economic*, and *social*, based on their work with a system they called the Information Lens. Their definition of cognitive filtering, the approach actually implemented by the Information Lens, is equivalent to the "content filter" defined earlier by Denning, and this approach is now commonly referred to as "content-based" filtering. Their most important contribution was to introduce an alternative approach that they called social (now also more commonly called collaborative) filtering. In social filtering, the representation of a document is based on annotations to that document made by prior readers of the document.

In the 1990s much work was done on collaborative filtering (CF). There were three systems that were considered to be the quintessential recommender systems. The Grouplens project (Miller, Albert, Lam, Konstan, & Riedl, 2003) initially was used for filtering items from the Usenet news domain. This later became the basis of Movielens. The Bellcore Video recommender system (Hill, Stead, Rosenstein, & Furnas, 1995), which recommended video films to users based on what they had rented before, and Ringo (Shardanand & Maes, 1995), which later was published on the Web and marketed as Firefly, used social filtering to recommend movies and music.

BACKGROUND

Filtering multimedia content is an extensive process that involves extracting and modeling semantic and structural information about the content as well as metadata (Angelides, 2003). The problem with multimedia content is that the information presented in any document is multimodal by definition. Attributes of different types of media vary considerably in the way the format of the content is stored and perceived. There is no direct way of correlating the semantic content of a video stream with that of an audio stream unless it is done manually. A content model of the spatial and temporal characteristics of the objects can be used to define the actions the objects take part in. This content model can then be filtered against a user profile to allow granular filtering of the content, allowing for effective ranking and relevancy of the documents.

Filtering has mainly been investigated in the domain of text documents. The user's preferences are used as keywords, which are used by the filters as criteria for separating the textual documents into relevant and irrelevant content. The more positive keywords contained in a document, the more relevant the document becomes. Techniques such as latent semantic indexing have found ways of interpreting the meaning of a word in different contexts to allow accurate filtering of documents using different syntax, but allow the same semantics to be recognised and understood.

Text documents adhere to the standards of the language they are written in. Trying to do the same for AV data streams, you are faced with the problem of identifying the terms in the content itself. The terms are represented as a series of objects that appear in the content, for example, a face in an image file. These terms cannot be directly related to the objects as there is no method of comparison, or if there is, it is complex to unlock. The title of the document and some information might be provided in the file description, but the actions and spatial and temporal characteristics of the objects will not be described to a sufficient level for effective analysis of relevancy.

MAIN THRUST OF ARTICLE

Information-filtering techniques have been applied to several areas including American football (Babaguchi, Kawai, & Kitahashi, 2001), digital television (Marusic & Leban, 2002), Web applications (Kohrs & Merialdo, 2000), and ubiquitous and pervasive device applications (Tseng, Lin, & Smith, 2002).

Filtering multimedia information requires different approaches depending on the domain and use of the information. There are two main types of multimedia information filtering: collaborative and content based. If the user wants a subjective analysis of content in order to find a recommendation based on their individual preference, then they use collaborative filtering, also known as social or community-based filtering. If, on the other hand, they require an objective decision to filter information from a data stream based on their information needs, then they use content-based filtering.

All of the above systems use either collaborative or content-based filtering or a combination of both (hybrid) as the techniques for recommending predictions on candidate objects. There are existing information-filtering models outside these classic techniques such as temperament-based filtering (Lin & McLeod, 2002), which looks at predicting items of interest based on temperament theory. It works on the same principle as social filtering. Unlike social filtering, the users are grouped on temperaments of the users and not on similar item selection.

Content-Based Filtering

Content-based filtering is suited to environments where the user requires items that have certain content features that they prefer. Collaborative filtering is unsuitable in this environment because it offers opinions on items that reflect preferences for that user instead of providing filtering criteria that tries to disseminate preferred content from a data stream based on a user's preference. Personalised video summaries are the perfect domain to use content-based filtering. The reason for this is that a user will be interested in certain content only within any video data stream. For example, when watching a football game, the user may only be interested in goals and free kicks. Therefore, users can state what content features and other viewing requirements they prefer and then filter the footage against those requirements.

The content-based approach to information filtering has its roots in the information retrieval (IR) community and employs many of its techniques. The most prominent example of content-based filtering is the filtering of text objects (e.g., mail messages, newsgroup postings, or Web pages) based on the words contained in their textual representations. Each object, here, text documents, is assigned one or more index terms selected to represent the best meaning of the document. These index terms are searched to locate documents related to queries expressed in words taken from the index language. The assumption underlying this form of filtering is that the "meaning" of objects and queries can be captured in specific words or phrases. A content-based filtering system selects items based on the correlation between the content of the items and the user's preferences as opposed to a collaborative filtering system that chooses items based on the correlation between people with similar preferences (van Meteren & Someren, 2000).

The main problem with content-based filtering is that it does not perform well in domains were the content of items is minimal and the content cannot be analysed easily by automatic methods of content-based retrieval (e.g., ideas and opinions). Users with eclectic tastes or who make ad hoc choices are given bad recommendations based on previous choices. For example, Dad, who usually buys classic rock CDs for himself, purchases a So Solid Crew album for his 12-year-old son. He may start getting recommendations for hardcore garage dance anthems every time he logs in. CF does not suffer this problem as it will rank on other users' recommendations of similar choices. Comparative studies have shown that collaborative-filtering recommender systems on the whole outperform content-based filtering.

Collaborative Filtering

A purely content-based approach to information filtering is limited by the process of content analysis. In some domains, until recently, the items were not amenable to any useful feature extraction with content-based filtering (such as movies, music, restaurants). Even for text documents, the representations capture only certain aspects of the content, and there are many others that would influence a user's experience, for example, in how far it matches the user's taste (Balabanovic, 2000).

Collaborative filtering is an approach to overcome this limitation. The basic concept of CF is to automate social processes such as "word of mouth." In everyday life, people rely on the recommendations from other people either by word of mouth, recommendation letters, and movie and book reviews printed in newspapers. Collaborative filtering systems assist and augment this process and help people in making decisions.

There are two main drawbacks to using collaborative filtering: the sparsity of large user-item databases and the first-rater problem (Rashid et al., 2002). Sparsity is a condition when not enough ratings are available due to an insufficient amount of users or too few ratings per user.

An example of sparsity is a travel agent Web site, which has tens of thousands of locations. Any user on the system will not have traveled to even 1% of the locations (possibly thousands of locations). If a nearest-neighbour algorithm is used, the accuracy of any recommendation will be poor as a sufficient amount of peers will not be available in the user-item database. The first-rater problem is exhibited when a new user is introduced that has not enough ratings. If no ratings have been given for an item or a new user has not expressed enough opinions, choices, or ratings, no predictions can be made due to the insufficient data available or bad recommendations will be made. In contrast, content-based schemes are less sensible to sparsity of ratings and the first-rater problem since the performance for one user relies exclusively on his or user profile and not on the number of users in the system.

Hybrid Filtering

Both content-based and collaborative filtering have disadvantages that decrease the performance and accuracy of the systems that implement them. If these methods are combined, then the drawbacks of one technique can be counteracted by the techniques of the other, and vice versa. There have been various implementations such as the following.

- By making collaborative recommendations, we can use others' experiences as a basis rather than the incomplete and imprecise content-analysis methods.
- By making content recommendations, we can deal with items unseen by others.
- By using the content profile, we make good recommendations to users even if there are no other users similar to them. We can also filter out items.
- We can make collaborative recommendations between users who have not rated any of the same items (as long as they have rated similar items).
- By utilizing group feedback, we potentially require fewer cycles to achieve the same level of personalisation.

User Profiles

In information filtering, a user's needs are translated into preference data files called user profiles. These profiles represent the users' long-term information needs (Kuflik & Shoval, 2000). The main drawbacks of using user profiles are creating a user profile for multiple domains and updating a user profile incrementally. The user profile can be populated by one or more of the following.

- *Explicit profiling*: This type of profiling allows users to let the Web site know directly what they want. Each user entering the site will fill out some kind of online form that asks questions related to a user's preferences (Eirinaki & Vazirgiannis, 2003). The problem with this method is the static nature of the user profile once it has been created. The stored preferences in the user profile cannot take into account the changing user's preferences.
- *Implicit user profiles*: This type of user profiles is created dynamically by tracking the user's behaviour pattern through automatic extraction of user preferences using some sort of software agent, for example, intelligent agents, Web crawlers, and so forth (Eirinaki & Vazirgiannis, 2003). All these usage statistics are correlated into a usage history that is an accurate interaction between the user and the system. This usage history is then analysed to produce a user profile that portrays the user's interests. The user profile can be updated every time the user starts a new session, making implicitly made profiles dynamic. The downside of this method is that the user initially will have to navigate and explore the site before enough data can be generated to produce an accurate profile.
- *Hybrid of implicit and explicit profiling*: The drawbacks of explicit and implicit profiling can be overcome by combining both methods into a hybrid. This allows the strong points of one technique to counteract the shortcomings of the other and vice versa. The hybrid method works by collecting the initial data explicitly using an online form. This explicitly created data is then updated by the implicit tracking method as the user navigates around the site. This is a more efficient method over both pure methods. In some instances, this hybrid method is reversed and the implicit tracking methods are used initially to produce a profile.
- *Stereotype profiling*: This can be achieved by data mining and analysis of usage histories over a period of visits. This provides accurate profiling for existing users with legacy data that is accurate. The disadvantage of this method is that it suffers from the same static nature as explicit profiling as the profile is created from archive data that might be obsolete, and therefore some updating might be necessary. The predefined user stereotype is a content-based user profile that has been created for a virtual user or group of users who have common usage and filtering requirements for consumption of certain material. The stereotyped profile will contain additional information about the stereotyped user such as demographic and social attributes. This additional information is then used

Table 1. Multimedia information filtering

	Description	Techniques Used	Advantages	Disadvantages	Future R&D
Information Filtering	filtering a dynamic information space using relatively stable user requirements	SDI systems recommender systems	allows user to constantly receive content they are interested in with minimal user effort	does not support ad hoc queries that are dynamic compared to the information space they are searching (information retrieval)	all of the below
Content-Based Filtering	filtering content from a data stream based on extracting content features that have been expressed in a content-based user profile	vector space model probabilistic/inference models latent semantic indexing	objective analysis of large and/or complicated (e.g., multimedia) sources of digital material without much user involvement	1. content dependent 2. hard to introduce serendipitous recommendations as approach suffers from "tunnel vision" effect	extracting semantics from the structure of the content automatically without human intervention
Collaborative Filtering	filtering items based on similarities between target user's collaborative profile and peer users/group	same as above	1. content independent 2. proves more accurate than content-based filtering for most domains of use enables introduction of serendipitous choices	1. sparsity: poor prediction capabilities when new item is introduced to database due to lack of ratings 2. new user: poor recommendations made to new users until they have enough ratings in their profiles for accurate comparison to other users	solving the sparsity and new-user problem finding other types of ratings schemes that do not use comparisons between users tastes (e.g., filtering using users temperament)
Hybrid Filtering	combines two or more filtering techniques	simple or rule based stereotype collaborative content based	to reduce weak points and promote strong points of each of the techniques used	weak points can outweigh strong points if the hybrid is created naively	using hybrid systems in domains where using one technique presents a large disadvantage/problem
User Profiles	log file containing user's preferences for consumption of content	content-based profiles collaborative profiles	1. user does not need to state preferences each and every time they use the system 2. user can maintain and update preferences with minimal effort compared to ad hoc methods	needs frequent updating or user preferences become stagnant	user profiles that are as ubiquitous and pervasive as the devices/systems that use them standardisation
Explicit User Profiles	user manually creates user profile by means of a questionnaire	questionnaires ratings	preference information gathered is usually of high quality	requires a lot of effort from user to update	collecting new user preferences that reduces user effort
Implicit User Profiles	system generates user profile from usage history of interactions between user and content	machine learning algorithms	minimal user effort required easily updateable by automatic methods	initially requires a large amount of interaction between user and content before an accurate profile is created	new machine learning algorithms for better accuracy when creating implicit user profiles
Hybrid User Profiles	combination of user profile techniques used to create a profile	explicit/implicit user profiles	to reduce weak points and promote strong points of each of the techniques used	N/A	finding effective strategies for deployment and use of hybrid profiles

to place new users to stereotyped profiles that match similar demographic and social traits. The new user without the need of any implicit or explicit tracking automatically inherits preference information.

FUTURE TRENDS

Content-based filtering in multimedia information filtering has one innate problem that researchers are trying to solve: How can we extract semantics automatically from structural content of the model? In collaborative filtering, the age-old problem of sparsity and the new-user problem are still the biggest hindrances to using this method of filtering. Sparsity is being solved presently by hybrid systems, and it appears that this will be favoured way of dealing with sparsity (Lin & McLeod, 2002). The most promising solutions appear to be collaboratively filtering, standardised content-based profiles, which allow flexibility for systems to use either pure content-based or collaborative filtering, or a hybrid of both interchangeably.

Current work on user profiles focuses on improving creation techniques such as improved machine learning algorithms that create implicit user profiles more rapidly so that they can be more reliable and accurate in a shorter amount of time. For explicit user profiling, there is the work on selecting items that increase the usefulness of initial ratings that we have already discussed. The main way forward here, though, appears to be hybrid user profiles that are initially explicitly created and then implicitly updated.

With the advent of digital television and broadband, consumers will be faced with a deluge of multimedia content available to them at home and at work. What they will require are autonomous, intelligent filtering agents and automated recommender systems that actively filter information from multiple content sources. These personalisation systems can then collaborate to produce ranked lists of recommendations for all purposes of information the user might require. The key to this kind of service is not in the implementation of these systems or the way they are designed, but rather on a standard metadata language that will allow systems to communicate without proprietary restrictions and aid in end-user transparency in the recommendation process.

CONCLUSION

In the coming years, as nearly all communication and information devices become digital, we will see the development of systems that will be able not only to recommend items of interest to us, but will be able to make minor decisions for us based on our everyday needs such as ordering basic shopping groceries or subscribing to entertainment services on an ad hoc basis. What is required is a model of the user that describes the user's preferences for a multitude of characteristics that define the user's information needs. This model can then be used to filter data and recommend information based on this complete view of the user's needs. This has been done for many years with text files using techniques such as content-based and collaborative filtering, but has always been a problem with multimedia as the content is diverse in terms of storage, analysis techniques, and presentation. In recent years, classical techniques used for text filtering have been transferred and used in the area of multimedia information filtering. New developments such as hybrid filtering and improved metadata languages have made filtering multimedia documents more reliable and closer to becoming a real-world application.

REFERENCES

20th Century Fox Pictures. (2002). *Minority report* [Motion picture]. 20th Century Fox Pictures.

Angelides, M. C. (2003). Guest editor's introduction: Multimedia content modelling and personalization. *IEEE Multimedia, 10*(4), 12-15.

Babaguchi, N., Kawai, Y., & Kitahashi, T. (2001). Generation of personalised abstract of sports and video. *IEEE Expo 2001*, 800-803.

Balabanovic, M. (2000). An adaptive Web page recommendation service. *First International Conference on Autonomous Agents*, 378-385.

Eirinaki, M., & Vazirgiannis, M. (2003). Web mining for Web personalization. *ACM Transactions on Internet Technology (TOIT), 3*(1), 1-27.

Hill, W., Stead, L., Rosenstein, M., & Furnas, G. (1995). Recommending and evaluating choices in a virtual community of use. *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*, 194-201.

Kohrs, A., & Merialdo, B. (2000). Using category-based collaborative filtering in the active Web museum. *IEEE Expo 2000*.

Kuflik, T., & Shoval, P. (2000). Generation of user profiles for information filtering: Research agenda. *Proceedings of 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 313-315.

Lin, C., & McLeod, D. (2002). Exploiting and learning human temperaments for customized information recommendation. *Internet and Multimedia Systems and Applications*, 218-223.

Malone, T. W., Grant, K. R., Turbak, F. A., Brobst, S. A., & Cohen, M. D. (1987). Intelligent information sharing systems. *Communications of the ACM, 30*(5), 390-402.

Marusic, B., & Leban, M. (2002). The myTV system: A digital interactive television platform implementation. *IEEE Expo 2002*.

Miller, B. N., Albert, I., Lam, S. K., Konstan, J. A., & Riedl, J. (2003). MovieLens unplugged: Experiences with an occasionally connected recommender system. *Proceedings of ACM 2003 International Conference on Intelligent User Interfaces (IUI'03)*.

Rashid, M., Albert, I., Cosley, D., Lam, S. K., McNee, S. M., Konstan, J. A., et al. (2002). Getting to know you: Learning new user preferences in recommender systems.

Shardanand, U., & Maes, P. (1995). *Social information filtering: Algorithms for automating "word of mouth."* Proceedings of the CHI-95 Conference, Denver, CO.

Tseng, B. L., Lin, C.-Y., & Smith, J. R. (2002). Video summarization and personalization for pervasive mobile devices. *SPIE* (Vol. 4676). San Jose.

van Meteren, R., & Someren, M. (2000). *Using content-based filtering for recommendation.* Retrieved from http://citeseer.nj.nec.com/499652.html

Wyle, M. F., & Frei, H. P. (1989). Retrieving highly dynamic, widely distributed information. In N. J. Belkin & C. J. van Rijsbergen (Eds.), *Proceedings of the 12th Annual International ACMSIGIR Conference on Research and Development in Information Retrieval* (pp. 108-115). ACM.

KEY TERMS

Collaborative filtering: aims at exploiting preference behaviour and qualities of other persons in speculating about the preferences of a particular individual

Content-based filtering: organizes information based on properties of the object of preference and/or the carrier of information

Hybrid filtering: a combination of filtering techniques in which the disadvantages of one type of filtering is counteracted by the advantages of another

Information filtering: filtering information from a dynamic information space based on a user's long-term information needs

Recommendation: a filtered list of alternatives (items of interest) that support a decision-making process

Recommender systems: assist and augment the transfer of recommendations between members of a community

User profile : a data log representing a model of a user that can be used to ascertain behaviour and taste preferences

M

Multiple Internet Technologies in In-Class Education

Mihir A. Parikh
University of Central Florida, USA

Neeraj Parolia
University of Central Florida, USA

INTRODUCTION

The Internet has a symbiotic relationship with academia. The Internet sprung from and is continually improved by academic research. In parallel, the Internet is also changing the way academia provides education and training. Most universities now disseminate administrative information to students through the Internet. However, despite this recent upsurge in the adoption of the Internet, educational institutes have yet to fully utilize the power of various Internet technologies. Other than the Web, educational institutes have largely ignored various Internet technologies, which can aid students in the learning process.

We have to go beyond the Web and leverage multiple Internet technologies to support in-class education. Alternate Internet technologies have to be integrated under a unifying framework to make classroom-based education more efficient and effective. We need to deploy a right combination of multiple Internet technologies with appropriate teaching methods and instructional material to improve education (Huang, 2001; Mahoney, 1998; Spooner et al., 1998; Sumner & Hostetler, 1999). Web-only education support has several inherent problems. We have to deploy the framework to alleviate these problems and improve learning effectiveness yielded by the new methodology.

Table 1. Classification of learning environments by Wilson (1996)

- Computer Microworld: Self-contained computer based learning environment.
- Classroom-based learning environment: Traditional educational setup involving students and teacher.
- Virtual learning environment: Telecommunications based learning environment in which students are dispersed over large geographic area.

EARLY EXPERIENCES WITH THE WEB

The Internet can provide valuable contributions to all three learning environments listed in Table 1. In the computer microworld environment, it can help distribute, maintain, and update training software and educational modules. In the classroom-based learning environment, it can help distribute course material, such as lecture notes and assignments, via course Web sites and provide e-mail-based communication between the instructor and students. In the virtual learning environment, it can replace the traditional telecommunications-based video conferencing network with a ubiquitous, multimedia network.

When the Web is used to support classroom instruction, several problems emerge. Some of these problems are listed in Table 2.

These problems create disappointment and prompt several instructors to reduce the use of the Web in their courses. To lessen these problems, off-the-shelf software products, like WebCT, TopClass, and BlackBoard, are used. However, at the time of this study, they also had their own problems such as server-based content management; they require efforts on the part of the student to check the Web site regularly and offer no support for off-line browsing. They also require reformatting of the content developed through commonly used software like Word and PowerPoint. While they do help technologically-challenged instructors to easily develop and maintain course Web sites, they do little to eradicate most of the above problems. We have to look for an alternate solution.

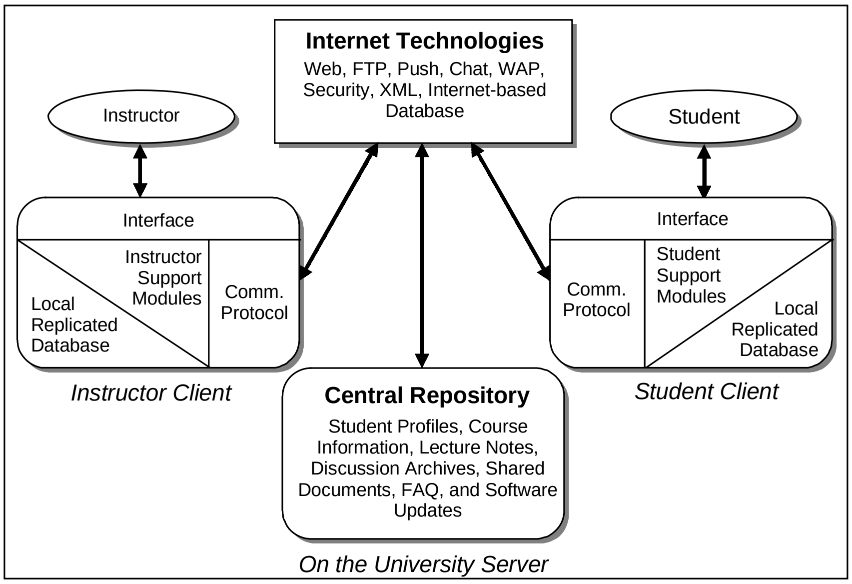
INTEGRATING INTERNET TECHNOLOGIES

The above problems relate primarily to the inherent limitations of the Web and insufficient utilization of other Internet technologies. It is easy to put documents on

Table 2. Problems with using the Web

Untimely Review of Material. The instructor regularly updates lecture notes and assignments on course Web sites, but they are not regularly reviewed by all students.	**Lack of Interactivity.** Interactivity needed for many learning activities and methods, such as group discussion, case study analysis, and real-time questions and answers are not well-developed on course Web sites.
No Confirmation Loop. The instructor does not always know who has reviewed the material and who has not.	**High Cost, No Reward.** Substantial costs are involved in developing Internet-compatible course material, in terms of time and efforts, but it brings little monetary or professional rewards for the instructor.
Wastage of Classroom Time. Significant portion of the classroom time goes in discussing and resolving technical problems.	**Varied Behavioral Response.** Some students display support for the new support technologies, while some students resist it.
Wastage of Instructor Time. The instructors usually spend substantial amount of time outside of class to develop and maintain course Web site and provide technical support to their students.	

Figure 1. Integrating multiple Internet technologies



course Web sites, but leveraging the full potential of the Internet requires integrating visual, aural, and textual material and providing nonlinear access to course material (Baer, 1998). Different Internet technologies will play an increasingly important role in the universities of the future (DosSantos & Wright, 2001).

This leads to the development and utilization of a novel, integrative model to support education (Figure 1, adapted from Parikh & Verma, 2002). This model goes beyond the Web to provide a unifying framework that can integrate and leverage various Internet technologies, such as the Web, FTP, chat, security, and Internet-based database in supporting education in the classroom-based learning environment. It has three main modules: central repository, which stores student and course-related information; instructor client, which assists the instructor in managing course information and administering the course; and student client, which assists students in accessing course information. Various Internet technologies connect these three modules and help perform information exchange task required for effective education.

NEW EDUCATION SUPPORT SYSTEM

Based on this model, a new easy-to-use education support system was developed and utilized it in eight sections of various types of courses in three semesters. It was

Table 3. Advantages of the framework

Collaborative Environment. Provides a collaborative environment, to facilitate free and easy interaction among students.	**Secure Delivery.** Enables secured delivery of course information, a significant part of which is usually an intellectual property of the instructor.
Instructor Visibility. Increases the visibility of instructors among students through synchronous and asynchronous communication.	**Mass Customization.** Using student profiles developed through the interactions with students, individual students are targeted with content customization
Confirmation Loop and Timely Feedback. Enables the instructor to track whether a student has reviewed the assigned material or not. Timely feedback is provided to the needy students.	**Ease of Use.** System modules are very simple to install and easy to use.
Increase in Efficiency. Reduces overall time spent by instructors in managing the course.	**Ease of Maintenance.** Push technology enables automatic transfer and installation on all clients without knowledge of instructors or students.
Reduced Unnecessary Meetings. Unnecessary student-instructor interactions are reduced because of FAQ and threaded discussion group databases.	**Empirical Findings.** In an empirical test, the system was found to be user friendly, to be useful, to increase convenience for students, and to provide accurate information which conformed to the needs of students in a timely manner.
Utilization of Preferred Technology. Enables the instructors to upload lecture notes and other course information in the original format.	

found that the new system was more advantageous in supporting all three types of learning activities (pre-, in, and post-class activities). The observations are in Table 3.

FUTURE TRENDS

Technology has become an integral part of learning environments. Computer hardware and software technologies provide support in computer microworld type learning environment and communications technologies provide support in the virtual learning environment. The model described previously was designed primarily for the classroom-based educational environment, but it can also be extended to the other two learning environments to simulate an interactive learning platform through synchronous and asynchronous interactivity.

On the technology front, two types of technologies are expected to have significant growth and influence in the upcoming years, peer-to-peer (p2p) technologies (e.g., Gnutella, Kazaa, etc.) and wireless combined with handheld technologies. These emerging technologies can also be utilized in the context of education under the same model discussed previously, as these technologies become omnipresent in the student community. Such technologies will involve multiple types of student client modules for various hardware types. However, all of these modules will synchronize with the single student profile stored on the central repository and replicated to all clients of the student. These technologies will further increase interactivity, facilitating the creation and sustenance of virtual communities that foster social relationships among the learners. These communities will further blur the boundaries of space and time within which education has been taking place.

Learning is not limited to the confines of academic institutions. Studies have shown that learning continues even in adulthood as individuals always adapt and learn through experiences (McCall et al., 1988). While the academic institutions are not yet at the stage of abandoning traditional degrees and adopting "learning contracts," they are preparing for the "life-long learning" as Alvin Toffler predicted. In close collaboration with corporations, they are embracing e-learning as a form of continuing education. Charles Handy (1989), a famous management guru, has foreseen that corporations would increasingly resemble universities or colleges in the years to come. We are already witnessing the beginning of this effect. U.S. corporations are spending over $60 billion annually on education with the growth of average 5% over the past decade (Prewitt, 1997). To benefit most from these efforts, going beyond the Web and leveraging all available technologies is necessary. The model discussed in this article can be the first right step in that direction.

CONCLUSION

Internet will bring about significant change in all aspects of education and training over the next decade (Aniebonam, 2000; Brandon & Hollingshead, 1999). A number of other users of Internet technologies to support university courses have suggested that they can be very effective (Jones & Rice, 2000; Stith, 2000). As the Internet is being increasingly integrated in education, it is imperative that we understand the limitations of using only one Internet technology, the Web. The Web is a powerful medium for delivering content or transferring knowledge. However, the core competency of educational institutions is developing knowledge, which can be done through intricate and robust networks and communities of students that last beyond the formal degree (Brown & Duguid, 1996). The Web falls short in this respect, probably because it is too broad and too open.

To provide more effective platform to support learning, we have to look beyond the Web. Many new Internet-based technologies have emerged recently and new ones continue to surface time and again. These technologies can provide complementary support to various educational activities that are not effectively supported by the Web. This article presented a case study of a system that integrated multiple Internet technologies, including the Web, to support learning. The system was indigenously developed with built-in flexibility to adapt to various types of courses. Further development and deployment of systems like this will provide the next frontier and drive the educational effort in the coming decades.

REFERENCES

Aniebonam, M.C. (2000, October). Effective distance learning methods as a curriculum delivery tool in diverse university environments: The case of traditional vs. historically black colleges and universities. *Communications of the Association for Information Systems*, 41–33.

Baer, W.S. (1998). Will the Internet transform higher education? The emerging Internet: Annual review of the Institute for Information Studies. Aspen, CO: The Aspen Institute. *http://www.aspeninst.org/dir/polpro/CSP/IIS/98/98.html*

Brandon, D.P., & Hollingshead, A.B. (1999, April). Collaborative learning and computer-supported groups. *Communication Education, 48*, 109–126.

Brown, J.S., & Duguid, P. (1996). Universities in the digital age. *Change, 28*(4), 11-19.

DosSantos, B., & Wright, A. (2001). *Information Services and Use, 21*(2), 53-64.

Handy, C. (1989). *The age of unreason*. Boston: Harvard Business School Press.

Huang, A.H. (2001). Problems associated with using information technology in teaching: A research proposal. *Proceedings of the Seventh Americas Conference on Information Systems* (pp. 39-40).

Jones, N.B., & Rice, M. (2000). Can Web-based knowledge sharing tools improve the learning process in an MBA consulting class? *The Journal, 27*(9), 100–104.

Mahoney, J. (1998). Higher education in a dangerous time: Will technology really improve the university? *Journal of College Admission, 24*(3), 161.

McCall, M.W., Jr., Lombardo, M.M., & Morrison, A.M. (1988). *The lessons of experience*. Lexignton, MA: D.C. Heath.

Parikh, M.A., & Verma, S.A. (2002). Utilizing Internet technologies to support learning: An empirical analysis. *International Journal of Information Management, 22*(1), 27-46.

Prewitt, E. (1997, January). What managers should know about how adults learn? *Management Update, 2*, 5.

Spooner, F., Spooner, M., Algozzine, B., & Jordan, L. (1998). Distance education and special education: Promises, practices, and potential pitfalls. *Teacher Education and Special Education, 21*(2), 121-131.

Stith, B. (2000). Web-enhanced lecture course scores big with students and faculty. *The Journal, 27*(8), 20–25.

Sumner, M., & Hostetler, D. (1999). Factors influencing the adoption of technology in teaching. *Proceedings of the Fifth Americas Conference on Information Systems* (pp. 951-953).

Wilson, B.G. (1996). *Constructivist learning environments: Case studies in instructional design*. Englewood Cliffs, NJ: Educational Technology Publication.

KEY TERMS

FTP: A protocol used to transfer files over a TCP/IP network (Internet, UNIX, etc.).

Handheld Technology: A computing device that can be easily held in one hand while the other hand is used to operate it.

Peer-to-Peer (P2P) Technology: Often referred to simply as *peer-to-peer*, or abbreviated *P2P*, a type of network in which each workstation has equivalent capabilities and responsibilities.

Push Technology: A data distribution technology in which selected data are automatically delivered into the user's computer at prescribed intervals or based on some event that occurs.

Replicated Database: A regular database in which tables, queries, and reports cannot be modified in design.

Video Conferencing: A video communications session among three or more people who are geographically separated.

Virtual Communities: A group of individuals who share a common interest via e-mail, chat rooms or newsgroups (threaded discussions). Members of a virtual community are self-subscribing.

WAP: A standard for providing cellular phones, pagers and other handheld devices with secure access to e-mail and text-based Web pages.

XML: (EXtensible Markup Language), an open standard, developed by the W3C, that involves formal syntax for adding structure and/or content information in a Web-based document. This subset of SGML defines data elements in a neutral way for easy interchange of structured data, such as mark-up tags, definitions, transmission validation, and interpretations across applications and organizations.

Music Score Watermarking

P. Nesi
University of Florence, Italy

M. Spinu
EXITECH S.r.L., Certaldo, Italy

INTRODUCTION

Music publishers, authors and/or distributors have high quantity of music scores in their archives. In classical music, the original music piece is normally kept in paper format, since its production goes back to many years ago. At present, only new light and popular music pieces are in symbolic notation formats. Light and popular music have a limited lifetime when compared with classical music pieces. The duration of the copyrights for that kind of music is about 60-80 years. Content owners are very cautious to transform their classical music pieces in digital format for e-commerce purposes, because they consider it as a highly risky process which could ultimately lose their copyright ownership. The situation is different when it comes to light and popular music, being market life shorter. According to content owners' opinion, e-commerce for music distribution cannot be accepted, unless adequate protection mechanisms are provided, as highlighted in WEDELMUSIC (www.wedelmusic.org) and MUSICNETWORK (www.interactivemusicnetwork.org). They accept to have their music protected only if it is possible to control while at the same time the users exploit content functionalities according to the established permissions and prices. To cope with these problems, mechanisms for protecting digital musical objects are used (see Table 1).

In this article, only problems and solutions for protecting and watermarking music scores are discussed.

Most music scores are still kept in paper format at publisher's archives. A first step to transform them into digital documents can be transforming them into images with a scanner. Another possible solution can be found in transforming them manually into symbolic music with a music editor. Obviously, this latter solution is very expensive, since the music has to be totally retyped. The use of very efficient Optical Music Recognition (OMR) software, similar to the Optical Character Recognition (OCR), seems to be quite unlikely in the next future. Currently, their recognition rate is close only to 90%, which makes this approach not too much reasonable when compared with retyping (www.interactivemusicnetwork.org, see assessment on the Working Group of Music Imaging).

Music images or symbolic music are obtained after music sheet digitalization. In the event of images, no further music manipulation is possible at the level of symbols. On the other hand, images can be easily viewed in any operating systems and with plenty of applications. The symbolic music gives several advantages in the score maintenance and manipulation; it allows the user to perform changes on the music, such as to justify it, change the page settings, add ornaments, accents, expressions, view single parts or the whole score, and so forth. The drawback consists in all these possible operations being performed only if the music editor is available: professional music sheets are produced by expensive and professional music editors.

It is well known that music sheets are distributed in paper format among musicians. Therefore, it seems that such digitizing process is useless. Practically speaking, Internet music sheet distribution, meaning from publishers to consumers, can only be achieved using digital formats. Distribution among users, as it occurs now with photocopies, could be made even via digital music sheets, as Napster did with audio files. Please note that on P2P (peer to peer) application there is also a quite significant

Table 1. Mechanisms for protecting digital musical objects

- encryption techniques to support any transferring of music objects;
- watermarking audio files in different formats;
- watermarking images of music score sheets;
- watermarking music sheets while they are printed from symbolic notation files.
- definition of digital rights management policies.

distribution of music scores (www.interactivemusicnetwork.org, read report on Music Distribution Models of the Working Group on Music distribution).

Whenever using digital formats, music could be converted again into paper (today musicians play music only from paper sheets).

BACKGROUND

The most relevant features for algorithms of score watermarking can be summed up into three categories (Monsignori, Nesi, & Spinu, 2003):

- *Content Requirements:*
  The embedded data may contain a simple identification code, which allows to recover the publisher and the distribution IDs simply by consulting a Web service. To this end, hiding about 100 bits is typically enough. The code can be encrypted, compressed and may include control and redundant bits to increase robustness.
- *Visual Requirements:*
  The watermark inserted in the printed music sheet has to be invisible for musicians or at least it should not bother musicians during their execution.
  The watermark has to be included in the music printed by the final user in any format if the music is available in symbolic format. Therefore, the watermark reading has not to depend on the availability of the original reference image of the music sheet.
- *Resistance Requirements:*
  The cost to remove watermark must be extremely expensive when compared to any regular purchase of the same music sheet.
  The watermark must resist against music sheet manipulation until the music printed becomes unreadable. Typically, five levels of photocopy are enough to make music unreadable or of a very bad quality.
  The watermark has to be readable when processing each single page or smaller part.

In addition, there are other parameters to be taken into account in order to analyze the technique capability.

- The amount of embedded information has a direct influence on watermark robustness. Typically, the hidden code is repeated several times in the same page; therefore, the bigger is the code, the lower is the number of times such code can be repeated, which means a decrease in the general robustness.
- Embedding strength " There is a trade-off between watermark embedding robustness and quality. Increased robustness requires a massive embedding of hidden bits. This increases music score degradation and watermark visibility.

Please note that watermarking images of scores or watermarking symbolic music lead to the same result: a watermarked music sheet. The watermarked music (symbolic or image) should be kept in some unchangeable digital file formats (like PDF) or in some formats difficult to change (PostScript), image format. The implementations of the algorithms for music watermarking in such two events are completely different (Busch, Nesi, Schmucker, & Spinu, 2002). In the first event, the watermarking is performed while the music score is printed by manipulating graphic primitives such as lines, music fonts, and so forth, and the process may generate a PostScript file or may send the information directly to the printer. In the latter case, the watermarking is performed by manipulating the B/W images.

In order to read the watermarked hidden code, the music sheet has to be scanned and the resulted image has to be elaborated with the watermark reader, to reconstruct the embedded code. The main advantages of distributing symbolic music sheets, instead of images are:

- Lower number of bytes for coding music, easier distribution, lower costs of download, and so forth;
- Higher quality of the printed music sheets, depending on the printer of the final user;
- Possibility of manipulating music notation for transposing, adding annotation, rearranging, and so forth; and
- Possibility of performing a direct music execution from symbolic format to produce MIDI or extended MIDI formats.

All of these features make the use of symbolic music more interesting for music distribution, and therefore its watermarking is very important for music protection.

APPROACHES

According to the user requirements, the printed music sheets must be produced at high resolution and quality. In appreciated music sheets, there is no noise, meaning that the information is in black and white, and therefore no space is left to hide information inside noise or in any kind of noise added-image. This means that the hidden code can be included only under the shape or in the position of music notation symbols. According to such purpose, some common elements of music sheets can be considered: staff lines, stems, note head, bar lines, and so forth.

Figure 1. Stem rotation approach

While stepping into such a direction, it is necessary to find a compromise between quality and watermark readability. Quality is very important for musicians and some minor changes could produce readability problems to musicians. They pay attention to the design of musical symbols, and any detectable variation may disturb the musician when playing. In general, the information to be hidden can be included in the changes considering both their presence and absence, for instance, coding 1 and 0 respectively. In some cases, the magnitude of the change can be used to hide more bits, for example in the orientation, the angle can be variable in order to add more bits.

Stem Rotation

The greatest problems of hiding information in the stem rotation (Busch, Rademer, Schmucker, & Wothusen, 2000) cope with the music score degradation and the low capacity in terms of hidden bits. As depicted in Figure 1, an untrained musician can identify that kind of changes in the music score. This method bothers many musicians when the music is read. In addition, the original music page is needed for watermark reading.

Beam Thickness Modification

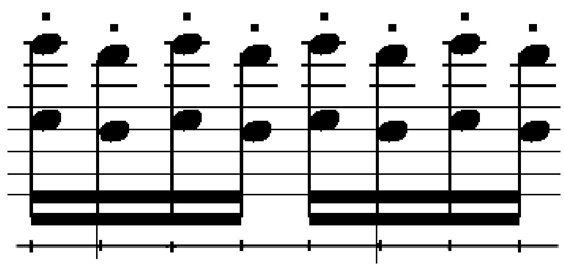
By modifying the orientation or thickness of beam lines, it is possible to hide only a few bits. Another important problem has to deal with the presence of beams which is not guaranteed in the music page. Musicians may easily detect the thickness variation when the beam is placed near a staff line. Furthermore, this method requires the original music page in order to perform the watermark reading.

Noteheads Shifting

The approach chosen by Schmucker, Busch, and Pant (2001) consists in shifting note heads (see Figure 2). The distance among notes has a musical significance. Therefore, in several cases, the approach may disturb the music reading. In Figure 2, the second chord from the left was moved to left, and musicians may detect the missed alignment of the chords. The displacement has been highlighted with the line below the staff and the gray lines. The movement of notes may generate problems when notes are marked with ornaments, accents, expressions, and so forth. In such cases, the movement becomes evident, thus creating a misalignment of notes with the markers. The idea can work things out and hide a significant code length, if there are enough noteheads in the score page.

If considering the main score, the shifted notes are quite easy to be detected by musicians reading them (according to the needs of simultaneity among parts/layers/voices), while it turns out to be quite invisible in single parts. Such a watermark is easy to be detected by musicians in regular groups of notes, provided that the distance among successive notes of the same beam is non-regular/periodic.

Figure 2. Shifting beamed notes approach



Different Fonts for the Same Music Symbol

According to this technique, different fonts for the selected music symbols are used to hide either 1 or 0, depending on the font used. This implies that the font has to be easily recognized during watermark reading. The approach was proposed for text watermarking by Maxemchuk and Low (1997).

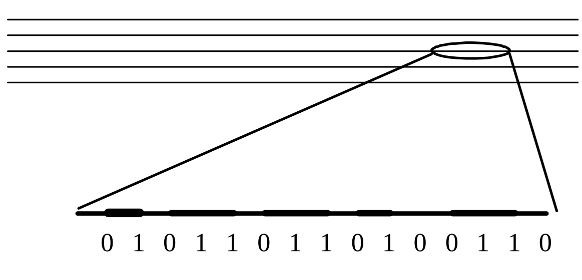
Watermarking Images of Music Sheets

The proposed methods are based on the possibility of storing information by exploiting the relationship of black and white pixels in image segments (i.e., a block) as information carrier (Funk & Schmucker, 2001). The method was elaborated upon Zhao and Koch method (1995) which is based on blocks of distinct size. The ratio of white and black pixels in certain block/area is used to embed a watermark. These areas are treated differently in the process of flipping pixels. The final idea is to embed the watermark only on the black pixels belonging to the staff lines. The fact that the pixel is on a line does not guarantee that it is on the staff line. For this purpose, only horizontal segment having a length greater than a fixed threshold was considered.

Line Thickness Modulation Approach

Figure 3 shows an example of the line modulation. It consists in modifying the lines' thickness in order to insert a binary code made up of several bits. Modulated lines can be easily noted if their presence is known, whereas they are not perceived if their presence is unknown (Monsignori, Nesi, & Spinu, 2001a, 2001b). This approach allows to hide a considerable number of bits in several instances per page, thus making the solution particularly suitable and robust to permit the watermark reading, even out of small parts of the music sheet. This approach has been used in the WEDELMUSIC Editor.

Figure 3. Staff lines thickness modification



The approach is robust with respect to staff bending, since the watermark is repeated on a large number of staff lines, and it can be read on bended lines. Moreover, a total of 108 bits can be hidden, and a certain number of CRC codes to increase the robustness has been added. This approach can be implemented only starting from the symbolic representation of music notation since the direct manipulation of staff lines on the image may introduce too much noise and produce line deformation.

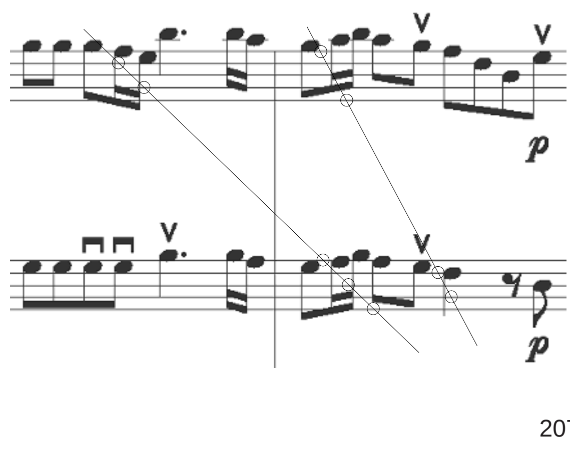
Line Mask Approach

This watermarking approach can be applied to images of music sheet or during the print out of a music score from a symbolic music notation file. The approach consists in marking some points in the music score for virtually hiding a number of lines connecting them (Monsignori et al., 2001a, 2001b). The position and the orientation of the hidden lines are used as the vehicle to hide the watermark code. In particular, the angle between the hidden line and the vertical axis has been used for hiding the information. The idea is not based on writing black lines on the music score (this may only lead up to destroy the music sheet). The points identifying the hidden line may be placed in the intersection among the hidden line and the staff lines, like it occurred with the points in Figure 4, highlighted with circles (in reality, they are interruption on the staff line). In the solution taken, groups of the lines contributing to encoding the same code start from common points. The method allows to hide a large number of bits for each page, and the code can be repeated several time increasing the robustness.

APPROACH VALIDATION

For the validation of these solutions two different phases have to be followed (Monsignori et al., 2003). First, the

Figure 4. Points chosen to be marked in the music score



validation has to be technically performed to assess the robustness against the attacks mentioned at the beginning of this article and to verify the effective coding of a large number of bits, repeated several times per page. As a second phase, the validation has to be focused on verifying the real applicability and acceptability of the solution from experts.

The experts' group has to cover the different needs: publishers, engravers, copyists, and many musicians which are the final users. Therefore, they are a very important category for the watermark validation. A specific watermark approach can be unacceptable for the musicians if the music sheet is not readable or annoying for the presence of evident changes. Typically, copyists are the most exigent. The validation has to request the assessment of a sequence of several different music score pages. Some of them are watermarked; others are not. Different levels of photocopy of the same watermarked or not watermarked music sheets have to be included. Different resolutions (dpi of the printer) of the same music sheets have also been used to assess the minimum acceptance level people involved in the validation. All music sheets were printed at the same magnitude, thus the dimension of the staff line was constant. Its value has been chosen according to that most commonly used in printed sheets.

Experts were informed about the main concepts of watermarking and not about these specific changes made in the music score. They have to perform the assessment individually, without being left with the possibility of a comparison with different pages of music and an exchange of opinions among one another.

FUTURE TRENDS AND CONCLUSIONS

As discussed in this article, the technology of music sheet watermarking is quite mature. Several algorithms have been tested and validated on real applications. The effective value of these solutions is similar to the watermark of Audio file. The presence of a specific watermark in the music sheet may be used to demonstrate the ownership of a music piece over the simple presence of textual fingerprints. In addition, the presence of the watermark can discourage people from any possible and intentional copying action of the music sheet for business purposes. The simple copying of the music sheet among friends is not prevented. The future trends of this technology are mainly in its application for monitoring the distribution of music sheets. In fact, the score watermark can be used for hiding code that can be detected during the simple distribution. This permits the content owners to set up specific services to control the data flow and thus to control and detect the passage of their digital items on the network.

REFERENCES

Busch, C., Nesi, P., Schmucker, M., & Spinu, M.B. (2002). Evolution of music score watermarking algorithm. In E.J. Delp III & P. Wong (Eds.), *Proceedings of the Real-Time Imaging V (EI12) IS\&T/SPIE 2002, Workshop on Security and Watermarking of Multimedia IV*, Vol.4675, San Jose, CA, USA, pp.181-193.

Busch, C., Rademer, E., Schmucker, M., & Wothusen, S. (2000). Concepts for a watermarking technique for music scores. In *Proceedings of 3rd International Conference on Visual Computing, Visual 2000*, Mexico City.

Funk, W., & Schmucker, M. (2001). High capacity information hiding in music scores. In *Proceedings of the International Conference on WEB Delivering of Music, WEDELMUSIC2001*, pp.12-19. Florence: IEEE Press.

Maxemchuk, N. F., & Low, S. (1997). Marking text documents. In *Proceedings of International Conference on Image Processing, ICIP97*, 3. Santa Barbara: IEEE Press.

Monsignori, M., Nesi, P., & Spinu, M.B. (2001a). A high capacity technique for watermarking music sheets while printing. In *Proceedings of IEEE 4th Workshop on Multimedia Signal Processing, MMSP2001*, pp.493-498. Cannes: IEEE Press.

Monsignori, M., Nesi, P., & Spinu, M.B. (2001b). Watermarking music sheet while printing. In *Proceedings of the International Conference on WEB Delivering of Music, WEDELMUSIC2001*, pp.28-35. Florence: IEEE Press.

Monsignori, M., Nesi, P., & Spinu, M.B. (2003). Technology of music score watermarking. In S. Deb (Ed.), *Multimedia systems and content-based image retrieval*. Hershey, PA: Idea Group Publishing, pp. 24-61.

Schmucker, M., Busch, C., & Pant, A. (2001). Digital watermarking for the protection of music scores. In *Proceedings of IS&T/SPIE 13th International Symposium on Electronic Imaging 2001, Conference 4314 Security and Watermarking of Multimedia Contents III*, 4314, pp.85-95, San Jose: SPIE Press.

Zhao, J., & Koch, E. (1995). Embedding robust labels into images for copyright protection. In *Proceedings of the International Congress on Intellectual Property Rights for Specialized Information, Knowledge and New Technologies*, pp.242-251, Vienna.

KEY TERMS

Fingerprinting: Used for calling the hidden serial numbers or anything else that should allow to the copyright owner to identify which reseller broke the license agreement. It is used for the multilevel document distribution.

Fragile Watermarking: Techniques that do not guarantee the watermark presence after few document manipulations.

Image Score: An image obtained from a page of music sheet, it can include a main score or a part.

Optical Music Recognition (OMR): Optical recognition of music, transcoding of an image score to a symbolic score format by using a specific algorithm, called OMR.

Robust Copyright Marking: A term used for the techniques that assure a watermark persistence also after the original document was changed in different ways (in the case of the images: cropping, resizing, brightness modification, etc.).

Staff Line: Each single line of the music score staff. The pentagram is made of five staff lines.

Steganography: Techniques that allow secret communication, usually by embedding or hiding the secret information (called embedded data) in other, unsuspected data. Steganographic methods are based on the assumption that the existence of the covert communication is unknown and they are mainly used in secret point-to-point communication between trusting parties. As a result, steganographic methods are usually not robust, that is the hidden information cannot be recovered after data manipulation.

Symbolic Score: A representation of the music notation in symbolic, including a description of music symbols and their relationships. This can be done in some formal specific format such as Finale, Sibelius, WEDELMUSIC, HIFF, SMDL, and so forth.

Watermark: The code hidden into a digital or analog object containing an ID (identification) code or other pieces of information. The watermark is used for identifying the fields of embedded data (serial numbers, logos, etc.) that tell us who is the owner of the object or supply an ID in order to identify data connected with the digital object.

Watermarking: Process of inserting a hidden code or message into a digital or analog object. As opposed to steganography, it has the additional notion of robustness against attacks. As the name suggests, the additional data (the watermark) is added in order to protect the digital document from copyright infringements. Even if the existence of the hidden information is known, it has to be hard for an attacker to destroy the embedded watermark without destroying the data itself.

Watermark Reading: Process of extracting the watermarked code into the watermarked object.

Natural Computing

Leandro Nunes de Castro
Catholic University of Santos (UniSantos), Brazil

INTRODUCTION

There are several artifacts developed by taking inspiration from natural organisms and phenomena. For instance, Velcro was inspired by a plant burr, bullet- proof vests were inspired by spider silk, sonars were inspired by bats, airplanes were motivated by the flying of birds, and the list goes on. The observation of nature has also allowed the development of several laws and theories that describe how parts of nature work; the laws of physics are a great example – laws of thermodynamics (conservation, entropy, and absolute zero); laws of movement (Newton's laws); laws of electromagnetism (Maxwell's laws); and others. They are used to explain the trajectories, attraction and other aspects of objects.

With the advent of computing, the way humans interact with nature changed dramatically. Nowadays, we not only observe, explain and use nature as an environment for living and providing us with food, we also use nature as a source of inspiration for the development of new technologies – new computing approaches, and we use computers to simulate and/or emulate natural phenomena as well. This scenario has changed even more remarkably over the last decade when researchers realized that it is actually possible to use natural mechanisms and means as a brand new computing paradigm, which is distinct from the well-known *in silico* computation. We are in the beginning of a new age: the natural computing age!

Background

Originally, the terminology *natural computing* was used to describe only those systems that employed natural means, such as DNA or RNA strands, to perform computation. Nowadays, the terminology has broadened to encompass three major areas: *computing inspired, or motivated, by nature*; 2) *the study of natural phenomena by means of computation;* and 3) *computing with natural means*. These three branches of natural computing have one aspect in common: a strong and computer-based or computer-oriented relationship with nature. This brief overview provides a taxonomy for natural computing based on these three sub-branches and lists the most important and well-known methods of each part.

COMPUTING INSPIRED BY NATURE

The first branch of natural computing is also the oldest one. With the discovery of many principles and theories about nature and the development of its several (theoretical) models, researchers from other areas realized that these could be useful as sources of inspiration for the development of computational systems for problem solving. In particular, computer scientists, engineers and others found many works on theoretical biology that could be used for the development of novel computing approaches. Computing inspired by nature has many subareas: 1) *artificial neural networks;* 2) *evolutionary algorithms;* 3) *swarm intelligence;* 4) *artificial immune systems;* and 5) others (models based on *growth, development, culture, etc.*).

A landmark work in the branch of biologically inspired computing was the paper by McCulloch and Pitts (1943), which introduced the first mathematical model of a neuron. This neuronal model, also known as artificial neuron, gave rise to a field of investigation of its own, the so-called artificial neural networks (Bishop, 1996; Fausett, 1994; Haykin, 1999; Kohonen, 2000).

Another computing approach motivated by biology arose in the mid 1960s with the works of I. Rechenberg (1973), Schwefel (1965), Fogel (Fogel et al., 1966), and Holland (1975). These works gave rise to the field of evolutionary computation (Bäck et al., 2000a, b; Bahnzaf & Reeves, 1998; Beyer, 2001; Fogel, 1998; Goldberg, 1989; Koza, 1992; Michalewicz, 1996; Mitchell, 1998), which uses ideas from evolutionary biology to develop (evolutionary) algorithms for search and optimization.

Swarm intelligence has two main frontlines: algorithms based on the collective behavior of social insects (Bonabeau et al., 1999), and algorithms based on sociocognition (Kennedy et al., 2001). In the first case, the collective behavior of ants and other insects has led to the development of algorithms for solving combinatorial optimization and clustering problems, among others. Algorithms based on sociocognition demonstrated effectiveness in performing search on continuous spaces.

Artificial immune systems borrow ideas from the immune system and its corresponding models to design computational systems for solving complex problems (Dasgupta, 1999; de Castro & Timmis, 2002; Timmis et al., 2003). This is also a young field of research that emerged around the mid 1980s. Its application areas range from biology to robotics. The other emerging types of algorithms inspired by nature are the cultural algorithms, the simulated annealing algorithm, the systems based on growth and development, the cells and tissues models, and various others (Aarts & Korst, 1989; Kochenberger & Glover, 2003; Kumar & Bentley, 2003; Paton, 1994; Paton et al., 2003).

STUDY OF NATURAL PHENOMENA BY MEANS OF COMPUTING

Differently from computing inspired by nature, studying natural phenomena by means of computing is a synthetic approach aimed at creating patterns, forms, behaviors, and organisms that do not necessarily resemble "life-as-we-know-it". It can result in completely new phenomena never observed in nature, but that possess enough features to be qualified as "natural" or "living". For example, artificial organisms can be created that do not bear any resemblance whatsoever to any known living being on Earth. The idea is to use computers to simulate and emulate natural phenomena in a non-deterministic fashion. The study of nature by means of computing has two main branches: 1) *artificial life* (ALife); and 2) the *computational, or fractal, geometry of nature*.

As put by C. Langton in his pioneering chapter on ALife:

"Artificial Life is the study of man-made systems that exhibit behaviors characteristic of natural living systems. It complements the traditional biological sciences concerned with the analysis *of living organisms by attempting to* synthesize *life-like behaviors within computers and other artificial media. By extending the empirical foundation upon which biology is based* beyond *the carbon-chain life that has evolved on Earth, Artificial life can contribute to theoretical biology by locating* life-as-we-know-it *within the larger picture of* life-as-it-could-be.*" (Langton, 1988, p. 1)*

In summary, ALife can be defined as the synthetic or virtual approach to the study of life-like patterns (forms), behaviors, systems, and organisms (Adami, 1998; Levy, 1992).

The computational visualization of (mathematical) models of natural structures and processes results in images, animations, and interactive systems useful as scientific, research and educational tools in computer science, engineering, biosciences and many other domains. This is the core of the computational geometry of nature: how to visualize models of natural phenomena. There are several techniques that can be used with this purpose, for example, cellular automata (Ilachinski, 2001; Wolfram, 1994), particle systems (Reeves, 1983), Lindenmayer systems (Lindenmayer, 1968), iterated function systems (Barnsley, 1988; Barnsley & Demko, 1985; Hutchinson, 1981), Brownian motion (Fournier et al., 1982; Voss, 1985), and so forth.

The applications of this branch of natural computing include computer-assisted landscape architecture, design of new varieties of plants, crop yield prediction, the study of developmental and growth processes, and the modeling and synthesis (and the corresponding analysis) of an innumerable amount of natural patterns and phenomena (Flake, 2000; Mandelbrot, 1983; Peitgen et al., 1992).

COMPUTING WITH NATURAL MEANS

In 1965, G. Moore observed that there is an exponential growth in the number of transistors that can be placed in an integrated circuit. According to the "Moore's law," there is a doubling of transistors in a chip every couple of years. If this scale remains valid, by the end of this decade silicon-based computers will have reached their limits in terms of processing power. One question that remains, thus, is what kind of medium, rather than silicon, can provide an alternative for the design and implementation of a computing device?

The last decade has seen the proposal and development of several of these alternative means of computing and their corresponding computing techniques. These approaches are mostly of two types: 1) the ones based on biomolecules, and 2) the ones based on quantum bits. The approaches that use biological molecules, such as DNA and RNA strands, or membranes for computing, are usually called *molecular computing* (Calude & Pâun, 2001; Gramß et al., 2001; Pâun et al., 1998; Pâun & Cutkosky, 2002; Sienko et al., 2003). By contrast, the approaches based on quantum bits constitute what is known as *quantum computing* or *quantum computation* (Hirvensalo, 2000; Nielsen & Chuang, 2000; Pittenger, 2000).

In the first case, biomolecules are used as means to store information and techniques from molecular (genetic) and chemical engineering are used to manipulate these molecules so as to process information. Note that this approach relies on the sophistication and efficiency of the genetic engineering techniques. Quantum computing, on the other hand, stores information in quantum bits and the evolution of the "quantum computer" follows the principles of quantum mechanics.

FUTURE TRENDS

In a few decades natural computing may be pervasive in our lives. Several artifacts will have technologies based on natural computing, such as control, optimization and data mining devices. Possibly computers will involve or be based upon biomolecules or quantum bits, and the way we interact with computers may also change considerably. The amount of human expertise necessary for the use of computers may decrease in an inverse proportion to the degree of automation of our machines. But until this stage is reached, much research efforts and accomplishments still have to be made.

CONCLUSION

Natural computing is thus the terminology used to refer to three types of systems: 1) computational algorithms for problem solving developed by taking inspiration from a natural process, theory, model or mechanism; 2) computational models for the simulation and/or emulation of natural systems and processes; and 3) novel computing paradigms that use media, other than silicon, to store and process information.

Although all these branches are quite young from a scientific perspective, several of them are already being used in our everyday lives. For instance, we now have "intelligent" washing machines, games, (virtual) pets, and so forth; research on ALife and the computational geometry of nature has allowed the creation of realistic models and the simulation of several plants and animal species, has aided the study of developmental processes, and so forth; and computing with natural means has given new insights into how to complement or supplement computer technology as known nowadays.

Natural computing is highly relevant for today's computer scientists and engineers because it has offered alternative, and sometimes brand new, solutions to problems yet unsolved or poorly resolved. It has also provided us with a new way of seeing, understanding and interacting with nature. Of course there is much to come and much to do in such a broad and young field as natural computing. However, we now know for sure that this is not only a promising field of investigation; its many applications and outcomes have been affecting our lives, even if this is not perceived by most people. And the best thing is that we are only in the beginning of the natural computing age!

REFERENCES

Aarts, E., & Korst, J. (1989). *Simulated annealing and Boltzman machines - A stochastic approach to combinatorial optimization and neural computing*. John Wiley & Sons.

Adami, C. (1998). *An introduction to artificial life*. Springer-Verlag/Telos.

Bäck, T., Fogel, D.B., & Michalewicz, Z. (2000a). *Evolutionary computation 1: Basic algorithms and operators*. Bristol: Institute of Physics Publishing (IOP).

Bäck, T., Fogel, D.B., & Michalewicz, Z. (2000b). *Evolutionary computation 2: Advanced algorithms and operators*. Bristol: Institute of Physics Publishing (IOP).

Banzhaf, W., & Reeves, C. (1998). *Foundations of genetic algorithms*. Morgan Kaufmann.

Barnsley, M.F. (1988). *Fractals everywhere*. Academic Press.

Barnsley, M.F., & Demko, S. (1985). Iterated function systems and the global construction of fractals. *Proc. of the Royal Soc. of London, A339*, 243-275.

Beyer, H.-G. (2001). *Theory of evolution strategies*. Springer-Verlag.

Bishop, C.M. (1996). *Neural networks for pattern recognition*. Oxford University Press.

Bonabeau, E., Dorigo, M., & Theraulaz, T. (1999). *Swarm intelligence: From natural to artificial systems*. New York: Oxford University Press.

Calude, C.S., & Pâun, G. (2001). *Computing with cells and atoms: An introduction to quantum, DNA, and membrane computing*. Taylor & Francis.

Dasgupta, D. (1999). *Artificial immune systems and their applications*. Springer-Verlag.

de Castro, L.N., & Timmis, J.I. (2002). *Artificial immune systems: A new computational intelligence approach*. Springer-Verlag.

Fausett, L. (1994). *Fundamentals of neural networks: Architectures, algorithms, and applications*. Prentice Hall.

Flake, G.W. (2000). *The computational beauty of nature*. MIT Press.

Fogel, D.B. (1998). *Evolutionary computation: Toward a new philosophy of machine intelligence*. IEEE Press.

Fogel, L.J., Owens, A.J., & Walsh, M.J. (1966). *Artificial intelligence through simulated evolution.* New York: Wiley.

Fournier, A, Fussell, D., & Carpenter, L. (1982). Computer rendering of stochastic models. *Comm. of the ACM, 25,* 371-384.

Goldberg, D.E. (1989). *Genetic algorithms in search, optimization, and machine learning.* Addison-Wesley Pub Co.

Gramß, T., Bornholdt, S., Groß, M., Mitchell, M., & Pellizzari, T. (2001). *Non-standard computation.* Wiley-VCH.

Haykin, S. (1999). *Neural networks: A comprehensive foundation.* Prentice Hall.

Hirvensalo, M. (2000). *Quantum computing.* Springer-Verlag.

Holland, J.H. (1975). *Adaptation in natural and artificial systems.* MIT Press.

Hutchinson, J. (1981). Fractals and self-similarity. *Indiana Journal of Mathematics, 30,* 713-747.

Ilachinski, A. (2001). *Cellular automata: A discrete universe.* World Scientific.

Kennedy, J., Eberhart, R., & Shi, Y. (2001). *Swarm intelligence.* Morgan Kaufmann Publishers.

Kochenberger, G.A., & Glover, F. (2003). *Handbook of metaheuristics.* Kluwer Academic Publishers.

Kohonen, T. (2000). *Self-organizing maps* (3rd ed.). Springer-Verlag.

Koza, J.R. (1992). *Genetic programming: On the programming of computers by means of natural selection.* MIT Press.

Kumar, S., & Bentley, P.J. (2003). *On growth, form and computers.* Academic Press.

Langton, C. (1988). Artificial life. In C. Langton (Ed.), *Artificial life* (pp. 1-47). Addison-Wesley.

Levy, S. (1992). *Artificial life.* Vintage Books.

Lindenmayer, A. (1968). Mathematical models for cellular interaction in development, parts I and II. *Journal of Theoretical Biology, 18,* 280-315.

Mandelbrot, B. (1983). *The fractal geometry of nature.* W.H. Freeman and Company.

McCulloch, W., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics, 5,* 115-133.

Michalewicz, Z. (1996). *Genetic algorithms + data structures = evolution programs* (3rd ed.). Springer-Verlag.

Mitchell, M. (1998). *An introduction to genetic algorithms.* The MIT Press.

Moore, G.E. (1965). Cramming more components into integrated circuits. *Electronics, 38*(8).

Nielsen, M.A., & Chuang, I.L. (2000). *Quantum computation and quantum information.* Cambridge University Press.

Paton, R. (Ed.). (1994). *Computing with biological metaphors.* Chapman & Hall.

Paton, R., Bolouri, H., & Holcombe, M. (2003). *Computing in cells and tissues: Perspectives and tools of thought.* Springer-Verlag.

Pâun, G., & Cutkosky, S.D. (2002). *Membrane computing.* Springer-Verlag.

Pâun, G., Rozenberg, G., & Saloma, A. (1998). *DNA computing.* Springer-Verlag.

Peitgen, H.-O, Jürgens, H., & Saupe, D. (1992). *Chaos and fractals: New frontiers of science.* Springer-Verlag.

Pittenger, A.O. (2000). *An introduction to quantum computing algorithms.* Birkhäuser.

Rechenberg, I. (1973). *Evolutionsstrategie: Optimierung technischer systeme nach prinzipien der biologischen evolution.* Frommann-Holzboog, Stuttgart.

Reeves, W.T. (1983). Particle systems – A technique for modeling a class of fuzzy objects. *ACM Transactions on Graphics, 2*(2), 91-108.

Schwefel, H.–P. (1965). *Kybernetische evolutionals strategie der experimentellen forschung in der stromungstechnik.* Diploma thesis. Technical University of Berlin.

Sienko, T., Adamatzky, A., & Rambidi, N. (2003). *Molecular computing.* MIT Press.

Timmis, J., Bentley, P.J., & Hart, E. (Eds.). (2003). Artificial immune systems. *Proc. of the International Conference on Artificial Immune Systems* (ICARIS 2003). Springer-Verlag.

Voss, R.F. (1985). Random fractals forgeries. In R.A. Earnshaw (Ed.), *Fundamental algorithms for computer graphics* (pp. 805-835). Berlin: Springer-Verlag.

Wolfram, S. (1994). *Cellular automata and complexity.* Perseus Books.

KEY TERMS

Analysis: The word analysis can have different meanings in different contexts. Within the natural computing domain, analysis corresponds to the investigation of a given phenomenon in order to obtain some information from it or to derive simplified computational systems (or theoretical models) that somewhat mimic the behavior of the natural system being analyzed.

Artificial Life (ALife): ALife can be defined as the synthetic or virtual approach to the study of life-like patterns (forms), behaviors, systems, and organisms, independently of the matter used for synthesis.

Computing Inspired by Nature: It embodies all approaches, mainly problem solving techniques, developed using ideas from, or inspired by, natural phenomena and/or their corresponding theoretical models. It is also sometimes referred to as computing with biological metaphors. Instead of trying to create accurate (theoretical) models, it usually results in high-level abstractions of natural phenomena.

Emergence: In the context of natural computing, an emergent phenomenon can be understood as the one whose global properties are neither possessed by, nor directly derivable from, any of its component parts. For instance, a single ant is a very simple insect with limited capabilities, but an ant colony is capable of performing complex tasks, such as nest building and organization.

Emulation: The *realization*, sometimes also called *emulation*, of a system or organism corresponds to a literal, material model that implements functions; it is a substantive functional device. Roughly speaking, a realization is judged primarily by how well it can function as an implementation of a design specification. A system or function is used to emulate or realize another, when one performs in exactly the same way as another. A typical example in computer science is the emulation of one computer by (a program running on) another computer.

Fractal Geometry: Fractal geometry is the geometry of the irregular shapes found in nature, and, in general, fractals are characterized by infinite details, infinite length, self-similarity, *fractal dimensions*, and the absence of smoothness or derivative.

Simulation: Simulations are metaphorical models that "stand for" something else, and may cover different levels of fidelity or abstraction, which can be performed by physical modeling, by writing a special-purpose computer program, or by using a more general simulation package that is probably still aimed at a particular kind of simulation. Computer simulation is pervasive in natural computing. It has been used to design problem-solving techniques that mimic the behavior of several biological phenomena, it has served to drive synthetic environments and virtual worlds, and it has been used to simulate "natural" computers.

Synthesis: Within the natural computing domain, synthesis corresponds to the act or process of developing or designing a computational system; an emulation or a simulation of a given natural phenomenon. It is usually an outcome of the analysis of the corresponding natural phenomenon.

Networks and Electronic Commerce Adoption in Small Businesses

Lejla Vrazalic
University of Wollongong, Australia

Robert MacGregor
University of Wollongong, Australia

Deborah Bunker
University of New South Wales, Australia

INTRODUCTION

The adoption and diffusion of electronic commerce (e-commerce) in small businesses remains a critical area of investigation in information systems (IS) literature. A number of studies (Miles, Preece, & Baetz, 1999; Overby & Min, 2001) have suggested that in order to accommodate a technologically uncertain and globally focussed economy brought on by the advent of e-commerce, many small businesses are turning toward some form of alliance or network where the locus of the impact of change is interorganisational rather than organisational. Alliances or networks are formed entities that have a defined set of shared values, roles, responsibilities, and governance. Through involvement in such networks, small businesses not only find a ready source of technical and marketing expertise, but the very nature of the network "buffers" the impact of global market turbulence. This would suggest that belonging to a network is an important indicator of successful e-commerce adoption. However, a number of authors (Drakopoulou-Dodd, Jack, & Anderson, 2002; Dennis, 2000; McBer & Company as cited in Dennis, 2000) have found that many small businesses avoid network arrangements. Despite the widespread existence of networks, no research studies to date have formally compared networked and nonnetworked small businesses in relation to e-commerce adoption. This article presents the results of an exploratory study that aims to correct this oversight.

BACKGROUND

According to the European Commission (2003), a *small business* is defined as an organisation that employs less than 50 employees. However, apart from size, small businesses are characterised by a number of other unique features that set them apart from their larger counterparts. These are summarised in Table 1. The presence of these unique features in small businesses creates a challenging environment in which to implement new technologies. In recent years, government organisations around the world have funded projects that assist small businesses in their adoption of internet and e-commerce technologies. At a broad level, *e-commerce* involves the application of Web-based information technologies toward automating business processes, transactions, and work flows (Kalakota & Whinston, 1997). Similarly, the government-funded projects are diverse and range from establishing a simple Internet presence to building virtual business networks (for examples see Dahlstrand, 1999; Damanpour, 2001; Jeffcoate, Chappell, & Feindt, 2002; Papazafeiropoulou, Pouloudi, & Doukidis, 2002; Smith, Boocock, Loan-Clarke, & Whittaker, 2002).

Despite early predictions that small business would benefit from e-commerce adoption, recent studies (Riquelme, 2002; Roberts & Wood, 2002; Barry & Milner, 2002) have shown that it is larger businesses that have more readily adopted this technology. As a result, research studies have begun to examine more closely the processes by which small businesses make their e-commerce adoption decisions. The driving forces or reasons behind e-commerce adoption, in particular, have been investigated in some detail. A summary of the outcomes of these investigations is shown in Table 2. The driving forces can be divided into two types: external and internal. External forces include the pressures and/or influences from parties outside the organisation such as customers, suppliers, and competitors, while internal forces are perceived benefits the small business believes it will achieve through e-commerce adoption.

Unlike previous technological initiatives, e-commerce is a "disruptive" innovation that radically transforms the way a company does business. For small businesses,

Table 1. Features unique to the small business sector

Unique Features of Small Businesses	Related Literature
Small businesses are product oriented, while large business is customer oriented.	Reynolds et al. (1994)
Decision making is intuitive, not based on detailed planning.	Bunker & MacGregor (2000) Reynolds et al. (1994)
Small businesses have a strong owner influence.	Bunker & MacGregor (2000) Reynolds et al. (1994)
Small businesses are riskier than large business and have higher failure rates.	Walker (1975) Brigham & Smith (1967)
Small businesses have difficulties obtaining finances.	Gaskill & Gibbs (1994) Reynolds et al. (1994)
Planning in small businesses is informal and does not entail exhaustive study.	Tetteh & Burn (2001) Miller & Besser (2000) Reynolds et al. (1994)
Small businesses have poor record keeping.	Markland (1974)
Small businesses are more reluctant to take risks.	Walczuch, Van Braven, & Lundgren (2000) Dennis (2000)
Family values intrude in small businesses because they are often family operated.	Dennis (2000) Bunker & MacGregor (2000) Reynolds et al. (1994)
Small businesses have a strong desire for independence and tend to avoid joint business ventures.	Dennis (2000) Reynolds et al. (1994)
Small businesses have centralised management.	Bunker & MacGregor (2000)
Small businesses have a lack of technical staff and IT expertise.	Martin & Matlay (2001) Bunker & MacGregor (2000) Reynolds et al. (1994)
Small businesses do not have control over the environment and face external uncertainty.	Hill & Stewart (2000) Westhead & Storey (1996)
Small businesses make limited use of technology.	Poon & Swatman (1997) MacGregor & Bunker (1996) Abell & Limm (1996)
Small businesses have a limited market share and often operate in niche markets.	Quayle (2002) Hadjimonolis (1999) Lawrence (1997)
Small businesses rely heavily on few customers.	Reynolds et al. (1994)
Small businesses have a narrow product/service range.	Bunker & MacGregor (2000) Reynolds et al. (1994)
Small businesses have practical but narrow skills and experience.	Bunker & MacGregor (2000) Reynolds et al. (1994)

adopting e-commerce has produced both positive and negative effects. Studies by Raymond (2001) and Ritchie and Brindley (2000) found that while e-commerce adoption has eroded trading barriers, this has often come at the price of altering or eliminating commercial relationships and exposing the business to external risks. Lawrence (1997), Tetteh and Burn (2001), and Lee (2001) contend that e-commerce adoption fundamentally alters the internal procedures within small businesses. Indeed, Lee adds that the biggest challenge to a small business is not to find the best e-commerce model, but to change the mindset of the organisation itself. For those that have developed an organisation-wide strategy, these changes can lead to an increase in efficiency in the firm. For those who have not developed this strategy, this can reduce the flexibility of the business (Tetteh & Burn) and often leads to a duplication of the work effort (MacGregor, Bunker, & Waugh, 1998). These are only some of a number of studies that have examined both the benefits and disadvantages of e-commerce adoption in SMEs. A summary of these studies can be found in Table 3.

Recent studies (Keeble, Lawson, Moore, & Wilkinson, 1999; Miles et al., 1999; O'Donnell, Gilmore, Cummins, & Carson, 2001; Overby & Min, 2001) have suggested that small businesses are increasingly turning toward formal and informal alliances or networking arrangements as a mechanism to overcome the disadvantages associated with e-commerce adoption. According to Achrol and Kotler (1999, p. 148), a network is an independent coalition of task- or skill-specialised economic entities (indepen-

Table 2. Summary of e-commerce adoption driving forces in SMEs

External Forces	Related Literature	Internal Forces	Related Literature
Pressure from customers to implement e-commerce	Power & Sohal (2002) Reimenschneider & Mykytyn (2000) PriceWaterhouseCoopers (1999)	*Reduction in costs*	Raisch (2001) Auger & Gallaugher (1997) Abell & Limm (1996)
Pressure from competitors to implement e-commerce	Raisch (2001) Poon & Strom (1997)	*Increase in revenues*	Lee (2001) Phan (2001) Abell & Limm (1996)
Pressure from suppliers to implement e-commerce	Raymond (2001) Reimenschneider & Mykytyn (2000) Lawrence (1997) MacGregor & Bunker (1996)	*Reach new customers/markets*	Power & Sohal (2002) Reimenschneider & Mykytyn (2000) Poon & Swatman (1997) Lawrence (1997)
Improved competitiveness	Raymond (2001) Turban, Lee, King, & Chung (2000) Reimenschneider & Mykytyn (2000)	*Improved customer service*	Power & Sohal (2002) Auger & Gallaugher (1997) Abell & Limm (1996) Senn (1996)
Availability of external support/assistance	Abell & Limm (1996)	*Improved marketing*	Power & Sohal (2002) Reimenschneider & Mykytyn (2000) Poon & Swatman (1997) Lawrence (1997)
		Internal efficiency	Porter (2001)
		Improved control and follow-up	Domke-Damonte & Levsen (2002) Poon & Joseph (2001) Reimenschneider & Mykytyn (2000) Auger & Gallaugher (1997)
		Improved lead time from order to delivery	Power & Sohal (2002) Reimenschneider & Mykytyn (2000) Abell & Limm (1996)
		Stronger relations with business partners	Raymond (2001) Evans & Wurster (1997) Poon & Swatman (1997)

dent firms or autonomous organisational units) that operate without hierarchical control but is embedded by dense lateral connections, mutuality, and reciprocity, in a shared value system that defines "membership" roles and responsibilities.

Viewed as "self-designing" partnerships, networks, as suggested by Eccles and Crane (as cited in Dennis, 2000), are a dynamic arrangement evolving and adjusting to accommodate changes in the business environment. Member organisations in networks have interconnected linkages that allow more efficient movement towards predetermined objectives than would be the case if they operated as a single separate entity. The advent of e-commerce technology has given rise to a new wave of research examining the role of networks in small businesses.

Early studies of small business networks (Gibb, 1993; Ozcan, 1995) concentrated on formal networks. Indeed, Golden and Dollinger (1993) in a study of small manufacturing firms concluded that few small firms were able to function without some form of interorganisational relationship having been established. They added that these interorganisational relationships were associated with successful strategic adaptation by small businesses. Later, many small business networks took a more semi-formal approach. Local or government agencies such as small business associations and chambers of commerce provided an umbrella in the form of advisory services that assisted in legal, financial, training, or technical advice. Individual members operated formally with the umbrella organisation but could interact informally with

Table 3. Summary of e-commerce benefits and disadvantages

E-Commerce Benefits	Related Literature	E-Commerce Disadvantages	Related Literature
Increased sales	Abell & Limm (1996)	*Higher costs*	Stauber (2000)
Access to new customers and markets	Quayle (2002) Ritchie & Brindley (2001) Raymond (2001) Sparkes & Thomas (2001)	*Increased computer/IT maintenance*	MacGregor et al. (1998)
Improved competitiveness	Vescovi (2000)	*Security concerns*	Ritchie & Brindley (2001)
Improved marketing	Sparkes & Thomas (2001) Vescovi (2000) Quayle (2002)	*Reduced flexibility of work*	Lee (2001) MacGregor et al. (1998) Lawrence (1997)
Lower administration costs	Quayle (2002) Poon & Swatman (1997) Abell & Limm (1996)	*Duplication of work processes*	MacGregor et al. (1998)
Lower production costs	Quayle (2002) Poon & Swatman (1997) Abell & Limm (1996)	*Dependence on e-commerce*	Sparkes & Thomas (2001)
Reduced lead time from order to delivery	Quayle (2002) Poon & Swatman (1997) Abell & Limm (1996)	*Deterioration of relations with business partners*	Raymond (2001) Stauber (2000)
Increased internal efficiency	Tetteh & Burn (2001) MacGregor et al. (1998)		
Improved relations with business partners	Poon & Swatman (1997)		
Improved quality of information	Quayle (2002) Poon & Swatman (1997) Abell & Limm (1996)		

fellow members. Recent research (Premaratne, 2001; Rosenfeld, 1996) suggests that informal or social linkages may provide a higher and more stable flow of information and resources in the small business environment. However, a comparison of networked and nonnetworked small businesses in relation to e-commerce specifically has not been carried out to date. As a result, an exploratory study of small businesses was undertaken in Sweden to determine whether organisations belonging to a network differed from those not formally networked in terms of e-commerce adoption driving forces, benefits, and disadvantages. The results of this study are briefly presented in the next section.

E-COMMERCE ADOPTION IN NETWORKED VERSUS NONNETWORKED SMALL BUSINESSES

Based on the background literature described above, a survey was developed and administered by mail to 1,170 randomly chosen small businesses in four regional areas of Sweden: Karlstad, Filipstad, Saffle, and Arvika. Regional areas are of particular interest because of historically high unemployment rates and the loss of labour to metropolitan centres. As a result, the growth of regional businesses has been a top priority for governments worldwide, and small businesses in particular play an important role in achieving sustainable development of regional communities (Keniry et al., 2003). The adoption of technologies such as e-commerce by small businesses contributes towards the achievement of this goal. Therefore, the focus of the study was primarily on e-commerce adoption in regional small businesses.

To profile the small businesses and in line with the study by Donckels and Lambrecht (1997), respondents were asked to respond to the questions shown in Table 4. Respondents were also asked whether their business was part of a network as defined by Achrol and Kotler (1999) and whether they had adopted e-commerce in their business. Those respondents who indicated that they had adopted e-commerce were asked to rate each of the driving forces, benefits, and disadvantages (as shown in Tables 2 and 3 above) across a 5-point Likert scale (1 being very

Table 4. Organisational profile questions

Survey Question	Possible responses
Size of the business	Single owner; 1 to 9 employees; 10 to 19 employees; 20 to 49 employees; 50 to 199 employees; More than 200 employees
Number of years in business	Less than 1 year; 1 to 2 years; 3 to 5 years; 6 to 10 years; 11 to 20 years; More than 20 years
Business sector	Industrial; Service; Retail; Finance; Other (Please Specify)
Market focus	Local; Regional; National; International

Table 5. Profile of survey respondents

NETWORKS	Networked Respondents	Nonnetworked Respondents
Total Respondents	**148**	**191**
E-Commerce Adopters	86	133
Nonadopters	62	58

E-COMMERCE	Adopters	Nonadopters	Missing
Total Respondents	**176**	**125**	**38**

SIZE OF BUSINESS	Single Owner	1-9 Employees	10-19 Employees	20-49 Employees
Total Respondents	**56**	**164**	**49**	**40**
E-Commerce Adopters	31	97	29	31
Nonadopters	25	67	18	9

NO. OF YEARS IN BUSINESS	<1 Year	1-2 Years	3-5 Years	6-10 Years	11-20 Years	> 20 Years
Total Respondents	**5**	**14**	**45**	**61**	**83**	**131**
E-Commerce Adopters	2	8	30	37	53	89
Nonadopters	3	6	15	24	30	42

BUSINESS SECTOR	Industrial	Service	Retail	Finance	Other
Total Respondents	**84**	**118**	**63**	**9**	**54**
E-Commerce Adopters	64	71	40	9	28
Nonadopters	20	47	23	0	26

MARKET FOCUS	Local	Regional	National	International
Total Respondents	**174**	**30**	**96**	**39**
E-Commerce Adopters	102	20	72	25
Nonadopters	72	10	24	14

unimportant in the decision-making process and 5 being extremely important in the decision-making process).

Responses were obtained from 339 small businesses giving a response rate of 29%. Of the 339 respondents, 148 indicated that they belonged to a network and have been termed *networked respondents*, while 191 indicated that they were not part of such an arrangement. These have been termed *nonnetworked respondents*. Table 5 provides a summary profile of the respondents.

A series of chi-square and two-tailed t-tests were performed on the data with interesting results*. Of the 339 respondents, 176 (or 52%) had adopted e-commerce, while 148 (or 43.7%) indicated that they belonged to a network. Table 6 shows a comparison of networked and nonnetworked small businesses and e-commerce adoption. While recent studies (Papazafeiropoulou et al., 2002; Riquelme, 2002; Tetteh & Burn, 2001) have concluded that adoption of e-commerce is best carried out in a networked environment, the data shown in Table 6 does not appear to support this view. It indicates that 64.9% of the

Table 6. Comparison of networked and nonnetworked small businesses and e-commerce adoption

	Networked	Nonnetworked
E-Commerce Adopters	61	115
E-Commerce Nonadopters	63	62
Missing	24	14

Significance Level .004

nonnetworked respondents have adopted e-commerce, while those respondents that were part of a networked structure were, at best, equivocal where e-commerce adoption is concerned.

A number of authors (Hawkins, Winter, & Hunter, 1995; Hyland & Matlay, 1997) have noted that the adoption of e-commerce appears to be significantly linked with the size of the business. Small businesses with fewer than 10 employees (also known as microbusinesses) were less likely to adopt e-commerce than small businesses with more than 10 employees. More recently, Fallon and Moran (2000) and Smith et al. (2002) have suggested that in order to get around this problem, microbusinesses were more likely to engage in some type of formal network. The results of the Swedish study show that nonnetworked small businesses are far more likely to adopt e-commerce than those part of a network. Thus, while microbusinesses might gravitate to some form of network, this is not borne out in their patterns of adopting e-commerce. A number of authors (Donckels & Lambrecht, 1997; Keeble et al. 1999; Schindehutte & Morris, 2001) have offered market focus as an indicator of e-commerce adoption. The Swedish study shows that market focus only appears to be a significant factor for e-commerce adoption with those organisations that operate within a network structure. No significant association was found for those who have stayed outside such an arrangement.

A comparison of means showed that only two driving forces had a significant difference between the two groups of respondents (networked and nonnetworked). These were the availability of external support or assistance and improved control and follow-up. While still below the median point (3), both were rated significantly higher in the nonnetworked group than by their networked counterparts. One possible explanation for this is that sufficient technical support and business know-how exists within the networked small businesses to satisfy its members. This would tend to support the views of Foy (as cited in Dennis, 2000), Keeble et al. (1999), and Overby and Min (2001) who suggest that many small businesses seek out a network to acquire skills that are absent in their own organisations.

In contrast to the views of Achrol and Kotler (1999) and Marchewka and Towell (2000), there were no significant differences in terms of benefits of e-commerce adoption between the networked and nonnetworked respondents. One possible explanation is that in both cases, respondents expected greater benefits from e-commerce adoption than was apparent. However, in a closer examination of the disadvantages of e-commerce adoption, five of these (higher costs, increased computer/IT maintenance, duplication of work processes, reduced flexibility of work, and security) showed a significant difference between the networked and nonnetworked respondents. In all cases, the nonnetworked group found a higher level of disadvantages from these factors than did the networked group. This would tend to support the views of Achrol and Kotler (1999) and Overby and Min (2000) that difficulties dissipate through a network more readily than they might in a single stand-alone unit.

FUTURE TRENDS

The results of the Swedish study raise a number of questions concerning the nature and role of networks in the small business arena and how these impinge upon the decision making and perception of the driving forces, benefits, and disadvantages of adopting e-commerce. The implications are significant because the results raise doubts about the validity and usefulness of government initiatives to promote networks between small businesses as a means of facilitating e-commerce adoption. The study shows that there are no considerable differences between networked and nonnetworked small businesses in terms of e-commerce adoption. Consequently, government initiatives in the networking arena may not necessarily lead to any substantial outcomes.

CONCLUSION

Previous research has suggested that by entering into a networking arrangement with other organisations, a small business (and a regional small business in particular) increases its prospects of survival in a dynamic global marketplace. Furthermore, networked small businesses are able to access technical resources which are not as readily available to their nonnetworked counterparts. This would seem to imply that belonging to a network facilitates the adoption of technologies such as e-commerce. However, the research presented in this article suggests otherwise. A comparison of networked and nonnetworked small businesses in regional areas has shown that the differences between the two groups are minimal where e-commerce driving forces and benefits are concerned. However, there do appear to be noticeable differences in the disadvantages experienced following e-commerce adoption. Networked small businesses are able to absorb the negative impact of e-commerce more readily than nonnetworked ones. The research presented here, however, is by no means an exhaustive study of the relationship between networks and e-commerce adoption. Rather, it is a first step toward a more comprehensive treatment of an issue that has significant implications for the way small businesses relate to their counterparts and embrace emerging technologies.

REFERENCES

Abell, W., & Lim, W. (1996). Business use of the Internet in New Zealand: An exploratory study. *Second Australian World Wide Web Conference*, Southern Cross University Press, Lismore, Australia (pp. 33-39).

Achrol, R. S., & Kotler, P. (1999). Marketing in the network economy. *Journal of Marketing, 63*, 146-163.

APEC Telecommunications Group and Business Facilitation Steering Group(1999). *SME electronic commerce study final report.* PriceWaterhouseCoopers.

Auger, P., & Gallaugher, J. M. (1997). Factors affecting adoption of an Internet-based sales presence for small businesses. *The Information Society, 13*(1), 55-74.

Barry, H., & Milner, B. (2002). SMEs and electronic commerce: A departure from the traditional prioritisation of training? *Journal of European Industrial Training, 25*(7), 316-326.

Brigham, E. F., & Smith, K. V. (1967). The cost of capital to the small firm. *The Engineering Economist, 13*(1), 1-26.

Bunker, D. J., & MacGregor, R. C. (2000). Successful generation of information technology (IT) requirements for small/medium enterprises (SMEs): Cases from regional Australia. *Proceedings of SMEs in a Global Economy*, 72-84.

Dahlstrand, A. L. (1999). Technology-based SMEs in the Goteborg region: Their origin and interaction with universities and large firms. *Regional Studies, 33*(4), 379-389.

Damanpour, F. (2001). E-business e-commerce evolution: Perspective and strategy. *Managerial Finance, 27*(7), 16-33.

Dennis, C. (2000). Networking for marketing advantage. *Management Decision, 38*(4), 287-292.

Domke-Damonte, D., & Levsen, V. B. (2002, Summer). The effect of Internet usage on cooperation and performance in small hotels. *SAM Advanced Management Journal, 67* (3), 31-38.

Donckels, R., & Lambrecht, J. (1997). The network position of small businesses: An explanatory model. *Journal of Small Business Management, 35*(2), 13-28.

Drakopoulou-Dodd, S., Jack, S., & Anderson, A. R. (2002). Scottish entrepreneurial networks in the international context. *International Small Business Journal, 20*(2), 213-219.

European Commission (2003). Definition of crafts and small enterprises. Retrieved February 15, 2004 from *http://europa.eu.int/comm/enterprise/entrepreneurship/craft/definition.htm*

Evans, P. B., & Wurster, T. S. (1997, September-October). Strategy and the new economics of information. *Harvard Business Review,* 75 (5), 70-82.

Fallon, M., & Moran, P. (2000). Information communications technology (ICT) and manufacturing SMEs. *Proceedings of the 2000 Small Business and Enterprise Development Conference,* 100-109.

Gaskill, L. R., & Gibbs, R. M. (1994). Going away to college and wider urban job opportunities take highly educated youth away from rural areas. *Rural Development Perspectives, 10*(3), 35-44.

Gibb, A. (1993). Small business development in central and Eastern Europe: Opportunity for a rethink. *Journal of Business Venturing, 8*, 461-486.

Golden, P. A., & Dollinger, M. (1993, Summer). Cooperative alliances and competitive strategies in small manufacturing firms. *Entrepreneurship Theory and Practice*, 17 (4), 43-56.

Hadjimonolis, A. (1999). Barriers to innovation for SMEs in a small less developed country (Cyprus). *Technovation*, 19(9), 561-570.

Hawkins, P., Winter, J., & Hunter, J. (1995). *Skills for* *graduates in the 21st century* (Rep. commissioned from the Whiteway Research). Cambridge: University of Cambridge, Association of Graduate Recruiters.

Hill, R., & Stewart, J. (2000). Human resource development in small organisations. *Journal of European Industrial Training*, 24(2/3/4), 105-117.

Hyland, T., & Matlay, H. (1997). Small businesses, training needs and VET provisions. *Journal of Education and Work*, *10*(2), 129-139.

Jeffcoate, J., Chappell, C., & Feindt, S. (2002). Best practice in SME adoption of e-commerce. *Benchmarking: An International Journal*, *9*(2), 122-132.

Kalakota, R., & Whinston, A. (1997). *Electronic commerce: A manager's guide*. Reading, MA: Addison-Wesley

Keeble, D., Lawson, C., Moore, B., & Wilkinson, F. (1999). Collective learning processes, networking and "institutional thickness" in the Cambridge region. *Regional Studies, 33*(4), 319-332.

Keniry, J., Blums, A., Notter, E., Radford, E., & Thomson, S. (2003). *Regional Business – A Plan for Action*. Department of Transport and Regional Services. Retrieved from *http://www.rbda.gov.au/ action_plan*, June 10, 2004.

Lawrence, K. L. (1997). Factors inhibiting the utilisation of electronic commerce facilities in Tasmanian small-to-medium sized enterprises. *Proceedings of the Eighth Australasian Conference on Information Systems*, 587-597.

Lee, C. S. (2001). An analytical framework for evaluating e-commerce business models and strategies. *Internet Research: Electronic Network Applications and Policy, 11*(4), 349-359.

MacGregor, R. C., & Bunker, D. J. (1996). The effects of priorities introduced during computer acquisition on continuing success with IT in small business environments. *Proceedings of the Information Resource Management Association International Conference*, 271-277.

MacGregor, R. C., Bunker, D. J., & Waugh, P. (1998). Electronic commerce and small/medium enterprises (SMEs) in Australia: An electronic data interchange (EDI) pilot study. *Proceedings of the 11th International Bled Electronic Commerce Conference*., 284-294.

Marchewka, J. T., & Towell, E. R. (2000). A comparison of structure and strategy in electronic commerce. *Information Technology and People, 13*(2), 137-149.

Markland, R.E. (1974). The role of the computer in small business management. *Journal of Small Business Management*, 12(1), 21-26.

Martin, L. M., & Matlay, H. (2001). "Blanket" approaches to promoting ICT in small firms: Some lessons from the DTI ladder adoption model in the UK. *Internet Research: Electronic Networking Applications and Policy, 11*(5), 399-410.

Miles, G., Preece, S., & Baetz, M. C. (1999). Dangers of dependence: The impact of strategic alliance use by small technology based firms. *Journal of Small Business Management*, 37(2), 20-29.

Miller, N. L., & Besser, T. L. (2000). The importance of community values in small business strategy formation: Evidence from rural Iowa. *Journal of Small Business Management*, *38*(1), 68-85.

O'Donnell, A., Gilmore, A., Cummins, D., & Carson, D. (2001). The network construct in entrepreneurship research: A review and critique. *Management Decision, 39*(9), 749-760.

Overby, J. W., & Min, S. (2001). International supply chain management in an Internet environment: A network-oriented approach to internationalisation. *International Marketing Review*, *18*(4), 392-420.

Ozcan, G. (1995). Small business networks and local ties in Turkey. *Entrepreneurship and Regional Development, 7*, 265-282.

Papazafeiropoulou, A., Pouloudi, A., & Doukidis, G. (2002). A framework for best practices in electronic commerce awareness creation. *Business Process Management Journal, 8*(3), 233-244.

Phan, D.D. (2001). E-business management: A business-to-business case study. *Information Systems Management*, Fall, 61-69.

Poon, S., & Joseph, M. (2001). A preliminary study of product nature and electronic commerce. *Marketing Intelligence & Planning, 19*(7), 493-499.

Poon, S., & Strom, J. (1997). *Small business use of the Internet: Some realities.* Proceedings of the Association for Information Systems Americas Conference, Indianapolis, IN.

Poon, S., & Swatman, P. (1997). The Internet for small businesses: An enabling infrastructure. *Proceedings of the Fifth Internet Society Conference*, 221-231.

Porter, M. (2001, March). Strategy and the Internet. *Harvard Business Review*, 79 (3), 63-78.

Power D.J., & Sohal A.S. (2002). Best practice in implementation and usage of electronic commerce: A comparative study of 10 Australian companies. *Benchmarking - Special Issue on "Electronic Commerce: A Best Practice Perspective*, 9(2), 190-208

Premaratne, S. P. (2001). Networks, resources and small business growth: The experience in Sri Lanka. *Journal of Small Business Management*, *39*(4), 363-371.

Quayle, M. (2002). E-commerce: The challenge for UK SMEs in the 21st Century. *International Journal of Operations and Production Management*, *22*(10), 1148-1161.

Raisch, W.D. (2001). *The eMarketplace: Strategies for Success in B2B*, New York: McGraw Hill.

Raymond, L. (2001). Determinants of Web site implementation in small business. *Internet Research: Electronic Network Applications and Policy*, *11*(5), 411-422.

Reimenschneider, C. K., & Mykytyn, P. P., Jr. (2000). What small business executives have learned about managing information technology. *Information & Management*, *37*, 257-267.

Reynolds, W., Savage, W., & Williams, A. (1994). *Your own business: A practical guide to success*. Melbourne: Nelson ITP.

Riquelme, H. (2002). Commercial Internet adoption in China: Comparing the experience of small, medium and large business. *Internet Research: Electronic Networking Applications and Policy*, *12*(3), 276-286.

Ritchie, R., & Brindley, C. (2000). Disintermediation, disintegration and risk in the SME global supply chain. *Management Decision*, *38*(8), 575-583.

Roberts, M., & Wood, M. (2002). The strategic use of computerised information systems by a micro enterprise. *Logistics Information Management*, 15(2), 115-125.

Rosenfeld, S. (1996). Does cooperation enhance competitiveness? Assessing the impacts of inter-firm collaboration. *Research Policy*, *25*(2), 247-263.

Schindehutte, M., & Morris, M.H. (2001) Pricing as entrepreneurial behavior. *Business Horizons*, 44(4), 41-48.

Senn, J. A. (1996, Summer). Capitalisation on electronic commerce. *Information Systems Management*, 13(3), 15-24.

Smith, A. J., Boocock, G., Loan-Clarke, J., & Whittaker, J. (2002). IIP and SMEs: Awareness, benefits and barriers. *Personnel Review*, *31*(1), 62-85.

Sparkes, A., & Thomas, B. (2001). The use of the Internet as a critical success factor for the marketing of Welsh agri-food SMEs in the 21st century. *British Food Journal*, 103(4), 331-347.

Stauber, A. (2000). *A survey of the incorporation of electronic commerce in Tasmanian small and medium sized enterprises*. Tasmanian Electronic Commerce Centre, Launceston, Australia.

Tetteh, E., & Burn, J. (2001). Global strategies for SME-business: Applying the SMALL framework. *Logistics Information Management*, *14*(1/2), 171-180.

Turban, E., Lee, J., King, D., & Chung, H. (2000). *Electronic commerce: A managerial perspective*. Englewood Cliffs, NJ: Prentice Hall

Vescovi, T. (2000). Internet communication: The Italian SME case. *Corporate Communications: An International Journal*, *5*(2), 107-112.

Walczuch, R., Van Braven, G., & Lundgren, H. (2000). Internet adoption barriers for small firms in The Netherlands. *European Management Journal*, *18*(5), 561-572.

Walker, E.W. (1975). Investment and capital structure decision making in small business. In Walker, E.W. (Ed) *The dynamic small firm: Selected readings*. Texas: Austin Press.

Westhead, P., & Storey, D. J. (1996). Management training and small firm performance: Why is the link so weak? *International Small Business Journal*, *14*(4), 13-24.

KEY TERMS

E-Commerce/Electronic Commerce: The application of Web-based information technologies toward automating business processes, transactions, and work flows (Kalakota & Whinston, 1997).

E-Commerce Driving Forces: Conditions and/or expectations that facilitate e-commerce adoption.

E-Commerce Benefits: Tangible and intangible business advantages achieved by adopting e-commerce.

E-Commerce Disadvantages: Difficulties or problems experienced by a business following e-commerce adoption.

Microbusiness: An organisation that employs less than 10 employees.

Network: An independently formed entity that has a defined set of shared values, roles, responsibilities, and governance, and is characterised by lateral connections, mutuality, and reciprocity between the network members.

Regional Area: A geographical area located outside major metropolitan centres.

ENDNOTES

* Owing to space restrictions, it is not possible to provide statistical tables here, however they are available.

Neural Networks for Automobile Insurance Pricing

Ai Cheo Yeo
Monash University, Australia

INTRODUCTION

In highly competitive industries, customer retention has received much attention. Customer retention is an important issue, as loyal customers tend to produce greater cash flow and profits, are less sensitive to price, bring along new customers and do not require any acquisition or start-up costs.

BACKGROUND

Various techniques have been used to analyse customer retention. Eiben, Koudijs and Slisser (1998) applied genetic programming, rough set analysis, Chi-square Automatic Interaction Detection (CHAID) and logistic regression analysis to the problem of customer retention modelling, using a database of a financial company. Models created by these techniques were used to gain insights into factors influencing customer behaviour and to make predictions on customers ending their relationship with the company. Kowalczyk and Slisser (1997) used rough sets to identify key factors that influence customer retention of a mutual fund investment company. Ng, Lui and Kwah (1998) integrated various techniques such as decision-tree induction, deviation analysis and multiple concept-level association rules to form an intuitive approach to gauging customers' loyalty and predicting their likelihood of defection.

Mozer and his co-researchers (2000) explored techniques from statistical machine learning to predict churn and based on these predictions to determine what incentives should be offered to subscribers of wireless telecommunications to improve retention and maximise profitability of the carrier. The techniques included logit regression, decision trees, neural networks and boosting. Besides Mozer and his co-researchers, others have also applied neural networks to churn prediction problems. Behara and Lemmink (1994) used the neural network approach to evaluate the impact of quality improvements on a customer's decision to remain loyal to an automanufacturer's dealership. Wray and Bejou (1994) examined the factors that seem to be important in explaining customer loyalty. They found that neural networks have a better predictive power than the conventional analytic techniques such as multiple regression. Smith, Willis and Brooks (2000) also found that neural networks provided the best results for classifying insurance policy holders as likely to renew or terminate their policies compared to regression and decision tree modelling.

PREDICTING RETENTION RATES

We have also used neural networks to learn to distinguish insurance policy holders who are likely to terminate their policies from those who are likely to renew in order to predict the retention rate prior to price sensitivity analysis. Policy holders of an Australian motor insurance company are classified into 30 risk groups based on their demographic and policy information using k-means clustering (Yeo, Smith, Willis & Brooks, 2001, 2003). Neural networks are then used to model the effect of premium price change on whether a policy holder will renew or terminate his or her policy. A multilayered feedforward neural network was constructed for each of the clusters with 25 inputs and 1 output (whether the policy holder renews or terminates the contract).

Several experiments were carried out on a few clusters to determine the most appropriate number of hidden neurons and the activation function. Twenty hidden neurons and the hyperbolic tangent activation function were used for the neural networks for all the clusters. A uniform approach is preferred to enable the straightforward application of the methodology to all clusters, without the need for extensive experimentation by the company in the future. Input variables that were skewed were log transformed.

Some of the issues we encountered in using neural networks to determine the effect of premium price change on whether a policy holder will renew or terminate his or her policy were:

- Determining the threshold for classifying policy holders into those who terminate and those who renew

- Generating more homogenous models
- Clusters that had too few policy holders to train the neural networks

Determining Threshold

The neural network produces output between zero and one, which is the probability that that a policy holder will terminate his or her policy. Figure 1 shows the probability of termination of Cluster 11. A threshold value is used to decide how to categorise the output data. For example a threshold of 0.5 means that if the probability of termination is more than 0.5, then the policy will be classified as terminated. Usually the decision threshold is chosen to maximise the classification accuracy. However, in our case we are more concerned with achieving a predicted termination rate that is equal to the actual termination rate. This is because we are more concerned with the performance of the portfolio (balancing market share with profitability) rather than whether an individual will renew or terminate his or her policy. The actual termination rate for cluster 11 is 14.7%. To obtain a predicted termination rate of 14.7%, the threshold was set at 0.204 (see Figure 1). The confusion matrix for a threshold of 0.204 is shown in Table 1. The overall classification accuracy is 85.3%.

Generating more homogeneous models

The confusion matrix provides the prediction accuracy of the whole cluster. It does not tell us how a given percentage change in premium will impact termination rates. To determine how well the neural networks were able to predict termination rates for varying amounts of premium changes, the clusters were then divided into various bands of premium as follows: decrease in premiums of less than 22.5%, premium decrease between 17.5% and 22.5%, premium decrease between 12.5% and 17.5% and so on.

Figure 1. Determining the threshold value of the neural network output

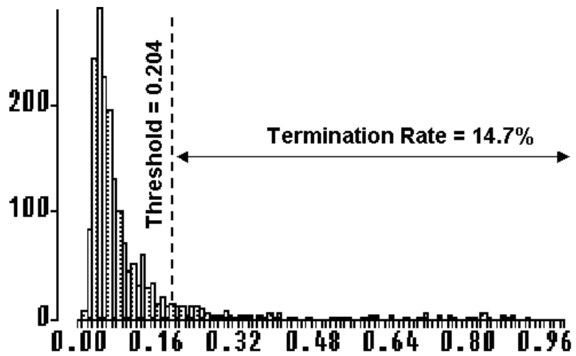


Table 1. Confusion matrix for cluster 11 with decision threshold = 0.204

	Classified as		
Actual	Terminated	Renewed	Total
Terminated	841 (50.0%)	842 (50.0%)	1,683
Renewed	845 (8.6%)	8,935 (91.4%)	9,780
Total	1,686	9,777	11,463
Overall Accuracy			85.3%

The predicted termination rates were then compared to the actual termination rates. For all the clusters the prediction accuracy of the neural networks starts to deteriorate when premium increases are between 10% and 20%. Figure 2 shows the actual and predicted termination rates for one of the clusters (Cluster 24).

In order to improve the prediction accuracy, the cluster was then split at the point when prediction accuracy starts to deteriorate. This is to isolate those policy holders with a significant increase in premium. It is believed that these policy holders behave differently due to a greater number of these policy holders who have upgraded their vehicles. Two separate neural networks were trained for each cluster. The prediction accuracy improved significantly with two neural networks as can be seen from Figure 3. The average absolute deviation decreased from 10.3% to 2.4%.

Combining Small Clusters

Some of the smaller clusters had too few policy holders to train the neural networks. We grouped the small clusters that had fewer than 7,000 policies. The criterion for grouping was similarity in risk. Risk in turn is measured by the amount of claims. Therefore the clusters were grouped according to similarity in claim cost. The maximum difference in average claim cost per policy was no more than $50. For the combined clusters, prediction ability is also improved by having two neural networks instead of one for each cluster.

PRICE SENSITIVITY ANALYSIS

Having trained neural networks for all the clusters, sensitivity analysis was then performed on the neural networks to determine the effect of premium changes on termination rates for each cluster. There are several ways of performing the sensitivity analysis:

One approach is based on systematic variation of variables (SVV). To determine the impact that a particular input variable has on the output, we need to hold all the other inputs to some fixed value and vary only the input

N

Figure 2. Prediction accuracy for one neural network model of cluster 24

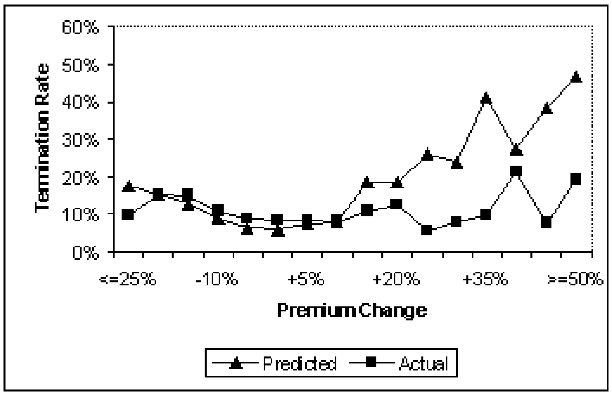
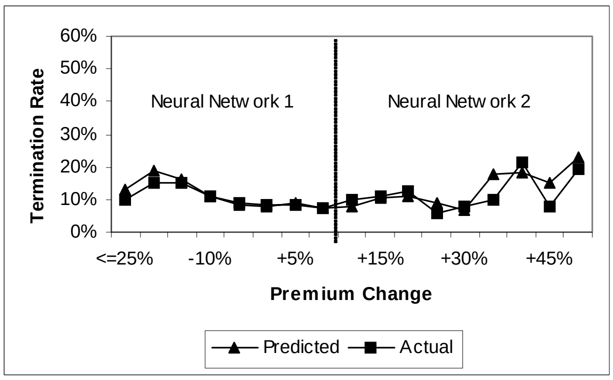


Figure 3. Prediction accuracy for two networks model of cluster 24



of interest while we monitor the change in outputs (Anderson, Aberg & Jacobsson, 2000; Bigus, 1996).

A more automated approach is to keep track of the error terms computed during the backpropagation step. By computing the error all the way back to the input layer, we have a measure of the degree to which each input contributes to the output error (Bigus, 1996).

The third approach is based on sequential zeroing of weights (SZW) of the connection between the input variables and the first hidden layer of the neural network (Anderson et al., 2000).

In their research, Anderson and his co-researchers (2000) found that neural networks are suitable not only for function approximation of nonlinear relationship but are also able to represent to a high degree the nature of input variables. The information generated about the variables using the SVV and SZW methods can serve as a guide to the interpretation of influence, contribution and selection of variables. We performed price sensitivity using the SVV approach by varying the premium information and holding all other inputs constant.

Table 2. Price sensitivity analysis – effect on termination rate (Cluster 24)

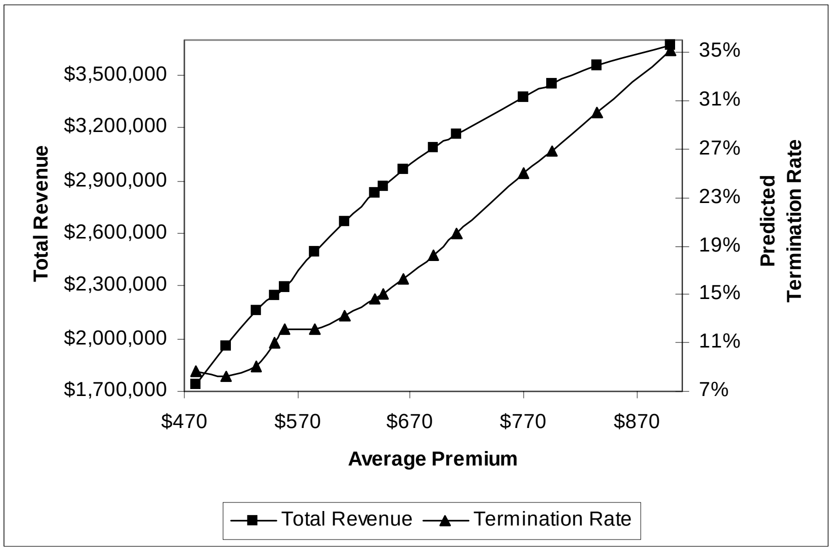
Scored Against	Average Change in Premium	Average Premium Amount ($)	Termination Rate
Neural Network 1	-8.3%	481	8.6%
	-3.3%	507	8.3%
	1.7%	533	9.1%
	6.7%	559	12.2%
Neural Network 2	11.7%	585	12.1%
	16.7%	612	13.2%
	21.7%	638	14.6%
	26.7%	664	16.3%
	31.7%	690	18.2%
	51.7%	795	26.9%
	71.7%	900	35.1%

Separate data sets were created from each "half" cluster with all variables remaining unchanged except the new premium and related variables (change in premium, percentage change in premium and ratio of new premium to new sum insured). For example, cluster 24 was split into two "halves"; policy holders with premium decreases or increases of less than 10% and policy holders with premium increases of more than 10%. Data sets with varying percentage changes in premium were created and scored against the trained neural networks to determine the predicted termination rates. Results of price sensitivity analysis for Cluster 24 are shown in Table 2 and Figure 4 as an example. They show that as premiums increase the termination rates also increase.

CONCLUSIONS AND FUTURE RESEARCH

We randomly selected a few clusters to carry out experiments to determine the appropriate neural network architecture for the prediction of retention rates. More experiments could be performed to find an appropriate neural network architecture for each of the clusters so that prediction accuracy could be improved. However, the insurance industry is unlikely to implement the proposed approach if it is difficult to implement. We have therefore adopted a straightforward approach that requires minimal fine tuning for specific data. While this research has been focused on premium pricing within the insurance industry, the methodology developed is quite general. In fact, the approach can be applied to any industry concerned with setting prices for products in competitive environments. This includes sectors such as retail and telecommunications.

Figure 4. Price sensitivity analysis - effect on termination rate and revenue (Cluster 24)



REFERENCES

Anderson, F.O., Aberg, M., & Jacobsson, S.P. (2000). Algorithmic approaches for studies of variable influence, contribution and selection in neural networks. *Chemometrics and Intelligent Laboratory Systems, 51*, 61-72.

Behara, R.S., & Lemmink, J. (1994). Modelling the impact of service quality on customer loyalty and retention: A neural network approach. *1994 Proceedings Decision Sciences Institute, 3*, 1883-1885.

Bigus, J.P. (1996). *Data mining with neural networks: Solving business problems—from application development to decision support.* New York: McGraw-Hill.

Eiben, A., Koudijs, A., & Slisser, F. (1998). *Genetic modelling of customer retention.* Paper presented at the Genetic Programming First European Workshop, EruoGP'98, Paris, France.

Kowalczyk, W., & Slisser, F. (1997). *Modelling customer retention with rough data sets.* Paper presented at the Principles of Data Mining and Knowledge Discovery. First European Symposium, PKDD '97, Trondheim, Norway.

Mozer, M.C., Wolniewicz, R., Grimes, D.B., Johnson, E., & Kaushansky, H. (2000). Predicting subscriber dissatisfaction and improving retention in the wireless telecommunication. *IEEE Transactions on Neural Networks, 11*(3), 690-696.

Ng, K.S., Lui, H., & Kwah, H.B. (1998). *A data mining application: Customer retention at the Port of Singapore Authority (PSA).* Paper presented at the 1998 ACM SIGMOD International Conference on Management of Data, Seattle, WA, USA.

Smith, K.A., Willis, R.J., & Brooks, M. (2000). An analysis of customer retention and insurance claim patterns using data mining: A case study. *Journal of the Operational Research Society, 51*(5), 532-541.

Wray, B., & Bejou, D. (1994). *An application of artificial neural networks in marketing: Determinants of customer loyalty in buyer-seller relationships.* Paper presented at the Proceedings of Decision Sciences Institute 1994 Annual Meeting, Honolulu, HI.

Yeo, A., Smith, K., Willis, R., & Brooks, M. (2001). Clustering technique for risk classification and prediction of claim costs in the automobile insurance industry. *International Journal of Intelligent Systems in Accounting, Finance and Management, 10*(1), 39-50.

Yeo, A.C., Smith, K.A., Willis, R.J., & Brooks, M. (2003). A comparison of soft computing and traditional approaches for risk classification and claim cost prediction in the automobile insurance industry. In V. Kreinovich (Ed.), *Soft computing in measurement and information acquisition* (pp. 249-261). Heidelberg: Physica-Verlag.

KEY TERMS

Activation Function: Transforms the net input of a neural network into an output signal, which is transmitted to other neurons.

Association Rules: Predict the occurrence of an event based on the occurrences of another event.

Backpropagation: Method for computing the error gradient for a feedforward neural network.

Boosting: Generates multiple models or classifiers (for prediction or classification), and to derive weights to combine the predictions from those models into a single prediction or predicted classification.

Chi-Square Automatic Interaction Detection (CHAID): A decision tree technique used for classification of a data set. CHAID provides a set of rules that can be applied to a new (unclassified) data set to predict which records will have a given outcome. CHAID segments a data set by using chi square tests to create multi-way splits.

Confusion Matrix: Contains information about actual and predicted classifications done by a classification system.

Decision Trees: Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset.

Deviation Analysis: Locates and analyses deviations from normal statistical behavior.

Genetic Programming: Search method inspired by natural selection. The basic idea is to evolve a population of "programs" candidates to the solution of a specific problem.

K-means Clustering: An algorithm that performs disjoint cluster analysis on the basis of Euclidean distances computed from variables and randomly generated initial seeds.

Logistic/Logit Regression: A technique for making predictions when the dependent variable is a dichotomy, and the independent variables are continuous and/or discrete.

Multi-Layered Feedforward Network: A layered neural network in which each layer only receives inputs from previous layers.

Multiple Concept-Level Association Rules: Extend association rules from single level to multiple levels. Database contents are associated together to the concepts, creating different abstraction levels.

Multiple Regression: A statistical technique that predicts values of one variable on the basis of two or more other variables.

Neural Network: Non-linear predictive models that learn through training and resemble biological neural networks in structure.

Neural Network Architecture: A description of the number of layers in a neural network, each layer's transfer function, the number of neurons per layer, and the connections between layers

Neuron: The basic processing element of a neural network.

Regression Analysis: A statistical technique used to find relationships between variables for the purpose of predicting future variables.

Rough Sets: Rough sets are mathematical algorithms that interpret uncertain, vague, or imprecise information.

Sequential Zeroing of Weights: A sensitivity analysis method that involves sequential zeroing of weights of the connection between the input variables and the first hidden layer of the neural network.

Statistical Machine Learning: An approach to machine intelligence that is based on statistical modeling of data.

Systematic Variation of Variables: A sensitivity analysis method whereby the input of interest is varied while holding all other inputs to some fixed value to determine the impact that particular variable has on the output.

Weights: Strength of a connection between two neurons in a neural network.

Neural Networks for Retail Sales Forecasting

G. Peter Zhang
Georgia State University, USA

INTRODUCTION

Forecasting of the future demand is central to the planning and operation of retail business at both macro and micro levels. At the organizational level, forecasts of sales are needed as the essential inputs to many decision activities in various functional areas such as marketing, sales, production/purchasing, as well as finance and accounting (Mentzer & Bienstock, 1998). Sales forecasts also provide basis for regional and national distribution and replenishment plans. The importance of accurate sales forecasts to efficient inventory management has long been recognized. In addition, accurate forecasts of retail sales can help improve retail supply chain operation, especially for larger retailers who have a significant market share. For profitable retail operations, accurate demand forecasting is crucial in organizing and planning purchasing, production, transportation, labor force, as well as after sales services.

Barksdale and Hilliard (1975) examined the relationship between retail stocks and sales at the aggregate level and found that successful inventory management depends to a large extent on the accurate forecasting of retail sales. Thall (1992) and Agrawal and Schorling (1996) also pointed out that accurate demand forecasting plays a critical role in profitable retail operations and poor forecasts would result in too many or too few stocks that directly affect revenue and competitive position of the retail business. The importance of accurate demand forecasts in successful supply chain operations and coordination has been recognized by many researchers (Chopra & Meindl, 2001; Lee et al., 1997).

Retail sales often exhibit both seasonal variations and trends. Historically, modeling and forecasting seasonal data is one of the major research efforts and many theoretical and heuristic methods have been developed in the last several decades. Different approaches have been proposed but none of them has reached consensus among researchers and practitioners. Until now, the debate has still not abated in terms of what is the best approach to handle the seasonality.

On the other hand, it is often not clear how to best model the trend pattern in a time series. In the popular Box-Jenkins approach to time series modeling, differencing is used to achieve stationarity in the mean. However, Pierce (1977) and Nelson and Plosser (1982) argue that differencing is not always an appropriate way to handle trend, and linear detrending may be more appropriate. Depending on the nature of the nonstationarity, a time series may be modeled in different ways. For example, a linear or polynomial time trend model can be used if the time series has a deterministic trend. However, if a time series exhibits a stochastic trend, the random walk model and its variations may be more appropriate.

In addition to controversial issues around the ways to model seasonal and trend time series, one of the major limitations of many traditional models is that they are essentially linear methods. In order to use them, users must specify the model form without the necessary genuine knowledge about the complex relationship in the data. This may be the reason for the mixed findings reported in the literature regarding the best way to model and forecast trend and seasonal time series.

One nonlinear model that has recently received extensive attention is the neural network model (NN). The popularity of the neural network model can be attributed to their unique capability to simulate a wide variety of underlying nonlinear behaviors. Indeed, research has provided theoretical underpinning of neural network's universal approximation ability. In addition, few assumptions about the model form are needed in applying the NN technique. Rather, the model is adaptively formed with the real data. This flexible data-driven modeling property has made NNs an attractive tool for many forecasting tasks, as data are often abundant, while the underlying data generating process is hardly known.

In this article, we provide an overview on how to effectively model and forecast consumer retail sales using neural network models. Although there are many studies on general neural network forecasting, few are specifically focused on trending or seasonal time series. In addition, controversial results have been reported in the literature. Therefore it is necessary to have a good summary of what has been done in this area and more importantly to give guidelines that can be useful for forecasting practitioners.

It is important to note that the focus of this article is on time series forecasting methods. For other types of forecasting methods used in retail sales, readers are referred to Green (1986), Smith et al. (1994), and Dominique (1998).

BACKGROUND

Neural networks are computing models for information processing. They are very useful for identifying the functional relationship or pattern in the retails sales and other time series data. The most popularly used neural network model in practice for retail sales is the feedforward multi-layer network. It is composed of several layers of basic processing units called neurons or nodes. For an in-depth coverage of NN models, readers are referred to Smith (1993) and Bishop (1995). A comprehensive review of the NNs for forecasting is given by Zhang et al. (1998).

Before it can be used for forecasting, the NN model must be built first. Neural network model building (training) involves determining the order of the network (the architecture) as well as the parameters (weights) of the model. NN training typically requires that the in-sample data be split into a training set and a validation set. The training set is used to estimate the parameters of some candidate models, among which the one that performs the best on the validation set is selected. The out-of-sample observations can be used to further test the performance of the selected model to simulate the real forecasting situations.

The standard three-layer feedforward NN can be used for time series forecasting in general and retail sales in particular. For one-step-ahead forecasting, only one output node is needed. For multiple-step forecasting, more output nodes should be employed. For time series forecasting, the most important factor in neural networks modeling is the number of input nodes, which corresponds to the number of past observations significantly auto-correlated with the future forecasts. In a seasonal time series such as the retail sales series, it is reasonable to expect that a forecasting model should capture the seasonal autocorrelation that spans at least one or two seasonal periods of, say, 12 or 24 for monthly series.

Therefore in modeling seasonal behavior, it is critical to include in the input nodes the observations separated by multiples of seasonal period. For example, for a quarterly seasonal time series, observations that are four quarters away are usually highly correlated. Although theoretically, the number of seasonal lagged observations that have autocorrelation with the future value can be high, it is fairly small in most practical situations, as empirical studies often suggest that the seasonal autoregressive order be one or at most two (Box & Jenkins, 1976).

There are many other parameters and issues that need to be carefully considered and determined in neural network model building for retail and other time series. These include data preparation, data division and sample size, network architecture in terms of number of hidden and input nodes, training algorithm, model evaluation criteria and so forth. Practical guidelines can be found in many references in the literature including Adya and Collopy (1998), Kaastra and Boyd (1996), and Zhang et al. (1998).

ISSUES AND LITERATURE REVIEW

Modeling seasonal and trend time series has been one of the main research endeavors for decades. In the early 1920s, the decomposition model along with seasonal adjustment was the major research focus due to Person's (1919) work on decomposing a seasonal time series. Different seasonal adjustment methods have been proposed and the most significant and popular one is the Census X-11 method developed by the Bureau of the Census in 1950s and 60s, which has evolved into the current X-12-ARIMA program. Because of the ad hoc nature of the seasonal adjustment methods, several model-based procedures have been developed. Among them, the work by Box and Jenkins (1976) on the seasonal ARIMA model has had a major impact on the practical applications to seasonal time series modeling. This model has performed well in many real-world applications and is still one of the most widely used seasonal forecasting methods. More recently, neural networks have been widely used as a powerful alternative to traditional time series modeling.

In neural network forecasting, little research has been done focusing on seasonal and trend time series modeling and forecasting. In fact, how to effectively model seasonal time series is a challenging task not only for the newly developed neural networks, but also for the traditional models. One popular traditional approach to dealing with seasonal data is to remove the seasonal component first before other components are estimated. Many practitioners in various forecasting applications have satisfactorily adopted this practice of seasonal adjustment. However, several recent studies have raised doubt about its appropriateness in handling seasonality. Seasonal adjustment has been found to lead into undesirable nonlinear properties, severely distorted data, and inferior forecast performance (Ghysels et al., 1996; Plosser, 1979). De Gooijer and Franses (1997) pointed out that "although seasonally adjusted data may sometimes be useful, it is typically recommended to use seasonally unadjusted data". On the other hand, mixed findings have also been reported in the limited neural network literature on seasonal forecasting. For example, Sharda and Patil (1992) found that, after examining 88 seasonal time series, NNs were able to model seasonality directly and pre-deseasonalization is not necessary. Alon et al. (2001) also found that NNs are able to "capture the dynamic nonlinear trend and seasonal patterns, as well as the interactions between them". However, Farway and Chatfield (1995) and Nelson et al. (1999), among others, found just the opposite. Their findings suggest that neural networks are not able to directly model seasonality and pre-deseasonalization of the data is necessary to improve forecasting performance.

Chu and Zhang (2003) compared the accuracy of various linear and neural network models for forecasting aggregate retail sales. Using multiple cross-validation samples, they found that the nonlinear models outperform their linear counterparts in out-of-sample forecasting, and prior seasonal adjustment of the data can significantly improve forecasting performance of the neural network model. The overall best model is the neural network built on deseasonalized time series data. While seasonal dummy variables can be useful in developing effective regression models for predicting retail sales, the performance of dummy regression models may not be robust. In addition, they found that trigonometric models are not useful in aggregate retail sales forecasting.

Zhang and Qi (2003) examined the issue of how to use neural networks more effectively in modeling and forecasting a seasonal time series with a trend component. The specific research questions they addressed are (1) whether neural networks are able to directly model different components of a seasonal and trend time series and (2) whether data pre-processing is necessary or beneficial. Instead of focusing solely on the seasonal component alone as in previous studies (e.g., Nelson et al., 1999), they took a systematic approach on the data preprocessing issue to study the relevance of detrending and deseasonalization. Using a large number of simulated and real time series, they evaluated the effect of different data preprocessing strategies on neural network forecasting performance. Results clearly show that without data pre-processing, neural networks are not able to effectively model the trend and seasonality patterns in the data, and either detrending or deseasonalization can greatly improve neural network modeling and forecasting accuracy. A combined approach of detrending and deseasonalization is found to be the most effective data preprocessing that can yield the best forecasting result. These findings are further examined and confirmed in Zhang and Qi (2002) with two monthly consumer retail sales—computer and computer software sales, and grocery store sales.

Therefore, it is recommended that data preprocessing is performed before building neural network models for retail sales data that contain seasonal and trend components. If the time series contains only the seasonal variation, deseasonalization should be the best choice. However, if both trend and seasonal fluctuations are present, a combined approach of detrending and deseasonalization should be used.

CONCLUSION

Neural networks become an important tool for retail sales forecasting. Due to the complex nature in model building, it is necessary to preprocess the data first. Several recent studies strongly suggest that for retail sales that contain both trend and seasonal variations, it is not appropriate to directly model sales time series with neural networks. Rather, both seasonal effect and trend movement can have significant effect in accurately modeling retail sales. In order to build the best neural network model, forecasters should use a combined approach of deseasonalization and detrending to remove the seasonal and trend factors first. Another benefit with both seasonality and trend removed is that more parsimonious neural networks can be constructed.

REFERENCES

Adya, M., & Collopy, F. (1998). How effective are neural networks at forecasting and prediction? A review and evaluation. *Journal of Forecasting, 17,* 481-495.

Agrawal, D., & Schorling, C. (1996). Market share forecasting: An empirical comparison of artificial neural networks and multinomial logit model. *Journal of Retailing, 72* (4), 383-407.

Alon, I., Qi, M., & Sadowski, R.J. (2001). Forecasting aggregate retail sales: A comparison of artificial neural networks and traditional methods. *Journal of Retailing and Consumer Services*, *8*(3), 147-156.

Barksdale, H.C., & Hilliard, J.E. (1975). A cross-spectral analysis of retail inventories and sales. *Journal of Business, 48*(3), 365-382.

Bishop, M. (1995). *Neural networks for pattern recognition*. Oxford: Oxford University Press.

Box, G.E.P., & Jenkins, G.M. (1976). *Time series analysis: Forecasting, and control.* San Francisco: Holden Day.

Chopra, S., & Meindl, P. (2001). *Supply chain management: Strategy, planning, and operation.* New Jersey: Prentice Hall.

Chu, C., & Zhang, G.P. (2003). A comparative study of linear and nonlinear models for aggregate retail sales forecasting. *International Journal of Production Economics,* forthcoming.

De Gooijer, J.G., & Franses, P.H. (1997). Forecasting and seasonality. *International Journal of Forecasting, 13*, 303-305.

Dominique, H.M. (1998). Order forecasts, retail sales, and the marketing mix for consumer durables. *Journal of Forecasting, 17,* 327-348.

Farway, J., & Chatfield, C. (1995). Time series forecasting with neural networks: A comparative study using the airline data. *Applied Statistics, 47*, 231-250.

Franses, P.H., & Draisma, G. (1997). Recognizing changing seasonal patterns using artificial neural networks. *Journal of Econometrics, 81*, 273-280.

Ghysels, E., Granger, C.W.J., & Siklos, P.L. (1996). Is seasonal adjustment a linear or nonlinear data filtering process? *Journal of Business and Economics Statistics, 14,* 374-386.

Green, H.L. (1986). Retail sales forecasting systems. *Journal of Retailing, 62,* 227-230.

Kaastra, I., & Boyd, M. (1996). Designing a neural network for forecasting financial and economic time series. *Neurocomputing, 10,* 215-236.

Lee, H.L., Padmanabhan, V., & Whang, S. (1997, Spring). The bullwhip effect in supply chains. *Sloan Management Review,* 93-102.

Mentzer, J.T., & Bienstock, C.C. (1998). *Sales forecasting management*. Thousand Oaks, CA: Sage.

Nelson, C.R., & Plosser, C. I., (1982). Trends and random walks in macroeconomic time series: some evidence and implications. *Journal of Monetary Economics, 10*, 139-162.

Nelson, M., Hill, T., Remus, T., & O'Connor, M. (1999). Time series forecasting using NNs: Should the data be deseasonalized first? *Journal of Forecasting, 18,* 359-367.

Persons, W.M. (1919). Indices of business conditions. *Review of Economics and Statistics, 1,* 5-107.

Pierce, D.A. (1977). Relationships - and the lack of thereof - between economic time series, with special reference to money and interest rates. *Journal of the American Statistical Association, 72,* 11-26.

Plosser, C.I. (1979). Short-term forecasting and seasonal adjustment. *Journal of the American Statistical Association, 74*, 15-24.

Sharda, R., & Patil, R.B. (1992). Connectionist approach to time series prediction: An empirical test. *Journal of Intelligent Manufacturing, 3,* 317-323.

Smith, M. (1993). *Neural networks for statistical modeling*. New York: Van Nostrand Reinhold.

Smith, S.A., McIntyre, S.H., & Dale, A.D. (1994). Two-stage sales forecasting procedure using discounted least squares. *Journal of Marketing Research, 31,* 44-56.

Thall, N. (1992). Neural forecasts: A retail sales booster. *Discount Merchandiser, 23* (10), 41-42.

Zhang, G., Patuwo, B.E., & Hu, M.Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting, 14,* 35-62.

Zhang, G.P., & Qi, M. (2002). Predicting consumer retail sales using neural networks. In K. Smith & J. Gupta (Eds.), *Neural networks in business: Techniques and applications*. Hershey, PA: Idea Group Publishing.

Zhang, G.P., & Qi, M. (2003). *Neural network forecasting of seasonal and trend time series*. Technical report. Department of Management, Georgia State University.

KEY TERMS

Box-Jenkins Approach: A very versatile linear approach that can model trend, seasonal, and other behaviors by using moving averages, autoregression, and difference equations.

Census II X-11: A method that systematically decomposes a time series into trend, cyclical, seasonal, and error components. It was developed by the Bureau of the Census of the Department of Commerce and is widely used in deseasonalizing economic data.

Deseasonalization: (Sometimes also called seasonal adjustment,) a process of removing seasonality from the time series. Most governmental statistics are seasonally adjusted to better reflect other components in a time series.

Detrending: A process of removing trend from the time series through either differencing of time series observations or subtracting fitted trends from actual observations.

Feedforward Neural Networks: A special type of neural networks where processing elements are arranged in layers and the information is one-directional from input layer to hidden layer(s) to output layer.

Neural Networks: Computing systems that are composed of many simple processing elements operating in parallel whose function is determined by network structure. They are used mainly to model functional relationship among many variables.

Seasonality: Periodic pattern that typically occurs within a year and repeats itself year after year. Most retail data exhibit seasonal variations that repeat year after year.

Trend: Long-term movement in a time series that represents the growth or decline over a extended period of time.

New Advancements in Image Segmentation for CBIR

Yu-Jin Zhang
Tsinghua University, Beijing, China

INTRODUCTION

The process of segmenting images is one of the most critical ones in automatic image analysis whose goal can be regarded as to find what objects are presented in images (Pavlidis, 1988). Image segmentation consists of subdividing an image into its constituent parts and extracting these parts of interest (objects). A large number of segmentation algorithms have been developed since the middle of 1960's (see survey papers and books, for example, Bovik, 2000; Fu & Mui, 1981; Lucchese & Mitra, 2001; Medioni, Lee, & Tang, 2000; Pal & Pal, 1993; Zhang, 2001), and this number continually increases from year to year in a fast rate. This number had attended, 10 years ago, the order of thousands (Zhang & Gerbrands, 1994). However, none of the proposed segmentation algorithms is generally applicable to all images, and different algorithms are not equally suitable for a particular application. Though several thousands of algorithms have been proposed, improvements for existing algorithms and developments for treating new applications are still going on.

Along with the progress of content-based image retrieval (CBIR) that searches images from a collection according to the content of images, especially with the object-based techniques, image segmentation has been involved more and more in this new research domain. By separating an image into different components and extracting the interest objects, more semantic meanings could be mined out from the image. Thus, the human intention could be fully represented and integrated into the search process. Due to the specific requirements and particular constrains of CBIR, not only have many common segmentation techniques proposed before the start of CBIR research in 1992 have been revised, but also various techniques specialized to the tasks of CBIR have been developed.

The remainder of this article is organized as follows: in the *BACKGROUND* section, image segmentation techniques are briefly summarized; and the specific requirements of CBIR to image segmentation are stated. In *MAIN THRUST*, a general trend for developing image segmentation techniques to fit the particular requirements of CBIR are discussed; some useful techniques for segmentation in the context of CBIR, such as meaningful region extractions and two-level segmentation are presented. In *FUTURE TRENDS*, some potential directions for further research are indicated. In *CONCLUSION*, several final remarks are given.

BACKGROUND

Classification of Image Segmentation Techniques

A formal mathematical-oriented definition for image segmentation can be found in Fu and Mui (1981). Generally speaking, it defines image segmentation as a process of treating every pixel in images to form non-overlap regions, each of which has certain properties different from that of others.

As many techniques for image segmentation have been proposed in the literature, the classification of segmentation algorithms becomes critical. Algorithm classification should be performed according to some suitable classification criteria; otherwise, the classification schemes might be non-consistent and/or non-complete (Fu & Mui, 1981; Pal & Pal, 1993).

Considering the classification of segmentation algorithms as a partition of a set (of algorithms) into subsets, an appropriate classification scheme should satisfy the following four conditions (Zhang, 1997; Gonzalez & Woods, 2002):

(1) Every considered algorithm must be classified into one subset.
(2) All subsets together can include all algorithms (form the whole set).
(3) Different algorithms in the same subset should have some common properties.
(4) The algorithms in different subsets should have certain distinguishable properties.

Two criteria commonly used for the classification of segmentation algorithms are based on two pairs of complementary attributes:

(1) Homogeneous property: either discontinual (inter-regions, at boundaries) or continual (intra-region);

(2) Processing strategy: either sequential (step-by-step) or parallel (simultaneous and independent).

Combining these two criteria, a classification scheme can distinguish the following four groups of segmentation algorithms (Zhang, 2001):

(1) Boundary-based sequential algorithms;
(2) Boundary-based parallel algorithms;
(3) Region-based sequential algorithms;
(4) Region-based parallel algorithms.

Recently, many new segmentation algorithms based on different theories, such as mathematical morphology, pattern recognition principle, neuron network, information theory, fuzzy logic, wavelet transform, genetic algorithms, and so forth, are proposed (Zhang 2001). All these algorithms can still be classified according to the above classification scheme.

In the context of image analysis aiming at automatically obtaining measurement data from objects, the most important criteria for judging the performance of segmentation algorithms are related to the accuracy of segmentation results, though some other criteria, such as processing complexity and efficiency as well as segmentation resolution of algorithms, are also considered (Zhang, 1996). From the accuracy point of view, the algorithms based on boundary representation and sequential strategy are often more powerful (especially for noisy images), according to their natures, than other algorithms.

Particularity of Image Segmentation in CBIR

Design of segmentation techniques for CBIR counts several new challenges. Two of them are indicated here.

In contrast to many applications of image analysis, which aim at obtaining objective measurements over objects in the images, CBIR is a somehow subjective process (Zhang, 2003). Instead of deriving accurate measures of object properties for further identification or classification, the primary concerns in CBIR are to separate required regions and to obtain more information related to the semantic of these regions.

From other side, typical image analysis applications are normally limited to particular domains, some *a priori* knowledge about the scene geometry, the sensor characteristics, lighting conditions, noise statistics, number and form of objects, and so forth, are available. Taking into consideration such kinds of information, the results of segmentation could be improved. While in CBIR using a generic database, none of these factors can be controlled; segmentation algorithms could not be specialized by taking into account this information.

Due to these particular aspects, the development of image segmentation techniques for CBIR should be guided by the purpose of segmentation in CBIR applications.

MAIN THRUST

General Ideas

With the progress in segmentation techniques, people have realized that a precise segmentation of objects in many cases is still beyond the capability of current computer techniques (Zhang, 2001). On the other hand, compared to some image analysis tasks that aim at obtaining accurate measurements from the segmented objects, the requirement for precise segmentation of objects can be somehow relaxed in the context of CBIR. As discussed previously, image retrieval is a subject-oriented process in which the precise object measurement might not be the must. In particular, for object-based image retrieval, the main purpose of segmentation here is for identifying the objects.

Several techniques taken into consideration of these particularities are proposed; some typical examples are discussed in the following.

Taking the homogeneous regions in images as the primitives (the "atomic" structures) and assuming these regions do not necessarily cover the whole images, an algorithm called perceptual region growing (PRG) has been proposed (Siebert, 1998). This algorithm combines region growing, edge detection, and perceptual organization principles together for a domain independent segmentation. The main stages of PRG consist of the following:

(1) Identify the seed pixels of each region from images;
(2) Merge adjacent pixels that have similar intensity values to these regions;
(3) Compute the gradient map of images and try to form closed boundaries with high gradient pixels;
(4) If a boundary is apparently formed, continuously growing its enclosed region to the limit of boundary.

Presuming that regions in images should integrate enough intrinsic variability to provide a better characterization of regions, an algorithm based on coarse region detection and fine description is suggested (Fauqueur & Boujemaa, 2002). The coarse determination of regions can alleviate the problems caused by over-segmentation that produces homogeneous and small regions, such as: it is hard to differentiate homogeneous regions; it is rarely that a small region to be visually salient in a scene. Since regions are detected coarsely and these regions should be more homogeneous than the whole images, finely describing the visual appearance is important to reveal even-detailed information in images, such as the granularity of color and the shades of hue.

Considering that the basis for retrieval queries would be the image semantics, the purpose of image segmentation should be to split the image into pieces for capturing a reduced and implicit portion of the full semantics (Jermyn, 2003). Thus, obtained semantics are successfully used for a psychovisual test with some interesting results that both the boundaries of regions and regions themselves are comparable in presenting the semantic meanings of images to human beings. The human visual system is quite tolerant to a number of factors, so the object extraction task can be simplified to a certain extent in this context.

Some details of two potential techniques — 1) extracting meaningful regions for object identification, and 2) two-level segmentation and matching — are described in the following as other examples.

Meaningful Region Extraction for Object Identification

The objective of image segmentation in CBIR applications is more related to the number and position of salient objects in images. So, one may only extract approximately their corresponding region, instead of to segment each object accurately (Luo et al., 2001). It is known that, during the process of visual perception and recognition, human eyes actively perform the problem-oriented and/or application-oriented selection and aggressively process the information from the visible world by moving and successfully fixating on the most informative parts of the image. These parts might not have the one-to-one correspondence to all objects, but they have some obvious common visual properties. These informative parts could be called "meaningful regions". Here, the definition of meaningful regions lies more on the general visual appearance, such as the form and position of big regions in images, while the exact boundary of region is less notable in the primary concern.

The "meaningful region" provides an effective visual representation of objects from the point of view of object identification. Though the "meaningful region" is not an exact representation of the objects, its semantic meaning could later be recovered based on some domain knowledge. On the other hand, robust extraction of "meaningful regions" could be accomplished in a relatively easy manner than accurate segmentation does. This makes the extraction of a "meaningful region" a feasible step for further object identification in CBIR using object-based techniques (Gao, Zhang, & Merzlyakov, 2000).

The general flowchart for meaningful region extraction is shown in Figure 1. First, the image to be treated is transformed or mapped into a visual feature space that combines the information of different visual features (such as HSI color features and Gabor texture features); an index of image complexity based on weighted density is used to select features for splitting images (Luo et al., 2001). Then, an unsupervised and adaptive clustering algorithm is applied to group pixels to form meaningful regions. The meaningful regions thus formed will be further identified with the help of a scene knowledge database, while those pixels that cannot form a meaningful region will be discarded. In such a way, the main object regions in images can be quickly captured, and their spatial relationship can be additionally established.

Two-level Segmentation and Matching

In CBIR applications, the search of images is based on the similarity matching between the query image and those database images. While image segmentation is first carried out, the matching will be performed between segmented regions. For generic images, as the number of objects in images are not known in advance (as indicated previously), over-segmentation (the image is divided too much in detail) and under-segmentation (the image is too coarsely divided) results are often created. Both will make the subsequent matching task less feasible (Wang & Wiederhold, 2001).

One method to solve the over-segmentation and under-segmentation problems in CBIR applications is to use a particular matching procedure to diminish the prerequisite for an accurate segmentation (Dai & Zhang, 2004). As the problem caused by over-segmentation and/or under-

Figure 1. Meaningful region extraction flowchart

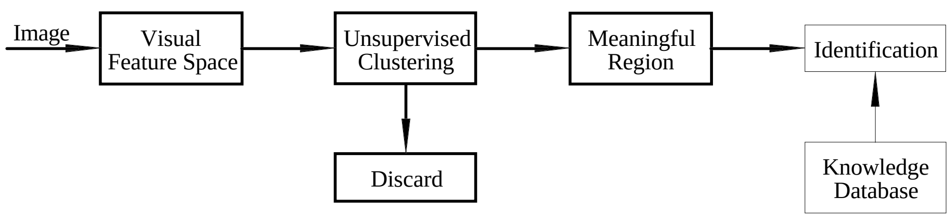


Figure 2. Two levels of segmentation results.

segmentation to the one-to-one matching can be considered as a problem of number inconsistency, the solution would be the permitting of a one-to-many or a many-to-one matching.

This adaptable matching strategy compensate for the problems caused by inaccurate segmentation. In practice, the image is first over-segmented, and then the resulted regions will be merged into under-segmented results. The over-segmented result is called detailed result and the under-segmented result is called rough result. Two groups of segmentation examples are shown in Figure 2, where for each group, the first is the original image, the second is the detailed result, and the third is the rough result.

Both over-segmented and under-segmented results are easily obtained over an exact result (Dai & Zhang, 2003). For example, the parameters in the segmentation algorithm can be biased first to perform the over-segmentation for producing detailed results. Then, by combining several regions in detailed results, a rough result corresponding to an under-segmented result can be obtained. In the retrieval process, the database images will be described based on the detailed result while the query images will be described by rough result. In other words, there are descriptions in two levels. The rough description for query images has fewer restrictions than the detailed description for database images does; this makes the query more flexible.

The similarity matching which follows the description is not computed by comparing region to region directly as the two-level description breaks the equivalence between regions in query and database images. On the other hand, the rough description of query images allows a many-to-one matching between the regions from database images and the regions from query image. This can also be considered as using the guidance of rough description for query images to group the database images with detailed descriptions. In such a way, the object matching is performed with fewer prerequisites for accurate image segmentation.

FUTURE TRENDS

To improve the performance of CBIR, different image segmentation techniques are to be enhanced:

(1) Describing objects as congruous to humans' sense as possible should be studied. As human beings are still far from knowing all of the cognitive details from the real world, how to automatically form semantic objects is a challenging task.

(2) Improving the robustness of meaningful region detection and extraction, especially with complicated images, by taking more characteristics of images into consideration.

(3) Using a more efficient feedback procedure to make the query process fast to follow the users' aspiration in the course of retrieval is an attractive research topic.

CONCLUSION

Efficient image segmentation should take into consideration the tasks that follow the segmentation step and also the goal of application in which segmentation is an indispensable stage. This has been clearly shown by the new advancements in image segmentation for CBIR. In general, further researches on image segmentation would be along the directions of improving concrete techniques for advanced image analysis and following the movements toward image understanding.

REFERENCES

Bovik, A.C. (2000). *Handbook of image and video processing*. Academic Press.

Dai, S.Y., & Zhang, Y.J. (2003). Color Image segmentation with watershed on color histogram and Markov random fields. *Proc. ICICS-PCM*, p0240 (pp.1-5).

Dai, S.Y., & Zhang, Y.J. (2004). Unbalanced region matching based on two-level description for image retrieval. *Pattern Recognition Letters*.

Fauqueur, J., & Boujemaa, N. (2002). Image retrieval by regions: coarse segmentation and fine color description. *Proc. International Conference on Visual Information System*.

Fu, K.S., & Mui, J.K. (1981). A survey on image segmentation, *Pattern Recognition, 13,* 3-16.

Gao, Y.Y., Zhang, Y.J., & Merzlyakov, N.S. (2000). Semantic-based image description model and its implementation for image retrieval. *Proceedings of the First International Conference on Image and Graphics*, 657-660.

Gonzalez, R.C., & Woods, R.E. (2002). *Digital Image Processing* (2nd ed.), Prentice Hall.

Jermyn, I. (2003) Psychovisual evaluation of image database retrieval and image segmentation. *ERCIM News, 55*, 20-22.

Lucchese, L., & Mitra, S.K. (2001). Color image segmentation: a state-of-the-art survey. *http://eecs.oregonstate.edu/~luca/proc_insa_01.pdf*

Luo, Y., Zhang, Y.J. & Gao, Y.Y., et al. (2001). Extracting meaningful region for content-based retrieval of image and video. *Proc. Visual Communications and Image Processing, 4310,* 455-464.

Medioni, G., Lee, M.S., & Tang, C.K. (2000). *A computational framework for segmentation and grouping*. Elsevier.

Pal, N.R., & Pal, S.K. (1993). A Review on image segmentation techniques, *Pattern Recognition, 26,* 1277-1294.

Pavlidis, T. (1988). Image analysis, *Ann. Rev. Comput. Sci., 3,* 121-146.

Siebert, A. (1998) Segmentation based image retrieval. *Proc. Storage and Retrieval for Image and Video Databases VI*, 14-24.

Wang, J.Z., Li, J., & Wiederhold, G. (2001). SIMPLIcity: Semantics-sensitive integrated matching for picture libraries. *IEEE Trans. PAMI. 23*(9), 947-963.

Zhang, Y.J. (1996). A survey on evaluation methods for image segmentation. *Pattern Recognition, 29*(8), 1335-1346.

Zhang, Y.J. (1997). Evaluation and comparison of different segmentation algorithms. *Pattern Recognition Letters, 18*(10), 963-974.

Zhang, Y.J. (2001). *Image segmentation*, Science Publisher.

Zhang, Y.J. (2003). *Content-based visual information retrieval*. Science Publisher.

Zhang, Y.J., & Gerbrands, J.J. (1994). Objective and quantitative segmentation evaluation and comparison. *Signal Processing, 39*(3), 43-54.

KEY TERMS

Accuracy: The agreement between the real measurement and some objective standard taken as the "ground truth." In a given measurement experimentation, the high accuracy could be achieved using correct sample design and measurement methods. An accurate measurement is also called an un-biased measurement.

Clustering: A process to group, based on some defined criteria, two or more terms together to form a large collection. In the context of image segmentation, clustering is to gather several pixels or groups of pixels with similar property to form a region.

Content-Based Image Retrieval (CBIR): A process framework for efficiently retrieving images from a collection by similarity. The retrieval relies on extracting the appropriate characteristic quantities describing the desired contents of images. In addition, suitable querying, matching, indexing and searching techniques are required.

HSI Model: A color model in which a color is described by three characteristics: hue (H), saturation (S), and intensity (I). They form a 3-D color space. HSI model is quite coherent with human perception. A closely related model is HSV model, in which value (V) takes the place of intensity.

Image Analysis: An important layer of image engineering, which is concerned with the extraction of information (by meaningful measurements with descriptive parameters) from an image (especially from interesting objects), with the goal of finding what objects (based on identification, recognition and classification) are present in the image.

Perceptual Organization: A process to group sensory primitives arising from a common underlying cause. Perceptual organization can also be defined as the ability to impose structural regularity on sensory data.

Precision: In terms of repeatability, the ability of the measurement process to duplicate the same object measurement and produce the same result. The high or low precision could be obtained is dependent upon the object of interest and in many cases could be controlled by working harder.

Segmentation Evaluation: A process to judge the performance of segmentation algorithms based on some defined quality criteria and/or ground truth in view to assess or reveal the property of algorithms in use. The inter-algorithm evaluation is also called segmentation comparison, while the intra-algorithm evaluation can be called segmentation characterization.

New Perspectives on Rewards and Knowledge Sharing

Gee-Woo (Gilbert) Bock
National University of Singapore, Singapore

Chen Way Siew
National University of Singapore, Singapore

Young-Gul Kim
KAIST, Korea

INTRODUCTION

Of the 260 responses from a survey of European multinationals, 94% believed that knowledge management requires employees to share what they know with others within the organization (Murray, 1999). Among the processes of knowledge management—creation, sharing, utilization and accumulation of knowledge—sharing is what differentiates organizational knowledge management from individual learning or knowledge acquisition.

However, the process of sharing knowledge is often unnatural to many. Individuals will not share knowledge that is regarded to be of high value and importance. In fact, the natural tendency for individuals is to hoard knowledge or look suspiciously at the knowledge of others. Thus, incentive schemes—where employees receive incentives as a form of compensation for their contributions—are common programs in many organizations. Such schemes have met their fair share of success as well as failure in the field of knowledge management. On the one hand, the carrot and stick principle used in Siemens' ShareNet project turned out to be a success (Ewing & Keenan, 2001). On the other hand, the redemption points used in Samsung Life Insurance's Knowledge Mileage Program only resulted in the increasingly selfish behavior of its employees (Hyoung & Moon, 2002).

Furthermore, despite the plethora of research on factors affecting knowledge sharing behavior, little concerns discovering effective ways to encourage individuals to voluntarily share their knowledge. Early studies on knowledge management began by trying to discover key factors pertaining to knowledge management in general, instead of knowledge sharing in particular, as summarized in Table 1. Although research on knowledge sharing started around the mid 1990s, it focused mainly on knowledge sharing at the group or organizational level in spite of the fact that knowledge itself actually originates from the individual. Even at the group or organizational level, most studies dealt with a specific knowledge type, such as best practices (Szulanski, 1996) or a specific context, such as between dispersed teams (Tsai, 2002). In addition, factors such as trust, willingness to share, information about the knowledge holder, and the level of codification of knowledge were considered in abstract. Although these factors are valuable, they require further empirical research before they could be used to explain the individual's fundamental motivation to share knowledge. Thus, this study aims to develop an understanding of the factors that support or constrain the individual's knowledge sharing behavior in the organization, with a special interest in the role of rewards. This is done according to Fishbein and Ajzen's (1975) Theory of Reasoned Action (TRA), a widely accepted social psychology model that is used to explain almost any human behavior (Ajzen & Fishbein, 1980).

BACKGROUND

Due to the fact that knowledge is a resource that is locked in the minds of humans, knowledge sharing does not occur with the sole implementation of information systems. As such, an investigation into the individual's motivation behind knowledge sharing behavior, coupled with a firm foundation in social psychology, should take precedence. Accordingly, the TRA is adopted so as to provide a well-established explanation for such volitional, rational, systematic decision logic as that of knowledge sharing.

The TRA assumes that human beings are usually rational in thinking, and would systematically use available information (Fishbein & Ajzen, 1975). In the TRA, the individual's attitude toward and subjective norm regarding a behavior jointly determine the behavioral intention that results in the individual's decision to engage in a specific behavior. In this study, we focus only on the

Table 1. Factors affecting knowledge management and knowledge sharing

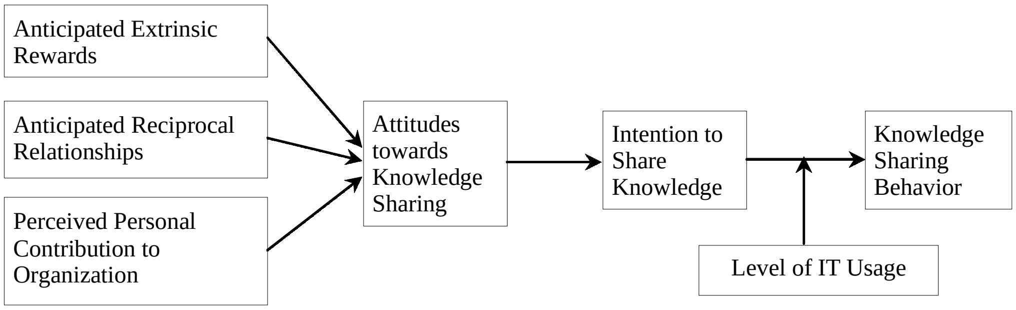
	Factors	References
Knowledge Management	Knowledge management system, Network, Knowledge worker, Clear vision and goals, Middle-up-down management, Organizational change, Monitoring and support, Knowledge infrastructure, Knowledge repository and map, Organizational culture, Top manager's support	Davenport, De Long, and Beers (1998); Davenport and Prusak (1998); Earl (1996); Nonaka and Takeuchi (1995); Ulrich (1998); Wiig (1997)
Knowledge Sharing	***The Group and Organizational Level*** Level of trust between groups, Arduous relationship between source and the recipient, Role of top managers, Characteristics of knowledge, Prior experience on knowledge transfer, Channel richness, Openness of the organization	Butler (1999); Gupta and Govindarajan (2000); Kogut and Zander (1993); Nelson and Cooprider (1996); Szulanski (1996); Wathne, Roos and Krogh (1996)
	The Individual Level Trust between individuals, Willingness to share, Information about the knowledge holder, Level of codification of knowledge	Hansen (1999); Kramer (1999); Moreland (1999); Stasser, Stewart, and Wittenbaum (1995); Tsai and Ghoshal (1998)

salient beliefs that affect the knowledge sharing attitude because knowledge sharing behavior is assumed to be motivated and executed mainly at the individual level. Since the TRA can be applied to almost any behavior, the nature of the beliefs operative for a particular behavior are left unspecified. Following the elicitation recommendations suggested by Fishbein and Ajzen (1975), free response interviews to elicit five to nine salient beliefs were conducted with chief knowledge officers (CKO) and chief information officers (CIO) of the subject population in April 1999. Once these salient beliefs surfaced, the research model was developed.

We propose three factors that are consistently emphasized throughout the interviews: anticipated extrinsic rewards, anticipated reciprocal relationships, and perceived personal contribution to the organization, as the antecedents of the attitudes towards knowledge sharing. According to the interdependence theory, individuals will behave according to rational self-interest. Knowledge sharing occurs when the rewards exceed the costs (Constant, Keisler & Sproull, 1994; Kelley & Thibaut, 1978), implying that anticipated extrinsic rewards will positively affect the individual's attitude. Concerning intrinsic rewards, the social exchange theory states that social exchanges entail unspecified obligations (Blau, 1967). As employees are seen to believe that their relationship with others can be improved through sharing knowledge, the anticipated reciprocal relationships positively affect the individual's attitude. In addition to these, the self-motivation theory (Deci, Connell & Ryan, 1989; Iaffaldano & Muchinsky, 1985; Schwab & Cummings, 1970) finds that feedback from others on shared knowledge can form a self-motivational factor and serve as another major determinant of the attitude toward knowledge sharing. Eisenberger and Cameron (1996) note that one's sense of competence actually increases due to the feedback concerning the quality of one's output. Employees who are able to link instances of past knowledge sharing with an understanding of how these actions contribute to others' work, and/or improvements in organizational performance are likely to develop more favorable attitudes toward knowledge sharing than employees who are unable to construct such linkages. Finally, following Fishbein and Ajzen's (1975) argument about the possibility of several external variables affecting intention to perform a behavior, we introduced an aspect of information technology (IT) into our model. Since IT is considered to be an important enabler in knowledge management (O'Dell & Grayson, 1998; Ruggles, 1998), we examined how the individual's level of IT usage affects knowledge sharing behavior.

Data were collected through the utilization of a survey. A total of 900 questionnaires were distributed in October and November 1999 to employees in 75 departments of four large government-invested organizations in South Korea. Of this total number, 861 responses were received, of which 467 were usable. We found that the anticipated reciprocal relationship provided for the individual's positive attitude towards knowledge sharing, and resulted in a positive influence of intention and behavior. However, contrary to many researchers' expectations, anticipated extrinsic rewards were found to have a negative effect on such an attitude.

Figure 1. Research model



FUTURE TRENDS

This negative correlation—which might prove important for future research—can be explained with the results of research in the pay-performance area. Kohn (1993) found that there is either no relationship or a negative relationship between rewards and performance, although many assume that people will do a better job if they are promised some form of reward. Kohn cited six reasons as to why rewards fail, three of which can also be considered within the knowledge-sharing context.

First, rewards are seen to have a punitive effect because, as compared to outright punishment, they are manipulative in nature. Not receiving an expected reward is seen to be indistinguishable from being punished. Both result in movement, but not motivation (Herzberg, 1968). Rewards are seen to destroy relationships because for any one winner, many others would feel that they have lost. When there exists a limited number of rewards, competition between employees will ensue. Second, rewards are at times used as a simpler alternative to addressing underlying issues, such as the lack of an ideal knowledge-sharing culture within the organization. The ideal culture mentioned should include providing useful feedback, social support, and room for self-determination. Third, rewards could be an undermining factor towards intrinsic motivation. Interest in knowledge sharing would decrease with an increase in one's perception of being controlled (Levinson, 1973). Employees might assume that the task at hand is not something they would want to do if they have to be bribed to do it. As such, with the increase in incentive offered, the negative perception towards the task at hand becomes greater.

Another explanation can also be found in organizational citizenship behavior literature. According to Katz and Kahn (1966), any critical voluntary behavior that is beyond the scope of one's job description is a direct result of one's identification with and internalization of individual and organizational values, rather than the involvement of any external factors. Furthermore, Constant et al. (1994) stated that experienced workers perceive the process of sharing knowledge as part of normal business activity. These workers hold a negative view of any extrinsic rewards given in return for sharing knowledge. With such strong support for the negative effect view of extrinsic rewards on the attitude towards knowledge sharing, would it be right to completely discard extrinsic rewards?

Eisenberger and Cameron (1996) found that extrinsic rewards could both positively and negatively influence motivation—knowledge sharing in this case. They find that rewards can be divided into two broad types, namely task-contingent and quality-dependent rewards. Quality-dependent rewards positively influence organization initiatives, as they do not reduce one's intrinsic motivation. In fact, due to the feedback concerning the quality of one's output, one's sense of competence actually increases. Task-contingent rewards, on the other hand, undermine any task because of their negative influence on intrinsic motivation. A possible design for such a scheme is a knowledge market where "buyers and sellers of knowledge negotiate a mutually satisfactory price for the knowledge exchanged" (Ba, Stallaert & Whinston, 2001, p. 232). In this way, the reward of individuals would be based on the usefulness of their knowledge, thus ensuring the creation of high-quality knowledge.

In addition, a reward that is less than what employees feel their performance justify could threaten their self-esteem. This is due to the fact that the self-ratings done by employees are usually higher than those done by the management. According to Meyer (1975), a common way for employees to cope with such a problem is to "downgrade the importance of the activity on which the threat is focused" (p. 44). Hence, an incentive scheme should be well-designed so as to reward individuals who are deemed deserving, as the rewarding of contributors aids in positively influencing the sharing of knowledge. Additionally, the scheme's design should discourage self-centered behavior, which is detrimental to the organization's health (Michailova & Husted, 2003); otherwise, the scheme would only produce temporary compliance, and might decrease an individual's intrinsic motivation (Deci, 1971,1972a). When intrinsic interests exist among employees for a particular task, the attachment of incentives to the performance of the task only results in the decrease of that interest (Deci, 1972b). "When pay becomes the important goal, the individual's interest tends to focus on that goal rather than on the performance of the task itself" (Meyer, 1975, p. 41), thus resulting in employees striving to increase incentives at the cost of output quality.

Furthermore and most importantly, an incentive scheme needs to be incorporated with proper organization norms (Markus, 2001). If the norms are not in place, employees will not share their knowledge even if there is in place, a comprehensive incentive scheme. According to O'Dell and Grayson (1998), "if the process of sharing and transfer is not inherently rewarding, celebrated and supported by the culture, then artificial rewards [will not] have much effect and can make people cynical" (p. 168). Husted and Michailova (2002) also stated that "unless knowledge sharing is built into the expectation of the individual and is reflected in the reward mechanism, sharing will not take place".

CONCLUSIONS

In summary, of the two views posed by past research, our recent study to discover the influence of extrinsic rewards in knowledge sharing supports the negative view. This implies that, when the management of an organization is motivated to embrace knowledge sharing but its employees are not, using incentives to influence knowledge sharing would only result in the employees placing emphasis on the incentives. This could result in the sharing of low-quality knowledge and undermine the whole knowledge-sharing effort. Furthermore, the continuous use of incentives "could actually be encouraging hoarding behavior and competitive actions, diminishing the free flow of knowledge in the organization" (Wasko & Faraj, 2000, p. 162). Therefore, extrinsic rewards should be coupled with other factors, such as organizational norms, to bring about benefits. The reconciliation of this disparity in views should provide new grounds for future research.

REFERENCES

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall.

Alavi, M., & Leidner, D.E. (2001). Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly*, *25*(1), 107-136.

Ba, S., Stallaert, J., & Whinston, A.B. (2001). Introducing a third dimension in information systems design-The case for incentive alignment. *Information Systems Research*, *12*(3), 225-239.

Blau, P. (1967). *Exchange and power in social life*. New York: Wiley.

Bock, G.W., & Kim, Y.G. (2002). Breaking the myths of rewards: An exploratory study of attitudes about knowledge sharing. *Information Resources Management Journal*, *15*(2), 14-21.

Bock, G.W., Zmud, R.W., Kim, Y.G., & Lee, J.N. (2003). *Determinants of the individual's knowledge sharing behavior: From the theory of reasoned action*. University of Minnesota: Minnesota Symposium on Knowledge Management, 3.

Butler, J.K. (1999). Trust expectations, information sharing, climate of trust, and negotiation effectiveness and efficiency. *Group & Organization Management*, *24*(2), 217-238.

Constant, D., Keisler, S., & Sproull, L. (1994). What's mine is ours, or is it? A study of attitudes about information sharing. *Information Systems Research*, *5*(4), 400-421.

Davenport, T.H., De Long, D.W., & Beers, M.C. (1998). Successful knowledge management projects. *Sloan Management Review*, *39*(2), 43-57.

Davenport, T.H., & Prusak, L. (1998). *Working knowledge*. Boston, MA: Harvard Business School Press.

Deci, E.L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, *18*(1), 105-115.

Deci, E.L. (1972a). Intrinsic motivation, extrinsic reinforcement, and inequity. *Journal of Personality and Social Psychology*, *22*(1), 113-120.

N

Deci, E.L. (1972b). The effects of contingent and noncontingent rewards and controls on intrinsic motivation. *Organizational Behavior and Human Performance*, *8*, 217-229.

Deci, E.L., Connell, J.P., & Ryan, R.M. (1989). Self-determination in a work organization. *Journal of Applied Psychology*, *74*(4), 580-590.

Earl, M.J. (1996). Knowledge as strategy. In L. Prusak (Ed.), *Knowledge in organizations*. Newton, MA: Butterworth-Heinemann.

Eisenberger, R., & Cameron, J. (1996). Detrimental effects of reward: Reality or myth? *American Psychologist*, *51*(11), 1153-1166.

Ewing, J., & Keenan, F. (2001). Sharing the wealth. *Business Week*, *3724*, EB36-EB40.

Fishbein, M., & Ajzen, I. (1975). *Beliefs, attitude, intention and behavior: An introduction to theory and research*. Philippines: Addison-Wesley Publishing Company.

Gupta, A.K., & Govindarajan, V. (2000). Knowledge flows within multinational corporations. *Strategic Management Journal*, *21*(4), 473-496.

Hansen, M.T. (1999). The search–transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Administrative Science Quarterly*, *44*(1), 82-111.

Herzberg, F. (1968). One more time: How do you motivate employees? *Harvard Business Review*, *46*(1), 53-62.

Husted, K., & Michailova, S. (2002). Knowledge sharing in Russian companies with western participation. *International Management*, *6*(2), 17-28.

Hyoung, K.M., & Moon, S.P. (2002). Effective reward systems for knowledge sharing. *Knowledge Management Review*, *4*(6), 22-25.

Iaffaldano, M.T., & Muchinsky, P.M. (1985). Job satisfaction and job performance: A meta-analysis. *Psychological Bulletin*, *97*(2), 251-273.

Katz, D., & Kahn, R.L. (1966). *The social psychology of organizations*. New York: Wiley.

Kelley, H.H., & Thibaut, J.W. (1978). *Interpersonal relations: A theory of independence*. New York: Wiley.

Kogut, B., & Zander U. (1993). Knowledge of the firm and the evolutionary theory of the multinational corporation. *Journal of International Business Studies*, *24*(4), 625-645.

Kohn, A. (1993). Why incentive plans cannot work. *Harvard Business Review*, *71*(5), 54-63.

Kramer, R.M. (1999). Social uncertainty and collective paranoia in knowledge communities: Thinking and acting in the shadow of doubt. In L.L. Thomson, J.M. Levine & D.M. Messick (Eds.), *Shared cognition in organizations, the management of knowledge*. London: LEA Inc.

Levinson, H. (1973). Asinine attitudes toward motivation. *Harvard Business Review*, *51*(1), 70-76.

Markus, M.L. (2001). Toward a theory of knowledge reuse: Types of knowledge reuse situations and factors in reuse success. *Journal of Management Information Systems*, *18*(1), 57-93.

McDermott, R., & O'Dell, C. (2001). Overcoming cultural barriers to sharing knowledge. *Journal of Knowledge Management*, *5*(1), 76-85.

Meyer, H.H. (1975). The pay-for-performance dilemma. *Organization Dynamics*, *3*(3), 39-50.

Michailova, S., & Husted, K. (2003). Knowledge-sharing hostility in Russian firms. *California Management Review*, *45*(3), 59-77.

Moreland, R.L. (1999). Transactive memory: Learning who knows what is work groups and organizations. In L.L. Thomson, J.M. Levine & D.M. Messick (Eds.), *Shared cognition in organizations, the management of knowledge*. London: LEA Inc.

Murray, P. (1999, March 8). How smarter companies get results from KM. *Financial Times*, 15.

Nelson, K.M., & Cooprider, J.G. (1996). The contribution of shared knowledge to IS group performance. *MIS Quarterly*, *20*(4), 409-429.

Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company*. New York: Oxford University Press.

O'Dell, C., & Grayson, C.J. (1998). If only we knew what we know: Identification and transfer of internal best practices. *California Management Review*, *40*(3), 154-174.

Ruggles, R. (1998). The state of notion: Knowledge management in practice. *California Management Review*, *40*(3), 80-89.

Schwab, D.P., & Cummings, L.L. (1970). Theories of performance and satisfaction. *Industrial Relations*, *9*(4), 408-430.

Stajkovic, A.D., & Luthans, F. (1998). Social cognitive theory and self-efficacy: Going beyond traditional motivational and behavioral approaches. *Organizational Dynamics*, *26*(4), 62-74.

Stasser, G., Stewart, D.D., & Wittenbaum, G.M. (1995). Expert roles and information exchange during discussion: The importance of knowing who knows what. *Journal of Experimental Social Psychology*, *31*(3), 244-265.

Stenmark, D. (2000). Leveraging tacit organizational knowledge. *Journal of Management Information Systems*, *17*(3), 9-24.

Szulanski, G. (1996). Exploring internal stickiness: Impediments to the transfer of best practice within the firm. *Strategic Management Journal*, *17*(Winter), 27-44.

Triandis, H.C. (1971). *Attitude and attitude change*. New York: John Wiley & Sons, Inc.

Tsai, W. (2002). Social structure of "coopetition" within a multiunit organization: Coordination, competition, and intraorganizational knowledge sharing. *Organization Science*, *13*(2), 179-190.

Tsai, W., & Ghoshal, S. (1998). Social capital and value creation: The role of intrafirm networks. *Academy of Management Journal*, *41*(4), 464-476.

Ulrich, D. (1998). Intellectual capital = Competence x commitment. *Sloan Management Review*, *39*(2), 15-26.

Wasko, M.M., & Faraj, S. (2000). "It is what one does": Why people participate and help others in electronic communities of practice. *The Journal of Strategic Information Systems*, *9*(2-3), 155-173.

Wathne, K., Roos, J., & Krogh, G. (1996). Toward a theory of knowledge transfer in a cooperative context. In G. Krogh & J. Roos (Eds.), *Managing knowledge*. London: Sage.

Wiig, K.M. (1997). Knowledge management: Where did it come from and where will it go? *Expert Systems with Applications*, *13*(1), 1-14.

KEY TERMS

Explicit Knowledge: Knowledge that has been captured and codified into manuals, procedures, and rules, and is easy to disseminate (Stenmark, 2000).

Extrinsic Rewards: Incentives that are mediated outside of a person, such as praises and monetary compensation (Deci, 1972b).

Implicit Knowledge: Knowledge that can be expressed in verbal, symbolic, or written form but has yet to be expressed (Bock, Zmud, Kim, & Lee, 2003).

Intrinsic Rewards: Incentives that are mediated within a person, such as satisfaction (Deci, 1972b).

Knowledge Management System: A knowledge repository, shared knowledge base or knowledge based system, which is a class of information systems developed to support and enhance the organizational processes of knowledge creation, storage / retrieval, transfer and application (Alavi & Leidner, 2001).

Knowledge Sharing: Voluntary activities of transferring or disseminating knowledge between people or groups in an organization (Bock, Zmud, Kim & Lee, 2003).

Organizational Norm: Organization culture or climate, which consists of the shared values, beliefs and practices of the people in the organization (McDermott & O'Dell, 2001).

Tacit Knowledge: Knowledge that cannot be easily articulated, and thus only exists in people's minds, and is manifested through their actions (Stenmark, 2000).

New SQL Standard in Database Modeling

Eric Pardede
La Trobe University, Australia

J. Wenny Rahayu
La Trobe University, Australia

David Taniar
Monash University, Australia

INTRODUCTION

Relational Database (RDB) is arguably the most widely used repository for database applications. Since the 1970s, we have witnessed the relational data model, from which the RDB is originated, evolving. The progress aims to answer the increasing requirement in database applications. One of them is the requirement to deal with complex structure of real world problems. Unlike its Object-Oriented Database (OODB) counterpart, the RDB, for example, does not have facilities to store large structured objects, semi-structured data, and so forth.

The question might have been answered with the release of new version *Structured Query Language* (SQL) (Fortier, 1999; Melton, Simon, & Gray, 2001). The new SQL has provided many new features including many new data types. Since SQL is used for database definition and manipulation for relational model, the extension has enriched the modeling capability in RDB.

A problem then arises. With the existence of the new data types in the new SQL, the whole database modeling processes have changed. So far, database designers still use the conventional method for RDB design and implementation. This method can create inefficient and incorrect use of the new data types.

This work aims to show how new SQL data types affect the database modeling processes. It highlights new opportunities and new research challenges, brought by the new standard.

BACKGROUND: HISTORY OF SQL

SQL was introduced in 1970 and has emerged as the standard language for RDB (Melton, Simon, & Gray, 2001). The 1992 revision, SQL2, has been widely used by all Relational Database Management System (RDBMS) products. In 1993, an attempt to develop a new standard was started since the RDBMS vendors had enhanced their existing relational products with Object-Oriented (OO) features. The existing standard had become somewhat obsolete because it provided no support for OO features.

Many of the vendors created their own language extension of SQL to retrieve and manipulate data such as POSTQUEL (Stonebraker, 1986), Starburst (Lindsay & Haas, 1990), and so forth. These are vendor-specific languages and still there is no standard that can be used and is acceptable to all vendors. For standardization purposes, a new SQL 1999 was developed.

SQL 1999 has been characterized as “OO SQL” and it has become the foundation for several DBMS such as Oracle8. Ironically, many believe that SQL2 will still be used in the future (Elmasri & Navathe, 2002), since many researchers and practitioners still have unsettled arguments on many SQL 1999 issues.

Further, the standardization body ANSI/ISO has started to review SQL 1999 and aims to release a new SQL4 version in few more years. At the time of writing, this version is still an ongoing work and no release date has been announced (Melton, 2002). SQL4 adds some features to SQL 1999, and it also reviews its previous versions.

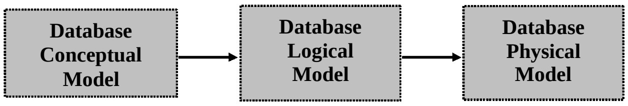
The new SQL has added some new data types to accommodate complex data structures. They are very useful to model real world problems. However, they have a big impact upon the database modeling and implementation.

DATABASE MODELING

Database modeling involves three main phases: conceptual model, logical model, and physical model (see Figure 1).

In the conceptual model level, the database designers capture the database user requirements. To model an RDB, the database designers can use many semantic data modelling such as ER, NIAM, EER, Functional Modeling, and so forth. With SQL4 data type extension, we cannot capture all features using traditional data modeling anymore.

Figure 1. Database modeling phases



The logical model links the conceptual model with the physical implementation. In RDB, this phase is started by transforming the conceptual model into logical design. It is followed by normalization, before we can come up with a set of relations that do not contain anomalies.

In traditional RDB, database designers are already familiar with the transformation methodology (Elmasri & Navathe, 2002), normalization rules (Codd, 1972), and so forth. However, this existing design was developed for simple data types. We do not know whether the old rules can be applicable to the new data types introduced in new SQL4 standard.

The physical model will be different based on how the database is implemented using DBMSs. In this phase, we are required to develop criteria in selecting DBMS. One of the most important criteria is the ability to implement all data type requirements.

For pure RDB that has simple and atomic data type, all DBMS products can be used for the implementation. These products are mainly developed based on SQL2 standard. With the existence of many new data types in SQL4, we need to search the newest DBMS version that can meet the requirement and design.

NEW SQL DATA TYPES

SQL4 classifies the data types into three main classes: predefined, constructed, and user-defined (Melton, 2002). It has few extensions from SQL 1999, but there are a large number of additional data types from SQL2, which is a pure relational model language. Figure 2 illustrates the complete data types supported by current SQL.

Predefined Data Types. Predefined or built-in data types are supported by original SQL. Even though predefined, sometimes the users are still required to determine certain parameters. They are atomic, and therefore very suitable for conventional relational model implementation.

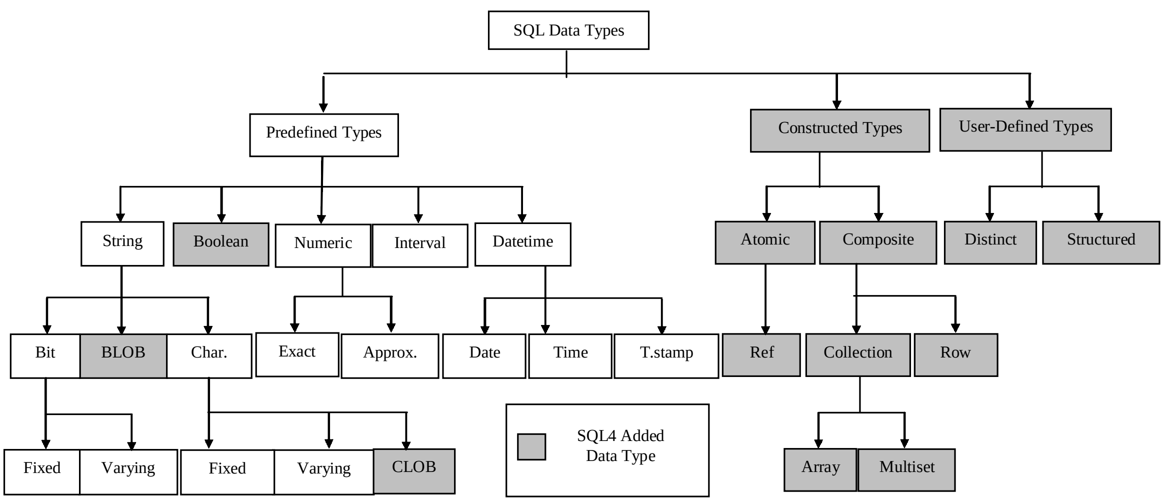
New SQL added Boolean and large object type (BLOB and CLOB). *Boolean* data type enables us to represent the true or false values instead of using a character value with permissible values "T" or "F" (Melton, Simon, & Gray, 2001). *BLOB* can hold a very large binary piece of data such as the digital representation of one's signature. *CLOB* can hold a very large, variably sized, and usually non structured group of characters such as one's resume.

Constructed Data Types. There are two categories of constructed data types: atomic and composite (Melton, 2002; Melton, Simon, & Gray, 2001). From the latest development, SQL4 supports one atomic constructed type and three composite types.

The atomic constructed data type is *REF* type. Its value can be used to address a site holding another value. The site pointed to can be another constructed data type or user-defined type in a typed table.

The first composite data type is *Row*. It contains a sequence of attribute names and their data types. Since it is defined like a flat table, a row type inside a table resembles a nested table. The next data type is *array*, which can hold composite elements of similar type with ordering semantic. Finally, *multiset* contains composite elements that can be duplicated and do not need an

Figure 2. SQL4 data types classification



ordering semantic.

User-Defined Data Types (UDT). This is the third data type provided by SQL4, and has affected the database design most significantly. Based on the number of attributes, UDT can be divided into distinct type and structured type.

Distinct type is formed by a single predefined data type. Its behaviour is different from the base type which it is derived from. The distinguished character of this type is the adherence to strong typing that guarantees correctness and intended type behaviour. *Structured* type is comprised of a number of attributes and routines. It enables users to define and to support the storage and manipulation of complex structure. Besides, structured type can be used to model richer relationships among classes such as inheritance, aggregation, and composition.

IMPACT TO DATABASE MODELING

In this section, we show how the existence of new data types by new SQL has affected the database modeling from conceptual model to the physical model or implementation. It highlights few problems, and thus, brings future research opportunities in the area.

Conceptual Model

It is widely-known that many database designers still use conventional data modeling such as ER for conceptual modeling purpose (Pfleeger, 2001). Furthermore, some even use them for non-relational model database.

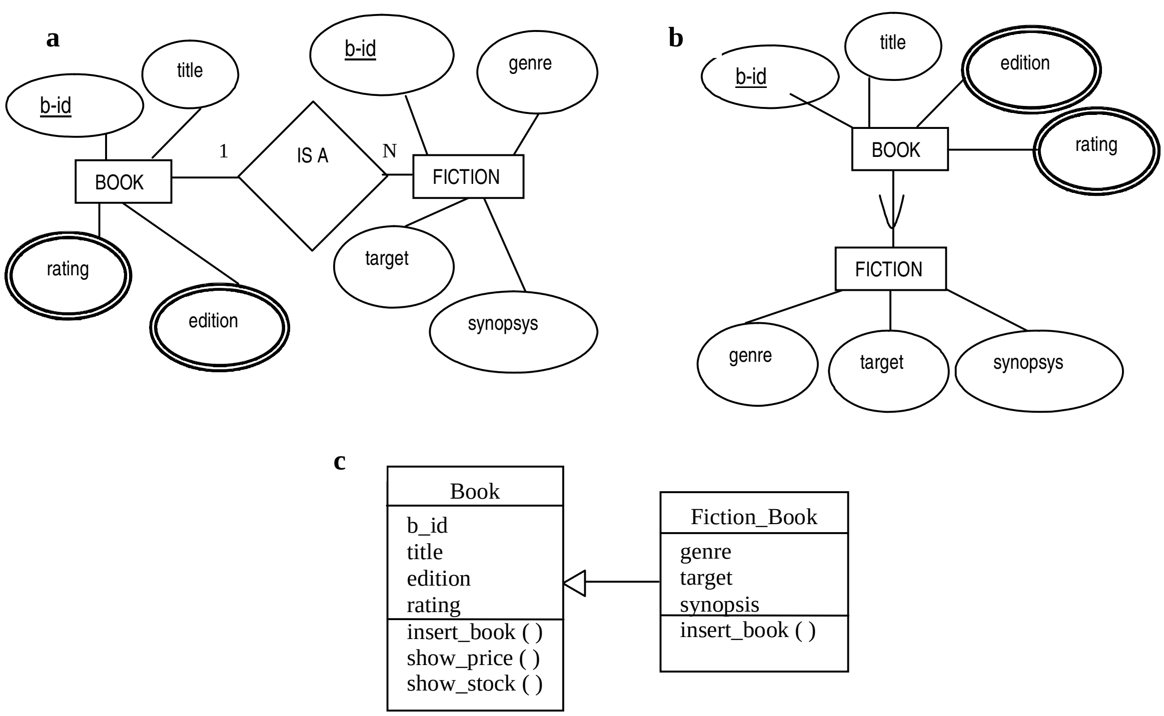
For example, we have two classes Book and Fiction (see Figure 3a) that have a simple 1:N relationship. This conceptual structure can be easily shown using traditional ER modeling. SQL4 recognizes the inheritance structure and now we want to model this relationship between Book and Fiction. Unfortunately, ER is not capable to show this relationship type.

To resolve the problem, we can use the Enhanced Entity Relationship (EER) Diagram (Elmasri & Navathe, 2002) (see Figure 3b). It can accommodate many SQL4 features, but not all of them. EER is applicable only for table or instance. Thus, it cannot include routines, which are embedded in object type.

We use Unified Modeling Language (UML) to resolve the issue (Booch, Rumbaugh, & Jacobson, 1999), since it can represent types and their routines (see Figure 3c). Nevertheless, there are many other SQL4 data type issues that cannot be captured in UML. For example, the representation of *edition* and *rating* attributes as collection types, and so forth.

This section shows that SQL4 affects the database conceptual model. We cannot use traditional data mod-

Figure 3. Conceptual modeling using ER, EER, UML



eling to capture different SQL features, including different new data types.

Logical Model

SQL4 new data types have changed the database design processes such as the transformation methodology and the normalization rules. Using conventional method will lead to an inefficient and incorrect database system. For illustration, we use SQL4 data types, written in bold, in Database Definition Language (DDL) example below.

```
CREATE TABLE Book
(b_id CHARACTER VARYING(5),
 title CHARACTER VARYING(50),
 publisher REF(Publisher_Type),
 author ROW (name CHARACTER_VARYING(20),
 city CHARACTER VARYING(15)),
 edition INTEGER ARRAY[10],
 rating INTEGER MULTISET)
```

REF type violates the conventional transformation method. Formerly, for referential integrity, we link the tables with foreign key (FK). As for the DDL example, a Publisher table is needed before the Book table can refer it. With REF type, we do not have to instantiate an object since this type can refer to an object type. The link is not done through value, but through the address.

Few problems arise with the existence of REF type, and it is a research area that can be extended further. To mention a few, problems are how to define integrity constraint toward types instead of table column, how to use REF type in different level of inheritance structure, and so forth.

Row type is not recognized in conventional model because the relation accepts only atomic values. This type was implemented as ordinary association relationship. With SQL4 row type, we can have a relation inside another relation.

The impact of row type in database design is significant. Some of the issues are: how to design a set of unique row in a tuple since we will need an attribute that take form as the row key; how to design the integrity constraint if the referring or referred attribute is actually a row attribute; how to avoid problems created by transitive, partial, or other undesired dependencies among the row attribute and simple attribute; and so forth.

Next, collection type such as array and multiset are the obvious example of how new SQL4 data types have violated the conventional design. Collection attribute violates the First Normal Form (1NF) of normalization rule (Codd, 1972) due to the existence of repeating groups. The collection would be implemented as separate tables. This practice has been followed although it has diminished the real semantic between the table and its collection attribute. With SQL4 collection types, we are able to store the repeating groups in a relation.

Collection types bring many design issues. Some of them are: how to accommodate unique collection in a tuple; how to transform different type of collection into tables; how to avoid database anomalies created by repeating groups and undesired dependencies brought by collection types; how to maintain the referential integrity if a collection is being referred, and so forth.

This brief section shows that new SQL4 standard has affected the logical model. The full research on each new type is required before we can use them with optimum benefit.

Physical Model

The physical model describes how the data is stored in the database system. The implementation should be committed to the design. For example, the DBMS must be able to

Table 1. DBMS types facilities

ORDBMS	New Data Type							
	Predefined Type		Constructed Type				User-Defined Type	
	Boolean	Large Object	REF	Row	Array	Multiset	Distinct	Structured
Informix SE	Y	Y	N	Y	List	Y	Opaque	Y
Oracle9	Y	Y	Y	Nested	Y	N	Y	Y
IBM DB2	N	Y	Y	N	N	N	Y	Y
UniSQL/X	N	Y	Y	N	List	Y	N	Y
Postgres 7.4	Y	N	N	N	Y	N	N	Y
SYBASE SQL AS 8.0.2	N	Y	N	N	N	N	N	Y

represent the level of abstraction in the design including high-level structures for complex data types and interrelationships.

One criterion of DBMS selection is the compliance with the standard. Table 1 summarizes different DBMS and whether they provide new SQL4 data types (Asiatek, 2003; IBM DB2, 2003; Informix, 2003; Oracle, 2003; Postgresql, 2003; Sybase, 2003).

The completeness of data type is an important issue on selecting DBMS. However, among the ones that facilitate the same data type, we still have to compare the pros and the cons. For example, we have to know how a DBMS implement the index for data type that requires ordering semantic such as array. In this phase, we also need to consider the query performance due to the existence of new type.

FUTURE TRENDS

With the demand of higher RDB capability, it is almost certain that SQL will add more data types in the future. These new data types will affect each database modeling phase as we have shown in this article.

The new SQL data types will enforce the DBMS vendors to comprise with the standard. If the SQL standardization body can keep up with the increasing demand of database community, the vendors will gradually reduce the use of proprietary language to the standard SQL.

Finally, richer RDB will suit more database application. Some applications that have not used RDBMS, in the future, can utilize many new data types provided by new SQL (Pardede, Rahayu & Taniar, 2003).

This article is a preliminary work to a more detailed research. For future work, one needs to investigate each database modeling phase. The processes include the conceptual data modeling, mapping, normalization, integrity constraint, implementation, and so forth, of each different data types.

CONCLUSION

New SQL standard has added many new data types for relational database. The standard impacts upon the whole phase in database modeling processes from conceptual model, logical model, to the physical model or implementation. This article shows the new data types and highlights how they have affected the database modeling processes.

REFERENCES

Asiatek. (2003). The specification of UniSQL (platforms, standards, and open interfaces) Available at *http://www.asiatek.com.tw/eng/unisql/prd_1_3.htm*

Booch, G., Rumbaugh, J., & Jacobson, I. (1999). *The unified modeling language users guide.*

Codd, E.F. (1972). Further normalization of the database relational model. *Data Base Systems*, (pp. 33-64).

Elmasri, R., & Navathe, S.B. (2002). *Fundamentals of database systems*.

Fortier, P. (1999). *SQL3 implementing the SQL foundation standard*.

IBM DB2. (2003).Available at *http://www.software.ibm.com/cgi-bin/db2www/library/*

Informix. (2003). Available at *http://www-3.ibm.com/software/data/informix/*

Lindsay, B.G., & Haas, L.M. (1990). Extensibility in the Starburst experimental database system. *IBM Symposium: Database Systems of the 90s*, (pp.217-248).

Melton, J., Simon, A.R., & Gray, J. (2001). *SQL: 1999 – Understanding relational language components.*

Melton, J. (ed.). (2002). Database language SQL – Part 2 foundation. ISO-ANSI WD 9072-2.

Oracle. (2003). *www.oracle.com*

Pardede, E., Rahayu, J.W., & Taniar, D. (2003). New SQL standard for object-relational database applications. *Proceedings of the 3rd International Conference on Standardization and Innovation in Information Technology*, (pp.191-203).

Pfleeger, S.L. (2001). *Software engineering: Theory and practice*.

Postgresql. (2003). Chapter 8. Data types. Available at *http://developer.postgresql.org/docs/postgres/datatype.html*

Stonebraker, M. (1986). Object management in Postgres using procedures. *Proceedings of the International Workshop on OODBS*, (pp.66-72).

Sybase. (2003). Sybase product manual – SQL anywhere studio. Available at *http://manuals.sybase.com/onlinebooks/*

KEY TERMS

Conceptual Model: A model that is concerned with the real world view and understanding of the data. It suppresses non-critical details in order to emphasize business rules and user objects.

Constructed Data Type: The data type that is formed by a number of predefined data type. This data type is also provided by the software products. The examples are LIST, ARRAY, and so forth.

Database Management System (DBMS): Software used to manage a database. It can be differentiated based on the data model such as Relational DBMS, Object-Oriented DBMS, Object-Relational DBMS, and so forth.

Database Modeling: The first step of database design, where the database designers define the data objects and their relationships to other data objects. Data modeling involves a progression from conceptual model to logical model to physical schema.

Data Type: The characteristic of a data element or a field that specifies what type of data it can hold. SQL classifies three main data types: predefined data type, constructed data type, and user-defined type.

Logical Model: Generalized formal structure in the rules of information science. For relational database, the logical model generally conforms to relational theory.

Physical Model: A logical model instantiated in a specific database management product in a specific installation.

Predefined Data Type: The primitive data type that is built-in by the software products. The examples are INTEGER, CHARACTER, and so forth.

Relational Database: A repository of data in the form of tables that have rows and columns to show the relationship between items. This database was based on relational model invented by E.F. Codd in 1970 (Codd, 1972).

Structured Query Langange (SQL): A standard interactive and programming language for getting information from and updating a database. It is developed by IBM in the mid-1970s for relational database management systems.

User-defined Type: Any data type formed by the users contained of any other data type and usually including the routines.

Newcomer Assimilation in Virtual Team Socialization

Shawn D. Long
University of North Carolina at Charlotte, USA

Gary F. Kohut
University of North Carolina at Charlotte, USA

Gaelle Picherit-Duthler
University of North Carolina at Charlotte, USA

INTRODUCTION

The way we work today is being transformed. Recent technological advances, combined with more flexible job design, have helped increase the number of people working in geographically and/or temporally dispersed environments. Increasing numbers of organizations have employees who are not physically present in the traditional organizations. Hence, more employees are working on teams that seldom, if ever, meet face-to-face. These "teleworkers" have the same work responsibilities as traditional employees, but they have the added challenges of managing or operating within the dynamics of these virtual teams.

Rapid developments in communication technology and the increased globalization of organizations have also greatly accelerated the growth and importance of virtual teams in the workplace. Virtual teams are becoming more commonplace because they are more efficient, less expensive, and less difficult to organize than traditional co-located teams (Larsen & McInerney, 2002; Lurey & Raisinghani, 2001; Piccoli & Ives, 2003). Although there are apparent advantages to organizing work virtually, the challenge for new member integration lies in the fact that team members communicate primarily via electronic mail, telephone, and videoconferencing or computer conferencing. This increased dependence on mediated communication significantly alters the traditional way new members are socialized to work teams. This article addresses the challenges of virtual team socialization with regard to newcomer assimilation. Particular attention is given to newcomer encounter as embedded processes of virtual team assimilation.

Effective communication is key to organizational and team socialization. How well an individual is socialized into a team may determine his or her success within the team, as well as the success of the team in achieving its goals and objectives. Team socialization and the communication practices associated with newcomer integration have been researched extensively (e.g., Brockmann & Anthony, 2002; Lagerstrom & Anderson, 2003) since Jablin (1982) first explored the multi-layered process. Socialization occurs when a newcomer to a team acquires the knowledge, behavior, and attitudes needed to participate fully as a member of that team. Jablin (1987) framed the stages of socialization as anticipatory socialization, organizational assimilation (encounter and metamorphosis), and organizational exit. Although there is an abundance of literature on traditional organizational socialization, research on virtual team socialization is beginning to emerge (Flanagin & Waldeck, 2004; Picherit-Duthler, Long & Kohut, 2004; Ahuja & Galvin, 2003).

BACKGROUND

Organizational assimilation is perhaps the most important stage of virtual team socialization. Such assimilation concerns the process by which individuals become integrated into the culture of an organization (Jablin, 1982). This stage consists of planned and unintentional efforts by the organization to "socialize" employees, and the attempts of organizational members to modify their roles and work environment to better fit their values, attitudes, and needs. As Jablin (1987) noted through the proactive and reactive communication of expectations, organizational roles are negotiated and socially created.

Organizational encounter is a time for newcomers to learn behaviors, values, and beliefs associated with their jobs and organizations (Schein, 1988). As a result of entering a new situation, newcomers want to clarify their situational identity through their work roles (Berlew & Hall, 1966; Feldman, 1976) or through securing approval of others (Graen & Ginsburgh, 1977; Katz, 1978; Wanous, 1980). To reduce uncertainty, newcomers search for information that allows them to adjust by defining the expectations of others and orienting their behavior to the behavior of others.

The speed that virtual teams form demands that workers deal with rapid change. Although research on teamwork suggests that teams function optimally after they have worked together, virtual teams may not have the luxury of establishing working relationships over an extended period of time (e.g., Furst, Blackburn & Rosen, 1999; Mark, 2001). Thus, it is vital for newcomers to quickly establish and develop relationships with others in the work setting, especially with peers and supervisors (Jablin, 2001).

Among other things, organizational relationships provide newcomers with support that facilitates the learning process, and reduces stress and uncertainty associated with adjusting to a new work environment (Jablin, 2001). Much of the research on relationship development in the organizational encounter stage focuses on information seeking and information giving (e.g., Boyd & Taylor, 1998), learning behaviors and attitudes through exchange activities (e.g., Comer, 1991), technical or social information (Comer, 1991; Morrison, 1995), and regulative and normative information (e.g., Galvin & Ahuja, 2001). Evidence suggests that formal and informal socialization practices may affect the level of organizational commitment (Berlew & Hall, 1966; Buchanan, 1974), longevity in the organization (Katz, 1978; Wanous, 1980), and satisfaction and feelings of personal worth (Feldman, 1976).

The next section examines the three central areas of relationship building in virtual teams: peer relationships, supervisory relationships, and mentoring relationships.

Peer Relationships

Anyone who has worked on a team project is immediately concerned with the interactions they might have with their teammates. Do these individuals meet the expectations the team has of them? Are they easy to get along with? Are they competent? Peers help newcomers integrate what may appear to be disjointed pieces of information (Van Maanen, 1984), and may communicate subtle values and norms that may not be well understood by supervisors. Newcomers have more contact with coworkers, and as a consequence, more opportunities to share information with them and develop relationships (Jablin, 2001; Comer, 1991; Teboul, 1994). Sias and Cahill (1998) proposed a variety of contextual factors, including shared tasks and group cohesion (e.g., Fine, 1986), physical proximity (e.g., Griffin & Sparks, 1990), lack of supervisor consideration (Odden & Sias, 1997), and life events outside the workplace, as well as individual factors, such as perceived similarity in attitudes and beliefs as well as demographic similarity (Adkins, Ravlin & Meglino, 1996; Duck, 1994; Glaman, Jones & Rozelle, 1996; Kirchmeyer, 1995), that may affect the development of relationships with peers.

Trust is a key factor in developing close relationships. However, due to the lack of physical proximity and the reliance on communication technologies, our understanding of trust in virtual teams is different from the trust in traditional teams. Meyerson, Weick, and Kramer (1996) coined the term "swift trust" to describe how virtual teams develop a different type of trust than in traditional teams. Due to the highly interdependent task orientation of the team, newcomers develop trust more quickly. Team members are able to develop trust in relationships on the basis of shared tasks rather than on the basis of similar demographics and/or physical proximity (Jarvenpaa & Leidner, 1999).

However, swift trust is not enough to develop close peer relationships. Team members face numerous challenges including: technological mistrust by both newcomers and established members, intuitive fear of the misuse of archived communication (e.g., e-mail trails), and the difficulty of sharing personal or non-work-related issues. Thus, virtual newcomers may be unable or unwilling to take advantage of the informal organizational development that appears central to organizational socialization in traditional teams. This clearly inhibits the development of close peer relationships in virtual teams which in turn may inhibit constructive team cohesion. Similarly, opportunities to understand organizational politics are reduced. Unless the communication among team members is open, power alliances form, allowing certain behaviors to take place such as social loafing, domination, and the formation of cliques. Groups or individuals may be alienated by these behaviors and may differ in their responses based on location or functional role; but the outcome is the same—limited effectiveness of the team, low commitment, low loyalty, and mistrust. Other sources of information such as supervisors and mentors may prove more helpful in recognizing and adapting to political nuances.

Supervisor Relationships

Supervisors are important for assimilating newcomers to organizations by helping build a shared interpretive system that is reflective of assimilation (Berlew & Hall, 1966; Feldman, 1976; Graen, 1976; Kozlowski & Doherty, 1989; Ostroff & Kozlowski, 1992; Schein, 1988). Supervisors frequently communicate with the newcomers, may serve as a role model, filter and interpret formal downward-directed management messages, have positional power to administer rewards and punishments, are a central source of information related to job and organizational expectations as well as feedback on task performance, and are pivotal in the newcomer's ability to negotiate his or her role (Ben-Yoav & Hartman, 1988; Jablin, 2001). According to Staples, Hulland, and Higgins (1998), workers who learned their communication practices by modeling their managers' behaviors had greater self-efficacy, better performance, and more positive job attitudes.

The supervisor-subordinate relationship may be even more important in virtual teams than in traditional teams because of the dislocated nature of the virtual structure. The relationship is complicated by the absence of a physical communication context that characterizes most traditional teams. The supervisor's coordination of virtual team activities may be more difficult because of the distinct nature of technological feedback, the lack of robust spontaneous information exchange between supervisor-subordinate, and the obvious reduction of face-to-face verbal and nonverbal communication cues. On the other hand, some findings suggest that the assessment of team member contributions may be more accurate in virtual rather than face-to-face environments. For example, Weisband and Atwater (1999) found that ratings of liking contributed less bias to evaluations of contribution for virtual groups than face-to-face groups. Similarly, Hedlund, Ilgen, and Hollenbeck (1998) found that leaders of computer-mediated teams were better able to differentiate quality of decisions compared to leaders in face-to-face teams.

It is important for newcomers to develop professional relationships with supervisors, but because virtual teams operate within a more limited timeframe than traditional teams, close relationships leading to friendships become more difficult. This can be a significant problem because the dynamics that help teams become effective require time to develop.

Regardless of whether the supervisor is part of the team or not, the effective supervisor-subordinate relationship will depend in large part on whether the organization uses a traditional approach to managing the virtual team. In traditional teams, supervisor-subordinate relationships are often characterized by hierarchically embedded roles and responsibilities, more formalized rules, procedures, and structures (McPhee & Poole, 2001). In virtual teams, however, there is a loosening of the rules and responsibilities in the supervisor-subordinate relationship. The virtual setting reduces tangible cues that distinguish the status and/or hierarchy of the team members. Thus, the supervisor-subordinate relationships in a virtual team may rely more on co-orientation, focusing on facilitating and supporting the socialization process. Such activities are also comprised in mentoring relationships that have been recognized as important to newcomers' adjustment to socialization efforts.

Mentoring Relationships

When discussing relationship building as part of the assimilation process, mentoring relationships must be considered. Mentors help facilitate newcomer adjustment to situations by offering advice, support, and if appropriate, coaching behaviors to accomplish goals. Wigand and Boster (1991) suggest that "mentoring speeds up socialization into the work role, encourages social interaction, provides an opportunity for high-quality interpersonal interactions, and enhances identification with and commitment to the organization" (p. 16). Mentoring helps team members to generate social capital or the quality of relationships that team members form in their work environments. Due to their expanded boundaries, virtual team members have greater access to a wider range of contexts than would be available in traditional face-to-face teams. For example, Kayworth and Leidner (2002) discovered that highly effective virtual team leaders act in a mentoring role and exhibit a high degree of empathy toward other team members.

Mentoring relationships can be formal and/or informal. Formal mentoring is a "deliberative pairing of a more skilled or experienced person with a lesser skilled or experienced one, with the agreed upon goal of having the lesser skilled or experienced person grow and develop specific competencies" (Murray & Owen, 1991, p. xiv). Several scholars (e.g., Allen, McManus & Russell, 1999; Heimann & Pittenger, 1996; Seibert, 1999) acknowledge that newcomers who participate in formal mentoring relationships in traditional organizations realize greater benefits than those who do not have formal mentoring. Specifically, participation in formal mentoring increases the newcomers' understanding of various organizational issues and increases their level of organizational and job satisfaction.

Informal mentoring relationships develop naturally at the discretion of the mentor and protégé, and exist as long as the parties involved experience sufficient positive outcomes (Jablin, 2001). As opposed to formal mentoring relationships, newcomers who participate in informal mentoring relationships are privileged to information not directly associated with the job role or organizational tasks. These indirect issues include organizational power and politics, more career-related support, "inside" information about various organizational issues and its members that exists on the "grapevine," and increased social interaction outside of the job. Virtual team members who have informal mentors use more direct and less regulative and contractual communication tactics to maintain relational stability than their formal mentored peers (Burke, McKenna & McKeen, 1991; Tepper, 1995). As trust,

Table 1. Guidelines for effective virtual team assimilation

- Organizational issues
  - Develop formal mentoring programs
  - Avoid monitoring electronic communication
  - Provide archive of team history
- Supervisory issues
  - Assist in relationship building
  - Facilitate rather than manage team
  - Encourage team socialization
- Team member issues
  - Actively seek information
  - Develop swift trust based on shared tasks
  - Share personal and non-work related issues

commitment, and identification with the virtual team develops for both the newcomer and more experienced workers, informal mentoring will naturally occur. Virtual teams benefit when barriers to communication are dismantled. This is especially true for newcomers who are uncertain about their roles and the norms of their virtual team.

Recognizing the value of both formal and informal mentoring relationships suggests that virtual teams benefit from such arrangements. Acknowledging the positive impact of mentoring on newcomer assimilation in traditional teams leads us to assume that mentoring will have the same positive impact on virtual team assimilation. However, due to the structural challenges of virtual teams, organizations should consider both formal and informal mentoring programs as tools to socialize newcomers to virtual teams.

In summary, virtual teams face many challenges and issues. The socialization process of team members can become an enigma when building virtual teams. Table 1 summarizes guidelines for organizations, supervisors, and team members to successfully assimilate newcomers into virtual teams.

FUTURE TRENDS

Although we acknowledge that some aspects of virtual team assimilation are similar to traditional team assimilation, they are not identical. What may differ is the speed and the structure of the assimilation process and how newcomers learn to identify with their virtual organization. It would be prudent to explore how newcomers are socialized to use technology and communicate with colleagues and supervisors.

The importance of communication and relationships make virtual teams a challenge for organizations that want to focus more on the outcome rather than the process of producing effective teams. Organizations need to focus more on factors related to internal team processes, such as trust and commitment, rather than outcome assessments such as cost, productivity, and effectiveness.

CONCLUSION

Organizations are turning to virtual teams as a way to remain competitive in the new century's turbulent environment that is characterized by globalization, mergers and acquisitions, and a dependence on information technologies. While great attention has been given to how to provide adequate virtual team infrastructure (e.g., hardware and software), little attention has been devoted to the "human-structure." We assert that how newcomers are assimilated in virtual teams is just as important as the software chosen to accomplish work.

If done correctly, virtual team assimilation may help lead to a strong sense of loyalty and commitment to the team and the organization, which in turn may lead to greater productivity. The responsibility of virtual team socialization resides with the newcomer as well as the organization and the supervisor. Our recommendations for establishing a formal mentoring program, transforming the role of supervisor from manager to facilitator, and supporting proactive behaviors for the newcomer should be identified as high priorities in the development and maintenance of virtual teams. In sum, organizations could obtain a greater return, such as loyalty and commitment, by focusing on newcomer assimilation.

REFERENCES

Adkins, C.L., Ravlin, E.C. & Meglino, B.M. (1996). Value congruence between co-workers and its relationship to work outcomes. *Group & Organization Management, 21*, 439-460.

Ahuja, M.K. & Galvin, J.E. (2003). Socialization in virtual groups. *Journal of Management, 29,* 161-185.

Allen, T.D., McManus, S.E. & Russell, J.E.A. (1999). Newcomer socialization and stress: Formal peer relationships as a source of support. *Journal of Vocational Behavior*, *54*, 453-470.

Ben-Yoav, O. & Hartman, K. (1988). Supervisors' competence and learning of work values and behaviors during organizational entry. *Journal of Social Behavior and Personality, 13,* 23-36.

Berlew, D.E. & Hall, D.T. (1966). The socialization of managers: Effects of expectations on performance. *Administrative Science Quarterly, 11*, 207-223.

Boyd, N.G. & Taylor, R.R. (1998). A developmental approach to the examination of friendship in leader-follower relationships. *Leadership Quarterly, 9*, 1-25.

Brockmann, E.N. & Anthony, W.P. (2002). Tacit knowledge and strategic decision making. *Group and Organization Management*, *27*, 436-455.

Buchanan, B. (1974). Building organizational commitment: The socialization of managers in work organizations. *Administrative Science Quarterly, 19*, 533-546.

Burke, R.J., McKenna, C.S. & McKeen, C.A. (1991). How do mentorships differ from typical supervisory relationships? *Psychological Reports, 68,* 459-466.

Comer, D.R. (1991). Organizational newcomers' acquisition of information from peers. *Management Communication Quarterly, 5*, 64-89.

Duck, S. (1994). *Meaningful relationships: Talking, sense, and relations*. Thousand Oaks, CA: Sage Publications.

Feldman, D.C. (1976). Contingency theory of socialization. *Administrative Science Quarterly, 21*, 433-452.

Fine, G.A. (1986). Friendships in the workplace. In V.J. Derlega & B.A. Winstead (Eds.), *Friendship and social interaction* (pp. 185-206). New York: St. Martin's.

Flanagin, A.J. & Waldeck, J.H. (2004). Technology use and organizational newcomer socialization. *Journal of Business Communication*, *41*, 137-165.

Furst, S., Blackburn, R. & Rosen, B. (1999). Virtual team effectiveness: A proposed research agenda. *Information Systems Journal*, *9*, 249-269.

Galvin, J.E. & Ahuja, M.K. (2001). Am I doing what's expected? New member socialization in virtual groups. In L. Chidambaram & I. Zigurs (Eds.), *Our virtual world: The transformation of work, play and life via technology* (pp. 40-55). Hershey, PA: Idea Group Publishing.

Glaman, J.M., Jones, A.P. & Rozelle, R.M. (1996). The effects of co-worker similarity on the emergence of affect in work teams. *Group & Organization Management*, *21*, 192-215.

Graen, G. (1976). Role-making processes within complex organizations. In M.D. Dunnette (Ed.), *Handbook of industrial/organizational psychology* (pp.1201-1245). Chicago: Rand McNally.

Graen, G. & Ginsburgh, S. (1977). Job resignation as a function of role orientation and leader acceptance: A longitudinal investigation of organizational assimilation. *Organizational Behavior and Human Performance, 19,* 1-17.

Griffin, E. & Sparks, G.G. (1990). Friends forever: A longitudinal exploration of intimacy in same-sex pairs and platonic pairs. *Journal of Social and Personal Relationships, 7,* 29-46.

Hedlund, J., Ilgen, D.R. & Hollenbeck, J.R. (1998). Decision accuracy in computer-mediated versus face-to-face decision-making teams. *Organizational Behavior & Human Decision Performance*, *76*, 30-47.

Heimann, B. & Pittenger, K.K.S. (1996). The impact of formal mentorship on socialization and commitment of newcomers. *Journal of Managerial Issues, 8*, 108-117.

Jablin, F.M. (1982). Organizational communication: An assimilation approach. In M.E. Roloff & C.R. Berger (Eds.), *Social cognition and communication* (pp. 255-286). Beverly Hills, CA: Sage Publications.

Jablin, F.M. (1987). Organizational entry, assimilation and exit. In F.M. Jablin, L.L. Putnam, K.H. Roberts & L.W. Porter. (Eds.), *Handbook of organizational communication: An interdisciplinary perspective* (pp. 679-740). Newbury Park, CA: Sage Publications.

Jablin, F.M. (2001). Organizational entry, assimilation, and disengagement/exit. In F.M. Jablin & L.L. Putnam (Eds.), *The new handbook of organizational communication: Advances in theory, research, and methods* (pp. 732-818). Thousand Oaks, CA: Sage Publications.

Jarvenpaa, S.L. & Leidner, D.E. (1999). Communication and trust in global virtual teams. *Organization Science, 10,* 791-815.

Katz, R. (1978). Job longetivity as a situational factor in job satisfaction. *Administrative Science Quarterly, 23*, 204-223.

Kayworth, T.R. & Leidner, D.E. (2002). Leadership effectiveness in global virtual teams. *Journal of Management Information Systems, 18,* 7-40.

Kirchmeyer, C. (1995). Demographic similarity to the work group: A longitudinal study of managers at the early career stage. *Journal of Organizational Behavior, 16,* 67-83.

Kozlowski, S.W.J. & Doherty, M.L. (1989). Integration of climate and leadership: Examination of a neglected issue. *Journal of Applied Psychology, 74*, 546-553.

Lagerstrom, K. & Anderson, M. (2003). Creating and sharing knowledge within a transnational team—the development of a global business system. *Journal of World Business*, *38,* 84-95.

Larsen, K.R.T. & McInerney, C.R. (2002). Preparing to work in the virtual organization. *Information and Management*, *39*, 445-456.

Lurey, J.S. & Raisinghani, M.S. (2001). An empirical study of best practices in virtual teams. *Information and Management, 38*, 523-544.

Mark, G. (2001). Meeting current challenges for virtually collated teams: Participation, culture, integration. In L. Chidambaram & I. Zigurs (Eds.), *Our virtual world: The transformation of work, play and life via technology* (pp. 74-93). Hershey, PA: Idea Group Publishing.

McPhee, R.D. & Poole, M.S. (2001). Organizational structures and configurations. In F.M.. Jablin & L.L. Putnam (Eds.), *The new handbook of organizational communication: Advances in theory, research, and methods* (pp. 503-542). Thousand Oaks, CA: Sage Publications.

Meyerson, D., Weick, K.E. & Kramer, R.M. (1996). Swift trust and temporary groups. In R.M. Kramer & T.R. Tyler (Eds.), *Trust in organizations: Frontiers of theory and research* (pp. 166-195). Thousand Oaks, CA: Sage Publications.

Morrison, E.W. (1995). Information usefulness and acquisition during organizational encounter. *Management Communication Quarterly, 9,* 131-155.

Murray, M. & Owen, M. (1991). *Beyond the myths and magic of mentoring: How to facilitate an effective mentoring program.* San Francisco: Jossey-Bass.

Odden, C.M. & Sias, P.M. (1997). Peer communication relationships and psychological climate. *Communication Quarterly, 45,* 153-166.

Ostroff, C. & Kozlowski, S.W.J. (1992). Organizational socialization as a learning process: The role of information acquisition. *Personnel Psychology, 45*, 849-87.

Piccoli, G. & Ives, B. (2003). Trust and the unintended effects of behavior control in virtual teams. *MIS Quarterly, 27*, 365-395.

Picherit-Duthler, G., Long, S. & Kohut, G. (2004). Newcomer assimilation in virtual team socialization. In S. Godar & S.P. Ferris (Eds.), *Virtual and collaborative teams: Process, technologies, & practice*. Hershey, PA: Idea Group Publishing.

Schein, E.H. (1988). Organizational socialization and the profession of management. *Sloan Management Review, 30*, 53-65.

Seibert, S. (1999). The effectiveness of facilitated mentoring: A longitudinal quasi-experiment. *Journal of Vocational Behavior, 54,* 483-502.

Sias, P.M. & Cahill, D.J. (1998). From coworkers to friends: The development of peer friendships in the workplace. *Western Journal of Communication, 62*, 273-299.

Staples, D.S., Hulland, J.S. & Higgins, C.A. (1998). A self-efficacy theory explanation for the management of remote workers in virtual organizations. *Journal of Computer-Mediated Communication, 3,* 758-776.

Teboul, J.C.B. (1994). Facing and coping with uncertainty during organizational encounter. *Management Communication Quarterly, 8,* 190-224.

Tepper, B.J. (1995). Upward maintenance tactics in supervisory mentoring and nonmentoring relationships. *Academy of Management Journal*, *38*, 1191-1205.

Van Maanen, J. (1984). Doing new things in old ways: The chains of socialization. In J.L. Bess (Ed.), *College and university organizations* (pp. 211-247). New York: New York University Press.

Wanous, J.P. (1980). *Organization entry: Recruitment, selection, and socialization of newcomers*. Reading, MA: Addison-Wesley.

Weisband, S. & Atwater, L. (1999). Evaluating self and others in electronic and face-to-face groups. *Journal of Applied Psychology, 4*, 632-639.

Wigand, R.T. & Boster, F.S. (1991). Mentoring, social interaction, and commitment: An empirical analysis of a mentoring program. *Communications, 16,* 15-31.

KEY TERMS

Co-Located Team: A traditional team that shares a common goal and works toward that goal in a face-to-face, same office environment.

Formal Mentoring: A deliberate pairing of a more skilled or experienced person with a lesser skilled or experienced one, with the agreed-upon goal of having the lesser skilled or experienced person grow and develop specific competencies.

Informal Mentoring: The non-assigned pairing of an experienced person who respects, guides, protects, sponsors, promotes, and teaches younger, less experienced personnel which develops naturally at the discretion of the mentor and protégé, and persists as long as the parties involved experience sufficient positive outcomes.

Organizational Assimilation: The processes by which individuals become integrated into the culture of an organization.

Organizational Encounter: A time for newcomers to learn behaviors, values, and beliefs associated with their jobs and organizations

Socialization: The acquisition of the knowledge, behavior, and attitude needed to participate fully as a member of a team.

Swift Trust: A type of trust that develops quickly on the basis of shared tasks rather than on the basis of similar demographics or physical proximity.

Virtual Team: A group of geographically and organizationally dispersed workers brought together across time and space through information and communication technologies.

Next–Generation ERP

Charles Møller
Aarhus School of Business, Denmark

INTRODUCTION

"ERP is dead - long live ERP II" is the title of a path breaking research note from Gartner Group (Bond et al., 2000). In this research note Gartner Group envision how the ERP vendors respond to market challenges and how ERP and ERP strategies evolve by 2005. Gartner Group defined ERP II as a transformation of ERP (Enterprise Resource Planning) and today the major vendors have adopted this concept in their contemporary ERP packages.

ERP (Enterprise Resource Planning) is an important concept to industry. Enterprises are increasingly implementing packaged ERP systems. A recent study confirmed that over 90% of the 500 largest Danish enterprises have adopted one or more ERP system. Further, the study found the systems to be of an average age of 2.8 years and decreasing (Møller, Kræmmergaard & Rotbøl, 2003).

ERP is a standardized software packaged designed to integrate the internal value chain of an enterprise (Klaus, Rosemann & Gable, 2000). The five major ERP vendors: (i) SAP; (ii) Oracle; (iii) Peoplesoft; (iv) SAGE; and (v) Microsoft Business Solutions control almost 50% of the ERP market (c.f. table 1) and consequently the corporate infrastructure is dominated by the design of these systems and the vendors.

According to Nah (2002) the American Production and Inventory Control Society (APICS) defines ERP as: "a method for the effective planning and controlling of all the resources needed to take, make, ship and account for customer orders in a manufacturing, distribution or service company". This definition expresses ERP as a tool but ERP is also a management vision and an agency of change and ERP has been attributed almost any good or bad that IT may bring about in business.

In the late 1990's the ERP hype was primarily motivated by companies rushing to prepare for Y2K (Calloway, 2000). Then after a short recession the adoption of ERP has continued. Davenport's sequel on enterprise systems (Davenport, 1998; Davenport, 2000; Davenport & Brooks, 2004) illustrates the changing business perspective on ERP and the ERP hype.

Davenport (1998) sums up the first wave of experiences from implementing ERP systems in a much cited paper on "putting the enterprise system into the enterprise", and points to the new potential business impact of the ERP systems. The discussion evolved over the first enthusiastic expectations, continued over a growing number of horror stories about failed or out-of-control projects, towards a renewed hype of expectations on e-business and SCM.

The ERP II concept is the software industry's perception of the new business challenges and the vision addresses the issues of e-business integration in the supply chain. ERP II is the next-generation ERP concept and in a few years from now the ERP II vision is going to be institutionalized into the infrastructure of most enterprises. This paper will portray the conceptual framework of ERP II.

BACKGROUND: EMERGENCE OF THE ERP CONCEPT

The ERP II concept may be understood by taking a closer look at the development of the ERP concept. Enterprise systems have often been explained through the historical evolution of ERP (Wortmann, 1998; Klaus, Rosemann & Gable, 2000; Chen, 2001). The concept of Enterprise Systems (ES) has evolved over almost fifty years, driven by:

Table 1. Top 5 worldwide ERP software application new license revenue market share estimates for 2002 (Source: Gartner Dataquest, June 2003)

Vendor	*2002 Market Share (%)*	*2001 Market Share (%)*
SAP AG	25.1	24.7
Oracle	7.0	7.9
PeopleSoft	6.5	7.6
SAGE	5.4	4.6
Microsoft Business Solutions	4.9	4.6
Others	51.1	50.3
Total Market Share	100.0	100.0

Table 2. Enterprise systems in retrospective

Decade	Concept	Function
50	Inventory Control Systems (ICS)	Forecast and inventory management
60	Material Requirement Planning (MRP)	Requirement calculations based on Bill-of-Material (BoM)
70	Manufacturing Resource Planning (MRP/II)	Closed-loop planning and capacity constraints
80	Computer Integrated Manufacturing (CIM)	Automation, Enterprise models
90	Enterprise Resource Planning (ERP)	Integrated processes

the changing business requirements, the new information technologies, and by the software vendor's ability to provide standardized solutions.

The fundamental structure of ERP has its origin in the fifties and in the sixties with the introduction of computers into business. The first applications were automating manual tasks such as book-keeping, invoicing and reordering. The early Inventory Control (ICS) systems and Bill of Material (BOM) processors gradually turned into the standardized Material Requirements Planning (MRP). The legacy of the IBM's early COPICS specifications can be found in the structure of the systems even today.

The development continued in the seventies and in the eighties with the MRP II and the CIM concept. During the 1970's MRP caught on like wildfire, and was fueled by the "MRP Crusade" of the American Production and Inventory Control Society (APICS). But gradually industry came to the understanding that neither of these concepts was able to meet the expectations. Even though the CIM ideas failed in many aspects the, the research, e.g. on IS development (ISD) and enterprise models provided the background for gradually integrating more areas into the scope and of the information systems (Wortmann, 2000). This development peaked in early nineties with the advent of the Enterprise Resource Planning (ERP) systems – often embodied in SAP R/3 (Bancroft, 1997) along with the other major vendors: Oracle, Peoplesoft, JD Edwards and Baan – the so called JBOPS. Although the ERP systems have other legacies like accounting, the prevailing planning and control philosophy is deeply rooted in manufacturing and in MRP.

FUTURE TRENDS: NEXT-GENERATION ERP

The ERP market experienced a hype based on the Y2K problem, but after Y2K the ERP market soured. It was doubted that traditional ERP could meet the e-business challenge (Mabert, Soni & Venkataramanan, 2001). New vendors of the "bolt-on" systems, for example, i2 Technology with SCM and Siebel with CRM emerged on the scene (Calloway, 2000) and Application Integration (EAI) became a critical issue (Evgeniou, 2002). New delivery and pricing methods such as ASP (Application Service Provider) and ERP rentals were conceived (Harell, Higgins & Ludwig, 2001) and the traditional ERP vendors were challenged.

The ERP II concept is a vision original conceived by Gartner Group in 2000. Gartner Group, who also put the name on the ERP concept, defines ERP II as, "a business strategy and a set of industry-domain-specific applications that build customer and shareholder value by enabling and optimizing enterprise and inter-enterprise, collaborative-operational and financial processes" (Bond et al., 2000).

ERP II builds on ERP and thus the concept excludes the "bolt-on" vendors like i2 or Siebel from this vision (Mello, 2001). AMR Research does not restrict their competing vision on Enterprise Commerce Management (ECM) to the ERP vendors and define ECM as, "a blueprint that enables clients to plan, manage, and maximize the critical applications, business processes and technologies they need to support employees, customers, and suppliers" (http://www.amrresearch.com/ECM). GartnerGroup has later resigned on this requirement and today ERP II is a framework which includes enterprise systems based on "Best of Breed" systems and EAI (Light, Holland & Willis, 2001) as well as "Single Vendor" solutions.

ERP II includes six elements that touch business, application and technology strategy: (i) the role of ERP II, (ii) its business domain, (iii) the functions addressed within that domain, (iv) the kinds of processes required by those functions, (v) the system architectures that can support those processes, and (vi) the way in which data is handled within those architectures. With the exception of architecture, these ERP II elements represent an expansion of traditional ERP. ERP II is essentially componentized ERP, e-business and collaboration in the supply chain (Bond et al., 2001).

Throughout the ERP industry the new philosophies of e-business was gradually incorporated into the legacy ERP systems, and system architectures were redesigned and modularized, such as what SAP did with their NetWeaver platform. Consequently the contemporary standard systems today incorporate the ERP II vision. The ERP industry survived the challenge and a recent market analysis does not render any signs of market fragmentation, but rather a consolidation. Today we see an ERP market consisting of one dominant actor, a handful of major vendors and a larger number of vendors of minor significance (also c.f. table 1).

Today all the major vendors have adopted the ERP II concept, either partly or to the full extent. The evolution has been driven by the emerging business requirements and by the possibilities offered by the new information technology exactly the same as the evolution of the ERP concept.

The new technologies are not necessarily inventions of the ERP vendors, but rather the technologies emerged as stand alone systems, and after a while they are adopted by the major vendors and then incorporated into the standard systems. For example, that happened to the application frameworks (.NET or J2EE), the databases (Oracle or MS SQL) or Decision-Support Systems (DSS). Business Intelligence (BI) is an example of an analytical DSS technology previously associated with add-ons, like Data warehouse systems based on OLAP tools, that are now integrated into to core of the standard databases. BI refers to a broad category of analytical applications that helps companies make decision based on the data in their ERP systems. Another example is the internet standards like XML, originally conceived outside the control of the major vendors but gradually adopted into the infrastructure of the ERP systems. Other examples are the Supply Chain Management (SCM) systems or the Customer Relationship Management (CRM) systems from third-party vendors like i2 or Siebel. Those third-party vendors experienced a short explosive growth, but then when the technologies are incorporated into the standard ERP system the potential business benefit increases.

There is an emerging pattern of stable generic application architectures which we choose to portray as the ERP II concept. Calloway (2000) and Weston (2003) have attempted to frame this overall development and partial aspects have further been dealt with, for example. by Wortmann (2000) and Møller (2003). The conceptual framework of ERP II illustrated in figure 1 consists of four distinct layers as exhibited in table 3: (i) the foundation layer, (ii) the process layer; (iii) the analytical layer and, (iv) the e-business layer consisting of the collaborative components.

Core Components

The foundation layer consists of the core components of ERP II. The core components shape the underlying architecture and provide a platform for the ERP II systems. ERP II does not need to be centralized or monolithic. One of the core components is the integrated database, which may be a distributed database. Another core component is the application framework likewise potential distributed. The integrated database and the application framework provide an open and distributed platform for ERP II.

Central Components

The process layer of the concept is the central components of ERP II and this layer reflects the traditional transaction based systems. ERP II is web based open and componentized. This is different from being web-enabled, and the ultimate ERP II concept may be implemented as a set of distributed web services.

Enterprise Resource Planning (ERP) is one of the central components in the ERP II conceptual framework. The backbone of ERP is the traditional ERP modules like

Table 3. Four layers of ERP II

Layer	*Components*	
Foundation	Core	Integrated Database (DB) Application Framework (AF)
Process	Central	Enterprise Resource Planning (ERP) Business Process Management (BPM)
Analytical	Corporate	Supply Chain Management (SCM) Customer Relationship Management (CRM) Supplier Relationship Management (SRM) Product Lifecycle Management (PLM) Employee Lifecycle Management (ELM) Corporate Performance Management (CPM)
Portal	Collaborative	Business-to-consumer (B2C) Business-to-business (B2B) Business-to-employee (B2E) Enterprise Application Integration (EAI)

financials, sales and distribution, logistics, manufacturing, or HR. ERP still makes up the backbone of ERP II along with the additional integrated modules aimed at new business sectors outside the manufacturing industries.

The ERP II concept is based on Business Process Management (BPM). ERP has been based on "best-practice" process reference models, but ERP II systems build on the notion of the process as the central entity and ERP II includes tools to manage processes: design (or orchestrate) processes, and to execute and to evaluate processes (Business Activity Monitoring). In ERP II redesigning processes will have effect in real-time.

The BPM component allows for ERP II to be accommodated to suit different business practices for specific business segments that otherwise would require problematic customization. ERP II further includes vertical solutions for specific segments like apparel and footwear or the public sector. Vertical solutions are sets of standardized pre-configured systems and processes with "add-ons" to match the specific requirements in , for example, a business sector.

Corporate Components

The analytical layer consists of the corporate components that extend and enhance the central ERP functions by providing decision support to manage relations and corporate issues. Corporate components are not necessarily synchronized with the integrated database and the components may easily be "add-ons" acquired by third-party vendors. The most common components such as Supply Chain Management (SCM) systems and Customer Relationship Management (CRM) systems are listed in table 3, illustrated in figure 1, and they will further be explained under the terms and definitions in the end of this paper.

In the future we can expect new breeds of corporate components. Product Lifecycle Management is already established as the R&D equivalence to ERP and Employee Lifecycle Management is emerging as an example of a people oriented component.

Collaborative Components

The e-business layer is the portal of the ERP II systems and this layer consists of a set of collaborative components. The collaborative components deal with the communication and the integration between the corporate ERP II system and actors like customer, business partners, employees, and even external systems.

The most common and generic components are listed in table 3 , illustrated in figure 1 and will further explained under the terms and definitions at the end of this paper.

CONCLUSION

ERP II is a vision of the next-generation ERP and the conceptual framework for ERP II is a generic model of the emerging architecture of the contemporary enterprise systems. ERP II is aimed at extending the reach of integration into the supply chain and the business benefits of the systems are only realized when the integration occurs (Davenport, Harris & Cantrell, 2004). Business managers will therefore need to consider their entire range of enterprise systems into a supply chain integration context, and future research will deal with inter-organizational integration based on the ERP II.

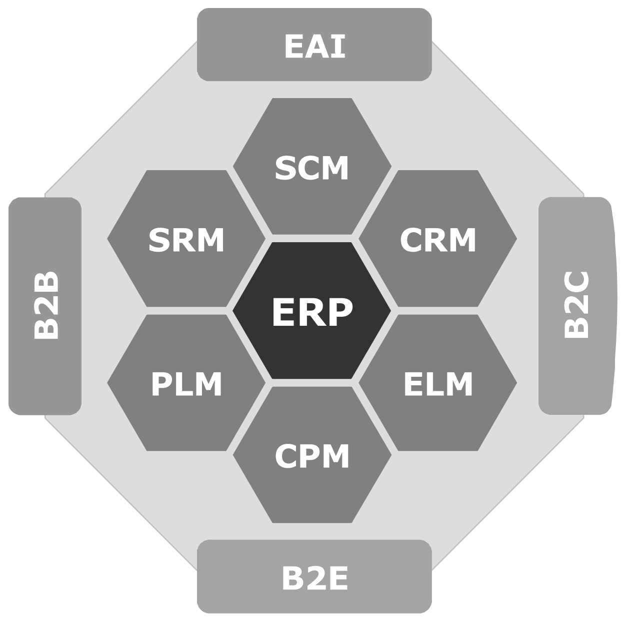
REFERENCES

Bancroft, N. H., Seip, H., & Sprengel, A. (1997). *Implementing SAP R/3: How to introduce a large system into a large organization*. Greenwich: Manning.

Bond, B., Genovese, Y., Miklovic, D., Wood, N., Zrimsek, B., & Rayner, N. (2000). *ERP is dead - Long Live ERP II* (No. Strategic Planning SPA-12-0420). GartnerGroup.

Callaway, E. (2000). *ERP - The next generation: ERP is WEB Enabled for E-business*. South Carolina: Computer Technology Research Corporation.

Figure 1. Conceptual framework for ERP II



Chen, I. J. (2001). Planning for ERP systems: Analysis and future trend. *Business Process Management Journal, 7*(5), 374.

Davenport, T. H. (1998, *July/August*). Putting the enterprise into the enterprise system. *Harvard Business Review*, 121-131.

Davenport, T. H. (2000). The future of enterprise system-enabled organizations. *Information Systems Frontiers, 2*(2), 163-180.

Davenport, T. H., Harris, J. G., & Cantrell, S. (2004). Enterprise systems and ongoing process change. *Business Process Management Journal, 10*(1), 16-26.

Davenport, T. H., & Brooks, J. D. (2004). Enterprise systems and the supply chain. *Journal of Enterprise Information Management, 17*(1), 8-19.

Evgeniou, T. (2002). Information integration and information strategies for adaptive enterprises. *European Management Journal, 20*(5), 486-494.

Harrell, H. W., Higgins, L., & Ludwig, S. E. (2001). Expanding ERP application software: Buy, lease, outsource, or write your own? *Journal of Corporate Accounting & Finance, 12*(5).

Klaus, H., Rosemann, M., & Gable, G. G. (2000). What is ERP? *Information Systems Frontiers, 2*(2), 141-162.

Light, B., Holland, C. P., & Wills, K. (2001). ERP and best of breed: A comparative analysis. *Business Process Management Journal, 7*(3), 216.

Mabert, V. A., Soni, A., & Venkataramanan, M. A. (2001). Enterprise resource planning: Common myths versus evolving reality. *Business Horizons, 44*(3), 69-76.

Mello, A. (2001, October 25). Battle of the labels: ERP II vs. ECM. *ZD TECH Update*.

Møller, C., Kræmmergaard, P., & Rotbøl, M. (2003). *Virksomhedssystemer i Danmark 2003 - En analyse af de 500 største danske virksomheders ERP systemer* (IFI working paper series no. 126 No. ISSN no. 1398-067X). Aarhus: Department of Information Science.

Møller, C. (2003). *ERP II extended enterprise resource planning*. Paper presented at the Seventh World Multi-Conference on Systemics, Cybernetics and Informatics, Orlando, US.

Nah, F. F.-H. (Ed.). (2002). *Enterprise resource planning solutions and management*. Hershey, PA: IRM Press.

Weston, E. C. T. (2003, November/December). ERP II: The extended enterprise system. *Business Horizons, 46*, 49-55.

Wortmann, J. C. (1998). Evolution of ERP systems. In U. S. Bititchi & A. S. Carrie (Eds.), *Strategic management of the manufacturing value chain*. Kluwer Academic Publishers.

KEY TERMS

Business to Business (B2B): E-procurement systems improves the efficiency of the procurement process by automating and decentralizing the procurement process. The traditional methods of sending Request for Quotes (RFQ) documents and obtaining invoices etc. are carried out over the web through purchasing mechanisms such as auctions or other electronic marketplace functions, including catalogues.

Business to Employee (B2E): Intranets or knowledge management systems provide the employee with an updated personalized portal to the enterprise on his desktop. The perspectives of the intranet and knowledge management systems increase in the context of the ERP II concept.

Consumer to Business (B2C): E-commerce systems deals with the carrying out of commercial transactions with businesses or with individual customers by using the Internet as an electronic medium. This requires an extensive infrastructure of which the main features are a catalogue, online ordering facilities and status checking facilities.

Corporate Performance Management (CPM): Sometimes Enterprise Performance Management (EPM) is an umbrella term that describes the methodologies, metrics, processes and systems used to monitor and manage an enterprise's business performance. Thus CPM provides the managements with an overall perspective on the business.

Customer Relationship Management (CRM): Systems facilitate the managing of a broad set of functions relating to managing customers relations that primarily include the categories of customer identification process and customer service management.

Enterprise Application Integration (EAI): Extranets provide the ERP II system with a portal and a platform for integration with other systems inside or outside the corporation. EAI provides the support for automating processes across various IT platforms, systems and organizations.

Employee Lifecycle Management (ELM): The integration of all aspects of information and knowledge in relation to an employee from the hiring to the retirement from the company. ELM enables enterprises to effectively manage their portfolio of competencies.

Enterprise Resource Planning (ERP): II systems are second generation ERP systems. ERP II extends on the ERP concept (Figure 1). The ERP II vision is framed by Gartner Group and in practice defined by the contemporary systems of the major ERP vendors.

Product Lifecycle Management (PLM): Including Product Data Management (PDM) enables enterprises to bring innovative and profitable products to market more effectively, especially in the evolving e-business environment. PLM enables enterprises to harness their innovation process through effective management of the full product definition lifecycle in their extended enterprises.

Supplier Relationship Management (SRM): The vendor side analogy to CRM aimed at the effective management of the supplier base. SRM facilitates the management of the supplier relations in its entire life-cycle.

Supply Chain Management (SCM): Systems provide information that assist in planning and managing the production of goods. For instance, SCM assists in answering questions such as where the good is to be produced, from which the parts are to be procured and by when it is to be delivered.

Non–Functional Requirements and UML Stereotypes

Guadalupe Salazar-Zárate
Technical University of Catalonia, Spain

Pere Botella
Technical University of Catalonia, Spain

INTRODUCTION

In Nuseibeh and Easterbrook (2000), an overview of the field of software and systems requirements engineering is presented. Therein is highlighted some key open-research issues for the future of the Requirements Engineering (RE). Some of the major challenges mentioned there, are related with the necessity of richer models for capturing and analyzing non-functional requirements. This paper draws some possible extensions of Unified Modeling Language (UML) (Booch, G., Jacobson, I. and Rumbaugh, J., 1998) in order to include non-functional requirements.

BACKGROUND

Within the Requirements Engineering processes (e.g., domain analysis, elicitation, modeling, validation, etc.), it is common to distinguish between functional and non-functional requirements. The relevance of functional requirements has been traditionally well-covered by the existing modeling techniques, where a lot of research has been done. However, non-functional requirements (NFRs for short), quality and constraint properties, are not usually covered by these modeling techniques. On the contrary of its functional counterpart and despite the critical role they have during system development, they have received little attention in the literature as mentioned by Chung, L., Nixon, B.A., Yu, E. and Mylopoulos, J. (2000). How can we model and reason about NFRs? How can these models be integrated with those modeling techniques? These are still some of the key challenges that need more research.

Functional and non-functional aspects regarding the external system behavior involve two different ways of evaluating and/or developing a given software system. On one hand, functional aspects are directly connected to *what the system does*, for example, the basic functions that a system (or a system component) must provide. On the other hand, non-functional aspects are related with how the system behaves with respect to some observable attributes such as performance, reliability, efficiency, reusability, portability, maintainability (i.e., some software qualities). To illustrate the wide variety of issues for software quality, in Chung et al. (2000, p.160) a list of non-functional requirements can be found.

There are two basic approaches to characterize non-functional requirements: *Product-oriented* and *Process-oriented* (Chung et al., 2000). The *product-oriented approach* basically focuses on the development of a formal framework so that a software product can be evaluated in relation to the highest degree it fulfills for non-functional requirements (constraints over non-functional properties). The *process-oriented approach* uses non-functional information to guide the development of software systems. Among the works dealing with this perspective of non-functionality, those of Chung et al. (2000) are without any doubt the most complete ones. Therein a NFR framework, to deal with diverse non-functional requirements to drive the design by justifying decisions during the software development process, is described. The framework also offers structured, graphical facilities for stating, managing, and inter-relating non-functional requirements while justifying decisions and determining their impact throughout the development process.

In Cysneiros and Leite (2001a, 2001b), and Cysneiros, L.M., Leite, J.C.S.P. and Neto, J.S.M. (2001) an approach that complements the work reported in Chung et al. (2000) is presented. Strategies can be found there that are concerned with the problem of how to identify and integrate non-functional requirements with functional requirements, in a process-oriented approach and using UML.

Another important language that focuses on non-functional requirements is the Goal-oriented Requirement Language (GRL) (ITU-T, URN Focus Group, 2002b), (http://www.cs.toronto.edu/km/GRL/). The GRL graphical language is used to support goal and agent-oriented modeling and reasoning of requirements. The GRL is built on the well-established NFR Framework (used for modeling NFRs) and the agent-oriented language i* (Yu, 1997) (used for the modeling, analysis, and reengineering of organizations and business processes).

OUR APPROACH TO NON-FUNCTIONAL REQUIREMENTS AND UML

UML has today become a standard modeling language for software systems development (Object Management Group, 2001). UML offers a graphical notation to create models. However, it is mainly focused on functional aspects of the software development.

In the following approach we outline some possible extensions of UML in order to include non-functional requirements. In the works by Botella, P., Burgués, X., Franch, X., Huerta, M. and Salazar, G. (2002), and Salazar-Zárate, G, Botella, P. and Dahanajake A. (2003), more detailed considerations about this topic can be found.

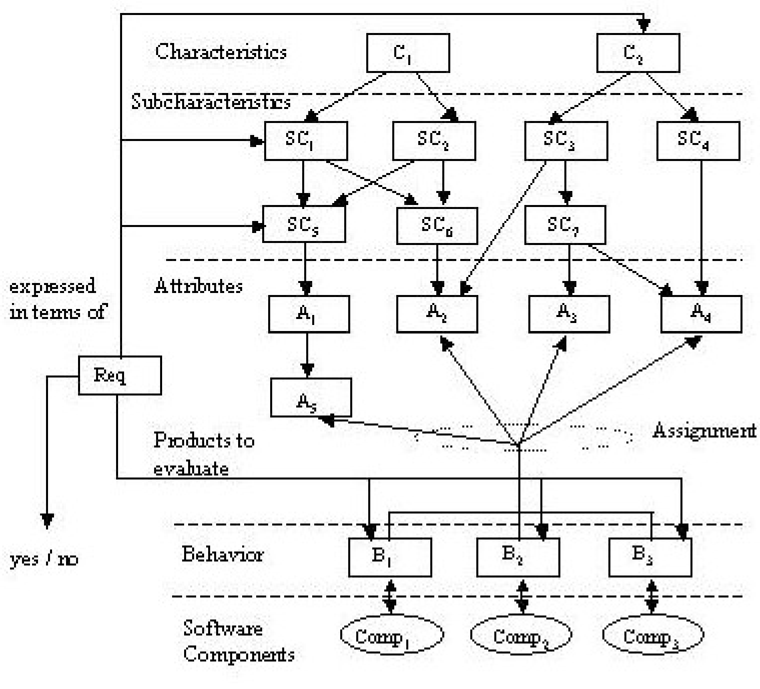
The International Standard ISO/EC 9126-1 (2001) can be used as a starting point to identify non-functional attributes of products that are potentially relevant to be modeled in a software development process. This standard provides a framework for software product quality, specifying quality characteristics to be used in the quality evaluation of software products. A *quality model* is defined by means of general *characteristics of software*, which are further refined into *subcharacteristics* in a multilevel hierarchy. In the standard, the characteristics of functionality, reliability, usability, efficiency, maintainability, and portability are being placed at the top of the hierarchy. Measurable *software attributes* appear at the bottom of the hierarchy. *Software quality metrics* allows developers to quantify up to what degree a software system meets non-functional requirements.

A language called *NoFun* (acronym for "NOn-FUNctional") is a notation that focuses on representing non-functional aspects of software systems, at the product level within the component-programming framework (Franch, 1998; Burgués & Franch, 2000; Botella et al., 2002). It is a formal language for description of software quality requirements using the ISO/IEC 9126 standard to summarize quality characteristics (see figure 1). Although the ISO/IEC 9126-1 (2001) replaces the previous version of 1991, the basic principles stay and we are still able to use the same layout therein described.

To achieve the goal of formalization within *NoFun*, three different kinds of capabilities are provided. First, there are modules for defining the different kind of concepts in the standard (characteristics, subcharacteristics and attributes). Second, values for these attributes may be given (in *behavior modules*) and bound to particular software components (i.e., the ones under evaluation). Third, additional constructions for representing quality requirements and assessment criteria are included.

In figure 1, an example of distribution of a quality model into modules is shown (extracted from Botella et al. (2002). Two characteristics defined in terms of four subcharacteristics appear in the upper part of the figure. Behavioral modules (denoted by the B_i), are abstractions of software components in the sense that they contain all the relevant information for quality evaluation. Quality requirements may be defined as restricting the values of

Figure 1. Layout of a quality model under NoFun language.



the quality entities. They are stated using operators over the quality entities, and these requirements may be categorized according to their importance (for more details see Botella et al., 2002).

In addition to those elements, an orthogonal concept is the one of refinement. Refinement allows the definition of quality models in an incremental manner, by specializing the more general ones. This kind of inheritance-like relationship is used to construct a structured representation of quality. As a result, quality models can be formulated first in a general way, and later refined in particular domains.

Using the formalizing ideas of the *NoFun* language and following a product oriented approach; we aim to extend the UML language (Booch et al., 1998) in order to capture non-functional information.

APPLYING UML STEREOTYPES FOR NFRS

Regarding non-functional information, UML offers notes to incorporate those aspects. UML notes may contain any information; and because of the different domains, types, and nature of the NF-information therein listed, specifying NF-attributes in this way may be insufficient if we want a systematic representation of the NF-information. Therefore, it is necessary to put this information in a more ordered form. One way to identify those kinds of attributes for visualizing non-functionality in UML diagrams is through the use of the layout described above with the *NoFun* language.

In order to incorporate non-functional requirements to UML, the following considerations have been devised. First, we propose to make use of stereotypes to represent the non-functionality. UML stereotypes allow us to create new kinds of building blocks that can be adapted to our specific problem. Second, those software quality characteristics and attributes defined in a layout as shown in figure 1 can be identified and some ad-hoc stereotypes can be also defined. Organizing collections of quality characteristics in this way gives us a generic language concept that can be instantiated later. There are various levels of granularity and, consequently, NF-information can be applied to a whole system, package, component or class in particular. Non-functional requirements tend to be properties of a system as a whole, but our approach attempts to be flexible in order to attach NF-information at every abstraction level, including a single component.

UML stereotypes to include non-functionality can be used in several ways:

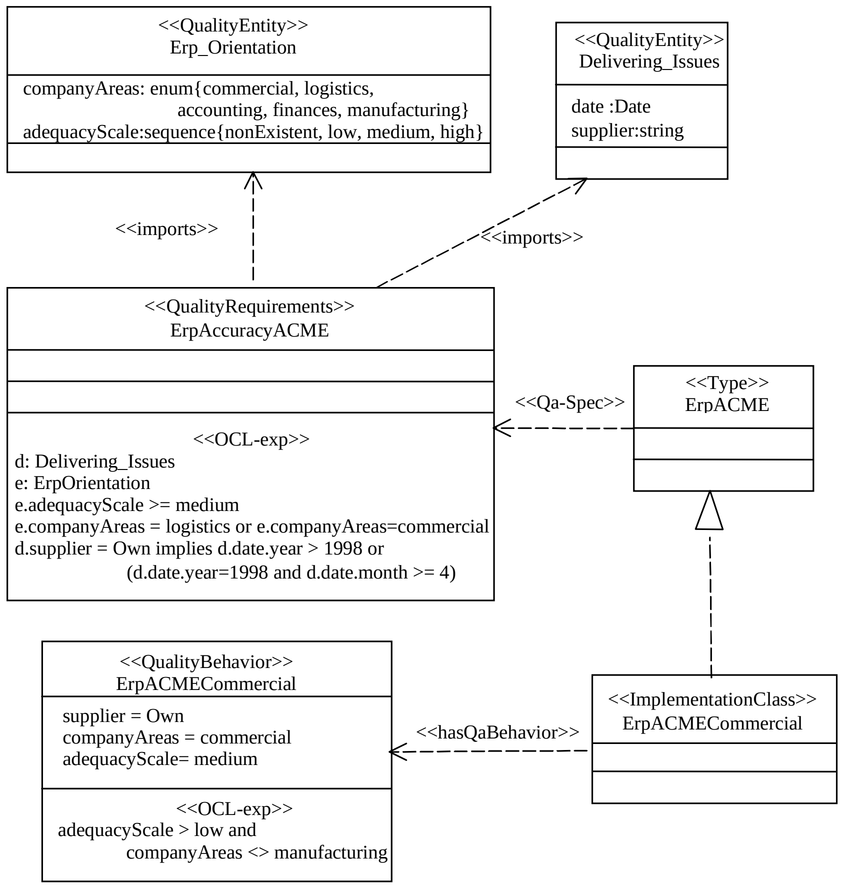
- In a *Class diagram*. At a conceptual level, one can create stereotyped classes <<NF-Entity>> to represent a specific quality characteristic. New stereotypes can be formulated and then refined into more specialized ones. In a more detailed level, the Object Constraint Language (OCL) (Warmer & Kleppe, 1999), or an adaptation of the NoFun language can be used to establish requirements in the form of constraints.
- In UML *packages*. One can group software elements under a stereotyped package, for example, to represent a collection of non-functional stereotyped classes. In this way, a framework for further design specifications is provided.
- In *Use Cases*. The non-functionalities can be seen as transformation rules causing new use cases to be added and/or the structure of the use case description to be changed or replaced so that the non-functional requirements are satisfied. New actors may also be identified in the transformation process. By adding stereotypes such as <<NF-uses>> or <<NF-extends>>, one can identify the use cases that involve non-functionalities. Constraints can be attached in a special script using OCL or NoFun.

To identify the non-functional concepts, the main idea is to use software components in an object-oriented approach. Hereafter, the software components can be seen as entities consisting in the definition of a class diagram that states both functional and non-functional characteristics. According to the analysis model proposed by Larman (1998), this approach focused initially on the conceptual model by using UML class diagrams. At a later stage, other models, such as essential use cases, system sequence diagrams (behavior), and the state diagrams, will be considered.

In a UML context, the functional requirements of the system are partially captured in a *Class Diagram* containing the classes and relationships that represents the key abstractions of the system under development. Figure 2 shows a particular case where non-functionality is applied to a whole system as represented with a single class. Following the organization described by Franch, X., Botella, P., Burgués, X. and Ribó, J.M. (1997), the functional specification part of a software system can be distinguished from its *implementation* part. In this way, it is possible to state that a specification module may have several different implementations. It is also possible to select the best implementation for a given specification according to the non-functional information. In Figure 2, it is represented by using the predefined UML stereotypes <<Type>> and <<ImplementationClass>>.

Stereotyped classes <<QualityEntity >> can be used to represent the modules for characteristics, subcharacteristics, and attributes. As a result, non-func-

Figure 2. Example showing non-functional stereotypes.



tional attributes are defined in this stereotyped class. In this way, one can create all of the stereotyped classes needed to describe the non-functionality. A kind of hierarchy between these attributes can be simulated with dependency relationships between the classes (<<imports>> indicates that the dependent class may import one or more modules of non-functional entities). In order to state constraints over the non-functional attributes, an extra compartment (titled <<OCL-exp>>) can be added to the stereotyped class.

A stereotype <<QualityRequirements>> can be declared for restricting the values of the quality entities. The stereotype <<QualityBehavior>> will define the *non-functional behavior* of a component implementation.

The use of UML stereotypes as a mechanism to incorporate non-functional elements has been outlined. It has been accomplished by describing new UML stereotypes for the class diagrams, package diagrams and use cases. In class diagrams the new stereotypes have been used to represent quality entities, in package diagrams they were used to represent collections of quality entities, and in use cases they may be applied to identify the non-functional information.

FUTURE TRENDS

The work here presented was motivated within the scope of the construction of quality models for software domains of our research group (http://www.lsi.upc.es/~gessi/). The work started by the group of Toronto (Mylopoulos, J., Chung, L. and Nixon, B.A., 1992; Chung & Nixon, 1995; Chung et al., 2000) is consolidated nowadays as a predominant research line in describing non-functional requirements. Both the UML approach of the work of Cysneiros and Leite (2001b), mentioned above, and the Goal-oriented Requirement Language (GRL) (ITU-T, URN Focus Group, 2002b) (http://www.cs.toronto.edu/km/GRL/) also mentioned before, are included under this line.

Currently there here is a project to create a language standardization to express user requirements in the scope of the telecommunications, denominated User Requirements Notation (URN) (ITU-T, URN Focus Group, 2002a), (http://www.usecasemaps.org/urn/urn-info.shtml). In this project, GRL is included for the expression of the non-functional requirements. The standardization group is mainly composed of researchers and practitioners from industry (Mitel, Nortel, KLOCwork, Rational, Telelogic, Phillips, and others) and academia (Carleton Univ. and Univ. of Ottawa for UCMs, Univ. of Toronto for GRL, and others).

CONCLUSION

Our approach presented here can be considered as a first attempt to include NFRs and distinguishing them from the FRs in the UML models and under the characteristic/subcharacteristic concepts of the standard ISO 9126-1.

Mainly using quality models (as defined by the ISO 9126-1) as a reference framework for the definition of nonfunctional requirements. As an example in Botella et al., (2003) we show the use of quality requirements based on quality models for the selection of ERP systems.

There are still many challenges to address the NFRs, and more strategies and tools are needed in order to offer alternatives to capture and analyze NFRs.

REFERENCES

Booch, G., Jacobson, I. & Rumbaugh, J. (1998). *The unified modeling language users guide*. Addison-Wesley Object Technology Series. Addison Wesley Longman, Inc.

Botella, P., Burgués, X., Franch, X., Huerta, M. & Salazar, G. (2002). Modeling non-functional requirements. In A. Durán & M. Toro (Eds.), *Applying requirements engineering* (pp. 13-33). Seville, Spain: Imprenta Catedral S.L.

Botella, P., Burgués, X., Carvallo, J.P., Franch, X., Pastor, J.A., & Quer, C. (2003). Towards a quality model for the selection of ERP systems. In A. Cechich, M. Piattini, & A. Vallecillo (Eds.), *Component-based software quality: Methods and technique* (pp. 225- 246). LNCS 2693: Springer-Verlag.

Burgués, X. & Franch, X. (2000, October). A language for stating component quality. *Proceedings of 14th Brazilian Symposium on Software Engineering (SBES), Joao Pressoa, Brasil, 69-84.*

Chung, L. & Nixon, B. A. (1995, April). Dealing with non-functional requirements: Three experimental studies of a process-oriented approach. In *Proceedings of the 17th International Conference on Software Engineering (ICSE 95), Seattle, WA, USA*, 25-36.

Chung, L., Nixon, B.A.,Yu, E. & Mylopoulos, J. (2000). *Non-functional requirements in software engineering*. Kluwer Academic Publishers.

Cysneiros, L.M., Leite, J.C.S.P. & Neto, J.S.M. (2001, April). A framework for integrating non-functional requirements into conceptual models. *Requirements Engineering Journal. 6*(2), 97-115.

Cysneiros, L.M. & Leite, J.C.S.P. (2001a, October). Driving non-functional requirements to use cases and scenarios. In *XV Simposio Brasileiro de Engenharia de Software*, Brazil, 7-20.

Cysneiros, L.M. & Leite, J.C.S.P. (2001b). Using UML to reflect non-functional requirements. In *Proceedings of the CASCON 2001, Toronto, Canada.*

Franch, X., Botella, P., Burgués, X. & Ribó, J.M. (1997, June). ComProLab: A component programming laboratory. In *Proceedings of 9th Software Engineering and Knowledge Engineering Conference (SEKE), Madrid, Spain*, 397-406.

Franch, X. (1998). Systematic formulation of non-functional characteristics of software. In *Proceedings of 3rd International Conference on Requirements Engineering (ICRE)*. Colorado Springs, USA: IEEE Computer Society, 174-181.

International Standard. (2001). ISO/IEC 9126-1 Software engineering–Product quality–Part 1: Quality Model, International Organization for Standardization. Geneva.

ITU-T, URN Focus Group. (2002a, February). Draft Rec. Z.150–User Requirements Notation (URN). Geneva. Available from URN Focus Group Web site, *http://www.UseCaseMaps.org/urn*

ITU-T, URN Focus Group (2002b), Draft Rec. Z.151 – Goal-oriented requirements language (GRL). Geneva. Available from URN Focus Group Web site, *http://www.usecasemaps.org/urn/urn-meetings.shtml#latest*

Larman C. (1998). *Applying UML and patterns. An introduction to object-oriented Analisys and Design*. New York: Prentice Hall, Inc..

Mylopoulos, J., Chung, L., & Nixon, B.A. (1992, June). Representing and using nonfunctional requirements; A process-oriented approach. *IEEE Transactions on Software Engineering, 18*(6), 483-497.

Nuseibeh, B. & Easterbrook, S. (2000, June). Requirements engineering: A roadmap. *Proceedings of International Conference on Software Engineering (ICSE-2000)*, Limerick, Ireland, ACM Press, 4-11.

Object Management Group (2001). *Unified modeling language specification, Version 1.4*. [Technical Report]. Available from Object Management Group Web site, *http://www.omg.org/technology/documents/formal/unified_modeling_language.htm*

Robertson S. & Robertson J. (1999). *Mastering the requirements process*. UK: Addison-Wesley.

Salazar-Zárate, G, Botella, P. & Dahanajake A.(2003). Introducing non-functional requirements in UML. In L. Favre (Ed.), *UML and the unified process* (pp. 116-128). Hershey, PA: Idea Group Inc.

Thayer, R. & Dorfman M. (Eds.). (1990). *System and software requirements engineering*. IEEE Computer Society Press.

Warmer, J. B. & Kleppe A. G. (1999). *The object constraint language: Precise modeling with Uml*. Reading, MA: Addison Wesley Longman, Inc.

Yu E. (1997, January). Towards modeling and reasoning support for early-phase requirements engineering. *Proceedings of the 3rd IEEE International Symposium on Requirements Engineering (RE'97), Washington D.C., USA*., 226-235.

Zave, P. (1997). Classification of research efforts in requirements engineering. *ACM Computing Surveys 29*(4), 315-321.

KEY TERMS:

i*: The i* modeling framework (pronounce eye-star) has been developed at the University of Toronto by Yu, Mylopoulos and Chung. It was developed for modeling and reasoning about organizational environments and their information systems (Yu, 1997, p.227)

ITU-T: The ITU Telecommunication Standardization Sector (ITU-T) is one of the three Sectors of the International Telecommunication Union (ITU). The function of ITU-T is to provide global telecommunication standards by studying technical, operating and tariff questions. The results of these studies are published as ITU-T Recommendations. (*http://www.itu.int/ITU-T/index.html*).

NFR: Non-functional requirement – in software system engineering, a software requirement that describes not what the system will do, but how the software will do it, for example, software performance requirements, software external interface requirements, software design constraints, and software quality attributes (Chung et al., 2000, p.6, based on the original given by Thayer & Dorfman, 1990).

NFR: Non-functional requirement. A property or quality that the product must have, such as an appearance, or a speed or accuracy property (Robertson & Robertson, 1999)

OCL: The Object Constraint Language is an expression language that enables one to describe constraints on object-oriented models and other object modeling artifacts (Warmer & Kleppe, 1999).

Quality evaluation: The systematic examination of the extent to which an entity is capable of fulfilling specified requirements (ISO/IEC 9126-1, 2001, p.20).

Quality Model: The set of characteristics and the relationships between them which provide the basis for specifying quality requirements and evaluating quality (ISO/IEC 9126-1, 2001, p.21).

Requirements Engineering: The branch of software engineering concerned with the real-world goals for, functions of, and constraints on software systems. It is also concerned with the relationship of these factors to precise specifications of software behavior, and to their evolution over time and across software families (Zave, 1997).

UML: The Unified Modeling Language is a standard graphical language for visualizing, specifying, constructing, and documenting the artifacts of a software-intensive system (Booch, et al., 1998).

N

Novel Indexing Method of Relations Between Salient Objects

Richard Chbeir
Bourgogne University, France

Kokou Yetongnon
Bourgogne University, France

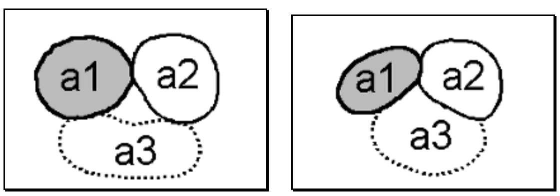
INTRODUCTION

During the last decade, a lot of work has been done in information technology in order to integrate image retrieval in the standard data processing environments (Rui, Huang & Chang, 1999; William & Grosky, 1997; Yoshitaka & Ichikawa, 1999).Relations between image objects are very frequently used to index the image-document (Peuquet, 1986). In medicine, for instance, the spatial data in surgical or radiation therapy of brain tumors are decisive because the location of a tumor has profound implications on a therapeutic decision (Chbeir, Amghar, & Flory, 2001). Hence, it is crucial to provide a precise and powerful system to express spatial relations. In the literature, three major types of spatial relations are proposed: metric (Peuquet, 1986), directional (El-kwae & Kabuka, 1999) and topological relations (Egenhofer, 1997; Egenhofer, Frank & Jackson, 1989).

In spite of all the proposed work to represent complex visual situations, several shortcomings exist in the methods of spatial relation computations. For instance, Figure 1 shows two different spatial situations of three salient objects that are described by the same spatial relations in both cases: topological relations: a1 Touch a2, a1 Touch a3, a2 Touch a3; and directional relations: a1 Above a3, a2 Above a3, a1 Left a2.

The existing systems do not have the required expressive power to represent these situations. Thus, in this article, we address this issue and propose a snapshot of our novel method that can easily compute several types of relations between salient objects with better expressions. The rest of this article is organized as follows. In the second section, we present related work in this domain. In the third section, we briefly present our method and discuss how it gives better results using spatial features. finally, conclusions and future directions are given in the fourth section.

Figure 1. Two different spatial situations



BACKGROUND

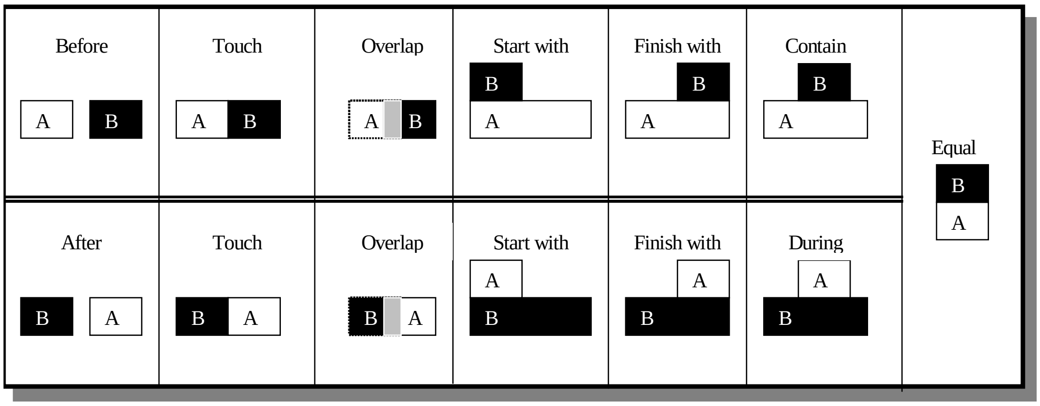
The problem of image retrieval is strongly related to image representation. The two different approaches used for the representation of images are: the metadata-based and the content-based approaches. Made with human assistance, the metadata-based representation uses alpha-numeric attributes to describe the context and/or the content of an image. The other approach for representation of images is using its low-level contents such as its color, texture, and shape (Remco & Mirela, 2000). The representations using low-level features are derived through feature extraction algorithms.

In both approaches, relations between either salient objects, shapes, points of interests, and so forth have been widely used in image indexing such as R-trees (Beckmann, 1990; Guttman, 1984), 2D-Strings (Chang & Jungert, 1991, 1997; Chang, Shi & Yan, 1987), and so forth. Temporal and spatial relations are the most used relations in image indexing.

To calculate temporal relations, two paradigms are proposed in the literature:

- The first paradigm consists of representing the time as a set of *instants:* $t_1,..t_i,..t_n$. Traditionally, only three temporal relations are possible between two objects: *before*, its symmetric relation *after*, and *equal*.
- The second paradigm considers the time as a set of intervals $[t_i, t_j]$. Allen relations (Allen, 1983) are often used to represent temporal relations between intervals. Allen proposes 13 temporal relations (Figure 2) in which six are symmetrical.

Figure 2. Allen relations



On the other hand, three major types of spatial relations are generally proposed in image representation (Egenhofer, Frank & Jackson, 1989):

- Metric relations: measure the distance between salient objects. For instance, the metric relation "far" between two objects A and B indicates that each pair of points A_i and B_j has a distance greater than a certain value δ.
- Directional relations: describe the order between two salient objects according to a direction, or the localisation of salient object inside images. In the literature, 14 directional relations are considered:
  - Strict: north, south, east, and west.
  - Mixture: north-east, north-west, south-east, and south-west.
  - Positional: left, right, up, down, front and behind.

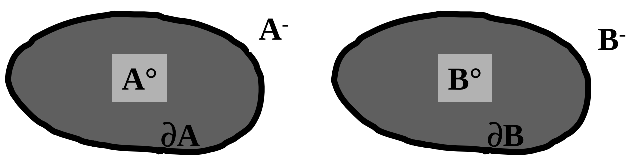
    Directional relations are *rotation variant* and there is a need to have referential base. Furthermore, directional relations do not exist in certain configurations.
- Topological relations: describe the intersection and the incidence between objects. Egenhofer (Egenhofer, 1997) has identified six basic relations: *Disjoint, Meet, Overlap, Cover, Contain, and Equal.* Topological relations present several characteristics that are *exclusive* to two objects (i.e., there is one and only one topological relation between two objects). Furthermore, topological relations have *absolute* value because of their constant existence between objects. Another interesting characteristic of topological relations is that they are transformation, translation, scaling, and zooming *invariant*.

PROPOSITION

The 9-Intersection model proposed by Egenhofer (Egenhofer, 1997; Egenhofer, Frank & Jackson, 1989) represents each shape "A" as a combination of three parts: *interior* $\mathbf{A}°$, *boundary* $\partial\mathbf{A}$ and *exterior* $\mathbf{A}^-$. The topological relations between two shapes are obtained by applying an intersection matrix between these parts (Figure 3). Each intersection is characterised by an empty ($\varnothing$) or non-empty ($\neg\varnothing$) value.

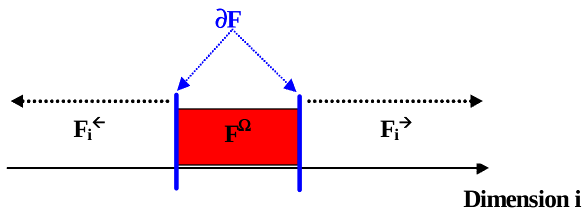
Our proposal represents an extension of this 9-Intersection model. It provides a general method for computing not only topological relations but also other types of relations such as temporal, spatial, and so forth. The idea shows that the relations are identified in function of features of shape, time, and so forth. The shape feature gives spatial relations; the time feature gives temporal relations, and so on. To identify a relation between two salient objects, we propose the use of an intersection matrix between sets of features. Let us first consider a feature F. We define its intersection sets as follows:

Figure 3. The 9-Intersection model: The topological relation between two shapes is based on the comparison of the three parts of each one.



$$R(\mathbf{A},\mathbf{B}) = \begin{pmatrix} \mathbf{A}°\cap\mathbf{B}° & \mathbf{A}°\cap\partial\mathbf{B} & \mathbf{A}°\cap\mathbf{B}^- \\ \partial\mathbf{A}\cap\mathbf{B}° & \partial\mathbf{A}\cap\partial\mathbf{B} & \partial\mathbf{A}\cap\mathbf{B}^- \\ \mathbf{A}^-\cap\mathbf{B}° & \mathbf{A}^-\cap\partial\mathbf{B} & \mathbf{A}^-\cap\mathbf{B}^- \end{pmatrix}$$

Figure 4. Intersection sets of one-dimensional feature



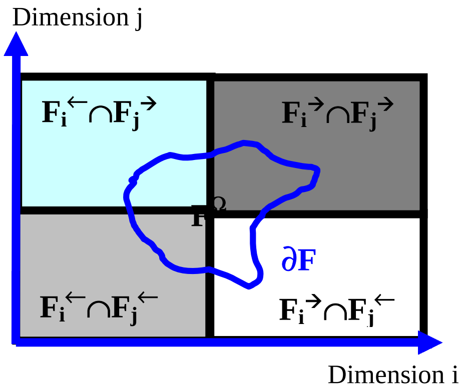
- The ***interior*** $\mathbf{F^{\Omega}}$**:** contains all elements that cover the interior or the core of F. In particular, it contains the barycenter of F. The definition of this set has a great impact on the other sets. F^{Ω} may be empty ($\emptyset$).
- The ***boundary*** **∂F:** contains all elements that allow determining the frontier of F. ∂F is never empty ($\neg\emptyset$).
- The ***exterior*** $\mathbf{F^{-}}$**:** is the complement of $F^{\Omega}\cup\partial F$. It contains at least two elements: $\perp$ (the minimum value) and ∞ (the maximum value). F^{-} can be divided into several disjoint subsets. This decomposition depends on the number of the feature dimensions. For instance, if we consider a feature of one dimension i (such as the acquisition time of an image), two intersection subsets can be defined (Figure 4):

- $\mathbf{F_i^{\leftarrow}}$ **(or inferior):** contains elements of F^{-} that do not belong to any other intersection set and inferior to ∂F elements based on i dimension.
- $\mathbf{F_i^{\rightarrow}}$ **(or superior):** contains elements of F^{-} that do not belong to any other intersection set and superior to ∂F elements based on i dimension.

If we consider a feature of two dimensions i and j (as the shape in a 2D space), we can define four intersection subsets (Figure 5):

- $\mathbf{F_i^{\leftarrow}\cap F_j^{\leftarrow}}$ **(or inferior):** contains elements of F^{-} that do not belong to any other intersection set and inferior to $\mathbf{F^{\Omega}}$ and ∂F elements based on i and j dimensions.
- $\mathbf{F_i^{\leftarrow}\cap F_j^{\rightarrow}}$**:** contains elements of F^{-} that do not belong to any other intersection set and inferior to $\mathbf{F^{\Omega}}$ and ∂F elements based on i dimension, and superior to $\mathbf{F^{\Omega}}$ and ∂F elements on the basis of j dimension.

Figure 5. Intersection sets of two-dimensional feature



- $\mathbf{F_i^{\rightarrow}\cap F_j^{\leftarrow}}$**:** contains elements of F^{-} that do not belong to any other intersection set and superior to $\mathbf{F^{\Omega}}$ and ∂F elements based on i dimension, and inferior to $\mathbf{F^{\Omega}}$ and ∂F elements on the basis of j dimension.
- $\mathbf{F_i^{\rightarrow}\cap F_j^{\rightarrow}}$ **(or superior):** contains elements of F^{-} that do not belong to any other intersection set and superior to $\mathbf{F^{\Omega}}$ and ∂F elements based on i and j dimensions.

More generally, we can determine intersection sets (2^n) of n-dimensional feature. In addition, we use a tolerance degree in the feature intersection sets definition in order to represent separations between sets. For this purpose, we use two tolerance thresholds:

- Internal threshold $\boldsymbol{\varepsilon}^{i}$ that defines the distance between F^{Ω} and ∂F,
- External threshold $\boldsymbol{\varepsilon}^{e}$ that defines the distance between subsets of F^{-}.

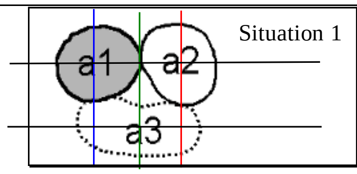
To calculate relation between two salient objects, we establish an intersection matrix of their corresponding feature intersection sets. Matrix cells have binary values:

- 0 whenever intersection between sets is empty,
- 1 otherwise

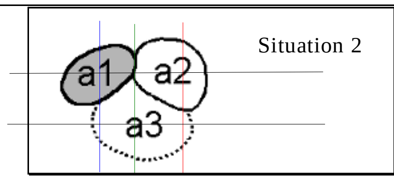
Figure 6. Intersection matrix of two objects A and B on the basis of two-dimensional feature

$A^{\Omega}\cap B^{\Omega}$	$A^{\Omega}\cap\partial B$	$A^{\Omega}\cap B_1^{\leftarrow}\cap B_2^{\leftarrow}$	$A^{\Omega}\cap B_1^{\leftarrow}\cap B_2^{\rightarrow}$	$A^{\Omega}\cap B_1^{\rightarrow}\cap B_2^{\rightarrow}$	$A^{\Omega}\cap B_1^{\rightarrow}\cap B_2^{\leftarrow}$
$\partial A\cap B^{\Omega}$	$\partial A\cap\partial B$	$\partial A\cap B_1^{\leftarrow}\cap B_2^{\leftarrow}$	$\partial A\cap B_1^{\leftarrow}\cap B_2^{\rightarrow}$	$\partial A\cap B_1^{\rightarrow}\cap B_2^{\rightarrow}$	$\partial A\cap B_1^{\rightarrow}\cap B_2^{\leftarrow}$
$A_1^{\leftarrow}\cap A_2^{\leftarrow}\cap B^{\Omega}$	$A_1^{\leftarrow}\cap A_2^{\leftarrow}\cap\partial B$	$A_1^{\leftarrow}\cap A_2^{\leftarrow}\cap B_1^{\leftarrow}\cap B_2^{\leftarrow}$	$A_1^{\leftarrow}A_2^{\leftarrow}\cap B_1^{\leftarrow}\cap B_2^{\rightarrow}$	$A_1^{\leftarrow}\cap A_2^{\leftarrow}\cap B_1^{\rightarrow}\cap B_2^{\rightarrow}$	$A_1^{\leftarrow}\cap A_2^{\leftarrow}\cap B_1^{\rightarrow}\cap B_2^{\leftarrow}$
$A_1^{\leftarrow}\cap A_2^{\rightarrow}\cap B^{\Omega}$	$A_1^{\leftarrow}\cap A_2^{\rightarrow}\cap\partial B$	$A_1^{\leftarrow}\cap A_2^{\rightarrow}\cap B_1^{\leftarrow}\cap B_2^{\leftarrow}$	$A_1^{\leftarrow}A_2^{\rightarrow}\cap B_1^{\leftarrow}\cap B_2^{\rightarrow}$	$A_1^{\leftarrow}\cap A_2^{\rightarrow}\cap B_1^{\rightarrow}\cap B_2^{\rightarrow}$	$A_1^{\leftarrow}\cap A_2^{\rightarrow}\cap B_1^{\rightarrow}\cap B_2^{\leftarrow}$
$A_1^{\rightarrow}\cap A_2^{\rightarrow}\cap B^{\Omega}$	$A_1^{\rightarrow}\cap A_2^{\rightarrow}\cap\partial B$	$A_1^{\rightarrow}\cap A_2^{\rightarrow}\cap B_1^{\leftarrow}\cap B_2^{\leftarrow}$	$A_1^{\rightarrow}A_2^{\rightarrow}\cap B_1^{\leftarrow}\cap B_2^{\rightarrow}$	$A_1^{\rightarrow}\cap A_2^{\rightarrow}\cap B_1^{\rightarrow}\cap B_2^{\rightarrow}$	$A_1^{\rightarrow}\cap A_2^{\rightarrow}\cap B_1^{\rightarrow}\cap B_2^{\leftarrow}$
$A_1^{\rightarrow}\cap A_2^{\leftarrow}\cap B^{\Omega}$	$A_1^{\rightarrow}\cap A_2^{\leftarrow}\cap\partial B$	$A_1^{\rightarrow}\cap A_2^{\leftarrow}\cap B_1^{\leftarrow}\cap B_2^{\leftarrow}$	$A_1^{\rightarrow}A_2^{\leftarrow}\cap B_1^{\leftarrow}\cap B_2^{\rightarrow}$	$A_1^{\rightarrow}\cap A_2^{\leftarrow}\cap B_1^{\rightarrow}\cap B_2^{\rightarrow}$	$A_1^{\rightarrow}\cap A_2^{\leftarrow}\cap B_1^{\rightarrow}\cap B_2^{\leftarrow}$

Figure 7. Identifed spatial relations using our method



$$R(a1, a2) = \begin{array}{|c|c|c|c|c|c|} \hline 0&0&1&1&0&0 \\ \hline 0&1&1&1&0&0 \\ \hline 0&0&1&0&0&0 \\ \hline 0&0&0&1&0&0 \\ \hline 1&1&0&1&1&0 \\ \hline 1&1&1&0&0&1 \\ \hline \end{array} \qquad R(a1, a3) = \begin{array}{|c|c|c|c|c|c|} \hline 0&0&0&1&0&0 \\ \hline 0&1&0&1&1&0 \\ \hline 1&1&1&1&0&0 \\ \hline 0&0&0&1&0&0 \\ \hline 0&0&0&1&1&0 \\ \hline 1&1&1&1&1&1 \\ \hline \end{array}$$



$$R'(a1, a2) = \begin{array}{|c|c|c|c|c|c|} \hline 0&0&1&1&0&0 \\ \hline 0&1&1&1&0&0 \\ \hline 0&0&1&0&0&0 \\ \hline 0&0&0&1&0&0 \\ \hline 1&1&0&1&1&0 \\ \hline 1&1&1&0&0&1 \\ \hline \end{array} \qquad R'(a1, a3) = \begin{array}{|c|c|c|c|c|c|} \hline 0&0&0&1&0&0 \\ \hline 0&1&0&1&1&0 \\ \hline 1&1&1&1&0&0 \\ \hline 0&0&0&1&0&0 \\ \hline 0&0&0&1&1&0 \\ \hline 1&1&1&0&1&1 \\ \hline \end{array}$$

For instance, for two-dimensional feature (such as the shape) of two salient objects A and B, we obtain the intersection matrix (see Figure 6).

Using our method, we are able to provide a high expression power to spatial relations that can be applied to describe images and formulate complex visual queries in several domains. For example, for Figure 1, which shows two different spatial situations between three salient objects a1, a2, and a3, our method expresses the spatial relations as shown in Figure 7. The relations R(a1, a2) and R'(a1, a2) are equal but the relations R(a1, a3) and R'(a1, a3) are clearly distinguished. Similarly, we can express relations between a2 and a3 in both situations.

Moreover, our method allows combining both directional and topological relation into one *binary* relation, which is very important for indexing purposes. There are no directional and topological relations between two salient objects but only one spatial relation. Hence, we can propose a 1D-String to index images instead of 2D-Strings (Chang, Shi & Yan, 1987).

CONCLUSION AND FUTURE TRENDS

In this article, we presented an overview of traditional relations used in image indexing. We also snapshot our method. This work aims to homogenize, reduce and optimize the representation of relations. It is not limited to spatial relations but it is also applicable to other types of relations (temporal, semantic, spatio-temporal, etc.).

However, in order to study its efficiency, our method requires more intense experiments in complex environment where a great number of feature dimensions and salient objects exist. Furthermore, we currently work on its integration in our prototype MIMS (Chbeir, Amghar & Flory, 2001) in order to improve image storage and retrieval. Future directions will address the use of these relations on video documents.

REFERENCES

llen, J.F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, *26*(11), 832-843.

Beckmann, N. (1990). The R*-tree: An efficient and robust access method for points and rectangles. *SIGMOD Record, 19*(2), 322-331.

Chang, S.K., & Jungert, E. (1991). Pictorial data management based upon the theory of symbolic projections. *Journal of Visual Languages and Computing, 2*(3), 195-215.

Chang, S.K., & Jungert, E. (1997). Human- and system-directed fusion of multimedia and multimodal information using the sigma-tree data model. *Proceedings of the 2nd International Conference on Visual Information Systems,* San Diego (pp. 21-28).

Chang, S.K., Shi, Q.Y., & Yan, C.W. (1987). Iconic indexing by 2-D strings. *IEEE-Transactions-on-Pattern-Analysis-and-Machine-Intelligence, PAMI-9, 3,* 413-428.

Chbeir, R., Amghar, Y., & Flory, A. (2001). A prototype for medical image retrieval. *International Journal of Methods of Information in Medicine, Schattauer, 3,* 178-184.

Egenhofer, M. (1997). Query processing in spatial query by sketch. *Journal of Visual Language and Computing, 8*(4), 403-424.

Egenhofer, M., Frank, A., & Jackson, J. (1989). A topological data model for spatial databases. Symposium on the Design and Implementation of Large Spatial Databases, Santa Barbara, CA. *Lecture Notes in Computer Science, 409,* 271-286.

El-kwae, M.A., & Kabuka, M.R. (1999). A robust framework for content-based retrieval by spatial similarity in image databases. *ACM Transactions on Information Systems, 17*(2), 174-198.

Guttman, A. (1984). R-trees: A dynamic index structure for spatial searching. *SIGMOD Record, 14*(2), 47-57.

Peuquet, D.J. (1986). The use of spatial relationships to aid spatial database retrieval. *Second International Symposium on Spatial Data Handling,* Seattle (pp. 459 – 471).

Remco, C.V., & Mirela, T. (2000). *Content-based image retrieval systems: A survey*. Technical Report UU-cs-2000-34. Department of Computer Science, Utrecht University.

Rui, Y., Huang, T.S., &Chang, S.F. (1999). Image retrieval: Past, present, and future. *Journal of Visual Communication and Image Representation, 10,* 1-23.

William, I., & Grosky. (1997). Managing multimedia information in database systems. *Communications of the ACM, 40*(12), 72-80.

Yoshitaka, A., & Ichikawa, T. (1999). A survey on content-based retrieval for multimedia databases. *IEEE Transactions on Knowledge and Data Engineering, 11*(1), 81-93.

KEY TERMS

Directional Relation: Describes the order between two salient objects according to a direction, or the localization of salient object inside images. In the literature, 14 directional relations are considered:

- Strict: north, south, east, and west.
- Mixture: north-east, north-west, south-east, and south-west.
- Positional: left, right, up, down, front and behind.

Image Indexing: Consists of assigning concise and significant descriptors to an image. Objects' shapes and positions are used in several indexing approaches where image is represented as a graph or tree (R-tree, B-tree, etc.).

Metric Relation: Measures the distance between salient objects. For instance, the metric relation "far" between two objects A and B indicates that each pair of points A_i and B_j has a distance grater than a certain value d.

Multimedia Document: Represents a document containing not only textual data but also multimedia ones such as images, videos, songs, and so forth.

Salient Object: Means interesting or significant object in an image (sun, mountain, boat, etc.). Its computing changes in function of the application domain.

Spatio-temporal Relation: A combination of two spatial and temporal relations into one relation used to index video-documents and objects' evolution between two images (e.g., tumor evolution).

Topological Relation: Describes the intersection and the incidence between objects. Six basic relations have been identified in the literature: *Disjoint, Meet, Overlap, Cover, Contain,* and *Equal*.

Object Database Benchmarks

Jerôme Darmont
Universite Lumiere Lyon 2, France

INTRODUCTION

The need for performance measurement tools appeared soon after the emergence of the first Object-Oriented Database Management Systems (OODBMSs), and proved important for both designers and users (Atkinson & Maier, 1990). Performance evaluation is useful to designers to determine elements of architecture and more generally to validate or refute hypotheses regarding the actual behavior of an OODBMS. Thus, performance evaluation is an essential component in the development process of well-designed and efficient systems. Users may also employ performance evaluation, either to compare the efficiency of different technologies before selecting an OODBMS or to tune a system.

Performance evaluation by experimentation on a real system is generally referred to as benchmarking. It consists in performing a series of tests on a given OODBMS to estimate its performance in a given setting. Benchmarks are generally used to compare the global performance of OODBMSs, but they can also be exploited to illustrate the advantages of one system or another in a given situation, or to determine an optimal hardware configuration. Typically, a benchmark is constituted of two main elements: a workload model constituted of a database and a set of read and write operations to apply on this database, and a set of performance metrics.

BACKGROUND

Object Database Benchmarks Evolution

In the sphere of relational DBMSs, the Transaction Performance Processing Council (TPC) issues standard benchmarks, verifies their correct application and regularly publishes performance test results. In contrast, there is no standard benchmark for OODBMSs, even if the more popular of them, OO1, HyperModel, and OO7, can be considered as *de facto* standards.

OO1, also referred to as the "Cattell Benchmark" (Cattell, 1991), was developed early in the nineties when there was no appropriate benchmark for engineering applications such as computer aided design (CAD), computer aided manufacturing (CAM), or software engineering (SE). OO1 is a simple benchmark that is very easy to implement. A major drawback of this tool is that its workload model is too elementary to measure the elaborate traversals that are common in many types of object-oriented applications.

The HyperModel Benchmark (Anderson et al., 1990), also referred to as the Tektronix Benchmark, possesses a richer workload model than OO1. This renders it potentially more effective than OO1 in measuring the performance of engineering databases. However, this added complexity also makes HyperModel harder to implement.

OO7 (Carey, Dewitt & Naughton, 1993) reuses the structures of OO1 and HyperModel to propose a more complete benchmark and to simulate various transactions running on a diversified database. It has also been designed to be more generic than its predecessors and to correct some of their known weaknesses. However, OO7 is even harder to implement than HyperModel.

OO1, HyperModel, and OO7, though aimed at engineering applications, are often viewed as general-purpose benchmarks. However, they feature relatively simple databases and are not well suited for other types of applications such as financial, telecommunication, and multimedia applications (Tiwary, Narasayya & Levy, 1995). Hence, many benchmarks were developed to study particular domains, such as client-server architectures (Schreiber, 1994), object clustering (Bancilhon, Delobel & Kanellakis, 1992; Darmont, Petit & Schneider, 1998; Gerlhof et al., 1996), object-relational systems (Carey, Dewitt & Naughton, 1993; Lee, Kim & Kim 2000), active databases (Zimmermann & Buchmann, 1995), workflow management (Bonner, Shrufi & Rozen, 1995), CAD applications (Kempe et al., 1995), or the study of views in an object-oriented context (Kuno & Rundensteiner, 1995). A fair number of these benchmarks are more or less based on OO1, HyperModel, or OO7.

An alternative to very specific benchmarks resides in generic and tunable benchmarks such as OCB (Darmont & Schneider, 2000). The flexibility and scalability of OCB is achieved through an extensive set of parameters that helps OCB simulate the behavior of the *de facto* standards in object-oriented benchmarking. Furthermore, OCB's generic model can be implemented within an object-relational system easily and most of its operations are relevant for such a system. Hence, it can also be applied in an object-relational context with few adaptations.

Finally, OCB has been recently extended to become the Dynamic Object Evaluation Framework (DOEF), which introduces a dynamic component in the workload (He & Darmont, 2003). Changes in access patterns indeed play an important role in determining the efficiency of a system or of key performance optimization techniques such as dynamic clustering, prefetching, and buffering. However, all previous benchmarks produced static access patterns in which objects were always accessed in the same order repeatedly. In contrast, DOEF simulates access pattern changes using configurable styles of change.

Issues and Tradeoffs in Benchmarking

Gray (1993) defines four primary criteria to specify a good benchmark: (1) *relevance:* it must concern aspects of performance that appeal to the largest number of potential users; (2) *portability:* it must be reusable to test the performances of different OODBMSs; (3) *simplicity:* it must be feasible and must not require too many resources; and (4) *scalability:* it must be able to be adapted to small or large computer systems or new architectures. Table 1 summarizes the characteristics of the main existing benchmarks according to Gray's criteria. It is important to note that these four criteria are in mutual conflict. For instance, the size and complexity of a relevant workload may come in conflict with its feasibility and possibly with portability requirements. Hence, it is necessary to find the right compromise regarding given needs.

The *de facto* standards in OODBMS benchmarking all aim at being generic. However, they all incorporate database schemas that are inspired by structures used in engineering software, which finally tailors them to the study of these particular systems. Adapting these benchmarks to another domain requires some work and a derived benchmark that takes into account specific elements often needs to be designed. Hence, their relevance decreases when they are applied in other domains but engineering. A solution to this problem is to select a generic benchmark that can be tailored to meet specific needs. However, there is a price to pay. Genericity is achieved with the help of numerous parameters that are not always easy to set up. Thus, the effort in designing a specific benchmark must be compared to the parameterization complexity of a generic benchmark.

There is also another, very different, "qualitative" aspect of benchmarking that has never really been considered in published benchmarks and that is important for a user to consider when selecting an OODBMS. Atkinson et al. (1992), Banerjee and Gardner (1995), and Kempe et al. (1995) all insist on the fact that system functionality is at least as important as raw performances. Hence, criteria concerning these functionalities should be worked out.

Finally, there is an issue that is not a scientific one. Carey, Dewitt and Naughton (1993) and Carey et al. (1994) pointed out serious legal difficulties in their benchmarking effort. Indeed, OODBMS vendors are sometimes reluctant to see benchmark results published. However, designing relevant benchmarks remains an important task and should still be carried out to help researchers, software designers or users evaluate the adequacy of any prototype, system or implementation technique in a particular environment.

CONCLUSION AND FUTURE TRENDS

The development of new object database benchmarks is now scarce, mainly because the first generation of OODBMSs failed to achieve any broad commercial success. This failure is largely due to the never-ending issue of poor performance compared to relational DBMSs, which are well optimized and efficient. However, with the development of object-oriented programming both off-line and online, the need for persistent objects remains. Object-relational systems are now used more and more frequently (to store XML documents, for instance). Thus, the experience that has been accumulated when designing object-oriented database benchmarks could be reused in this context. The challenge for object store designers is now to produce efficient systems, and sound benchmarks could help them achieve this goal.

Table 1. Comparison of existing benchmarks with Gray's criteria

	Relevance	Portability	Simplicity	Scalability
OO1	– –	+ +	+ +	–
HyperModel	+	+	–	– –
OO7	+ +	+	– –	–
OCB	+ +	+	–	+ +

Strong point: +　Very strong point: ++　Weak point: –　Very weak point: – –

REFERENCES

Anderson, T.L., Berre, A.J., Mallison, M., Porter, H.H., & Scheider, B. (1990). The HyperModel benchmark. *International Conference on Extending Database Technology*, Venice, Italy (pp. 317-331).

Atkinson, M.P., Birnie, A., Jackson, N., & Philbrow, P.C. (1992). Measuring persistent object systems. *5th International Workshop on Persistent Object Systems*, San Miniato (Pisa), Italy (pp. 63-85).

Atkinson, M.P., & Maier, D. (1990). Perspectives on persistent object systems. *4th International Workshop on Persistent Object Systems*, Martha's Vineyard, USA (pp. 425-426).

Bancilhon, F., Delobel, C., & Kanellakis, P. (Eds.). (1992). *Building an object-oriented database system: The story of O_2*. Morgan Kaufmann.

Banerjee, S., & Gardner, C. (1995). Towards an improved evaluation metric for object database management systems. *OOPSLA 95 Workshop on Object Database Behavior, Benchmarks and Performance,* Austin, TX.

Bonner, A.J., Shrufi, A., & Rozen, A. (1995). Benchmarking object-oriented DBMSs for workflow management. *OOPSLA 95 Workshop on Object Database Behavior, Benchmarks and Performance*, Austin, TX.

Carey, M.J., Dewitt, D.J., Kant, C., & Naughton, J.F. (1994). A status report on the OO7 OODBMS benchmarking effort. *SIGPLAN Notices*, *29*(10), 414-426.

Carey, M.J., Dewitt, D.J., & Naughton, J.F. (1993). The OO7 benchmark. *ACM SIGMOD International Conference on Management of Data*, Washington, USA (pp. 12-21).

Carey, M.J., Dewitt, D.J., & Naughton, J.F. (1997). The BUCKY object-relational benchmark. *ACM SIGMOD International Conference on Management of Data*, Tucson, AZ (pp. 135-146).

Cattell, R.G.G. (1991). An engineering database benchmark. In J. Gray (Ed.), *The benchmark handbook for database transaction processing systems* (pp. 247-281). Morgan Kaufmann.

Darmont, J., Petit, B., & Schneider, M. (1998). OCB: A generic benchmark to evaluate the performances of object-oriented database systems. *6th International Conference on Extending Database Technology*, Valencia, Spain. *LNCS, 1377,* 326-340.

Darmont, J., & Schneider, M. (2000). Benchmarking OODBs with a generic tool. *Journal of Database Management, 11*(3), 16-27.

Gerlhof, C., Kemper, A., Kilger, C., & Moerkotte, G. (1996). On the cost of monitoring and reorganization of object bases for clustering. *SIGMOD Record*, *25*(3).

Gray, J. (Ed.). (1993). *The benchmark handbook for database and transaction processing systems* (2nd ed.). Morgan Kaufmann.

He, Z., & Darmont, J. (2003). DOEF: A dynamic object evaluation framework. *The 14th International Conference on Database and Expert Systems Applications*, Prague, Czech Republic. LNCS, 2736, 662-671.

Kempe, J., Kowarschick, W., Kießling, W., Hitzelgerger, R., & Dutkowski, F. (1995). Benchmarking object-oriented database systems for CAD. *6th International Conference on Database and Expert Systems Applications,* London, UK. *LNCS, 978,* 167-176.

Kuno, H., & Rundensteiner, E.A. (1995). Benchmarks for object-oriented view mechanisms. *OOPSLA 95 Workshop on Object Database Behavior, Benchmarks and Performance*, Austin, TX.

Lee, S., Kim, S., & Kim, W. (2000). The BORD benchmark for object-relational databases. *11th International Conference on Database and Expert Systems Applications*, London, UK. *LNCS, 1873*, 6-20.

Schreiber, H. (1994). JUSTITIA: A generic benchmark for the OODBMS selection. *4th International Conference on Data and Knowledge Systems in Manufacturing and Engineering*, Shatin, Hong Kong (pp. 324-331).

Tiwary, A., Narasayya, V.R., & Levy, H.M. (1995). Evaluation of OO7 as a system and an application benchmark. *OOPSLA 95 Workshop on Object Database Behavior, Benchmarks and Performance,* Austin, TX.

Zimmermann, J., & Buchmann, A.P. (1995). Benchmarking active database systems: A requirement analysis. *OOPSLA 95 Workshop on Object Database Behavior, Benchmarks and Performance,* Austin, TX.

KEY TERMS

Benchmark: A standard program that runs on different systems to provide an accurate measure of their performance.

Database Benchmark: A benchmark specifically aimed at evaluating the performance of DBMSs or DBMS components.

Object-Oriented Database: A database system offering DBMS facilities in an object-oriented programming environment. Data are natively stored as objects.

Object-Relational Database: A database system where the relational model is extended with object-oriented concepts. Data are still stored in relational structures.

Performance Metrics: Simple or composite metrics aimed at expressing the performance of a system.

Synthetic Benchmark: A benchmark in which the workload model is artificially generated, as opposed to a real-life workload.

Workload Model: In a database benchmark, a database and a set of read and write operations to apply on this database.

Object Oriented Software Metrics

Jana Polgar
Monash University, Australia

INTRODUCTION

Software measurement is considered to be an efficient means to monitor the quality of software projects, predict cost of maintenance, assess reusability of components, provide prediction of faults in similar projects, and contribute to improvement of the software development process. This chapter surveys software metrics literature with particular focus on object-oriented metrics, and metrics for measuring code complexity. Furthermore, we provide a critical view of software metrics and their usability.

BRIEF SURVEY OF OBJECT-ORIENTED METRICS

Since 1995, the trend toward incorporating *measurement theory* into all software metrics has led to identification of scales for measures, thus providing some perspective on dimensions. The most common scale types based on measurement theory are as follows:

- *Ordinal scale*: An ordered set of categories (often used for adjustment factors in cost models based on a fixed set of scale points)
- *Interval scale*: Numerical values, where the difference between each consecutive pair of numbers is an equivalent amount, but there is no "real" zero value
- *Ratio scale*: Elements are assigned numbers such that differences and ratios between the numbers reflect differences and ratios of the attribute
- *Nominal scale*: A set of categories into which an item is classified.
- *Absolute scale*: Elements are assigned numbers such that all properties of the numbers reflect analogous properties of the attribute

Fetchke (1995) and Zuse (1994) analyzed the properties of object-oriented software metrics on the basis of measurement theory. The underlying notion of measurement theory is based on intuitive or empirical existence of relationships among objects within our Universe of Discourse. These relationships can then be formally described in a mathematically derived formal relational system. They also investigated how and under what conditions the software measures may be viewed as ordinal, ratio, nominal, and interval. They admitted that these scale types present very little meaning with regard to maintainability and "error-proneness" of the application.

The contribution of Zuse and Fetchke's work is in the introduction of specific perspectives of measures. They emphasize preciseness of definition of scales as well as definition of an attribute that is measured.

The *axiomatic approach* was proposed by Weyuker (1988). This framework is based on a set of nine axioms, as listed in Table 1.

In Weyeker's metric proposal, we observe the formalization of structural inheritance complexity metrics. Property 9 means that splitting one class into two classes can reduce the complexity. The experience supports argument by Chidamber and Kemerer (1994) that the complexity of interaction may even increase when classes are divided.

Fenton and Pfleger (1997) used the term "software metrics" to describe the following artifacts:

- A number derived, usually empirically, from a process or code [for example, lines of code (LOC) or number of function points]
- A scale of measurement (The example used in Fenton's book is nominal scale or classification.)
- An identifiable attribute used to provide specific functionality (an example is "portability" or class coupling metric)
- Theoretical or data-driven model describing a dependent variable as a function of independent variables (an example can be the functional relationship between maintenance effort and program size)

These descriptions typically lead to widespread confusion between models and their ability to predict desired software characteristics, thus their suitability in being used for estimation purposes.

The metrics of Chidamber and Kemerer, summarized in Table 2, also have foundation in measurement theory. The authors do not base their investigation on the extensive structure. The criticism by Churcher and Sheppard (1994) points to the ambiguity of some metrics, particularly WMC. Hitz and Montazeri (1996) and Fetchke (1995) showed that CBO does not use a sound empirical relation

Table 1. Weyuker's axioms

Axiom	Name	Description
1	Noncoarseness	$(\exists P)(\exists Q)(\mu(P) \neq \mu(QQ))$
2	Granularity	Let c be non-negative number. Then there is finite number of class with the complexity = c
3	Nonuniquesniss	There is distinct number of classes P and Q such that $\mu(\mathrm{P}) = \mu(\mathrm{Q})$
4	Design detail matter	$(\exists P)(\exists Q)(P \equiv Q and \mu(P) \neq \mu(Q))$
5	Monotonicity	$(\forall P)(\forall Q)(\mu(P) \leq \mu(P+Q) and \mu(Q) \leq \mu(P+Q))$
6	Non-equivalence of interaction	a) $(\exists P)(\exists Q)(\exists R)\mu(P) = \mu(Q) and \mu(P+R) \neq \mu(Q+R)$ b) $(\exists P)(\exists Q)(\exists R)\mu(P) = \mu(Q) and \mu(R+P) \neq \mu(R+Q)$
7	Interaction among statements	Not considered among objects
8	No change on renaming	If P is renaming of Q then $\mu(P) = \mu(Q)$
9	Interaction CAN increase complexity	$(\exists P)(\exists Q)(\mu(P) + \mu(Q) < \mu(P+Q))$

Table 2. Chidamber and Kemerer metrics (Chidamber, 1994)

Weighted Methods per Class **(WMC)**	$WMC = \sum_{i=1}^{n} c_i$ where c_i is the static complexity of each of the n methods
Depth of Inheritance Tree **(DIT)**	With multiple inheritance the max DIT is the length from node to the root
Number of Children **(NOC)**	Number of immediate subclasses
Coupling Between Object Classes **(CBO)**	Number of other classes to which a particular class is coupled. CBO maps the concept of coupling for a class into a measure.
The Response for a Class **(RFC)**	The size of response set for a particular class.
The lack of Cohesion metric **(LCOM)**	$LCOM = \lvert P \rvert - \lvert Q \rvert if \lvert P \rvert > \lvert Q \rvert = 0 otherwise$

system, particularly, that it is not based on the extensive structures. Furthermore, LCOM metric allows representation of equivalent cases differently, thus introducing additional error.

Coupling measures form an important group of measures in the assessment of dynamic aspects of design quality. Coupling among objects is loosely defined as the measure of the strength of the connection from one object to another. The approaches of different authors mostly differ in definition of the measured attribute—coupling among classes. Table 3 provides a summary of differences in definitions. Some of the attributes may be known only too late in development.

Two aspects impact coupling between classes: the frequency of messaging between classes (cardinality and multiplicity of objects derived from these classes) and type of coupling. The discussion in Eder and Kappel (1994) distinguishes among three types: interaction coupling, component coupling, and inheritance coupling. The degree of coupling is based on defining a partial order on the set of coupling types. The low end is described by small and explicit interrelationships, and the high end of the scale is assigned to large, complex and implicit interrelationships. The definition is subjective and requires empirical assignment of values in order to be used as a software quality indicator.

Cohesion is defined as a degree to which elements in a class belong together. The desirable property of a good design is to produce a highly cohesive classes. Comparison of different frameworks and thorough discussion can be found in Briand's work (Briand, Daly, & Wurst, 1997). Eder (Eder & Kappel, 1994) provided a comprehensive

Table 3. Comparison of attribute definition for coupling.

Attribute definition	Eder et al (Eder., 1994)	Hitz & Montazeri (Hitz, 1995)	Briand et al. (1999)
Public attribute visibility	X		
Method references attribute		X	
Method invokes method	X	X	X
Aggregation	X		X
Class type as a parameter in method signature or return type	X		X
Method's local variable is a class type	X		
A method invoked from within another method signature or return type	X		
Inheritance	X		
Method receives pointer to method			X

Source: Briand, Daly, and Wurst (1999).

framework that requires semantic analysis of classes and methods. The metrics of Chidamber define LCOM as the number of disjoint sets created by intersection of the n sets. The definition in Eq. 1 does not state how the inheritance of methods and attributes is treated with regard to method override and overload and the depth of the inheritance tree.

$$LCOM = \begin{Bmatrix} |P| - |Q|, if\ |P| > |Q| \\ 0, otherwise \end{Bmatrix} \qquad (1)$$

In Henderson-Sellers (1996), the cohesion measure is based on the number of attributes referenced by a method:

$$LCOM = \frac{\frac{1}{a}\sum_{j=1}^{a}\mu\left(A_j\right) - m}{1 - m} \qquad (2)$$

where a means number of attributes, m denotes methods, and $m(A_j)$ denotes the measure that yields 0 if each method in the class references all attributes, and 1 if each method in a class references only single attributes.

Complexity Measures

The modularized code encapsulating object interactions is characterized by class hierarchies. One of the problems in software development is that some kinds of behavior or functionality *crosscut* or are *orthogonal* to classes in many object-oriented components, and they are not easily modularized to a separate class. Examples of such behavior include the following: synchronization and concurrency, performance optimization, exception handling and event monitoring, coordination and interaction protocols, and object views. These changes in code lead to chaotic development, additional subclassing, and restructuring of the class hierarchies. They typically result in increased complexity of the code and disorder.

Cyclomatic complexity is the most widely used member of a class of static software metrics. Cyclomatic complexity as a broad measure of soundness and confidence for a program was introduced by Thomas McCabe in 1976. It measures the number of linearly independent paths through a program module. This measure provides a single ordinal number that can be compared to the complexity of other programs. Cyclomatic complexity is often referred to simply as program complexity, or as McCabe's complexity (McCabe, 1994).

Entropy-based complexity measures are based on the theory of information. The approach taken by Davis and LeBlanc (1988), who quantified the differences between *anded* and *neted* structures, is an unbiased estimate of the probability of occurrence of event m. This measurement is based on chunks of FORTRAN and COBOL code.

In 1976, Belady and Lehman elaborated on the law of increasing entropy: the entropy of a system (level of its unstructuredness) increases with time, unless specific work is executed to maintain or reduce it. An increase of entropy can result in severe complications when a project has to be modified and is generally an obstacle of maintenance.

The use of entropy as a measure of information content has been around since 1992, when it was introduced by Harrison (1992). Harrison's software complexity metric is based on empirical program entropy. A special symbol, reserved word, or a function call is considered as an operator (Zweben & Hasltead, 1979). (It is assumed that they have certain natural probability distributions.) The probability p_i of i^{th} most frequently occurring operator is defined as

$$p_i = {f_i}/{N_i} \tag{3}$$

where f_i is the number of occurrences of the i^{th} operator, and N_i is total number of nonunique operators in the program.

The entropy is defined as

$$H = -\sum_{i=1}^{Ni} p_i \log_2 p_i \tag{4}$$

Harrison's *Average Information Content Classification* (*AICC*) measure definition is shown in Eq. (5):

$$AICC = -\sum_{i=1}^{Ni} \frac{f_i}{N_i} \log_2 \frac{f_i}{N_i} \tag{5}$$

Harrison assessed the performance of these entropic metrics in two commercial applications written in C language with total number of lines of code over 130,000.

The work of Bansiya et al. (1999) introduced a similar complexity measure—*Class Definition Entropy* (*CDE*)—replacing the operators of Harrison with name strings used in a class. The assumption that all name strings represent approximately equal information is related to the possible error insertion by misusing the string. The metric has been validated on four large projects in C++, and results have been used to estimate the *Class Implementation Time Complexity* measure.

CRITICAL VIEW OF OBJECT-ORIENTED METRICS

Many metrics deal predominantly with static characteristics of code. Hitz and Montazeri (1995) clearly distinguished the difference between *static* and *dynamic class method invocation*: *number of methods invoked by a class compared to frequency of method invocation*. A metrics suite capable of capturing the dynamic behaviors of objects in regard to coupling and complexity was presented by Yacoub (Yacoub, Ammar, Hany, & Robinson, 1999). The dynamic behavior of an implementation is described by a set of scenarios. The *Export* and *Import Object Coupling* metrics are based on the percentage of message exchange between class instances (objects) to the total number of messages. The *Scenario Profiles* introduce the estimated probability of the scenario execution. The complexity metrics are aimed predominantly at assessment of stability of active objects as frequent sources of errors.

Obvious criticism of the Henderson-Sellers metrics include the typical interaction among objects, for example, treatment of inheritance, and in particular, the access to superclass attributes (private and protected) or method-to-method calls. Many metrics show dimensional inconsistencies, or their results are derived from correlative or regression analysis. As reported in Gursaran and Gurdev (2001), experienced object-oriented designers found memory management and run time errors to be more problematic and difficult to deal with.

Class size problems represent confusing effects with regard to the validity of object-oriented metrics. The confounding effect of class size has been reported by Khaled, Benlarbi, Nishith, and Shesh (2001), provoking some doubts with regard to the validity of the software metrics currently being used as early quality indicators. The relationship of high coupling factor to faults proneness seems to support the hypothesis that large class size may have a negative impact on quality and occurrence of faults.

Harrison's entropic metrics are intended to order programs according to their complexity. Because entropy provides only the ordinal position, thus restricting the way of usage, we cannot measure the "distance" between two programs.

SOME REMARKS ON OBJECT-ORIENTED METRICS

Object-oriented code is characterized by class hierarchies, which are shared structures. Often, some additional subclassing, modification to existing classes, the restructuring of the hierarchy, and changes in the visibility of attributes and sometimes even methods are made. Given these changes and possibly a lack of comprehensive documentation and time constraints, we assume that class hierarchies will become the subject of disorder and exhibit entropic tendencies in the actual code as well as in content. In some mobile agent applications, we observed that the probability that a subclass will not consistently extend the content of its superclass is increasing with the depth of hierarchy.

The entropy metrics are useful in ranking different modules and symbols with regard to their complexity. The single-valued measure of complexity is appealing as a quality indicator. However, as also discussed in Fenton's book (1997), the results may not be suitable for use in prediction models or as guidance for improving the quality of the product.

Finally, we need to establish a framework for measuring product quality by associating each metric or a set of metrics with a qualifier, such as maintainability, reusability, etc. Furthermore, for each metric, a range of threshold and "desirable" values must be established depending on application objectives and framework focus.

REFERENCES

Bansiya, J., Davis, C., & Etzkorn, L. (1999). An entropy-based complexity measure for object-oriented designs. *Theory and Practice of Object Systems*, *5*(2), 11–118.

Belady, L. A., & Lehman, M. M. (1976). A model of a large program development. *IBM Systems Journal*, *15*(3), 225–252.

Briand, L., Daly, J., & Wurst. (1999). A unified framework for coupling measurement in object oriented systems. *IEEE Transactions on Software Engineering, 25*(1), 99–121.

Briand, L., Daly, J., & Wurst. (1997). A unified framework for cohesion measurement in object oriented systems. *Technical Report*, ISERN-97-05.

Chidamber, S., & Kemerer, C. (1994). A metric suite for object oriented design. *IEEE Transactions on Software Engineering*, *20*(6), June, 476–549.

Churcher, N., & Shepperd, M. (1994). A metric suite for object oriented design. *IEEE Transactions on Software Engineering*, *20*(6), June, 476–549.

Davis, J. S., & LeBlanc, R. J. (1988). A study of the applicability of complexity measures. *IEEE Transactions on Software Engineering*, *14*(9), 1366–1371.

Eder, J., & Kappel, G. (1994). Coupling and cohesion in object oriented systems. Technical Report. University of Klagenfurt.

Fenton, N., & Pfleger, S. L. (1997). *Software metrics: A rigorous and practical approach*. International Thomson Computer Press.

Fetchke, T. (1995). Software Metriken bei der Objectorientierten Programmierung. Diploma Thesis. GMD, St. Augustine.

Gursaran, & Gurdev, R. (2001). On the applicability of Weyuker Property 9 to object oriented structural inheritance complexity metrics. *IEEE Transactions on Software Engineering*, *27*(4), April.

Harrison, W. (1992). An entropy-based measure of software complexity. *IEEE Transactions of Software Engineering, 18*(11), November, 1025–1029.

Henderson-Sellers, B. (1996). *Object-oriented metrics measures of complexity*. New York: Prentice Hall PTR.

Hitz, M., & Montazeri, B. (1995). Measuring product attributes of object-oriented systems. In *Proceedings of the Fifth European Software Engineering Conference (ESEC'95*), Barcelona, Spain (pp. 124–136).

Hitz, M., & Montazeri. (1996). B. Chidamber and Kemerer's metric suite: A measurement theory perspective. *IEEE Transactions on Software Engineering, 22*(4), 270–276.

Khaled, E., Benlarbi, S., Goel, N., & Rai, S. N. (2001). The confounding effect of class size on validity of object-oriented metrics. *IEEE Transactions on Software Engineering*, *27*(7).

McCabe, T. J., & Watson, A. H. (1994, December). Software complexity. *Crosstalk, Journal of Defense Software Engineering*, *7*(12), 5–9.

Weyuker, E. J. (1988). Evaluating software complexity measures. *IEEE Transactions on Software Engineering, 14*(9), 1357–1365.

Yacoub, S. M., Ammar, H. H., & Robinson, T. (1999). Dynamic metrics for object oriented designs. In *Proceedings of the Sixth International Symposium on Software Metrics (METRICS'99)* (pp. 50–61). Boca Raton, FL, November 4–6.

Zuse, H. (1994). *Software complexity metrics/analysis*. In J. Marciniak (Ed.), Encyclopedia of software engineering (Vol. 1, pp. 131–166). New York: John Wiley & Sons.

Zweben, S., & Hasltead, M. (1979). The frequency distribution of operators in PL/I programs. *IEEE Transactions of Software Engineering, SE-5*(March), 91–95.

KEY TERMS

Axiom: This is a generally accepted proposition or principle, sanctioned by experience; a universally established principle or law that is not necessarily true.

Cohesion: This is the degree to which elements within a class are related to one another and work together to provide well-bounded behavior.

Complexity: This is the degree to which a system or component has a design or implementation that is difficult to understand and verify. The first and still classic measure of complexity is that introduced by Kolmogorov, which is the shortest computer program capable of generating a given string.

Coupling: Coupling between classes is a measure of strength of association established by the communication link between two objects.

Cyclomatic Complexity: This is a broad measure of soundness and confidence for a program. This measure was introduced by Thomas McCabe in 1976.

Information Theory: In this method, probability and ergodic theory are employed to study the statistical characteristics of data and communication systems and coding theory, which uses mainly algebraic and geometric tools to contrive efficient codes for various situations. Introduced by Claude Shannon in 1948, it was a revolutionary new probabilistic way of thinking about communication and the first truly mathematical theory of entropy.

Measurement: When measuring some attribute of a set of elements, we can go through the process of assigning numbers or other symbols to the elements in such a way that relationships of the numbers or symbols reflect relationships of the attribute being measured.

Measurement Theory: This is a branch of applied mathematics that is useful in measurement and data analysis. The fundamental idea of measurement theory is that measurements are not the same as the attribute being measured.

Object-Oriented Software Reuse in Business Systems

Daniel Brandon
Christian Brothers University, USA

INTRODUCTION

"Reuse [software] engineering is a process where a technology asset is designed and developed following architectural principles, and with the intent of being reused in the future" (Bean, 1999). "If programming has a Holy Grail, wide-spread code reuse is it with a silver bullet. While IT has made and continues to make laudable progress in our reuse, we never seem to make great strides in this area" (Grinzo, 1998). "The quest for that Holy Grail has taken many developers over many years down unproductive paths" (Bowen, 1997). This article is an overview of software reuse methods, particularly object oriented, that have been found effective in business systems over the years.

BACKGROUND

Traditional software development is characterized by many disturbing but well documented facts, including:

- Most software development projects "fail" (60%) (Williamson, 1999).
- The supply of qualified IT professionals is much less than the demand (*www.bls.gov*, 2003).
- The complexity of software is constantly increasing.
- IT needs "better," "cheaper," "faster" software development methods.

Over the years, IT theorists and practitioners have come up with a number of business and technical methods to address these problems and improve the software development process and results thereof. Most notable in this sequence of techniques are: CASE (Computer Aided Software Engineering), JAD (Joint Application Development), Prototyping, 4GL (Fourth Generation languages), and Pair/Xtreme programming. While these methods have often provided some gains, none have provided the improvements necessary to become that "silver bullet". CASE methods have allowed development organizations to build the wrong system even faster; "wrong" in the sense that requirements are not met and/or the resulting system is neither maintainable nor adaptable. JAD methods tend to waste more of everyone's time in meetings. While prototypes can help better define user requirements, the tendency (or expectation) that the prototype can be easily extended into the real system is very problematic. The use of 4GL languages only speeds up the development of the parts of the system that were easy to make anyway, while unable to address the more difficult and time consuming portions. Pair programming has some merits, but stifles creativity and uses more time and money.

The only true "solution" has been effective software reuse. Reuse of existing proven components can result in the faster development of software with higher quality. Improved quality results from both the use of previous "tried and true" components and the fact that standards (technical and business) can be built into the reusable components (Brandon, 2000). There are several types of reusable components that can address both the design and implementation process. These come in different levels of "granularity," and in both object oriented and non-object oriented flavors.

"Software reuse received much attention in the 1980s but did not catch on in a big way until the advent of object oriented languages and tools" (Anthes, 2003). In Charles Darwin's theory of species survival, it was the most adaptable species that would survive (not the smartest, strongest, or fastest). In today's fast moving business and technical world, software must be adaptable to survive and be of continuing benefit. Object oriented software offers a very high degree of adaptability. "Object technology promises a way to deliver cost-effective, high quality and flexible systems on time to the customer" (McClure, 1996). "IS shops that institute component-based software development reduce failure, embrace efficiency and augment the bottom line" (Williamson, 1999). "The bottom line is this: while it takes time for reuse to settle into an organization – and for an organization to settle on reuse – you can add increasing value throughout the process" (Barrett, 1999). We say "object technology," not just adopting an object oriented language (such as C++ or Java), since one can still build poor, non-object oriented, and non-reusable software even using a fully object oriented language.

TYPES AND APPLICATIONS OF REUSE

Radding defines several different types of reusable components (Radding, 1998), which form a type of "granularity scale":

- GUI widgets - effective, but only provide modest payback
- Server-Side components - provide significant payback but require extensive up-front design and an architectural foundation.
- Infrastructure components - generic services for transactions, messaging, and database … require extensive design and complex programming
- High-level patterns - identify components with high reuse potential
- Packaged applications - only guaranteed reuse, … may not offer the exact functionality required

An even lower level of granularity is often defined to include simple text files, which may be used in a number of code locations such as "read-me" and documentation files, "help" files, Web content, business rules, XML Schemas, test cases, and so forth. Among the most important recent developments of object oriented technologies is the emergence of design patterns and frameworks, which are intended to address the reuse of software design and architectures (Xiaoping, 2003). The reuse of "patterns" can have a higher level of effectiveness over just source code reuse. Current pattern level reuse includes such entities as a J2EE Session Façade or the .Net Model-View-Controller pattern.

Reusing code also has several key implementation areas: application evolution, multiple implementations, standards, and new applications. The reuse of code from prior applications in new applications has received the most attention. However, just as important is the reuse of code (and the technology embedded therein) within the same application.

Application Evolution

Applications must evolve even before they are completely developed, since the environment under which they operate (business, regulatory, social, political, etc.) changes during the time the software is designed and implemented. This is the traditional "requirements creep". Then after the application is successfully deployed, there is a constant need for change.

Multiple Implementations

Another key need for reusability within the same application is for multiple implementations. The most common need for multiple implementations involves customizations, internationalization, and multiple platform support. Organizations whose software must be utilized globally may have a need to present an interface to customers in the native language and socially acceptable look and feel ("localization"). The multiple platform dimension of reuse today involves an architectural choice in languages and delivery platforms.

Corporate Software Development Standards

Corporate software development standards concern both maintaining standards in all parts of an application and maintaining standards across all applications. "For a computer system to have lasting value it must exist compatibly with users and other systems in an ever-changing Information Technology (IT) world" (Brandon, 2000). As stated by Weinschenk and Yeo, "Interface designers, project managers, developers, and business units need a common set of look-and-feel guidelines to design and develop by" (Weinschenk, 1995). In the area of user interface standards alone, Appendix A of Weinschenk's book presents a list these standards; there are over 300 items (Weinschenk, 1997). Many companies today still rely on some type of printed "Standards Manuals".

EFFECTIVE SOFTWARE REUSE

Only about 15% of any information system serves a truly original purpose; the other 85% could be theoretically reused in future information systems. However, reuse rates over 40% are rare (Schach, 2004). "Programmers have been swapping code for as long as software has existed" (Anthes, 2003). Formal implementation of reuse in various forms of software reuse has been a part of IT technology since the early refinements to 3GL's (Third Generation Languages). COBOL had the "copy book" concept where common code could be kept in a separate file and used in multiple programs. Almost all modern 3GLs have this same capability, even today's Web based languages like HTML and JavaScript on the client side, and PHP (on the server side). HTML has "server side includes," JavaScript has ".js" and ".css" files, and PHP has "require" files (".inc"). Often used in conjunction with these "include" files is the procedure capability where some code is compartmentalized to perform a particular task and where code can be sent arguments and possibly also return arguments. In different 3GLs this might be called "subroutines," or in modern languages, "functions". A function "library" is a separate file of one or more functions, and depending on the language may be pre-compiled.

Object oriented methods are concerned with the design and construction of modules of code that can exhibit certain characteristics. The key characteristics are encapsulation, composition, inheritance (generalization and specialization), and polymorphism. The code modules are typically called "object types" at the design level and "classes" at the implementation level. These classes contain both form (data) and functionality (functions). Encapsulation involves public "access functions" media access to the private data of a class to preserve both data security and integrity. A class can be "composed" of other classes, and this provides one form of code reuse. A class can also be derived from another class (a more general class), and this provides another form of reuse; derived classes (more specific classes) inherit the form and functionality of their base class and can add additional form or functionality and also modify functionality; this is the polymorphism aspect.

As an example of object oriented application, consider a large GUI application that may have hundreds of windows. Suppose the corporate design is to have each window show the company name on the top left and the date/time on the top right. Later a new boss decides he wants the company name on the top right and the date/time on the top left. If the application had been written in a non-object oriented language, then we would have to go back and modify hundreds of windows. If instead we had used an environment that supported inheritance (such as C++, Java, or PHP) and had derived all our windows from a base window class, then we would only have to change the one base class. If we used only an "include file" from which to get the company name, that would allow us to easily change the company name, but not to change the place where the name appeared in the window (unless we had clairvoyantly foreseen such a possibility, and designed a name location parameter in our file).

In most organizations, software reusability is still a goal that is very elusive; as said by Bahrami, "a most difficult promise to deliver on" (Bahrami, 1999). Radding stated: "Code reuse seems to make sense, but many companies find there is so much work involved, it's not worth the effort. …In reality, large-scale software reuse is still more the exception than the rule" (Radding, 1998). Bean in "Reuse 101" states that the current decreased "hype" surrounding code reuse is likely due to three basic problems (Bean, 1999):

- Reuse is an easily misunderstood concept.
- Identifying what can be reused is a confusing process.
- Implementing reuse is seldom simple or easy to understand.

Grinzo also lists several reasons and observations on the problem of reuse (Grinzo, 1998), other than for some "difficult to implement but easy to plug-in cases" such as GUI widgets: a "nightmare of limitations and bizarre incompatibilities," performance problems, "thorny psychological issues" involving programmers' personalities, market components that are buggy and difficult to use, fear of entrapment, component size, absurd licensing restrictions, or lack of source code availability.

Schach lists and describe the impediments to reuse as:

- Too many IS professionals would rather rewrite than reuse.
- Many IS professionals are skeptical on the safety and quality of using components built for other applications.
- Reuse artifacts are not stored and cataloged in a useful manner.
- Reuse can be expensive (building a reuse process, maintaining the reuse library, and verifying the fitness of the artifact for the new application).

Some organizations try to promote software reusability by simply publishing specifications on class libraries that have been built for other in-house applications or that are available via third parties, some dictate some type of reuse, and other organizations give away some type of "bonus" for reusing the class libraries of others (Bahrami, 1999). But more often than not, these approaches typically do not result in much success.

"It's becoming clear to some who work in this field that large-scale reuse of code represents a major undertaking" (Radding, 1998). " An OO/reuse discipline entails more than creating and using class libraries. It requires *formalizing* the practice of reuse" (McClure, 1996).

There are generally two key components to formalizing an effective software reuse practice both within an application development and for new applications (Brandon, 2002). These components are:

1. Defining a specific information technology architecture within which applications would be developed and reuse would apply.
2. Defining a very specific object oriented "reuse foundation" that would be implemented within the chosen IT architecture

Once the technical issues of an architecture and a reuse foundation are addressed, then management issues need to resolved, namely: "procedures, disciplines, and tools for tracking, managing, searching, and distributing software assets" (Anthes, 2003). Procedures and disciplines have to be formulated and enforced as part of the software development management and organizational "culture". Software can be built or procured to handle the searching, tracking, cataloging, and distribution issues.

FUTURE TRENDS

Several future trends will be key to the continuing success of software reuse. The first trend is for companies to adopt the necessary management principles that foster reuse, including programming incentives. Programmers must be rewarded for reusing software, and also for producing software that can be reused.

Another trend is for software to be developed that facilitates "library" functions for reusable components. Vendors and products in this area today include: CMEE (Component Manager Enterprise Edition) from Flashline (http://www.ejbean.com 2004), Logidex byLogicLibrary, and Component Manager by Select Business Solutions (*www.selectbs.com,* 2004). International standards like UML (Unified Modeling Language) and RAS (Reusable Asset Specification) (www.rational.com, 2001) will make these tools more interoperable. As well as "library" type tools, other technologies and tools will facilitate reuse that is based on standards such as "Web services" and other messaging technologies that will let software be reused "where it sits".

Still another trend is for companies to devote programming resources specifically to developing reusable component, while other programmers use the components built by this devoted group.

CONCLUSION

"If you want reuse to succeed, you need to invest in the architecture first" (Radding, 1998). "Without an architecture, organizations will not be able to build or even to buy consistently reusable components". In terms of general IT architectures for business systems, there are historically several types, such as: central computer, file services, two or three tier client server, and two or three tier Internet (browser) based. Various transaction processing and database vendors have their own "slants" on these basic approaches, which may depend upon how business logic and the database are distributed.

Today companies are mainly interested in the last of these categories. Internet based applications are becoming the preferred way of delivering software based services within an organization (intranets), to the worldwide customer base via browsers and "net appliances" (Internet), and between businesses (extranets). Vendor independent and "open" architectures are often preferred, and the "multiple platform" dimension of reusability is handled by using server generated HTML and JavaScript (via Java or PHP programs on the server).

As has been concluded by several authors, "A reuse effort demands a solid conceptual foundation" (Barrett, 1999). One such foundation was presented in Brandon (2002). It is based on the key object oriented principles of inheritance and composition. By establishing this foundation, an organization can effectively begin to obtain significant reusability since programmers must inherit their class from one of the established classes and they must only compose their class of the established pre-built components.

REFERENCES

Athens, G. (2003, July 28). Code reuse gets easier. *Computerworld,* 24-26.

Bahrami, A. (1999). *Object oriented systems development.* Irwin McGraw Hill.

Barrett, K., & Schmuller, J. (1999, October). Building an infrastructure of real-world reuse. *Component Strategies,* 1-5.

Bean, J. (1999, October). Reuse 101. *Enterprise Development.*

Bowen, B. (1997). Software reuse with Java technology: Finding the Holy Grail. *www.javasoft.com/features/1997/may/reuse.html*

Brandon, D. (2000). An object oriented approach to user interface standards. In *Challenges of information technology in the 21st century*. Hershey, PA: Idea Group Inc.

Brandon, D. (2002). Achieving effective software reuse for business systems. In *Successful software reengineering.* Hershey, PA: Idea Group Inc.

Grinzo, L. (1998, September,).The unbearable lightness of being reusable. *Dr. Dobbs Journal.*

McClure, C. (1996). Experiences from the OO playing field. *Extended Intelligence.*

Radding, A. (1998, November 9). Hidden cost of code reuse. *Information Week,* 1a-8a.

Reifer, D. (1997). *Practical software reuse.* Wiley Computer Publishing.

Schach, S. (2004). *Introduction to object oriented analysis and design.* Irwin McGraw Hill.

Weinschenk, S., Jamar, P., & Yeo, S. (1997). *GUI design essentials.* John Wiley & Sons.

Weinschenk, S., & Yeo, S. (1995). *Guidelines for enterprise wide GUI design.* John Wiley & Sons.

Williamson, M. (1999, May). Software reuse. *CIO Magazine.*

www.bls.gov (2003). US Department of Labor, Bureau of Labor Statistics.

www.ejbean.com/products/related/flashline_cm2.html (2004). Component Manager, Flashline.

www.rational.com/rda/ras/preview/index.htm (2001). Reusable Asset Specification, Rational Rose.

www.selectbs.com (2004). Component manager, select business solutions.

Xiaoping, J. (2003). *Object oriented software development using Java.* Addison Wesley.

KEY TERMS

Class: A program construct representing a type of thing (abstract data type) that includes a definition of both form (information or data) and functionality (methods).

Composition: A new class in an objected programming language that is composed of other classes.

Encapsulation: The ability to insulate data in a class so that both data security and integrity is improved.

Framework: A software foundation that specifies how a software system is to be built. It includes standards at all levels, both internal construction and external appearance and behavior.

Function: A programming construct where code that does a particular task is segregated from the main body of a program; the function may be sent arguments and may return arguments to the body of the program.

Implementation: The code placed inside of methods. For some languages this code is pre- compiled or interpreted.

Include: Some code stored separately from the main body of a program, so that this code can be used in many programs (or multiple places in the same program).

Inheritance: A feature of object oriented languages that allows a new class to be derived from another class (a more general class); derived classes (more specific classes) inherit the form and functionality of their base class.

Interface: The specification for a method ("what" a method does); how that function is called from another program. Interfaces are provided in source form as opposed to implementations, which are secure. This allows one to use a method without regard for "how" that method is coded. It also allows multiple implementations of the same interface.

Libraries: A group of functions and/or classes stored separately from the main body of the main program; an "include" file consisting of functions and/or classes.

Method: A function defined inside of a class.

Packages: Similar to a library, but just containing classes.

Patterns: A software library for a common business scenario. A framework may be a design framework (possibly expressed in UML [Unified Modeling Language]) or an implementation framework (possibly in C++, Java, or PHP).

Polymorphism: The ability of object oriented programs to have multiple implementations of the same method name in different classes in an inheritance tree. Derived classes can override the functionality defined in their base class.

Reuse: Reuse (software) is a process where a technology asset (such as a function or class) is designed and developed following specific standards, and with the intent of being used again.

Separation: The separation of what a method does (interface) from how the method does it (implementation).

O

Observations on Implementing Specializations within an IT Program

Erick D. Slazinski
Purdue University, USA

INTRODUCTION

With a projected 2.26 million additional jobs to fill in various computer fields by the year 2010, there are and will continue to be ample job opportunities in the computer industry. However, the computer field is far too broad for one individual to be an expert in the entire field. Therefore it may be more useful for students to have the opportunity to concentrate their studies in a specific interest area within a broader Information Technology (IT) degree.

IT educators throughout the United States (US) have paid attention to the needs and demands of the IT industry. To address the need for IT graduates with specialized skills, many of the leading universities have created programs which allow undergraduate students to specialize or focus their studies.

This chapter will discuss findings on the state of IT programs with regards to their course offerings. One area of specialization, or track, is presented as an example. It will be noted that even within a specialty area, there can be further specializations. In addition to supporting the students pursuing the specialty area, general knowledge courses must also be offered to those pursuing other specialty areas.

BACKGROUND

The Bureau of Labor Statistics reported 2.9 million computer-related jobs in 2000, with an expected 4.89 million computer jobs by the year 2010. Considering new jobs as well as replacements, over 2.26 million additional people will be needed to fill these jobs (Hecker, 2001). The fluid nature of the IT industry makes generalizations difficult. Therefore, skills are often categorized or grouped together into skill sets (or job descriptions). The most common clustering of skills has been summarized in Table 1.

Of these pathways (or specialization tracks), two of the top occupations (as predicted by the US Dept of Labor) are systems analysis and database administrators (which have been grouped in the EDC Information Support and Service Pathway). See Table 2 for a listing of the top growth occupations.

WHERE ARE THE SPECIALIZED IT PROGRAMS?

Published curriculum from the institutes who attended the Conference for IT Curriculum (CITC) II held in April of 2002 were used as the sample set. The conference attendees were primarily IT educators from around the US, who had an interest in IT curriculum issues. An IT curriculum is focused on the application of technologies to solve problems. To differentiate, a traditional Computer Science curriculum is focused on algorithm design.

Table 3 illustrates, that out of the 28 programs studied, 50% (14) had some specialization available for students. Of the 14 programs that offered specializations, 45% (6) of those offered at least a database specialization similar to our sample track.

NEED FOR A DATABASE TRACK

The same data from the Bureau of Labor Statistics (Hecker, 2001) indicates there were 106,000 jobs for database

Table 1. Summary of educational pathways

Network Systems Pathway
Information Support and Service Pathway
Programming and Software Development Pathway
Interactive Media Pathway
From EDC (2002)

Table 2. Percentage change in employment, projected 1998-2008

Occupation	Percent change
Computer engineers	108
Computer support specialists	102
Systems analysts	94
Database administrators	77

From U.S. Dept. of Labor

administrators (DBAs) in 2000, with a projected 176,000 openings to fill by the year 2010. In addition to DBAs, there are also database professionals who specialize in database architecture and database programming.

ANATOMY OF A DATABASE TRACK

Though there are various job titles given to database activities in the workplace, such as these listed on the ITWORKS-OHIO website: Data Analyst, Database Administrator, Database Analyst, Database Developer, and Database Specialist—many others exist in the marketplace. However, based on the author's opinion, when the job descriptions are examined, one can find that there are really three, inter-related roles. Figure 1 illustrates these roles using the structure of a house as an example. The DBA (as the foundation) keeps a database available, secure, and healthy. Without the DBA, the rest of the database team could not function. The database developer (as the framing of the house) encodes the business logic in the database. Additionally the database devel-

Table 3. Programs and specializations

INSTITUTION NAME	SPECIALAZATION	DATABASE SPECIALIZATION
Ball State	NO	
Bentley	NO	
Brigham-Young University (BYU)	NO	
BYU-Hawaii	NO	
BYU-Idaho	NO	
Capella	YES	NO
Drexel	YES	YES
Florida State University	YES	NO
Georgia Southern	NO	
George Mason University	YES	NO
Hawaii at Manoa	NO	
Houston	NO	
Indiana University	YES	NO
Indiana University Purdue University at Indianapolis	YES	NO
Macon State	YES	YES
New Jersey Institute of Technology	YES	NO
Northern Alabama	YES	NO
Pace Univerisy	NO	
Pennsylvania College of Technology	YES	NO
Purdue University	YES	YES
Purdue University - Calumet	YES	YES
Rochester Institute of Technology	YES	YES
Southern Alabama	YES	YES
State University of New York (SUNY) Morrisville	NO	
Towson University	NO	
University of Baltimore	NO	
University of Cincinnati-Clermont	NO	
USCS	NO	

oper often develops the interface layer between the system software and the database engine. Lastly, the database architect (as the roof of the house) is often a senior staff member. Typical duties include determining the amount of business logic to be encoded in the database, developing the database design (including distribution of data, data flows, etc.) and often overseeing the implementation of the database. Like the components of a house, all three roles are co-dependant and necessary. In addition to the specific duties, each member must maintain several lines of communication within the development team and organization.

The database administrator must often work with the system/network administrators when it comes to the purchasing of hardware, determining network traffic and bandwidth requirements, coordinating the use of backup devices, etc. They often work with the database architects in providing recommendations for physical implementation issues. For the database developer, they must interact with the database architect (they are implanting the design from the architect) as well as the system developers and ensure that the developers have access to the data they need while maintaining the integrity of the database. The database architect must work with the software architect. Together they determine how to partition the business logic of the system. The database architect works with the DBA in determining the best physical implementation of the supporting database. And lastly the architect often coordinates the activities of the database developers.

Based on these responsibilities and communication channels, the database track was designed to produce students prepared for these multi-faceted jobs. During the first two years of our program, all students gain a broad overview of the Information Technology field, taking introductory courses in programming, Internet technologies, architectures, telecommunications, database, and systems analysis and design

- **Introduction to Application Development:** Introduction to system development using MS Access.
- **Introduction to Computer Programming:** Application development using Visual Basic.
- **Internet Foundations and Technologies:** Web page development using XHTML.
- **Information Technology Architectures:** Explores the history, architecture, and development of the Internet and the World Wide Web.
- **Systems Software and Networking:** Introduction to data communications and Network Operating systems.
- **Programming for the Internet:** Internet application development using scripting languages.
- **Database Fundamentals:** Normalization, SQL, and application interfaces to databases.
- **Systems Analysis and Design Methods:** Introduction to information systems development.

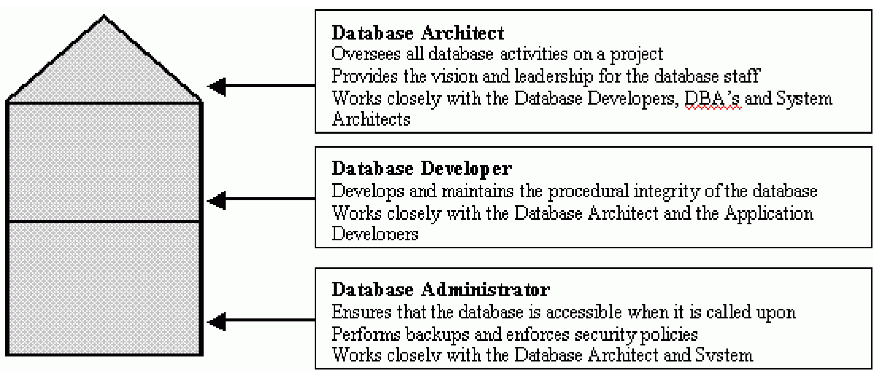
COURSES IN THE DATABASE TRACK

Students specializing in the database area must complete an additional 25 credit hours of technical courses focused on getting the database student prepared for one of the three roles described above. The progression of the courses is shown in Figure 2. Then it will be shown how these courses, in combination with other technical electives, prepare our students for future jobs.

A brief description of these database courses follows.

- **Introduction to Application Development (all students):** Introduces the development of information systems through the use of a database. Topics include business information systems, system and application development, database management

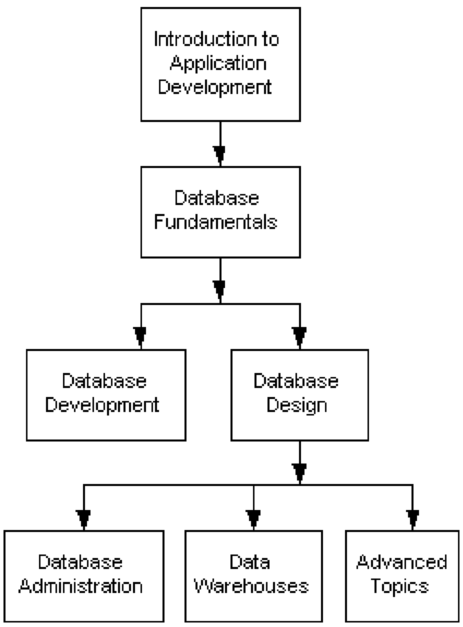
Figure 1. Database roles



systems, problem solving, logic, data types, and programming using database technology. Given a database design and application requirements, students design, construct, and test a personal computer information system.

- **Database Fundamentals (all students):** Looks at relational database concepts, including data design, modeling and normalization. Students use SQL to query, define, populate, and test a database. Expands on previous courses by accessing databases from programs and the Web, and discusses practical issues that database developers must handle.
- **Database Development:** Explores some of the programmatic extensions to SQL supported by leading Relational Database Management Systems (RDBMS) vendors. Topics include stored procedure and trigger design and implementation, query optimization to enhance performance, and data transformation to enhance interoperability of data.
- **Database Design and Implementation:** Deals with advanced design techniques and physical issues relating to enterprise-wide databases. Topics include advanced normalization, data distribution, distributed database design and replication, storage estimation and allocation, usage analysis, partitioning of very large tables, metadata analysis, data conversion, and load techniques.
- **Database Administration (elective):** Explores tools and techniques for managing an organization's database technology. Topics include database architecture, database technology installation, database creation and maintenance, RDBMS operations and troubleshooting, and database performance tuning.
- **Data Warehousing (elective):** Studies the design and implementation of data warehouses (including data marts and operational data stores) using current database technologies. Topics include data modeling for warehouses, data warehousing infrastructure and tool selection, data exploration, data synthesis and reduction, organizational metadata, and data warehouse administration.
- **Advanced Topics in Database Technology (elective):** Explores contemporary issues in the database arena. These issues may be related to new or breakthrough concepts, technologies, or techniques.

Figure 2. Database courses



Rounding Out the Students

Since database development has a significant analysis and design component, students are required to take a Systems Requirements Discovery and Modeling course, offered by our systems analysis group. Likewise many database personnel are often relegated to a distinct group with a project that has management requirements, so a Project Management course is also required. To support the further specialization into the database architect, database programmer and database administrator roles, two separate selective areas - database management selectives and database technology selectives - were defined. The students must choose a minimum of six credit hours (two courses) in each of the selective areas. The database management selectives are the elective courses listed with the database course flow in Figure 2 (above) – database administration, data warehousing and advanced topics in database management, as well as any graduate level database course offered during the student's senior year. The database technology selectives provide additional areas of exploration on topics that piqued the students' interest in their first two years. Again they are allowed to choose any two courses from the following:

- **Object-Oriented Programming:** use object-oriented programming languages (Java) in the development of modern, business applications.
- **Advanced Design Techniques:** advanced study of system design methods and techniques used by systems analysts to develop information systems.
- **Software Development Methodologies:** methodologies and practices commonly used in contemporary software development projects.

- **Enterprise Application Development:** component development and reuse, distributed object technologies, multi-tier applications.
- **Senior Software Development Project:** integrates the software development technologies and techniques taught in prior courses.
- **E-Commerce:** components of e-commerce.
- **Automatic identification and data capture:** real-time data feeds.
- Any other database management selective course(s) not already taken.

MODEL CURRICULA

At the time this chapter was written, there are at least three curricula models available from which to build from: the IEEE/ACM Computing Curricula; the EDC Academic Foundations and Pathways (EDC, 2002) and the ACM SIG ITE Curriculum Proposal. These models share more similarities than differences and selecting a model may be more a matter of preference of one set of authors over another. Considering the most mature model, the author has chosen to compare the specified database track against the IEEE / ACM Computing Curricula 2001.

The most recent IEEE / ACM Computing Curricula report, available at the ACM Web site, identifies Information Management (IM) as a knowledge area within their Computing Curricula 2001. The IM area of knowledge includes 14 components. Three are considered core, and 11 are viewed as electives. The courses in our database track provide significant coverage of the IM areas identified. Additionally, a Database Administration course that is beyond the model is offered. Table 4 shows the mapping from the IM components to our DB-track courses.

FUTURE TRENDS

For our University, the new curriculum implementing the track concept started in Fall 2001. In Spring 2003 the first cohort of Freshmen (who entered the program with the tracks in place) were required to select their track of interest. However, as news of the program's availability became known, students under the old plan of study transferred to the new plan of study. Additionally, transfer students were admitted under the new plan of study. These students plus an enthusiastic faculty caused the transition to the new course models to occur immediately after the new program (with tracks) was ratified by the University curriculum committee instead of waiting until the Spring 2003 semester.

This caused several registration issues which our counselors had to handle: multiple plans of studies; multiple course equivalencies (as courses were being brought online and retired); and pre-requisite issues. Of these the issue of tracking pre-requisites has been the most challenging because certain courses exist in multiple curriculums whose content has changed every time the curriculum has changed. Often counselors would have to

Table 4. Comparison of model curricula to database track courses

IM Knowledge Areas	DB Track
IM1: Information Models and Systems (core)	In 1st year course
IM2: Database Systems (core)	In 1st year course
IM3: Data Modeling (core)	2nd year: Database Fundamentals 3rd year: Database Design
IM4: Relational Databases (elective)	2nd year: Database Fundamentals 3rd year: Database Design
IM5: Database Query Languages (elective)	2nd year: Database Fundamentals (not OO queries)
IM6: Relational Databases Design (elective)	2nd year: Database Fundamentals 3rd year: Database Design
IM7: Transaction Processing (elective)	2nd year: Database Fundamentals 3rd year: Database Development
IM8: Distributed Databases (elective)	3rd year: Database Design 4th year Data Warehousing
IM9: Physical Database Design (elective)	3rd year: Database Design
IM10: Data Mining (elective)	4th year Data Warehousing
IM11: Information Storage and Retrieval (elective)	4th year Data Warehousing (partial)
IM12: Hypertext and Hypermedia (elective)	Candidate for 4th year Adv Topics
IM13 Multimedia Information and Systems (elective)	Candidate for 4th year Adv Topics
IM14: Digital Libraries (elective)	Candidate for 4th year Adv Topics

have professors validate that a student had the appropriate pre-requisite knowledge. This often led to questions such as "who taught this pre-requisite course?" or "did you cover this topic?"

CONCLUSION

Since the IT field is too broad to be comprised of only generalists, the IT workforce is becoming increasingly specialized. In response to this, IT educators have started offering areas of specialization or tracks that meet industry's needs. However, designing and implementing a tracked curriculum is a demanding task. The implementation costs may be higher than first estimated. If students are to get through a curriculum with multiple tracks in a reasonable time frame, the number of courses that need to be offered every semester can increase. This increase will require more resources could put a drain on other departmental offerings (such as service/graduate courses).

REFERENCES

Bachelor of Science in Information Technology. (2002). Retrieved July 13, 2002 from the World Wide Web at: *http://ite.gmu.edu/bsit/degree.htm*

Computer and Information Sciences. (2004). Retrieved June 30, 2004 from the World Wide Web at: *http://wwwnew.towson.edu/cosc//*

Computer Information Systems Programming and Database Processing Associate of Applied Science Degree. (2002). Retrieved July 13, 2002 from the World Wide Web at: *http://www.pct.edu/degrprog/pd.shtml*

Department of Computer Information Systems. (2004). Retrieved June 30, 2004 from the World Wide Web at: *http://www2.una.edu/business/cis.html*

Department of Information Systems. (2002). Retrieved July 13, 2002 from the World Wide Web at: *http://www.byui.edu/Catalog/2002-2003/_jim2.asp?departmentID=1888#680*

Education Development Center, Inc. (EDC). (2002, August) Information Technology Career Cluster Initiative Academic Foundations and Pathway Standards ValidationStudies. Retrieved November 29, 2004 from: *http://webdev2.edc.org/ewit/materials/ITCCIVSFinal.pdf*

Finkelstein, L., and Hafner, C. (2002). *The Evolving Discipline(s) of IT (and their relation to computer science): A Framework for Discussion*. Retrieved June 30, 2004 from the World Wide Web at: *http://www.cra.org/Activities/itdeans/resources.html*

Free On-Line Dictionary of Computing (FOLDOC). (2004). Retrieved June 30, 2004 from the World Wide Web at: *http://wombat.doc.ic.ac.uk/foldoc/foldoc.cgi*

Georgia Southern University – College of Information Technology. (2004). Retrieved June 30, 2004 from the World Wide Web at: *http://cit.georgiasouthern.edu/*

Hecker, D. (2001, November). Occupational employment projections to 2010. *Monthly Labor Review*. Retrieved January 6, 2002 from the World Wide Web at: http://*www.bls.gov/opub/mlr/2001/11/art4full.pdf*

Indiana School of Informatics – Undergraduate Program. (2004). Retrieved June 30, 2004 from the World Wide Web at: *http://www.informatics.indiana.edu/academics/undergrad.asp*

Information Systems. (2002). Retrieved July 13, 2002 from the World Wide Web at: *http://soc.byuh.edu/is/course.htm*

Information Systems and Operations Management. (2004). Retrieved June 30, 2004 from the World Wide Web at: *http://www.bsu.edu/web/catalog/undergraduate/programs/Programs02/isom02_cb.html*

Information Technology – Macon State College. (2002). Retrieved July 13, 2002 from the World Wide Web at: *http://www.maconstate.edu/it/it-bachelors.asp*

IT Concentrations – Information Technology Program. (2002). Retrieved July 13, 2002 from the World Wide Web at: *http://www.it.njit.edu/concentrations.htm*

Joint Computer Society of IEEE and Association for Computing Machinery. (2001, August 1). *Computing Curricula 2001 – Steelman Draft (August 1, 2001)*. Retrieved January 6, 2002 from the World Wide Web at: *http://www.acm.org/sigcse/cc2001/steelman/*

Morrisville State College – CIT. (2002). Retrieved July 13, 2002 from the World Wide Web at: *http://www.cit.morrisville.edu/index.html*

Occupational Area Definitions for Information Systems and Support. (2002). Retrieved July 11, 2002 from the World Wide Web at: *http://www.itworks-ohio.org/ISSdefinit.htm*

Purdue University Computer Technology. (2002). Retrieved July 13, 2002 from the World Wide Web at: *http://www.tech.purdue.edu/cpt/*

Required Courses for BSIS year 2000-02 Entrants. (2002). Retrieved July 13, 2002 from the World Wide Web at:

http://www.cis.drexel.edu/undergrad/bsis/required_2000+_3.asp#dms

School of Computer and Information Sciences. (2002). Retrieved July 13, 2002 from the World Wide Web at: *http://www.southalabama.edu/bulletin/cis.htm*

School of Computer Science and Information Systems at Pace University. (2004). Retrieved June 30, 2004 from the World Wide Web at: *http://csis.pace.edu/csis/index.html*

School of Technology – BS in Information Technology. (2002). Retrieved July 13, 2002 from the World Wide Web at: *http://www.capella.edu/aspscripts/schools/technology/bs_general.asp*

SIGITE. (2002, Sept 27). *IT Curriculum Proposal – Four Year Degrees*. Retrieved November 29, 2003 from the World Wide Web at: *http://site.it.rit.edu/*

Specialization in Computer Information Systems. (2002). Retrieved July 13, 2002 from the World Wide Web at: *http://business.ubalt.edu/DegreePrograms/ungrad_prog/special.html#CIS*

Undergraduate Programs in ICS. (2004). Retrieved June 30, 2004 from the World Wide Web at: *http://www.ics.hawaii.edu/academics/degree_programs/undergrad/index.jsp*

University of Cincinnati Undergraduate Programs Information Systems. (2004). Retrieved June 30, 2004 from the World Wide Web at: *http://www.uc.edu/programs/viewprog.asp?progid=1492*

U.S. Department of Labor Bureau of Labor Statistics. (2003). *Working in the 21st Century*. Retrieved November 29, 2003 from the World Wide Web at: *http://www.bls.gov/opub/working/home.htm*

USCS Computer Information Systems. (2004). Retrieved June 30, 2004 from the World Wide Web at: *http://www.uscs.edu/academics/cas/mcs/Catalog_info/ba_cis.html*

KEY TERMS

DBA: The title database administrator (DBA) represents an IT professional who ensures the database is accessible when it is called upon, performs maintenance activities, and enforces security policies.

IT Discipline: The ***intellectual gap*** in our educational frameworks for students who are interested in computing careers but find computer science too narrow, mathematical, and physical-science oriented, while MIS is insufficiently deep in technical content and too focused on traditional business topics and culture (Finkelstein, 2002).

Knowledge Area: Represents a particular sub-discipline that is generally recognized as a significant part of the body of knowledge that an undergraduate should know (IEEE, 2001).

Relational Database Management System (RDBMS): A suite of programs which typically manages large structured sets of persistent data, offering ad hoc query facilities to many users, which is based on the relational model developed by E.F. Codd (FOLDOC).

Selective List: A set of courses, grouped together for the purpose of filling an educational skill gap.

Systems Analysis: The design, specification, feasibility, cost, and implementation of a computer system for business (FOLDOC).

Systems Design: The approach used to specify how to create a computer system for business (FOLDOC).

Track: A series of courses designed around a topical area which is structured in a manner to efficiently develop a student's skill set.

Obstacles to SMEs for E-Adoption in the Asia Pacific Region

Sushil K. Sharma
Ball State University, USA

Nilmini Wickramasinghe
Cleveland State University, USA

INTRODUCTION

The e-commerce revolution has affected all organizations. Of particular interest is its effect on small and medium-sized enterprises (SMEs), because we can observe an interesting duality, where these companies are most likely to benefit from the opportunities afforded by e-commerce (because through e-commerce, it is possible to level the playing field), and yet they appear to be slowest in embracing many of these e-commerce possibilities. On several social and economic grounds, SMEs are of overwhelming importance in most of the Asia Pacific/Pacific region. In fact, they comprise well over 90% of all enterprises in this region and provide employment for over half of the region's workforce (Wattanapruttipaisan, 2002; APEC, 2001). Typically, the SME sector accounts for upwards of 90% of all firms outside the agricultural sector of East and Southeast Asia, and of Japan as well (Alphonso, 2001; Regnier, 2000; Tambunan, 2000; Wattanapruttipaisan, 2002; A Report Prepared for Asia Pacific Foundation, 2002).

Given the importance of SMEs to this region, it becomes crucial to understand the obstacles they are facing. First, we must understand why e-commerce adoption has been slow, if it occurs at all, for these SMEs, and then suggest changes to policy that may enable the alleviation of such obstacles, and hence, the successful embracing of e-commerce for these SMEs. What follows is a brief synopsis that serves to outline the key factors hindering SMEs' participation in e-commerce and the obstacles to SMEs for e-adoption in the Asia Pacific region.

BACKGROUND

Electronic commerce (e-commerce) succinctly stated refers to any transaction completed over a computer-mediated network that involves the transfer of ownership or rights to use goods or services (Tambunan, 2000); i.e., e-commerce can be defined as the conduct of buying, selling, and financially transacting by electronic means. E-commerce has been around for many years, particularly in the form of electronic data interchange (EDI). However, it is the Internet that is bringing it to the fore (Afuah & Tucci, 2000). The key aspect of e-commerce today is doing business over the Internet. E-commerce levels the playing field and lets small organizations compete with large organizations. The concept of e-commerce is about using the Internet to do business better and faster as well as shrinking traditional barriers of distance and time (Afuah & Tucci, 2000). Despite the downturn in the Internet economy represented by the crash of many dot-com companies, several forecasts continue to predict huge potential in global e-commerce over the next several years (Turpin et al., 1998). With an estimated 188 million users at year-end 2002, the Asia Pacific/Pacific region constitutes the world's second largest Internet market, ahead of North America (eMarketer Inc., 2002). South Korea, China, India, Singapore, Malaysia, and Indonesia are leading in ICT proliferation and e-commerce offerings as compared to other countries (Hill, 1998). ICT use is related to general economic indicators, such as gross domestic product (GDP), yet it is also affected by a wide range of other factors, including government policies, telecommunications costs, and social aspects (Beal, 2001; A Study Report on Thailand, 2001). Despite a sluggish economy in the Asia Pacific region, consumer demand for devices such as computers, mobile phones, and personal digital assistants (PDAs) has remained consistent throughout the region, and Internet usage is rising[1,2] (Debroy, 2002; UNCTAD, 2002; PricewaterhouseCoopers, 1999), yet the adoption of e-commerce by SMEs has been notably slow.

As the usage of the Internet is growing, the e-commerce and e-business initiatives in Asia Pacific are growing at a very fast rate. eMarketer projects that business-to-business (B2B) e-commerce revenues will grow to exceed $300 billion in the Asia Pacific/Pacific region by 2004 (eMarketer, 2002, www.emarketer.com). Despite such a high growth of e-commerce in Asia Pacific/Pacific and other Asia Pacific countries, the percentage of revenues generated through business-to-consumer (B2C) e-commerce transactions are still very low as compared to developed countries (Beal, 2001).

SMEs and E-Commerce

The adoption of e-commerce technologies is important for the ongoing survival of SMEs. The most significant business benefits for SMEs from the use of e-commerce are described below (Arunachalam, 1995; Sengenberger et al., 1990, Sharma et al., 2003, Afuah & Tucci, 2000):

- Better service quality—E-commerce can improve the quality of the marketing, sales, support, and procurement processes of SMEs by delivering more accurate, timelier, and more complete information to the point of sale, point of decision, or point of support. Service can also be delivered over a broader geographic and temporal base worldwide or anytime, anywhere.
- Reduced service costs—The self-service Internet-based e-commerce model is characterized by relatively low variable transaction costs, with increasing cost-effectiveness as the scale of activity increases. At higher volumes of activity, Internet commerce channels for sales and support services are likely to be low-cost channels.
- Increased revenue—Extended geographic sales channels and improved service quality may lead directly to increased market share, improved competition among brands, and greater revenue.
- Reduced time to complete a business transaction—By communicating electronically, the time required to place and confirm an order can be compressed by hours or, in some cases, days or weeks. This shortens the lead time for product delivery. As a result, it may be possible to reduce parts or finished goods inventories or receive critical products more rapidly to gain a competitive advantage.
- Reduced administrative costs—The cost of processing purchase requisitions, purchase orders, and payments can be dramatically reduced, as can invoice and bill presentation costs. The accuracy of business transactions is improved, increasing customer satisfaction; reducing transaction, auditing, and administrative expenses; as well as reducing the costs of expedited manufacturing and shipping costs to correct erroneous or late orders.
- Improved return on capital—By shortening the "product turn" and payment cycles, enterprises can reduce the amount of material requirements and the time for which funds must be committed to pay for the production of inventory, materially lowering capital requirements.
- Increased return on investment through better planning—More timely planning and information allows for more efficient acquisition and scheduling of capital equipment, reducing unit costs, increasing the return on investment, and ensuring a better match between manufacturing capacity, production, and market demand.
- Increased time devoted—SMEs now can devote more time to their products and their customers and lose less time in unproductive administrative jobs, such as time savings associated with electronic purchasing and materials management.
- Enhanced marketing—SMEs can enhance their local and international marketing efficiency through electronic catalogues and advertising through the Internet.
- Enhanced innovation capacity—SMEs can dramatically enhance their innovation capacity through technology watch, reverse and concurrent engineering, rapid prototyping, and distance manufacturing.
- Increased bidding opportunity—SMEs can bid on larger and international projects through networking with other partners worldwide and even with competitors.

OBSTACLES OR BARRIERS TO THE ADOPTION OF E-COMMERCE

The potential opportunities and benefits of e-commerce for SMEs include strengthening customer relationships, reaching new markets, optimizing business processes, reducing costs, improving business knowledge, attracting investment, and creating new products and services (APEC, 1999). Specifically, e-commerce represents an opportunity for SMEs to compensate for their traditional weaknesses in areas such as gaining access to new markets and gathering and diffusing information on a broad and international scale. Despite this, the following are the various obstacles that are experienced by Asia Pacific SMEs for their e-adoption that we have found through our research.

Lack of Awareness among SMEs about E-Commerce

This first and basic obstacle for adoption of e-commerce usage among SMEs in Asia Pacific is lack of awareness of e-commerce and the availability and access to the telecom infrastructure at a reasonable cost. This finding was confirmed by the APEC Report (2002). Many SMEs are found to be not aware of the developments taking place and the role they could play in this new marketplace. At times, the persons involved in SMEs found e-commerce literature too technical and complicated. Unless governments or other agencies simplify technical information for them, SMEs find it difficult to get involved in e-commerce concepts and implementation. Many Asia Pacific countries also found the English language a barrier, because the people involved conduct their business using local languages. The lack of awareness is closely linked to the fact that SMEs in Asia Pacific are usually slower in adopting new technologies given the often high investments necessary. Many of the Asia Pacific SMEs are also found to take less risks and to not experiment (APEC, 1999; Hall, 2002).

Lack of Critical Mass of Customers, Suppliers, and Business Partners

Another obstacle we found that was experienced by Asia Pacific SMEs is the lack of a critical mass among customers, suppliers, and business partners. The lack of a critical mass of these stakeholders is due to either ignorance or fears of a lack of security and privacy using electronic means. The e-commerce infrastructure is poor in Asia Pacific countries, and online shopping is not yet popular among the masses. Due to low use of e-commerce, there is not a large enough mass of customers and suppliers for e-commerce. This acts as a discouraging factor for SMEs to jump into the e-commerce revolution. There are not many marketplaces that could attract SMEs to take advantage (APEC, 1999). Until sufficient numbers of their main local customers or suppliers participate in online commerce activities, there is little incentive for individual SMEs to become engaged in e-commerce. In Asia Pacific countries, SMEs cited several factors contributing to low levels of customer e-commerce use, including language barriers and low levels of English fluency, lack of comfort and familiarity with e-commerce technologies, a cultural preference for more traditional trade practices involving face-to-face contact between buyer and seller, and continued preference for the use of cash in transactions. SMEs fear doing business with international marketplaces due to differing cultural backgrounds and a fear of being deceived due to lack of knowledge of new technologies.

Trust and Confidence

Lack of trust and confidence in various aspects of the electronic marketplace was identified as another main obstacle to the growth of the e-commerce market in general, and for SME engagement in e-commerce in particular. The security issue is perceived as very important across the Asia Pacific region, and the majority of SMEs have a fear of electronics. Security may not be a serious problem, but due to the low levels of technology diffusion and awareness among SMEs, it is still a psychological barrier for SMEs, as confirmed in various reports (APEC, 1999; A Report Prepared for Asia Pacific Foundation, 2002; DRPAD, **1999;** Beal, 2000; Turpin, 2000; APCTT, 2000, 2001)**.** Many of these SMEs do not have technical backgrounds and are not convinced that the technology standards, such as encryption, will protect them. Due to such perceived security fears, SMEs are not willing to use electronic payment systems. Credit cards and e-checks are naturally, then, a distant dream for many of them. Security, legal, and liability issues were frequently identified as important concerns of participating SMEs in Asia Pacific. Asia Pacific SMEs do not trust e-commerce or the technical infrastructure to support it. Trust and confidence in a sound legal framework and security standards are necessary to make e-commerce happen on a larger scale (APEC, 1999; Hall, 2002).

Confidence in Legal and Regulatory Framework

The lack of a comprehensive and acceptable legal and regulatory framework is an issue for Asia Pacific SMEs. Most of the responses indicated that SMEs want government interventions for developing an appropriate legal framework for protecting them against any frauds or disputes. Many of these Asia Pacific countries still do not have laws for electronic contracts, invoices, and other types of documentation in place. E-commerce demands several legal and liability issues to be addressed before it is widely accepted by SMEs and others in the Asia Pacific. Conducting business through electronic networks raises numerous legal questions that include the legal status and enforceability of electronic contracts, the legal jurisdiction of international e-commerce transactions, intellectual property rights and copyright protection for digital content, privacy of personal data, and validity of electronic "evidence" in legal disputes (APEC, 1999; Wade, 1990). All of these concerns about legal and liability issues are important to participating SMEs. Most of the Asia Pacific countries still do not have legal and regulatory infrastructure in place and have not addressed these issues. Unless these issues are addressed, Asia Pacific SMEs may not choose e-commerce as a medium for their businesses.

Taxation

Asia Pacific SMEs are concerned about taxation issues. Taxation processes in these countries are not transparent and are often subject to the discretion of evaluators. There are many malpractices in avoiding taxes, and SMEs do not want to have other countries' tax laws, which they anticipate could be more stringent and tough. SMEs also lack the guidance of lawyers, because many of them cannot afford the high fees necessary for sound legal advice on these matters. The application of existing taxation on commerce conducted over the Internet should be consistent with the established principles of international taxation, should be neutral with respect to other forms of commerce, should avoid inconsistent national tax jurisdictions and double taxation, and should be simple to administer and easy to understand (Hall, 1995, 2000).

Lack of Knowledge of E-Commerce

Asia Pacific SMEs lack extensive knowledge of e-commerce technologies, and that is one of the big obstacles for their participation and engagement in e-commerce. Due to lack of knowledge of e-commerce technologies, there is an internal resistance to change, and skepticism of the benefits of e-commerce among SMEs. E-commerce demands fundamental shifts in business strategies, operations, and technologies. Many participating SMEs indicated that they have limited access to information about the business models and technologies that form the basis of e-commerce success. Lack of knowledgeable staff in SMEs is also responsible for nonadoption of e-commerce. An issue often cited by participating Asia Pacific SMEs was the general lack of success stories available to demonstrate that e-commerce can be successfully implemented by firms that are similar in some way to their own. Asia Pacific SMEs do not have many success stories that could motivate other SMEs to adopt e-commerce.

Information Infrastructure Access, Quality, and Cost

Despite the cost reduction of Internet-based technology, implementing e-commerce solutions still represents a considerable and costly challenge for most SMEs in the Asia Pacific region. Large corporations with more funding, more attainable skills, and strengths in building solid business strategies could afford e-commerce deployment. But most SMEs, typically with less cash, and a shortage of information technology (IT) staff and necessary infrastructure, are not able to afford e-commerce. Insufficient access to an appropriate information infrastructure of suitable quality, and at reasonable cost, is a fundamental barrier to SME adoption and use of e-commerce in Asia Pacific (Hall, 1995). The information infrastructure required for e-commerce involves reliable telecommunications links and Internet services being available to SMEs. The level of information infrastructure to support e-commerce applications differs among the Asia Pacific countries from very poor to moderate. E-commerce applications greatly cut enterprise procurement costs and bolster the efficiency of business operations. However, low-level computerization and inferior e-commerce infrastructure reduced Internet contributions to the majority of Asia Pacific SMEs. Our research shows that although 70% of SMEs have connected with the Internet, most simply opened a home page and an e-mail address.

Payment and Delivery Systems

Although Asia Pacific's Internet players are rushing to promote e-commerce as a creative business mode for the new economy, few can break through the last-mile barriers against online purchasing. These barriers have been attributed to lack of a standard credit system and an effective express network, two critical elements for the operation of online shopping and business transaction. Credit card and off-line payment, or a combination of the two, are by far the most used means of payment. Only few SMEs (those that are technology developers) support electronic cash, electronic checks, and micropayments. Asia Pacific SMEs do not have appropriate supporting payment and distribution systems in place for the adoption of e-commerce. The majority of customers and suppliers do not use credit cards for payment in Asia Pacific countries. Therefore, the development of electronic payment systems, including the use of credit cards, is limited. A relatively low credit card penetration among consumers acts as a domestic constraint to B2C e-commerce development. In many cases, it the "mindset" that inhibits the adoption of online credit card payment. The credit card culture is not there in the society, and hence, credit card companies and consumers shy away from using credit cards for transactions. The financial network of banks and other institutions do not have their databases integrated and available online (APEC, 1999; A Report Prepared for Asia Pacific Foundation, 2002; DRPAD, **1999;** A Study Report on Thailand, 2001; Beal, 2000; Turpin, 2000; APCTT, 2000, 2001)**.**

CONCLUSION

This synopsis serves to highlight that SMEs face many obstacles with respect to e-adoption. Furthermore, we note that governments have an important role to play in encouraging SMEs to embrace e-commerce in Asia Pacific countries. SMEs in Asia Pacific countries need to be made more competitive by promoting the development and use of e-commerce. In particular, governments must address four key areas so that high e-business potential will ensue for these SMEs; namely, they must address the areas of business infrastructure, regulatory or commerce infrastructure, user infrastructure, and telecommunications infrastructure. It is only through the systematic efforts of government and policy makers to encourage e-commerce adoption by SMEs and at the same time ensure that the appropriate safeguards and controls are in place, that SMEs in the Asia Pacific region will begin to embrace e-commerce. E-commerce is critical to the long-term success and sustainability of SMEs in this region. It is imperative that the full e-business potential that e-commerce affords can be utilized and maximized.

REFERENCES

A Report by Development Research and Policy Analysis Division (DRPAD). (1999). *Promotion of trade and investment through electronic commerce in Asia Pacific.*

A Report Prepared for Asia Pacific Foundation. (2002). *E-commerce and Asia Pacific*. Castle Asia Pacific.

A Study Report on Thailand. (2001). *SMEs and e-commerce*. Asia Pacific Foundation.

Afuah. A., & Tucci, C. (2000). *Internet business models*. Boston, MA: McGraw-Hill Irwin.

Alphonso, O. M., & Myrna, R. C. (Eds.). (2001). *Bridging the gap—Philippine SMEs and globalization*. Manila: Small Enterprises Research and Development Foundation.

APEC. (1999). *SME electronic commerce study*. Final report.

APEC. (2001). *The new economy and APEC*. Report to the Economic Committee of APEC.

Arunachalam, V. (1995). EDI: An analysis of adoption, uses, benefits and barriers. *Journal of Systems Management*, *46*(2), 60–65.

Beal, T. (2000). SMEs and the World Wide Web: Opportunities and prospects. In A. Moha Asri (Ed.), *Small and medium enterprises in Asia Pacific Pacific, Vol. III: Development prospects* (pp. 102–134). Commack, NY: Nova Science Publishers.

Beal, T. (2001). Patterns of Internet development in the New Economy: Implications for Asia Pacific SMEs. In A. Moha Asri (Ed.), *Small and medium enterprises in the information age: Asia Pacific perspectives*. Leeds, UK: Wisdom House.

Debroy, B. (2002). Information and communications technology and development: An Indian perspective. *Bulletin on Asia-Pacific Perspectives 2001/02* (Bangkok, ESCAP).

eMarketer Inc. *http://live.emarketer.com/*

Hall, C. (1995). Investing in intangibles: Improving the effectiveness of government SME advisory services in the APEC region. In *Report of APEC Symposium on HRD for SMEs*, APEC, China Productivity Centre, Chinese Taipei.

Hall, C. (2002, August). *Profile of SMEs and SME issues in APEC 1999–2000*. Final draft report.

PricewaterhouseCoopers. (1999). *SME electronic commerce study*. Report to the APEC Telecommunications Working Group and Business Facilitation Steering Group.

Regnier, P. (2000). *Small and medium enterprises in distress—Thailand, the East Asian crisis and beyond*. Aldershot: Gower Publishing.

Sengenberger, W., Loveman, G., & Piore, M. (1990). *The re-emergence of small enterprises: Industrial restructuring in industrial countries*. Geneva: International Institute of Labour Studies.

Sharma, S. K., Wickramansinghe, N., & Gupta, J. N. D. (2003). Working paper. *What should SMEs do to succeed in todays knowledge-based economy?*

Tambunan, T. T. H. (2000). *Development of small-scale industries during the new order government in Indonesia*. Aldershot: Gower Publishing.

The Proceedings of the Regional Seminar on Electronic Commerce was organized by the APT, Asia Pacific and Pacific Center for Transfer of Technology (APCTT) and Department of Telecommunication (DoT), Government of India, December 11–14, 2001.

The Proceedings of the Seminar on Electronic Commerce Strategies for the Asia Pacific/Pacific was organized by the APT and Asia Pacific and Pacific Center for Technology Transfer of Technology (APCTT) and was sponsored by the Department of Telecommunications Services (DTS) and the Department of Telecommunication (DoT), Government of India, August 8–10, 2000.

Turpin, T., Spence, H., Garrett, J. S., & Marsh, A. (1998). South-East Asia Pacific and the Pacific Rim. In H.J. Moore (Ed.), *World Science Report* (pp. 212–236). Amsterdam: UNESCO Publishing, Elsevier.

Turpin, T (2001) *The role of SMEs in the diffusion of technology among East Asian economies* in Charles Harvie and Boon Chye-Lee (eds.) Small and Medium Sized Enterprises in East Asia in the aftermath of the Financial

Crisis Vol. 1 - Overview, issues, policy and prospects, Edward Elgar, Aldershot (in press).

Turpin, T. (2000). *SMEs and Technology Diffusion in Asia-Pacific Economies: Indonesia as a Case-Study* in Advances in International Business, 2000, published proceedings of the 2000 Annual Conference of Academy of International Business Southeast Asia Region, Editor, Oliver H. M. Yau, City University of Hong Kong, Hong Kong, SAR, People´s Republic of China, 23-26.

Turpin, T. & Xielin, L. (2000). Balanced development: The challenge for science, technology and innovation policy. In C. Harvie (Ed.), *Contemporary Development and issues in China's economic transition*. New York: Macmillan.

UNCTAD. (2002). *Trade and development report 2002*. Geneva; New York: United Nations.

Urata, S. (2000). Policy recommendations for SME promotion in Indonesia. Report to the Coordination Ministry of Economy, Finance and Industry.

Wade, R. (1990). *Governing the market: Economic theory and the role of government in East Asia Pacificn industrializations*. Princeton, NJ: Princeton University Press.

Wattanapruttipaisan, T. (2002). Promoting SMEs development: Some issues and suggestions for policy consideration. *Bulletin on Asia-Pacific Perspectives 2002/03* (pp. 57–67).

KEY TERMS

Dot-Com: A company with operations that are entirely or primarily Internet based or, more specifically, a company with a business model that would not be possible if the Internet did not exist. Dot-coms often deliver all their services over an Internet interface, but products might be delivered through traditional channels as well. Dot-coms are often divided into two categories: those that provide products and services for consumers (B2C) and those that provide products and services to other businesses (B2B).

E-Commerce: The conduct of buying and selling products and services by businesses and consumers over the Internet. "E" simply means anything done electronically, usually via the Internet. E-commerce is the means of selling goods on the Internet, using Web pages.

SMEs: As per European Commission definition, small and medium-sized enterprises, referred as "SMEs," are defined as enterprises that have fewer than 250 employees, and have either an annual turnover not exceeding 40 million EURO, or an annual balance sheet total not exceeding 27 million EURO, as shown below.

	Medium-sized	**Small**	**Micro-enterprise**
Maximum number of employees	250	50	10
Maximum turnover (in million ECU)	40	7	—
Maximum balance sheet total (in million ECU)	27	5	—

ENDNOTES

1 http://www.india-today.com/ctoday/20000901/telecom.html

2 http://www.atimes.com/atimes/China/EA18Ad06.html

Offshore Software Development Outsourcing

Stephen Hawk
University of Wisconsin - Parkside, USA

Kate Kaiser
Marquette University, USA

OFFSHORE SOFTWARE DEVELOPMENT OUTSOURCING EVOLUTION

Until the global economic downturn of the new millennium, demand for information technology (IT) professionals exceeded supply mostly due to specific skill sets such as integrating legacy applications with Web development, project management, telecommunications, mobile commerce, and enterprise resource planning. More firms are turning externally not only to local vendors but also to services across the globe (Carmel, 1999). Staff supplementation from domestic contractors has evolved to a sophisticated model of partnering with offshore/ nearshore software development firms. Many of these relationships evolved from a short-term project need for select skills to a long-term commitment of resources, cultural diversity efforts, and dependencies that integrate vendors as partners.

The most pervasive IT project, Year 2000 (Y2K), had constraints of skill sets, time, and budget. IT managers had to look at many alternatives for achieving compliance. Firms that planned as early as the mid-1990s had time to experiment and build new relationships. With governmental sanction and support, some countries and their business leaders recognized the competitive advantage their labor force could offer (O'Riain, 1997; Heeks, 1999; Trauth, 2000). An unusual need for services because of Y2K, economic disparity within a global workforce, and proactive efforts of some governments led to the fostering of offshore software development.

Early companies to outsource offshore were software vendors. Managing offshore development smoothly took years and a certain type of project management expertise in addition to a financial commitment from executives. The activity involved new applications and integrating software development with existing domestically built applications. The Y2K investment and intense cultural communication paid off for firms willing to work through the challenges. Not only did initial offshore projects provide a solution to the skill shortage, they also yielded substantial cost savings when compared to outsourcing the work domestically. Such factors resulted in these relationships continuing past Y2K, where some companies now regard their offshore arm as partners.

The IT outsourcing market was estimated at over US$100 billion by 2001 (Lacity and Willcocks, 2001). Although outsourced IT services can include call centers and facilities management, this discussion focuses on outsourcing software development. "Offshore" software development typically refers to engaging workers from another continent. Examples are U.S. companies using Indian contractors. "Nearshore" software development refers to vendor firms located in nearby countries often on the same continent, for example Dutch firms engaging Irish software developers. For our purposes, discussion of offshore software development issues is assumed to apply to nearshore outsourcing, since most of the issues are the same.

Most firms already have had experience supplementing their staff. Dealing with offshore firms, however, is relatively new for firms whose main business is not software development. Distance, time zones, language and cultural differences are some key issues that differentiate offshore software development from the use of domestic contractors or consultants (Carmel, 1999). Nearly 1 million IT-jobs will move offshore from the United States by 2017 (Gaudin, 2002).

The most common reasons for outsourcing are cost reduction, shortages of IT staff, reduced development time, quality of work, and internationalization (see Table 1).

APPROACHES TO OFFSHORE SOFTWARE DEVELOPMENT

The approaches taken to offshore development describe the basic features of how firms have structured their relationships with offshore software developers. The basic dimensions include the type of organization that provides offshore services, the nature of the working relationship that clients have with that organization, and whether offshore staff are on a defined project or in a staff augmentation role.

Table 1. Reasons for engaging in offshore software development

- Cost reduction - due primarily to lower wages and secondarily from tax incentives. Accounting for indirect costs, such as travel and teleconferencing, overall average savings may be close to 25% (Overby, 2001).
- Access to a global pool of talent - quicker staffing thereby enabling faster time to delivery.
- 24×7 Productivity – the global delivery model allows a "follow-the sun" approach for round-the-clock handing-off of work to developers in different time zones.
- Quality – rigorous requirements of certifications (CMM, ISO) or cultural dedication to detail and quality.
- Localization - adapting software to local requirements may be best handled by resources in or near the target location.

Type of Organization Providing Offshore Development

An important dimension for describing the relationship with foreign software developers would be the basic types of organizations with which client firms form an outsourcing arrangement. The basic options are to contract directly with an offshore IT services firm, to engage a domestic IT services firm with offshore capabilities, or to set up a wholly owned software development center in an offshore location.

- Direct contract with an offshore firm – The client firm contracts directly with the offshore software firm. This option represents a range of alternatives.

  • Long-distance contracting – the offshore firm maintains no permanent staff in the country of their clients. Clients and offshore firms use long distance communications technologies such as phone, e-mail, and short-term visits.
  • Limited presence – Offshore firms locate some staff in client countries by setting up satellite offices in one or more cities. Most of the work is performed offshore, but a local project manager and a small local team answers questions and function as liaisons to solve problems.
  • Integrated on-site – the export firm provides a mix of extensive onsite expertise to handle business and requirements analysis, system design, and project management, and coordinates with offshore staff, normally working closely with client management. As many as 70 firms from Ireland have set up subsidiaries in the U.S. with Irish government financial support (Cochran, 2001).

- Contract with a domestic IT services firm with offshore capabilities – Many domestically-based IT services firms have the ability to send work offshore. This may take the form of an IT firm that acts as an intermediary, or a firm that itself owns an offshore satellite location. Some domestic firms employ both approaches. An intermediary acts as a liaison with offshore software firms, and already has subcontracting or partnering arrangements with them. The IT services firm works with clients in much the same way as other domestic IT services firms, but differs in that it uses offshore resources for some of the software development. Intermediaries may negotiate a cut of the cost savings their clients achieve by going offshore. The advantage in giving up the full savings of dealing directly with outsourcers is that the intermediaries buffer clients from most aspects of dealing with offshore providers. For small to midsize companies new to outsourcing, brokers offer the big benefit of offshore projects—good work done more cheaply, but without cross-cultural, long–distance, vendor management headaches.

- Wholly owned Satellite – The firm operates in an offshore location to directly hire IT professionals to become its employees. This is model is almost exclusively used by companies that sell software or by companies whose products have a large software component.

Length of Relationship and Type of Project

The working relationship between clients and vendors can be described by whether the relationship is short or long term, and the type of work the vendor performs.

Relationships with an offshore IT services firms can range from a few months to many years. Short-term relationships may be a single-hit project with no plan for follow-up projects. Long-term relationships may become important business partnerships where an offshore firm is responsible for an ongoing stream of projects.

The nature of work given to offshore vendors can range from well-defined projects to relatively unstructured ones. Typical examples of well-defined projects are technical design, implementation, maintenance, or software conversion. In other cases, vendors may take on significant responsibility for project initiation, requirements determination, and project management involving new systems or major enhancements to existing systems that have strategic impact to the firm.

Carmel and Agarwal (2002) propose a four-stage model of offshore outsourcing that can be described in terms of time frame and nature of work. They describe firms moving from short-term structured engagements to longer-term structured engagements, and finally to long-term unstructured engagements as a maturation process. A company, however, may have outsourcing efforts that fall into different categories. For example an organization may engage offshore firms for both an ongoing stream of structured and unstructured projects, while at the same time evaluating other vendors, or outsourcing special needs projects on a one-time basis.

Defined Project vs. Staff Augmentation

The previous discussion of offshore development focused on contracting with the outsourcer for the delivery of a defined product. It's possible to tap some outsourcing firms as sources for augmenting staff. Companies contract outsourcing services on the basis of developer time instead of a fixed price software deliverable. Offshore staff may travel to the client's site and be managed closely by the client or may remain in the offshore location managed remotely. Offshore firms in this case have limited responsibility for managing projects.

SOFTWARE EXPORTING COUNTRIES

Table 2 lists dominant players by order of reported revenues, industry software organizations, and their Web sites. Many reports of exports are not standardized. It is not clear what is included in sales figures and who is reporting them. Even if these figures were comparable, the projection for evaluating different countries is not. Some countries are producing more software with fewer

Table 2. Information Technology Industry Organizations by Country (Enterprise Ireland promotes Irish exports, including software services.)

India	National Association of Software and Services Companies -NASSCom	www.nasscom.org
Philippines	Information Technology Association of the Philippines - ITAP	www.itaphil.org
Ireland	Enterprise Ireland*	www.enterprise-ireland.com
Israel	Israeli Association of Software Houses – I.A.S.H.	www.iash.org.il
Russia	RUSSOFT - The National Software Development Association of Russia	www.russoft.org
China	Software Offshore Business Union of Shanghai - SOBUS Ministry of Information Industry - MII	www.sobus.com.cn www.mii.gov.cn/mii/index.html
Canada	Information Technology Association of Canada	www.itac.ca
Mexico	AMITI - Mexican Association for the Information Technologies Industry	www.amiti.org.mx/

developers than others. The types of applications delivered may not compare. Converting a legacy customer database is not similar to customizing an enterprise system. More importantly, within a country there may be many development vendors and generalizations that are not applicable to each company.

CRITICAL ISSUES OF OFFSHORE SOFTWARE DEVELOPMENT

Challenges associated with offshore development are different than domestic development. In most cases the issues lead to creative solutions that some companies have tackled and incorporated to their advantage. Special circumstances of offshore development present factors that practitioners have found to be of special importance in achieving success.

Quality Certification

The Software Engineering Institute's Capability Maturity Model (CMM) (CM-SEI, 2002) or the International Organization for Standardization's ISO9001: 2001 standard (ISO, 2002) serves as an independent evaluation of the quality of a vendor's development processes. Certification should not be taken at face value. Investigation can clarify its meaning. When it was acquired? Was it qualified by project or by subsidiary? Who provided the evaluation? Was it self-assessed? There are a number of questions to raise.

Onsite Presence

Despite the ability to operate virtually, the tenuous nature of offshore development is optimized when face-to-face contact, handshaking, and social/professional relationships nurture the engagement. The limited presence and integrated on-site approaches address the distance resistance.

Managing Globally

The downside of being productive 24×7 is that some part of the team is often not available for real-time contact because they are not at work at the same time. Travel by someone is a necessity in almost all offshore arrangements. The nature of the relationship will dictate how many and how often, from where to where. Sometimes this may involve extended stays.

Trust and Confidence

Relationship management is the overriding success factor when minimal levels of other factors are in place. IT services firms accent their marketing on these traits. For offshore firms to be seen as important business partners and take on work beyond routine development work, clients need to have a high level of trust and confidence in the outsourcers' abilities and intentions.

Political Environment

The stability of the offshore country's government prevents some firms from exploring offshore outsourcing (Nicholson and Sahay, 2001). Israel's thriving development curtailed with the Palestinian war. The Pakistan-India border disputes caused the U. S. State Department to ban travel to India. Although firms contract with a software firm and not the country, political environment of the country can influence the feasibility of working with firms located there.

Intellectual property concerns are usually dissuaded due to protection by law. Most countries have laws that are comprehensive enough to deal with software as ownership. However, having laws in place does not mean that governments enforce them.

Cultural Awareness and Integration

Close working relationships between offshore vendors and clients from another culture may require significant time, not only for becoming aware of each others' cultures, but to integrate it in practice. Firms with close working relationships not only achieve integration in software methods and delivery but also achieve integration in social aspects of team building.

Nationalism

Organizations that hire offshore developers may have to justify why they are not hiring domestically when many people are unemployed. Issuing of U.S. H1-B visas raised concern as many Americans became unemployed. The IT community was sensitive to the impact of hiring foreigners with special skills. Although visas are no longer granted as they once were, many organizations have had to deal with the same issue that arises with offshore software development. Human rights and workers' ability to earn a living wage are not comparable to client countries' workers—but no published reports of abuse or exploitation surfaced. Offshore software development offers governments ways to gain power and enhance their standing in the world.

Capacity of the Workforce

The availability of human resources to deliver IT demand is debatable. There is mixed evidence about the ability to meet demand (Overby, 2001; Carmel and Agarwal, 2002). Determining capacity is tricky. One must not only review population numbers but the educational, telecommunications, and managerial infrastructure that allows growth.

Risk Management

The most obvious threat relates to the political environment. Because some firms perceive security breaches as more vulnerable using offshore developers, outsourcers probably take extra precaution to sell their services. Y2K and September 11th reinforced the need for business continuity planning.

Risk management can also affect selection of firms in part on the basis of location. Developing relationships with a geographically diverse portfolio of outsourcers could minimize the risk of regional political and economic disturbances.

FUTURE TRENDS

Increased Offshore Outsourcing of Higher-Level System Life Cycle Tasks

Greater experience in using offshore developers on the part of client firms coupled with the maturation of technologies and methodologies for global software delivery allow firms to outsource more high-level development and realize the benefits of outsourcing across a wider variety of IT work.

Emergence of Major Contenders to India

India has been by far the major destination for offshore software development outsourcing, with Israel and Ireland usually considered "major players." China, Philippines, and Russia are often noted as potential contenders. Outsourcing industries are emerging in a number of unexpected locations such as Viet Nam, Mexico, Ukraine, and Iran due to low entry costs.

Increasing Use of Offshore Workers by Domestic Software Firms

To remain competitive, domestic IT outsourcing firms increasingly use offshore developers either through wholly-owned development centers offshore, or by subcontracting or partnering with foreign outsourcing firms (Overby, 2001).

Increased Sensitivity to Outsourcing Domestic Jobs

Reduced employment in importing countries creates a potential for a backlash. Ensuing restrictive changes in visas for foreign workers could limit higher-level work outsourced offshore since it requires greater onsite presence.

Increased Use of Third-Party Offshore Software Development Suppliers

Rather than bear project management costs directly, more importers will engage third-part consultants to act as a buffer and minimize relationship costs

CONCLUSIONS

Offshore software development outsourcing potentially provides considerable benefits. Some businesses are in the early stages of offshore experimentation, while others have formed mature business partnerships with offshore software firms. Many businesses will outsource more IT work offshore in order to cut costs and reduce time to market in addition to the benefits of quality. An increasing share of the IT work being outsourced is found at higher levels. These forces, however, may be compensated by factors such as political agenda of domestic employment, a desire to avoid negative public relations, and visa limitations that increase the challenges for offshore firms to provide on-site staff.

REFERENCES

Carmel, E. (1999). *Global Software Teams: Collaborating Across Borders and Time Zones.* New York: Prentice-Hall.

Carmel, E., and Agarwal, R. (2002). The Maturation of Offshore Sourcing of Information Technology Work. *MIS Quarterly Executive.* 1(2), 65-77.

CM-SEI. (2002). Carnegie Mellon Software Engineering Institute's Capability Maturity Model® for Software (SW-CMM®). Retrieved December 1, 2003 from the World Wide Web at: *http://www.sei.cmu.edu/cmm/cmm.html*

Cochran, R. (2001). Ireland: A Software Success Story. *IEEE Software*. 18(2), 87-89.

Gaudian, S. (2002, November 19). Nearly 1 Million IT Jobs Moving Offshore. *Datamation*. Retrieved December 1, 2003 from the World Wide Web at: *http://itmanagement.earthweb.com/career/article.php/1503461*

Heeks, R.B. (1999). Software Strategies in Developing Countries. *Communications of the ACM*. 42(6), 15-20.

ISO. (2002). International Organization for Standardization's "Selection and Use of the ISO 9000:2000 family of standards" includes ISO-9001. Retrieved December 1, 2003 from the World Wide Web at: *http://www.iso.org/iso/en/iso9000-14000/iso9000/selection_use/selection_use.html*

Lacity, M. C., and Willcocks, L.P. (2001). *Global Information Technology Outsourcing: In Search of Business Advantage*. New York: Wiley.

Nicholson, B., and Sahay, S. (2001). Some Political and Cultural Issues in the Globalisation of Software Development: Case Experience from Britain and India. *Information and Organization,* 11(1).

O'Riain, S. (1997). The Birth of a Celtic Tiger. *Communications of the ACM*. 40(3), 11-16.

Overby, C. S. (2001, September). The Coming Offshore Service Crunch. *Forrester Report*, 1-21.

Trauth, E.M. (2000). The *Culture of an Information Economy: Influences and Impacts in The Republic of Ireland*. Netherlands: Kluwer, Dordrecht.

KEY TERMS

Capability Maturity Model (CMM): A methodology used to evaluate an organization's software development process. The model describes a five-level evolutionary path of increasingly organized and systematically more mature processes.

Global Delivery Model: To distribute and manage software development across multiple global locations. GDM requires infrastructure, project management, cultural sensitivity, and process guidelines to support communication and coordination between locations.

Intellectual property rights: Laws and enforcement mechanisms to afford the creator of an intellectual property (e.g., software) the means of controlling how their work is used, and ensuring that the creator is properly rewarded and recognized for their work**.**

ISO 9001:2001: Provides standards used to assess an organization's ability to meet its customer's requirements and achieve customer satisfaction. Software firms use it as an alternative to CMM.

Localization: The process of adapting software to a particular language, culture, and desired local "look-and-feel". This may include local sensitivities, geographic examples, and adhering to local legal and business requirements.

Staff Augmentation: When an employee of an outsourcing firm is lent to a client company for work. Normally a staff augmentation employee works at the client's worksite with the technical direction coming from the client company.

On the Relativity of Ontological Domains and Their Specifications

E. Kapetanios
Semantic Technologies, Switzerland

INTRODUCTION

Business knowledge is considered the most valuable resource of organizations. The exchange of business knowledge within an organization as well as among different cooperating organizations is considered a major promotion of know-how and better coordination of various organization units, as well as an increase of the value of an organization itself.

In the era of "information overload" and an ever increasing "networked world," however, where the premium of business and IT is placed on communications and distributed computing and networks, semantics play a crucial role. To this extent, the usage of *semantic technology* should encourage not only *information globalism*, but also *information particularism*, in which:

- individuals or groups define, generate, and access information in ways that make sense to them,
- information is accessible across the business by integrating different and, very often, controversial points of view, and
- semantic heterogeneity is not treated as a "necessary evil," but also encouraged by the technology and solutions.

Meeting these challenges is considered a key issue in order to enable real knowledge exchange and cope with inherited diversities such as cultural differences, vaguely formulated or uncertain specifications, relativity of concept validity, and so forth. This goal becomes more crucial when, for example, advanced scientific (biology, physics, chemistry, etc.) or engineering application domains in large business settings and environments are concerned.

BACKGROUND

Knowledge models, or *ontologies*, are a necessary precondition to any semantic application (Bresciani & Fontana, 2002; Kapetanios, 2002; Kapetanios, Baer, & Groenewoud, 2003; Kapetanios, Baer, Groenewoud, Mueller, Novosad, 2004; Sugumaran & Storey, 2002). Ontologies have emerged over the last 10 years as a core technology and fundamental data model for knowledge systems.

They enable various advanced functions, for example, smart search (Andreasen, Motro, Christiansen, & Larsen, 2002), and are the foundation of the (emerging) semantic Web (Berners-Lee, Hendler, & Lassila, 2001; Davis, Fensel, D., Hendler A. J, Lieberman H., Wahlster W., & Harmelen, 2002; Fensel, Hendler, Lieberman, & Wahlster, 2002). Ontology-enabled semantic technologies show great promise for the next generation of more capable information technology solutions because they can solve some problems much more simply than before and make it possible to provide certain capabilities that have otherwise been very difficult to support.

The current state of the art, however, on representing and using ontologies has grown out of several efforts that started in the 1980s. Back then, KL-ONE was the most influential of the frame-based representation languages; it allowed for the representation of categories and instances, with inheritance of category properties, and a formal logic (Baader, Calvanese, McGuinness, Nardi, & Patel-Schneider, 2003) for expressing the meaning of properties and categories. At about the same time, rule-based systems were a promising technology. The NASA-sponsored C-Language Integrated Production System (CLIPS) became a de facto standard for building and deploying rule-based systems.

The Knowledge Interchange Format (KIF) and its accompanying translation tool Ontolingua were developed to allow knowledge to be shared among these different efforts, and provided the capability to translate knowledge bases in one representation language to another. These languages were ahead of their time. As a result, they have remained largely within academia, gaining little commercial support.

With the advent of the World Wide Web and the acceptance of XML (extended markup language) as a de facto standard for the representation of information on the Web, ontology efforts joined in. An early project at the University of Maryland produced SHOE (simple HTML [hypertext markup language] ontology extension), a system for expressing ontologies in XML and marking up Web pages with ontology-based annotations. Many of the ideas from this work made it into the World Wide Web Consortium (W3C; http://www.w3c.org) proposal for the Resource Description Framework (RDF) language (http:/

/www.w3c.org/RDF). Moreover, languages such as DAML (Defense Advanced Research Projects Agency - DARPA - or agent markup language), DAML+OIL (Ontology Inference Layer), and OWL (Web ontology language; http://www.w3c.org/2001/sw/WebOnt), which became a foundation for the W3C Web ontology language, are built on RDF.

However, the major assumption underlying all currently available ontology description formalisms and languages has been the "common view" of a conceptualization of an application domain (Berners-Lee et al., 2001; Davis et al., 2002; Fensel, Hendler, Lieberman, & Wahlster, 2002). This is also reflected in currently available ontology development tools, which are a key factor in adoption of semantic technology. Many tools are still offered as research prototypes, but many others have begun to be commercialized (they are often commercial versions of their direct research counterparts). Standard compliance and support for RDF and OWL is growing. Due to the different forms of RDF, ontology tools still do not interoperate well.

It turns out that the assumption of a common view is an unrealistic one given the ever increasing networked world, which results in a wide range of user perspectives and points of view of the same concepts within, for example, different organizational units or cultural and social environments. Information and knowledge exchange should pay tribute to these diversities in order to increase information quality and interpretability, as well as to strengthen collaboration among various user communities or organizational units and processes.

PRESENTING ONTOCONTEXT

The most outstanding difference between OntoContext and other ontology description languages and editors is twofold.

- The support of concept relativity in terms of diverse perspectives, viewpoints, natural languages, and so forth, which are reflected in both the ontology description language and the tool itself.
- Functioning as a collaborative platform and a database-driven approach reflecting concept relativity such that large-scale ontologies are enabled and shared by a wide range of end users.

Conceptulization According to Relativism

OntoContext addresses relativity and perspectiveness within an ontology in a twofold manner:

1) an ontology *external context*, such as bounding the representation of concepts to particular user environments such as natural language, roles of organization units, and so forth, and
2) an ontology *internal context*, enabling the expression of holding conditions or constraints under which concepts, business rules, and their definition, even in an agreed-upon ontology, become relevant.

Examples of an ontology internal context are given by

- the consideration of the validity of properties, value domains, or classification hierarchies under specific conditions,
- the definition of the membership of instances to classes according to the distinction between *necessary* and *sufficient* conditions,
- the definition of percentages for the membership of instances to classes,
- the definition of classes according to the distinction between prototypes and well-established properties, and
- the naming of concepts by using more than once the same name within the same ontology description name space (no unique name assumption).

As far as the database-driven availability and management of large-scale ontologies is concerned, this strengthens their practicability and usage in multiuser environments, and enables the implementation and coordination of authoring or editing facilities among various user communities, in addition to querying and sharing of business knowledge.

As far as the human-computer interaction facilities are concerned, they are addressed by a graphical user interface, which

- is free of knowledge-based formalisms and concepts,
- enables interaction with the business knowledge in more than one natural language,
- enables adaptation to particular user environments with respect to ontology external context, and
- enables the presentation of the ontology internal context issues, as stated above.

Tool Description and Architecture

The concept-relativity issues as described above are reflected in the OntoContext tool, which is an ontology development tool. However, the outstanding difference between OntoContext as an ontology development tool and other ones is

a) the human-computer interaction, and
b) the collaboration mechanism,

which both reflect the language properties of *conceptual relativism*. Therefore, the tool enables

1) specification and description of application domain semantics by novice users rather than knowledge engineers, that is, it is free of knowledge formalisms,
2) specification and description of application domain semantics by paying tribute to different perspectives and cultural differences, for example, natural languages,
3) querying and sharing of contextualized application domain semantics for better coordination and collaboration among organization units as well as among organization processes,
4) convergence of common aspects through sharing and exchange of organization and process knowledge, and
5) exporting or importing ontologies in languages such as RDF, OWL, and so forth, as proposed for the semantic Web.

These features are supported by the architectural issues of the tool, which rely on a *database-driven realization* for the needs of persistency, maintenance, and querying, as well as on the *inference services* underlying an intelligent human-computer interaction mechanism for the needs of guiding the end user for both querying and ontology specification.

Comparisons

Having said this, it is worth referring to other ontology development tools, as far as known to the author, in order to raise the major differences. The following table gives an overview of those ontology development tools with respect to some of the issues according to enabling collaboration and relativism, as described above.

As stated in Table 1, most of the currently available ontology development tools do not support collaboration. They are meant to be single-user tools and collaboration is only enabled through the exchange of ontology description files (no database for sharing of views and querying of ontologies).

It is also stated that none of these tools, even those collaborative ones, do support relativism as far as different natural languages, points of views, and user settings are concerned.

FUTURE TRENDS

Many of the issues for conceptualization according to relativism have been tried out in large- and small-scale projects in real-world applications. In particular, smart querying and semantic data collection forms, as well as concept-based search mechanisms, have been applied for a Second Opinion (Decision Support) System for querying medical guidelines, for collecting and querying multiple collections of databases as related to a Swiss national registry for clinical trials and studies for acute

Table 1. Overview of ontology development tools with respect to OntoContext's main features

Tool/Vendor	Collaboration	Natural Language	Points of View	User Bounded
Protègè 2000/Stanford USA	No/Single User	Only One	Only One	No
OilEd/U. of Manchester	No/Single User	Only One	Only One	No
OntoEdit/Ontoprise	No/Single User	Only One	Only One	No
LinkFactory Workbench/Language and Computing	Some support for collaboration	Only One	Only One	No
K-Infinity/Intelligent Views	Yes/Multiuser	Only One	Only One	No
Cerebra Construct/Network Inference	Yes/Multiuser	Only One	Only One	No

myocardial infarctions, and for querying of collected data from demining actions.

In all these application settings, external relativity has been practiced due to user-bounded validity of concepts from the underlying ontology and the application of the relevant constraints. Internal relativity has been practiced due to current states of concept descriptions such that the validity of interconnected concepts is related to that current state of concept description. This has been extensively practiced in semantic data collection forms.

For making concept descriptions, however, relative to conditions and constraints beyond those found in other ontology description languages such as OWL, which address slot or value constraints in terms of classes, a more generic constraints schema needs to be elaborated and addressed within an ontology description language. This copes with constraints as posed to class or instance definitions, adaptable value domains, and so forth.

Furthermore, in a collaborative setting for ontology specification and querying, constraints need to be addressed as first-class citizens, such that they remain consistent in large-scale ontology development. In addition, embedding of multiple viewpoints and different natural languages needs to be efficiently handled while reasoning.

CONCLUSION

The importance of increasing the value of an organization through intra- as well as interorganizational knowledge exchange and coordination of processes has been stressed many times in the past. The contribution and usage of IT, however, in such a process remained quite limited, despite the technological advantages and wide use of information technology.

Ontology-enabled semantic technologies promise to add value to business settings and environments. In particular, ontology development tools emerged recently, which put the emphasis on extrapolation of knowledge-based formalisms and provided some common view across all units or user communities. However, business settings in an ever increasing networked world with large-scale ontology development efforts forces thinking toward the accommodation of semantic heterogeneity and cultural diversity within ontologies rather than dealing with a single common view.

Sharing different perspectives and viewpoints, regardless of the user's natural language in collaborative business environments, will enable knowledge exchange and a better understanding of processes and data.

REFERENCES

Andreasen, T., Motro, A., Christiansen, H., & Larsen, H.-L. (Eds.). (2002). Flexible query answering systems. In *Lecture notes in artificial intelligence* (Vol. 2522). Copenhagen, Denmark: Springer Verlag.

Ankolekar, A., Burstein, M., Hobbs, J., Lassila, O., Martin, D.L., McDermott, D., McIlraith, S.A., Narayanan, S., Paolucci, M., Payne, T.R., & Sycara, K. (2002). *DAML-S: Web service description for the semantic Web.* First International Semantic Web Conference (ISWC 02), Sardinia, Italy.

Baader, F., Calvanese, D., McGuinness, D., Nardi, D., & Patel-Schneider, P. (2003). *The description logic handbook.* UK: Cambridge University Press.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic Web. Scientific American, *284*(5), 34-43.

Bresciani, P., & Fontana, P. (2002). Knowledge-based query system for biological databases. In *Lecture notes in artificial intelligence: Vol. 2522, International conference on flexible query answering systems (FQAS).* Copenhagen, Denmark: Springer Verlag.

Davis, J., Fensel, D., & Harmelen, F. (Eds.). (2002). *Towards the semantic Web: Ontology-driven knowledge management.* Wiley.

Fensel, D., Hendler A. J, Lieberman H., Wahlster W. (Eds.). (2002). *Spinning the semantic Web: Bringing the World Wide Web to its full potential.* MA: MIT Press.

Kapetanios, E. (2002). A semantically advanced querying methodology for medical knowledge and decision support. In R Gruetter (Ed.), *Knowledge media in healthcare: Opportunities and challenges.* Hershey, PA: Idea Group Publishing.

Kapetanios, E., Baer, D., & Groenewoud, P. (2003). *Simplifying syntactic and semantic parsing of NL based queries in advanced application domains.* International Conference on Applications of Natural Languages to Information Systems (NLDB), Burg, Germany.

Kapetanios, E., Baer, D., Groenewoud, P., & Mueller, P. (2002). *The design and implementation of a meaning driven data query language.* The 14th International Conference on Scientific and Statistical Databases (SSDBM), IOS Press, Edinburgh, Scotland.

Kapetanios, E., Baer, D., Groenewoud, P., Mueller, P., & Novosad, L. (2004). Querying the unknown: A semantics driven query language for advanced application domains. *Journal of Computational Methods in Sciences and*

Engineering [Special issue on intelligent systems]. Accepted for publication, to appear. PA: ISO Press.

Kapetanios, E., & Groenewoud, P. (2002). *Query construction through meaningful suggestions of terms.* Fifth International Conference on Flexible Query Answering Systems (FQAS), Copenhagen, Denmark.

Sugumaran, V., & Storey, C. V. (2002). *An ontology based framework for generating and improving DB design.* Seventh International Conference on Applications of Natural Language to Information Systems, Stockholm, Sweden.

Zhdanova, A. V. (2002). *Automatic identification of European languages.* Seventh International Conference on Applications of Natural Language to Information Systems, Stockholm, Sweden.

KEY TERMS

Concept-Based Search: Search over a corpus of documents or databases can proceed using a search that matches concepts rather than matching words. The value of concept-based search increases with the technical complexity of the domain of search.

Intelligent (Smart) Querying: Querying that is driven by some kind of inference engine or mechanism for refining, formulating, or completing query construction.

KL-ONE: A knowledge representation system in the late 1970s that inspired the development of a number of frame-based representation systems. These systems embraced the ideas of frames as structured descriptions, as well as the idea of differentiating between terminological and assertional aspects of knowledge representation and the central role of subsumption and classification inferences.

Knowledge Interchange Format (KIF): KIF is a language designed for use in the interchange of knowledge among disparate computer systems (created by different programmers, at different times, in different languages, and so forth). KIF is *not* intended as a primary language for interaction with human users (though it can be used for this purpose). The purpose of KIF is roughly analogous to that of Postscript. Postscript is commonly used by text and graphics formatting systems in communicating information about documents to printers.

Ontology: It is the name given to a structure of knowledge used as a means of knowledge sharing within a community. Aristotle's notion of Organon might be the earliest example of an ontology in this sense. An ontology, however, as software technology is a "data skeleton" that can be developed for a variety of data-based and knowledge-based systems. It consists of a collection of concepts and relationships between concepts. Generally, each concept is represented by a favored term (word or phrase).

Ontology Creation (Editor) Tool: A tool that enables ontology specification and creation through some collaborative authoring. It usually requires users to be trained in knowledge representation and predicate logic. They are supposed to offer a server-based environment with support for consistency checking of interconnected ontologies and a collaborative environment for model review and refinement.

OWL: It has been designed to meet the need for a Web ontology language. OWL is part of the growing stack of W3C recommendations related to the semantic Web. OWL adds more vocabulary for describing properties and classes: among others, relations between classes (e.g., disjointness), cardinality (e.g., "exactly one"), equality, richer typing of properties, characteristics of properties (e.g., symmetry), and enumerated classes.

RDF: The Resource Description Framework is a framework for representing information in the Web. The underlying structure of any expression in RDF is a collection of triples, each consisting of a subject, a predicate, and an object. A set of such triples is called an RDF graph. This can be illustrated by a node and directed-arc diagram, in which each triple is represented as a node-arc-node link (hence the term graph).

Semantic Technology: It is a software technology that allows the meaning of and associations between information to be known and processed at execution time. For a semantic technology to be truly at work within a system, there must be a knowledge model of some part of the world that is used by one or more applications at execution time.

Semantic Web: It provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF) and other Web ontology description languages (OWLs), which integrate a variety of applications using XML for syntax and Uniform Resource Identifiers (URIs) for naming.

SHOE: Simple HTML ontology extensions (SHOE) is a knowledge representation language that allows Web pages to be annotated with semantics.

O

One Method for Design of Narrowband Low-Pass Filters

Gordana Jovanovic-Dolecek
National Institute of Astrophysics Optics and Electronics, INAOE, Mexico

Javier Díaz-Carmona
National Institute of Astrophysics Optics and Electronics, INAOE and Technology Institute of Celaya (ITC), Celaya, Mexico

INTRODUCTION

Stearns and David (1996) states that "for many diverse applications, information is now most conveniently recorded, transmitted, and stored in digital form, and as a result, digital signal processing (DSP) has become an exceptionally important modern tool." Typical operation in DSP is digital filtering. Frequency selective digital filter is used to pass desired frequency components in a signal without distortion and to attenuate other frequency components (Smith, 2002; White, 2000). The pass-band is defined as the frequency range allowed to pass through the filter. The frequency band that lies within the filter stop-band is blocked by the filter and therefore eliminated from the output signal. The range of frequencies between the pass-band and the stop-band is called the transition band and for this region no filter specification is given.

Digital filters can be characterized either in terms of the frequency response or the impulse response (Diniz, da Silva, & Netto, 2000). Depending on its frequency characteristic, a digital filter is either low-pass, high-pass, band-pass or band-stop filters. A low-pass (LP) filter passes low frequency components to the output while eliminating high-frequency components. Conversely, the high-pass (HP) filter passes all high-frequency components and rejects all low-frequency components. The band-pass (BP) filter blocks both low- and high-frequency components while passing the intermediate range. The bands-stop (BS) filter eliminates the intermediate band of frequencies while passing both low- and high-frequency components.

In terms of their impulse responses, digital filters are either infinite impulse response (IIR) or finite impulse response (FIR) digital filters. Each of four types of filters (LP, HP, BP, and BS) can be designed as an FIR or an IIR filter (Grover & Deller, 1999; Ifeachor & Jervis, 2001; Oppenheim & Schafer, 1999).

The design of a digital filter is carried out in three steps (Ingle & Proakis, 1997):

- Define filter specification;
- Approximate given specification;
- Implement digital filter in hardware or software.

The topic of filter design is concerned with finding a magnitude response (or, equivalently, a gain) which meets the given specifications. These specifications are usually expressed in terms of the desired pass-band and stop-band edge frequencies w_p and w_s, the permitted deviations in the pass-band (pass-band ripple) R_p, and the desired minimum stop-band attenuation A_s, (Mitra, 2001). Here, we consider the specifications given in dB. In this case, the gain in decibels is

$$\text{Gain in dB} = 20\log_{10}\left|H(e^{j\omega})\right|. \tag{1}$$

Figure 1 illustrates a typical magnitude specification of a digital low-pass filter.

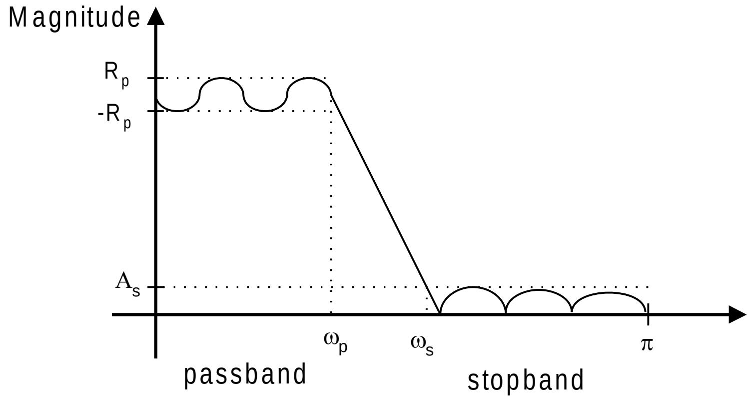
Due to their complexity, narrowband low-pass FIR filters are difficult and sometimes impossible to implement using conventional structures (Milic & Lutovac, 2002). The interpolated finite impulse response (IFIR) filter proposed by Neuvo, Cheng-Yu, and Mitra (1984) is one efficient realization for the design of narrowband FIR filters.

The IFIR filter $H(z)$ is a cascade of two filters,

$$H(z) = G(z^M)I(z), \tag{2}$$

where $G(z^M)$ is an expanded shaping or model filter, $I(z)$ is an interpolator or image suppressor, and M is the interpolator factor. In this manner, the narrowband FIR prototype filter $H(z)$ is designed using lower order filters, $G(z)$ and $I(z)$. For more details on the IFIR structure, see Neuvo, Cheng-Yu, and Mitra (1984) and Jovanovic-Dolecek (2003).

Figure 1. Low-pass filter magnitude specification



An increase in the interpolation factor results in the increase of the interpolation filter order as well as in the decrease of the shaping filter order.

The design goal in Jovanovic-Dolecek and Diaz-Carmona (2003) is to decrease the shaping filter order as much as possible, and to efficiently implement the high order interpolator filter. To do so, we propose to use a sharpening recursive running sum (RRS) as an interpolator in the IFIR structure.

BACKGROUND

Sharpening Technique

The sharpening technique, which was first proposed by Kaiser and Hamming (1984), attempts to improve both the pass-band and stop-band of a linear FIR filter by using multiple copies of the same filter based on the Amplitude Change Function (ACF). An ACF is a polynomial relationship of the form $H_0(w) = P[H(w)]$ between the amplitude responses of the overall and the prototype filters, $H_0(w)$ and $H(w)$, respectively. The improvement in the pass-band, near H=1, or in the stop-band, near H=0, depends on the order of the tangency of the ACF m and n at H=1 or H=0, respectively. The expression proposed by Kaiser and Hamming for an ACF giving the mth and nth order tangencies at H = 1 and H = 0, respectively, is given by

$$H_0 = H^{n+1} \sum_{k=0}^{m} \frac{(n+k)!}{n!\,k!}(1-H)^k$$
$$= H^{n+1} \sum_{k=0}^{m} C(n+k,k)(1-H)^k , \qquad (3)$$

where $C(n+k,k)$ is the binomial coefficient.

Hartnett and Boudreaux (1995) proposed an extension of this method by introducing the slope d of tangency at H=0 and the slope s of tangency at H=1. Samadi (2000) proposed an explicit formula for the filter sharpening polynomial.

We use the following ACF polynomials:

1) $P\left[H(\omega)\right] = 3H^2(\omega) - 2H^3(\omega); \delta = \sigma = 0; m = n = 1.$

2) $P\left[H(\omega)\right] = 6H^2(\omega) - 8H^3(\omega) + 3H^4(\omega); \delta = \sigma = 0; m = 2; n = 1.$

3) $P\left[H(\omega)\right] = 10H^2(\omega) - 20H^3(\omega)$
$+15H^4(\omega) - 4H^5(\omega); \delta = \sigma = 0; m = 3, n = 1.$

(4)

The plot of each one of these ACF is shown in Figure 2. Note that the third ACF has the best passband improvement and the smallest stopband attenuation, whereas, the first ACF has the worst pass-band improvement but the best stopband attenuation.

The three sharpening structures are presented in Figure 3, where the resulting number of multipliers per output sample (MPS) is equal to three, four and five for the first, second and third structure, respectively.

SHARPENING RRS FILTER

The simplest low-pass FIR filter is the moving-average (MA) filter. Its impulse response $g(n)$ is given by,

$$g(n) = \frac{1}{M} \sum_{k=0}^{M-1} g(n-k) . \qquad (5)$$

Figure 2. Plot of the three ACF's

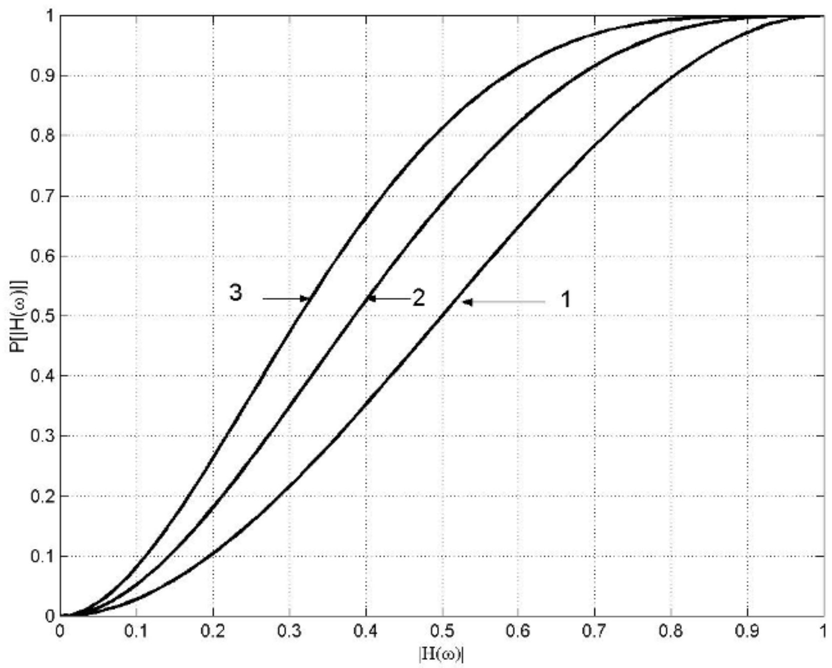


Figure 3. Sharpening structures

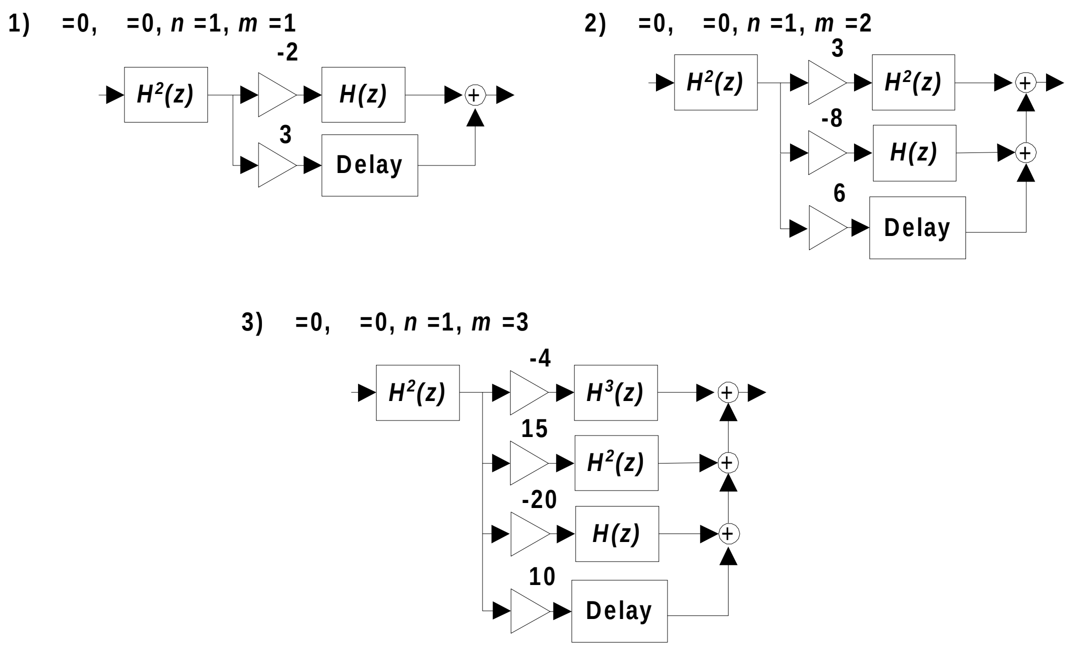
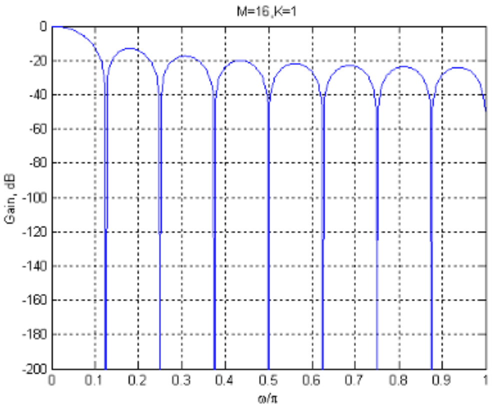
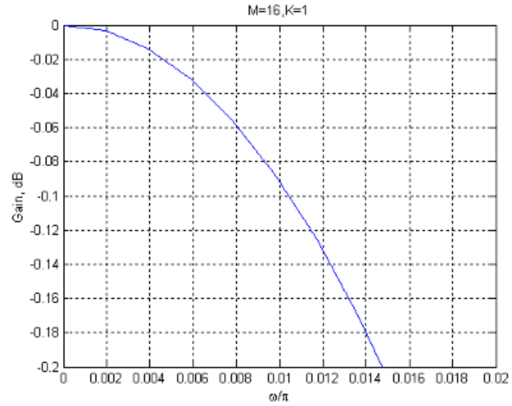


Figure 4. Magnitude responses of the RRS filter with M=16 and K=1, 2, and 3

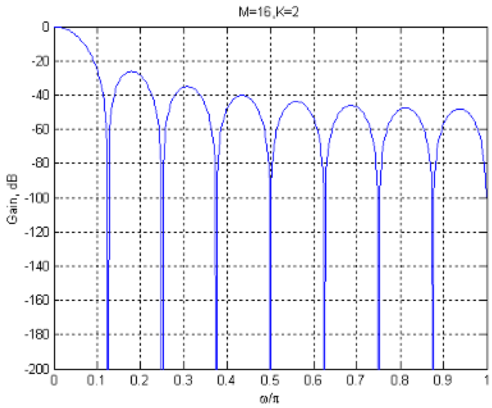
a. Overall magnitude response



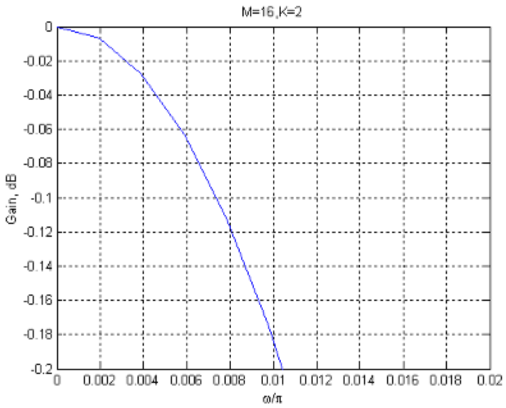
b. Pass-band zoom



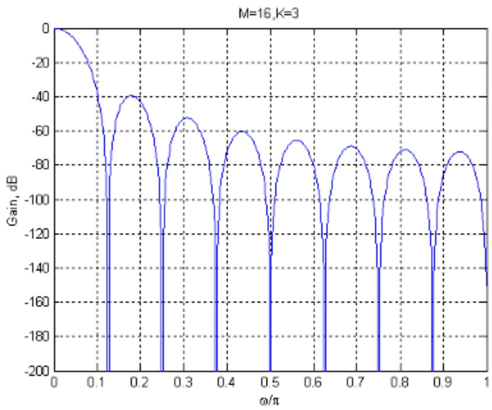
c. Overall magnitude response



d. Pass-band zoom



e. Overall magnitude response



f. Pass-band zoom

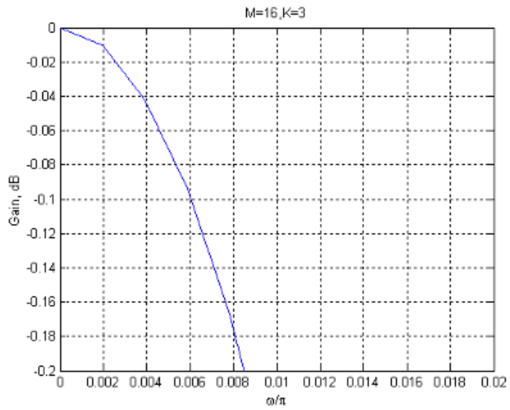
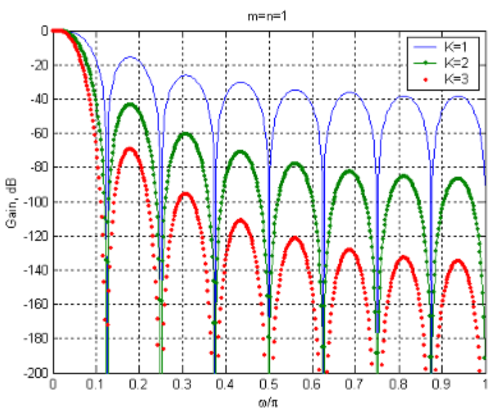
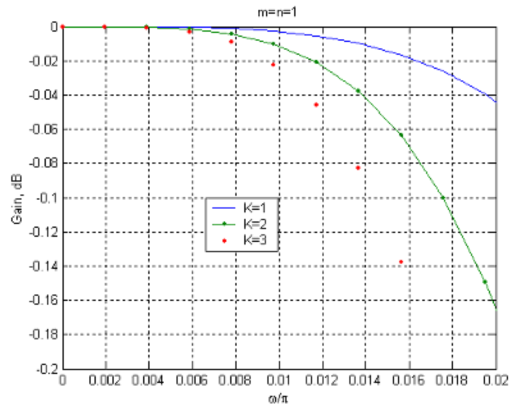


Figure 5. Magnitude responses of the sharpened RRS filters with M=16 and K=1, 2, and 3

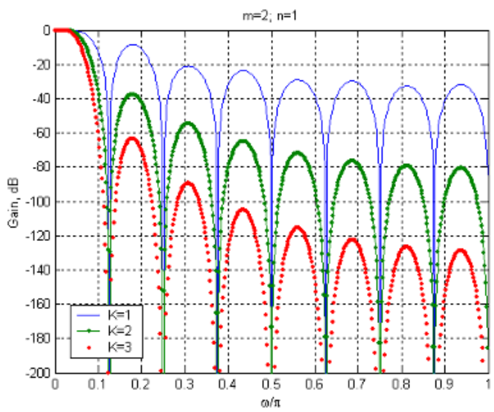
a. Overall magnitude response



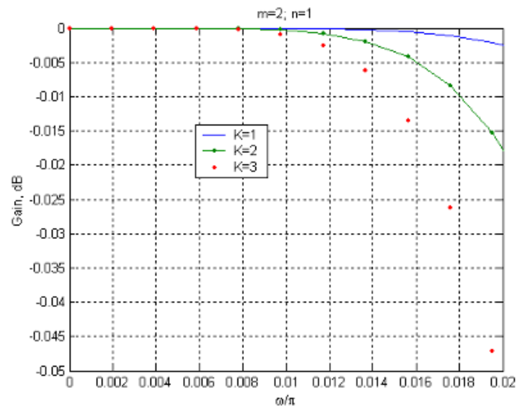
b. Pass-band zoom



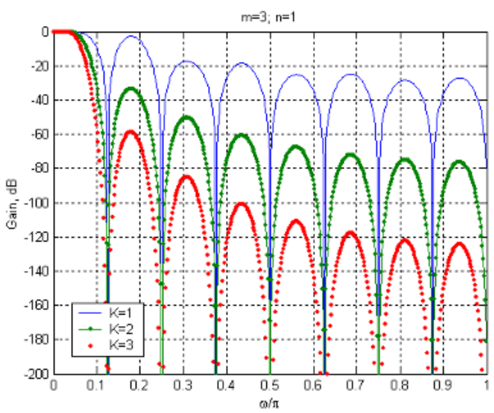
c. Overall magnitude response



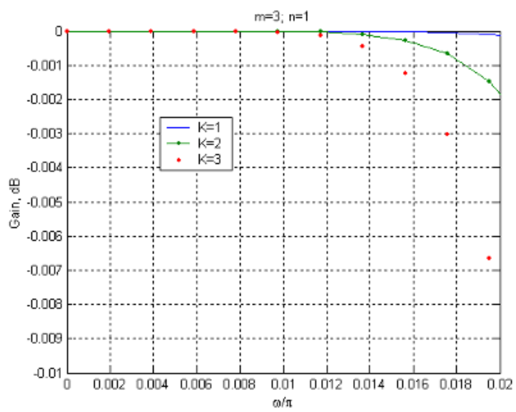
d. Pass-band zoom



e. Overall magnitude response



f. Pass-band zoom



Since all impulse response coefficients are equal to 1, the filter requires no multiplications.

Its transfer function is given by

$$G(z) = \frac{1}{M}[1 + z^{-1} + ... + z^{-(M-1)}] = \frac{1}{M}\sum_{k=0}^{M-1} z^{-k} \quad . \ (6)$$

A more convenient form of the above transfer function for realization purposes is given by

$$H_{RRS}(z) = [G(z)]^K = \left[\frac{1}{M}\frac{1 - z^{-M}}{1 - z^{-1}}\right]^K, \qquad (7)$$

also known as a recursive running-sum filter (RRS) (Mitra, 2001). The scaling factor $1/M$ is needed to provide a dc gain of 0 dB, and K is the number of the cascaded sections of the filter. The magnitude response has nulls at integer multiples of $(2\pi / M)$ as shown in the following relation

$$\left|H_{RRS}(e^{j\omega})\right| = \left| \left[\frac{\sin(\omega M/2)}{M\sin(\omega/2)}\right]^K \right|. \qquad (8)$$

This characteristic provides natural alias-rejections introduced by expanding the shaping filter $G(z)$ by factor M. The stop-band attenuation is improved as a result of the increased number of stages K. Figure 4 illustrates the overall magnitude responses and the corresponding pass-band details of the RRS filter for M=16 and K=1, 2, and 3. We can notice that the characteristics have a high pass-band droop which limits the application of the RRS filters. In order to improve the characteristics we propose to use the sharpening polynomials (4). Figure 5 shows the improved magnitude responses for the RRS filter with M=16 and the cascade of 2 and 3 RRS filters using sharpening polynomials (4).

DESIGN PROCEDURE

This section describes the design of a narrowband lowpass filter using an IFIR structure, where the interpolator filter is the sharpening recursive running (RRS) filter. In order to suppress the mirror images of the expanded model filter, the RRS frequency nulls must be centered at each mirror image. This implies that the length of the RRS filter must be equal to the interpolator factor M.

The design procedure is outlined in the following steps:

1. Choose the interpolation factor M and design the model filter $G(z)$, using the specifications:

$$\omega_p^G = M\omega_p, \qquad \omega_s^G = M\omega_s,$$

$$R_p^G = \frac{R_p}{2}, \qquad A_s^G = A_s \quad , \qquad (9)$$

   where ω_p^G and ω_s^G are the pass-band and the stop-band edge frequencies, R_p^G is the maximum pass-band ripple and A_s^G is the minimum stop-band attenuation of the filter $G(z)$.
2. Design the K-stage RRS filter $I_{RRS}(z)$ of the length M.
3. Choose an ACF and apply the filter sharpening to the filter $I_{RRS}(z)$. The resulting filter is denoted as $I_{shRRS}(z)$.
4. Cascade the model filter $G(z^M)$ and the sharpening RRS filter $I_{shRRS}(z)$.
5. If the resulting stop-band filter specification A_s is not satisfied, go to Step 2 and increase the number of stages K.
6. If the pass-band filter deviation is higher than R_p go to Step 3 and change ACF.

The number of stages in the RRS filter K controls the resulting filter stop-band attenuation and the filter sharpening technique controls the filter pass-band width. The design algorithm is illustrated with following example.

Example:

A linear-phase narrowband lowpass FIR filter is designed to meet these specifications: normalized pass-band edge at w_p=0.01, normalized stop-band edge at w_s=0.015, maximum pass-band ripple R_p =0.086dB, and minimum stop-band attenuation A_s=-60 dB.

The linear-phase FIR filter designed by Parks McClellan method would have the order N=1017, which translates into 509 multiplications per output sample (MPS). Using M=18 in the IFIR structure, the orders of filters G(z) and I(z), both designed by Parks McClellan method, are 66 and 69 respectively, thereby requiring a total of 68 MPS.

By choosing the first ACF given in Equation (4), the interpolation factor M is 21, and the number of the stages of the RRS filter K is 2. The order of the model filter $G(z)$ is N_G= 57. The resulting structure is shown in Figure 4. The all-band magnitude response for the designed filter and the pass-band zoom are shown in Figure 5.

Figure 6. Resulting structure in example 1 using the first ACF

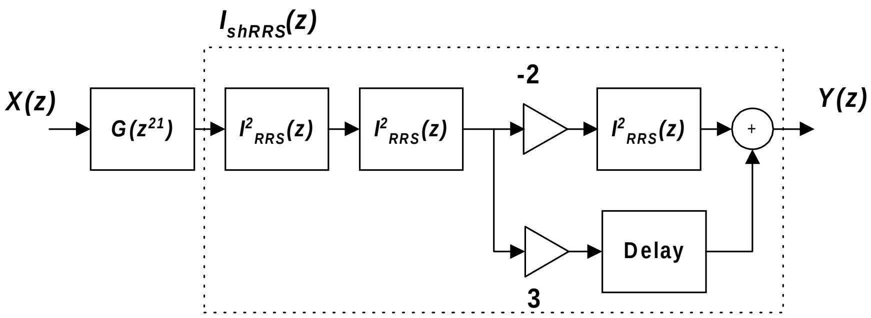


Figure 7. Magnitude response for example 1 and first ACF

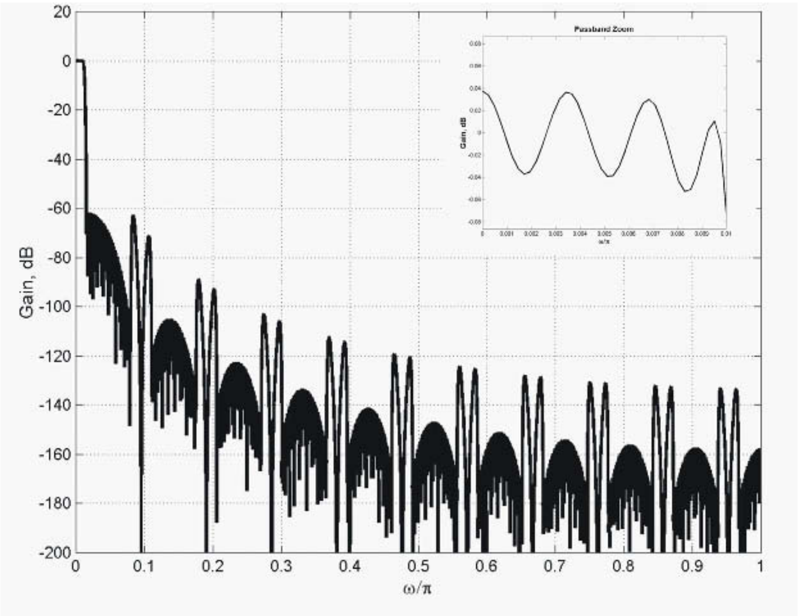


Table 1. Design parameters and number of MPS for example 1

ACF	M	N_G	L	MPS
1	21	57	2	31
2	31	40	3	24
3	37	34	4	22

If instead, the second ACF in Equation (4) is used as a sharpening filter polynomial, it would be possible to use even higher values for the interpolation factor M, such as M=31. In this case, the RRS filter has 3 stages, and the resulting order of the filter $G(z)$ is 40. A further order reduction of the filter $G(z)$ is obtained by using the third ACF in Equation (4). The interpolation factor M is then 37, and the number K of the stages of RRS filter is equal to 4. As a result, the model filter has order N_G= 34. The design parameters for all three cases and the resulting number of MPS are summarized in Table 1. Note that the number of MPS is decreased as a result of the increase in the complexity of the ACF.

FUTURE TRENDS

The use of IFIR filter design techniques is now a rather well-developed topic in the field of digital filters. Nonetheless, an analytical technique for obtaining a design employing the minimum number of filter-tap multipliers has not previously appeared and presents an important research task (Mehrnia & Willson, 2004). The future research also includes a multistage design and finding an optimal choice of the corresponding interpolation factors.

Another important issue that needs to be addressed is how to choose the sharpening polynomials (as a trade-off between the values of tangencies m and n and the number of cascaded basic RRS filters) that will yield an improved characteristic of the RRS filter and a decreased complexity of the overall design.

CONCLUSION

An overview of the design method for narrowband linear-phase FIR filter with sharp transition band is presented. The method reduces the complexity of an IFIR structure by using a high value of the interpolation factor which in turn makes the order of the model filter low.

Furthermore, the resulting high order interpolation filter is replaced by a recursive running sum (RRS) filter, which does not need any multiplications. The sharpening technique, based on the amplitude change function (ACF), improves the magnitude characteristic of the RRS filter.

Three amplitude change function (ACF) with the tangency orders $n = 1$, and $m = 1, 2, 3$, are used. As a result, the total number of multipliers per output sample is significantly decreased.

REFERENCES

Diniz, P.S.R., da Silva, E.A.B., & Netto, S.L. (2002). *Digital signal processing, system analysis and design*. Cambridge: Cambridge University Press.

Grover, D., & Deller, J.R. (1999). *Digital signal processing and the microcontroller*. New Jersey: Prentice Hall, Inc.

Hartnett, R.J., & Boudreaux, G.F. (1995, December). Improved filter sharpening. *IEEE Transactions on Signal Processing*, *43*, 2805-2810.

Ifeachor, E.C., & Jervis, B.E. (2001). *Digital signal processing: A practical approach* (2nd ed.). NJ: Prentice Hall.

Ingle, V.K., & Proakis, J.G. (1997). *Digital signal processing using MATLAB*. Boston: PWS Publishing Company.

Jovanovic-Dolecek, G. (2003). Design of narrowband highpass FIR filters using sharpening RRS filter and IFIR structure. In J. Peckham & S.J. Lloyd (Eds.), *Practicing software engineering in the 21st century*, (pp.272-294). Hershey, PA: Idea Group Publishing.

Jovanovic-Dolecek, G., & Diaz-Carmona J. (2003). One method for design of narrowband lowpass filters. In J. Peckham & S.J. Lloyd (Eds.), *Practicing Software Engineering in the 21st Century*, (pp.258-271). Hershey, PA: Idea Group Publishing.

Kaiser, J.F., & Hamming, R.W. (1984, October). Sharpening the response of a symmetric non-recursive filter by multiple use of the same filter. *IEEE Transactions on Acoustics Speech and Signal Processing*, *ASSP-25*, 415-422.

Mehrnia, A., & Willson, Jr., A.N. (2004). On optimum IFIR filter design. In *Proceedings of the International Conference ISCAS 2004*, Vancouver, Canada, May 2004, III (pp.13-136).

Milic, L., & Lutovac, M. (2002). Efficient multirate filtering. In G. Jovanovic-Dolecek (Ed.), Chapter IV of *Multirate systems: Design and applications*. Hershey: Idea Group Publishing.

Mitra, S.K. (2001). *Digital signal processing: A computer-based approach* (2nd ed.). New York: McGraw-Hill.

Neuvo, Y., Cheng-Yu, D., & Mitra, S. (1984, June). Interpolated finite impulse response filters. *IEEE Transactions on Acoustics Speech and Signal Processing*, *32*, 563-570.

Oppenheim, A.V., & Schafer, R.W. (1999). *Discrete-time signal processing* (2nd ed.). New Jersey: Prentice-Hall, Inc.

Samadi, S. (2000, October). Explicit formula for improved filter sharpening polynomial. *IEEE Transactions on Signal Processing*, *9*, 2957-2959.

Smith, S. (2002). *Digital signal processing: A practical guide for engineers and scientists*. NY: Newnes.

Stearns, S.D., & David, R. (1996). *Signal processing algorithms in MATLAB*. New Jersey: Prentice-Hall, Inc.

White, S. (2000). *Digital signal processing: A filtering approach*. Delmar Learning.

KEY TERMS

ACF: An ACF (Amplitude change function) is a polynomial relationship of the form $H_0(w) = P[H(w)]$ between the amplitude responses of the overall and the prototype filters, $H_0(w)$ and $H(w)$, respectively.

Digital Filter Design: The process of deriving the transfer function of the filter is called digital filter design. It is carried out in three steps: Definition of filter specification, Approximation of given specification, and Implementation of digital filter in hardware or software.

Expanded Filter: Expanded filter $G(z^M)$ is obtained by replacing each delay z^{-1} in the filter $G(z)$ with M delays z^M. In the time domain this is equivalent to inserting M-1 zeros between two consecutive samples of the impulse response of $G(z)$.

FIR Filter: Digital filter with finite impulse response. This filter is always stable and can be designed to have linear phase. (Filters with the symmetric impulse response). The main disadvantage is high complexity.

Frequency Selective Filters: Digital filters which pass desired frequency components in a signal without distortion and attenuate other frequency components. A low-pass (LP) filter passes low frequency components to the output while eliminating high-frequency components. Conversely, a high-pass (HP) filter passes all high-frequency components and rejects all low-frequency components. A band-pass (BP) filter blocks both low- and high-frequency components while passing the intermediate range. A bands-stop (BS) filter eliminates the intermediate band of frequencies while passing both low- and high-frequency components.

IIR Filter: Digital filter with infinite impulse response. This filter must be checked for the stability and does not have linear phase. This filter has lower complexity than its FIR counterpart.

Interpolator: The filter $I(z)$ which is used to eliminate the unwanted spectrum introduced by expansion of the model filter in an IFIR structure. For the low-pass prototype filter design, this filter is a low-pass filter with lower complexity than the prototype filter.

Model or Shaping Filter: The filter $G(z)$ in the IFIR structure which has M times higher both the pass-band and the stop-band frequencies, than the prototype filter. Since both the stop-band and the pass-band increase M fold so does their difference, that is, the transition band. As a consequence of increased transition band the order of the filter decreases, and in turn the overall complexity decreases as well.

MPS: Measure of the computational complexity of a digital filter expressed in terms of multipliers per output sample.

Pass-band: Pass-band is defined as the frequency range allowed to pass through the filter.

Passband Ripple: The permitted deviation in the pass-band.

Stop-band: The frequency band that is blocked by the filter and therefore eliminated from the output signal.

Stopband Attenuation: The desired minimum attenuation in the stopband.

Transition Band: The range of frequencies between the pass-band and the stop-band is called the transition band, and for which no filter specification is given.

One-to-One Video-Conferencing Education

Hock Chuan Chan
National University of Singapore, Singapore

Bernard C.Y. Tan
National University of Singapore, Singapore

Wei-Ping Tan
National University of Singapore, Singapore

INTRODUCTION

Technologies have become a critical component affecting teaching and learning effectiveness. Technologies enable informal computer-mediated interaction among instructors and students, for example, through electronic mail or bulletin boards. The Internet has changed the dynamics of teaching and learning by enabling distance direct personal tutoring, whereby a tutor (a private or personal instructor) provides personal additional instruction and attention to a student. The cost of video-conferencing using standard personal computers and off-the-shelf software involves a low set-up cost and very low usage fee (local phone charges). Its low cost can lead to a proliferation of its use. Students are no longer physically constrained in their quest for tutors. It is important to research the factors that may facilitate or hinder learning via Internet video-conferencing capabilities.

A case study was conducted, through multiple data collection methods, with two tutors and three students in Singapore. The impacts of four critical factors (system characteristics, mode characteristics, social presence, and media richness) on the effectiveness of teaching and learning were studied. This study helps to fill a gap in knowledge that arises because prior studies tend to concentrate on big virtual classroom settings.

BACKGROUND

Earlier studies on the use of information technology for education have focused on student usage of learning software, or on group interactions. For example, Sankar et al. (1997) investigated two different ways of delivering lectures with video-conferencing, without any instructor-student interaction. There are very few studies on one-to-one distance education. Graesser and Person (1994) found that the quality of student questions was correlated with grades in one-to-one tutoring sessions on mathematics. It was found that collaborative problem solving, prompt question answering, and clarity of explanation using examples contributed significantly to learning effectiveness (Graesser et al., 1995). Hume et al. (1996), studying one-to-one tutoring effectiveness, found that hints encouraged students to engage in active cognitive processes that promoted long-term retention and deeper understanding. Chi (1996) found that certain types of interaction between tutors and students, during one-to-one tutoring sessions, could produce deep learning.

The body of literature on one-to-one distance education motivates this research effort in two ways. First, like all the existing studies, this study seeks to identify factors that may enhance the effectiveness of teaching and learning in such an environment. In this study, effectiveness is measured by asking instructors to indicate their perceived ability to teach and asking students to indicate their perceived ability to learn via distance education, relative to traditional face-to-face education sessions. Second, while the results of all the existing studies alluded to the importance of communication between the tutor and the instructor, this issue has never been directly investigated. Therefore, this study focuses on identifying factors that may impact the communication process between the instructor and the student, thereby affecting distance learning effectiveness.

One-to-one distance education is examined in the context of desktop video-conferencing because the economy and prevalence of desktop video-conferencing facilities are likely to make it a dominant mode of distance education in the future (Rhodes, 2001). Table 1 presents four critical factors that can affect the success of using desktop video-conferencing facilities for education.

System Characteristics

Every desktop video-conferencing facility has both hardware and software components. A digital camera and a video card (in some products) are needed to capture images. A microphone, a sound card, and speakers are

Table 1. Factors affecting teaching and learning effectiveness

Factor	Key aspects of factor
System characteristics	Hardware, software, and bandwidth
Mode characteristics	Usefulness, challenge, attractiveness, and clarity
Social presence	Sociability, warmth, and personal focus
Media richness	Multiple cues and interactivity

needed to capture and project voices. Many windows are needed to display the captured images (of the tutor or the student), the chat window, and other applications such as Word. In addition, the software should facilitate document sharing. Bandwidth limitations on the Internet and processing speed could lead to grainy pictures and a lack of synchronization between video and audio signals (Tackett, 1995), thereby affecting teaching and learning effectiveness.

Mode Characteristics

Four key perceptual characteristics of a mode of instruction determine its effectiveness: usefulness, challenge, attractiveness, and clarity (Champness & DeAlberdi, 1981; Sankar et al., 1995). Usefulness refers to how much the mode of instruction is perceived to be appropriate for the learning task. Challenge is the extent to which the mode of instruction is able to facilitate learning of difficult concepts. Attractiveness is the extent to which the mode of instruction is perceived to be lively, exciting, and interesting. Clarity refers to the extent with which the mode of instruction is perceived to allow comprehensible communication.

Social Presence

Social presence is defined as the extent to which a communication medium allows the actual physical presence of the communicating partners to be conveyed, and how far it allows communicating parties to socialize with each other, feel the warmth of each other, and exchange messages that are personal in nature (Markus, 1994; Short et al., 1976). Short et al. (1976) rank the following five communication media in order of decreasing social presence: face-to-face, television, multi-speaker audio, telephone, and business letter. The literature suggests that desktop video-conferencing may enable people to transmit more warmth and sociability than the telephone. Video images can help people who have just met recently to become more familiar with one another (Czeck, 1995). Facial signals can allow the instructor to assess student understanding.

Media Richness

Media richness is defined as the extent to which a communication medium can facilitate shared understanding (Daft et al., 1987). Rich media enable people to overcome equivocality, the existence of different interpretations. Variables typically used to gauge media richness include multiple cues and interactivity. Trevino et al. (1987) provide this ranking, in decreasing order of media richness: face-to-face, telephone, and printed documents. Face-to-face meeting has interactivity and immediate feedback so that mutual understanding between communicating parties can be checked and differences in interpretations reconciled. This medium also carries visual, verbal, and textual cues. Printed documents have neither interactivity nor immediate feedback. Desktop video-conferencing facilities may be less rich than face-to-face meetings because interactivity may be lower (Rice, 1992). Interactivity between the instructor and the student is a critical success factor for distance education (Milhem, 1996; Millbank, 1994).

RESEARCH METHODOLOGY

A case study research approach is appropriate for examining a phenomenon in its natural setting, and it leads to in-depth answers (Yin, 2002). A prior survey with undergraduates and high school students revealed that most of them had no experience with distance education via desktop-video-conferencing facilities (Tan & Chan, 1998). Since this mode of learning was relatively new, the case study research methodology was employed. Five subjects (two tutors and three students) volunteered for this study. All the one-to-one tutoring sessions involved two individuals interacting with each other. However, the primary interest of this study was to identify factors that influence the communication efforts of an individual (tutor or student), thereby impacting his or her ability to teach or learn. Prior studies on one-to-one tutoring sessions have also focused on the perceptions of an individual, rather than the interaction between two individuals (Chi, 1996). The two tutors were undergraduates from a large university. They each had about 2 years of experience with one-to-one tuition in a traditional face-to-face context. The three students were attending high school.

Table 2. Summary of ratings

Subject	System characteristics	Method characteristics	Social presence	Media richness	Overall experience
Student Y	Fair	Fair	Poor	Poor	Poor
Student X	Neutral	Neutral	Poor	Poor	Fair
Tutor B	Neutral	Good	Fair	Neutral	Neutral
Tutor A	Fair	Good	Neutral	Neutral	Good
Student Z	Good	Excellent	Excellent	Excellent	Excellent

They had been receiving one-to-one tuition from the two tutors for a year at their respective homes. All the tutors and students had no prior experience with distance education of any kind. The topic of instruction was mathematics at a high school level.

All tuition sessions were conducted using desktop video-conferencing tools and involved a tutor and a student. All personal computers used for desktop video-conferencing tuition sessions in this study were connected to the Internet via either a local area network or a 56 kbps modem. This speed was more reflective of distance education, where students and tutors are connected through the telephones. Microsoft NetMeeting was installed to facilitate communication. A digital camera and a video card were used to capture video signals. A microphone, a sound card, and two speakers were used to capture and project voice signals. For each subject (tutor or student), data from five such desktop video-conferencing tuition sessions were collected. To preserve the realism of the natural tuition context, there was minimal control over the instructional process and materials (Yin, 2002). Multiple methods for data collection including on-site observation, computer and video recording, and interviews were used (Miles & Huberman, 1994; Yin, 2002).

DATA ANALYSES

For each case, all the data were analysed for the four main factors (system characteristics, mode characteristics, social presence, and media richness) and perceptions of effectiveness in teaching or learning. Analysis results were verified with the subjects. The subjects also contributed additional insights and highlighted issues that were of particular concern to them. Overall, two subjects were in favour of the video-conferencing tutoring, while three provided unfavorable feedback. The evidence suggests that subjects were affected by system characteristics. Those who experienced technical difficulties in starting up their systems and had more problems with visual and voice signals tended to form poorer perceptions of their tuition sessions. Although they were clearly frustrated when they encountered system problems, subjects were generally optimistic when asked to comment on the future of such applications.

The results on mode characteristics suggest that subjects tend to have better perceptions of their tuition sessions if they could understand complex concepts through these means, or if the ideas exchanged were clear enough to them. Ways to enhance exchange of complex ideas include having “some methods to decompose a complex idea into several simpler ideas before communication” as suggested by a tutor. Overall, subjects felt this mode of education had increased their interest in mathematics because it was “interesting,” “amazing” and “fun”.

Results on social presence and media richness suggest that subjects tend to form better perceptions of their tuition sessions if they could feel the presence of the other party, if they could send personal messages, or if they could get replies quickly. People who have developed shared understanding on how to work with each other through the new technology are less affected by changes in social presence and media richness (Lee, 1994). Therefore, tutors and students are likely to be able to free themselves from such restrictions if “we have enough experience with such tuition sessions”.

It is evident from Table 2 that the four factors are reasonably good predictors of success with distance education employing one-to-one desktop video-conferencing. As each of these four factors improves, the overall experience is likely to be enhanced. Specifically, the strongest discriminatory factor for success seems to be media richness, while the weakest discriminatory factor seems to be system characteristics. Given their predictive power, these four factors could serve as a basis for future research efforts on distance education involving large-scale surveys.

FUTURE TRENDS

While the use of desktop video-conferencing facilities for one-to-one tuition sessions has its limitations, it has a promising future too. The subjects have some technical and procedural suggestions on how to improve the quality of such tuition sessions in the future (see Table 3).

Future research can implement or fine-tune software tools to put in place procedural routines. Future research can also replicate this study with different topics of

Table 3. Suggestions for improvement

Category	Suggestions
Technical	▪ Use pen-based interface to draw pictures and write formulae. ▪ Replace the white board with shared document to increase shared space. ▪ The system should capture all the transmitted information for reference. ▪ Consider using ICQ instead of NetMeeting to speed up connection time. ▪ Bypass the Internet if other connections are available (e.g., direct phone).
Procedural	▪ Should speak slower with regular intervals to enhance audio quality. ▪ Close video window to improve overall transmission quality. ▪ Use multiple means to make explanation of complex concepts clearer. ▪ Solicit feedback at regular intervals. ▪ Respond quickly to suggestions or queries. ▪ Try to engage in more small talk or share jokes more frequently. ▪ Should meet face-to-face before using the technology.

instruction (e.g., languages or sciences) and different levels of instruction (e.g., middle school or college). This study has focused solely on tutors and students from Singapore but future research efforts can examine the same phenomenon in a multi-cultural context. Research has shown that culture does moderate the impact of technologies on human behavior (Tan et al., 1998), particularly since distance learning via desktop video-conferencing capabilities can match tutors and students from different cultures. Also, as network speed improves, further studies could be made to ascertain whether better speed has any significant positive effects on the other factors.

CONCLUSION

This study demonstrates the promise of desktop video-conferencing technologies as a means for conducting one-to-one distance education. Given the widespread availability of low-cost personal computers and Internet access, it is plausible that this mode of distance education may partially replace face-to-face tuition sessions. More importantly, the benefits of the technological capabilities examined in this article can extend beyond one-to-one tuition sessions to larger-scale distance education efforts. Although this study focuses on desktop video-conferencing technologies, other technologies must not be neglected. In addition to seeing each technology as a solution to an existing problem, it is also fruitful to examine how a collection of information and communication technologies may open up new and exciting possibilities for education.

Information and communication technologies have permeated many aspects of education. These technologies will continue to impact education in the future. An example of how such technologies may alter the future of education is a course on Global Project Coordination, jointly conducted by National University of Singapore, Stanford University, and Swedish Royal Institute of Technology. In this course, students from the three universities enrol in the same global class. Faculties from the three universities take turns to give weekly lectures to the entire global class via a three-way video-conferencing facility, based on real-time multicast technologies. Students also form global teams to work on large-scale projects sponsored by the industry. As part of this course, faculties and students employ a wide range of technologies to communicate with and learn from each other. The desktop videoconferencing capabilities investigated here can certainly facilitate such learning efforts by helping to breach geographical barriers among faculties and students of such a global class.

Although this article focuses on desktop video-conferencing technologies as a means for distance education, other existing and emerging technologies based on the Internet must not be neglected. In the 21st century, when information and communication technologies are likely to play a critical role in enabling effective education, knowledge accumulated on existing and emerging technologies can guide us in terms of what technologies are appropriate under what circumstances. Rather than seeing each technology as a solution to an existing problem, it is more fruitful to examine how the collection of information and communication technologies may complement each other to open up new and exciting possibilities for educating people.

REFERENCES

Champness, B., & DeAlberdi, M. (1981). *Measuring subjective reactions to teletext page design*. NSF Grant DAR-7924489-A02. New York, NY: Alternate Media Centre, New York University.

Chi, M.T.H. (1996). Constructing self-explanations and scaffolded explanations in tutoring. *Applied Cognitive Psychology, 10*(Special Issue), S33-S49.

Czeck, R. (1995). Videoconferencing: Benefits and disadvantages to communication. *http://www.ils.unc.edu/cscw*

Daft, R.L., Lengel, R.H., & Trevino, L.K. (1987). Message equivocality, media selection, and manager performance: Implications for information systems. *MIS Quarterly, 11*(3), 355-366.

Graesser, A.C., & Person, N.K. (1994). Question asking during tutoring. *American Educational Research Journal, 31*(1), 104-137.

Graesser, A.C., Person, N.K., & Magliano, J.P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology, 9*(6), 495-522.

Hume, G., Michael, J., Rovick, A., & Evens, M. (1996). Hinting as a tactic in one-to-one tutoring. *Journal of the Learning Sciences, 5*(1), 23-47.

Lee, A.S. (1994). Electronic mail as a medium for rich communication: An empirical investigation using hermeneutic interpretation. *MIS Quarterly, 18*(2), 143-157.

Markus, M.L. (1994). Electronic mail as the medium of managerial choice. *Organization Science, 5*(4), 502-527.

Miles, M.B., & Huberman, A.M. (1994). *Qualitative data analysis: An expanded sourcebook.* Newbury Park, CA: Sage.

Milhem, W.D. (1996). Interactivity and computer-based interaction. *Journal of Education Technology Systems, 23*(3).

Millbank, G. (1994). *Writing multimedia training with integrated simulation.* Writers' Retreat on Interactive Technology and Equipment. Vancouver, BC: University of British Columbia.

Rhodes, J. (2001). *Videoconferencing for the real world: Implementing effective visual communications systems.* Focal Press.

Rice, R.E. (1992). Task analyzability, use of new media, and effectiveness: A multi-site exploration of media richness. *Organization Science, 3*(4), 475-500.

Sankar, C.S., Ford, F.N., & Terase, N. (1997). Impact of videoconferencing in teaching an introductory MIS course. *Journal of Educational Technology Systems, 26*(1), 67-85.

Sankar, C.S., Kramer, S.W., & Hingorani, K. (1995). Teaching real-world issues: Comparison of written versus annotated still image case study. *Journal of Educational Technology Systems, 24*(1), 31-53.

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications.* New York, NY: John Wiley.

Tackett, R. (1995). Pain worth the gain? *Network World, 42*(12).

Tan, B.C.Y., Wei, K.K., Watson, R.T., Clapper, D.L., & McLean, E.R. (1998). Computer-mediated communication and majority influence: Assessing the impact in an individualistic and a collectivistic culture. *Management Science, 44*(9), 1263-1278.

Tan, W.P., & Chan, H.C. (1998). A TAM-based assessment of videoconferencing for remote tutoring. *Proceedings of the Fourth Annual Association for Information Systems Americas Conference* (pp. 1094-1096).

Trevino, L.K., Lengel, R.H., & Daft, R.L. (1987). Media symbolism, media richness, and media choices: A symbolic interactionist perspective. *Communication Research, 14*(5), 533-575.

Yin, R.K. (2002). *Case study research: Design and methods* (3rd ed.). Newbury Park, CA: Sage.

KEY TERMS

Interactivity: The level of interaction among communication partners.

Media Richness: Media richness is the extent to which a communication medium can facilitate shared understanding among the communicating partners.

Mode Characteristics: Mode characteristics refer to the characteristics of a mode of instruction. The relevant characteristics identified for this study are: usefulness, challenge, attractiveness and clarity.

Microsoft NetMeeting: "NetMeeting lets you hold video conference calls, send text messages, collaborate on shared documents, and draw on an electronic whiteboard over the Internet or an intranet." (www.microsoft.com)

One-to-One Tutoring: Unlike a classroom setting with one instructor and many students, this is tuition (teaching-learning interaction) between one tutor and one student.

Social Presence: In this study, social presence refers to the extent to which a communication medium projects the physical presence of the communicating partners.

Video-Conferencing: Video-conferencing is the use of information and communications technology to allow people at different locations to hear, see and speak to one another. It often includes the sharing of documents and textual communications.

Online Academic Libraries and Distance Learning

Merilyn Burke
University of South Florida, USA

Bruce Lubotsky Levin
University of South Florida, USA

Ardis Hanson
University of South Florida, USA

BRIEF HISTORY OF DISTANCE LEARNING

Historically, distance learning or distance education began as little more than "correspondence courses," which promised an education in one's own home as early as 1728 (Distance Learning, 2002). By the 1800s the concept of distance education could be found in England, Germany and Japan (ASHE Reader on Distance Education, 2002).

In 1933, the world's first educational television programs were broadcast from the University of Iowa and in 1982, teleconferencing began (Oregon Community Colleges for Distance Learning, 1997), often using videotaped lectures, taped-for-television programs and live programming, adding a human dimension. Students and faculty were now able to interact with each other in real time, enhancing the learning process by allowing student access to teachers across distances.

ACADEMIC DISTANCE LEARNING & THE VIRTUAL LIBRARY

Distance learning can be defined by the fact that the student and the instructor are separated by space. The issue of time is moot considering the technologies that have evolved allowing real-time access. Today, universities around the world use various methods of reaching their remote students. With the use of technology, access becomes possible, whether it is from campuses to remote sites, or to individuals located in their own homes.

The development of course instruction, delivered through a variety of distance learning methods (e.g., including Web-based synchronous and asynchronous communication, e-mail, and audio/video technology), has attracted major university participation (Burke, Levin & Hanson, 2003). These electronic learning environment initiatives increase the number of courses and undergraduate/graduate degree programs being offered without increasing the need for additional facilities.

During the 2000-2001 academic year, the NCES (National Center for Education Statistics) estimated in the United States alone there were 3,077,000 enrollments in all distance education courses offered by 2-year and 4-year institutions, with an estimated 2,876,000 enrollments in college-level, credit-granting distance education courses, with 82% of these at the undergraduate level (Watts, Lewis & Greene, 2003, p. iv). Further, the NCES reported that 55% of all 2-year and 4-year U.S. institutions offered college-level, credit-granting distance education courses, with 48% of all institutions offering undergraduate courses, and 22% of all institutions at the graduate level (ibid, p. 4). It is clear that distance education has become an increasingly important component in many colleges and universities, not only in the United States, but also worldwide.

Although educational institutions create courses and programs for distance learners, they often omit the support component that librarians and accrediting organizations consider critical. It is recommended that courses be designed to ensure that students have "reasonable and adequate access to the range of student services appropriate to support their learning" (WICHE, Western Interstate Commission for Higher Education). Further, courses should incorporate information literacy skills within the course or in class assignments to ensure skills for lifelong learning (American Library Association, 1989; Bruce, 1997).

Distance learning (DL) students are unlikely to walk into the university's library for instruction on how to use the resources, from print to electronic journals, as well as services such as electronic reserves and interlibrary loan. The elements of any successful distance-learning program must include consideration of the instructors and the students, both of whom have needs that must be examined and served.

With imaginative use of technology, libraries have created "chat" sessions, which allow 24/7 access to librarians who direct students to the resources that are available online or through interlibrary loan. In addition, librarians assist faculty in placing materials on electronic reserve so that their students can access the materials as needed. Libraries have become more willing to provide mail services and desk top delivery of electronic articles to their distance learning students and, when that is not possible, refer their students to local libraries to take advantage of the interlibrary loan system. Online tutorials have been created to help students learn how to access these resources, while other libraries have specific departments that assist their distance education students and faculty. The role of the library in this process is one of support, both for the students and the faculty.

CHANGES IN DISTANCE LIBRARIANSHIP

Of all of the "traditional" library functions, such as materials provision, electronic resources, and reciprocal borrowing available to the distance learner, there remains a significant gap in service, that of reference. Although chat lines and other 24/7 services are available, these services simply do not provide the DL student the same quality of service that the on-campus student gets when he or she consults with a librarian in person. Newer versions of distance learning course software provide external links to resources, but do not yet include reference service by e-mail and live chat sessions in their basic packages. It will be the responsibility of the library to make these services easily available and known to the distant learner, whose contact to the institution may not include information about the library and its resources. Proactive planning by the library with those who are responsible for distance education can ensure that the students are made aware of what is available for them in the library.

Recently, libraries have been looking at e-commerce business models as a functional way to serve their clientele in reference services, as today's "customers" are savvier, and businesses have become more sophisticated in responding to customers' needs. Libraries can use these models to provide the services for DLs whose level of skills has risen with the increased use of the Internet. Coffman (2001) discusses the adaptation of such business tools as customer relations management (CRM) software, such as the Virtual Reference Desk, Webline, NetAgent, and LivePerson. These programs are based upon the "call center model," which can queue and route Web queries to the next available librarian. A quick visit to the LSSI Web site (Library Systems and Services, L.L.C, *http://www.lssi.com*) allows a look into the philosophy of offering "live, real-time reference services". LSSI's "Virtual Reference Desk" allows librarians to "push" Web pages to their patrons' browser, escort patrons around the Web and search databases together, all while communicating with them by chat or phone (*www.lssi.com*). Many of these systems provide the capability to build a "knowledge base" that can track and handle a diverse range and volume of questions. These collaborative efforts, with a multitude of libraries inputting the questions asked of them and creating FAQs (frequently asked questions lists), provide another level of service for the distance learner (Wells & Hanson, 2003).

These systems have great potential, and while they show tremendous possibilities, they need more work to make them more functional for library use. Chat sessions are problematic when the patron is using his or her phone line to connect to the computer, and libraries must look to the emerging technology to find solutions to such issues to prevent becoming obsolete.

Another direction is the development of "virtual reference centers," which would not necessarily have to be located in any particular physical library. Current collaboratives among universities have created consortial reference centers accessible anywhere and anytime. The reference center librarian could direct the student to the nearest physical resource or to an online full-text database based upon the student's educational profile (e.g., university, student status, and geographic location). Although the physical library may indeed become a repository for books and physical items, the reference component may no longer be housed within that particular building.

An example of support is Toronto's Ryerson Polytechnic University (Lowe & Malinski, 2000) infrastructure, which is based upon the concept that, in order to provide effective distance education programs and resources, there must be a high level of cooperation between the university, the departments involved, and the library. At Ryerson, the Continuing Education Department studied what types of support the students needed and identified technical, administrative, and academic help as three major areas of concern. Technical help was assigned to the university's computing services; administrative help was available on the Web and through telephone access, and academic help included writing centers, study skill programs, and library services. Ryerson's philosophy encompassed the concept that synchronization of all these components would assist in making the student's experience richer and give the student a higher degree of success

The library and the distance education unit worked to provide connectivity to resources that were important to the classes being taught online or at-a-distance. It is these types of library involvement that can make distance learning an even more successful and enriching experience. When a university system, as a whole, embraces a collaboration of all its components, both the students and the university reap the rewards.

CONCLUSIONS

Distance education will only continue to grow. In order to support this educational initiative, academic libraries must establish a supporting framework and commitment to those services traditionally provided by libraries such as lending books and answering reference questions in person or by telephone, plus new services such as "live chat" and desk top delivery of articles that are unique to the virtual environment. Faculty and students in distance learning courses depend on the academic library for their resources and services, and the library must be able to deliver materials to students or assist them in finding alternate sources in a timely manner. Libraries need to be able to identify their DL students using the necessary resources to verify information. Help desks, chat rooms, e-mail programs, and live reference all contribute to the support of the distance learning programs. Since DL students may never visit a library's physical facility, it is important to provide information on how best to access the library virtually.

Faculty members also require library support for their courses. For example, materials may be scanned and placed on the Web or videos may be "streamed" for online access. In order to digitize and make these items accessible, faculty need information on the correct use of copyrighted materials. It is also advisable to put into place an action plan to implement a program for distance learning and a method for assessing that program once it is in place.

FUTURE TRENDS

As distance learning continues to flourish, research will be needed to examine the effective implementation and ongoing management of distance education. While several issues emerge as salient, such as the social aspects of communication in the networked environment, and the integrity of Web-based course resources, it is the role of libraries in support of distance education that must be considered. Although much has been written about the social isolation of distance work, recent advances in groupware technologies have enhanced an individual's ability to stay connected for both work and social exchange through the use of synchronous and asynchronous remote communication (Li, 1998; Watson, Fritz et al., 1998). However, the increased use of technology suggests that formal and extensive training on both distance technology and team communications are necessary (Venkatesh & Speier, 2000).

Libraries, often overlooked in this process, have to be far more assertive in the distance learning process. Libraries can be a center of technical and administrative help along with the traditional academic role that they have normally held. The growing DL field allows librarians to redefine their roles, and request monies for advanced technological necessary to become as "virtual" as the classes being taught. In addition, to serve the ever-increasing DL population, library education must now include the course work that will provide future librarians the training necessary to serve this ever-expanding population.

REFERENCES

American Library Association. (1989). Presidential Committee on Information Literacy. *Final Report.* Chicago: American Library Association.

Bruce, C. (1997). *Seven faces of information literacy*. Adelaide, South Australia: AUSLIB Press.

Burke, M., Levin, B.L., & Hanson, A. (2003). Distance learning. In A. Hanson & B.L. Levin (Eds.), *The building of a virtual library* (pp.148-163). Hershey, PA: Idea Group Publishing.

Coffman, S. (2001). Distance education and virtual reference: Where are we headed? *Computers in Libraries, 21*(4), 20.

Distance Learning. (2002). 1728 advertisement for correspondence course. Retrieved March 8, 2002, from *http://distancelearn.about.com/library/timeline/bl1728.htm*

Li, F. (1998). Team-telework and the new geographical flexibility for information workers. In M. Igbaria & M. Tan (Eds), *The virtual workplace* (pp. 301-318). Hershey, PA: Idea Group Publishing.

Lowe, W., & Malinksi, R. (2000). Distance learning: Success requires support. *Education Libraries, 24*(2/3), 15-17.

Oregon Community Colleges for Distance Learning. (1997). The strategic plan of the Oregon Community Colleges for Distance Learning, distance learning history, current status, and trends. Retrieved March 8, 2003, from *http://www.lbcc.cc.or.us/spoccde/dehist.html*

Venkatesh, V., & Speier, C. (2000). Creating an effective training environment for enhancing telework. *Interna-*

tional Journal of Human Computer Studies, 52(6), 991-1005.

Watts, T., Lewis, L., & Greene, B. (2003). Distance education at degree-granting postsecondary institutions: 2000–2001. Washington, D.C.: National Center for Education Statistics. *http://nces.ed.gov/pubs2003/2003017.pdf*

Wells, A.T., & Hanson, A. (2003). E-reference. In A. Hanson & B.L. Levin (Eds.), *The building of a virtual library* (pp.95-120). Hershey, PA: Idea Group Publishing.

WICHE (Western Cooperative for Educational Telecommunications). (n.d.). *Balancing quality and access: Reducing state policy barriers to electronically delivered higher education programs*. Retrieved September 2, 2003, from *http://www.wcet.info/projects/balancing/principles.asp*

KEY TERMS

Chat: A real-time conferencing capability, which uses text by typing on the keyboard, not speaking. Generally between two or more users on a local area network (LAN), on the Internet, or via a bulletin board service (BBS).

CRM (Customer Relationship Management): This term refers to how a company interacts with its customers, gathers information about them (needs, preferences, past transactions), and shares these data within marketing, sales, and service functions.

Desk Top Delivery: Using electronic formats to send articles to users.

Distance Learning/Distance Education: Taking courses by teleconferencing or using the Internet (together with e-mail) as the primary method of communication.

Electronic Reserves: The electronic storage and transmission of course-related information distributed by local area networks (LANs) or the Internet. Also known as e-reserves; in addition to displaying items on a screen, printing to paper and saving to disk are often allowed.

Internet: A worldwide information network connecting millions of computers. Also called the Net.

Link-rot: The name given to a link that leads to a Web page or site that has either moved or no longer exists.

Next Generation Internet (NGI): Currently known as Abilene, the next generation Internet refers to the next level of protocols developed for bandwidth capacity, quality of service (QOS), and resource utilization.

Real-time: Communication that is simultaneous; see **synchronous.**

Social Aspects of Communication: A social process using language as a means of transferring information from one person to another, the generation of knowledge among individuals or groups, and creating relationships among persons.

Streaming Video: A technique for transferring data as a steady and continuous stream. A browser or plug-in can start displaying the data before the entire file has been transmitted.

Synchronous and Asynchronous Communication: Synchronous communication is when messages are exchanged during the same time interval (e.g., Instant Messenger™).

Asynchronous communication is when messages are exchanged during different time intervals (e.g., e-mail).

Virtual Library: More than just a means of collocating electronic resources (full-text materials, databases, media, and catalogues), a virtual library also provides user assistance services, such as reference, interlibrary loan, technical assistance, and so forth.

Voice over Internet Protocol (VoIP): A protocol that enables people to use the Internet as the transmission medium for telephone calls.

Web (World Wide Web): A global system of networks that allows transmission of images, documents, and multimedia using the Internet.

Online Communities and Community Building

Martin C. Kindsmüller
Berlin University of Technology, Germany

Sandro Leuchter
Berlin University of Technology, Germany

Leon Urbas
Berlin University of Technology, Germany

INTRODUCTION

"Online community" is one of today's buzzwords. Even though superficially it is not hard to understand, the term has become somewhat vague while being extensively used within the e-commerce business. Within this article, we refer to online community as being a voluntary group of users who partake actively in a certain computer-mediated service. The term "online community" is preferred over the term "virtual community," as it denotes the character of the community more accurately: community members are interacting online as opposed to face to face. Furthermore, the term "virtual community" seems too unspecific, because it includes other communities that only exist virtually, whereas an online community in our definition is always a real community in the sense that community members know that they are part of the community.

Nevertheless, there are other reasonable definitions of online community. An early and most influencing characterization (which unfortunately utilizes the term "virtual community") was coined by Howard Rheingold (1994), who wrote: "…virtual communities are cultural aggregations that emerge when enough people bump into each other often enough in cyberspace. A virtual community is a group of people […] who exchanges words and ideas through the mediation of computer bulletin boards and networks" (p. 57). A more elaborated and technical definition of online community was given by Jenny Preece (2000), which since then, has been a benchmark for developers. She stated that an online community consists of four basic constituents (Preece, 2000, p. 3):

1. Socially interacting people striving to satisfy their own needs.
2. A shared purpose, such as interest or need that provides a reason to cooperate.
3. Policies in the form of tacit assumptions, rituals, or rules that guide the community members' behavior.
4. A technical system that works as a carrier that mediates social interaction.

Not explicitly mentioned in this characterization but nevertheless crucial for our aforementioned definition (and not in opposition to Preece's position) is voluntary engagement.

BACKGROUND

Just because everybody is now talking about them, online communities are, historically seen, neither an implication of the World Wide Web — which dates back to 1991 (Berners-Lee et al., 1992) — nor dependent on the Internet as a transport infrastructure. In fact, online communities emerged at times when ARPAnet—the predecessor of the Internet — was still restricted to military-funded institutions. They were based on computerized bulletin boards first introduced by Christensen and Suess (1978). Their system was called CBBS (computerized bulletin board system) and followed the idea of a thumbtack bulletin board hosted electronically on a computer. Other computer hobbyists were able to connect with their home computers via a dial-up modem connection and could "pin" messages to a shared "board." The first online communities developed through other participants responding to those messages, creating ongoing discussions. At that time, computer hobbyists and scientists were more or less the only ones who owned computers and modems. Therefore, most topics on CBBS were within the realm of computers, but in the long run, the discussions broaden. Within the 1980s, similar systems appeared that were now subsumed as BBS (bulletin board system). The most well known were "The Well" (Whole Earth 'Lectronic Link) and FidoNet (Rheingold, 2000).

Apparently, at the very same time, the technological and social environment was ready for online communities, as there were at least two other independent developments concerning this matter:

1. The Usenet was invented by computer science students at Duke University and the University of North Carolina, using a simple scheme by which these two computer communities could automatically exchange information via modems at regular intervals.
2. The first MUDs appeared at the University of Essex (UK) creating playful and imaginative online communities. MUDs (short for Multi-User Dungeon/Dimension/Domain) are computer-implemented versions of text-based role-playing games, in which multiple persons can take virtual identities and interact with one another. Early MUDs were adventure games played within old castles with hidden rooms, trapdoors, etc.

Nowadays, most online communities are using the Internet as carrier, and most of them are Web based, using HTTP as a protocol for transportation and the DHTML standard for presentation. But there are still communities that employ other systems and protocols, like newsreaders using NNTP and mail-groups using SMTP or IRC (Internet relay chat) based chatting systems (IRC). Some online communities even use multiple systems and protocols to communicate and cooperate.

ONLINE COMMUNITIES

The conditions in pure online communities highly differ from a computer-mediated communication situation within a company. Whereas employees in a computer-supported cooperative work (CSCW) context usually meet online as well as face-to-face, members of online communities have, as a general rule, never met each other. Working in a highly standardized company context, employees have to focus on task fulfillment within a certain time frame. Superiors evaluate their achievements, and they are accordingly paid by the company. Online communities thrive on volunteers. Usually none of the community members can be forced to do something, and there are no tangible incentives. Basic research in motivation psychology (Franken, 2001) even shows that incentives tend to be counterproductive.

Community members usually show a high degree of intrinsic motivation to participate actively in the development of an online community. It is still open to discussion where this motivation comes from. Simple rules like "It's all based on trying to maximize the potential personal benefit" seem to fail, as long as one has a simplistic concept of the term "personal benefit." As the attention-economy-debate (i.e., Aigrain, 1997; Ghosh, 1997; Goldhaber, 1997) shows that personal benefit is a complex entity if one relates it to online activities in the World Wide Web.

The likelihood of taking an active part in a community increases with the potential personal benefit that could be gained within that community. This is directly related to the quality of the contents offered. As, e.g., Utz (2000) stated, the likelihood of submitting high quality contributions increases with the quality and the manifoldness of the existing entries. Appropriate solutions of these quality assurance aspects are rating systems.

A "killer-feature" for such an application generates immediate benefit for a user as soon as he or she contributes to the community, even without anybody else contributing. In addition to such a feature, or even as a partial replacement, one can follow best practices. After analyzing numerous well-working online communities, Kollock (1999) came to the conclusion that there are basically two states of motivation: self-interest (what seems to be the common motivation found) and altruism. Self-interest as a motivational state is linked to expectation of reciprocity: people are willing to help or cooperate with others if they can expect a future quid pro quo.

A widely discussed issue in the context of community building is the so-called public goods dilemma: if people can access public goods without restriction, they tend to benefit from these goods and, therefore, from others' contributions without contributing in the same way. If, on the other hand, most members of a community are led into temptation, the public good will vanish (Kollock & Smith, 1996). The main problem is to keep the balance between the individual and common interest: an individually favorable and reasonable behavior turns out to be harmful for the others, and in the long run, disastrous for the community (Axelrod, 1984; Ostrom, 1990).

Owing to these circumstances, it is not surprising that a great deal of all online community building projects fail, even though much effort has been put into these projects due to the high profit opportunities within the field as, for instance, Hagel and Armstrong (1997) predicted.

ONLINE COMMUNITY BUILDING

Recipe-based fabrication of online communities is, at least, a bold venture if not an illusionary enterprise. Social relationships and group momentum are particularly hard to predict. As Rheingold (2000) explicated, online communities grow organically and tend to follow their own rules. Therefore, controlling efforts always have to be adjusted to the current group context. Nevertheless, some well-approved principles could be derived from findings that were discussed in the last paragraph.

According to Kollock (1999), cooperation within an online community can only be successful if individuals:

1. Can recognize each other, i.e., they are not operating anonymously within the community.
2. Have access to each others interaction history.
3. Share the presumption of a high likelihood of a future encounter within the online community.

This leads to the conclusion that online communities have to offer possibilities of creating and managing relationships by supporting continuous interaction between their members. Therefore, it is advantageous if the system has a memory, in the sense that every community member and every item stored in the system holds a personal history.

People tend to act from self-interest if they are aware that their actions have effects on their reputations: high-quality contributions, impressive knowledge, and the perception of being willing to help others enhance the prestige of the community member. Although altruism as a motivational state for taking part in an online community is less common in comparison with self-interest, it is still frequent enough to be addressed if one thinks about community building. People with altruistic motivation try to meet the needs of the group or certain group members. This motivational state can be satisfied by establishing a public space where these needs can be stated, communicated, and discussed.

Considering the public goods dilemma, it is essential to introduce a role concept to clearly communicate the borderline between being in a group and out of a group. To get full access to all group resources, one has to join the group. Several functionalities are only accessible for registered and authorized users. The commitment that is required to join the group leads to a comprehensible demarcation between members and nonmembers, who, in turn, facilitate the togetherness of the group and the identification of the members within the group. Three further selective measures address the public goods dilemma: personal presence, personal responsibility, and personal history. Anonymity and lack of continuity among the members promotes egoistic behavior. Therefore, modifying actions should be tagged with users' login names, which, in turn, should be associated with records of personal data. Tagging entries and actions with user login names makes it easy to recognize people and enhances the constitution of personal relationships and online cooperation among the community members. Seeing all modifying actions supports the actors' personal responsibility. If the system has a memory of every action, this gives a personal history to every community member, as well as to the community's artifacts. Information about past interactions of the members again increases personal responsibility, whereas information about interaction with the artifacts facilitates getting up-to-date within a new area of interest and familiarizing new members with the online community's etiquette and who-is-who.

"Content is king" is commonplace for virtually all Web-based efforts. This is notably true for online communities operating on a user-as-editors base. To implement a reasonable quality assurance system, it is crucial to apply technical as well as social means. Technically, this can be done by employing a content rating system. Employing a "tiger team" of highly motivated volunteers can, on the other hand, help the online community to start up by delivering good content. For all content producers, it has to be as easy as possible to feed new content into the system. The best way of avoiding barriers is through continuous usability testing.

Introducing dedicated and active moderators seems to be the most important step to nourish motivation of the community members. Moderators can enhance group activities and increase the efficiency of the group. They are responsible for communicating the group standards (etiquette), acting as confessors for new community members, and helping in preserving continuity. Rojo (1995) showed that an active moderator can, to some extent, compensate for the lack of active members in an online community.

As opposed to face-to-face communities, where only members can start a conversation, online communities can initiate communication. This opportunity should be brought into play by implementing change awareness functions as software agents that collect relevant information for users and present it in e-mails and personalized portal pages. An important item in using profiles for personalized services is keeping profiles up-to-date. Experience shows that interests continuously change over time. Profile setting dialogues are often accessed once and then forgotten. Thus, there is risk for personalized services to decrease subjectively in quality over time. Hence, it is important to monitor user behavior and let the agents ask from time to time if interpretations of observations of, for example, changing interests, are correct.

The open-source movement has become very successful in recruiting new developers who start their own projects or join existing software development efforts. Today, most if not all open-source software development communities use online services for cooperation. In the remainder of this section, the application of the requirements for software supporting online community building is demonstrated with examples of open-source software development communities:

- **"Killer-feature":** Open-source projects are often founded to solve an existing problem, i.e., the killer-feature is the product of the online community. When others join the project, the work becomes even more effective.

- **Recruitment:** Open-source communities produce software systems that are not only intended for use by themselves but also for external clients. These external users, i.e., users not actively involved in the open-source online community, can be made active developers that modify the source and give it back to the project. To foster this process of developer recruitment, online communities should provide transparent rules for becoming actively involved. The projects that do not seem to be hermetic have better chances of growing their developer base.
- **Transparency:** The most important possibility of gaining transparency is through a public archive of the project's mailing lists. Because it is often not easy to scan large e-mail archives, open-source communities should provide text documenting guidelines and standards.
- **Policy:** The Debian community that maintains an open-source Linux distribution is a good example of a growing community that has given itself a substantial set of roles, rules, and guidelines. They achieve transparency of their standards by publication of documents on their Web server — other projects, such as UserLinux, use a Wiki for that purpose which makes such standards more vivid and activates community members' attendance.
- **Trust:** Debian has a twofold quality assurance system. There are no anonymous additions to the system, and approved maintainers electronically sign all modifications. Bugs are reported by all users. The bug lists are available to the public.
- **Cooperation and usability:** Cvs is a configuration management software for source code trees in distributed development teams. cvs is an example of cooperation software and its usability issue: Based on rcs (revisions management for single files), it is efficient to use for everyday tasks in open-source software development.
- **Awareness:** Workflows can provide awareness. Examples include automated e-mail distribution of users' bug reports and e-mail notifications of cvs commits.

This exploration into open-source projects and their online coordination and cooperation tools reveals that a voluntary community approach works, and the infrastructures and supporting tools of these projects can be taken as a best practice reference case.

FUTURE TRENDS

Recently, the term "socialware" was proposed for software systems dedicated to enhance social relations. According to Hattori et al. (1999), socialware are systems which aim to support various social activities by means of computer networks. This is done by linking people with others, fostering communication within a community and connecting the community's information. Initially intended for CSCW systems that are used by stable communities, the socialware approach seems suitable for implementing software for online communities as well. It uses rules of interpersonal communication and transfers these structures into software. The technical concept associated with socialware is a multiagent system architecture. The CSCW functionality is achieved through coordination and cooperation of a distributed set of software entities (agents). Users of a community system have personal agents for gathering and exchanging information, visualizing contexts, and supporting decisions. Personal agents and the users they belong to are seen as personal units. Personal units interact with community agents that have the function of providing shared information and mediating communication between personal units. This approach also makes it possible to link several partially overlapping online communities.

A current development in online communities is the transformation of the virtuality of computer networks into the real world. There are different enabling technologies for mobile and ad hoc communities. An important factor is localizability in cellular phone networks or with global positioning systems (GPSs). Using the localization information as part of the application context allows for mobile communities. They are often based on asynchronous communication, like Internet online communities. An example for such a mobile community is the petrol station price comparison community. In 2000, the German Research Center for Information Technology offered car drivers a location aware service for obtaining the locations of petrol stations together with their prices, in exchange for other stations' current petrol prices.

The availability of new short-range radio networking technologies, such as Bluetooth or wireless LAN, enables new synchronous mobile communities. This gives users the ability to connect devices ad hoc (i.e., without a server infrastructure), permitting mobility and interaction. As with other Internet online communities, besides professional applications such as disaster management (Meissner et al., 2002), game playing is an important technology driver, e.g., pervasive group games are being developed (Pennanen & Keinänen, 2004) that could build up social structures in some ways comparable to online communities.

CONCLUSION

Advanced software solutions like the aforementioned socialware approach can help to build and maintain stable online communities. In the long run, though, it is not the technology, it is the people that make an online community work. The most advanced technology is neither sufficient nor, as early BBS/MUD approaches show, necessary for stable online communities. People will always make creative use of technology by using it in other ways than were originally intended by the designers. This will, once in a while, generate possibilities for new online communities.

Nevertheless, the most important factor for successful online communities is providing awareness about changes in the communities' databases to members. Awareness functions provide an understanding of the others' activities and the communities' goals and progress to relate and evaluate the users' own activities accordingly (Dourish & Bellotti, 1992).

REFERENCES

Aigrain, P. (1997). Attention, media, value and economics. *First Monday, 2*(9). Retrieved March 15, 2004, from http://www.firstmonday.dk/issues/issue2_9/aigrain/

Axelrod, R. (1984). *The evolution of cooperation*. New York: Basic Books.

Berners-Lee, T. J., Cailliau, R., Groff, J. -F., & Pollermann, B. (1992). World-Wide Web: The information universe. *Electronic Networking: Research, Applications and Policy, 2*(1), 52–58.

Christensen, W., & Suess, R. (1978). Hobbyist computerized bulletin board. *Byte Magazine, 3*(11), 150–158.

Dourish, P., & Bellotti, V. (1992). Awareness and coordination in shared work spaces. *In Proceedings ACM Conference on Computer-Supported Cooperative Work CSCW'92* (pp. 107–114), Toronto, Canada.

Franken, R. E. (2001). *Human motivation* (5th ed.). Pacific Grove, CA: Brooks/Cole.

Ghosh, R. A. (1997). Economics is dead. Long live economics! A commentary on Michael Goldhaber's "The Attention Economy." *First Monday, 2*(5). Retrieved March 15, 2004, from http://www.firstmonday.dk/issues/issue2_5/ghosh/

Goldhaber, M. H. (1997). The attention economy and the Net. *First Monday, 2*(4). Retrieved March 15, 2004, from http://www.firstmonday.dk/issues/issue2_4/goldhaber/

Hagel, J., & Armstrong, A. G. (1997). *Net gain: Expanding markets through virtual communities*. Boston, MA: Harvard Business School Press.

Hattori, F., Ohguro, T., Yokoo, M., Matsubara, S., & Yoshida, S. (1999). Socialware: Multi-agent systems for supporting network communities. *Communication of the ACM, 42*(3), 55–61.

Kollock, P. (1999). The economies of online cooperation. Gifts and public goods in cyberspace. In M. A. Smith & P. Kollock (Eds.), *Communities in cyberspace*. London, UK: Routledge.

Kollok, P., & Smith, M. A. (1996). Managing the virtual commons: Cooperation and conflict in computer communities. In S. Herring (Hrsg.), *Computer-mediated communication: Linguistic, social, and cross-cultural perspectives* (pp. 109–128). Amsterdam, The Netherlands: John Benjamins.

Meissner, A., Luckenbach, T., Risse, T., Kirste, T., & Kirchner, H. (2002). Design challenges for an integrated disaster management communication and information system. *First IEEE Workshop on Disaster Recovery Networks (DIREN 2002)*, June 24, 2002, New York City. Retrieved October, 04, 2004 from: *http://comet.columbia.edu/~aurel/workshops/diren02/IEEE_DIREN2002_Meissner_DesignChallenges.pdf*

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. New York: Cambridge University Press.

Pennanen, M., & Keinänen, K. (2004). Mobile gaming with peer-to-peer facilities. *ERCIM News, 57*, 31–32.

Preece, J. (2000). *Online communities: Designing usability and supporting sociability*. Chichester, UK: John Wiley & Sons.

Rheingold, H. (1994). A slice of life in my virtual community. In L. M. Harasim (Ed.), *Global networks: Computers and international communication* (pp. 57–80). Cambridge, MA: MIT Press.

Rheingold, H. (2000). *The virtual community: Homesteading on the electronic frontier* (revised edition). Cambridge, MA: MIT Press.

Rojo, A. (1995). *Participation in scholarly electronic forums*. Unpublished Ph.D. thesis, Ontario Institute for Studies in Education, University of Toronto, Canada. Retrieved March 14, 2004, from http://www.digitaltempo.com/e-forums/tabcont.html

Utz, S. (2000). Identifikation mit virtuellen Arbeitsgruppen und Organisationen. In M. Boos, K. J. Jonas, & K.

Sassenberg (Eds.), *Computervermittelte Kommunikation in Organisationen*. Göttingen: Hogrefe.

KEY TERMS

Community Building: All activities related to building and maintaining online communities.

CSCW (Computer-Supported Cooperative Work): Software tools and technology as well as organizational structures that support groups of people (typically from different sites) working together on a common project.

Online Community: An online community is a voluntary group of active users that partake actively in a certain computer-mediated service.

Socialware: Socialware aims to support various social activities on a network. Rules of interpersonal communication are used and transferred into community software.

UaE (User-As-Editors) Approach: The community members are responsible for supplying new content and for the quality assurance of existing content, as well as for creating and maintaining the etiquette of the community.

Virtual Community: A featureless and, therefore, often misleading term usually regarded as synonymous to online community. The term "online community" is preferable, as it denotes the character of the community more accurately.

Wiki: Internet service based on HTTP and HTML providing "open editing" of Web pages with a Web browser. Hyperlinks between documents are supported with simple textual references. By default, everybody is allowed to edit all available pages.

Online Learning as a Form of Accommodation

Terence Cavanaugh
University of North Florida, USA

INTRODUCTION

An estimated three billion people, representing approximately half of the planet's population, are in some way affected by disabilities, which includes an estimated 150 million from the United States of America (Half the Planet, 2001). According to the *Twenty-Third Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act* (U.S. Department of Education, 2002a), concerning students with special needs between the ages of three and 21, the U.S. and its outlying areas are currently serving educationally more than 6,272,000 students classified as having a disability. The inclusion model, in which a special needs student participates in the "regular" classroom, has become the current classroom education standard. Today's special needs students have increasing impacts on the general education teacher as, during the past 10 years, the percentage of students with disabilities served in schools and classes with their non-disabled peers has gradually grown to over 90% in 1998 (U.S. Department of Education, 2000b). Because of the large and increasing number of special needs students, assistive educational technology is growing in importance. The population of postsecondary students with disabilities has increased over the past two decades, and currently there are approximately one million persons in postsecondary institutions who are classified as having some form of disability (U.S. Department of Education, 2000b). In 1994, approximately 45% of the adult population who reported having a disability had either attended some college or had completed a bachelor's degree or higher, as compared to only 29% in 1986 (National Center for Educational Statistics, 1999a).

BACKGROUND

Changes in the Population of Schools

While the makeup of the student population (K-20) has changed, because more students have been classified as having a disability and are now included in the general educational population, so too have the possibilities of the educational setting changed. For the 1999-2000 school year, the number of U.S. students with disabilities served was 588,300 preschool children and 5,683,707 students ages 6 through 21, representing an increase of 2.6% over the previous year (U.S. Department of Education, 2002a). Instructors now have on hand instructional tools that include forms of interactive telecommunication, such as the Internet and two-way video communication, as options for the delivery of instruction. While schools may not have been planning, designing, and creating distance learning courses and programs to meet the needs of students with disabilities, many students' needs were met through such a delivery system nonetheless. Electronic learning in and of itself is a form of instructional accommodation. Additionally, a range of assistive technology can support the student in the distance learning environment. The online class can be an assistive technology tool that students can use who would otherwise not be able to participate in a classroom for physical, health, or other reasons.

The number of students with disabilities is growing in the online education environment. A 1999 Canadian study of students with disabilities attending community colleges and universities found that an overwhelming majority of respondents (95%) indicated that they used a computer in their education situation, to the most noted reason for using the Internet was for doing research (Fichten, Asuncion, Barile, Fossey & De Simone, 2000). Thompson's (1998) summarizing report states that approximately 5% of the undergraduates at Open University of the United Kingdom have disabilities, with their population increasing at a rate of approximately 10% per year. The growth is ascribed to the convenience of home study and the ability of technology to overcome barriers to learning for students with disabilities. According to the U.S. Department of Education's (2002a) National Postsecondary Student Aid Study of 1999-2000, more than 8% of all undergraduates took at least one distance learning course, and 9.9% of those students identified themselves as having some form of disability.

Accommodations or Modifications Needed for Disabled Access

There is a difference between accommodations and modifications for students with special needs. Accommodations are considered to be provisions made in how a student accesses and/or demonstrates learning. The term accommodations focuses on changes in the instruction, or how students are expected to learn, along with changes in methods of assessment that demonstrate or document what has been learned. The use of an accommodation does not change the educational goals, standards, or objectives, the instructional level, or the content, and provides the student with equal access and equal opportunity to demonstrate his or her skills and knowledge (State of Florida, Department of State, 2000). Accommodations assist students in working around the limitations that are related to their disabilities and allow a student with a disability to participate with other students in the general curriculum program. Accommodations can be provided for: instructional methods and materials; assignments and assessments; learning environment; time demands and scheduling; and special communication systems. By comparison a modification is a change in what a student is expected to learn and demonstrate. The use of a modification for a student changes the standard, the instructional level, or the content to be learned by the student (Beech, 2000).

According to the Assistive Technology Education Network (ATEN) of Florida (2000), instructors of any classes that have students with disabilities should provide students with:

- opportunities to interact with others,
- varied models of print use,
- choices—and then wait for the student to respond,
- opportunities to communicate, and
- expectations that students will communicate, this may require the use of an alternate or augmentative form of communication.

Online instruction, especially through asynchronous Internet presentation, provides all of ATEN's requested opportunities for students. In a "traditional" course, a teacher or professor would be in a classroom, with the students sitting at tables or desks, and there would be lectures, demonstrations, possibly videos and slideshows, handouts, and readings. In an online course in an asynchronous course model, these interactions could still take place, but without the limitations of specific time and location (Picciano, 2001). In such a distance learning course, the main interactions between the student and the instructor take place using course Web pages, streaming audio and video, forums, e-mail, and online books. Assistive tools that a student with special needs may require could be more easily applied in the online environment; such assistive tools may include speech-to-text programs, environmental control devices, or assistive hearing devices. Additionally the asynchronous course design allows the students to access the information at the course Web site and learn at a time convenient to them (Barron, 1999). Within online course sections there could be forums or discussions in which students can participate, allowing each and every student the opportunity and appropriate time to develop and share responses, again without the time restrictions of the standard class period.

The Law, the IEP, and Education

Federal laws and their directives charge that each student classified as having any form of disability have an individual education plan (IEP) developed specifically for that student, that assistive technology devices and services must be considered, and that the student must be taught in the least restrictive environment (Individuals with Disabilities Education Act, 1992). The IEP will be developed by a team of people including teachers, administrators, counselors, parents, outside experts (as needed), and often the student. Distance learning can be considered an adapted form of instruction that through the use of telecommunication technology (usually the Internet) allows a student to participate in a class, meeting the classification of assistive technology. While some students with special needs may choose distance learning courses because these courses provide the necessary accommodations or educational modifications that they need in order to function in that form of "classroom," that in and of itself is not enough. It is up to educators to make sure that the accommodations and educational modifications necessary for these students to function in our classrooms exist or can be made available to these students as they need them. The educational charge extends to ensuring that distance learning classes are also accessible. These distance learning courses or situations must be designed, accommodated, or modified to allow students with special needs to be able to effectively participate.

Distance Learning and Students with Disabilities

A recent survey of seven open enrollment distance learning schools (state, public or private, or college/university) that offered Internet-based instruction may indicate trends in the current status of distance learning programs and special needs students. The distance learning population of the responding schools ran from 300 to 5,000 full- or part-time students, with an average of approximately 1,000 students. Most schools indicated that they did not have records or tracking methods for identifying students with disabilities. Schools that did identify these students indicated that special needs populations ran between 2% and 10%. With the exception of the responding university school, all the K12 distance learning schools indicated that their teachers have participated in IEPs for students. Half of the respondent schools indicated that they currently had students taking courses as part of the student's IEP, as recommended or required by the student's home school IEP team. The schools also indicated that they did not participate as IEP team members, but that the school or distance learning environment was written in as a service in the student's IEP. A consistent thread in the responses was that the distance education schools were sure that they had special needs students, but that they were not identified. When identifying their accommodations for students with special needs, all of the schools responded that they were willing to make accommodations, and the most common accommodation occurring in their programs was extending required time on tests and assignments. When questioned about the virtual school's faculty, only half of the respondents indicated that they had both counselors and exceptional education teachers on staff. Others indicated that they depended on counselors or other support personal to work with the student's home school. Interestingly, all responding virtual schools did indicate that distance learning instructors have already had, or currently have access to, training concerning accommodations for educating special needs students. Only two of the responding distance learning schools indicated that their Web-based instructional pages were compliant with either national or international accessibility guidelines, with one indicating that it was in the process of becoming so.

Hospital/Homebound Students

According to U.S. government statistics, there are more than 26,000 students classified as hospital/homebound students across the nation (U.S. Department of Education, 2002a). How these students are being served at a distance from their "home" school is a distance learning strategy question. The classic hospital/homebound program has a visiting teacher who acts as intermediary between a student's regular teacher and the student. The hospital/homebound teacher contacts the student's classroom teacher or teachers to collect assignments and directions to deliver to the hospital/homebound student, and visits the student to provide instruction and assistance. The more common approaches for hospital/homebound education are the "teleclass" phone model for secondary students and hospital/home visitation for the elementary level. In phone-based or teleclass instruction, all students taking a course dial into a common number, and a teacher provides oral instruction. Students have the opportunity to ask questions and interact through voice (G.A. Ball, personal phone communication, October 2, 2002). This instruction by phone service qualifies as a true distance learning program, as education is provided to the students through a telecommunication system. Other school systems also use online instruction, audio/videotaped instruction, and CD-ROMs, where the districts provide the needed hardware and software; some district school systems have even placed fax machines in students' homes for students to use to receive and submit assignments (C. Bishop, personal e-mail communication, January 2003).

FUTURE TRENDS

Online Education

Distance learning is a growing educational option. The numbers of both public and private distance learning intuitions are on the rise, along with the numbers of students attending. An instructional example would be the Florida Virtual School, going from 227 students in 1998 to 8,200 in 2002, an increase of 3,612% in just five years. School systems and students need flexible options to increase educational success, and using distance learning as an alternative can do that. We are already seeing experiments with asynchronous distance learning being used as an alternative for the hospital/homebound, but research still needs to be done. Virtual schools need to track the special needs students and identify effective methods that are working with them. Hospital/homebound programs need to investigate options for students who must be out of school for extended periods of time and determine the effectiveness of online skill-building education versus asynchronous module instruction of courses. Also, can distance learning assist such programs by allowing small or rural districts to band together and use common certified teachers, who are in short supply and are needed for No Child Left Behind requirements, to be shared between districts? In addition to human concerns, the future brings hardware and software issues, since most of the course management systems (CMSs) are not "disabled" accessible according to US 508 or W3C accessibility guidelines. Schools need to either create their own accessible course systems or apply pressure so that the currently available systems become compliant with access guidelines.

CONCLUSION

Distance Learning as an Accommodation

Electronic learning in and of itself is a form of accommodation, and there is also a range of assistive technology that can support the student in the distance learning environment. The online class can be an effective assistive technology tool that students can use who would otherwise not be able to participate in a classroom for physical, health, or other reasons. While distance learning may not be appropriate for every course or accommodating to every form of disability, it does provide a viable instructional option for many. As the inclusive education of all students occurs more frequently within the standard K12 classroom and through the electronic environment at the college level, it is reasonable to expect that more students with disabilities will participate in online education. While many accommodations are already available online, instructors need to insure that online courses should adhere to accessibility guidelines as proposed by state, national, or international organizations to allow all students access. These educational environments should be designed for all students, even those students who may need modifications, accommodations, and assistive technology.

REFERENCES

Assistive Technology Education Network of Florida (ATEN). (2000). *Assistive technology: Unlocking human potential through technology*. Presentation at the University of South Florida, USA.

Barron, A. (1999). *A teacher's guide to distance learning*. Tampa, FL: Florida Center for Instructional Technology.

Beech, M. (2000). *Accommodations and modifications*: *What parents need to know*. Florida Developmental Disabilities Council, Inc. ESE10753.

Fichten, C.S., Asuncion, J.V., Barile, M., Fossey, M. & De Simone, C. (2000, April). *Access to educational and instructional computer technologies for postsecondary students with disabilities: Lessons from three empirical studies*. EvNet Working Paper. Retrieved from evnet-nt1.mcmaster.ca/network/workingpapers/jemdis/jemdis.htm.

Half the Planet. (2001). *Half the Planet foundation information*. Retrieved from www.halftheplanet.com.

Individuals with Disabilities Education Act. (1992). Pub. L. No. 101-476. Retrieved from frWebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=105_cong_public_la.

National Center for Educational Statistics. (1999a). *Students with disabilities in postsecondary education: A profile of preparation, participation, and outcomes*. Retrieved from nces.ed.gov/pubs99/1999187.pdf.

Picciano, A.G. (2001). Distance learning: Making connections across virtual space and time. Upper Saddle River, NJ: Prentice-Hall.

State of Florida, Department of State. (2000). *Developing quality individual educational plans*. Document ESE9413, Bureau of Instructional Support and Community Services, Florida Department of Education.

Thompson, M.M. (1998). Distance learners in higher education. *Global Distance Education Net*. Retrieved from wbWeb5.worldbank.org/disted/Teaching/Design/kn-02.html.

U.S. Department of Education. (2000b). To assure the free appropriate public education of all children with disabilities. *Twenty-Second Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act*. Retrieved from: www.ed.gov/offices/OSERS/OSEP/Products/OSEP2000AnlRpt/index.html.

U.S. Department of Education. (2002a). *Twenty-Third Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act*. Retrieved from www.ed.gov/offices/OSERS/OSEP/Products/OSEP2001AnlRpt/index.html.

U.S. Department of Education. (2002b). The percentage of undergraduates who took any distance education courses in 1999-2000, and among those who did, the percentage reporting various ways in which the courses were delivered. *1999-2000 National Postsecondary Student Aid Study (NPSAS:2000)*. NEDRC Table Library. Retrieved from nces.ed.gov/surveys/npsas/table_library/tables/npsas22.asp (number of disabled students taking distance learning).

KEY TERMS

Accommodations: Provisions made in how a student accesses and/or demonstrates learning. The term focuses on changes in the instruction, or how students are expected to learn, along with changes in methods of assessment that demonstrate or document what has been learned. The use of an accommodation does not change the educational goals, standards, or objectives, the instructional level, or the content, and provides the student with equal access and equal opportunity to demonstrate his or her skills and knowledge.

Assistive Technology: "…any item, piece of equipment, or product system, whether acquired commercially off the shelf, modified, or customized, that is used to increase, maintain, or improve functional capabilities of individuals with disabilities…." (20 U.S.C. 1401 (33)(250))

Asynchronous: Communications between the student and teacher which do not take place simultaneously.

Disabled Student: From the *U.S. Federal Register*: child/student has been evaluated as having mental retardation, a hearing impairment including deafness, a speech or language impairment, a visual impairment including blindness, serious emotional disturbance (hereafter referred to as emotional disturbance), an orthopedic impairment, autism, traumatic brain injury, an other health impairment, a specific learning disability, deaf-blindness, or multiple disabilities, and who, by reason thereof, needs special education and related services (IEP or 504).

Inclusion: A classroom design where all students should take part and attend "regular" classes. Generally, an ESE and regular education teacher work together with the same group of students, including students with disabilities and general education students. Both of the teachers share the responsibility for all of the students.

Individualized Education Program (IEP): A written statement for each child with a disability that is developed, reviewed, and revised in accordance with this section. (20 U.S.C. 1414 (d)(1)(A)) (Individuals with Disabilities Education Act, 1997)

Modification: A change in what a student is expected to learn and demonstrate. The use of a modification for a student changes the standard, the instructional level, or the content to be learned by the student.

Specific Learning Disability: Term meaning a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, that may manifest itself in an imperfect ability to listen, think, speak, read, write, spell, or to do mathematical calculations, including conditions such as perceptual disabilities, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia.

Teleclass: Voice-only communications linking two or more sites. A standard method used is to connect multiple telephone lines for an audio conference through a phone bridge. A telephone bridge where the conference is established by having all of the distant sites call in to a common bridge telephone number.

Ontology-Based Query Formation and Information Retrieval

Sheng-Uei Guan
National University of Singapore, Singapore

INTRODUCTION

With the introduction of new technologies such as WAP, HSCSD, GPRS, UMTS, and Bluetooth, it is believed that the e-commerce arena will sooner or later merge its applications with handheld devices to create more opportunities for the birth of mobile commerce (m-commerce). However, m-commerce is largely unrealized to date because there still does not exist a single killer application that can attract wireless users to use wireless services. According to a recent survey by Gartner, Inc., besides the importance of coverage of wireless network and pricing issues, the wireless Internet and data services is the next crucial factor that attracts users to use wireless service. As such, there is a need to improve the data services over the wireless network. One of these services is the information retrieval service.

Most electronic product information retrieval systems are still not efficient enough to cater to the increasing needs of customers. This is especially serious in the m-commerce arena, where the bandwidth of mobile devices is low and large data would not be possible. Thus, the discovery of new information retrieval techniques is inevitable.

The main objective of this article is three-fold: 1) to research the use of ontology to assist users in shaping their product enquiries; 2) to study the use of genetic algorithms and agents in query optimization; and 3) to develop information retrieval services for the m-commerce arena. The article proposes a methodology for efficient query formation for product databases and for effective information retrieval systems, which includes the evaluation of retrieved documents to enhance the quality of results that are obtained from product searches.

This article discusses the usage of ontology to create an efficient environment for m-commerce users to form queries. The establishment of a method that combines keyword searches with using ontology to perform query formation tasks further allows a more flexible m-commerce environment for users. Also, with the use of genetic algorithms, it is hoped that query effectiveness can be achieved, at the same time saving computational time.

BACKGROUND

Definition of Ontology

In artificial intelligence, ontology is defined as a design of a conceptualization to be reused across multiple applications (Fensel, 2000; Braga, Werner & Mattosso, 2000; Hendler, 2001). A conceptualization is a set of concepts, relations, objects, and constraints that define a semantic model of some domain of interest. In other words, ontology is like the structure that describes or encodes the conceptualization in any relational aspect (McGuinness, 1998; Karp, 2000).

Literature Survey

In this section, a survey of present query formation methods and information retrieval methods will be discussed.

Unlike in e-commerce, query information using keywords alone in m-commerce is unrealistic, as mobile devices are too small and keypads are not suitable for typing. Moreover, it may be difficult for the user when vocabulary of subject is unfamiliar. Thus, relevance feedback is still the main technique for query modification.

Relevance feedback technique has been investigated for more than 20 years in various information retrieval models, such as the probabilistic model and vector space model (Boughanem, Chrisment & Tamine, 1999; Salton, 1989). It is based on randomly changing the set of query terms, as well as the weights associated with these terms, according to the document retrieved and judged during the initial search.

A lot of research (Boughanem et al., 1999; Yang & Korfhage, 1994; Kraft, Petry, Buckles & Sadasivan, 1994; Kouichi, Taketa & Nunokawa, 1999) has been done on how genetic algorithms (GAs) can be used in information retrieval. One popular approach is query restructuring, which is used to improve the efficiency and effectiveness of the queries formed. GAs actually extend the concepts of relevance feedback. The difference is that genetic algorithms use more than one query and compare the fitness among these queries. The fittest query will survive in the end. Thus, this article focuses on extending the concepts of using genetic algorithms in query restructuring.

Fitness Functions

There are a number of measures of query fitness used in previous works, namely precision and recall retrieved (Kraft et al., 1994; Salton & McGill, 1983), average search length (Losee, 1991), and average maximum parse length (Losee, 1991).

Precision is the percentage of documents retrieved that are relevant, while *recall* measures the percentage of the relevant documents retrieved (Kraft et al., 1994; Salton & McGill, 1983). These two tend to be inversely proportional, so that one is traded for another in most situations. *Average search length* is the average number of documents or text fragments examined in moving down a ranked list of documents until arriving at the average position of a relevant document (Losee, 1988, 1996). Evaluating the performance of a filtering or retrieval process with average search length provides a single number measure of performance. *Average maximum parse length* is the average (over a set of sentences) of the largest number of terms in a parse for each sentence. There are also measures that combine both average search length and average maximum parse length.

Typically, present methods had only dealt with the relevance of the document retrieved. This is reasonable but inefficient, because it is rather difficult to indicate the relevance of a document when the number of documents could be very large. This article measures the relevance of queries instead of documents retrieved. Based on this, efficiency will be improved significantly as the number of queries will be much smaller than the number of documents retrieved, which is ideal for mobile devices.

The Proposed Approaches

Both keyword- and ontology-based approaches have their advantages and disadvantages. Ontology provides the structure, context, and visual aid, while keyword provides a direct search mechanism. Both approaches are relevant for mobile commerce because they save time in browsing and searching, which is very much required by mobile users who are always on the move. Thus, by combining keyword queries with ontology, it is possible to achieve a better and more effective query formation. Before ontology terms are accessed to form the queries, there will be a keyword search to find the required ontology term. For example, "ps2" can be hidden in the node "mouse" when presented in the ontology. The user will not be able to know where "ps2" can be found intuitively without eyeballing the ontology. With the help of keyword search, the term "ps2" can be found easily.

In forming queries, there can be a high chance that the vocabulary used by the user to describe a query does not exactly match the vocabulary used by a query system (Preece, 1999). This will result in getting insufficient information. Therefore, restructuring dealing with domain ontology relationships might be useful. These relationships involve semantic links such as hyponyms and synonyms (Braga et al., 2000). Here, using synonyms is an adequate option to restructure queries because it correctly broadens the scope of search even to the extent of different languages.

When too little information is retrieved, the use of synonym or hyponym might be necessary in order to relax the constraints of the query. However, this approach has a major disadvantage. By relaxing the constraints of a query using synonym or hyponym to increase the number of documents retrieved, one could actually deface the meaning of the original query such that it could drift away from the user's intention. This concern can be alleviated by having user feedback along the process. Also, we have considered relaxing constraints step-by-step. This option can better eliminate the chances of constructing far-fetched queries from the use of genetic algorithms.

MAIN THRUST OF THE ARTICLE

Prototype Design and Implementation

Query Formation Using Ontology

Query formation will be done with the aid of tree ontology. Following the tree path will help form the requirements of a query. This allows forming a query easily. An illustration of the query formation process is shown in Figure. 1. As can be seen from this illustration, using ontology helps the user to save several steps by forming a query using the ontology path that is selected. Thus, it can be claimed that forming queries using ontology is actually more efficient than using keywords.

Combining Keywords and Ontology

The design of parallel combination is rather straightforward. Ontology does not cover everything. Thus, besides having ontology for the user to click on when forming a query, there should be some fields present for the user to fill in. When these fields are being filled in, they can replace the use of ontology either partially or completely. For a serial combination, keywords are used to look for

Figure 1. Illustration of using ontology to form queries

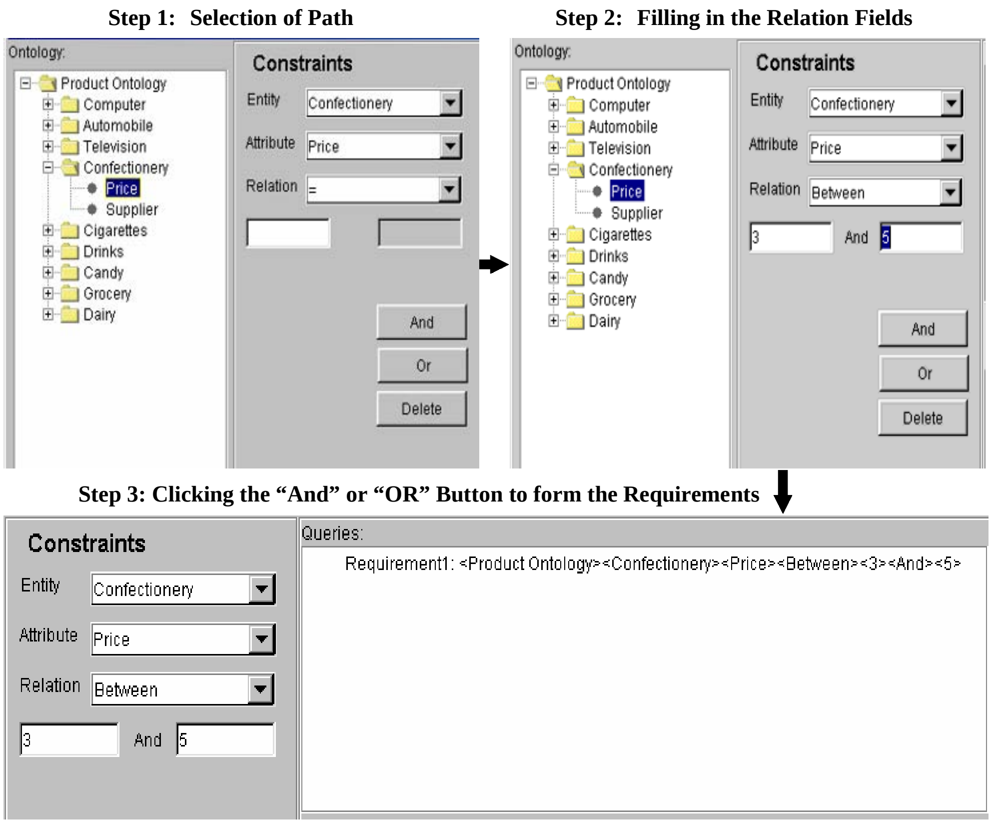
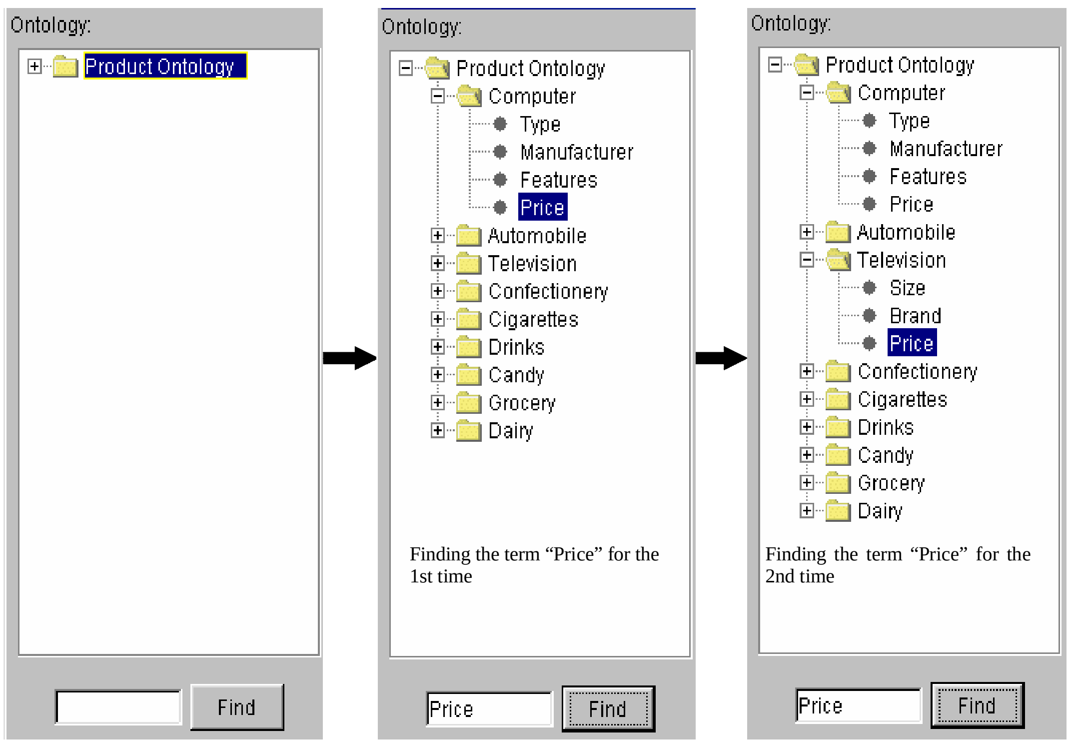


Figure 2. Illustration of the sequence of events for finding ontology terms



ontology terms in the ontology. This is necessary because when the ontology is too large, search for an ontology term by manual clicking becomes difficult. Thus, there would be a field that allows the user to highlight the terms in the ontology itself, as shown in Figure. 2. From this illustration, it can be seen that using keywords to search ontology terms in the ontology creates an efficient environment and context for the user.

Information Retrieval

Using the query formed by the query formation application, an application searches the databases to retrieve information. Intuitively, this application would first do a normal search before allowing the user to proceed with a genetic algorithm search. This is because a genetic algorithm search would definitely take a much longer time than a normal search because of its expensive iterations. The retrieval results are presented to the user; if he is not satisfied, he can then choose to proceed with a genetic algorithm search.

Genetic Algorithm

If the user requests the use of a genetic algorithm, the system will request some input from the user to perform genetic algorithm computation. The system then creates a population of queries from the original query. Basically, genetic algorithm will mutate the queries according to the synonyms of the terms in the ontology.

The Fitness Function

The main concern of using genetic algorithms is the design of the fitness function. In this application, three major elements are used to define the fitness function, namely the fitness of the number of documents retrieved (f_d), the fitness of the average quality of the query results (f_q), and the overall correlation for the query (f_r). The fitness of each chromosome is calculated as follows:

$$\boldsymbol{Fitness} = |f_r \cdot (f_d + f_q)|$$

| . | indicates that the fitness function is normalized to form a population distribution function.

The calculation of the value of f_d is not straightforward. Let i be the ideal number of documents specified by the user. If the user does not know the value of i, the default value will be 20. Using the value of i as the mean value, two *band pass filters*-like functions, namely the triangular and Gaussian functions, are used to create a more flexible mapping from the number of documents retrieved (*d*) to f_d. The triangular function gives a constant drop in 'gain' (decrease in fitness) from the 'center frequency' (the mean). This function is good when the user wants to give an equal amount of demerits for every document that is away from his expected or ideal number. The Gaussian function is a more robust or high-ordered 'band pass' such that its 'bandwidth' could be specified by the value of the standard deviation. Thus, this function is useful when the user wants to give heavy demerits to queries that do not fall near his expected or ideal number.

Only requirements that are specified by numerical constraints will have the value f_q. Here, it is required that the numerical values are summed up and averaged. Then, they are normalized. The signs "<" and ">" formed during the query indicate the direction in which the quality is favored.

Another interesting portion that contributes to the fitness function is the correlation of the synonyms (f_r) with the ontology terms. A value from 0 to 1 is assigned to each relation between the ontology terms and their synonyms. When a requirement is <Television><Price><<><2000>, the value or f_r will be the product of all the discrete correlations. Also, the user should be able to edit the correlation values to his preference. When there are many requirements in the query, these requirements will be linked with an "OR" or "AND" term.

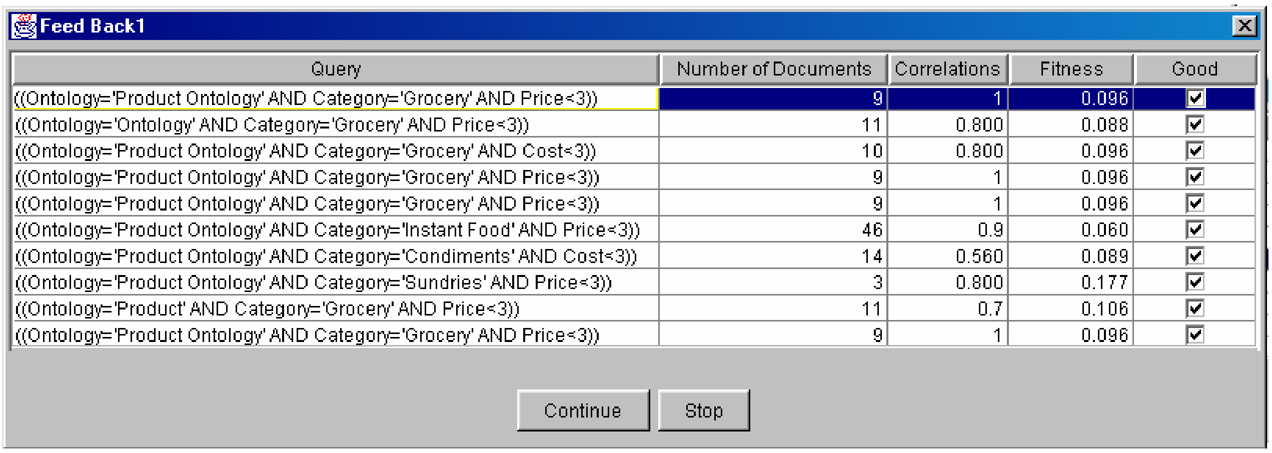
Mutation and Crossover

The concept of mutation is to replace some terms with synonyms when parsing the results. Basically, the mutants are the terms that are included in each query. These terms are mutated randomly according to the synonyms so that new populations will be formed. Crossover will only interchange the different genes between two different chromosomes. A one-point crossover will be performed. This is also done randomly.

Feedback and Selection of Survival

The survivors are selected according to their overall fitness in the roulette-wheel selection manner. However, before this is done, the system will prompt for feedback from the user. The feedback will show the user some quality of each query. From this quality metric, the user may choose to kill queries that do not meet his or her requirements. Figure 3 shows a screenshot of a feedback presented to the user. If the user is satisfied with the results, he or she can choose to end the genetic algorithm by clicking on the "Stop" button. In this way, he or she can look at the retrieved results immediately.

Figure 3. Screenshot of a feedback frame



Query	Number of Documents	Correlations	Fitness	Good
((Ontology='Product Ontology' AND Category='Grocery' AND Price<3))	9	1	0.096	☑
((Ontology='Ontology' AND Category='Grocery' AND Price<3))	11	0.800	0.088	☑
((Ontology='Product Ontology' AND Category='Grocery' AND Cost<3))	10	0.800	0.096	☑
((Ontology='Product Ontology' AND Category='Grocery' AND Price<3))	9	1	0.096	☑
((Ontology='Product Ontology' AND Category='Grocery' AND Price<3))	9	1	0.096	☑
((Ontology='Product Ontology' AND Category='Instant Food' AND Price<3))	46	0.9	0.060	☑
((Ontology='Product Ontology' AND Category='Condiments' AND Cost<3))	14	0.560	0.089	☑
((Ontology='Product Ontology' AND Category='Sundries' AND Price<3))	3	0.800	0.177	☑
((Ontology='Product' AND Category='Grocery' AND Price<3))	11	0.7	0.106	☑
((Ontology='Product Ontology' AND Category='Grocery' AND Price<3))	9	1	0.096	☑

Prototype Testing and Evaluation of Genetic Algorithms

Effectiveness of the Genetic Algorithm

It is believed that the effectiveness of the genetic algorithm chosen is mainly determined by its supremacy in query effectiveness amplification. This is because its evolution power allows more retrieval results. The system was tested with a product list database. The effectiveness was measured by testing a series of queries with and without using genetic algorithms. For example, a query “<Product Ontology><Grocery><Price><<> <3>” can only retrieve nine items when a normal search is performed, but can retrieve 98 items when a genetic algorithm is performed. Table 1 shows the other results obtained by other queries.

By comparing the results shown in Table 1, it is obvious that using a genetic algorithm does, in fact, retrieve more items than using a normal search.

Effect of the Fitness Function

The fitness function in a genetic algorithm determines how well it can optimize a query. The OntoQuery system tested various fitness functions to improve the power of the genetic algorithm. The usage of triangle or Gaussian functions to evaluate the fitness for the number of documents retrieved suggested some ways to counter the “too many or too few retrieved documents” dilemma in typical search engines.

Table 1. Results showing effectiveness of GAs

Query Formed	Without GA	With GA
<Product Ontology><Drinks><Price><<><2>	25	95
<Product Ontology><Diary><Price><<><3>	3	92
<Product Ontology><Candy><Price><<><3>	13	178
<Product Ontology><Confectionery><Price><<><3>	20	248
<Product Ontology><Confectionery><Supplier> <Contains><Ho>	1	52

Efficiency of the Genetic Algorithm

Although using genetic algorithms allows a more flexible and effective platform in retrieving information, there is no doubt that they trade off efficiency due to their expensive iterations. Thus, the only study that can be made here is about their improvement over relevance feedback. In relevance feedback, query expansion is achieved by modifying a query. Similarly, genetic algorithms extend the relevance feedback techniques with an additional rule, the survival of the fittest.

In this research, the efficiency of the system is measured as follows:

$$e\ (\text{Efficiency}) \approx \frac{E}{t}$$

$$\approx \frac{D}{I}$$

where E denotes the effectiveness of the system.
t denotes the time taken for the system.
D denotes the number of relevant documents retrieved.
I denotes the number of iterations.

Efficiency is formulated as above because it is believed that the number of documents retrieved is linearly proportional to the effectiveness of the system. Also, the number of iterations is directly related to the time taken to retrieve the results.

CONCLUSION AND FUTURE WORKS

In summary, this research work investigated the OntoQuery system within an m-commerce agent framework against current query formation and information retrieval systems from extant work which were not intended for m-commerce.

The prototype implementation results showed that querying formation using an ontology approach is efficient, as it provides a friendly environment to the user using mobile devices. In addition, by combining the keyword and ontology approaches, a more efficient and effective way of forming queries could be achieved. Thus, the objective to propose efficient query formation for product databases is successful.

It was found that genetic algorithms are able to optimize queries effectively. Also, using genetic approaches, we have proposed and tested various fitness functions for searching product databases. Moreover, adding feedback to the system helps it to cater to the needs of the user more closely.

Considering typical e-commerce or m-commerce users using mobile devices for product enquiry tend to have some time constraints and they need quick response before making decisions, our work suggests a feasible approach. With the use of ontology, product enquiry can be formed easily and quickly. With the help of genetic algorithms and agents, and the technique of query optimization such as query restructuring, futile enquiries can be made productive. And finally, the use of genetic algorithms also helps improve the quality of product information retrieved.

REFERENCES

Boughanem, M., Chrisment, C. & Tamine, L. (1999). Genetic approach to query space exploration. *Information Retrieval, 1*(3), 175-192.

Braga, R.M.M., Werner, C.M.L. & Mattosso, M. (2000). Using ontologies for domain information retrieval. *IEEE Proceedings of the 11th International Workshop on Database and Expert Systems Applications* (pp. 836-840).

Fensel, D. (2000). The semantic Web and its language. *IEEE Intelligent Systems, 15*(6), 67-73.

Hendler, J. (2001). Agents and the semantic Web. *IEEE Intelligent Systems, 16*(2), 30-37.

Karp, P.D. (2000). An ontology for biological function based on molecular interactions. *Bioinformatics, 16*(3), 269-285.

Kouichi, A.B.E., Taketa, T. & Nunokawa, H. (1999). An efficient information retrieval method in WWW using genetic algorithm. *Proceedings of the International Workshops on Parallel Processing* (pp. 522-527).

Kraft, D.H., Petry, F.E., Buckles, B.P. & Sadasivan, T. (1994). The use of genetic programming to build queries for information retrieval. *Proceedings of the 1st IEEE Conference on Computational Intelligence* (pp. 468-473).

Losee, R.M. (1988). Parameter estimation for probabilistic document retrieval models. *Journal of the American Society for Information Science, 39*(1), 1-16.

Losee, R.M. (1991). An analytic measure predicting information retrieval system performance. *Information Processing and Management, 27*(1), 1-13.

Losee, R.M. (1996). Learning syntactic rules and tags with genetic algorithms for information retrieval and filtering: An empirical basis for grammatical rules. *Information Processing & Management, 32*(2), 185-197.

McGuinness, D.L. (1998). Ontological issues for knowledge-enhanced search. (FOIS'98) (pp. 302-316).

Preece, A., Hui, K., Gray, A., Marti, P., Bench-Capon, T., Jones, D. & Cui, Z (1999). The KRAFT architecture for knowledge fusion and transformation. In M. Bramer, A. Macintosh & F. Coenen (Eds.), *Research and development in intelligent systems XVI* (pp. 23-38). Berlin: Springer-Verlag.

Salton, G. (1989). *The transformation, analysis and retrieval of information by computer*. Reading, MA: Addison-Wesley.

Salton, G. & McGill, M. (1983). *Introduction to modern information retrieval*. New York: McGraw-Hill.

Yang, J.J. & Korfhage, R.R. (1994). Query modification using genetic algorithms in vector space models. *International Journal of Expert Systems: Research and Applications, 7*(2), 165-191.

KEY TERMS

Crossover: In genetic algorithms, it is the process of combining features of a chromosome with other chromosome(s).

Evolutionary Algorithm: An algorithm incorporating aspects of natural selection or survival of the fittest.

Fitness Function: In genetic algorithms, it is a measure of how well a chromosome can perform in certain environments (functions).

Genetic Algorithm: An evolutionary algorithm that generates each individual from some encoded form known as "chromosomes" or "genome."

Mutation: In genetic algorithms, it is defined as a change in form or qualities of chromosomes.

Ontology: Looks for semantic and ontological primitives and concepts to describe aspects or parts of "the world."

Restructuring: Replacing query terms using synonyms so that logical operators and numerical constraints can be restructured.

Roulette-Wheel Selection: A roulette wheel selects the chromosomes used in reproduction. The wheel is the fitness array, and the marble is a random unsigned integer less than the sum of all fitness in the population. To find the chromosome associated with the marble's landing place, the algorithm iterates through the fitness array; if the marble value is less than the current fitness element, the corresponding chromosome becomes a parent. Otherwise, the algorithm subtracts the current fitness value from the marble and then repeats the process with the next element in the fitness array.

Open Source Software Development Model

Luyin Zhao
Philips Research, USA

Fadi P. Deek
New Jersey Institute of Technology, USA

INTRODUCTION

The open source movement can be traced back to the hacker culture in the ’60s and ’70s. In the early 1980s, the tenet of free software for sharing was explicitly raised by Richard Stallman, who was working on developing software systems and invited others to share, contribute, and give back to the community of cooperative hackers. Stallman, together with other volunteers, established the Free Software Foundation to host GNU (Gnu’s Not Unix, a set of UNIX-compatible software system). Eric Raymond, Stallman’s collaborator, is the primary founder of the Open Source Initiative. Both communities are considered the principal drivers of open source movement.

A number of worldwide, online communities for open source development have been established since then to facilitate the development of open source software. For example, the Open Source Development Network (OSDN) is one of the largest organizations for such purpose. One of its subordinates, sourceforge.net, is hosting nearly 70,000 projects and 700,000 registered developers and users at the present time. The recent wave of sponsoring open source projects by commercial companies is another significant phenomena. Big IT players such as IBM, Sun, and HP have realized the importance and benefits to “open” their source. From other perspectives, open source applications have been expanding into various domains, including education, the Internet, office management, programming, communication, and even the medical domain. The Apache server powers half of all Web servers worldwide, far more than Microsoft and Netscape combined. Table 1 lists some typical application domains and well-known open source software.

Table 1. A list of popular open source software products

Operating system: Linux, FreeBSD
Internet: Apache Server, Mozilla
Communication: sendMail, OpenSSL
Programming: Perl, Tcl/Tk, GNU
Office: OpenOffice

BACKGROUND

With the successful delivery of many software products, the open source development model has been attracting increased interests from both practitioners and researchers. The open source development model can be characterized by its fast evolution, distributed development, and extensive user collaboration. It is a simple mean of releasing software with free source code, but one that brings a series of new social and technical challenges, including licensing, distributed development, project management, commercial adoption, and user collaboration. A number of research methods have been adopted to investigate the phenomena of open source development, including general descriptive discussion (Raymond, 1999; Hars & Ou, 2002; Cubranic & Booth, 1999; Augustin, Bressler & Smith, 2002), case studies (Mockus, Fielding & Herbsleb, 2002; Lakhani & Hippels 2003), and surveys (Zhao & Elbaum, 2003; Lakhani & Hippel, 2003). As usual, different research methods employ unique ways of investigation but have potential weaknesses. For example, case studies tend to focus on a few large open source projects such as Apache, Linux, and Mozilla, but lack of comprehensive observations on common issues for a broader spectrum of open source projects. Survey research is able to cover a large number of projects, however it may ignore details or specific issues for individual open source projects.

The open source model can dramatically affect changes in the traditional ways of software development. Table 2 lists some relevant issues that arise with open source development.

.

KEYS TO UNDERSTANDING THE OPEN SOURCE DEVELOPMENT MODEL

Despite the debates, critiques, and the evident enthusiasm regarding the Open Source Development Model, without any doubt, it is becoming a recognized paradigm that competes with the traditional methods of software development and is expected to grow in the future. To understand this phenomena, several essential attributes

Table 2. Issues that arise with the Open Source Development Model

- *Reliability:* Whether software developed under a model different from traditional carries same or higher reliability.
- *Licensing:* Licenses under which open source software products are distributed. How terms and conditions are phrased.
- *Release Management:* The rapid evolutionary open source model requires unique version control, release distribution, and management.
- *Quality Assurance:* What traditional QA methods and open-source-specific mechanisms (e.g., user contribution) are leveraged to insure quality.
- *Adoption:* How open source products are adopted by people and organizations outside of the open source community.
- *Documentation and Maintenance:* Due to the voluntary nature, attention should be paid on later development tasks such as documentation and maintenance offered to users.
- *Project Management:* Efforts are needed to keep the project stable and attract as many contributing users as possible to improve the software.
- *Collaboration Tools:* Without effective collaborative tools and environments, it is impossible to complete projects as a virtual team.

of the Open Source Development Model must be observed, including motivation to contribute, development process and quality assurance, communication methods, and user collaboration.

- *Motivations to Contribute*—It is sensible to understand the open source model by trying to understand the motivations of contributors who are willing to develop software and give it away for free. Zhao and Elbaum (2003) discovered that open source projects are started mainly for personal needs, community needs, and company needs. In a more focused study on this topic, Hars and Ou (2001) identified two types of motivations that account for people's participation in open source projects. These are internal factors that include intrinsic motivation, altruism, and community identity, and external factors that include future rewards (i.e., selling products), human capital, self-marketing, peer recognition, and personal need. Lakhani and Hippel (2003) also find that an opportunity to lead is also a motivation for developers to contribute. A recent survey study (Zhao & Deek, 2004), designed specifically for open source users, found that in user groups, most of the users are motivated to contribute by personal needs, followed by having fun and employer needs. Also, the same study verified that learning is a major benefit that users receive by contributing.
- *Communication Methods*—Open source development is a success story of collaborative development through computer-mediated virtual communities such as sourceforge.net. That is also a reason why open source projects attract contributors from all over the world. Web-based collaborative tools enable developers and users from different locations to coordinate and collaborate on development tasks. However, current collaboration environments are still largely facilitating technically oriented tasks (e.g., version control) instead of socially oriented ones (e.g., group decision making). Future trends of computer-mediated collaboration methodologies in open source development are likely to expand towards the socially oriented direction.
- *Development Process and Quality Assurance*—Unlike with traditional or commercial software development, open source projects are less formally organized in terms of following certain process models. Similarly, quality assurance activities are carried out in a less formal ways (Zhao & Elbaum, 2003). The rapid release and evolution of software raises some discussions on the analogous relationship with the eXtreme Programming model. Although there are a large variation in terms of how the evolutionary processes are managed in different projects, Zhao and Elbaum (2003) present many interesting findings relating to the lifecycles of a broad range of open source projects. For example,

43% of projects have a new release every month, around 75% of projects use configuration management tools, and when projects become more mature, the tendency of serving different users changes gradually. There is a belief, however, within the open source development community that extensive user inspection and contribution is a good remedy for the lack of formal development processes. However, we expect to see formal quality assurance activities in open source development, as more commercial IT players join in the open source community.

- *Extensive User Collaboration*—In a recent study, Zhao and Deek (2004) found that open source users are highly knowledgeable in software development at a level that is comparable to "core" developers. In other words, the projects are supported by a much larger group of experienced developers in the "background." However, this type of support occurs more often in large open source projects. The majority of open source projects are small in scale, lack sufficient user support, and are like smaller, start-up companies in the commercial world. The key for the growing of these small projects is to attract more developers and users; to achieve this goal, it is critical to understand motivations of users and use mechanisms to fulfill their needs (e.g., providing learning opportunities).

FUTURE TRENDS

There is no doubt that the open source development process is becoming more and more mature when researchers and practitioners gain deeper insights into this unique model. Future research on this topic includes attracting user contributions, providing better communication support, knowledge sharing within open source communities, overcoming language barriers, providing learning opportunities for participants, and the impact of commercial software vendor involvement.

CONCLUSION

The open source development model is an evolving effort. Participants with various skills, interests, needs, and geographic locations are able to take part and collaborate to produce powerful and free software, although there remains a host of issues to be understood and resolved by researchers and practitioners. We urge the readers to view and understand this model from four key aspects: motivation, collaboration, process, and user collaboration.

REFERENCES

Augustin, L., Bressler, D. & Smith, G. (2002). Accelerating software development through collaboration, *Proceedings of the 24th International Conference on Software Engineering* (pp. 559-563).

Cubranic, D. & Booth, K.S. (1999). Coordinating open source software development. *Proceedings of IEEE 8th International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprise* (pp. 61-69).

Hars, A. & Ou, S. (2001). Working for free? Motivations of participating in open source projects. *Proceedings of the 34th Annual Hawaii International Conference on System Sciences* (pp. 2284–2292).

Lakhani, K.R. & Hippel, E. (2003). How open source software works: "Free" user-to-user assistance. *Research Policy, 32*(6), 923-943.

Mockus, A., Fielding, R.T. & Herbsleb, J.D. (2002). Two case studies of open source software development: Apache and Mozilla. *ACM Transactions on Software Engineering and Methodology, 11*(3), 309-346.

Open Source Initiatives. *www.opensource.org.*

Raymond, E.S. (1999). Linux and open source success. *IEEE Software, 16*(1), 85-89.

Zhao, L. & Elbaum, S. (2003). Software quality assurance under the open source model. *Journal of Systems and Software, 66*(1), 65-75.

Zhao, L. & Deek, F. (2004). User collaboration in open source software development. *Electronic Markets, 14*(2), 89-103.

KEY TERMS

Computer-Mediated Communication: The process by which people create and exchange information using networked systems that facilitate encoding, transmitting, and decoding of messages.

Open Source: There is no universally accepted definition, but the one given by the Open Source Initiative summarized the concept as: free distribution, source code, derived works, integrity of the author's source code, no discrimination against persons or groups, no discrimination against fields of endeavor, distribution of license, license must not be specific to a product, license must not restrict other software, and license must be technology neutral.

Open Source Developer: Software developers who are core members of open source projects. They either initiated projects or joined in subsequently as major contributors.

Open Source Users: Individuals interested in open source software and willing to contribute by different means. Unlike open source developers, open source users are more heterogeneous in background, less committed to projects, but provide major impetus to software improvement.

Software Process: Methodologies offering guidelines for improving software engineering performance by requiring developers to adopt a disciplined, structured approach to building software systems.

Software Quality Assurance: Methods and approaches to ensure that a software system meets specified requirements, and customer or user needs or expectations.

User Collaboration: Voluntary collaborative efforts from users in developing open source software. The collaboration includes looking for bugs, suggesting new features, reviewing/submitting source code, documentation, and offering project administration assistance.

Optical Music Recognition with Wavelets

Susan E. George
University of South Australia, Australia

INTRODUCTION: SUPER-IMPOSED OBJECTS

The aim of optical music recognition (OMR) is to "recognise" images of music notation and capture the "meaning" of the music. When OMR is successful it will be able to automatically extract a logical representation of printed or handwritten music captured in an image. There are a variety of reasons why OMR is required. Chiefly, it is convenient for swift input of music notation and might be subsequently edited, performed, used as a search or other. There are many stages before that final high-level interpretation can be made and recognition of the primitive symbols contained in the notation is primary. One of the biggest challenges in OMR is the super-imposition of music notation symbols – notes and other – upon stave lines in the music image. This article examines a general-purpose knowledge-free method in the wavelet transform, to deal with super-imposition in images of typeset music.

Super-imposition arises when notes and other music symbols are placed upon the stave line (or staff), making "ink" of the symbol overlap with "ink" of the stave line. There are various reasons why isolating the super-imposed object is so difficult within OMR. Firstly, image capture may have introduced perturbations; stave lines in the image are rarely parallel, horizontal, equidistant, of constant thickness, or even straight. Secondly, other symbols within the music (such as beams, phrase marks and others) can be mistaken for stave lines and hence lines mis-located. Thirdly, there is only one piece of "ink" for objects that are super-imposed upon the lines and stave lines have to be extracted, leaving the symbol intact — or conversely the symbol has to be segmented from the stave line, having identified its location within the stave.

The OMR field has taken two basic approaches to dealing with super-imposed objects. These approaches are (i) the removal of stave lines and (ii) the recognition/segmentation of music symbols. Neither of these methods has met with complete success. When focusing upon the segmentation, difficulties specific to music notation arise, including (i) the variant size of symbols (e.g., phrase markings, slurs) rendering template matching inappropriate and (ii) the various ways of typesetting a particular musical sound are potentially infinite, and again template matching would not suffice for all the possibilities.

Since the 1960s there have been various approaches to dealing with super-imposed objects. All of these approaches are knowledge-based techniques, in that they are assuming some information is known about the format of music images in order to locate the stave lines or isolate symbols. Blostein and Baird (Blostein & Baird, 1992) present a critical survey of problems and approaches to music image analysis. Here we summarise some of the approaches to OMR from the initial attempts in the 1960s.

- Pruslin (1967) - remove thin horizontal lines (by thresholding the length of continuous vertical runs of pixels, assuming that a figure for "line thickness" was known)
- Prerau (1975) – remove stave lines and restore those parts that coincide with symbols (using a contour trace from the edge of each stave line to locate the start of symbols that might be placed on those lines).
- Nakamura et al. (1979) – remove stave lines using line tracker, calculating a least-squares fit for stave lines computing threshold to erase the stave lines and achieve segmentation.
- Andronico and Ciampa (1982) - remove just the exposed stave lines.
- Aoyama and Tojo (1982) – detect stave lines by using the horizontal histogram of black pixels within each block of low resolution scan lines and using a threshold to find sets of five peaks. Coarse segmentation was then undertaken, removing obvious, exposed sections of stave line by examining the vertical run-lengths involved. Fine segmentation detected black and white noteheads by searching for overlapping pixel runs.
- Mahoney (1982) - isolate symbols from stave lines, interpolating between end-points where there were gaps.
- Matsushima et al. (1985) – detect stave lines by a short bar-like filter that operated down equi-spaced columns in the image, simultaneously with the scanning process.
- Roach and Tatem (1988) – detect stave lines with a line-tracking algorithm calculating line direction and thickness at each pixel in a grey-scale image by

passing a window of appropriate proportions over the image. They have also considered the specific challenges of handwritten music.

- Brainbridge and Bell (1997) – determine presence of stave lines by following the wobble in groups of pixels and remove with heuristics.
- Lin and Bell (Lin et al., 2000) - propose colour printing and scanning technology to help deal with the issue of super-imposed objects.
- Choudhury, DiLauro et al. (2001) – remove staff lines using projection profiles, remove text by heuristics, vertical components of stems and barlines and horizontal components of note heads, recognising the remaining symbols with K-NN classifier combing the glyphs again.
- Droettboom et al. (Droettboom et al., 2002) - describe the use of classifiers for symbols recognition, providing tools for the creation of a simple heuristic classifier, a template-based image matching and a k-nearest neighbour learning classifier.

All these approaches have met with some success and various limitations. Perhaps the biggest limitation is that they are knowledge-based, requiring specific knowledge about music to isolate the symbols, or follow stave lines. Implicitly, knowledge-based techniques are constrained by the knowledge that they may draw upon. For example, a symbol finding algorithm can only find symbols that it knows about, and should a new symbol be encountered within the music notation, the recognition is likely to fail. Similarly, a stave line tracking algorithm is constrained by heuristics about how stave lines generally appear, or in what ways they may be corrupted. When knowledge is explicitly utilised, at some point, a novel input will be encountered that falls outside the scope of the knowledge-based method (this is especially the case when dealing not with "printed" documents, but handwritten ones). Hence the importance and appeal of general-purpose knowledge-free methods.

BACKGROUND: WAVELETS

Wavelets are a relatively recent mathematical approach extending some of the principles of Fourier analysis for the study of periodic signals, decomposing a signal into its component parts. Wavelet theory is well suited for complex signals, where the frequency structure changes throughout the signal (i.e., non-periodic signals). Since very few signals or images are truly periodic, wavelet techniques are ideal. Wavelets permit the signal to be viewed so that the large-scale fluctuations are emphasised (with the small detail as noise), or such that the small fluctuations are emphasised (with the larger scale fluctuations as background). Interest in wavelets is also stimulated by the fact that some wavelets may be implemented in an extremely computationally efficient manner.

The wavelet is defined by the "mother wavelet" from which other wavelets in the "family" can be generated. All members of a particular family share the same basic wavelet shape that is shifted or dilated (*i.e.*, they are made tall and skinny or short and fat). The specific parameters are translation ('b') and contraction ('a'). Members of an individual wavelet family y ^{a,b} (x) can be defined by (1):

$$\psi^{a,b}(x) = |a|^{-1/2}\,\psi\left(\frac{x-b}{a}\right) \qquad (1)$$

Equation 1 generates a one-dimensional function that can be plotted. There are a number of families of wavelets including Harr, Daubechies, Coifmanns, Mexican Hat, Meyer and Morlet. The coefficients in the wavelet transform are an important part of the wavelet expression since (combined with the generating mother wavelet) they compact the signal and can be used to approximate a function. Instead of having to store every value of the function, it is only necessary to store the coefficients, perhaps at various levels of decomposition, and from these the signal can be obtained. Two-dimensional wavelets can be applied to two-dimensional signals; that is, the wavelets are applied to images.

In terms of general image processing, wavelets decompose images into their high and low pass components in row and column-wise directions, thus filtering an image into its component parts. There are four combinations of low pass filters (H1 and H2) and high pass filters (G1 and G2) in the vertical (H1 and G1) and horizontal (H2 and G2) directions. These decompositions are typically represented in quadrant diagrams where the top left quadrant contains an approximation of the original image, the top right contains the horizontal components, the bottom left the vertical components and the bottom right the diagonal components.

When considering wavelet image decomposition, it is worth observing that the music image possesses certain properties, including: (i) the music stave lines – which present a strong horizontal component to the image signal and which are (generally) regular across the image page, (ii) other music notation symbols - which present a more vertical component and are more local perturbations within the image; wavelets are particularly useful for such local irregularities within an image, and some wavelets may be more suitable than others for the regular/irregular patterns that occur within a music image.

WAVELETS IN OMR ANALYSIS

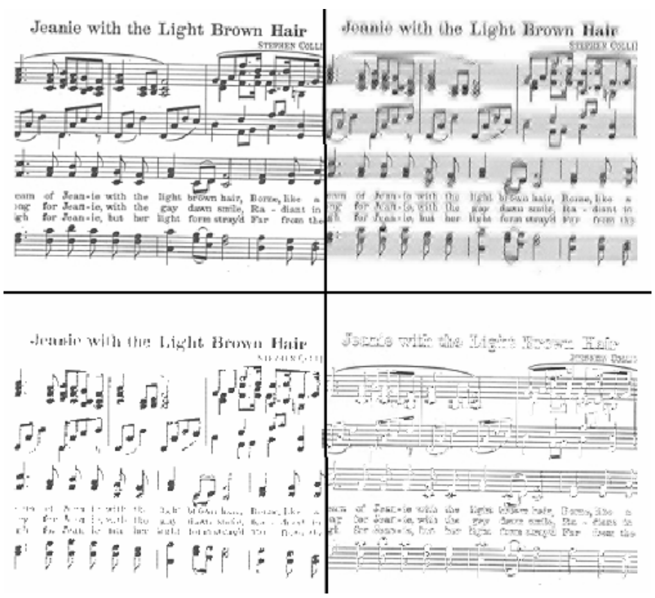
Wavelets represent a general-purpose knowledge-free method of image filtering that is capable of separating super-imposed objects. They provide a fresh (theoretical) perspective on the problem of super-imposed objects in music notation; recognising the duality of the segregation task that exists with stave line removal/symbol extraction. The wavelet provides a unified theoretical framework within which to deal with super-imposed objects in OMR.

Recent results (George, 2004) have found that they are able to segregate the super-imposed components for stave line removal and symbol segmentation, finding that the Coifmann family are most suitable for emphasising vertical image components and the Daubechies for the horizontal. Evaluation is conducted at a pixel-based level using a truth representation of the image to evaluate the accuracy with which symbols are identified.

Figure 1 illustrates a 256 x 256 binary image filtered with the Daubechies wavelet. The original image appears in the top left; the filtered image (top right) has been blurred in the horizontal direction, making the stave lines and other horizontal components less distinct. The bottom left quadrant contains the thresholded transformed image, showing how the cut-off has removed the lighter blurred pixels from the image (these are precisely the stave lines, leaving chiefly the notes in the filtered image). The bottom right quadrant illustrates the difference between the original and the thresholded transformed image, showing the areas of overlap between notes and stave lines.

One of the first questions that arise with the application of wavelet transforms to super-imposed objects within music is the question of evaluation. That is, the question of "how can we know whether a wavelet transform has been of any benefit to OMR?" Fundamental to evaluating accuracy is some target "truth representation". For images, the target representation is another image and a pixel-based measure can be made comparing the filtered image with the target, eventually attaining a sensitivity/specificity measure for how well the filtering has "diagnosed" black/white pixels in the image.

Figure 1. Example of wavelet filtering



Locating Stave Lines

We can demonstrate that stave lines can better be detected in a wavelet filtered version compared to the original image. Various music samples were captured, including: stave lines interspersed with text, small font inserted staves, incomplete staves and skewed staves, as illustrated in Figure 2. These examples present special challenges to conventional stave line finding methods.

Stave lines are located in the original image and the wavelet transform using a simple pixel histogram along the rows to locate where the peaks and troughs of black pixels are. The pixel histograms of the original and transformed image are compared. The presence of peaks in the histograms indicates where there is a strong horizontal signal in the image. Ideally, for locating stave lines there will be five such peaks close together. This characteristic peak is most clearly seen in the filtered images rather than the original images, suggesting that whatever stave line finding algorithms are used, the filtered image provides a clearer starting base. For complete details of the experiments see George (George, 2004).

Segmenting Symbols

Similarly, we can demonstrate that the wavelet can help in isolating the symbols superimposed upon the lines. Three music fragments were selected, containing a varying amount of detail across the image, as illustrated in Figure 3. From these music fragments six images were generated, capturing the image at 150, 250 and 350dpi in binary and grayscale modes. For each of these images a truth representation was constructed by hand, separating the stave lines from the music symbols that appear upon them.

The Daubechies-3 wavelet transform was applied to each of the images in turn, isolating the horizontal and vertical components of the image and the filtered image compared with the truth representation containing the segmented symbols. The results suggest that image 2 (containing just a few bars) was the easiest image to deal with since the average sensitivity was 0.65 and specificity was 0.75 over all the image capture parameters. The hardest image was image 3, with average sensitivity of

Figure 2. Examples for stave line locating

0.31 and specificity of 0.78 over the image capture parameters. However, it was determined that a pixel-based evaluation may be misleading since pixels in the filtered image may simply be shifted left or right and cause a low index in the quantification, although in reality the filtering has benefited the isolation of symbols from the horizontal components.

CONCLUSION

We have found that it is possible to provide a unified theory for dealing with super-imposed objects in OMR, since both stave line location and symbol segmentation can be addressed by the same technique – that of wavelet filtering. As a knowledge-free, general-purpose image processing technique, wavelet filtering makes it possible to emphasise certain features of the image using the image decompositions.

We observed that (i) the horizontal pixel histograms made stave lines easier to detect in the filtered image. Stave line finding algorithms may want to use the filtered image rather than the original to improve upon their results; (ii) the accuracy of detection for segmented music symbols appeared better in low resolution binary images, but this was mainly due to the amount of "ink" involved in the pixel-based accuracy evaluation (which was made between the truth representation and the filtered image). Nevertheless, filtering the image to isolate the non-stave line components emphasised a feature that was not so obvious in the original image.

At present, wavelet filtering is likely to be most practically useful as a pre-processing method to emphasise certain features of the images for conventional OMR methods. However, there is scope that it may be used in isolation as a knowledge-free OMR procedure for super-imposed objects that can isolate notation symbols from stave lines. We consider that further work is required to continue the investigation of the wavelet as a knowledge-free, general-purpose image processing technique for OMR that is able to provide a unified theory for super-imposed objects.

Figure 3. Examples for symbol segmentation

REFERENCES

Andronico, A., & Ciampa, A. (1982). On automatic pattern recognition and acquisition of printed music. *Proceedings of the International Computer Music Conference,* Venice, Italy (pp. 245-270).

Aoyama, H., & Tojo, A. (1982). *Automatic recognition of printed music.* IECE (Japan) Technical Report, PRL 82-85, 33-40. (Translated from the original Japanese.)

Bainbridge, C., & Bell, T.C. (1997). Dealing with superimposed objects in optical music recognition. *6th International Conference on Image Processing and its Applications.*

Blostein, D., & Baird, H.S. (1992). A critical survey of music image analysis. In H.S. Baird, H. Bunke & K. Yamamoto (Eds.), *Structure document image analysis* (pp. 405-434). Berlin, Heidelberg: Springer-Verlag.

Choudhury, G.S., DiLauro, T., Droettboom, M., Fujinaga, I., & MacMillan, K. (2001, February). Deriving searchable and playable digital formats from sheet music. *D-Lib Magazine*, *7*(2).

Droettboom, M., MacMillan, K., Fujinaga, I., Choudhury, G.S., DiLauro, T., Patton, M., & Anderson, T. (2002). *Using Gamera for the recognition of cultural heritage materials.* Submitted for consideration to Joint Conference on Digital Libraries.

George, S.E. (2004). Wavelets for dealing with superimposed objects in the recognition of music notation. In *Visual perception of music notation.* Hershey, PA: Idea Group Publishing.

Lin, K., & Bell, T. (2000). Integrating paper and digital music information systems. *International Symposium on Music Information.* Retrieved March 20, 2002, from *http://ciir.cs.umass.edu/music2000/papers.html*

MathWorks. (1994). *MATLAB reference guide.* The MathWorks Inc.

Mahoney, J.V. (1982, May). *Automatic analysis of music score images.* BSc dissertation. Massachusetts Institute of Technology.

Matsushima, T. et al. (1985). Automated recognition system for musical score. *Bulletin of Science and Engineering Research Laboratory Waseda University, 112,* 25-52.

Nakamura, Y., Shindo, M., & Inokuchi, S. (1979). *Input method of musical note and realization of folk-music database.* IECE (Japan) Technical Report, PRL 78-73, 41-50. (Translated from the original Japanese.)

Polikar, R. (1996). The wavelet tutorial part 1, Fundamental concepts & an overview of the wavelet theory. Retrieved February 8, 2003, from http://www.public.iastate.edu/~rpolikar/WAVELETS/WTpart1.html

Prerau, D.S. (1975, January). DO-RE-MI: A program that recognizes music notation. *Computers and the Humanities, 9*(1), 25-29.

Pruslin, D.H. (1967, January). *Automatic recognition of sheet music.* ScD dissertation. Massachusetts Institute of Technology.

Roach, J.W., & Tatem, J.E. (1988). Using domain knowledge in low-level visual processing to interpret handwritten music: An experiment. *Pattern Recognition, 21,* 33-44.

Saha, S. (2001) Image compression - from DCT to wavelets: A review. ACM Crossroads. Retrieved February 8, 2003, from http://www.acm.org/crossroads/xrds6-3/sahaimgcoding.html

Vidakovic, B. (1999). Statistical modeling by wavelets. *Wiley Series in Probability and Statistics.*

KEY TERMS

Knowledge-Free Image Processing: A general method of dealing with an image that does not require specific information or data. The image is simply filtered, typically uniformly, and so transformed or processed without the need for knowledge-based methods.

Optical Music Recognition (OMR): OMR is the recognition of images of music notation in such a way that the "meaning" is captured.

Stave lines/Stafflines: The five horizontal lines upon which symbols of music notation are placed.

Super-imposed Object: Super-imposition is an image processing problem that occurs in music notation as music symbols are placed upon the stave lines. Separating the symbol from the lines becomes an important task in most approaches to recognition. Approaches either attempt to remove the stave lines or segment the music symbols.

Wavelets: Wavelets are a relatively recent mathematical approach extending some of the principles of Fourier analysis for the study of periodic signals, decomposing a signal into its component parts. Wavelets permit the signal to be viewed so that the large scale fluctuations are emphasised (with the small detail as noise), or such that the small fluctuations are emphasised (with the larger scale fluctuations as background).

Organization and Management Issues in End User Computing

Jason C.H. Chen
Gonzaga University, USA

Robert W. Holt
Gonzaga University, USA

D. Bruce Sun
California State University at Long Beach, USA

INTRODUCTION

End-user computing (EUC) or as it is commonly termed, end-user development (EUD), is a concept and capability granted by advancement in technology that allows participants in a business environment to utilize information technology (IT) by developing applications of their own. Traditional methodology required a software program to be developed by trained programmers in the analysis and design stages of the systems development life cycle, where a user had to accept the program as an individual entity with unalterable characteristics. EUC/EUD now enables this person to customize the program around his/her specific demands. The framework of EUC establishes empowerment and capabilities so that anyone can develop entire information systems (IS) with little or no help from professional systems analysts or programmers, along with accessing data and creating reports (Laudon & Laudon, 2003). EUC/EUD is a topic in the IT environment that cherishes a progressive history spanning from the mid-1970's to what it as today.

BACKGROUND

During the 1970's, the concept of management information systems (MIS) began which grounded the importance of utilizing IT as a strategic implementation tool in changing business environments. The event produced a two-fold outcome where generalized perceptions about computers changed to include the relevance and effective usage of data to direct most business decisions, and technological ideology shifted from an information specific denomination to a belief towards support management. (Charr, 1988). Once support management became recognized, a paradigm occurred where decision support systems (DSS) became the focal point of MIS integration. The new paradigm conceptualized the computer as a necessary tool for the decision making process in accordance with its data storing properties.

EUC has become increasingly more available, due to the induction of less complicated programming languages, termed fourth-generation languages (4GL's), that provide users who may not have a sufficient skill and knowledge of programming expertise, to develop programs, customized to their individual needs. These specific languages signify simplicity, since most users already have an intuitive understanding of the logic and terminology. They are frequently referred to as user-friendly and nonprocedural, which determines that the languages must be given less structured instructions to achieve the same result as earlier capabilities (Charr, 1988). Their relevance to EUC is significant because earlier languages, such as assembly or procedural language, required a high level of education to interpret the meaning of codes and to produce the desired outcome with the program.

From a technology perspective, when the last three decades are crucially examined on how EUC has gained its prominence today, three defining events are attributable for this outcome (Charr, 1988). First, is the success of computer engineering to derive programs which grant users, whom are not skilled computer programmers, to manipulate and maintain the software; second, computer training is now a skill that everyone must have to enter the job market, so everyone must demonstrate proficiency in this area; third, the huge growth in the evolution of the use of computers has drastically decreased costs of hardware and software (Charr, 1988). When these three forces are evolved, it is obvious how EUC has radically gained in recognition and how progressive computer technology has grown to what we know and use today.

The computer has developed into a commodity as common as any household item, like the food processor or even the vacuum cleaner. Despite slightly higher costs, people all over the world are finding the investment increasingly beneficial. The number of things one can accomplish in the comfort of one's home, even with a little PC, is incredibly large. Hence, it is no surprise that the term EUC has earned its position in the world of computer jargon.

Management Issues of EUC

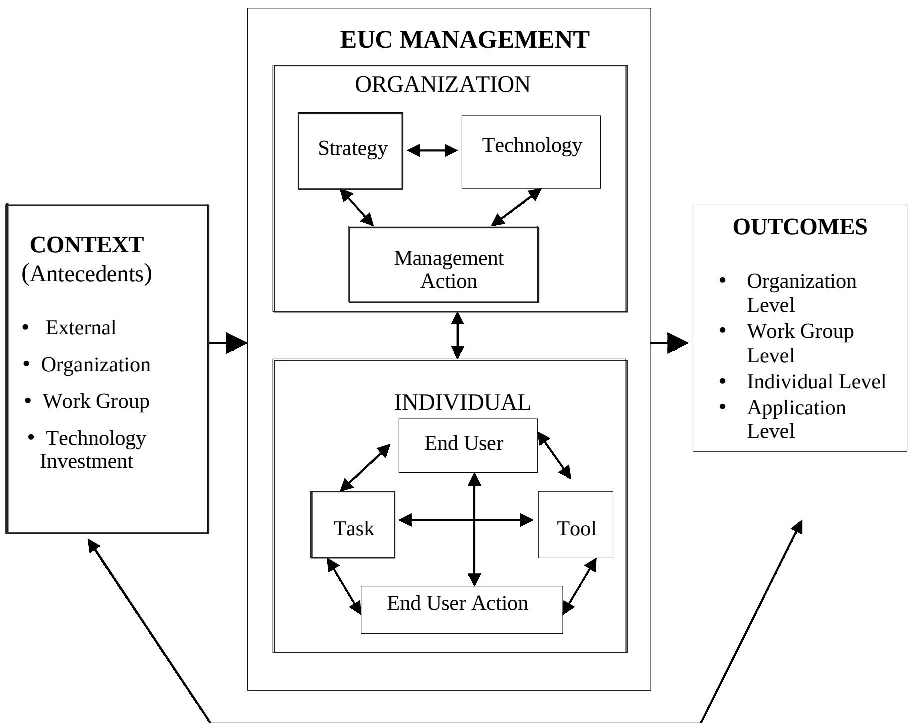
Even with the popularity of EUC, the greatest debate substantiated through this capability is: how much will a business benefit from this process, and can certain disadvantages be posed to limit the effectiveness of EUC implementation? This paper is structured to further provide a background on the evolvement of EUC, based in part on research conducted by Brancheau and Brown (1993), to establish a framework which will finally provide the necessary information and address important management issues and perspectives on EUC.

EUC Management Model

EUC can be defined as "the adoption and use of IT personnel outside the IS department to develop software applications in support of organizational tasks" (Brancheau & Brown, 1993). The content of Brancheau and Brown's study was generated from different journal articles and upon conclusion of their findings they developed a model, incorporating all of the aspects identified with EUC (see Figure 1).

The model includes three major components: antecedents (contexts), behavior (EUC management) and consequences (outcomes). EUC management is further divided into two customized sections, organizational EUC management and individual EUC management. The organizational part of the model represents the planning, organizing, staffing, directing, controlling, supporting, and coordinating functions associated with EUC management. As the model depicts, three relationships occur involving strategy, technology, and management action. Four factors are incorporated within the individual EUC management: the end user, task, tools and end-user action. The antecedents incorporate those factors pertaining to external, organization, workgroup, and technology investment activities as related to the business. Through the integration of EUC, both individual and organizational management styles, four levels of outcomes occur, organization, workgroup, individual and application (Moore & Powell, 2002). With the background pertaining to this model, we propose that four management issues should be addressed: 1) what are the characteristics associated with End-User Computing? 2) what are the necessary tools needed? 3) what are the advantages gained through EUC? and 4) what are the possible risks and security issues associated with EUC?

Figure 1. EUC Management Model



What are the Characteristics Associated with EUC?

The characteristics of the EUC can be explored in three areas: applications, development skills and prototyping. The details are explained in the following sections.

End-User Computing and Applications

Early on, EUC was identified as a spreadsheet-oriented operation, repetitively calculating large quantities of data and then later graduated to incorporate operational systems throughout entire organizations (O'Donnell & Sanders, 2003). Approximately ten years later, from the mid-1990's and beyond when graphical user interfaces and 4GL's advanced in complexity, EUC was elevated to include large database systems, known as data warehouses, and Internet Web sites (O'Donnell & Sanders, 2003). The impact was delivered throughout the EUC establishment, where individual applications can now include a variety of group applications. EUC initially began as only an inter-organizational intermediary, but it can now be able to accommodate a wider spectrum of implementations, consisting of inter-organization and organization-to-customer based systems on the Internet (O'Donnell & Sanders, 2003; Christoff, 1990).

End-User System Development Skills

It is vitally important before a DSS is designed, that the end-user is first identified by the end-user developer. The end-user's applications have to be analyzed initially, based upon what they will do with the software and the technical knowledge that they possess. This establishment will govern how complex the program will be for the end-user's and customize it to fit their individual needs. End-user developers consist of three types (from the least to the most in complexity and technical skill level): menu-level developers, command-level developers, and end-user programmers (O'Donnell & Sanders, 2003).

Prototype Development

The characteristic that defines one of EUC's unique advantages is "flexibility," which basically explains its purpose in the technical software world and is encapsulated by the implementation of prototyping. To define prototyping's specific parameters, a prototype software program must meet these qualifications: quick to create, easy of change, inexpensive to build, and inexpensive to discard (Carr, 1988; Gelbstein, 2003). Advantages of prototyping are severely weakened if any of these requirements are not successfully met.

What are the Necessary Tools Needed?

End-users must have the necessary tools available to them, so that they can develop specific software satisfying their individual needs. The categories of tools consist of three major types geared toward specific requirements within the business. The three varieties include: office productivity tools, workgroup computing tools and application development tools (Bruni & Stigliano, 2003).

The objective of office productivity tools is to enhance the productivity of individuals, by providing computer support for common office tasks. Typical with the commonalities of the business sector, these tools include all of the relevant software applications, such as word processing, database management, spreadsheets, presentation graphics and other similar programs. With a click of the mouse button, the series of functions will orchestrate, illustrating a custom application on the user's worksheet.

In the office environment, the effectiveness of groups is widely understood with workgroup computing tools to support their specific needs. Specifically, the tools "support the efforts of groups of users working on related tasks," unlike office productivity tools, which are specific to a single user (Bruni & Stigliano, 2003). Examples of tools under this category include, but are not limited to: email, group calendars, workflow systems and other similar applications, providing for effective communication, among employees.

The final division of development tools that are vital for the successful implementation of EUC in a business are application tools. Application tools consist of four types: application generators, fourth-generation languages, application development environments and scripted web pages, functioning to "develop reusable programs with which end users create custom applications" (Bruni & Stigliano, 2003). Essentially, these are the most crucial tools of the three categories, since they support the capability for the formal development and maintenance of computer-based applications, along with software problem resolution and customer support.

What are the Advantages Gained Through EUC?

In today's environment, every corporate advantage that can be gained, invites the potential for greater revenue achievement, promoting a greater capture of market share. In respect to obtaining distinct advantages, EUC encourages another important behavioral application in addition to "flexibility," by increasing a program's "usability." There are many advantages for a usable software: reduced training costs, reduced support and service costs, reduced error costs, increased productivity of users, increased customer satisfaction, and increased maintainability (Hohmann, 2003; Holsapple & Whinston, 1996).

Based on the premise of EUC that the developer is also the user, a key understanding to what the end-user ultimately wants with the product can be achieved. Once the groundwork is laid out, product development can proceed as usual, with fewer modifications and/or errors.

What are the Possible Risks and Security Issues Associated with EUC?

Potential risks are involved in a system development with its integration into the company. Four main areas are: poorly aimed systems, poorly designed and documented systems, inefficient use of organizational resources, and loss of security (O'Donnell & Sanders, 2003; McLeod & Schell, 2001). It is logical to determine how each of these risks can limit success upon inviting EUC into the business atmosphere. The purpose of EUC is to mediate the software development problems that are associated with a poorly aimed, designed, and documented computing system, to provide the ultimate end-user with certain specified capabilities (Lucas, 2000). EUC is a system to provide for that allowance of software customization and enhanced programming communication.

It is evident that in today's world, there is always technical evolution occurring, between computer security systems, increasing in complexity to counteract the continual updated attempts by software hackers to disrupt the software information. This kind of "predator & prey" relationship is an ongoing process and certainly impacts and threatens the business and the EUC software applications. What is essential to know is that the consequences of a business not having EUC software can greatly outweigh the potential security aspect that looms over all types of software programs. Only the business itself can determine that decision and judge the possible outcomes it may encounter, depending upon which decision is chosen.

Costs/Benefits Analysis of EUC

The costs of EUC are evident and intensive studies have been completed to examine and determine what these costs entail. The research conducted by James, Ang, Joo-Eng and Shaw (2003) uncovers deep concerns involved in this type of application. They discovered that users without adequate technical knowledge did not conduct activities that required a greater degree of technicality. Furthermore, due to these inadequacies, these users "expected more from the PC, viewing it as a task enhancing tool" (James et al., 2003). These findings illustrate a very essential point relating to EUC which is regardless of the software program's abilities, if the employees lack the proper expertise to tap into these aspects of the program, the benefits of EUC will never be recognized.

No matter how well designed, constructed, and tested a system or application may be, errors or bugs will inevitably occur. The systems developed by the end users must be maintained, and the end users' knowledge and skills must be updated periodically as part of the maintenance. Along with training and maintenance, the end-user system will be operational if the organization moves to the next-level programming language or update its software applications and hardware.

The benefits of EUC software and its applications are profound, giving modern businesses the distinct edge on their competitors. This provides for a greater level of independence with each employee, so that they do not have to directly rely on a software specialist to solve these problems.

One specific characteristic of EUC is that end-users may work directly with software development specialists to organize and design applications which include attributes needed in the program but allows them the "flexibility" to make any necessary manipulations in the future. This benefit substantiates in a business only having to purchase one program which will adapt to a variety of applications, saving time and money.

PERSPECTIVES ON EUC

The applications of EUC are judged at two distinct levels based on which area in the software development process is effected, either the end-user or the IS specialist. Both of these perspectives are outlined below, detailing the importance of this capability for both constituents.

The End User Perspective

In what perspective do end users view EUC? Response to the growth in EUC has been extremely positive, especially from the perspective of the common end user. From Word Processors to Database Management Systems, and from Decision Support Systems to Executive Information Systems, EUC has broken down all barriers between the layman and the computer. The dependence on computers has grown tremendously. The end user perspective is inclined towards formal as well as informal means of computation.

The IS Specialist Perspective

From this angle, EUC has also had ample support. IS specialists initially would solve the layman's problems from start to end, generating the required output on the computer. With the growth in EUC, end users began gaining independence and so the IS specialists had to adopt the job of training end users, helping them gain some control over computing functions. Information Centers have cropped up in almost all computer-using organizations, where IS specialists perform a variety of functions, from hardware and software purchasing and maintenance to end user support and training functions. It has been argued many times that EUC might be a threat to IS professionals, but upon reflection, it seems unlikely. The end user would be someone whose career focus would be something different from technology itself. He would, however, be using technology as a career and would constantly be a source of assistance to the common end user. Hence, it is important to note the role played by the IS specialist in EUC and conclude that the two categories complement, rather than oppose each other (Alter, 1999).

FUTURE TRENDS AND CONCLUSION

With the continuing growth in technology and in the capacities of the computer, the trend for the future seems to be towards more emphasis on user-friendliness and ease of handling (Morre & Powell, 2002). EUC has an extremely positive future, even considering the general technology trends - the technology connecting the information resource to the end user. In the future, organizations would be better off moving towards an investment in IT as an integral part of the organizational constitution rather than considering it a functional cost. With such an attitude developing in the business world, EUC has a bright future with many expanding horizons. This would be due to the fact that IS/IT would become such an integral part of the creation and existence of organizations, and that ultimately its role would not have to be questioned or analyzed.

As it has been described in this article, EUC has an essential application associated with software development and customization that greatly increases a businesses competitive advantage upon implementation. Specifically, it reduces a business cost in relation to the fewer number of programs that as a result need to be purchased to maintain the corporation's vitalities. It seeks to further reduce time and complexities through the provision of manipulating software applications around an employee or department's individual needs. Although providing employees with this specific type of software poses the possibility for security problems, this risk is always a given threat with all software programs. The success of EUC is weighed heavily upon many variables, and various tools have to be available to expand on all of its capabilities. EUC software development can be made more efficiently, when the software vendor can establish a close link of communication with the final end-user and specifically identify and outline their requests. Perspectives on end-user computing have certainly been well supported among both the end-users and the IS specialists, rooting its existence in present day business strategies. Although its history has evolved in less than three decades, end-user computing has established itself in the IT environment and will only further progress to ever increasing heights in the future.

REFERENCES

Alter, S. (1999). *Information systems: A management perspective*. Addison-Wesley.

Brancheau, J. & Wetherbe, J. (1993). The adoption of spreadsheet software: Testing innovation diffusion theory in the context of end-user computing. *Information Systems Research, 1*(2), 115-143.

Brown, C. & Bostrom, R.(1993). Organization designs for the management of end-user computing: Reexamining the contingencies. *Journal of Management Information Systems, 10*(4), 183-211.

Bruni, M. & Stigliano J. (2003). End-user computing tools. In *Encyclopedia of Information Systems* (Vol. 2, pp. 127-139).

Carr, H. (1998). *Managing end user computing.* Prentice Hall.

Christoff, K. A. (1990). *Managing the information center.* Scott, Foresman and Company.

Gelbstein, E. (2003). End-user computing, managing. *Encyclopedia of Information Systems* (Vol. 2, pp. 115-126).

Hohmann, L. (2003). Usability: Happier users mean greater profits. *Information Systems Management*, 66-76.

Holsapple, C. W. & Whinston, A. B. (1996). *Decision support systems.* West Publishing Company.

James, S.K. Ang, L-P., Joo-Eng & Shaw, N. (2003). Understanding the hidden dissatisfaction of users toward end-user computing. *Journal of End User Computing, 15*(2), 1-22.

Laudon, K.C. & Laudon, J.P. (2003). *Essentials of management information systems.* Prentice Hall.

Lucas, H.C. Jr. (2000). *Information technology for management* (7th Ed). Irwin McGraw-Hill.

McLeod, R. Jr. & Schell, G. (2001). Management Information Systems (8th Ed). Prentice-Hall.

Morre, Jo E. & Powell, A. (2002). The focus of research in end user computing: Where have we come since the 1980's. *Journal of Organizational and End User Computing*, *14*(1).

O'Donnell, J.B. & Sanders, L.G. (2003). End-user computing concepts. In *Encyclopedia of Information Systems (Vol.* 2, pp. 101-113).

KEY TERMS

Database Management System (DBMS): A software program (or group of programs) that manages and provides access to a database.

Decision Support Systems (DSS): A computer-based information systems whose purpose is the support of (not replacement) decision-making activities.

End-User Computing (EUC): Direct, hands-on use of information systems by end user whose jobs go beyond entering data into a computer or process transactions.

End-user Development: The development of information systems by end users rather than information system specialists.

Hacker A very knowledgeable computer user who uses his or her knowledge to invade other people's computer through the computer network.

Information Systems (IS): A computer-based system helps people deal with the planning for, development, management, and use of information technology tools to help them perform all tasks related to their information needs.

Information Technology (IT): A computer-based tool (hardware and software) that people use to work with information and support the information-processing needs of an organization.

Organizational Hypermedia Document Management Through Metadata

Woojong Suh
Inha University, Korea

Garp Choong Kim
Inha University, Korea

INTRODUCTION

Web business systems, the most popular application of hypermedia, typically include a lot of hypermedia documents (hyperdocuments), which are also called Web pages. These systems have been conceived as an essential instrument in obtaining various beneficial opportunities for CRM (customer relationship management), SCM (supply chain management), e-banking or e-stock trading, and so forth (Turban et al., 2004). Most companies have made a continuous effort to build such systems. As a result, today the hyperdocuments in the organizations are growing explosively.

The hyperdocuments employed for business tasks in the Web business systems may be referred to as organizational hyperdocuments (OHDs). The OHDs typically play a critical role in business, including the forms of invoices, checks, orders, and so forth. The organization's ability to adapt the OHDs rapidly to ever-changing business requirements may impact on business performance. However, the maintenance of the OHDs increasing continuously is becoming a burdensome task to many organizations; managing them is as important to economic success as is software maintenance (Brereton et al., 1998).

An approach to solve the challenge of managing OHDs is to use metadata. Metadata are generally known as data about data (or information about information). Concerning this approach, this article first reviews the previous studies and discusses perspectives desirable to manage the OHSs and then provides metadata classification and elements. Finally, this article discusses future trends and makes a conclusion.

BACKGROUND

The hyperdocument is a special type of digital document based on the interlinking of nodes such as multimedia components and sets of data elements derived from databases. For digital document, metadata have typically been employed for the access to media- and application-specific documents, such as for information discovery (Anderson & Stonebraker, 1994; Glavitsch et al., 1994; Hunter & Armstrong, 1999). Also, most of the previous studies on metadata for hyperdocuments have also been interested in information discovery from a content-oriented perspective (Lang & Burnett, 2000; Li et al., 1999; Karvounarakis & Kapidakis, 2000). Especially, a set of hyperdocument metadata, the Dublin Core (Dublin Metadata Core Element Set) (Weibel et al., 1995; Weibel & Koch, 2000), has been paid attention to as a standard for Web information resources and also focuses on the information discovery. However, besides this perspective, for the OHDs metadata, the organizational perspectives also need to be considered to satisfy various managerial needs of organizations.

First, a process-oriented perspective needs to be considered. It is also pointed out that the perspective needs to be reflected on defining metadata of corporate digital documents (Murphy, 1998). OHDs as corporate digital documents are closely related to business tasks and information for them in an organization. Generally, corporate documents are produced in undertaking an organizational process (Uijlenbroek & Sol, 1997); furthermore, most businesses are based on, or driven by, document flow (Sprague, 1995). Thus, documents and business processes may be considered simultaneously in the analysis of a corporate information system (Frank, 1997). In this context, the OHDs may affect the speed of communications to perform business process. Accordingly, the OHDs should be designed to support collaboration among workers in business processes. Also, the OHDs can be rapidly improved to fit ever-changing business requirements.

Second, the metadata for OHDs are to be considered from a technical perspective. The system resources linked to the OHDs, such as program files and data components dynamically cooperated, are a considerable part of the organizational assets. The links between such resources and OHDs are very complex. Accordingly, managing the resources and the links through metadata can result in the efficient use of the organizational asset; the metadata related to the technical components can help developers change and improve the OHDs more efficiently.

Third, in the long term, the metadata role of OHDs should be extended toward organizational memory (OM), because organizational digital documents are a major source of OM (Murphy, 1998; Sprague, 1995). The OM techniques concentrate on managing an organization's information or knowledge of the past (Stein & Zwass, 1995; Wijnhoven, 1998). The metadata for OHDs need to play a critical role in managing a variety of histories in terms of business functions, communication mechanisms, technical artifacts, and contents. The memory may provide knowledge to support various decisions for controlling communication mechanisms in a business process, linking to the previous responsible workers, or maintaining the hypermedia applications.

Considering all the perspectives discussed previously, metadata roles for OHDs can be summarized in three levels--operation, system, and organization--as shown in Table 1. In fact, we believe that these roles can also be applied to other kinds of corporate digital documents.

METADATA CLASSIFICATION AND ELEMENTS FOR OHDS

Metadata classification can be perceived as a fundamental framework for providing metadata elements. According to the our perspective on the metadata for OHDs described in the previous section, the following categories of metadata need to be considered:

- *Content-dependent Metadata*: These metadata are used to enable understanding of the content of documents. The metadata include information that depends on (i) the content directly, and (ii) semantic meanings based on the content of the document indirectly.
- *Workflow-dependent Metadata*: These metadata provide information about workflow related to an organizational hyperdocument. These metadata are concerned with process-related factors such as workers, tasks, or business rules.
- *Format-dependent Metadata*: These metadata describe information on formats related to organizational hyperdocuments as well as hypermedia components such as nodes, anchors, interface sources, and database attributes.
- *System-dependent Metadata*: These metadata provide information concerned with storage- and software-related information on system resources such as hyperdocuments, interface sources, and databases.
- *Log-dependent Metadata*: These metadata describe information on the history and the status of organizational hyperdocuments.

Content-dependent metadata are essential for discovering information in OHDs. Workflow-dependent metadata can contribute to increasing the ability to control business processes through the efficient adaptation of OHDs to ever-changing organizational requirements. Format-dependent metadata can provide an understanding of the hypermedia features in terms of structures and operational mechanisms, so that they can be useful in the technical maintenance of OHDs. System-dependent

Table 1. Metadata roles for OHDs

Level	Metadata Roles
Operation	• Easy and fast access • Increased accuracy
System	• Interoperability under heterogeneous environment • Document maintenance • Document distribution
Organization	• Increased reusability of information and knowledge resources • Increased capacity of business management • Increased organizational memory

Table 2. Metadata elements of OHDs

Classifications	Elements
Content-dependent	[Document] Title, Description, Document Domain Name, Conceptual Attribute Name [Anchor] Name [Data Node] Title [Interface-Source] Name
Workflow-dependent	Task Domain Name, Task, Agent Domain Name, Agent Object Name, Business Rule
Format-dependent	[Document] Type [Anchor] Type [Node] Type, [Data Node] Property [Interface-Source] Property [DB] Physical Attribute Type
System-dependent	[Document] File Name, H/W Name, Location Path, S/W Technology [Data Node] File Name, H/W Name, Location Path [Interface-Source] File Name, Storage, Location Path [Database] Name, H/W Name, Location Path, Table Name, Table Type, Physical Attribute Name, DBMS Name
Log-dependent	Document Number, Version Number, Loading Date, Withdrawal Date, Update Date, Update Description, Director, Operator

metadata can also play a critical role in technical maintenance by providing information on hardware and location, and software technologies applied to the hyperdocuments. This meta-information is essential for sharing and reusing system resources. Finally, log-dependent metadata may contribute to organizational memories. Thus, the metadata in this category should be specified in order to capture the history of OHDs. According to this classification, detailed metadata elements may be specified under the classification suggested in this article, as shown in Table 2.

Content-dependent classification consists of elements that enable users to understand the content of the hyperdocuments. The document domain may be in terms of content and roles. The conceptual attributes, as data items represented on a hyperdocument, are connected to a corporate database. Interface sources are primarily multimedia components such as image or animation that are represented on interfaces.

A node, an essential factor of hypermedia, has been defined as the fundamental unit of hypertext (Nielsen, 1993), fragments of information (Fluckiger, 1995), or basic information containers (Schwabe & Rossi, 1994). This article defines a node as any navigational object with hyperlinks. An object may a type of media, such as image, sound, animation, or a hyperdocument itself. Nodes may be categorized into two types from the perspective of their properties: document nodes and data nodes. Nodes are also of two types from the perspective of link directions: source and destination. The fundamental definitions for nodes are summarized in Table 3.

An interface may consist of one or more hyperdocuments. Accordingly, a document node, a hyperdocument, can be either only a part of an interface or an interface itself. From these definitions, the element of node type in format-dependent metadata may take a document node or data node as its value.

The information of a hyperdocument in terms of a process can be obtained effectively by the use of a document-based workflow concept. The workflow con-

Table 3. Types of nodes

Perspectives	Types	Descriptions
Properties	Document Node	A unit of an HTML document, which may be a whole interface or a part of it.
	Data Node	A unit of multimedia data that may be accessed from a document node.
Link Direction	Source Node	Nodes that can access a current node.
	Destination Node	Nodes that a current node can access.

cept typically includes common essential factors in terms of a unit of a work, a tool of a work, and a person for a work (Lee & Suh, 2001). In the document-based workflow approach, an OHD is regarded as a tool of a work. A task, as a work unit consisting of a workflow, may be described as operations or descriptions of human actions with a hyperdocument. An agent refers to a person who performs the task, and is expressed by hierarchical status in an organization. An agent domain can be defined as a group of agent objects having common tasks or organizational objectives. The agent domain can be typically conceived as a department of an organization. The task domain is a set of tasks corresponding to an agent domain.

The format-dependent metadata are concerned with type or properties of hyperdocuments, anchors, nodes, interface sources, and databases. The types of anchors can be static or dynamic depending their value. The definitions of these types are as follows:

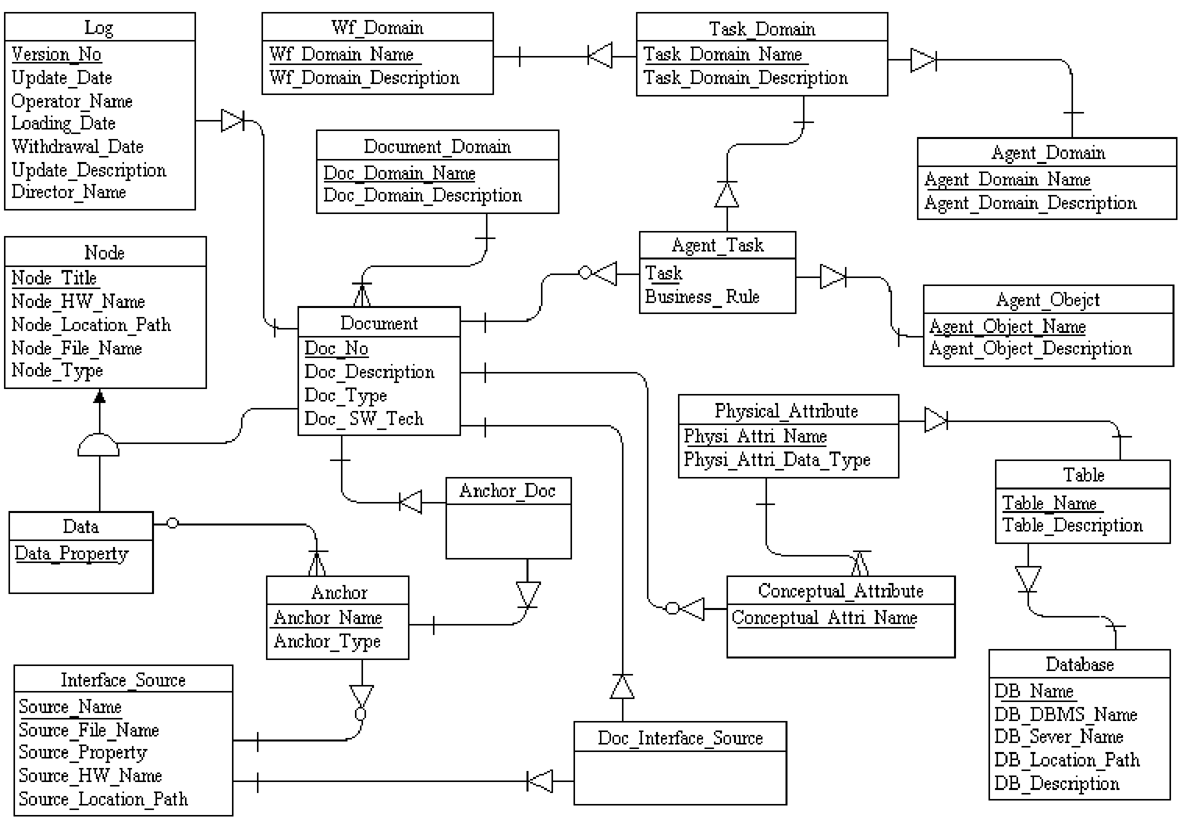
- *Static anchor:* One fixed in a hyperdocument.
- *Dynamic anchor:* One generated by data stored in a database; that is, it refers to data transformed into and represented as an anchor when the data are accessed by a hyperdocument according to any event that occurs as a function or another anchor.

The types of OHDs can be categorized into three: control, processing, and referential, according to their roles in a hypermedia application. These types are defined as follows:

- *Control Type:* Hyperdocuments that typically guide users to other hyperdocuments of processing or referential types. Homepages or index pages are examples of this type.
- *Processing Type:* Hyperdocuments that typically contain data attributes connected with a database in the style of a form.
- *Referential Type:* Hyperdocuments that provide supplementary information about work instructions, business rules, news, or products.

Properties of interface sources are multimedia properties such as images or animation. The properties of data nodes are the same as those of interface sources. The

Figure 1. Metadata DB schema of OHDs



physical attribute type of a database implies the data properties of the attribute.

System-dependent metadata focus on storage-related information. The storage-related information can be found in various applications, but they are not integrated, so it is difficult to create a synergetic effect. However, if metadata of all the components of hypermedia systems, such as hyperdocuments, data nodes, interface sources, and databases, are integrated into a repository, it is possible to manage a hypermedia system effectively. Software technology is a major factor in determining the capacity and characteristics of a system. Recently, for example, a variety of emerging software technologies, such as ASP (Active Server Page), Java scripts, Visual Basic scripts, or Perl, have had a considerable impact on the improvement of hypermedia systems. Accordingly, the information on software technologies applied to a hyperdocument may contribute to the maintenance of a hypermedia application.

Log-dependent metadata are used for tracing the history of hyperdocuments for the maintenance of their system. Although there may be log information captured automatically by an operating system or an application program, it is typically separated, so it is difficult to obtain a synergetic effect in maintenance. Furthermore, it is insufficient to maintain a hypermedia system effectively. Therefore it is necessary to capture the information about changes of hyperdocuments synthetically. Some hyperdocuments may be operated temporally, depending on their purposes. Accordingly, version- or time-related information should be managed. The loading date is a date pushing a hyperdocument into its system. The withdrawal date is a date removing a hyperdocument from the system for the expiration of its periodic role or for updating. Information on responsible operators and directors for a hyperdocument may be required for responsible management, or questions by new staff members.

The metadata elements in Table 2 can be designed as shown in Figure 1. The schema was implemented in a meta-information system for the maintenance of OHDs (Suh & Lee, 2002). The system consists of two main modules: metadata management and a supporting module. The metadata management module is responsible for metadata handling such as creating, editing, or deleting. The supporting module serves two types of functions: searching an OHD and reporting its meta-information.

FUTURE TRENDS

Concerning the challenge to cope with a flood of OHDs, today content management system (CMS) is drawing great attention; the market size of content management solutions are forecasted to grow explosively (WinterGreen Research, Inc., 2003). CMSs help organizations develop Web applications, and support content lifecycle including the following phases: creation, approval, publishing, deployment, delivery and removal. Now, CMSs are conceived as an essential infrastructure for not only developing but also maintaining Web applications; most Web projects include implementing CMSs along with Web applications. Furthermore, CMSs are often incorporated into the existing infrastructures of Web applications to support the maintenance of the applications.

The lifecycle of content is supported primarily by the engines of workflow and personalization on the basis of a repository that includes the metadata as well as actual content. The metadata are a critical component required to operate those engines and other functions of CMSs, so their importance has been addressed (Perry, 2001). In fact, current CMS solutions originate from some different backgrounds, so that the functional features and metadata schema are also different from each other. Accordingly, in the near future, the metadata standard may be required in the CMS solution industry as it has been in other application domains of information technologies so far.

CONCLUSION

Recently, many organizations have expanded their business workplaces through Web business systems. Organizational hyperdocuments (OHD), critical information resources, are growing explosively in such organizations. Accordingly, the maintenance of these documents is becoming a burdensome task. For this challenge, this article offers an approach based on metadata required from the managerial perspectives as well as content-oriented. The managerial perspectives address the adaptability of Web business systems to the ever-changing business requirements and the efficiency of developers' works. The presented metadata are expected to help manage organizational memory in the long term through implementation into a metadata-based system.

REFERENCES

Anderson, J.T., & Stonebraker, M. (1994). Sequoia 2000 metadata schema for satellite images. *ACM SIGMOD Record*, *23*(4), 42-48.

Brereton, P., Budgen, D., & Hamilton, G. (1998). Hypertext: The next maintenance mountain. *IEEE Computer*, *31*(12), 49-55.

Fluckiger, F. (1995). *Understanding networked multimedia: Applications and technology*. Englewood Cliffs, NJ: Prentice Hall.

Frank, U. (1997). Enhancing object-oriented modeling with concepts to integrate electronic documents. *Proceedings of the 30th Hawaii International Conference on System Sciences, 6,* 127-136.

Glavitsch, U., Schauble, P., & Wechsler, M. (1994). Metadata for integrating speech documents in a text retrieval system. *ACM SIGMOD Record, 23*(4), 57-63.

Hunter, J., & Armstrong, L. (1999). A comparison of schema for video metadata representation. *Computer Networks, 31,* 1431-1451.

Karvounarakis, G., & Kapidakis, S. (2000). Submission and repository management of digital libraries, using WWW. *Computer Networks, 34,* 861-872.

Lang, K., & Burnett, M. (2000). XML, metadata and efficient knowledge discovery. *Knowledge-Based Systems, 13,* 321-331.

Lee, H., & Suh, W. (2001). A workflow-based methodology for developing hypermedia information systems. *Journal of Organizational Computing and Electronic Commerce, 11*(2), 77-106.

Li, W., Vu, Q., Agrawal, D., Hara, Y., & Takano, H. (1999). PowerBookmarks: A system for personalizable Web information organization, sharing, and management. *Computer Networks, 31,* 1375-1389.

Murphy, L.D. (1998). Digital document metadata in organizations: Roles, analytical, approaches, and future research directions. *Proceedings of the 31st Hawaii International Conference on System Science, 2,* 267-276.

Nielsen, J. (1993). *Hypertext and hypermedia.* Boston: Academic Press Professional.

Perry, R. (2001). Managing the content explosion into content-rich applications. *Internet Computing Strategies, Report, 6*(2). Yankee Group.

Schwabe, D., & Rossi, G. (1994). *From domain models to hypermedia applications: An object-oriented approach.* Technical Report MCC 30/94. Dept. of Information, PUC-Rio.

Sprague, R.H. (1995, March). Electronic document management: Challenges and opportunities for information systems managers. *MIS Quarterly, 19,* 29-49.

Stein, E.W., & Zwass, V. (1995). Actualizing organizational memory with information systems. *Information Systems Research, 6*(2), 85-117.

Suh, W., & Lee, H. (2002). Managing organizational hypermedia document: A meta-information system. In K. Siau (Ed.), *Advance topics in database research* (vol. 1, pp. 250-266). Hershey, PA: Idea Group Publishing.

Turban, E., Lee, J., King, D., & Chung, H.M. (2004). *Electronic commerce 2004: A managerial perspective.* Prentice Hall.

Uijlenbroek, J.J.M., & Sol, H.G. (1997). Document based process improvement in the public sector: Aettling for the second best is the best you can do. *Proceedings of the 30th Hawaii International Conference on System Sciences, 6,* 107-117.

Weibel, S., Godby, J., Miller, E., & Daniel, R. (1995). OCLC/NCSA metadata workshop report. Retrieved March 21, 2004, from *http://www.oasis-open.org/cover/metadata.html*

Weibel, S., & Koch, T. (2000). The Dublin Core metadata initiative: Mission, current activities, and future directions. *D-Lib Magazine, 6*(12). Retrieved March 21, 2004, from *http://www.dlib.org/dlib/december00/weibel/12weibel.html*

Wijnhoven, F. (1998). Designing organizational memories: Concept and method. *Journal of Organizational Computing and Electronic Commerce, 8*(1), 29-55.

WinterGreen Research, Inc. (2003). *Content management market opportunities, market forecasts, and market strategies, 2004-2009.* Report.

KEY TERMS

Guided Tours: A navigation type that leads users to a predefined trail of nodes without freely explorative navigation, using, for example, previous and next anchors.

HTML: HyperText Markup Language. It is markup language using tags in pairs of angle brackets, for identifying and representing the Web structure and layout through Web browsers; it is not a procedural programming language like C, Fortran, or Visual Basic.

Hyperspace: Information spaces interlinked together with hypermedia structures. Concerning World Wide Web, cyberspace is sometimes used instead of hyperspace.

META Tag: A type of HTML tags with a word, “META”. Web “spiders” read the information contained within META tags to index Web pages. The information consists of keywords and a description of its Web site.

Open Hypermedia: As a more generalized concept addressing interoperability among hypermedia services, it has the following characteristics: it should be platform independent and distributed across platforms, and users should be able to find, update, make links to, and exchange the information, unlike hypermedia titles on CD-ROM.

SGML: Standard Markup Language. Document standard from ISO (reference ISO 8897). It is a meta-language that can define document logical structure by using Document Type Definition (DTD) component. An example of document types defined using the DTD of SGML is HTML.

XML: eXtensible Markup Language. It is quite different from HTML in that XML gives document authors the ability to create their own markup. XML is flexible in creating data formats and sharing both the format and the data with other applications or trading partners, compared with HTML.

Organizing Multimedia Objects by Using Class Algebra

Daniel J. Buehrer
National Chung Cheng University, Taiwan, R.O.C.

WEB-BASED DEVELOPMENT

Web-based applications (Web services and service-oriented architectures) can be run via a Web-based browser. There are several approaches to writing such Web-based applications. A lightweight approach is suitable for hand-held devices. In this approach, a Java servlet or a JSP page (Java 2 Platform, JSP), or an ASP application (Microsoft .NET, ASP) generates HTML (Hypertext Markup Language), XHTML, or XML documents (W3C Semantic Web Activity, XHTML, XML) to be displayed by the browser. Most browsers use an anchored URLs extension (e.g., .doc, .jpg, .xml, etc.) to choose an appropriate plug-in to display the URL when it is clicked. Besides displaying text and multimedia, Web servers and/or browsers can also execute Java applets or scripting languages to read and/or change persistent data. Previously, about 98% of these data were stored in relational or object-relational databases. However, recently more of these data are being stored in XML-based documents. Often these documents have an associated "schema" declaring the nesting of tags and the types of primitive values, or an "ontology" (Everett et al., 2002, Hunter, 2003) declaring classes, attributes, and relations that are used in the document.

Databases, XML documents, and knowledge bases have traditionally been only a small part of a company's programming environment. The goal of most development environments is to use UML or entity-relation diagrams to analyze and design suitable classes, design patterns, and active objects from other existing ones. Web-based development environments (Stal, 2002) help us to find suitable classes, design patterns, and active objects on the Web, perhaps for a fee, or perhaps for a limited development time. That is, in the future, the development environment itself will be built on a distributed knowledge base.

Will there be a programming language, and will it still have explicit calls to SQL? Currently, scripting languages have their own unique means of accessing SQL databases, and the whole area of Web-based programming and multimedia networking has become quite messy (Weinstein & Gelman, 2003). Although J2EE's Container-Managed Persistency (CMP), JDO, and many object-relational development environments provide ways to hide SQL calls, security, and transactions, many applications may still choose to use the richer functionality of direct JDBC.

Web-based development environments, however, are forcing some changes in the way that persistent data are stored. Heterogeneous databases can be said to have failed to provide a sufficient platform for sharing data because of the difficulty in integrating so many SQL meta-data definitions. There was no "class hierarchy" to organize the definitions of the SQL tables, and small incompatibilities in SQL definitions often caused major headaches. Although ODBC and JDBC are now SQL standards, they permit various implementations of the meta-data, and they provide no standard way of "including" meta-data definitions from other "superclass" tables. In order to counter this lack of a class hierarchy in object-relational databases, the Semantic Web Project is proposing to use XML-based documents to share data and their declarations.

How will the Web-based development environments and their applications access Web databases? Will they access SQL, XML, or both? The answer is almost certain to be both, as well as other spreadsheets, natural language documents, and multimedia.

ISSUES IN INTEGRATING XML AND RELATIONAL DATABASES

As long as the application-tier, Web-tier or client-tier programs access the data via a standard interface, it makes little difference to the programmer if the persistent data are stored in SQL databases or XML files. This interface should take care of transactions (Ahamad & Chelliah, 2002) and security, sequencing set/get/add/remove calls from multiple users and doing load balancing and fail-over across multiple servers.

J2EE from the Java Community Process and Microsoft's .NET have the two major candidates for this programmer interface. One goal of these platforms is to hide the SQL and XML calls behind standardized set/get/add/remove calls for enterprise components. However, SQL programming and XML programming are currently fairly incompatible. There are several XML query languages for querying XML documents, and XSLT can be used to reformat XML documents. There are numerous XML schema mechanisms to enforce some type-checking, but higher-level schema mechanisms, such as RDF-Schema and OWL Ontologies (W3C RDF, OWL), are really needed before attribute and relation assignments can be strongly type-checked. OWL declarations essentially take the place of Corba IDL declarations, OMG ODL declarations, or SQL meta-data tables, by declaring ontology namespaces, a class IS-A hierarchy, attribute types and relation domains/ranges.

CLASS ALGEBRA'S ROLE

Class algebra (Buehrer, 2001; Buehrer & Chien, 2003; Buehrer & Lee, 1999; Buehrer, Tse-Wen & Chih-Ming, 2001; Buehrer, Tse-Wen, Chih-Ming & Hou, 2001; Buehrer & Wang, 2003) can serve as the theoretical basis for type-checking both OWL-based and SQL-based set/get/add/remove calls. Each SQL table is considered to be a class, and each line of the table is an object. The primary key of the object is used, along with the JNDI database service name and the table name, as a unique object identifier on the network. Each join of a foreign key to the associated primary key corresponds to a binary relation in class algebra.

For example, suppose that the user wants to find all images of Tom and his dog. One could locate Tom either by using the query @People{firstname= "Tom"} to find all people named Tom, or by following relations from his school, friends, or other objects to locate Tom. Then his pet dog(s) can be found by following the "hasPet" relation and limiting the answer to dogs. The intersection of the images containing Tom and the images containing his dog will be the images containing both Tom and his dog, as given by the query "query1@= @People{firstname= "Tom"}; query2@=query1.inImages @* (query1.hasPet @* @Dog) . inImages", where "inImages" is the relation from objects to the images that contain them and @* is the class intersection operator. These class algebra queries can be translated into SQL or XQuery, depending on where the data are stored. An interactive environment allows the user to search the IS-A hierarchy for a particular class, and then recursively constrain some of the attribute or relation ranges. The user can then combine various subqueries with Boolean operations, as given previously. The major advantage of class algebra queries is that the range is known. For example, query1 is People and query2 is Images. Therefore, all subclasses and their attributes and relations can be displayed when creating further queries. Moreover, a query may make use of the containment algorithm to optimize (Beneventano, Bergamaschi & Sartori, 2003) the processing of the query.

There are many details of the definition of class algebra that may be incompatible with various programming language typing constraints. For instance, in order to satisfy the laws of Boolean algebra, class algebra permits multiple inheritance, unlike Java and C#. Class algebra has tried to follow the Ontology Web Language, OWL, as much as possible. For example, attributes, relations, and methods are defined relative to an ontology name space rather than relative to a class. An object (or class) is permitted to be equivalent to other objects (or classes) that have different object (class) identifiers. For instance, anything known about "007" can be added to anything known about "James Bond," since they are names of the same concept (i.e., class). This facilitates sharing of equivalent objects and class definitions among organizations. The normalization process will replace such object (class) identifiers by a unique representative, and attribute and relation values will be merged together for this canonical oid (cid).

In relational algebra, the result of any operation is a relation. In class algebra, the result of any operation is a class. A class's "extent" is a DBTable, which is a collection of DBObjects. These are different from transient programming language objects, which can only be made persistent by serializing them as the value of an attribute of a DBObject. DBTables can have listeners that are guaranteed to hear updates to their contained DBObjects. The calls to get/set/add/remove DBObjects are implicitly queued as necessary to preserve ACID (atomicity, consistency, isolation, durability) properties of transactions.

A DBTable has methods for sorting, grouping, histograms, and report generation. It also has a "select" operation, which can choose a subset of objects that satisfy a given class algebra constraint, represented as a string (unlike J2EE "find" methods, which must be compiled). The constraint is first normalized to a "simplest" form. The simplest form is guaranteed to be computable in $O(n^3)$ time since it involves a simple sort of a transitive closure of Horn clauses that contain no explicit variables. A Prolog definition of the normalization process is available at *http://www.cs.ccu.edu.tw/~dan/fuzzyProlog.txt*.

As well as using the normalization procedure to optimize queries, the class "intents" (i.e., normalized constraints) are used to organize the definitions of classes into the class hierarchy. That is, the normalized class definitions are used to recognize when one class is a superclass of (i.e., contains) another, or when two class definitions are equivalent or have empty intersections.

Table 1. Advantages of semantic Web over object-relational databases

- Standardized ontology declarations are readable by both humans and computers
- Nested inclusions of classes, attributes, relations, and methods
- More compatible with UML, E-R development environments
- SQL tables become classes of class algebra
- Inter-database references use type-checked URIs
- Following a URI can be more efficient than doing a join
- Has implicit cascade deletes based on XML document nesting
- B-Tree and R-Tree indices can still be used for XML documents
- Knowledge-based rules can create implicit tables
- No limits on size of strings, integers, dollars, or floats
- Fuzzy-valued attributes, relations, and rules
- Java methods are sharable on servers and clients
- Security can use Horn clause rule-based policies

CONCLUSION

As persistent storage mechanisms become distributed across the Web, it becomes imperative that a class hierarchy mechanism be used to organize the meta-data of SQL tables. Relational algebra can be extended to class algebra to provide a theoretical basis for such a class hierarchy. This requires some changes of perspective. An SQL table then represents a collection of objects rather than a relation. Only binary relations are permitted, and these correspond to joins from foreign keys to primary keys, and vice versa for inverse relations. The JNDI name of a database service and a class (i.e., table) name can be added to a foreign key to let it represent a network-wide object identifier.

Using class algebra, a standard component get/set/add/remove interface can be used to update persistent data, whether that data reside on SQL databases, XML documents, or other files. The details regarding type-checking, security, caching, and transactions can be hidden beneath these calls.

If class algebra is integrated into the programming language, as is sometimes done with SQL, then type-checking can be done statically, as in J2EE or .NET. Otherwise, some overhead must be paid for the dynamic execution when type checking assignments or when de-serializing attribute values. The positive side of using dynamic type-checking is that there is no need for recompiling applications when new subclasses or relations are declared.

A static J2EE or .NET environment can be used to develop a class algebra service for DBTables and DBObjects, which can dynamically include other classes, relations, and methods from the network. In this approach an appropriate serialization write/read method must be called to make persistent/transient copies of attribute values of type “String”. The database then handles its own DBTable and DBObject type-checking, caching, transactions, update notifications, load-balancing, and so forth, without having to worry about any particular language’s primitive type declarations and serialization procedures.

REFERENCES

Ahamad, M., & Chelliah, M. (2002, September/October). Flexible robust programming in distributed object systems. *IEEE Transactions on Knowledge and Data Engineering, 14*(5).

Beneventano, D., Bergamaschi, S., & Sartori, C. (2003, March). Description logics for semantic query optimization in object-oriented database systems. *ACM Transactions on Database Systems, 28*(1).

Buehrer, D.J. (2001). Organizing multimedia objects using a class algebra database. In T.K. Shih (Ed.), *Distributed multimedia databases: Techniques and applications* (ch. 20, pp. 318-326). Hershey, PA: Idea Group Publishing.

Buehrer, D.J., & Chien, L.R. (2003, October 26-29). Knowledge creation using class algebra. *2003 Int’l. Conference on Natural Language Processing and Knowledge Engineering,* Beijing, China (pp.108-113).

Buehrer, D.J., & Lee, C.-H. (1999, September 22-25). *Class algebra for ontology reasoning. Proc. of TOOLS Asia 99 (Technology of Object-Oriented Languages and Systems, 31st International Conference),* Nanjing, China (pp. 2-13). IEEE Press.

Buehrer, D.J., Tse-Wen, L., & Chih-Ming, H. (2001). Abia Cadabia: A distributed, intelligent database architecture. In M. Syed & O. Baiocchi (Eds.), *Intelligent multimedia, computing, and communications.* New York: John Wiley and Sons.

Buehrer, D.J., Tse-Wen, L., Chih-Ming, H., & Hou, M. (2001). The containment problem for fuzzy class algebra. In C.H. Dagli et al. (Eds.), *Intelligent engineering systems through artificial neural networks* (vol. 11, pp. 279-284). New York: ASME Press.

Buehrer, D.J., & Wang, T.-Y. (2003, September 12). The Cadabia Database Project. *The 14th Workshop on Object-Oriented Technology and Applications,* Ywan Jr University (pp. 385-392).

Everett, J.O. et al. (2002, February). Making ontologies work for resolving redundancies across documents. *Communications of the ACM, 45*(2).

Hunter, J. (2003, January). Enhancing the semantic interoperability of multimedia through a core ontology. *IEEE Transactions on Circuits and Systems for Video Technology, 13*(1), 49-58.

HyperText Markup Language (HTML) Home Page. *http://www.w3.org/MarkUp/*

Java 2 Platform, Enterprise Edition. *http://java.sun.com/j2ee/*

Java Server Pages: Dynamically Generated Web Content (JSP). *http://java.sun.com/products/jsp/*

Microsoft ASP.NET. *http://www.asp.net/*

Microsoft .NET. *http://www.microsoft.com/net/*

Stal, M. (2002, October). Web services: Beyond component-based computing. *Communications of the ACM, 45*(10).

W3C OWL Web Ontology Language Overview. *http://www.w3.org/TR/2003/CR-owl-features-20030818/*

W3C RDF Resource Description Framework. *http://www.w3.org/RDF/*

W3C Semantic Web Activity. *http://www.w3.org/2001/sw/news*

W3C XHTML™ 1.0 The Extensible HyperText Markup Language (2nd ed.). *http://www.w3.org/TR/xhtml1/*

W3C XML Extensible Markup Language. *http://www.w3.org/XML/*

Web Services and Service-Oriented Architectures. *http://www.service-architecture.com/index.html*

Weinstein, S., & Gelman, A.D. (2003, June). Networked multimedia: Issues and perspectives. *IEEE Communications Magazine, 41*(6), 138-143.

KEY TERMS

Class Algebra: This fuzzy Boolean algebra can be used either to state necessary conditions (declarations) or sufficient conditions (queries). Because of the decidability of the containment of one class algebra expression by another, any set of class algebra expressions forms an IS-A hierarchy.

Closed World Assumption: The assumption that any goal that is not provable is false. For example, if the knowledge base cannot prove that there is a flight from Taipei to New York, then you can assume that there is no such flight. This assumption is used by most knowledge bases. For class algebra, which is decidable (i.e., membership in a class expression is either true or false for a given object), the closed world assumption corresponds to the "true-complement ~" operator. A "pseudo-complement -" operator is used in proving that something has negative evidence, but the pseudo-complemented predicates are treated like positive literals in terms of Horn clauses. The maximum proofs of positive and negative evidence produce a fuzzy interval for any class algebra expression.

Horn Clauses: A set of clauses in conjunctive normal form (i.e., a conjunction of disjunctions of literals) is a Horn set of clauses if each clause contains at most one positive predicate. That is, all literals are positive when written as Prolog clauses (Head:-Body), since the negated literals become a conjunction of positive predicates when put into the Body of the clause. When resolved with positive facts, such clauses can only result in new positive facts, and the complete set of such results is enumerable for Horn clauses, and is a unique "initial" (i.e., smallest) model for the clauses.

Initial Model: The unique set of unit consequents of a set of Horn clauses.

IS-A Hierarchy: This is a directed acyclic graph, where each node is a class, and each edge represents a "subclassOf" relationship from a class to one of its superclasses. In class algebra, the classes are labeled by normalized class algebra expressions.

Literal: A predicate or a complemented predicate.

Normalized Class Algebra Expression: A Boolean expression involving the basic class algebra predicates. The sorted leaf nodes of the non-subsumed proof trees basically form a Karnough map, which covers all assignments that could make the query true.

Subsumption: For clauses in conjunctive normal form, a clause C subsumes (i.e., contains) a clause D if there exists a substitution s such that the set of predicates of Cs are a subset of the predicates in D. The subsumed clause D may be safely deleted, as it adds nothing to the logical meaning of the conjunctive normal form in a Boolean logic.

Outsourcing Information Technology in Australia

Nicholas Beaumont
Monash University, Australia

INTRODUCTION

Outsourcing information technology (IT) has been defined as passing IT functions previously performed in-house to outside contractors (Hancox & Hackney, 2000). Outsourcing has spawned variants, especially insourcing—two departments of one organization formalizing their supply relationship.

Lonsdale and Cox (2000, pp. 445-449) summarize the history of outsourcing, noting that it substitutes for once fashionable enthusiasms for conglomeration, horizontal integration, and vertical integration. Williamson and Winter (1993) stress the role of transaction costs in all business decisions asking: Given that the market allocates resources efficiently, why do firms exist? They introduce the "asset specificity" concept; some assets have a narrow range of applications and have few or no potential buyers or sellers. The assets are inoperable without a cluster of peripheral assets, and high levels of skill and knowledge in the firm and its suppliers. There is no perfect market for such assets; investment in them is only practical in a planned environment that provides a guaranteed market for a high volume of very specific outputs and guaranteed supply of specific components.

Dealing in imperfect markets involves costs (such as identifying appropriate vendors, verifying their competence, communicating changing requirements to them, and monitoring their performances) and risks (of vendors failing or being unable to meet specifications). Deciding whether to outsource depends on comparisons of the long-term costs and risks of different modes of supply. A long-term outsourcing relationship with a trusted vendor (incorporating some price signals) may be preferable to securing supply by backward integration or developing an in-house capacity (possibly creating administrative complexity). An organization will always outsource some activities (water and electricity supply), but retain others such as strategic planning.

BACKGROUND

Outsourcing information processing (IP) functions has been facilitated by Web technologies that make it easy, quick, and cheap to transfer internal data (e.g., payroll data or delivery instructions) to a vendor. The vendor can analyze data and instantly transmit results to the client, employees' banks, and the tax office, eliminating time delay problems. Physical locations of the vendor and client are irrelevant for IP tasks; some kinds of outsourcing can be performed in low wage countries and/or a centralized facility that provides economies of scale. This is exemplified by the Asian call center and the payroll specialist whose volume (derived from many clients) allows exploitation of a high fixed-cost, low variable-cost technology and expensive expertise.

These fundamental considerations are manifest in many different motivations for outsourcing. A small organization might outsource a payroll application primarily to reduce costs by accessing economies of scale, but IT outsourcing decisions can be complex because IT applications tend to be integrated with each other and permeate the firm, touching most activities performed. The criteria used to evaluate outsourcing decisions may be multidimensional and intangible; besides cost savings (sometimes hard to ascertain), the firm must consider possible effects on customer service and its own staff; the potential advantages of access to expertise; and whether it might become uncomfortably dependent on the service provider (Kern & Willcocks, 2000) or hamper adoption to a changing environment.

THE AUSTRALIAN MARKET

In 2003, IDC (Allen & Langby, 2003) estimated that the Australian IS outsourcing market was approximately AU$3.85 billion in 2002 and would be AU$5.5B in 2007, a compound annual growth rate (CARG) of 7.4%. IT outsourcing represents "…over 10% of the total IT market and is growing faster than the total IT market." The market leaders—IBM GSA, EDS, and CSC—accounted for approximately 65-75% of the Australian IS outsourcing services market. Growth has slowed, the market is becoming more mature and competitive, clients are seeking shorter contracts and outsourcing specific tasks instead entrusting all their IT activities to a single vendor. Motivations include lower costs, higher quality, and focus on core competencies. Clients considering outsourcing were inhibited by potential loss of control, possible inability to resume running applications in-house, the complexity of protracted negotiations, loss of intellectual property, and security issues.

In 2001, Gartner predicted that (i) the value of the Australian IT services market would grow from $AU20 billion in 2001 to $AU 50 billion in 2004. (ii) By 2005, 25% of Australian organizations will not own their IT services or infrastructure, and (iii) the value of the Australian business process outsourcing market was $AU5.5 billion in 2001 and will grow to $AU15.4 billion by 2004 (Brown, 2001).

Robertson (2001) opined that "revenue from the Australian IT outsourcing market was $A1.3 billion in 1997, and projections to 2004 will push it up to $A5.3 billion, a CAGR of 16%." The increased use of IT outsourcing in Australia is exemplified by recent outsourcing decisions, some of which are listed in Table 1.

In 1998 the Australian federal government decided to force government departments to outsource their IT requirements. This initiative was a failure (Connors, 2001); the suppliers did not meet the users' requirements (wanting to apply one template to similar but distinct applications); the costs were excessive and the initiative was at least temporarily abandoned. In 1995, the South Australian state government initiated an ambitious "whole of government" 10-year/$AU1 billion IT outsourcing project with EDS. There are different views of the direct and indirect benefits of this contract to the state, but the consensus is that great effort from both parties produced marginal benefits. The more experienced state government would expect more benefit from any future outsourcing contract (Connors, 2004).

The few academic studies of Australian IT outsourcing include: Hurley and Costa (2001), a set of Australian cases; Costa (2001), an outsourcing literature review set in an Australian context; Beaumont and Costa (2002), a report on IT outsourcing; and Beaumont and Sohal (2004), a report on outsourcing in general. The author recently interviewed 10 senior executives of Australian vendors; insights from these interviews are used in this article.

REASONS FOR OUTSOURCING

Many writers have proposed research frameworks (Hancox & Hackney, 2000), criteria (Mamaghani, 2000; Smith & Rupp, 2003), and schemes for analyzing or assessing outsourcing (Bazinet, Kahn & Smith, 1998). Earl (1996) stresses the importance of considering intangible and tangible criteria in major business decisions. Some commonly cited reasons for using IT outsourcing are given in Table 2. Reasons for not using IT outsourcing and risks inherent in IT outsourcing are summarized in Table 3. Barthelemy (2001) and Barthelemy and Adsit (2003) give more details. The most common cause of outsourcing project failure is *project risk*. Problems arise when potential users of outsourcing do not carefully define their requirements, do not properly prepare for negotiations with potential vendors, or do not put enough effort into monitoring the outsourcing arrangement. An Australian contract with a call center did not specify some of the services that the client offered its customers. Another

Table 1. Some recent outsourcing contracts

Client	Vendor	Tasks Outsourced	Total Value (AU$)*	Duration (years)*
Commonwealth Bank of Australia (King, 1998)	EDS	IT	5B	10
Bank of Queensland	EDS	IT, call centers, and credit card processing	480M	10
Telstra	IBM GSA		4B	
Australian Taxation Office	EDS	IT	490M	
Five Government Agencies	CSC	IT	160M	
Two Government Agencies (Beer, 1999)	IBM GSA		351M	
Qantas	Oracle/ IBM GSA	IT	200M	
AMP (McFarlan & Nolan, 1995)	KAZ	IT/Insurance	70M	6
BHP-Billiton	CSC	IT	1B pa	10
South Australian Government	EDS	IT	700M	10
Westpac	EDS	IT	1B	10

Table 2. Reasons for adopting outsourcing

Reason	Comment
Reduce processing costs	By accessing economies of scale or expertise.
Focus	Outsource non-core activities and concentrate on areas of competence. Quinn and Hilmer (1994) note that identifying core activities may not be straightforward.
Access expertise	Especially in building new systems.
Avoid cultural difficulties	The IT department has anomalous working conditions.
Financial	Change from fixed to variable cost, sell assets, or manipulate accounts.
Benchmark internal operations	Get rid of an IT department that is hard to manage and/or unproductive.
Quality	Improve the IT services provided.
Resources	To meet uneven demand and/or access resources not available internally.
Risk avoidance	Avoid financial uncertainty through a fixed cost per transaction. Avoid the risks inherent in managing a project.
Ideological	The current Australian government has an ideological commitment to outsourcing, opining that the private sector is intrinsically more efficient than the public sector.

organization did not specify a reporting system in enough detail to demonstrate that the vendor was failing to deliver the sought services. Further references to reasons for and impediments to outsourcing are given in Beaumont and Costa (2002) and Beaumont and Sohal (2004).

FUTURE TRENDS

The relationship between vendor and client can (in theory) range from an arms-length and legalistic arrangement (appropriate where the vendor's performance can be unambiguously measured) to an intensely cooperative arrangement. In interviews, Australian vendors very firmly opined that taking a legalistic view would be extremely short sighted. The first contract with a client is an opportunity to demonstrate the competence and commitment prerequisite to having the contract renewed and assuming responsibility for other of the client's business processes. Clients expect to share the benefits of improving technology and exploit vendors' expertise. This reflects two of clients' primary outsourcing goals: reduced costs and access to expertise.

Australian vendors stress the importance of negotiating detailed service-level agreements (SLAs), noting that some of their clients have used specialist advice to

Table 3. Reasons for not outsourcing

Dependence	Becoming dependent on the vendor.
Confidentiality	Keep confidential data in-house.
Intellectual property	Do not give other parties access.
Loss of distinctive competencies	Outsourcing may atrophy in-house skills.
Loss of flexibility	A three-year contract may reduce the ability to adjust to changes in the environment or exploit new technology.
Personnel and change problems	Redundant IT staff may have to be dismissed or redeployed.
Information asymmetry	It is difficult for the client to determine whether the prices paid are reasonable.
Project risk	Poor methodology, negotiation, and monitoring.
Vendor incompetence	
Relationship costs	The costs of each stage of the outsourcing cycle, e.g., preparing and analyzing requirements and negotiating and monitoring vendor performance.

drive hard bargains. Comprehensive SLAs are necessary, but not sufficient, for a fruitful relationship. Competent vendor performance generates trust and ultimately "partnering" (see definitions below). Many participants in Australian outsourcing projects emphasize that negotiating and implementing outsourcing agreements is extremely stressful. A typical history tells of euphoria when protracted negotiation climaxes with signing the contract and is followed by disillusionment six months later. Usually, both parties recognize the need to repair the relationship, and more negotiations usually result in a workable relationship emerging.

Many small and medium enterprises (SMEs) retain IT departments afflicted by diseconomies of scale. Such departments find it difficult to keep current with evolving technology, tend to reinvent wheels, and are internally unpopular essentially because they do not have the skills or resources to develop and maintain the reliable systems expected and demanded by the organization. The assumption that systems should be developed in-house is too rarely examined in Australia. Although all organizations' IT systems look different, they reflect a finite number of different business processes. It would be rational for most SMEs to abandon internal systems development and either outsource IT or adopt a package such as SAP that encompasses all common business processes and can be tailored to individual firms' needs.

SURVEYS OF IT OUTSOURCING IN AUSTRALIA

Beaumont and Costa's (2002) paper on IT outsourcing in Australia was based on a survey of CFOs and CIOs, supplemented by interviews with six executives heavily involved in outsourcing. The research sought information on demographics, the kinds of activities outsourced, reasons for and obstacles to IT outsourcing, and the nature of outsourcing relationships. Some findings follow.

IT outsourcing use is not related to industry classification or organizational size. It is positively related to the size of the IT budget. The frequency of outsourcing is strongly dependent on the activity; for example, hardware maintenance is frequently outsourced, but "analysis and IT strategy" is rarely outsourced. Reasons for outsourcing are broadly consistent with findings from North America. Table 4 summarizes the results and includes a comparison with findings from a North American survey (Collins & Millen, 1995). The measurements of anticipated and actual benefits were made using a five-point Likert scale anchored by 1 (not important) and 5 (very important).

Another survey exploring Australian corporations' attitudes to outsourcing (not just IT outsourcing) was conducted in late 2001 (Beaumont & Sohal, 2004). Of the 162 responses, 93 reported outsourcing at least one as-

Table 4. Anticipated and actual objectives (ranked)

Rank based on anticipated benefits	Objective	Comparison with ranks given in Collins & Millen (1995, Table 4)	Mean (anticipated)	Mean (actual)
1	Access to skills	May be correlated with personnel cost savings (2)	2.65	2.27
2	Improved service quality	Synonymous (3)	2.48	2.06
3	Focus on core business	Synonymous (1)	2.34	1.97
4	Defined service levels	An aspect of improved quality of IS services (3)	2.34	1.89
5	Additional flexibility	Synonymous (4)	2.31	1.94
6	Access to technology	Synonymous (5)	2.25	1.98
7	Improved performance	An aspect of improved quality of IS services (3)	2.29	1.84
8	Cost savings	Reflected by personnel cost saving (2); technology cost saving (8); stabilize IS costs (7)	1.89	1.64
9	Change fixed cost basis	Not measured	1.72	1.52

pect of IT. Some aggregated responses of these 93 informants are reported here.

Table 5 gives the IT functions that were outsourced. The intensity reflects the frequency with which these functions were outsourced (a five-point scale). Clearly, development of systems and Web sites were the activities most often outsourced.

Respondents were asked to assess the benefits obtained from outsourcing (23 possible benefits were nominated). The responses were factor analyzed to obtain common themes. These common themes were, in decreasing order of importance: 1) cost savings; 2) improved performance as experienced by customers (e.g., improved user satisfaction and better service levels); 3) access to skills and technology; and 4) (surprisingly) the elimination of internal power struggles, "removal of an unsatisfactory department," and "cultural differences" were often cited.

Reluctance to outsource was primarily attributed to difficulties in ascertaining what, if any, savings could be made. Less frequently cited reasons were a fear of becoming dependent on the vendor, lack of confidence in vendors, and a fear of loss of control.

Respondents were asked how they chose vendors. The results were not amenable to factor analysis, but criteria commonly cited were: cost; proven ability ("track record"); willingness to "adapt to our business"; reputation, competence, and "understanding of our industry"; and the vendor's commitment to a long-term relationship.

OFFSHORING

Offshoring (having work outsourced to low-wage countries such as India) has become controversial in Australia and other first-world countries (Beaumont, 2004). Australian companies will be unable to ignore the opportunity to reduce the cost of systems development, data entry, and transaction processing by up to 80%. The Indian workforce is well educated with proven IT skills (McLaughlin, 2003); China, the Philippines, and perhaps New Zealand are other emerging locations. There have been calls for offshoring in Australia and the USA to be regulated to prevent IT jobs being transferred to India. Such fears were exacerbated by a Telstra vendor considering transferring some IT work to India. Several other large Australian companies have considered offshoring, some retreating because of adverse public reactions.

It is difficult for a client or governments to control where IT work gets done. Some very large organizations (Microsoft, IBM, HP) have organized their own support and software development sites in India and China. There is a bright side for Australia and perhaps New Zealand, whose citizens have high levels of education, excellent English, cultures akin to those of America and Europe, excellent technical skills, and labor costs distinctly lower than those of America. IBM and Oracle already have Australian centers supporting international activities.

Table 5. Kinds of IT outsourcing

Application	Number responding	Intensity
Development of IT applications	92	3.5
Development of Web sites	88	3.5
Security of IT	86	2.9
IT infrastructure	85	3.2
IT data center	82	2.7

CONCLUSION

The use of IT outsourcing in Australia is growing at about 10% per year. Initially adopted uncritically with some unfortunate results, it is now used more soberly in a strategic context. Establishing a fruitful outsourcing arrangement requires sustained effort from both parties; clearly written SLAs reflecting the client's objectives are a necessary, but not sufficient condition. There is a tendency to trial vendors by writing smaller, shorter term contracts for parts of the IT task and demanding demonstrated benefits before extending the relationship. Political problems may arise when cost competition forces Australian organizations to consider outsourcing IT work to third-world countries.

REFERENCES

Allen, P. & Langby, M. (2003). *Australia IS outsourcing forecast and analysis, 2003-2007*. Report AU221212. Sydney: IDC.

Barthelemy, J. (2001). The hidden costs of IT outsourcing. *Sloan Management Review, 42*(3), 60-69.

Barthelemy, J. & Adsit, D. (2003). The seven deadly sins of outsourcing. *Academy of Management Executive, 17*(2), 87-100.

Bazinet, C.G., Kahn, S.A. & Smith, S.J. (1998). Measuring the value of outsourcing. *Best's Review, 98*(12), 97-100.

Beaumont, N.B. (2004). Offshoring: Political dynamite or an unstoppable force? *Corporate Outsourcing, 2*(April-May), 40-41.

Beaumont, N.B. & Costa, C. (2002). Information technology outsourcing in Australia. *Information Resources Management Journal, 15*(3), 14-31.

Beaumont, N.B. & Sohal, A. (2004). Outsourcing in Australia. *International Journal of Operations and Production Management*, (in press).

Beer, S. (1999). Outsourcing 'likely to end in failure'. *Australian Financial Review,* (April 19), 34.

Brown, P. (2001). IT rides on shift to outsourcing. *The Australian,* (April 3), 23.

Collins, J.S. & Millen, R.A. (1995). Information systems outsourcing by large American industrial firms: Choices and impacts. *Information Resources Management Journal, 8*(1), 5-13.

Connors, E. (2001). Humphry leaves IT strategy in tatters. *Australian Financial Review,* (January 15), 4.

Connors, E. (2004). IT sourcing in transition. *Australian Financial Review*, (March 2), 30.

Costa, C. (2001). Information technology outsourcing in Australia: A literature review. *Information Management & Computer Security, 9*(5), 213-224.

Earl, M. (1996). The risks of IT outsourcing. *Sloan Management Review, 37,* 26-32.

Greaver, M.F. (1999). *Strategic outsourcing: A structured approach to outsourcing decisions and initiatives.* New York: Amacom.

Hancox, M. & Hackney, R. (2000). IT outsourcing: Frameworks for conceptualizing practice and perception. *Information Systems Journal, 10*(3), 217-237.

Hurley, M. & Costa, C. (2001). *The blurring boundary of the organisation.* Melbourne: KPMG Consulting (Australia).

Kern, T. & Willcocks, L. (2000). Contracts, control and 'presentation' in IT outsourcing: Research in thirteen UK organizations. *Journal of Global Information Management, 8*(4), 15-27.

King, J. (1998). Companies team up for equity outsourcing. *Computerworld, 32*(11), 16.

Lonsdale, C. & Cox, A. (2000). The historical development of outsourcing: The latest fad? *Industrial Management & Data Systems, 100*(9), 444-450.

Mamaghani, F. (2000). Selecting an outsourcing vendor for information systems. *International Journal of Management, 17*(3), 334-343.

McFarlan, F.W. & Nolan, R.L. (1995). How to manage an IT outsourcing alliance. *Sloan Management Review, 36*(2), 9-22.

McLaughlin, L. (2003). An eye on India: Outsourcing debate continues. *IEEE Software, 20*(3), 114-117.

Quinn, J.B. & Hilmer, F.G. (1994). Strategic outsourcing. *Sloan Management Review, 35*(4), 43-55.

Robertson, R. (2001). Size dominates the IT outsourcing market. *Australian Financial Review*, (February 28), 3.

Smith, A.D. & Rupp, W.T. (2003). Application service providers: An application of the transaction cost model. *Information Management & Computer Security, 11*(1), 11.

Strassmann, P.A. (1997). *The squandered computer: Evaluating the business alignment of information technologies.* New Canaan, CN: Information Economics.

Williamson, O.E. & Winter, S.G. (1993). *The nature of the firm: Origins, evolution, and development.* New York: Oxford University Press.

KEY TERMS

Backsourcing or Repatriation: Taking work that had been outsourced back in-house.

Co-Sourcing: Used instead of *outsourcing* to emphasize that the outsourcing arrangement is based on cooperation (or "partnering") between the parties. Partnering is often used to signify a highly cooperative arrangement, but Australian law only recognizes written contracts (not "a spirit of cooperation") and "de facto" relationships.

Insourcing: Formalization of the supply relationship between two departments of the same organization.

Outsourcing: "The act of transferring some of a company's recurring internal activities and decision rights to outside providers as set forth in a contract" (Greaver, 1999, p 3). Some disasters have brought outsourcing into disrepute, and arrangements have been relabeled *technology services partnerships* or *managed services contracts.*

Outsourcing Cycle: The stages of an outsourcing project: identification of a need; definition of project objectives; choice of vendor; negotiation of a Service-Level Agreement; implementation; performance monitoring; incremental modification of the agreement; and renewal or cancellation of the agreement.

Out-Tasking: The vendor provides resources (computer time or labor) and does work precisely defined by the client (Strassmann, 1997).

Service-Level Agreement: A contract defining the services to be provided by a vendor and service levels; procedures for measuring performance; payments and penalties and procedures for renegotiation.

Strategic Outsourcing: Outsourcing considered in strategic context (Quinn & Hilmer, 1994). An arrangement attractive on purely cost grounds might be rejected because it reduced the client's ability to react to a changing environment or entailed unhealthy dependence on the vendor.

ENDNOTE

* Where known

Overcoming Barriers in the Planning of a Virtual Library

Ardis Hanson
University of South Florida, USA

INTRODUCTION

What is a virtual library? Gapen (1993) states:

The virtual library has been defined as the concept of remote access to the contents and services of libraries and other information resources, combining an on-site collection of current and heavily used materials in both print and electronic form, with an electronic network which provides access to, and delivery from, external worldwide library and commercial information and knowledge sources. In essence the user is provided the effect of a library which is a synergy created by bringing together technologically the resources of many, many libraries and information resources (p.1).

A decade later, the same definition holds true.

BACKGROUND

The implementation of a university-wide virtual library inevitably causes significant changes within each department across the library and campus. However, real ownership comes about only with a commitment from management to 1) educate and train staff in the new technologies and procedures, 2) manage the change process as candidly as possible, and 3) disseminate information as openly and widely as possible. Each one of these commitments is essential to the change process.

When creating a virtual library, all participants must clearly define and realize institutional commitment to the plan. To do so effectively requires an examination of the organizational and cultural change agents that will affect the implementation. The cultural constructs by which the various types of librarians and staff view the worlds of work and patron must be considered in the redesign of the current environment. Although much has been written about creating virtual libraries, very little is known about the working conditions or institutional and organizational practices that make virtual libraries most usable by library faculty and staff.

Preparing a proposal for a university-wide virtual library requires a new level of collaborative visioning, planning, and implementing among the participating departments/libraries. This process depends upon a reliable and advanced networked infrastructure and on staff, facility, and financial decisions focused on true programmatic cooperation. It is well cited in the change management literature that without commitment from administration, the belief of the individual, the team, or the organization in the relevance of the project wavers or devolves to indifference to the project. Organizational commitment to the project and, therefore, to change is a key factor in determining the success of this type of project.

COMMITMENT TO CHANGE

As libraries respond to the rapid evolution of the information environment, organizational change must occur for successful adaptation. The literature provides some insight into variables important to organizational change. An integral link exists between successful leadership, human resources, and organizational structure. If any one component has a weakness, it will negatively impact the other two. A key component to success is the ability to show flexibility and rapid response to change. This is mirrored in the qualifications necessary for staff to work effectively in a rapidly evolving environment. The literature also points to two key components in creating an effective organizational structure: 1) flattening of the existing structure and 2) development of committees that cross departmental lines.

Since open discussion of barriers is an effective strategy for implementation and adaptation of change (Beer & Eisenstat, 2000), the library as organization must be examined with a critical eye to identify strengths, weakness, ideologies, and strategic staff. Plans to resolve challenging organizational or staff problems must be devised and implemented.

Further, project outcome is shown to depend upon the behavior and resources available during the development process, and these depend upon the level of commitment. Additional factors to consider include the possibility of budgetary constraints, top management's technological non-sophistication, and the reluctance to draw up a very detailed operating plan. Finally, barriers on systems implementation included hesitation in accepting a system, lack of standard formats, and unrealistic priorities and expectations from the parties involved. These challenges are particularly true in the design of a virtual library where resources and technical standards are constantly evolving.

When identifying problems, it is essential to determine which problems are undesirable situations but impossible to correct. This allows the project team to acknowledge the problem, but not to spend too much time on it, allowing the team to focus on answers or actions to potentially correctable issues. Common problems identified include 1) the disparate geographic locations of the libraries (especially with multi-campus or departmental libraries within a university or college); 2) the lack of parity in equipment and technologies among departments within the library or among libraries; 3) budget issues; 4) staff involvement in the project; 5) difficulty in selling the virtual library as a critical library project to university administration, teaching and research faculty, and library faculty and staff; 6) unrealistic expectations by management and staff; and 7) resistance to change.

MANAGING CULTURAL CHANGE

The coming of a virtual library will significantly impact the culture of libraries. Janz and Prasarnphanich (2003, p. 353) believe that "corporate culture determines values, beliefs, and work systems that could encourage or impede learning (knowledge creation) as well as knowledge sharing." Policies, procedures, and behavior are a visible manifestation of these assumptions that most members of a culture never question or examine. The members of a culture may not even be aware of their own culture until they encounter a different one.

The cultural constructs by which the various types of librarians and staff view the worlds of work and patron should be considered in the redesign of the current environment. Very little is known about the working conditions or institutional and organizational practices that make digital libraries most usable by library faculty and staff. Most of the emphasis on users has been on the library patron or on the technology, not on the work environment of the library staff. Much of the literature discusses change at the technological level, again not at the human level. Examples include the use of Standard Generalized Markup Language (SGML) hyperlinks in bibliographic records, metadata, collection development with Internet sources, online reference services, end user electronic document delivery, electronic full-text reserves, and networked paper-less administration.

Models that make explicit connections between a focal technology (such as a digital library) and its immediate users should be used to review the ecology of social relationships with other social groups and organizations in which the technology is developed, adopted and used.

IMPLEMENTATION ISSUES

To promote ownership of the virtual library plan among the library's administration, faculty, and staff, "actions" are developed for each of the areas under design within the virtual library. At institutions that have successfully implemented a virtual library, work groups and task forces were instituted to review workflow, policies and procedures, overlap, and to create implementation plans to effect organizational change and major projects driven by the new technologies (Steele, 2000). Early virtual libraries created multi-disciplinary teams across academic disciplines and support services to implement their plans or created large groups who use phased plans covering a number of years (Anderson, 1995; Birmingham, Drabenstott, Warner, & Willis, 1994).

Alternative organization design principles replace individual job design with task design (e.g., task forces and teams). These principles are an excellent way to handle the enormity of the proposed design process. They replace "one best way" thinking with choices for organizational and technological solutions (Sarin & McDermott, 2003). Such solutions can cope with more complexity and unpredictability when people in the teams are multi-skilled and delegated more autonomy and responsibility (Hoegl, Parboteeah, & Munson, 2003; Man & Lam, 2003).

Implementing a virtual library will create cultural and organizational changes, particularly in technology and staffing. It is critical that library faculty and staff be assured of their importance and value within the new organizational structure. Staff development programs need to be made available and designed in such a way that staff members feel the new skills will enhance their ability to perform their jobs and accommodate changing work functions. Administrators and department heads need to be proactive and provide these programs to staff in an appropriate time frame. The expectation that "just-in-time" training is "good enough" is unwarranted. Often, this kind of training approach increases staff stress and decreases morale, although too much advanced training can be counterproductive as well.

As technology and a growing reliance on "virtual" materials and services increase, service points and their job functions should be analyzed. This excerpt from the USF Virtual Library Blueprint (Metz-Wiseman et al., 1996) acknowledges the importance of recognizing and managing the change process as well as emphasizing the need for staff development:

"This process of change will require creativity, a more global point of view, flexibility, and adaptability. These will be the qualities of highly valued staff.… Staff will need to understand not only their own role within the organization, but also the structure of other departments and the relationships between departments." (p.26)

Therefore, in designing a virtual library, there needs to be a more global understanding of library processes, cross training and shifting staff between departments, continuous training to remain current with technology, and time for staff to develop projects and procedures across departmental boundaries.

ROLES OF AND FOR THE LIBRARY ADMINISTRATOR

As stated earlier, any recommended actions in the implementation of a virtual library require long-term commitment by the libraries' administrations (upper- and middle-management) to their staff. Strategies to continue the change process, identified on both an individual and administrative level for professional development, include: funding for each library faculty member and staff to attend conferences, training and cross-training sessions on current library operations for staff both inside and outside the involved departments; utilization of campus training programs; an ongoing evaluation program to ensure that training is available for new technologies, workflow, and responsibilities; creation of a new technologies committee whose members are responsible for training their departments; and up-to-date documentation on policies and procedures.

Inter-institutional collaboration is difficult to define and implement, and even more difficult to implement well. Four additional infrastructure elements - attention to continuing planning, process, people, and administrative elements- are essential to a project of this kind. This becomes even more important with a conscious commitment to maintaining tradition while implementing a virtual workplace (Black & Edwards, 2000).

For administrators, the single most important factor for the success of such a system is a vision that is meaningful to the director, his or her faculty, and staff, and can be advocated within larger institutional settings as well as the library. An administrator needs to examine his or her own areas and skills so he or she can effectively foster an environment that supports the innovations in technology and services. Within a larger context, a library administrator must be confident that the larger university or college atmosphere is willing to allow him or her the opportunity to initiate collaborative, inter-institutional activities and decision-making processes. This is especially critical in today's academic environment, which requires that any large-scale initiative be integrated within the mission and core values of the university, i.e., the support of scholarship and the development and continuance of a sophisticated and individualized learning environment.

UNIVERSITY INVESTMENT

The key to this scholarly environment is the development of organizational, instructional, and informational infrastructures that 1) capitalize on the technology, and 2) reinforce the values and the identities of the institution (Heikkila & Isett, 2004). Critical technical challenges in building a virtual library include:

- parity in networking, workstation, and desktop requirements;
- a single, uncomplicated point of access for the library's electronic resources;
- standard, staff-utilized software packages;
- measurable information applications that are scalable, efficient, and interoperable;
- multiple simultaneous access with minimal down/lag time; and
- technologies and services that are user-friendly and provide privacy, security, stability, mobility, and ubiquity.

Institutions must be able to combine resources in new ways, gain additional resources, dispose of superfluous resources, and to do this repeatedly and rapidly if they are to be successful (Daniel & Wilson, 2003). Existing and planned voice, data, and image networks; institutional information systems; personal computers and workstations for students, staff, and faculty; and strategic plans for information and computing technologies at the university or college level should be considered in the creation of any virtual library plan. Conversely, the virtual library should be incorporated into university and college planning for technology and information systems. Both the library and larger administration should ensure that the virtual library is part of any extension of existing technologies and plans and future technologies used at the university. This is particularly crucial as wireless communications move into student dorms, classroom buildings, student centers, and outside space as viable a "library place" as the physical building.

FUTURE TRENDS

Three additional issues of importance are especially relevant to libraries and their user communities: content, access, and support. All three issues identified are extremely dependent upon funding. Money buys research resources (content), telecommunications infrastructure (access) necessary to deliver the content, and pays staff to run the infrastructure (support). A quick look at trends in higher education indicates that traditional funding sources remain flat or are decreasing; public and state mandates call for more accountability in the spending of state dollars; and student expectations demand more sophisticated services and greater access to data. Any virtual library plan needs to develop sustainable funding mechanisms.

It is important for library administration to make a long-term commitment to inter-institutional collaboration. The single most important factor in shared decision making is a common vision that is meaningful to the dean/director, his or her faculty, and staff. This vision must be one that is shared by the administrative staff and strenuously advocated within their individual institutional settings as well as within the larger context of the university. The role of the campus library may not be to lead the development of new information technologies, but to establish an atmosphere and a process that will promote the integration of these new technologies with each other and with the mission and core values of the university and its campuses.

CONCLUSION

The real revolution in information technology is about communication, not computation. The essential catalyst for change must be in how our cooperative efforts are communicated both internally and externally. More and more teams and groups are functioning as a community-in-practice. This type of work group evolves from three characteristics: 1) valuation of work roles, 2) the degree of participation in "peripheral" learning permitted under working conditions such as conferences, workshops, and networking opportunities, and 3) opportunities for participation in innovative implementations. Learners, according to Brown and Duguid (1991), grow through practice. "The central issue in learning is *becoming* a practitioner not learning about *practice*." As participants in an innovative "community-of-practice," work group practices are fluid and ignore many of the traditional assumptions about librarianship, librarians, and libraries. As members of the committee become more comfortable as a "community," such candidness promotes a high level of cooperation between all the committee members, despite disagreements, and can help maintain the level of commitment to the group and the project development. Consequently, organizational knowledge arises from group knowledge as well as from individual knowledge, as is also indicated by the recent literature on communities of practice (Sabherwal & Becerra-Fernandez, 2003).

REFERENCES

Anderson, G. (1995). MIT: The distributed library initiative: Collaboration, vision, prototyping. *Publications of Essen University Library, 18*, 61-89.

Beer, M. & Eisenstat, R.A. (2000). *The silent killers of strategy implementation and learning. Sloan Management Review, 41*(4), 29-40.

Birmingham, W.P., Drabenstott, K.M., Warner, A.J., & Willis, K. (1994) *The University of Michigan Digital Library: This is not your father's library*. Retrieved 23 August 2003 from http://www.csdl.tamu.edu/DL94/paper/umdl.html.

Black, J. A. & Edwards, S. (2000). Emergence of virtual or network organizations: Fad or feature. *Journal of Organizational Change Management, 13*(6), 567-576.

Brown, J.S. & Duguid, P. (1991). Organizational learning and communities-in-practice: Toward a unified view of working, learning, and innovation. *Organization Science, 2*(1), 40-57.

Daniel, E. M. &Wilson, H. N. (2003). The role of dynamic capabilities in e-business transformation. *European Journal of Information Systems, 12*(4), 282-296.

Gapen, D. K. (1993). The virtual library: Knowledge, society and the librarian. In L.M. Saunders (ed.), *The virtual library: Visions and realities* (pp. 1-14). London: Meckler Publishing.

Heikkila, T. & Isett, K. R. (2004). Modeling operational decision making in public organizations - An integration of two institutional theories. *American Review of Public Administration, 34*(1), 3-19.

Hoegl, M., Parboteeah, K. P., & Munson, C. L. (2003). Team-level antecedents of individuals' knowledge networks. *Decision Sciences, 34*(4), 741-770.

Janz, B. D. & Prasarnphanich, P. (2003). Understanding the antecedents of effective knowledge management: The importance of a knowledge-centered culture. *Decision Sciences, 34*(2), 351-384.

Man, D. C. & Lam, S. S. K. (2003). The effects of job complexity and autonomy on cohesiveness in collectivistic and individualistic work groups: A cross-cultural analysis. *Journal of Organizational Behavior, 24* (8), 979-1001.

Metz-Wiseman, M., Silver, S., Hanson, A., Johnston, J., Grohs, K., Neville, T., Sanchez, E., & Gray, C. (1996). *The USF Libraries Virtual Library project: A blueprint for development*. Tampa, FL: University of South Florida Libraries. [ERIC Document ED418704].

Sabherwal, R. & Becerra-Fernandez, I. (2003). An empirical study of the effect of knowledge management processes at individual, group, and organizational levels. *Decision Sciences*, *34*(2), 225-260.

Sarin, S. & McDermott, C. (2003). The effect of team leader characteristics on learning, knowledge application, and performance of cross-functional new product development teams. *Decision Sciences*, *34*(4), 707-739.

Steele, N.O. (2000). Success factors for virtual libraries. *Econtent,* 23(5), 68-71.

KEY TERMS

Change Management: The application of many different ideas from the engineering, business, and psychology fields that focus on observable, measurable business elements. These elements, which can be changed or improved, include business strategy, processes, systems, organizational structures, and job roles.

Community-In-Practice: A term coined by Brown and Duguid (1991) regarding how groups of individuals become a team able to effectively participate in innovative implementations.

"Just-In-Time" Training (JIT): The philosophy of providing training at the exact point when one needs the knowledge/skills to perform a task or job function.

Metadata: Data that describes the content, data definition and structural representation, extent, quality, availability, status, and administration of a dataset. Examples include the MARC record in a library catalogue or information in the meta-tags of an HTML or SGML document.

Organizational Culture: A pattern of behavior, such as cross-individual behavioral consistency. For example, when people say that culture is *"the way we do things around here,"* they mean that there is a consistent way people perform tasks, solve problems, and resolve conflicts. Organizational culture is also defined as the informal values, norms, and beliefs that control how individuals and groups in an organization interact with each other and with people outside the organization.

Virtual Library: Services and resources that are provided in a digital or electronic format and accessed via a computer either in a library or remotely from another workstation.

Wireless Communications: To communicate with or send communications by wireless means, relating to radio or communication by radiotelegraphy or radiotelephony.

Packet Inter-Arrival Distributions in Computer Network Workloads

Dennis Guster
St. Cloud State University, USA

David Robinson
St. Cloud State University, USA

Paul Safonov
St. Cloud State University, USA

INTRODUCTION

The past decade could be classified as the "decade of connectivity"; in fact it is commonplace for computers to be connected to a LAN, which in turn is connected to a WAN, which provides an Internet connection. On an application level this connectivity allows access to data that even 5 years earlier were unavailable to the general population.

This growth has not occurred without problems, however. The number of users and the complexity/size of their applications continue to mushroom. Many networks are over- subscribed in terms of bandwidth, especially during peak usage periods. Often network growth was not planned for, and these networks suffer from poor design. Also the explosive growth has often necessitated that crisis management be employed just to keep basic applications running. Whatever the source of the problem, it is clear that proactive design and management strategies need to be employed to optimize available networking resources (Fortier & Desrochers, 1990).

BACKGROUND

Obviously, one way to increase network bandwidth is to increase the speed of the links. However, this may not always be practical due to cost or implementation time. Furthermore, this solution needs to be carefully thought out because increasing speed in one part of a network could adversely effect response time in another part of that network. Another solution would be to optimize the currently available bandwidth through programming logic. Quality of service (QOS) and reservation bandwidth (RB) are two popular methods currently being utilized. Implementation of these optimization methods is rarely simple and often requires a high degree of experimentation if they are to be effectively configured (Walker, 2000). This experimentation can have detrimental effects on a live network, often taking away resources from mission critical applications.

THE BENEFITS OF SIMULATION

Therefore, the most efficient way to ascertain the potential benefit and derive baseline configuration parameters for these optimization methods is through simulation or mathematical modeling. Simulation can be very effective in planning a network design. For example, what if network link #3 was increased to 100Mbs? Would workstations on that link experience an improvement in response time? What would happen to workstations on the other part of the total network? Another approach to ascertain if a given network will exceed its capacity is based on network calculus (Cruz, 1991; Le Boudec, 1998). In this method the characteristics (such as speed, maximum packet size, peak rate) of the network architecture are analyzed and performance bounds are defined. The goal then is to devise control/management programs (such as QOS and RB) that will keep the workload within those defined bounds. There are numerous applications of this control/management logic, such as Cruz (1995), Cruz and Tsai (1996), Firoiu, Le Boudec, Towsley and Zhang (2002), and Vojnovic and Le Boudec (2002). These control/management programs have proved very effective under a variety of circumstances, but are influenced by the packet interarrival rate as well. So therefore, if a network designer is contemplating invoking one of these options, simulation could be used to test how the option in question would improve performance on his/her system, provided an adequate method could be found to describe the distribution within that network.

Simulation has been used for many years in network design; however, the time and cost of its use have often been prohibitive. In recent years, new windows-based point and click products such as Comnet III (and its successors Network & Simscript) have eliminated the drudgery and the cost of writing simulations via a command line interface (CACI, 1998). Under Comnet III the appropriate devices are selected, connected together and their characteristics defined. There is still a limiting factor in this process: the definition of the distribution of the packet inter-arrival rates.

The theoretical model often used to describe computer networking is the Poisson. This model may have been adequate for some of the first single tier, single protocol networks. However, it lacks validity in today's hierarchically complex multi-protocol networks. In the classical Poisson process model (such as M/M/1), when the number of arrivals follows a Poisson probability distribution, then the time between arrivals (inter-arrival time) follows a decaying exponential probability distribution. A number of studies confirm that the actual inter-arrival distribution of packets is not exponential as would be expected in the classical model (Guster, Robinson & Richardson, 1999; Krzenski, 1998; Partridge, 1993; Vandolore, Babic & Jain, 1999).

The inter-arrival distribution selected can have a major impact on the results of the simulation (Guster, Safonov, Hall & Sundheim, 2003; Guster, Sohn, Robinson & Safonov, 2003). In a study by Krzenski (1999) that analyzed the simulated performance on a shared Ethernet network, 12 different inter-arrival distributions were tried within the same simulation problem. These included the gamma distribution, which is a generalization of the exponential distribution, allowing for a modal inter-arrival time (the most commonly occurring time between arrivals) to be moved out away from the very short, nearly instantaneous time occurring with the exponential distribution. Another distribution among the 12 was an integer distribution, whereby equal probabilities are assigned to different values that are equally spaced throughout the possible inter-arrival times. Among the 12 distributions, there were vast discrepancies in the results. For example, the number of collision episodes varied from 310 with a gamma distribution to 741 with an integer distribution. These results further support the need to have the correct distribution in simulations designed to provide design and management feedback about computer networks.

The frustration of the past work and the need for additional research is best summarized by Partridge (1993):

"…We still do not understand how data communication traffic behaves. After nearly a quarter of a century of data communication, researchers are still struggling to develop adequate traffic models. Yet daily we make decisions about how to configure networks and configure network devices based on inadequate models of data traffic. There is a serious need for more research work on non-Poisson queuing models."

A number of different strategies have been employed in the development of models used to describe packet inter-arrival rates (Van Slooten, 2002). Perhaps the most valid is to record all of the packet arrival times for the time period desired and use that to generate the distribution. The advantage of this strategy is accuracy, but it often requires massive amounts of data to be recorded and processed. To lessen this burden, often a representative sample from the time period is used. However, validating the sample period is often difficult, especially if the file size is not large. Known distributions have been used with limited success (Guster & Robinson, 1994, 2000; Robinson & Juckel, 2000). For simple networks exponential distributions provide some promise; however, they fail to deal with the intricacies of complex multi-protocol networks. Tabular distributions, in which one column describes the interval and a second column describes the probability of a value from that interval occurring, offer a moderate degree of accuracy, but they take time to derive and their sophistication is related to the number of rows included. Regression and ANOVA have been used in some cases, but lack the ability to describe the peaks and valleys associated with packet arrival data. Time series deals with these variations better but still lacks the sophistication needed and requires relatively complex models to even come close (Guster & Robinson, 1993). Packet trains are very effective in describing packet traffic from a single session (such as telnet), but lack the complexity to deal with multiple concurrent sessions on the same network.

FUTURE TRENDS

Two non-Poisson queuing models have offered a degree of promise. One method involves viewing the observation interval as containing several independent Poisson processes rather than as a single exponential distribution. In a study by Guster, Robinson and Richardson (1999), actual data were analyzed and shown to contain three Poisson processes of differing characteristics. During the first phase activity was increasing. During the second phase activity was decreasing. The last phase followed a classical Poisson model. For each phase, a power law process model was fit to the data, indicating the nature of the changing traffic intensity ($b>1$ – increasing intensity, $b<1$ – decreasing intensity, $b=1$ – constant intensity). These data were taken over a 24-hour period. The three phases had widely different levels of traffic intensity. In a shorter time frame, for example 10 minutes, one is less likely to see differences in intensity that dramatic. Thus, the power law process model is most appropriate for longer time frames.

A second strategy focuses on the influence any given data point has on later data points. In other words, does knowing the magnitude of the packet inter-arrival rate at any point in the time interval make it easier to predict the next inter-arrival time? Historically, there has been a tendency to apply statistical treatments such as regression, ANOVA, or time series analysis to categorize or forecast inter-arrival trends (Frieberger, 1972). These methods have proved to be limited in accuracy in modeling inter-arrival rates and require large, truly representative databases to calculate these values. Therefore, the attractiveness of methodologies such as Markov chains that use a more independent technology is apparent (Robert & Le Boudec, 1996). Specifically, a Markov chain is a tool for modeling how processes behave over time. We classify the process (in this case, the inter-arrival times of packets) into categories called "states". The assumption behind Markov Chains is that the probability for what state will be observed next depends only on the previous state. In this context, a process depends only on the most recent inter-arrival time to determine the probability of subsequent inter-arrival times increasing or decreasing

Figure 1. Frequency counts of packet inter-arrival times for ethernet data

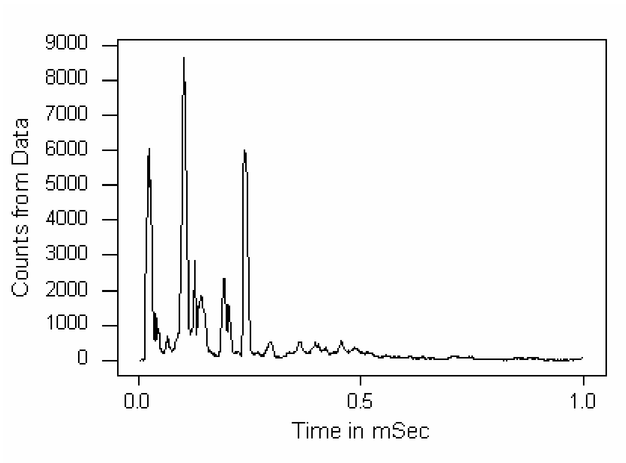
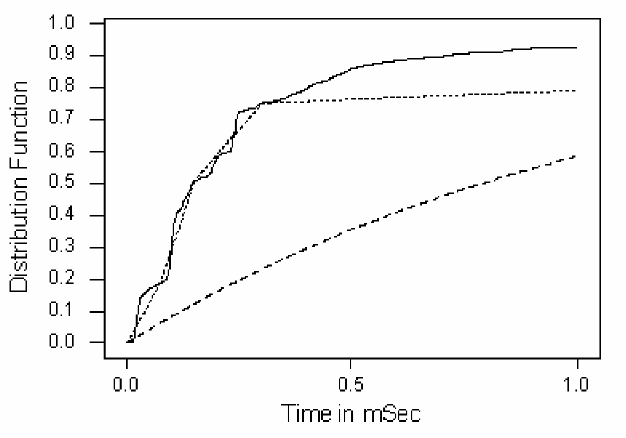


Figure 2. Density function for the actual data, the Poisson



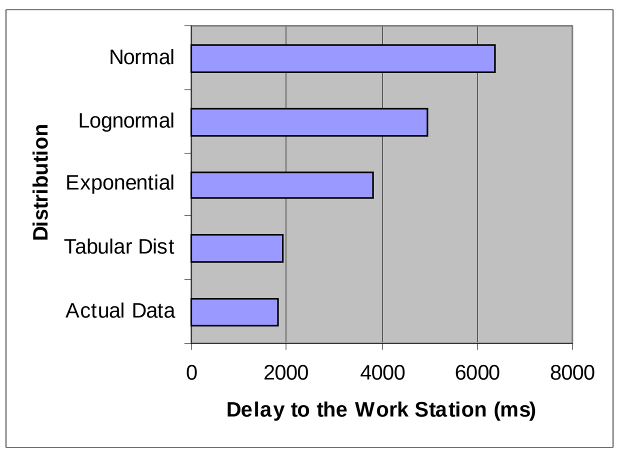
(Guster & Robinson 1994, 2000).

To provide the reader with a visual description of packet inter-arrival times from a complex multi-protocol network, a frequency plot is provided in Figure 1. It is the massive peaks and valleys from 0 to about 0.25 milliseconds that make modeling a packet distribution so difficult.

There are fairly great discrepancies in the density plots of the various distributions. For example, using the actual data from Figure 1, two attempts to model the cumulative distribution of the inter-arrival times are graphed with the actual distribution. Figure 2 depicts the actual data with the solid line, the Poisson model with long dashed line and the Markov model with the short dashed line. From the data in Figure 2 it is clear that the Markov model fits quite a bit better than the classic Poisson model.

To illustrate the effect not having the correct distribution can have on the results of a simulation, a 10-minute sample of packet traffic (packet inter-arrival times) was fed into a database inquiry simulation program. The delay in milliseconds from inquiry to response from that database was recorded. From that sample a tabular distribution was devised and placed into the same simulation and the results recorded. Then the same simulation was run three more times using exponential, lognormal and normal distributions with the appropriate means and standard deviations derived from the original 10-minute sample. The normal distribution is the classical mound shaped probability distribution used in statistical applications. The lognormal distribution is the result of exponentiating values from a normal distribution, giving a more realistic, asymmetric distribution for the inter-arrival times. Both the normal and lognormal distributions, like the gamma distribution, have modal values away from the very short, nearly instantaneous times occurring with the exponential distribution. The results are depicted in Figure 3. The tabular distribution was the closest to the actual, but a

Figure 3. Work station delay observed by varying the packet inter-arrival distribution



slight overestimate. All three of the other distributions were overestimates by quite a significant magnitude. From these results it is clear that the validity of any network simulation involving packet traffic is dependent upon using the appropriate distribution.

CONCLUSIONS

As the need to make the most of available network bandwidth increases, the importance of having valid simulation techniques available to test optimization methodologies increases. The key to this validity is an inter-arrival distribution that is truly representative of the actual data. This article explored several alternatives in selecting this distribution.

First, the actual data obviously offer the best accuracy, but is often impractical due to its massive size. Furthermore, selecting a representative sample is often challenging due to widespread variation over time exhibited by most network traffic.

Second, the exponential distribution (or any other known distribution), which according to queuing theory should be appropriate, does not fit the data well in a number of studies.

Third, the power law process has limited value in relatively short time interval studies. It does, however, offer more promise in data sets involving very large time spans.

Fourth, the Markov process has exhibited somewhat promising results. In fact, the visual plots (Figure 2) reveal a much closer fit than the Poisson model. However, Markov models have displayed limitations in the middle time range (.3-5 milliseconds), which make them far from a perfect choice (Van Slooten, 2002).

Therefore, much additional work is needed. A simple visual inspection of Figure 1 reveals the complexity of the data and subsequently the difficulty in determining a mathematical model that would truly represent them. The results exhibited by the Markov process are encouraging, and studies that examine more sophisticated Markov-related models should be encouraged.

The importance of obtaining the appropriate distribution for network traffic should not be underestimated. Without it, simulations designed to ascertain network design efficiency, network performance, and the effectiveness of software optimization techniques generate invalid results.

REFERENCES

CACI. (1998). *Comnet III reference manual*. CACI Products Inc., La Jolla, CA.

Cruz, R. (1991, January). A calculus for network delay. *IEEE Transactions on Information Theory*.

Cruz, R. (1995, August). Quality of service guarantees in virtual circuit switched networks. *IEEE Journal of Selected Areas in Communication, special issue on Advances in the Fundamentals of Networking*.

Cruz, R., & Tsai, J. (1996, February). COD: Alternative architectures for high speed packet switching. *IEEE/ACM Transactions on Networking*.

Firoiu, V., Le Boudec, J., Towsley, D., & Zhang, Z. (2002, September). Theories and models for Internet quality of service. Proceedings of the IEEE, *90*(9), 1565-1591.

Fortier, P.J., & Desrochers, G.R. (1990). *Modeling and analysis of local area networks.* CRC Press.

Frieberger, W. (1972). *Statistical computer performance evaluation.* Academic Press.

Guster, D., & Robinson, D. (1993, May 24-26). *The application of Box-Jenkins time series analysis to performance problems in computer networks.* A paper presented at the Information Resources Management Association Conference, Salt Lake City, UT.

Guster, D., & Robinson, D. (1994, April 29-30). *Markov chains as a predictor of performance decay in a PC-based LAN environment.* A paper presented at the Small College Computing Symposium, Winona, MN.

Guster, D., & Robinson, D. (2000, November 19-21). *Using Markov chains to analyze the inter-arrival distributions of ATM and ethernet traffic in computer networks.* A paper presented at the 2000 DSI Annual Meeting, Orlando, FL.

Guster, D., Robinson, D., & Juckel, A. (2000, May 21-25). *Differences in the interarrival rate distributions between ATM and high-speed ethernet and their implications on computer network performance.* A paper presented at the 11th Annual Information Resource Management Association International Conference, Anchorage, AK.

Guster, D., Robinson, D., & Richardson, M. (1999, April 22-24). Application of the power law process in modeling the inter-arrival times of packets in a computer network. *Proceedings of the 30th Annual Meeting of the Midwest Decision Sciences Institute,* Springfield, IL.

Guster, D., Safonov, P., Hall, C., & Sundheim, R. (2003). Using simulation to predict performance characteristics of mirrored hosts used to support WWW applications. *Issues in Information Systems, 4*(2).

Guster, D., Sohn, C., Robinson, D., & Safonov, P. (2003). A comparison of asynchronous transfer mode (ATM) and high speed ethernet and the network design implications to a business organization. *Journal of Information Technology and Decision Making, 2*(4).

Krzenski, K. (1998). *Analysis of the predictive process request-response modeling in a hypermedia environment.* Masters thesis. St. Cloud State University.

Krzenski, K. (1999). *The effect of varying the packet interarrival distribution in the simulation of ethernet computer networks.* Unpublished graduate research project. St. Cloud State University.

Le Boudec, J. (1998). Application of network calculus to guaranteed service networks. IEEE Transactions on Information Theory, *44*(3).

Partridge, C. (1993). The end of simple traffic models. Editor's note. *IEEE Network, 7*(5).

Robert, S., & Le Boudec, J. (1996, October). On a Markov modulated chain with pseudo-long range dependences. Performance Evaluation, 27-28, 159-173.

Vandolore, B., Babic, G., & Jain, R. (1999). *Analysis and modeling of traffic in modern data communications networks.* A paper submitted to the Applied Telecommunication Symposium.

Van Slooten, K. (2002). A comparison of stochastic models for the inter-arrival times of packets in a computer network. *Optimal information modeling techniques.* Hershey, PA: Idea Group.

Vojnovic, M., & Le Boudec, J. (2002, October). Stochastic bound on delay for guaranteed rate nodes. IEEE Communications Letters, *6*(10), 449-451.

Walker, J. (2000). *Testing and tuning QoS for network policies.* Technical Paper. Net IQ Corporation.

KEY TERMS

Inter-Arrival Distribution: The probability density function that describes likely and unlikely inter-arrival times for packets.

Inter-Arrival Time: The amount of time that elapses after the receipt of a packet until the next packet arrives.

Markov Chain: A model that determines probabilities for the next event, or "state," given the result of the previous event.

M/M/1 Model (exponential/exponential with one server): The queuing model that assumes an exponential distribution for inter-arrival times, an exponential distribution for service times, and a single server.

Packet: A finite stream of bits sent in one block with header/and trailer bits to perform routing and management functions.

Power Law Model: A generalization of the classical Poisson model, allowing for changes in the intensity of the arrivals.

Quality of Service (QOS): A method of marking certain packets for special handling to ensure high reliability or low delay.

Reservation Bandwidth: Reserving a portion of the available bandwidth for a given protocol or application, which ensures its network access will not be adversely affected by massive network traffic generated by other protocols or applications.

Parallel and Distributed Multimedia Databases

P

S. Geisler
University of Clausthal, Germany

O. Kao
University of Paderborn, Germany

INTRODUCTION

Sensing and processing of multimedia information is one of the basic traits of human beings. The development of digital technologies and applications allows the production of huge amounts of multimedia data. The rapidly decreasing prices for hardware such as digital cameras/camcorders, sound cards and the corresponding displays led to wide distribution of multimedia-capable input and output devices in all fields of the everyday life, from home entertainment to companies and educational organisations. Thus, multimedia information in terms of digital pictures, videos, and music can be created intuitively and is affordable for a broad spectrum of users.

An important question in this context is related to the archiving of the acquired information. The old-fashioned albums with pictures from holidays, children, special occasions, and so forth are replaced by photo-CDs and DVDs. Analogously, digital videos are edited, valorised by including meta-information (occasion, place, date …) and archived on DVDs. If a particular scene, image, or sound file is needed, then one can use its memory to find the corresponding medium. This type of organisation is surely not applicable to large multimedia archives, which often exist in industrial and educational sectors and where Petabytes worth of multimedia data are produced year for year. All this information has to be systematically collected, registered, stored, organised, and classified. Therefore, in many branches professional archives for such multimedia information are established, such as document management systems, digital libraries, photo and video archives used by public authorities, corporations, broadcasting and TV companies, as well as archives for satellite and surveillance photos. The scope and spread of such systems grow day by day and lead to new demands for efficient retrieval of the archived information based on user-specific description of the sought image, video or audio.

The search for a medium similar to the given one is, due to the complexity of multimedia information, a very challenging problem and requires a number of novel mechanisms. Beside the search procedures, also methods to formulate queries, and ways to visualise the results have to be provided. Moreover, the search has to be performed efficiently in order to achieve acceptable response times for the user. Therefore, a combination of modern multimedia archives with powerful parallel and distributed architectures described in this article is mandatory for the integration of multimedia retrieval into real-world applications.

BACKGROUND

The necessity for organisation and retrieval of multimedia data led to development of a large number of prototypes and operational multimedia database management systems, which manage the multimedia data in terms of storage, annotation, and retrieval. In the early years this task was tended to by existing database management systems (DBMS) with multimedia extensions. The basis for representing and modelling multimedia data in such systems is so-called Binary Large Objects (BLOBs), which store images, video and audio sequences without any formatting and analysis done by the system. The media are saved in the current form in the database and their additional information – called meta-information – is inserted into the database tables. Typically, the file name, categories and additional key words entered by the user serve as meta-information. Once the user submits a key word about the sought media, the blocks with meta-information are searched using the existing database functions and compared with the input. In case of a key word match, the corresponding media is presented.

These extensions reflect a certain aspect of multimedia database systems, but this approach does not satisfy the requirements of multimedia archives, as the manual annotation of the media is too time-consuming and not applicable in real-world applications. Furthermore, key words are not sufficient to represent content of images or videos entirely (*An image says more than 1000 words*). Therefore, the media annotation and retrieval has to be content-based; that is, features describing the multimedia content have to be extracted automatically from the media itself and compared to the corresponding features of the sample medium. The functionality of such a multimedia database is well defined by Khoshafian and Baker, (1996):

"A multimedia database system consists of a high performance DBMS and a database with a large storage capacity, which supports and manages, in addition to alphanumerical data types, multimedia objects with respect to storage, querying, and searching."

The DBMS is already provided by traditional databases and therefore will not be discussed in the following sections. Instead, the focus is set on the mechanisms for multimedia retrieval and high-performance implementation.

MULTIMEDIA RETRIEVAL

The content-based annotation of multimedia data requires the integration of additional information, which can be classified into the following categories:

- Technical information describes details of the recording, conversion, and storage. Examples: filename, resolution, compression, frame rate.
- Extracted attributes are features that are deduced by analysing the content of the media directly. Examples: average colour or colour histograms of an image, camera motion in videos, pitch in audio files.
- Knowledge-based information links the objects, people, scenarios, and so forth detected in the media to entities in the real world.
- World-oriented information encompasses information on the producer of the media, the date and location, language, and so forth.

Technical and world-oriented information can be modelled with traditional data structures. The knowledge-based information assumes semantic analyses of the media, which is nowadays still not possible in general. However, many recent research efforts in this direction promise the applicability of semantic information in the future (Zhao & Grosky, 2002).

Most of the existing multimedia retrieval systems are specialised to work on media of a limited domain, for example news (Christel & Hauptman, 2002; Yang & Chairsorn, 2003), American football (Li & Sezan, 2002), or integrate general retrieval algorithms like face, speech or character recognition. They use features extracted from the media content to annotate and retrieve the multimedia objects, which are usually related to the colour, edge, texture, layout properties in case of images or consider object motion in case of videos or specific tone sequences for audios. Table 1 gives an overview of several prominent, specialised systems, which introduced main research retrieval concepts to the scientific community. Meanwhile, many of these systems became a part of commercial products or a part of a general multimedia archive. A survey is provided in Venters and Cooper (2000).

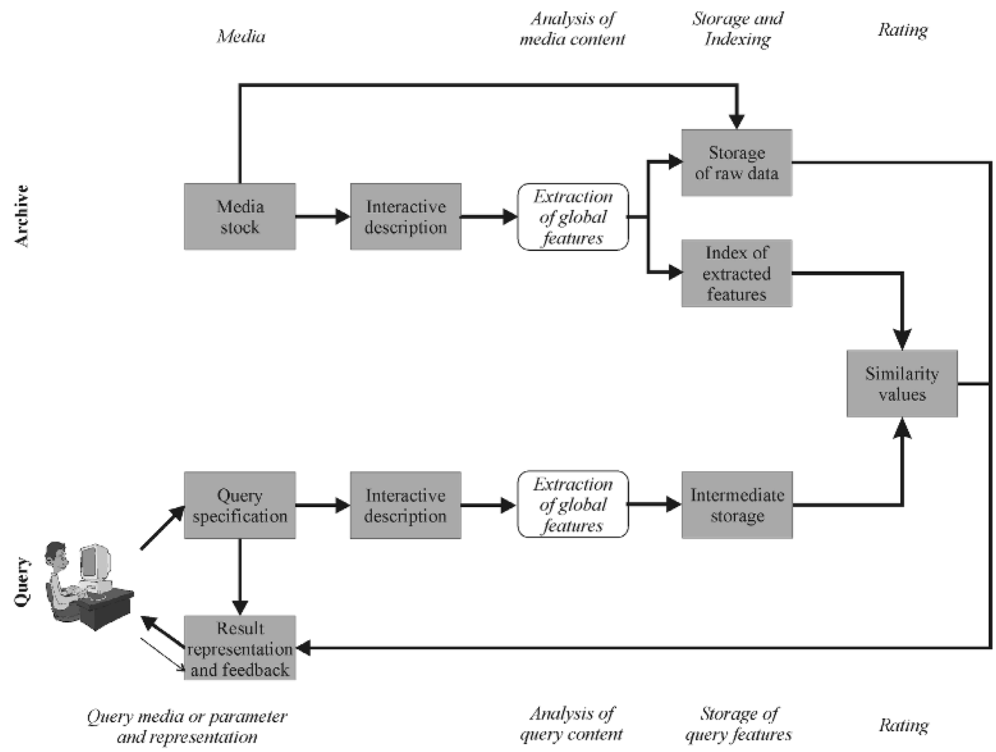
In the following, the retrieval workflow for multimedia data will be depicted by considering images as an example. The user has a specific image in mind and starts a query for this specific image or for similar samples. The user can for example browse the data set or give a suitable key word. However, for content-based similarity search, sophisticated interfaces are necessary:

- Query by pictorial example: the user supplies the system with a complete sample image, which is similar to the sought one.
- Query by painting: the user sketches the looked-for image with a few drawing tools (Rajendran & Chang, 2000).
- Selection from standards: lists of sample instances – called standards – can be offered for individual features.
- Image montage: the image is composed of single parts similar to a mosaic.

Table 1. Prominent examples for image, video and audio databases.

- Image databases
  - Qbic (Flickner, Sawhney et al., 1995)
  - Photobook (Pentland, Picard & Sclaroff, 1994)
  - Surfimage (Nastar, Mitschke et al., 1998)
- Audio databases
  - VARIATIONS (Dunn & Mayer, 1999)
  - MUSART (Birmingham, Dannenberg et al., 2001)
- Video databases
  - VideoQ (Chang, Chen et al., 1998)
  - Virage Video Engine (Hampapur, Gupta et al., 1997)
  - CueVideo (Ponceleon, Srinivasan et al., 1998)

Figure 1. General retrieval workflow



All approaches have individual advantages as well as shortcomings; thus a suitable selection depends on the domain. However, the query by pictorial example is one of the most often applied methods since it provides the greatest degree of flexibility.

In the next step the query image is compared to all archived images in the database based on the extracted features. Each feature emphasises one or more aspects of the image, usually related to colour, texture and layout. These features are extracted off-line (at creation time) for the archived images and online (at query time) for the given sample image. Thus, after the analytical phase, the images are represented by consistent feature vectors, which can be directly compared using similarity metrics or functions. For this purpose well-known metrics such as the (weighted) Euclidian Distance or specially developed and adapted metrics like Earth Mover Distance can be applied (Rubner & Tomasi, 2000).

The result of this comparison is a similarity value for the query and the analysed image. The process is repeated for all *n* images in the database, resulting in a similarity ranking. The first *k* entries, *k* being a user-defined constant, represent the *k* best hits, whose raw data are then displayed. Figure 1 describes the global workflow.

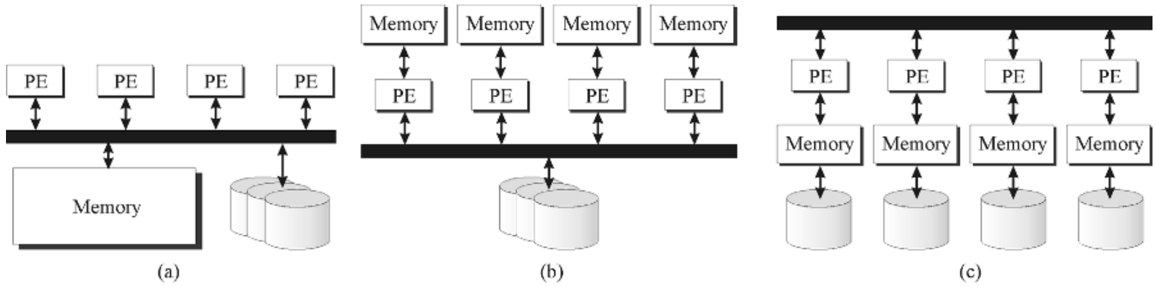
The comparison process can be accelerated by using index structures. These contain a-priori extracted features and are organised in a way that the comparisons can be focused to a certain area around the query. A novel approach and an overview of often-applied index structures are provided in Tuncel and Ferhatosmanoglu (2002). A survey of specialised retrieval techniques is provided for example in Rui (1999).

HIGH-PERFORMANCE COMPUTING FOR MULTIMEDIA RETRIEVAL

The database architecture – software and hardware – is decisive for the efficiency and thus for the usability of a multimedia archive. Because of their high storage and computing requirements, multimedia databases belong to those applications which rapidly hit the limits of the existing technology.

The widespread client/server architectures are – in their usual form – not fit for multimedia database implementations. Firstly, a centralised organisation of a media server requires immense storage and computation resources. With a growing number of user queries and data

Figure 2. Classification of parallel architectures: a) shared everything; b) shared disk; c) shared nothing



to be organised, such a centralised system will quickly reach the borders of its capabilities and the quality of service is no longer fully sustainable.

One possible solution for this problem is offered by distributed and parallel architectures, where multiple processing elements (PEs) work cooperatively on an efficient solution of a large problem. The data and the programs are spread over several nodes, so that the processing is accelerated (parallel processing) or the path to the user is shortened (distributed processing, e.g., video-on-demand servers).

There are many possible ways for the organisation of such architecture; the well-known are shared everything, shared disk, and shared everything architectures shown in Figure 2.

Due to the shared utilisation of the hard disks, buses, and memory in case of shared disk and shared everything architecture respectively, the data throughput in these architectures is rather low and the response time long. Experimental measurements executed in Bretschneider, Geisler and Kao (2002) showed that shared nothing architectures – in particular cluster-based systems – with an appropriate data distribution across the multiple nodes are most suitable for the realisation of an efficient multimedia retrieval. A node is defined as an entity consisting of a PE, a memory, a storage resource, and a network adapter.

The database running on this architecture can be denoted as a parallel and distributed database: The data are distributed over all nodes, but logically combined in a single entity, making it look like a single database system from the outside. The only difference noticeable to the user is the performance improvement of the system, which results from utilising the parallelism in the computer network. Further characteristics concern the optimisation of queries, controlling parallelism, recovery, integrity, and security. All these aspects are already well defined for conventional database systems, and have been analysed in detail. However, the possibility of a geographically separated data distribution introduces new constraints and requires additional communication and synchronisation mechanisms. The problems grow more complex when heterogeneous computer architectures, long processing times per data set – as in multimedia applications –, hardware and software failures, and so forth have to be considered.

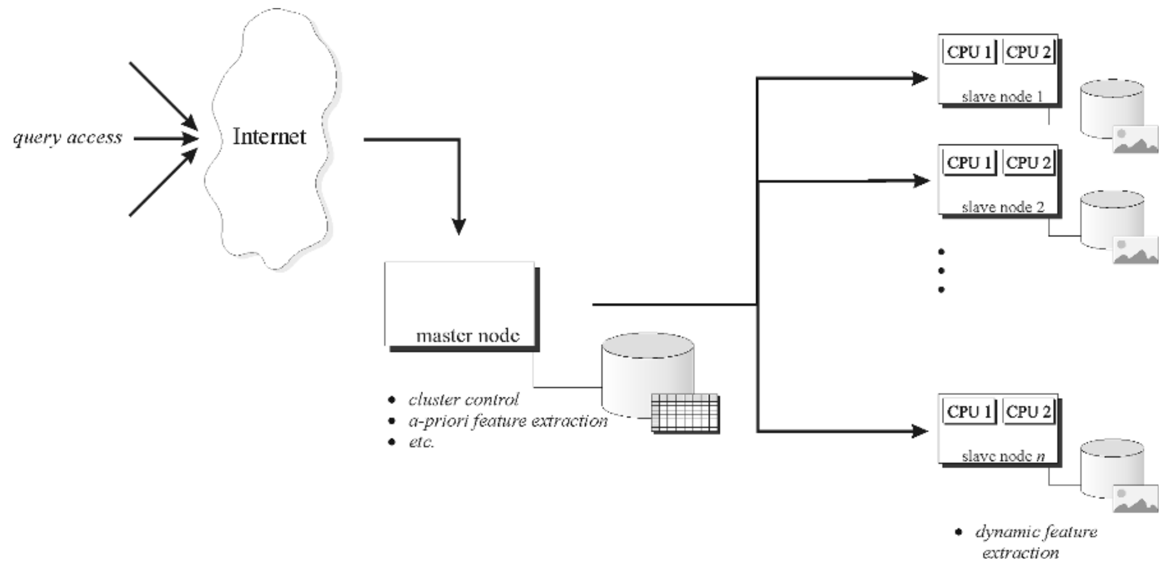
As already noted, the data distribution is a crucial efficiency aspect and has to fulfil numerous, sometimes even conflicting requirements: The data needed by an operation should – if possible – all be on one node (*data locality*). On the other hand, as many operations as possible should be processed in parallel, that is, the data should be distributed evenly among all available nodes. A drawback of the broad distribution is given by the time-consuming data transfer between the individual nodes, which affects the performance significantly. Thus a general distribution is not applicable; hence it has to be tailored for the current application. One possible solution will be presented in the following by considering the cluster-based multimedia database Cairo as an example (Kao & Stapel, 2001).

The general Cairo architecture shown in Figure 3 consists of:

- Query stations host the Web-based user interfaces for the access to the database and for the visualisation of the retrieval results.
- Master node controls the cluster, receives the query requests and broadcasts the algorithms, search parameters, the sample media, features, and so forth to the compute nodes. Furthermore, it unifies the intermediate results of the compute nodes and produces the final list with the best hits.
- Compute nodes perform the image processing and comparisons. Each of these nodes contains at least one partition and executes all operations with the data stored on the local devices. The computed intermediate results are sent to the master node.

The distribution of the media stock across the individual cluster nodes is created according to the following requirements:

Figure 3. Schematic of the Cairo cluster architecture



- Similar storage sizes of the partitions and thus an even distribution of the media across the individual nodes,
- Computation reduction for the retrieval, and
- Minimising the communication between the cluster nodes.

Cairo uses a content-independent, size-based strategy for the available media stock that creates a set of partitions $P = P_1 \cup P_2 \cup \ldots \cup P_n$ where $P_i \cap P_j = \emptyset$ and size(P_i) $\approx$ size(P_j) for all i,j, i$\neq$j. The current image under investigation is assigned to the node with the least storage space used. This process is repeated until all images are distributed to the partitions.

The processing of a partition P_i is executed per media instance. The individual operations are independent of one another, so the order of execution is irrelevant. This initial partitioning makes it possible for all nodes to have uniform processing times, assuming a homogenous, dedicated execution platform. The management overhead depends on the operator applied and the structure of the partial results. However, this time is usually neglectable, compared to the media processing times.

The advantage of a static distribution is that expensive computations and data transfers are not necessary during runtime and the partitioning can be manually optimised and adapted to a given application. On the other hand, short-term reactions to variable workloads among the nodes are not possible. A dynamic distribution of the data is done during runtime and considers the current workload of the nodes, as well as the number and the structure of pending queries. Idle times can be minimised by continuously re-distributing the data, and increasing the total throughput. Unfortunately, the long communication time for multimedia data often eliminates all performance advantages gained through the parallel processing.

After the partitions are created, these are distributed evenly over the cluster nodes, allowing parallel retrieval by executing the same operations on all nodes with the local media subset. The implementation is based on the following components:

- Transaction manager encompasses the analysis of the transformations to be executed and the order of the operations.
- Distribution manager receives a list of algorithms to be executed and a set of media instances as input. Then the programs for the analysis and comparison are compiled and sent to the computing managers.
- Computing manager controls the execution of the extraction algorithms with the local data. The process runs on each cluster node and supervises the communication with the master node.
- Result manager: The data partitioning in disjoint subsets results in a sub-ranking per node that need to be unified by the result manager.
- Update manager realises the media insertion in the database. The raw data are transformed in a uniform format and tagged with a unique identifier. All procedures for feature extraction are then applied. Finally, the cluster node, on whose hard disk the raw data are to be stored, is determined.

The described organisation was successfully implemented and evaluated in real-world scenarios. The minimised communication between the nodes allows a nearly full utilisation of the available compute resources and leads to a linear speedup; that is, the response time is shortened by a factor equal to the number of included nodes.

Other performance relevant parts of a multimedia database system are still the subject of research. Examples are given by the disk scheduler for multimedia storage (Huang & Huang, 2004) and scheduling for broadcasting videos and audios (Yoshihisa & Tsukamoto, 2003).

FUTURE TRENDS

Despite the immense development of multimedia in recent years, the efficient organisation and retrieval of multimedia data still remains a large scientific challenge. The growing number of multimedia data and new applications fields are the driving forces.

The consideration of semantic information is necessary to realise a human-like retrieval approach for images, video and audio and to integrate multimedia components in the semantic Web for ubiquitous access. Furthermore, the current restriction of the most databases to a single media type will be eliminated by methods for multimodal retrieval. For example, the editor of a newscast needs access to videos from news agencies, images from photographs, audios files from radio and text from newspaper articles. Merging the information into a single database requires methods to classify all these media instances and measure the content-based similarity between them. Speech recognition for example can be used to compare audio or video files with textual information. Multimodal access accommodates different user profiles or device capabilities and can help to receive improved query results.

Not only different media types but also media data from different providers have to be merged to create a large distributed multimedia warehouse or a multimedia grid. MPEG-21 is designed for the content description and digital rights management. Load balancing, data distribution and quality of service considerations are necessary to achieve high performance and consumer satisfaction in a very heterogeneous environment. Finally, the multimedia information has to be accessible for mobile devices using standards such as UMTS.

CONCLUSION

Traditional databases offer a search for certain media instance based on describing key words. The main invention of multimedia databases is the possibility to search directly on the media content. However, this time-consuming process has to be performed in parallel to achieve acceptable response times. The cluster architecture scales well even for a high number of processing elements, but if and only if the underlying data partitioning strategy is suitable for the given application.

Further research is needed to give the user feedback on how to improve the query and how to supply the system with feedback about the relevance of delivered results. A further aspect is the improvement of the database design, especially data distribution in parallel environments. Furthermore, advanced methods for semantic content description have to be exploited.

REFERENCES

Bainbridge, D. et al. (1999). Towards a digital library of popular music. *Proceedings of the 4th ACM Conference on Digital Libraries* (pp. 161-169).

Birmingham, W.P. et al. (2001). MUSART: Music retrieval via aural queries. *Int. Symposium on Music Information Retrieval (ISMIR)*.

Bretschneider, T., Geisler, S., & Kao, O. (2002). Simulation-based assessment of parallel architectures for image databases. *Proceedings of the Conference on Parallel Computing* (pp. 401-408). Imperial College Press.

Chang, S.-F. et al. (1998). A fully automated content based video search engine supporting spatio-temporal queries. *IEEE Trans. CSVT*, *8*(5), 602-615.

Christel, M., Ng, H., & Wactlar, A.H. (2002). Collages as dynamic summaries for news video. *Proceedings of ACM Multimedia '02* (pp. 561-569).

Dunn, J.W., & Mayer, C.A. (1999). VARIATIONS: A digital music library system at Indiana University. *Proceedings of the 4th ACM Conference on Digital Libraries* (pp. 12-19).

Flickner, M. et al. (1995). Query by image content the QBIC system. *IEEE Computer Magazine*, *28*(9), 23-32.

Hampapur, A. et al. (1997). Virage Video Engine. *Proceedings of SPIE Storage and Retrieval for Image and Video Databases*, *3022*, 188-197.

Huang, Y.-F., & Huang, J.-M. (2004). Disk scheduling on multimedia storage servers. *IEEE Transactions on Computers*, *53*(1), 77-82.

Kao, O., & Stapel, S. (2001). Case study: Cairo – a distributed image retrieval system for cluster architectures. In

T.K. Shih (Ed.), *Distributed multimedia databases: Techniques and applications* (pp. 291-303). Hershey, PA: Idea Group Publishing.

Khoshafian, S., & Baker, A.B. (1996). *Multimedia and imaging databases*. Morgan Kaufmann Publishers.

Li, B., & Sezan, M.I. (2002). Event detection and summarization in American football broadcast video. *Proceedings SPIE Storage and Retrieval for Media Databases, 4676*, 202-213.

Nastar, C., Mitschke, M., Meilhac, C., & Boujemaa, N. (1998). Surfimage: A flexible content-based image retrieval system. *ACM Multimedia'98 Conference Proceedings* (pp. 339-344).

Pentland, A., Picard, R., & Sclaroff, S. (1994). Photobook: Tools for content-based manipulation of image databases. *Proceedings of SPIE Storage and Retrieval for Image and Video Databases II, 2185*, 34-47.

Ponceleon, D., Srinivasan, S., Amir, A., Petkovic, D., & Diklic, D. (1998). Key to effective video retrieval: Effective cataloging and browsing. *ACM Multimedia'98 Conference Proceedings* (pp. 99-107).

Rubner, Y., Tomasi, C., & Guibas, L.J. (2000). The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision, 40*(2), 99-121.

Rui, Y., Huang, T., & Chang, S. (1999). Image retrieval: Current techniques, promising directions and open issues. *Journal of Visual Communication and Image Representation*, *10*(4), 39-62.

Tuncel, E., Ferhatosmanoglu, H., & Rose, K. (2002). VQ-index: An index structure for similarity searching in multimedia databases. *ACM Multimedia 2002* (pp. 543-552).

Venters, C.C., & Cooper, M. (2000). A review of content-based image retrieval systems. *Technical Report jtap-054*. University of Manchester.

Yang, H., Chaisorn, L., Zhao, Y., Neo, S.-Y., & Chua, T.-S. (2003). VideoQA: Question answering on news video. *ACM Multimedia 2003* (pp. 632-641).

Yoshihisa, T., Tsukamoto, M., & Nishio, S. (2003). Scheduling methods for broadcasting multiple continuous media data. *ACM MMDB '03* (pp. 40-47).

Zhao, R., & Grosky, W.I. (2002). Bridging the semantic gap in image retrieval. In *Distributed multimedia databases: Techniques & applications* (pp. 14-36). Hershey, PA: Idea Group Publishing.

Zhao, R., & Grosky, W.I. (2002). Narrowing the semantic gap – Improved text-based Web document retrieval using visual features. *IEEE Transactions on Multimedia, 4*(2), 189-200.

KEY TERMS

A-Priori Feature Extraction: Analysis and description of the media content at the time of insertion in the database. The gained information is stored in a database and enables content-based retrieval of the corresponding media object without actually accessing the latter.

Cluster: Parallel architecture that contains multiple "standard" computers connected via a high performance network that work together to solve the problem.

Dynamic Feature Extraction: Analysis and description of the media content at the time of querying the database. The information is computed on demand and discarded after the query was processed.

Feature Vector: Data that describe the content of the corresponding multimedia object. The elements of the feature vector represent the extracted descriptive information with respect to the utilised analysis.

Load Balancing: Techniques to distribute the tasks over the single processors in parallel systems in a way that idle time is minimized.

Multimedia Database (after Khoshafian & Baker): A multimedia database system consists of a high performance database management system and a database with a large storage capacity, which supports and manages, in addition to alphanumerical data types, multimedia objects with respect to storage, querying, and searching (Khoshafian & Baker, 1996).

Quality of Service (QoS): The collective term for all demands on the recording and the replaying procedures, which refer to generating and maintaining a continuous data stream.

Query by Example/Sketch/Humming: Methods to formulate queries in multimedia databases. The user provides an example media file and the result of the database system is a set of similar media files. If no example is available the user can draw a sketch or hum a melody.

Retrieval: Accessing stored information from the database.

Personal Information Privacy and Internet Technology

Edward J. Szewczak
Canisius College, USA

INTRODUCTION

Concerns about the collection of personal information by Internet technology and the possibility of misuse of that information are a primary reason why people limit their use of the Internet and are even limiting the success of e-commerce (Szewczak, 2004). Various uses of technology that collect and/or disseminate personal information include corporate and government databases, e-mail, wireless communications, clickstream tracking, hardware and software watermarks, and biometric devices. The main challenge to personal information privacy is the surreptitious monitoring of user behavior on the Internet without the user's consent and the possible misuse of the collected information resulting in financial and personal harm to the user. Our focus is primarily on Internet use in the United States of America, though clearly the technology is global in nature and poses challenges and issues for societies around the world.

PERSONAL INFORMATION AND INTERNET USE

The results of a 1998 survey conducted by Louis Harris & Associates, Inc. revealed that worries about protecting personal information ranked as the top reason people generally are avoiding the Web (Hammonds, 1998). A survey by NFO Interactive (*www.nfoi.com*) found that the safekeeping of online consumer personal information was the main reason people chose not to shop online. The misuse of credit card data for activities such as identity theft is a major concern (*Consumer Reports*, 2003).

Some recent events serve to confirm these concerns. Italian Privacy Commissioner Stefano Rodota ordered Infostrada to temporarily shut down its free ISP service because it required users to disclose their age, health status, sexual habits and political, labor and religious preferences in order to qualify for the service. Infostrada said that the information was required for marketing purposes (*http://www.privacytimes.com/NewWebstories/oxymoron_prv_2_23.htm)*. Failed Internet companies such as Boo.com, Toysmart.com and CraftShop.com have either sold or have tried to sell customer data that may include phone numbers, credit card numbers, home address and statistics on shopping habits, even though they had previously met Internet privacy monitor Truste's criteria for safeguarding customer information privacy. The rationale for the selling was to appease creditors (Sandoval, 2000). The Foundation for Taxpayer and Consumer Rights purchased the Social Security numbers and home addresses of the CIA director and the U.S. Attorney General on the Internet for $26 (Kerr, 2003).

WHAT IS PRIVACY?

In his excellent study on privacy in the information age, Cate (1997) adopted the definition of privacy as "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others" (Westin, 1967, p. 7). Westin/Cate's definition is interesting because it allows for flexibility in discussing privacy within the context of the Internet. Whereas many people worry about divulging personal information electronically, other people seem more than willing to give it away, trading their personal information for personal benefits such as free shipping and coupons (Kuchinskas, 2000). Personalized service is the main benefit. A Web site can save a shopper time and money by storing and recalling a user's tastes and buying habits (Baig, Spepanek & Gross, 1999). ISPs are willing to allow Web users cheaper access to the Internet provided the users are amenable to having their online behavior tracked for marketing purposes by specialized software (Angwin, 2000). Spyware installs itself on computers when users download programs, and then tracks each click the user makes (Hagerty & Berman, 2003).

TECHNOLOGICAL CHALLENGES TO PRIVACY

Government regulators and enforcement officials have to consider a host of technological challenges to personal information privacy on the Internet.

Corporate and Government Databases

The practice of gathering personal information about customers and citizens by corporations and governments is well established. Software is available that is dedicated to analyzing data collected by company Web sites, direct-mail operations, customer service, retail stores, and field sales. Web analysis and marketing software enable Web companies to take data about customers stored in large databases and offer these customers merchandise based on past buying behavior, either actual or inferred. It also enables targeted marketing to individuals using e-mail. Governments routinely collect personal information from official records of births, deaths, marriages, divorces, property sales, business licenses, legal proceedings, and driving records. Many of the databases containing this information are going online (Bott, 2000). MGM Mirage's online gambling business is aided by online screening programs wherein a bettor's location is determined using Internet mapping software and personal details checked against six external databases (King, 2003).

The deregulation of the financial services industry has made it possible for banks, insurance companies, and investment companies to begin working together to offer various financial products to consumers. Personal financial information that was kept separate before deregulation can now be aggregated. In fact the ability to mine customer data is one of the driving forces behind the creation of large financial conglomerates. Services can be offered to customers based on their information profiles. Large credit bureaus such as Equifax and Trans Union have traditionally been a source of information about a person's credit worthiness. Their databases contain information such as a person's age, address, and occupation. Credit bureaus have begun to sell personal information to retailers and other businesses (*Consumer Reports,* 2000a).

Like personal financial information, medical information is for most people a very private matter. Despite this fact, there is a wealth of personal medical data in government and institutional databases. As *Consumer Reports* (2000b, p. 23) notes:

The federal government maintains electronic files of hundreds of millions of Medicare claims. And every state aggregates medical data on its inhabitants, including registries of births, deaths, immunizations, and communicable diseases. But most states go much further. Thirty-seven mandate collection of electronic records of every hospital discharge. Thirty-nine maintain registries of every newly diagnosed case of cancer. Most of these databases are available to any member of the public who asks for them and can operate the database software required to read and manipulate them.

Much of personal health information that is available to the public is volunteered by individuals themselves, by responding to 800 numbers, coupon offers, rebate offers and Web site registration. Much of the information is included in commercial databases like Behavior-Bank sponsored by Experian, one of the world's largest direct-mail database companies. This information is sold to clients interested in categories of health problems, such as bladder control or high cholesterol.

E-mail

E-mail accounts for 70% of all network traffic, yet only 10% of it is protected by security measures. Thus it is susceptible to tampering and snooping (Armstrong, 2000; Weingarten & Weingarten, 2002). In many companies, employee e-mail communications are routinely monitored. Despite the fact that most companies had policies alerting employees that they were subject to monitoring, 25% surveyed had fired employees based on evidence collected during monitoring (Seglin, 2000). Hackers can also be a problem. Programs can be surreptitiously installed that monitor a user's keystrokes. The keystrokes can be sent across the Internet to a computer that logs everything that is typed for later use (Glass, 2000).

Employees' invasion of privacy claims have not been upheld in the United States courts, which argue that, since employers own the computer equipment, they can do whatever they want with it (McCarthy, 2000).

Wireless Communications

Wireless advertising poses a host of challenges for privacy advocates. Wireless service providers know customers' names, cell phone numbers, home and/or office addresses, and the location from where a customer is calling as well as the number a customer is calling. Each phone has a unique identifier that can be used to record where in the physical world someone travels while using the cell phone (Petersen, 2000). The Federal Communications Commission requires cell phone service providers to be able to identify the location of a caller who dials 911, the emergency number. Since a cell phone service provider can track the location of a 911 call, it can track the location of any other call as well.

Clickstream Tracking

Tracking employee behavior on the Internet has become common practice. Software programs have been specifically designed to monitor when employees use the Internet and which sites they visit (McCarthy, 1999b).

Internet companies monitor Internet user behavior by a number of means, primarily to gather data about shopping and buying preferences with a view toward developing "user profiles". These technological means include capturing and examining environment variables, cache memory, and cookies (*http://www.cnil.fr/traces*).Environment variables contain data about a user's system configuration and site last visited. These variables include a user's domain name, system address, IP (Internet Protocol) address, operating system version, browser version, and URL of the last site visited. These data are transmitted with each packet of data transmitted to an Internet server, where they are extracted by a CGI (Common Gateway Interface) script (program). This capturing of data happens without the user's explicit consent.

Cache memory is a commonly available technology that was developed to improve file download time and maximize network performance. When a user requests access to a Web site, the browser checks a directory on the user's hard drive to see if the Web page had been loaded previously. If not, it records the page on the user's hard disk and displays it on the screen. The next time the same request is made, the browser reads the page from the hard disk and saves network resources by not requiring an additional server transmission. The cache memory is not hidden and may be accessed by an outside observer using another computer, thereby revealing the sites that have been visited.

Cookies are text files created by a Web server and stored on a user's hard disk. A cookie is a set of fields that a user's computer and a server exchange during a transaction. The server may change or suppress the contents of a cookie it has created. Web servers work with ad placement companies that resell advertising space from popular sites. These companies maintain large databases in which are recorded details about who looks at which pages. When a user connects to a Web site, the browser checks the cookies on the hard drive. If a cookie matches the site's URL, the browser uploads the cookie to the Web site. With the information contained in the cookie, the site can run programs that personalize site offerings and/or track the user's activity while online.

Hardware and Software Watermarks

Hardware and software identifiers (watermarks) can also be used to identify individual users.

In 2000 Intel Corporation announced it would include a unique Processor Serial Number (PSN) in its new Pentium III microprocessor chips. The rationale for the PSN was that it was to be used for authentication purposes in e-commerce insofar as the PSN would be linked to a person's real-world identity. Though Intel later decided not to include a PSN in its chips, the possibility of a privacy challenge was evident (*http://www.bigbrotherinside.org*).

Every Ethernet card used in computer communications has its own MAC (Medium Access Control) address, a 48-bit number sent in the header of every message frame. As the Ethernet standard evolves into a wide-area communications protocol, this identifier may become of increasing concern to Internet users intent on protecting their privacy (Dornan, 2000).

Microsoft Corporation includes a unique numeric identifier into every copy of its Office program. When a Microsoft Office document is created, it is watermarked with this unique identifier (*http://www.forbes.com/Forbes/99/1129/6413182s1.htm*).

Biometric Devices

Various devices are available that identify people through scans of their faces, hands, fingers, eyes, or voice. Biometric devices create a statistical profile by assessing a number of biological characteristics. Using biometrics it is possible to scan millions of faces and other characteristics into a computer database. Privacy advocates object to the fact that much of the measurement taking happens without the knowledge or explicit cooperation of a subject, which can lead to abuses of the technology (Stepanek, 2000). However, seven states have chosen to use Viisage Technology's facial recognition systems to lessen identify theft. Scientists working for the U.S. Commerce Department's National Institute of Standards and Technology have recommended that a combination of fingerprint and facial technologies be used to secure U.S. borders (Armstrong, 2003).

FUTURE TRENDS

The issue of personal information privacy and the Internet continues to be debated within the community of Internet users. Privacy is a social issue, generally speaking. How the personal information privacy debate is ultimately resolved will be decided by the values inherent in a society. Since the position of the privacy advocates differs so markedly from the position of the technology growth advocates, and since privacy issues have been addressed in court and precedents established in state and common law, it seems likely that the personal information privacy debate will be resolved in the world's legislatures and resulting laws enforced in the courts (Lessig, 1999). The U.S. has seen the passage of the Health Insurance Portability and Accountability Act that prohibits unauthorized disclosures of personal medical information, punishable by a fine up to $250,000 and 10 years in jail (Lueck, 2003). Other nations will have to adopt their own measures to ensure the privacy of their citizens' personal information. Since the Internet is a global technology, perhaps a global effort will be needed to adequately address these issues.

CONCLUSION

Various uses of technology that collect and/or disseminate personal information include corporate and government databases, e-mail, wireless communications, clickstream tracking, hardware and software watermarks, and biometric devices. Clearly the challenges to personal information privacy posed by the various forms of Internet technology are not the result of the technology itself. Rather it is the uses of the technology that pose the threat to the integrity of personal information privacy. In particular, the surreptitious monitoring of user behavior without the user's consent and the possible misuse of the collected information pose the biggest threats. The various nations of the world must come to grip with these challenges in their own way, perhaps working in unison. Otherwise the personal information privacy of their citizens will be continuously at risk.

REFERENCES

Angwin, J. (2000, May 1). A plan to track Web use stirs privacy concern. *The Wall Street Journal*, B1f.

Armstrong, I. (2003, May). Biometrics finding a niche at last. *SCMagazine*, 26-28.

Armstrong, L. (2000, July 10). Someone to watch over you. *Business Week*, 189-190.

Baig, E.C., Stepanek, M., & Gross, N. (1999, April 5). Privacy. *Business Week*, 84-90.

Bott, E. (2000, March). We know where you live, work, shop, bank, play…and so does everyone else! *PCComputing*, 80-100.

Cate, F.H. (1997). *Privacy in the information age*. Washington, D.C.: Brookings Institution Press.

Dornan, A. (2000, June). Internet indiscretions. *Network Magazine*, 100-104.

Glass, B. (2000, June 6). Keeping your private information private. *PC Magazine*, 118-130.

Hagerty, J.R., & Berman, D.K. (2003, August 27). New battleground in Web privacy war: Ads that snoop. *The Wall Street Journal*, A1f.

Hammonds, K.H. (Ed.). (1998, March 16). Online insecurity. *Business Week*, 102.

Kerr, J.C. (2003, August 28). Group buys data on Ashurst. Tenet to expose privacy flaws. *The Buffalo News*, A-5.

King, R. (2003, April). Hedging their bets. *Business 2.0*, 91.

Kuchinskas, S. (2000, September 12). One-to-(N)one. *Business 2.0*, 141-148.

Lessig, L. (1999). *Code and other laws of cyberspace*. New York: Basic Books.

Lueck, S. (2003, March 19). Tough new law helps to guard patient privacy. *The Wall Street Journal*, D1f.

McCarthy, M.J. (1999, October 21). Now the boss knows where you're clicking. *The Wall Street Journal*, B1f.

McCarthy, M.J. (2000, April 25). Your manager's policy on employee's e-mail may have a weak spot. *The Wall Street Journal*, A1f.

Petersen, A. (2000, July 24). Coming to phone screens: Pitches, privacy woes. *The Wall Street Journal*, B1f.

Sandoval, G. (2000, July 1). Sensitive data on customers being sold by failed e-retailers. *The Buffalo News*, A9.

Seglin, J.L. (2000, August 8). Who's snooping on you? *Business 2.0*, 200-203.

Stepanek, M. (2000, May 8). Are they selling your face? *Business Week*, 106E6.

Szewczak, E.J. (2004). Personal information privacy and EC: A security conundrum? In M. Khosrow-Pour (Ed.), *E-commerce security: Advice from experts* (pp. 88-97). Hershey, PA: Idea Group Publishing.

Weingarten, M., & Weingarten, A. (2002, January). Email tampering – This time, the good guys win. *Business Communications Review*, 46-48.

Westin, A.F. (1967). *Privacy and freedom*. New York: Atheneum.

Westin, A.F. (2000a, May). Big browser is watching you! *Consumer Reports*, 43-50.

Westin, A.F. (2000b, August). Who knows your medical secrets. *Consumer Reports*, 22-26.

Westin, A.F. (2003, October). Stop thieves from stealing you. *Consumer Reports*, 12-17.

KEY TERMS

Biometrics: The use of technological devices to identify people through scans of their faces, hands, fingers, eyes or voice.

Cache Memory: A technology developed to reduce file download time and maximize network performance.

Clickstream Tracking: The use of software to monitor when people use the Internet and what sites they visit.

Cookies: Text files created by a Web server and stored on a user's hard disk that contain data about which Web sites have been visited.

Environment Variables: System variables whose values contain data about a user's system configuration and Web site last visited.

Identity Theft: The stealing and use of a person's identity through the acquisition of personal information without that person's knowledge or permission.

Personal Information: Information about, or peculiar to, a certain person or individual.

Privacy: The claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others (Westin).

Spyware: Software that installs itself on computers when programs are downloaded and that tracks each user click, usually without the user's knowledge or permission.

Personal Internet Usage and Quality of Worklife

Pruthikrai Mahatanankoon
Illinois State University, USA

INTRODUCTION

The management of Internet usage in the workplace is becoming one of most important productivity concerns of the modern workplaces. Through proper use of Internet technology, organizations reap the benefits of efficient communication and information sharing among employees, managers, and organizations' customers. The Internet is an essential necessity, as global business expansion requires networked organizations that can understand and predict future market trends, analyze competitors' movements, and identify vital environmental factors that could be affecting their competitive position. For this reason the Internet has provided many added benefits for organizations, and modern workplaces have equipped their employees with access to the Internet and provided individual e-mail accounts.

Not until recently, Internet usage in the workplace has received a negative view from the business community. The issue of employees spending their work time on personal activities is not new to management. In some ways, spending time on personal telephone conversations, taking longer break times, or chatting with colleagues in the office is similar to personal Internet usage. Unfortunately, publicized cases of Internet abuse in the workplace, such as pornography, employee harassment, information leakage, and software piracy, have generated different and increasingly more serious ethical, legal, and productivity concerns for many organizations.

A study conducted in the year 2000 by NFO Interactive indicated that one out of three employees spends at least 2.5 hours of work time per week on non-work-related Internet surfing. Recent online industry studies suggest that such Internet usage in the workplace costs U.S. organizations approximately $1 billion annually in lost productivity (Foster, 2001). When Internet usage policy (IUP) or broad communication regarding proper Internet usage between employees and management does not work, stronger practical solutions are implemented to handle Internet abuse, such as reprimanding employees, deploying Internet monitoring and filtering software to restrict non-work-related Web site access, even restricting the hours of access. Although these strategies have shown improvement in employees' productivity and reduction in occurrences of Internet abuses, a thorough understanding of their impact on employees' job satisfaction and quality of work life still requires further investigation.

Evidence shows that Internet filtering and monitoring tools have led to employee dissatisfaction and other productivity debates. Urbaczewski and Jessup (2002) suggest that employee motivation generally plays a significant role in productivity, performance, and satisfaction in the workplace, but employee dissatisfaction increases with any type of organizational monitoring, including Internet monitoring. Management finds it difficult to maintain a workable balance between proper behavioral controls and employee empowerment. Because both stringent control and little control generally lead to abuse, management of Internet usage in the workplace requires developing a balance of human resource strategies with individual psychological considerations to maintain and improve employee satisfaction and well-being overall. To identify the appropriate balance, it is important to understand and examine the various Internet usage behaviors.

BACKGROUND OF INTERNET USAGES BEHAVIORS

The literature related to non-work-related Internet usage in the workplace can be classified into three categories: pathological, personal, and abusive. Influenced by various individual and social psychological aspects, each category deserves a research area of its own, involving different sets of determinants and consequences related to each behavioral category.

Pathological Internet Usage (PIU) involves excessive Internet usage from people who use the Internet as a means of coping with their personal problems or current personal difficulties (David, 2001). Psychology literature defines any extensive usage of the Internet where users have no control over their behaviors as "Internet addiction" (Greenfield, 1999; Young, 1998a). Internet addiction is currently being viewed as similar to substance abuse or a gambling addiction (Young, 1998b). PIU causes a specified number of psychological symptoms, such as altered moods, denial of responsibilities, guilt, and craving (Morahan-Martin & Schumacher, 2000). Brenner (1997) claims that Internet addicts have a higher tolerance level, withdrawal symptoms, and craving for the Internet as compared to normal Internet users. However, the addictive symptoms happen only to a minority of obsessive and/or compulsive users (Griffiths, 2000). In general workplace settings, occurrences of PIU or Internet addition are atypical; organizations can deal with problem employees specifically. Research finds that normal employees utilize the Internet for its usefulness, information content, ease of use, and communication (Stevens, Williams & Smith, 2000), whether the access is work-related or not.

Personal Web Usage in the Workplace (PWU) is defined as voluntary online Web behaviors during working time where employees use any of the organization's Internet resources for activities outside current customary job/work requirements (Anandarajan & Simmers, 2002). These activities include any leisure and/or personal use of the Internet at work; for example, using the Internet to escape from work by spending work time on the Internet searching for news and entertainment information, shopping online, booking personal vacations, and using a personal e-mail account to send messages to friends and family (Mahatanankoon, Anandarajan & Igbaria, 2004). In some cases, personal Web usage is also now referred to as cyberloafing (Lim, 2002), cyberslacking, or Internet abuse (Anandarajan, 2002; Young & Case, 2004).

Internet abuse is a general term often used mostly by practitioners; it refers to any wrongful or improper use of the Internet in the workplace. Behaviors related to Internet abuse are often more severe in nature, such as viewing pornography, harassing other employees, downloading illegal software, moonlighting, playing online games, online chatting, and so forth.

Although *Internet abuse* and *pathological Internet usage* lead to lost productivity, wasted or overused network bandwidth, and other possible legal liabilities, low to moderate non-work-related usage may in fact lead to positive consequences. There are studies suggesting that some *personal Web usage* in the workplace can be a healthy recipe for a productive work life. In fact, satisfied and productive employees frequently use the Internet (Stanton, 2002). Restricting online recreational activities may prevent employees from using the Internet to its full potential. Therefore, organizations should take actions both to empower and educate employees about the balance between work and play (Oravec, 2002). Others recommend organizations should not restrict personal Web browsing and take the necessary training steps to "facilitate the transfer of learning from the play domain to work-related tasks" (Belanger & Van Slyke, 2002, p. 65). Table 1 summarizes the possible benefits and potential drawbacks of *personal Web usage* at work.

ENHANCING QUALITY OF WORK LIFE

As evidence encourages low to moderate non-work-related behaviors in the workplace, managers and practitioners are left with new issues to maintain an effective workplace Internet management strategy. The solutions deal more or less with understanding the relationship between employees' well-being and employee job performance, and filtering out Web sites that could potentially lead to unproductive employees or legal liabilities.

Because personal Web usage in the workplace may influence the well-being of motivated employees, organizations must take precautions against the restriction of PWU. In some cases, employee training or education, together with effective communication between management and employees, can reduce the impact of Internet abuse. Too much PWU or too little PWU leads to Internet abuse (Anandarajan, 2002). When Internet monitoring is necessary to enforce appropriate social norms, an organizational Internet usage policy (IUP) should be based on maintaining employee job performance without lowering the positive quality of work life. Management can thus create a healthy workforce through use of an "*adaptive Internet monitoring and filtering policy* (AIMF)" (refer to Figure 1).

Adaptive strategy implies that in order to improve employee well-being, organizations should allocate time for PWU when employees have performed up to management expectations or their job requirements. The Adaptive Internet Monitoring and Filtering policy (AIMF) requires a reciprocal sense of respect and fulfillment of an organizational and employee psychological contract. Employees themselves need to have self-discipline and a strong work ethic. In some cases, changing employee behaviors requires organizational support and commitment that is established through Internet training, education, and peer influence. Organizations must also take employees' personal needs and job characteristics into consideration, including the organizational culture, the technological infrastructure, and individual employee roles and statuses. These factors dictate the amount of Internet usage activities performed by each employee. However, the purposes of using the Internet and its usage norms (whether at home or at work) are indeed "co-evolving" (Kraut & Kiesler, 2003), so it is difficult for organizations to examine the relationships between various non-work-related Internet activities, and any individual and organizatonal outcomes.

Management must decide on the tradeoffs between employee job satisfaction, job performance, and organizational liability. In other words, performing certain PWU

Table 1. Benefits and drawbacks of personal Web usage in the workplace

Benefits	Drawbacks
• Increase job satisfaction • Reduce stress • Increase job performance • Increase employee empowerment and learning • Improve social integration and collaboration	• Productivity loss, both for organizations and employees • Wasted network bandwidth • Possible legal liability • Information leakage • Overpaid employees; lost wages

activities can increase an employee's job satisfaction by amplifying their *intrinsic* demands and increase job performance by reducing stress; however, such behaviors can also lead to higher organizational risks and operating costs. One recommended strategy is to examine each PWU activity based on job satisfaction and job performance dimensions (Mahatanankoon & Igbaria, 2004). Figure 2 shows a 2x2 workplace Internet monitoring and filtering decision grid.

The grid has two dimensions: employee well-being and job performance. Employee well-being is related to job satisfaction, intrinsic motivation, and the quality of work life. Job performance involves job productivity, extrinsic motivation, and efficiency and effectiveness at work. Different PWU activities influence the two dimensions distinctly, suggesting possible actions that organizations can employ when deciding which PWU activities should be permitted and which activities should be restricted. Organizations should permit PWU activities that fall within Area 1 (high well-being/high performance) and strongly restrict activities in Area 4 (low well-being/low performance). Examples of PWU in Area 1 include learning about educational training classes related to work, visiting professional Web sites, reading current news/events to fulfill an employee's career or promotional goals, and searching for new business tools (Anandarajan, Devine & Simmers, 2004), including activities related to personal e-commerce (Mahatanankoon & Igbaria, 2004). However, PWU activities in Area 4 fall into Internet abuse categories, such as visiting pornography sites, downloading music or illegal software, online chatting, or playing games online. These activities should be strictly monitored and restricted, as they create lower employee work performance and job satisfaction, as well as increase organizational legal liabilities.

PWU activities in Area 2 (low well-being/high performance) and Area 3 (high well-being/low performance) fall into gray areas. The trade-offs between performance and job satisfaction should be based on management judgment (Mahatanankoon & Igbaria, 2004). In other words, an Internet monitoring/filtering policy should rely on general norms, cultural fits, and the objectives of each organization. The productive use of the Internet in the workplace falls within Area 2, including those uses commonly related to work, such as accessing organizational intranet/extranet sites; sending and sharing information with coworkers; visiting customers, suppliers, and/or competitors' Web sites, and so forth. To the contrary, Internet usages in Area 3 are generally related to leisure and recreational activities, which tend to reduce stress but do not have a direct impact on job performance - for example, using personal Web-based e-mails for personal communication with friends and family, surfing the Internet without any purpose, exploring personal hobbies and interests, reading general news and sport scores, and so forth. Taking the AIMF perspective, as activities

Figure 1. Adaptive Internet monitoring and filtering policy

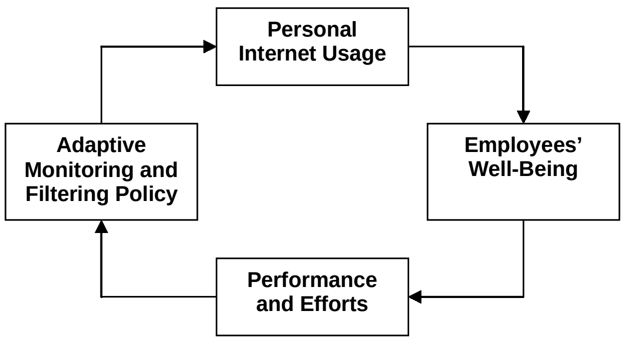
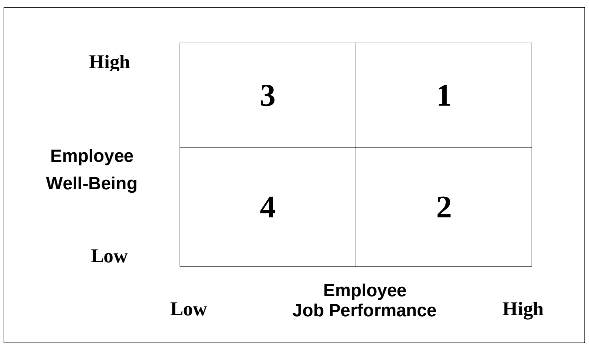


Figure 2. Workplace Internet monitoring and filtering decision grid



in Area 2 (low well-being/high performance) intensify, organizations should be ready to permit Area 1 (high well-being/high performance) and Area 3 (high well-being/low performance) activities to expand.

FUTURE RESEARCH

Monitoring strategy should align with organizational culture (Urbaczewski, 2004). Earlier research into the area of "junk computing" can be related to PWU. Junk computing refers to the use of organizational information technology that does not align with organizational goals (Guthrie & Gray, 1996). According to the definition, examples of junk computing are games, personal work, junk reporting, excessive unproductive computer work, and excessive e-mails. The new challenge for Internet-enabled workplaces is to find the right balance between work and play. Future studies need to examine:

- positive/negative relationships between various non-work-related Internet activities and job satisfaction or work performance;
- suitable attitudinal changes that enhance work/play ethics of employees;
- technological strategies to use to identify problem employees, and the proper tactics to reprimand valued employees;
- the moderating effects of cultural differences on non-work-related behaviors; and
- possible determinants that can lead to a different spectrum of Internet behaviors.

CONCLUSION

The Internet has been become one of the most technological necessary tools in today's workplace. The broad scope of its usefulness and ease of use make the Internet most beneficial to employee daily work activities as well as other personal activities. This article identifies the impact of personal Internet usage on employee job satisfaction and performance, and recommends several strategies that management can implement to increase employee well-being. Balancing strategies - such as a workplace Internet monitoring and filtering decision grid, and an adaptive Internet monitoring and filtering policy - will help researchers and practitioners to better understand the Internet usage patterns of employees and assist them in implementing better Internet usage policies that fit the workplace environment, employees' personal agendas, and management goals.

REFERENCES

Anandarjan, M., Devine, P. & Simmers, C.A. (2004). A multidimensional scaling approach to personal Web usage in the workplace. In M. Anandarajan & C.A. Simmers (Eds.), *Personal Web usage in the workplace: A guide to effective human resources management* (pp. 61-78). Hershey, PA: Idea Group Inc.

Anandarajan, M. (2002). Internet abuse in the workplace. *Communications of the ACM*, *45*(1), 53-54.

Anandarajan, M. & Simmers, C. (2002). Factors influencing Web access behavior in the workplace: A structural equation approach. In M. Anandarajan (Ed.), *Internet usage in the workplace: A social, ethical and legal perspective* (pp. 44-66). Hershey, PA: Idea Group Publishing.

Belanger, F. & Van Slyke, C. (2002). Abuse or learning? *Communications of the ACM*, *45*(1), 64-65.

Brenner, V. (1997). Psychology of computer use: XLVII. Parameters of Internet use, abuse and addiction: The first 90 days of the Internet Usage Survey. *Psychological Reports*, *80*(3), 879-882.

David, R.A. (2001). A cognitive-behavioral model of pathological Internet use. *Computers in Human Behavior*, *17*(2), 187-195.

Foster, M. (2001). Be alert to the signs of employees Internet addiction. *National Public Accountant*, *46*(9), 39-40.

Greenfield, D.N. (1999). Virtual addiction: Help for Netheads, cyberfreaks, and those who love them. Oakland, CA: New Harbinger Publications.

Griffiths, M. (2000). Does Internet and computer "addiction" exist? Some case study evidence. *CyberPsychology & Behavior*, *3*(2), 211-218.

Guthrie, R. & Gray, P. (1996). Junk computing: Is it bad for an organization? *Information Systems Management*, *13*, 23-28.

Kraut, R. & Kiesler, S. (2003). The social impact of Internet use. *Psychological Science Agenda*, *16*(3), 8-10.

Lim, K.G. (2002). The IT way of loafing on the job: Cyberloafing, neutralizing, and organizational justice. *Journal of Organizational Behavior*, *23*(5), 675-694.

Lin, J. C.-C. & Lu, H. (2000). Towards an understanding of the behavioral intention to use a Web site. *International Journal of Information Management*, *20*(3), 197-208.

Mahatanankoon, P., Anandarajan, M. & Igbaria, M. (2004). Development of a measure of personal Web usage in the workplace. *CyberPsychology & Behavior*, *7*(1), 93-104.

Mahatanankoon, P. & Igbaria, M. (2004). Impact of personal Internet usage on employee's well-being, In M. Anandarajan & C.A. Simmers (Eds.), *Personal Web usage in the workplace: A guide to effective human resources management* (pp. 246-263). Hershey, PA: Idea Group Inc.

Morahan-Martin, J. & Schumacher, P. (2000). Incidence and correlates of pathological Internet use among college students. *Computers in Human Behavior*, *16*(1), 13-29.

Oravec, J.A. (2002). Constructive approach to Internet recreation in the workplace. *Communications of the ACM*, *45*(1), 60-63.

Stanton, J.M. (2002). Company profile of the frequent Internet user. *Communications of the ACM*, *45*(1), 55-59.

Stevens, P.M., Williams, K.P. & Smith, M.C. (2000). Organizational communication and information processes in an Internet-enabled environment. *Psychology & Marketing*, *17*(7), 607-632.

Urbaczewski, A. & Jessup, L.M. (2002). Does electronic monitoring of employee Internet usage work? *Communications of the ACM*, *45*(1), 80-83.

Urbaczewski, A. (2004). Monitoring strategies for Internet technologies, In M. Anandarajan & C.A. Simmers (Eds.), *Personal Web usage in the workplace: A guide to effective human resources management* (pp. 141-157). Hershey, PA: Idea Group Inc.

Young, K. & Case, C.J. (2004). Internet abuse in the workplace: New trends in risk management. *CyberPsychology & Behavior*, *7*(1), 105-111.

Young, K. (1998a). Caught in the Net: How to recognize the signs of Internet addiction - and a winning strategy for recovery. New York: John Wiley & Sons.

Young, K.S. (1998b). Internet addiction: The emergence of a new clinical disorder. *CyberPsychology & Behavior*, *1*(3), 237-244.

KEY TERMS

Cyberloafing: Any voluntary act of employees using their organization's Internet access during office hours to surf non-work-related Web sites for non-work purposes, and access non-work-related e-mail.

Internet Abuse: Any wrongful or improper use of the Internet in the workplace.

Internet Filtering and Monitoring Software: Software tools used for reducing occurrences of Internet abuse by blocking inappropriate Web sites and identifying frequently visited Web sites.

Internet Usage Policy (IUP): An organizational policy handed down to employees that governs the use of the Internet in a specific workplace. The goals of an IUP, if properly written and implemented, are to help organizations communicate proper Internet usage behaviors, lessen employees' perceived expectation of privacy, and reduce costly litigation that may occur from the use of Internet monitoring and filtering software.

Junk Computing: The use of organizational information technology that does not align with organizational goals.

Pathological Internet Use (PIU): Excessive Internet usage from people who use the Internet as a means of coping with their personal problems or current personal difficulties.

Personal Web Usage in the Workplace (PWU): Voluntary online Web behaviors during work time where employees use any of the organizations' resources for activities outside current customary job/work requirements.

Quality of Work Life: Workplace hygiene factors that support the well-being and job satisfaction of employees.

Perturbations, Accuracy and Robustness in Neural Networks

Cesare Alippi
Politecnico Di Milano, Italy

Giovanni Vanini
Politecnico Di Milano, Italy

INTRODUCTION

A robustness analysis for neural networks, namely the evaluation of the effects induced by perturbations affecting the network weights, is a relevant theoretical aspect since weights characterise the "knowledge space" of the neural model and, hence, its inner nature.

In this direction, a study of the evolution of the network's weights over training time (training perturbations) allows the researcher for shedding light on the mechanism behind the generation of the knowledge space. Conversely, the analysis of a specific knowledge space (fixed configuration for weights) provides hints about the relationship between knowledge space and accuracy. This aspect is particularly relevant in recurrent neural networks, where even small modifications of the weight values are critical to performance (e.g., think of the stability of an intelligent controller comprising a neural network and issues, leading to robust control).

Robustness analysis must also be taken into account when implementing a neural network (or the intelligent computational system) in a physical device or in intelligent wireless sensor networks. Behavioral perturbations affecting the weights of a neural network abstract uncertainties such as finite precision representations, fluctuations of the parameters representing the weights in analog solutions (e.g., associated with the production process of a physical component), aging effects or more complex and subtle uncertainties in mixed implementations.

In this article, we suggest a robustness/sensitivity analysis in the large, that is, without assuming constraints on the size or nature of the perturbation; as such, the small perturbation hypothesis becomes only a sub-case of the theory. The suggested sensitivity/robustness analysis can be applied to all neural network models (including recurrent neural models) involved in system identification, control signal/image processing and automation-based applications without any restriction to study the relationship between perturbations affecting the knowledge space and the induced accuracy loss.

ROBUSTNESS ANALYSIS: THE STATE OF THE ART

The sensitivity/robustness issue has been widely addressed in the neural network community with a particular focus on specific neural topologies. In particular, when the neural network is composed of linear units, the relationship between perturbations and the induced performance loss can be obtained in a closed form (Alippi & Briozzo, 1998). Conversely, when the neural topology is non-linear we have either to assume the small perturbation hypothesis or particular assumptions about the stochastic nature of the neural computation, for example, see Alippi and Briozzo (1998), Pichè (1995), and Alippi (2002b); unfortunately, such hypotheses are not always satisfied in real applications. Another classic approach requires expanding the neural computation with Taylor around the nominal value of the trained weights. A subsequent linearized analysis follows which allows the researcher for solving the sensitivity issue problem (Pichè, 1995). This last approach has been widely used in the implementation design of neural networks where the small perturbation hypothesis abstracts small errors introduced by finite precision representations of the weights (Dundar & Rose, 1995; Holt & Hwang, 1993). Again, the validity of the analysis depends on the validity of the small perturbation hypothesis.

Differently, other authors avoid the small perturbation assumption by focusing the attention on very specific neural network topologies and/or by introducing particular assumptions regarding the distribution of perturbations, internal neural variables and inputs as done for Madalines neural networks (Stevenson, Winter, & Widrow, 1990; Alippi, Piuri, & Sami, 1995).

Some other authors tackle the robustness issue differently by suggesting techniques leading to neural networks with improved robustness ability by acting on the learning phase (e.g., see Alippi, 1999) or by introducing modular redundancy (Edwards & Murray, 1998); though, no robustness indexes are suggested there.

A ROBUSTNESS ANALYSIS IN THE LARGE

In the following, we consider a generic neural network implementing the $\hat{y}(x) = f(\hat{\theta}, x)$ function where $\hat{\theta}$ is the weight vector of the trained neural network.

In several neural models, and in particular in those related to system identification and control, the relationship between the inputs and the output of the system is captured by considering a regression vector φ, which contains a limited time-window of actual and past inputs, outputs, and, possibly, predicted outputs. Of particular interest are those models which can be represented by means of the model structures $\hat{y}(t) = f(\varphi)$ where function $f(\cdot)$ is a regression-type neural network, characterised by N_φ inputs, N_h non-linear hidden units and a single effective linear/non-linear output (Hassoun, 1995; Hertz, Krogh, & Palmer, 1991; Ljung, 1987; Ljung, Sjoberg, & Hjalmarsson, 1996).

The presence of a dynamic in the data can be modelled by a suitable number of delay elements which may affect inputs (time history on external inputs *u(t)*) system outputs (time history on $y(t)$) on predicted outputs (time history on $\hat{y}(t)$) or residuals (time history on $e(t) = \hat{y}(t) - y(t)$). Where it is needed, $y(t)$, $\hat{y}(t)$ and $e(t)$ are vectorial entities, a component for each independent distinct variable.

Several neural model structures have been suggested in the literature, which basically differ in the regression vector; examples are the NARMAX structures which can be obtained by considering both past inputs and outputs:

$$\varphi = [u(t), u(t-1), \cdots, u(t-n_u), y(t-1), \cdots, \\ y(t-n_y), ..., e(t-1), \cdots, e(t-n_e)]$$

and the NOE ones which process only the past inputs and

$$\varphi = [u(t), u(t-1), \cdots, u(t-n_u), \hat{y}(t-1), \cdots, \hat{y}(t-n_y)].$$

Static neural networks, such as classifiers, can be obtained by simply considering external inputs

$$\varphi = [u(t), u(t-1), \cdots, u(t-n_u)].$$

We denote by $\hat{y}_\Delta(x) = f_\Delta(\hat{\theta}, \Delta, x)$ the mathematical description of the perturbed computation and by $\Delta \in D \subseteq \Re^p$ a generic p-dimensional perturbation vector, a component for each independent perturbation affecting the network weights of model $\hat{y}(x)$. The perturbation space D is characterised in stochastic terms by providing the probability density function pdf_D.

To measure the discrepancy between $\hat{y}_\Delta(x)$ and $y(x)$ or $\hat{y}(x)$ we consider a generic loss function $U(\Delta)$. A common example for U is the mean square error (MSE) loss function

$$U(\Delta) = \frac{1}{N_x} \sum_{i=1}^{Nx} (y(x_i) - \hat{y}_\Delta(x_i))^2 \tag{1}$$

but a generic Lebesgue measurable loss function with respect to D can be taken into account (Jech, 1978). The formalization of the impact of perturbation on the performance function can be simply derived as:

Definition: Robustness Index

We say that a neural network is robust at level $\bar{\gamma}$ in D, when the robustness index $\bar{\gamma}$ is the minimum positive value for which

$$U(\Delta) \le \bar{\gamma}, \forall \Delta \in D. \tag{2}$$

Immediately, from the definition of robustness index, we have that a generic neural network NN1 is more robust than another NN2 iff $\bar{\gamma}_1 < \bar{\gamma}_2$; the property holds independently from the topology of the two neural networks.

The main problem related to the determination of the robustness index $\bar{\gamma}$ is that we have to compute $U(\Delta), \forall \Delta \in D$ if we wish a tight bound. The $\bar{\gamma}$-identification problem is, therefore, intractable from a computational point of view if we relax all assumptions made in the literature as we do. To deal with the computational aspect we associate a dual probabilistic problem to (2):

Robustness Index: Dual Problem

We say that a neural network is robust at level $\bar{\gamma}$ in D with confidence η when $\bar{\gamma}$ is the minimum positive value for which

$$\Pr(U(\Delta) \le \bar{\gamma}) \ge \eta \quad \text{holds} \quad \forall \Delta \in D. \tag{3}$$

The probabilistic problem is weaker than the deterministic one since it tolerates the existence of a set of perturbations (whose measure according to Lebesgue is 1-η) for which $u(\Delta) > \bar{\gamma}$. In other words, not more than 100η% of perturbations $\Delta \in D$ will generate a loss in performance larger than $\bar{\gamma}$. Probabilistic and deterministic problems

P

are "close" to each other when we choose, as we do, $\eta \to 1$.

The non-linearity with respect to Δ and the lack of a priori assumptions regarding the neural network do not allow computing (2) in a closed form for the general perturbation case. The analysis, which would imply testing $U(\Delta)$ in correspondence with a continuous perturbation space, can be solved by resorting to probability according to the dual problem and by applying randomized algorithms (Alippi, 2002a; Bai, Tempo, & Fu, 1997; Tempo & Dabbene, 1999; Vidyasagar, 1996; Vidyasagar, 1998) to solve the robustness/sensitivity problem.

RANDOMIZED ALGORITHMS AND PERTURBATION ANALYSIS

In the following, we denote by $p_\gamma = \{U(\Delta) \le \gamma\}$ the probability that the loss in performance associated with perturbations in D is below a given (but arbitrary) value γ.

Probability p_γ is unknown, it cannot be computed in a form for a generic U function and neural network topology, and its evaluation requires exploration of the whole perturbation space D. Anyway, the unknown probability p_γ can be estimated by sampling D with N independent and identically distributed samples Δ_i (intuitively a sufficiently large random sample explores the space); extraction must be carried out according to the pdf of the perturbation. For each sample Δ_i we then generate the triplet

$$\{\Delta_i, U(\Delta_i), I(\Delta_i)\}, i = 1, N \qquad \text{where}$$

$$I(\Delta_i) = \begin{cases} 1 \text{ if } U(\Delta_i) \le \gamma \\ 0 \text{ if } U(\Delta_i) > \gamma \end{cases}. \tag{4}$$

The true probability p_γ can now be simply estimated by means of the frequency as

$$\hat{p}_N = \frac{1}{N}\sum_{i=1}^{N} I(\Delta_i). \tag{5}$$

Of course, when N tends to infinity, $\hat{p}_N$ somehow converges to p_γ. By introducing an accuracy degree ε on the difference $|p_\gamma - \hat{p}_N|$ and a confidence level $1-\delta$ (which requests that the $|p_\gamma - \hat{p}_N| \le \varepsilon$ inequality is satisfied at least with probability $1-\delta$), our problem can be formalized by requiring that the inequality

$$\Pr\{|p_\gamma - \hat{p}_N| \le \varepsilon\} \ge 1-\delta \tag{6}$$

is satisfied $\forall \gamma \ge 0$. In other words (6) states that the true probability and its estimates must be very close (they may differ not more than ε) and that the statement must be true with high probability. Of course, we wish to control the accuracy and the confidence degrees of (6) by allowing the user to choose the most appropriate values for the particular need. Finally, by extracting a number of samples from D according to the Chernoff (1952) bound

$$N \ge \frac{\ln \frac{2}{\delta}}{2\varepsilon^2}, \tag{7}$$

we have that $\Pr\{|p_\gamma - \hat{p}_N| \le \varepsilon\} \ge 1-\delta$ holds for $\forall \gamma \ge 0, \forall \delta, \varepsilon \in [0,1]$

As an example, by considering a 5% in accuracy and 99% in confidence we have to extract 1060 samples from D; with such choice we can approximate p_γ with $\hat{p}_N$ introducing the maximum error 0.05 ($\hat{p}_N - 0.05 \le p_\gamma \le \hat{p}_N + 0.05$) and the inequality holds at least with probability 0.99.

The Chernoff bound grants that the dual probabilistic problem related to the identification of the robustness index $\bar{\gamma}$ can be solved with a polynomial complexity algorithm in the accuracy and the confidence degrees independently from the number of weights of the neural model network. In fact, from (6) if accuracy ε and confidence δ are small enough we can confuse p_γ and $\hat{p}_N$ by committing a small error. As a consequence, the dual probabilistic problem requiring $p_\gamma \ge \eta$ becomes $\hat{p}_N \ge \eta$. We surely assume ε and δ to be small enough in subsequent derivations.

The final algorithm, which allows for testing the robustness degree $\bar{\gamma}$ of a generic neural network, can be summed up as:

1. Select ε and δ sufficiently small to have enough accuracy and confidence;
2. Extract from D, according to its pdf, a number of perturbations N as suggested by (7);
3. Generate the indicator function $I(\Delta)$ and generate the estimate $\hat{p}_N = \hat{p}_N(\gamma)$ according to (5); and
4. Select the minimum value γ_η from the $\hat{p}_N = \hat{p}_N(\gamma)$ function so that $\hat{p}_N(\gamma_\eta) = 1$ is satisfied $\forall \gamma \ge \gamma_\eta$. γ_η is the estimate of the robustness index $\bar{\gamma}$.

Note that with a simple algorithm, we are able to estimate in polynomial time the robustness degree $\bar{\gamma}$ of a generic neural network. The accuracy in estimating $\bar{\gamma}$ can be made arbitrarily good by considering a larger number of samples as suggested by the Chernoff's bound.

SOME EXPERIMENTAL RESULTS

To shed light on how the methodology can be used to test the robustness of a neural network, we focus the attention on a regression-type experiment. In particular, we consider the simple error-free function

$$y = -x \cdot \sin(x^2) + \frac{e^{-0.23 \cdot x}}{1+x^4}, \quad x \in [-3,3]$$

and we approximate it with a regression-type neural network having 10 hidden units. Learning is then perfected with a Levenberg-Marquardt training algorithm (Hassoun, 1995).

We then wish to test the robustness of the trained neural network in the large, that is, by not assuming the small perturbation hypothesis. Perturbations affecting weights and biases of the neural network are defined in a perturbation space subject to a uniform distribution. In particular, a perturbation Δ_i affecting a generic weight w_i must be intended as a relative perturbation with respect to the weight magnitude according to the multiplicative perturbation model $w_{i,p} = w_i(1+\Delta_i)$, $\forall i = 1, n$ (e.g., see Edwards & Murray, 1998). As such, a t% perturbation implies that Δ_i is drawn from a symmetrical uniform distribution of extremes $\left[-\frac{t}{100}, \frac{t}{100}\right]$; a 5% perturbation affecting weights and biases composing vector $\hat{\theta}$ hence requires that each weight/bias is affected by an independent perturbation extracted from the $[-0.05, 0.05]$ interval and applied to the nominal value according to the multiplicative perturbation model. By applying the algorithm suggested in the previous section, we determine the $\hat{p}_\gamma = \hat{p}_\gamma(\gamma)$ functions corresponding to the 1%, 3%, 5%, 10%, 30% perturbations. Results are given in Figure 1 where we considered $\varepsilon = 0.02, \delta = 0.01$ leading to N=6624.

From its definition, $\bar{\gamma}$ is the smallest value for which $\hat{p}_\gamma = 1$, $\gamma \geq \bar{\gamma}$; in the figure, if we consider the 5% perturbation case, $\bar{\gamma}$ assumes a value around 7. We observe that by increasing the strength of perturbation (i.e., by enlarging the extremes of the uniform distribution characterizing the pdf of D) γ increases. In fact, stronger perturbations have a worse impact on the performance loss function since the error-affected neural network diverges from the error-free one. Conversely, we see that small perturbations, for example, the 1% one, induce a very small loss in performance since the robustness index γ_η is very small.

Another interesting use of the suggested methodology allows the neural network designer to test different neural models and identify, for instance, the most robust one.

To this end, we consider a set of performance-equivalent neural networks, each of which able to solve the application with a performance tolerable by the user. All neural networks are characterized by a different topological complexity (number of hidden units).

The $\hat{p}_\gamma = \hat{p}_\gamma(\gamma)$ curves parameterized in the number of hidden units are given in Figure 2 in the case of 1% perturbation. We see that by increasing the number of hidden units $\bar{\gamma}$ decreases. We immediately realize that neural networks with a reduced number of hidden units are, for this application, less robust than the ones possessing more degrees of freedom. For this application large networks provide, in a way, a sort of spatial redundancy: information characterizing the knowledge space of the neural networks is distributed over more degrees of freedom.

It should be clear, anyway, that robustness is a property of the identified neural model and strongly depends on the neural network complexity, the envisaged training algorithm, the training starting point and the training data samples. Hence, for a generic application, it is not possible to assert that by increasing the network complexity (i.e., by increasing the number of hidden units) we improve the robustness of the obtained model.

Figure 1: $\hat{p}_\gamma$ as a function of γ for the 10 hidden units neural network

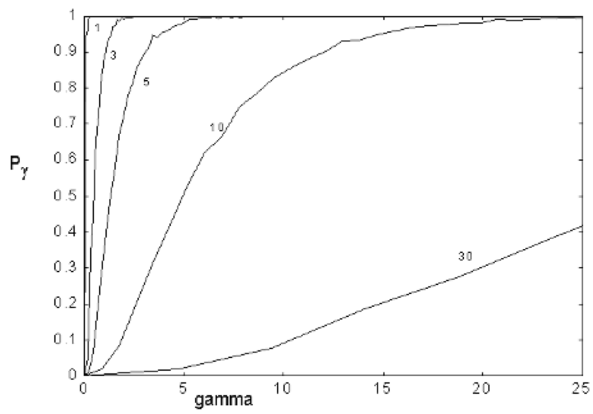
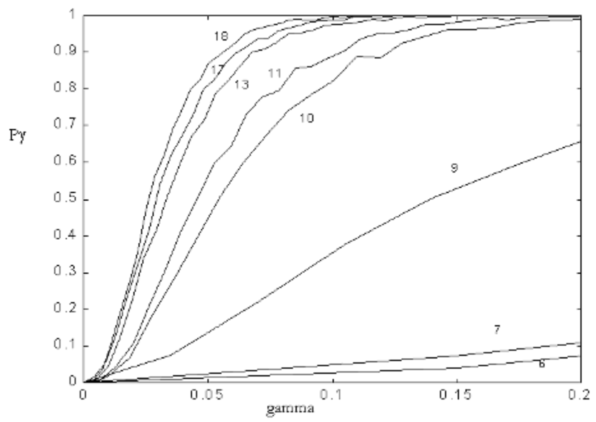


Figure 2: $\hat{p}_\gamma$ over γ and parameterized in the number of hidden unit



CONCLUSION

A robustness analysis in the large, i.e., by not assuming any hypothesis about the perturbation strength, can be addressed for a generic neural network model and accuracy loss figure of merit by resorting to randomization. In reality, the robustness analysis is even more general and can be considered for any Lebesgue-measurable computation (basically, all functions involved in signal/image processing are Lebesgue-measurable).

Hence, by considering trained neural networks, we can easily estimate the effects induced by perturbations affecting a generic neural network by considering a probabilistic approach. The robustness index, which can be used to investigate the relationships between knowledge space and neural network accuracy, can be computed with an effective poly-time algorithm which spouses Monte Carlo and learning theories sampling methods.

REFERENCES

Alippi, C. (1999). Feedforward Neural networks with improved insensitivity abilities. *ISCAS '99. Proceedings of the 1999 IEEE International Symposium on Circuits and Systems, 1999*, 5, 359-362.

Alippi, C. (2002a). Randomized algorithms: A system level, poly-time analysis of robust computation. *IEEE Transactions on Computers, 51*(7), 740-749.

Alippi, C. (2002b). Selecting accurate, robust and minimal feedforward neural networks. *IEEE Transactions on Circuits and Systems: Part I, Fundamental theory and applications, 49*(12), 1799-1810.

Alippi, C., & Briozzo, L. (1998). Accuracy vs. precision in digital VLSI architectures for signal processing. *IEEE Transactions on Computers, 47*(4), 472-477.

Alippi, C., Piuri, V., & Sami, M. (1995). Sensitivity to errors in artificial neural networks: A behavioural approach. *IEEE Transactions on Circuits and Systems: Part 1, 42*(6), 358-361.

Bai, E., Tempo, R., & Fu, M. (1997). Worst-case properties of the uniform distribution and randomized algorithms for robustness analysis. *Proceedings of. IEEE American Control Conference*, 11, 861-865.

Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, *23*, 493-509.

Dundar, G., & Rose, K. (1995). The effects of quantization on multilayer neural networks. *IEEE Transactions on Neural Networks, 6*(11), 1446-1451.

Edwards, P.J., & Murray, A.F. (1998). Towards optimally distributed computation. *Neural Computation, 10*(4), 987-1005.

Hassoun, M.H. (1995). *Fundamentals of artificial neural networks*. The MIT Press.

Hertz, J., Krogh, A., & Palmer, R.G. (1991). *Introduction to the theory of neural computation*. Addison-Wesley Publishing Company.

Holt, J., & Hwang, J. (1993). Finite precision error analysis of neural network hardware implementations. *IEEE Transactions on Computers, 42*(3), 281-290.

Jech, T. (1978). *Set theory*. Academic Press [Pure and Applied Mathematics].

Ljung, L. (1987). *System identification. Theory for the user*. Prentice-Hall.

Ljung, L., Sjoberg, J., & Hjalmarssoon, H. (1996). On neural networks model structures in system identification. In *Identification, adaptation, learning*. NATO ASI.

Piché, S. (1995). The selection of weights accuracies for Madalines. *IEEE Transactions on Neural Networks, 6*(2), 432-445.

Stevenson, M., Winter, R., & Widrow, B. (1990). Sensitivity of feedforward neural networks to weights errors. *IEEE Transactions on Neural Networks, 1*(1), 71-80.

Tempo, R., & Dabbene, F. (1999). Probabilistic robustness analysis and design of uncertain systems. *Progress in Systems and Control Theory*, *25*, 263–282.

Vidyasagar, M. (1996). *A theory of learning and generalisation with applications to neural networks and control systems*. Springer-Verlag, Berlin.

Vidyasagar, M. (1998). Statistical learning theory and randomized algorithms for control. *IEEE Control Systems Magazine*,*18*(6), 69-85.

KEY TERMS

Knowledge Space: The space defined by the neural network weights.

NARMAX: Nonlinear auto regressive moving average with external inputs model. The dynamic model receives past input, output and prediction error values to estimate the current output.

Neural Network Weights: The free parameters of the neural model.

NOE: Nonlinear output error model. The dynamic model receives past input and predicted output values to estimate the current output.

Perturbation: A behavioral entity affecting the weights of a neural network.

Poly-Time Algorithm: An algorithm whose complexity evolves polynomially with respect to the envisaged complexity parameter.

Regression-Type Neural Network: A neural network with one hidden layer and a unique linear output neuron used to approximate a mathematical function.

Robustness: A property possessed by a system. A system is robust with respect to a perturbation when the perturbation effect on the system performance is tolerable according to a predefined threshold.

Randomized Algorithms: A probabilistic sampling technique for exploring a space combining learning theories and Monte Carlo approaches.

Small Perturbation Hypothesis: The strength of the envisaged perturbations is small enough to guarantee effectiveness for a linearized sensitivity analysis.

Wireless Sensor Network: A network of distributed sensors linked by a wireless connection.

Policy Frameworks for Secure Electronic Business

Andreas Mitrakas
Ubizen, Belgium

INTRODUCTION

Terms conveyed by means of policy in electronic business have become a common way to express permissions and limitations in online transactions. Doctrine and standards have contributed to determining policy frameworks and making them mandatory in certain areas such as electronic signatures. A typical example of limitations conveyed through policy in electronic signatures includes certificate policies that Certification Authorities (CAs) typically make available to subscribers and relying parties. Trade partners might also use policies to convey limitations to the way electronic signatures are accepted within specific business frameworks. Examples of transaction constraints might include limitations in roles undertaken to carry out an action in a given context, which can be introduced by means of attribute certificates. Relying parties might also use signature policies to denote the conditions for the validation and verification of electronic signatures they accept. Furthermore, signature policies might contain additional transaction-specific limitations in validating an electronic signature addressed to end users. Large-scale transactions that involve the processing of electronic signatures in a mass scale within diverse applications rely on policies to convey signature-related information and limitations in a transaction. As legally binding statements, policies are used to convey *trust* in electronic business. Extending further the use of policy in transaction environments can enhance security, legal safety, and transparency in a transaction. Additional improvements are required, however, in order to render applicable terms that are conveyed through policy and enforce them unambiguously in a transaction. The remainder of this article discusses common concepts of policies and certain applications thereof.

BACKGROUND

An early example of a transaction framework is open EDI (Electronic Data Interchange) that aims at using openly available structured data formats and is delivered over open networks. While the main goal of open EDI has been to enable short-term or *ad hoc* commercial transactions among organisations (Kalakota & Whinson, 1996), it has also aimed at lowering the entry barriers of establishing structured data links between trading partners by minimising the need for bilateral framework agreements, known as interchange agreements. One specific requirement of open EDI is to set up the operational and contract framework within which a transaction is carried out. Automating the process of negotiating and executing agreements regarding the legal and technical conditions for open EDI can significantly lower the entry barriers, especially for non-recurrent transactions (Mitrakas, 2000).

Building on the model for open EDI, the Business Collaboration Framework is a set of specifications and guides, the centre of which is the UN/CEFACT; it aims at further lowering the entry barriers of electronic commerce based on structured data formats. The need for flexibility and versatility to loosely coupled applications and communication on the Internet has led to the emergence of Web services. A Web service is a collection of protocols and standards that are used to exchange data between applications. While applications can be written in various languages and run on various platforms, they can use Web services to exchange data over the Internet.

In Web services, using open standards ensures interoperability. These standards also include formal descriptions of models of business procedures to specify classes of business transactions that all serve the same goal. A trade procedure stipulates the actions, the parties, the order, and the timing constraints on performing actions (Lee, 1996). In complex business situations, transaction scenarios typically might belong to a different trade partner that each one owns a piece of that scenario. Associating a scenario with a trade partner often requires electronic signatures. When a trade partner signs with an electronic signature, she might validate or approve of the way that individual procedural components might operate within a transaction. The signatory of an electronic document or a transaction procedure depends on the performance of complex and often opaque-to-the-end-user systems.

Trust in the transaction procedures and the provision of services is a requirement that ensures that the signatory eventually adheres to transparent contract terms that cannot be repudiated (Mitrakas, 2003). Policy is seen as a way to formalise a transaction by highlighting those aspects of a transaction that are essential to the end user (Mitrakas, 2004). The immediate effect of using policies to convey limitations is that the party that relies on a signed transaction adheres to the limitations of that policy. Policy is, therefore, used to convey limitations to a large number of users in a way that makes a transaction enforceable. While these limitations are mostly meaningful at the operational or technical level of the transaction, they often have a binding legal effect and are used to convey contractual terms. Although these terms are not necessarily legal by nature, they are likely to have a binding effect. Sometimes they can be more far reaching by constraining relying parties that validate electronic signatures. Limitations might be mandated by law or merely by agreement, as in the case of limitations of qualified signatures according to European Directive 1999/93/EC on a common framework for electronic signatures (ETSI TS 101 456).

POLICY CONSTRAINTS IN ELECTRONIC BUSINESS

Electronic signatures have been seen as a lynchpin of trust in electronic transactions. The subject matter of current electronic signature regulation addresses the requirements on the legal recognition of electronic signatures used for non-repudiation and authentication (Adams & Lloyd, 1999). Non-repudiation is addressed in both technical standards such as X.509 and legislation. Non-repudiation addresses the requirement for electronic signing in a transaction in such a way that an uncontested link to the declaration of will of the signatory is established. Non-repudiation is the attribute of a communication that protects against a successful dispute of its origin, submission, delivery, or content (Ford & Baum, 2001). From a business perspective non-repudiation can be seen as a service that provides a high level of assurance on information being genuine and non-refutable (Pfleeger, 2000). From a legal perspective non-repudiation, in the meaning of the Directive 1999/93/EC on a common framework on electronic signatures, has been coined by the term, *qualified signature*, which is often used to describe an electronic signature that uses a secure signature creation device and is supported by a qualified certificate. A qualified signature is defined in the annexes of the directive and is granted the same legal effect as hand-written signatures where law requires them in the transactions.

Policies aim at invoking trust in transactions to ensure transparency and a spread of risk among the transacting parties. Policies are unilateral declarations of will that complement transaction frameworks based on private law. Policies can be seen as guidelines that relate to the technical organizational and legal aspects of a transaction, and they are rendered enforceable by means of an agreement that binds the transacting parties.

In Public Key Infrastructure (PKI), a CA typically uses policy in the form of a certification practice statement (CPS) to convey legally binding limitations to certificate users, being subscribers and relying parties. A CPS is a statement of the practices that a CA employs in issuing certificates (ABA, 1996). A CPS is a comprehensive treatment of how the CA makes its services available and delimiting the domain of providing electronic signature services to subscribers and relying parties. A certificate policy (CP) is sometimes used with a CPS to address the certification objectives of the CA implementation. While the CPS is typically seen as answering "how" security objectives are met, the CP is the document that sets these objectives (ABA, 2001). A CP and a CPS are used to convey information needed to subscribers and parties relying on electronic signatures, in order to assess the level of trustworthiness of a certificate that supports an electronic signature. By providing detailed information on security and procedures required in managing the life cycle of a certificate, policies become of paramount importance in transactions. Sometimes, a PKI Disclosure Statement (PDS) distils certain important policy aspects and services the purpose of notice and conspicuousness of communicating applicable terms (ABA, 2001). The Internet Engineering Task Force (IETF) has specified a model framework for certificate policies (RFC 3647).

Assessing the validity of electronic signatures is yet another requirement of the end user, most importantly, the relying parties. A signature policy describes the scope and usage of such electronic signature with a view to address the operational conditions of a given transaction context (ETSI TR 102 041). A signature policy is a set of rules under which an electronic signature can be created and determined to be valid (ETSI TS 101 733). A signature policy determines the validation conditions of an electronic signature within a given context. A context may include a business transaction, a legal regime, a role assumed by the signing party, and so forth. In a broader perspective, a signature policy can be seen as a means to invoke trust and convey information in electronic commerce by defining appropriately indicated trust conditions.

In signature policies it is also desirable to include additional elements of information associated with certain aspects of general terms and conditions to relate with the scope of the performed action as it applies in the transaction at hand (Mitrakas, 2004). A signature policy might, therefore, include content that relates it to the general conditions prevailing in a transaction, the discreet elements of a transaction procedure as provided by the various parties involved in building a transaction, as well as the prevailing certificate policy (ETSI TS 102 041).

Trade parties might use transaction constraints to designate roles or other attributes undertaken to carry out an action within a transaction framework. Attribute certificates are used to convey such role constraints and are used to indicate a role, a function, or a transaction type constraint. Attribute policies are used to convey limitations associated with the use and life cycle of such attributes (ETSI TS 101 058).

Processing signed electronic invoices is an application area of using policies. By means of a signature policy, the recipient of an invoice might mandate a specific signature format and associated validation rules. The sender of the invoice might require that signing an invoice might only be carried out under a certain role; therefore, an attribute certificate issued under a specific attribute policy might be mandated. This attribute policy complements the certification practice statement that the issuer of electronic certificates makes available. It is expected that certificate policies shall influence the requirements to make a signature policy binding (Mitrakas, 2003).

BINDING POLICIES IN ELECTRONIC BUSINESS

Communicating and rendering policies binding has been an issue of significant importance in electronic transactions. Inherent limitations in the space available for digital certificates dictate that policies are often conveyed and used in a transaction by incorporating them by reference (Wu, 1998). Incorporation by reference is to make one message part of another message by identifying the message to be incorporated, providing information that enables the receiving party to access and obtain the incorporated message in its entirety, expressing the intention that it be part of the other message (ABA, 1996). The incorporation of policies for electronic signatures into the agreement between signatory and recipient can take place by referencing the intent to use such policy in transactions. When the recipient accepts the signed document of the signatory, he implicitly agrees on the conditions of the underlying signature policy. In practice, incorporating policy into the agreement between signatory and recipient can also be effected by:

- Referring to a policy in a parties' agreement that explicitly refers to such policy.
- Accepting a signed document and implicitly agreeing on the conditions of the underlying policy, although this option might be more restrictive in case of a dispute.

An issue arises with regard to how and under which conditions a particular policy framework can be incorporated into an agreement of a signatory in a way that binds a relying party, regardless of its capacity to act as consumer or business partner. Incorporation of contract terms into consumer contracts and incorporation of contract terms into business contracts follow different rules. Incorporation by reference in a business contract is comparatively straightforward, whereas in a consumer contract stricter rules have to be followed as mandated by consumer protection regulations. Limitations to the enforceability of legal terms that are conveyed by means of policy are applied as a result of consumer protection legislation. In Europe, consumer protection legislation includes the Council Directive 93/13/EC on unfair terms in consumer contracts, Directive 97/7/EC on the protection consumers in distance transactions, and Directive 1999/44/EEC on certain aspects of the sale of consumer goods and associated guarantees (Hoernle, Sutter & Walden, 2002). In an effort to proactively implement these legal requirements, service providers strive to set up specific consumer protection frameworks (GlobalSign, 2004).

Sometimes the scope of the underlying certificate policy frameworks is to equip the transacting parties with the ability to use a certificate as evidence in a court of law. It is necessary to also provide transacting parties with assurance that allows a certificate to be admitted in legal proceedings and that it provides binding evidence against the parties involved in it, including the CA, the subscriber, and relying parties (Reed, 2000). Qualified electronic signatures in the meaning of Directive 1999/93/EC establish a rebuttable presumption that reverses the burden of proof. In other words the court may at first admit a signature that claims to be qualified as an equivalent of a handwritten signature. The counter-party is allowed to prove that such signature does not meet the requirements for qualified signatures, and could therefore be insecure for signing documents requiring a handwritten signature (UNCITRAL, 2000). To further answer the question of admissibility, it is necessary to examine the admissibility of electronic data as evidence in court, which is a matter that has been addressed in Directive 2000/31/EC on electronic commerce. Consequently, electronic data can be admitted as evidence as long as certain warranties are provided with regard to the production and retention of such data. In assessing the reliability of a certificate, a Court will have to examine the possibility of a certificate being the product of erroneous or fraudulent issuance, and if is not, the Court should proclaim it as sufficient evidence against the parties involved within the boundaries of conveyed and binding policy.

FUTURE TRENDS

While case law is expected to determine and enhance the conditions of admissibility and evidential value of policy in transactions based on electronic signatures, additional technological features such as the use of object identifiers (OIDs) and hashing are expected to further enhance the certainty required to accept policies. Remarkably, to date there has been little done to address in a common manner the practical aspects of identifying individual policies and distinguishing among the versions thereof. Additionally, mapping and reconciling policy frameworks in overlapping transactions also threaten transactions, which are based on the use and acceptance of varying terms. A typical hard case might involve for example overlapping policy conditions, which apply to certificates issued by different CAs. The situation is exacerbated if those CAs do not have the means to recognise one another, while they issue certificates that can be used in the same transaction frameworks (ETSI TS 102 231). Although such certificates may well be complementary to a transaction framework, the varying assurance levels they provide might threaten the reliability of the transaction framework as a whole. The immediate risk for the transacting parties can be an unwarranted transaction environment that threatens to render otherwise legitimate practices unenforceable. Reconciling the methods used across various electronic signing environments is likely to contribute to creating trust in electronic business.

An additional area of future attention may address policy frameworks related to the application layer in a transaction. As present-day requirements for transparency are likely to be further raised, it is expected that online applications will increasingly become more demanding in explaining to the end user what they do and actually warranting the performance. To date general conditions and subscriber agreements cover part of this requirement; however, it is further needed to provide a comprehensive description of the technical features and functionality of the online application. In electronic business, consumers and trade partners are likely to benefit from it. Policies for the application layer are likely to become more in demand in electronic government applications, where the requirement for transparency in the transaction is even higher than in electronic business. Finally, specifying policies further to meet the needs of particular groups of organisations is an additional expectation. Again in electronic government it is expected that interoperability will be enhanced through best practices and standards regarding policy in specific vertical areas.

CONCLUSION

While policies emerge as a necessary piece in the puzzle of invoking trust and legal safety in electronic transactions, policy frameworks can still have repercussions that reach well beyond the scope of single transaction elements and procedures in isolated electronic business environments. Formal policy frameworks require additional attention to ensure that apparently loosely linked policy elements do not threaten to overturn the requirements of transaction security and legal safety, which are the original objectives of using policy frameworks. Electronic transaction frameworks for diverse application areas can benefit from the processing of data on the basis of policy-invoked constraints among the parties involved. Large-scale processing that requires policy to convey operational and legal conditions in electronic transactions benefits from a combination of policy instruments, including certificate polices, signature policies, attribute certificate policies, and so forth, to enhance the outlining of the transaction framework and allow the transacting parties to further rely on electronic business for carrying out binding transactions.

NOTE

The views expressed in this article are solely the views of the author.

REFERENCES

American Bar Association. (1996). *Digital signature guidelines.* Washington, DC.

American Bar Association. (2001). *PKI assessment guidelines.* Washington, DC.

Adams, C. & Lloyd, S. (1999). *Understanding public key infrastructure.* Macmillan Technical Publishing, London.

ETSI TS 101 733. (2000). *Electronic signature formats.* Sophia-Antipolis.

ETSI TS 101 456. (2001). *Policy requirements for CAs issuing qualified certificates.* Sophia-Antipolis

ETSI 102 041. (2002). *Signature policy report, ETSI.* Sophia-Antipolis.

ETSI TS 101 058. (2003). *Policy requirements for attribute authorities.* Sophia-Antipolis.

ETSI TS 102 231. (2003). Provision of harmonized trust service provider status information. Sophia-Antipolis.

Ford, W. & Baum, M.(2001). *Secure electronic commerce* (2nd edition). Englewood Cliffs, NJ: Prentice-Hall.

GlobalSign. (2004). *Certification practice statement.* Retrieved from.*http://www.globalsign.net/repository*

Hoernle, J., Sutter, G. & Walden, I. (2002). Directive 97/7/EC on the protection of consumers in respect of distance contracts. In A. Lodder & H.W.K. Kaspersen (Eds.), *eDirectives: Guide to European Union Law on e-commerce*. Kluwer Law International, The Hague.

IETF RFC 3647. (2003). *Internet X.509 public key infrastructure—certificate policies and certification practices framework.* Retrieved from http://www.faqs.org/rfcs/rfc3647.html

ITU-T Recommendation X.509, ISO/IEC 9594-8. *Information technology—open systems interconnection—the directory: Public-key and attribute certificate frameworks.* Draft revised recommendation. Retrieved from *http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail ?CSNUMBER =34551 &ICS1=35*

Kalakota, R. & Whinson A. (1996). *Frontiers of electronic commerce*. Boston: Addison-Wesley.

Lee, R. (1996). *InterProcs: Modelling environment for automated trade procedures*. User documentation, EURIDIS, WP 96.10.11, Erasmus University, Rotterdam.

Mitrakas, A. (2003). Policy constraints and role attributes in electronic invoices. *Information Security Bulletin, 8*(5).

Mitrakas, A. (2004). Policy-driven signing frameworks in open electronic transactions. In G. Doukidis, N. Mylonopoulos & N. Pouloudi (Eds.), *Information society or information economy? A combined perspective on the digital era*. Hershey, PA: Idea Group Publishing.

Mitrakas, A. (2000). Electronic contracting for open EDI. In S.M. Rahman & M. Raisinghani (Eds.), *Electronic commerce: Opportunities and challenges.* Hershey, PA: Idea Group Publishing.

Pfleeger, C. (2000). *Security in computing.* Englewood Cliffs, NJ: Prentice-Hall.

Reed, C. (2000). *Internet law: Text and materials.* Butterworths.

United Nations. (2000). *Guide to enactment of the UNCITRAL uniform rules on electronic signatures.* New York.

Wu, S. (1998). Incorporation by reference and public key infrastructure: Moving the law beyond the paper-based world. *Jurimetrics, 38*(3).

KEY TERMS

Certification Authority: An authority such as GlobalSign that issues, suspends, or revokes a digital certificate.

Certification Practice Statement: A statement of the practices of a certificate authority and the conditions of issuance, suspension, revocation, and so forth of a certificate.

Electronic Data Interchange (EDI): The interchange of data message structured under a certain format between business applications.

Incorporation by Reference: To make one document a part of another by identifying the document to be incorporated, with information that allows the recipient to access and obtain the incorporated message in its entirety, and by expressing the intention that it be part of the incorporating message. Such an incorporated message shall have the same effect as if it had been fully stated in the message.

Public Key Infrastructure (PKI): The architecture, organization, techniques, practices, and procedures that collectively support the implementation and operation of a certificate-based public key cryptographic system.

Relying Party: A recipient who acts by relying on a certificate and an electronic signature.

Signature Policy: A set of rules for the creation and validation of an electronic signature, under which the signature can be determined to be valid.

Portable Portals for M-Commerce

Irvine Clarke III
James Madison University, USA

Theresa Flaherty
James Madison University, USA

INTRODUCTION

In the new decade, the call for information technology will be information, anytime, anyplace, and on any device. Accordingly, e-commerce is poised to witness an unprecedented explosion of mobility, creating a new domain of mobile commerce. Mobile commerce, or m-commerce, is the ability to purchase goods anywhere through a wireless Internet-enabled device (e.g., cellular phone, pager, PDA, etc.). Mobile commerce refers to any transaction with monetary value that is conducted via a mobile network. It will allow users to purchase products over the Internet without the use of a PC. Non-existent a year ago, m-commerce is now the buzzword of the marketing industry (King, 2000).

Over the past few years, e-commerce has become increasingly reliant upon portals to attract and retain users. Portals are the preferred starting point for searches that provide the user easily customizable architecture for finding relevant information. Portals provide the valuable gateways for getting users to their desired destinations. Today, about 15% of all Web page view traffic goes through the top nine portals, making them some of the most valuable land on the Web (Monohan, 1999). This heavy traffic flow gives the Web-based portal a unique position in corporate e-commerce strategy—with even greater potential influence for mobile applications. For mobile devices, these portals take on increased significance, as consumers are unwilling to spend long periods "surfing" on these inherently less user-friendly wireless devices. By the year 2006, 25 million people are expected to be dedicated wireless portal users (Carroll, 2000).

As m-commerce success will likely depend upon maintaining consumer utilization of these gateways, the companies that leverage the unique characteristics of wireless devices will gain exploitable advantages in the mobile marketplace. Due to current technological limitations and varying mobile consumer behavior patterns, portals developed for mobile devices must emphasize differing characteristics than conventional Web-based portals. As such, many portals may be unsuited for application in the mobile world.

"Traditional portals are not providing information that is specific enough for the user of a mobile portal. They are not able to incorporate location-specific information nor do they have the data and knowledge of each customer that the mobile operator has." (Durlacher Research, 2000, p. 65)

BACKGROUND

Mobile devices have been the fastest adopted consumer products of all time with, last year, more mobile phones shipped than automobiles and PCs combined (Chen, 2000). By 2003, there will be 1.4 billion mobile phones worldwide, and half will be Internet-enabled (Zabala, 2000). "The wireless world is a parallel universe almost as large as the Net, and the two are beginning a fascinating convergence," said Swapnil Shah, director of Inktomi Europe, a search engine and caching solutions company (Rao, 2000, p. 1). It is predicted that this emergence of mobile commerce will happen even faster than the development of e-commerce—in roughly the time between the invention of the first Web browser and now (Schenker, 2000). "If you look five to 10 years out, almost all of e-commerce will be on wireless devices," says Jeff Bezos, chief executive and founder of Amazon.com (McGinity, 2000, p. 1).

The potential of m-commerce is considerable for those willing to develop mobile-specific business models. However, as m-commerce matures, current mobile operators will rely less upon usage fees and increasingly derive revenues from content and services. Additionally, m-commerce is going to bring about a massive change in the way users consume products and services. As Cindy Dahm, European director for Phone.com, stated:

"It is key that commerce companies recognize m-commerce as a completely unique service. Cell phone users are more impatient than Internet users. The paradigm here is not surfing; all services for the mass market have to be pitched at users in such a seamless way that they need not even be aware that they are accessing the Net." (Rao, 2000, p. 2)

Those best able to provide value-added user experiences, through content aggregation and portal development, will achieve long-term success. Merely extending the current Internet presence will not be enough. "Mobile Internet customers will be more demanding. They will want personalized service to meet their individual wants and needs, and will no longer be satisfied with being a mass market" (KPMG, 2000, p. 2). Providers must take advantage of the characteristics which distinguish m-commerce from e-commerce to develop truly unique and compelling services rather than replicating current e-commerce models.

What is a Portable Portal?

The word portal is derived from the Latin *porta*, or gate, through which something will pass, in an effort to get to another place. In the traditional sense of the word, the portal is not the desired end-state. Rather, a portal is a necessary or convenient place one must go to get to the desired location. For example, the airport is not the desired location for most people, rather a necessary portal through which they must pass to obtain transportation to another location. Similarly, portals assist by directing the transport of the Web user to the ultimate location of their choice. Mobile portals, sometimes referred to as "portable portals," are typically developed to assist wireless users in their interactions with Web-based materials. Today, most mobile portals are being formed by syndicating content providers into a centralized source of personal productivity information. Mobile portals are often modeled by aggregating applications (e-mail, calendar, instant messaging, etc.) and content from various providers in order to become the user's prime supplier for Web-based information. Mobile portals differ from traditional Web-based portals by a greater degree of personalization and localization. Representative objectives for such mobile portals may be to attract the desired viewer, and build valuable customer relationships, through value-added content and community services, to augment the overall wireless Internet experience and build long-term customer loyalty.

"Like portals of the wired world, wireless portals have a degree of control over what users see on the Internet, so the portal provider can charge service providers and advertisers high fees. Given the projected penetration rates of mobile devices and mobile consumers' increased reliance on portal services, many observers expect wireless portals to be as highly valued on the stock market as their established Internet equivalents." (Barnett, Hodges & Wilshire, 2000, p. 166)

Established portal players, such as AOL (www.aol.com/product/anywhere.adp), Yahoo! (mobile.yahoo.com), and MSN (mobile.msn.com) have recognized the potential impact of the mobile Internet and have created mobile portals targeting U.S. subscribers. However, many of the traditional portal players are experiencing difficulties adapting to the mobile world. The mobile portals emerging today are, in many ways, a stripped-down version of traditional Web portals, without an understanding of the special requirements of mobile users. Consequently, these offering are unacceptable to mobile Internet users.

"The mobile portal strategy of the traditional portal players often lacks in-depth understanding of national mobile markets and of the specific local dynamics involved in building businesses in a territory. In addition, the differences between a fixed and more mobile portal model are non-trivial and, as such, lessons learned in the past are not necessarily directly transferable." (Durlacher Research Ltd., 2000, p. 69)

Current portal strategy is based on a traditional paradigm of consumers as passive receivers of communication efforts, with the portal provider holding control of the "when" and "where" of information. With wireless Internet-enabled devices, consumers now have more discretion of "when" and "where" the information that is available, creating a demand for a specialized portal offering.

Unique Advantages of Mobile Portals

The mobility afforded wireless devices will shape mobile portals into a disparate entity from conventional portals. The advantages of mobile devices provide a greater offering of *value-for-time* to users. That is, by accessing the Internet through mobile devices, users will be able to realize additional value allowances for any specified period of time that fixed-line users will not be able to achieve. Information will now truly become available anytime, anyplace, and on any wireless device. Portable portals differ from traditional portals on four dimensions: ubiquity, convenience, localization, and personalization.

(1) **Ubiquity:** Mobile devices offer users the ability to receive information and perform transactions from virtually any location on a real-time basis. Thus, mobile portal users can have a presence everywhere, or in many places simultaneously, with a similar level of access available through fixed-line technology. Communication can take place independent of the user's location. Mobile portals, for example, can leverage this advantage of ubiquity by

providing alert notifications, such as for auctions, betting, and stock price changes, which are specified by the user as an important part of relevant personal content. As such, the real-time, everywhere presence of mobile portals will offer capabilities uniquely beneficial to users. Industries that are time and location sensitive, such as financial services, travel, and retail, are likely to benefit from portals exploiting this value-added feature of m-commerce.

(2) **Convenience:** The agility and accessibility provided from wireless devices will further allow m-commerce to differentiate its abilities from e-commerce. People will no longer be constrained by time or place in accessing e-commerce activities. Rather, m-commerce could be accessed in a manner which may eliminate some of the labor of life's activities and thus become more convenient. For example, consumers waiting in line or stuck in traffic will be able to pursue favorite Internet-based activities or handle daily transactions through m-commerce portals. Consumers may recognize a higher level of convenience which could translate into an improved quality of life.

(3) **Localization:** Knowing the geographical location of an Internet user creates a significant advantage for m-commerce over e-commerce. Location-based marketing, via global positioning technology (GPS), will soon be available in all mobile devices. Through GPS technology, service providers can accurately identify the location of the user so that m-commerce providers will be better able to receive and send information relative to a specific location. Since mobile devices, such as cell phones, are almost always on, vendors will know the location of their customers and can deliver promotions based on the likely consumer demands for that location.

Location-specific information leverages a key advantage a mobile portal has over a traditional Web portal by supplying information relevant to the current geographic position of the user. Portable portal providers will be able to both push and access information relevant to the user's specific location. Portable portals may serve as points of consolidation of consumer information and disseminate the relevant information for a particular location. This can be based on profile data built on the user's past behavior, situation, profile, and location. As such, real-time discounting may become the "killer application" for m-commerce.

(4) **Personalization:** Mobile devices are typically used by a sole individual, making them ideal for individual-based and highly personalized target marketing efforts. Mobile devices offer the opportunity to personalize messages to various segments, based upon time and location, by altering both sight and sound. New developments in information technology and data mining make tailoring messages to individual consumers practical and cost effective. Value-added becomes maximized for those portals that are best able to implement the four distinguishing capabilities of ubiquity, convenience, localization, and personalization. Mobile portals will become differentiated based upon their abilities to integrate and actuate the four advantages which are germane to mobile devices.

CONCLUSION

Mobile portal development is likely to parallel the growth in overall m-commerce. New and even more innovative applications than those discussed here will arise as more people connect to the Web through wireless devices and additional content becomes available. As other functionalities such as file attachments, faster network speeds, Bluetooth, speech recognition, and e-commerce security features are added, more users will be attracted to this intriguing mobile marketplace. M-commerce is still not without its limitations. The problems it must overcome include: uniform standards, ease of operation, security for transactions, minimum screen size, display type and bandwidth, billing services, and the relatively impoverished Web sites. In the short term, portable portals must be mindful to operate within evolving technological constraints, while providing the most amount of flexibility possible. In the end, these problems, much like the initial problems associated with e-commerce, will also be overcome. Long-term mobile portal success is likely to come from consumer-driven, rather than technology-based models.

REFERENCES

Barnett, N., Hodges, S. & Wilshire, M. (2000). M-commerce: An operator's manual. *McKinsey Quarterly, 3*, 162-171.

Carroll, K. (2000). Portable portals. *Telephony, 238*(10), 10-11.

Chen, P. (2000). Broadvision delivers new frontier for e-commerce. *M-commerce*, (October 25).

Durlacher Research Ltd. (2000). Mobile commerce report. Retrieved from *www.durlacher.com*.

King, G. (2000). The m-marketing model dilemma: Pushmeorpullyou?. *Newsbeat ChannelSeven.com*, (December 7). Retrieved from *www.channelseven.com/newsbeat/2000features/news20001207.shtml*.

KPMG. (2000). Creating the new wireless operator. Retrieved from *www.kpmg.com*.

McGinity, M. (2000). The Net/wireless meeting of the minds. *Inter@ctive Week*, (March 6), 2.

Monohan, J. (1999). Portal puzzle. *Banking Strategies*, *75*(6), 148-158.

Rao, M. (2000). EM-wire: E-commerce, m-commerce poised for rapid take-off in Europe. *Electronic markets: The International Journal of Electronic Commerce & Business Media*, (April 6). Retrieved from *www.electronicmarkets.org/electronic markets/electronicmarkets.nsf/pages/emw_0004_cell.html*.

Schenker, J. (2000). Europe's high-tech start-ups have stolen a march on Silicon Valley by launching new mobile services. *Time Europe*, (February 7), 1.

Zabala, H. (2000). M-commerce, the next big thing?. *Asian Business*, (June), 34-35.

KEY TERMS

Localization: Through GPS technology, service providers can accurately identify the location of the user so that m-commerce providers will be better able to receive and send information relative to a specific location.

M-Commerce: The ability to purchase goods anywhere through a wireless Internet-enabled device (e.g., cellular phone, pager, PDA, etc.). Mobile commerce refers to any transaction with monetary value that is conducted via a mobile network.

Mobile Networks: A system that allows wireless devices to seamlessly interact with traditional telecommunications and other wireless devices.

Personalization: Developing tailored approaches with relevant material packaged for the particular wireless user.

Portals: A gate through which something will pass, in an effort to get to another place. In the traditional sense of the word, the portal is not the desired end-state. Rather, a portal is a necessary or convenient place one must go to get to the desired location.

Ubiquity: Mobile devices offer users the ability to receive information and perform transactions from virtually any location on a real-time basis.

Wireless Communication: Interaction through a wireless device (e.g., cellular phone, pager, PDA, etc.).

P

Principles of Advanced Database Integrity Checking

Hendrik Decker
Instituto Tecnológico de Informática, Spain

INTRODUCTION

Integrity constraints (hereafter, sometimes simply 'constraints') are formal representations of conditions for the semantic correctness of database records. In science, constraints are usually expressed in declarative knowledge representation languages such as datalog or predicate logic. In commercial databases, they are usually expressed by distinguished SQL statements.

BACKGROUND

Integrity has always been regarded as an important issue for database management, as attested by many early publications (e.g., Fraser, 1969; Wilkes, 1972; Eswaran & Chamberlin, 1975; Hammer & McLeod, 1975; Nicolas, 1978; Hammer & Sarin, 1978; Codd, 1979; Bernstein, Blaustein & Clarke, 1980; Nicolas, 1982; Bernstein & Blaustein, 1982); later ones are too numerous to mention. To express database semantics as invariants, that is, properties persisting across updates, had first been proposed by Minsky (1974). Florentin (1974) suggested to express integrity constraints as predicate logic statements. Stonebraker (1975) proposed to formulate and check integrity constraints declaratively as SQL-like queries.

Referential integrity, a special case of functional dependencies (Armstrong, 1974), has been included in the 1989 SQL ANSI and ISO standards (McJones, 1997). The SQL2 standard (1992) introduced the CHECK option and the ASSERTION construct as the most general means to express arbitrary integrity constraints declaratively in SQL (Date & Darwen, 1997). In the 1990s, uniqueness constraints, foreign keys, and complex queries involving EXISTS and NOT became common features in commercial database products. Thus, arbitrarily general integrity constraints could now be expressed and evaluated in most relational databases.

Integrity constraints may involve nested quantifications over huge extents of several tables. Thus, their evaluation can easily become prohibitively costly. Most SQL databases offer efficient support only for the following three simple kinds of declarative constraints:

- Domain constraints, i.e., restrictions on the permissible range of scalar attribute values of tuples in table columns, including options for default and null values.
- Uniqueness constraints, as enforced by the UNIQUE construct on single columns, and UNIQUE INDEX and PRIMARY KEY on any combination of one or several columns in a table, preventing multiple occurrences of values or combinations thereof.
- Foreign key constraints, for establishing an identity relationship between columns of two tables. For instance, a foreign key on column emp of relation works_in may require that each emp value of works_in must occur in the emp_id column of table employee, where the referenced columns (here, emp_id) must be a primary key.

For more general constraints, SQL database manuals usually ask the designer to renounce declarative constructs and instead resort to procedural triggers and stored procedures. However, declarativity does not need to be sacrificed in order to obtain efficiency. One approach developed to that end, in the framework of predicate logic and datalog, was *soundcheck* (Decker, 1986). In the spirit of the latter, a translation of integrity constraints expressed in predicate logic into SQL is described in Decker (2003).

SIX PHASES OF SIMPLIFIED INTEGRITY CHECKING

Below, the *soundcheck* approach for simplifying the evaluation of integrity constraints is outlined as a succession of six phases. Except Step I, proposed in Decker (1987), this approach originates in Nicolas (1982). All or part of it is used in one way or another in most known methods for integrity checking. It can be easily implemented in SQL (Decker, 2003). In the next section of this article, Steps I-VI are illustrated with an example. The six phases are then discussed in general.

I Generate the difference between the old and the new state
II Skip idle updates
III Focus on relevant integrity constraints
IV Specialize relevant constraints
V Optimize specialized constraints
VI Evaluate optimized constraints

AN EXAMPLE OF SIMPLIFIED INTEGRITY CHECKING

For illustrating Steps I-VI, consider an update of a relational database with tables for workers and managers, defined as follows.

```
CREATE TABLE(worker(CHAR name, CHAR department))
CREATE TABLE(manager (CHAR name)).
```

Now, suppose there is an integrity constraint requiring that no worker is a manager, expressed by the SQL condition:

```
NOT EXISTS (SELECT . FROM worker, manager WHERE
worker.name = manager.name).
```

If the number of workers and managers is large, then checking whether this constraint is violated or not can be very costly. The number of facts to be retrieved and tested is in the order of the cardinality of the cross product of worker and manager, whenever the constraint is checked. Fortunately, however, the frequency and amount of accessing stored facts can be significantly reduced by taking Steps I-VI. Before walking through them, a possible objection at this stage needs to be dealt with.

SQL programmers might feel compelled to point out that the constraint above is probably much easier checked by a trigger such as:

```
CREATE TRIGGER ON worker FOR INSERT :
IF EXISTS
(SELECT * FROM inserted, manager WHERE
inserted.name = manager.name)
ROLLBACK.
```

Its evaluation would only need to access manager and a cached relation inserted containing the row to be inserted to worker, but not the stored part of worker. However, it is easily overlooked that the sample integrity constraint also requires implicitly that somebody who is promoted to a manager must not be a worker, thus necessitating a second trigger for insertions into manager. In general, each occurrence of each atom occurring in a constraint requires a separate trigger, and it is by far not always as obvious as in the simple example above how they should look like. Apart from being error-prone, hand-coded triggers may also bring about unpredictable effects of mutual interactions that are hard to control. Hence, hand-coding triggers, as recommended in many database manuals, hardly seem advisable.

Now, let INSERT INTO worker VALUES ('Fred', 'sales') be an update. Then, running Steps I through VI means the following:

I Generate difference between old and new state

The explicit update INSERT INTO worker VALUES ('Fred', 'sales') may have implicit update consequences on database views, the definition of which involves worker. The set of explicitly and implicitly updated facts constitutes the difference Δ between old and new database state. Each fact in Δ may violate integrity. Thus, Δ must be generated, and each fact in Δ needs to be run through Steps II-VI. For example, suppose a view containing all workers entitled to some benefit, for example, if they work in some distinguished department d, and a constraint C on that view. Then, C needs to be evaluated only if Fred's department is d; otherwise, no additional constraint needs to be checked.

II Skip idle updates

If Fred already has been a worker (e.g., in some other department) before the INSERT statement was launched, then it is not necessary to check again that he must not be a manager, since that constraint has already been satisfied before.

III Focus on relevant integrity constraints

Unless II applies, the constraint that no worker must be manager is clearly relevant for the given update and hence must be checked. Any integrity constraint that is not relevant for the insertion of rows into the worker table needs not be checked. For instance, a constraint requiring that each department must have some least number of workers is not relevant for insertions, but only for deletions in the worker table. Also, constraints that do not involve worker need not be checked.

IV Specialize relevant constraints

For the given INSERT statement, the WHERE clause of the SQL condition:

```
EXISTS (SELECT * FROM worker, manager WHERE
worker.name = manager.name)
```

can be specialized to a much less expensive form:

EXISTS (SELECT * FROM worker, manager WHERE worker.name = 'Fred' AND worker.name = manager.name)

V Optimize specialized constraints

Clearly, the specialized condition in IV can be optimized to

EXISTS (SELECT * FROM manager WHERE name = 'Fred')

VI Evaluate optimized constraints

Evaluation of the query whether Fred is a manager means looking up a single fact in a stored relation. That of course is much less costly than evaluating all integrity constraints in their full generality, as would be necessary without having done I-V.

The example above is very simple. However, by running Steps I-VI, the same proportions of simplification of integrity checking are obtained systematically for arbitrarily complex constraints.

PRINCIPLES OF SIMPLIFIED INTEGRITY CHECKING

Below, the six phases of simplified integrity checking are again walked through, abstracting general principles from the example in the previous section.

Let D be a relational database, where views, if any, are defined by datalog queries possibly involving negation. Further, let IC be a predicate logic sentence representing some integrity constraint in D, and fact be a relational tuple to be either inserted or deleted. Then, doing I through VI means the following:

I Generate the difference between the old and the new state

For each view v in D and each occurrence A of an atom in the definition of v which matches fact (i.e., fact and A have the same relation name and their column values can be unified), the insertion of fact may cause an implicit insertion in v if A is not negated, and an implicit deletion in v if A is negated. Conversely, the deletion of fact may cause an implicit deletion in v if A is not negated, and an implicit insertion in v if A is negated. Recursively, each implicit update may cause further implicit updates, in case of views defined by other views. The original explicit update and all implicitly updated facts constitute the difference Δ. Each fact in Δ has to be generated and run through II-VI.

II Skip idle updates

Usually, identical copies of a tuple in a table or view are considered redundant ("idle") and should be avoided. More precisely, an insertion of a fact is idle if that fact is already present in the table or derivable from a view definition, and a deletion of a fact is idle if a (view-defined or explicit) copy of that fact will persist after the update. For each such idle update, all of III-VI can be skipped.

III Focus on relevant integrity constraints

The general rule for focusing integrity checking on those constraints that are relevant is as follows: IC is relevant for the insertion (resp., deletion) of fact, and thus needs to be checked, if and only if there is an atom A with negative (resp., positive) polarity in IC which unifies with fact.

Polarity is defined as follows: A has positive polarity in A; if A has positive (resp., negative) polarity in some formula B, then A has positive (resp., negative) polarity in the formulas $A \wedge B$, $A \vee B$, $B \rightarrow A$ and also in the universal and the existential closure of B; A has negative (resp., positive) polarity in $\neg A$ and in $A \rightarrow B$.

IV Specialize relevant constraints

The focus obtained in III can be further narrowed by specializing variables to ground values in updated facts. The general rule is as follows: for IC, fact, and A as in III, let ϕ be a most general unifying substitution of fact and A. Then, IC can be specialized to ICθ, where the substitution θ is obtained from ϕ by restricting ϕ to the universally ($\forall$ -) quantified variables in A that are not dominated by an existential quantifier $\exists$ (i.e., no $\exists$ occurs on the left of $\forall$ in IC).

By definition, the quantification of a variable in a formula is obtained by moving either all negations innermost or all quantifiers outermost, such that logical equivalence is preserved (for details, see Decker, 2003; Nicolas, 1982).

V Optimize specialized constraints

The following rewrites should be applied to specialized constraints obtained from IV. If fact is to be inserted (resp., deleted), replace all occurrences of fact by true (resp., false) and recursively apply obvious simplifications: replace $\neg$true by false and $\neg$false by true; replace each of true $\rightarrow$ B, true $\wedge$ B, false $\vee$ B by B; true $\vee$ B and false $\rightarrow$ B by true; false $\wedge$ B by false. For more on such simplifying optimizations, see Demolombe (1989).

VI Evaluate optimized constraints

For evaluating constraints resulting from I-V, built-in database query optimization facilities and access plan improvements should be exploited as far as available. If applicable, also other techniques for evaluating integrity constraints should be used, for example, query containment techniques (Gupta, Sagiv, Ullman & Widom, 1994) and semantic query optimization (Godfrey, Gryz & Zuzarte, 2001).

Neither of the latter two approaches has originally been meant to be applied specifically to simplified constraints as obtained from I-V, but both can be adapted for use in Step VI.

FUTURE TRENDS

In the future, the importance of database integrity is likely to grow further, hand in hand with increasing concerns for the quality, reliability, and dependability of data. Also the declarativity of data integrity is going to become more important, because of the growing complexity of data and applications, the development and maintenance of which would otherwise become too troublesome. The SQL99 standard has tried to do justice to this trend by proposing norms for non-scalar data types, recursive view definitions, and triggers, which however have hardly been taken up uniformly by database vendors (Gorman, 2001).

Commercial databases lag behind the state of the art of integrity checking in academia. Besides a lack of support for the simplification of ASSERTION statements, also the support for declarative constraints on views is scant, if not completely missing in many database products. Technologies still to be transferred from theory to practice are aplenty (see, e.g., Decker, 1998). In practice, integrity support in distributed databases is even worse than for singleton systems (e.g., declarative referential integrity across remote tables is not available), in contrast to a vast literature on this topic (e.g., Ibrahim, 2002; Bright & Chandy, 2003). Also the convergence of distributed databases and agents technology is likely to benefit from advanced approaches of handling integrity constraints efficiently (Sadri & Toni, 2000). Both theory and practice need to advance in terms of integrity support for replicated data. And, beyond distributed databases, there is the problem of the integrity of data on the Web, which certainly is going to receive more attention in the future (for a primer, see Aldana, Yagüe & Gómez, 2002).

CONCLUSION

This article has outlined basic principles for checking semantic integrity constraints that are more advanced than the usual SQL support for constraints related to data types, attribute values, and referential integrity. The principles have been described by examples using SQL syntax, and then more generally in the language of predicate logic.

A systematic translation of constraints expressed in predicate logic into SQL is described in Decker (2003). There, also, a fully automated and provably correct mechanism for translating declarative constraints into efficient triggers is specified. It takes advantage of all simplifications outlined above. Other work related to similar topics is documented in Cochrane, Pirahesh, and Mattos (1996), Ceri, Cochrane, and Widom (2000), and Date (2000).

REFERENCES

Aldana, J., Yagüe, I. & Gómez, L. (2002). Integrity issues in the Web: Beyond distributed databases. In J. Doorn & L. Rivero (Eds.), *Database integrity: Challenges and solutions* (pp. 293-321). Hershey, PA: Idea Group Publishing.

Armstrong, W. (1974). Dependency structures of data base relationships. *Proceedings of IFIP "74* (pp. 580-583). Amsterdam: North-Holland.

Bernstein, P. & Blaustein, B. (1982). Fast methods for testing quantified relational calculus assertions. *Proceedings of SIGMoD'82* (pp. 39-50). ACM Press.

Bernstein, P., Blaustein, B. & Clarke, E. (1980). Fast maintenance of semantic integrity assertions using redundant aggregate data. *Proceedings of the 6th VLDB* (pp. 126-136). IEEE Computer Society.

Bright, J. & Chandy, J. (2003). Data integrity in a distributed storage system. *Proceedings of PDPTA'03* (pp. 688-694). CSREA Press.

Ceri, S., Cochrane, R. & Widom, J. (2000). Practical applications of triggers and constraints: Successes and lingering issues. *Proceedings of the 26th VLDB* (pp. 254-262). San Francisco: Morgan Kaufmann.

Cochrane, R., Pirahesh, H. & Mattos, N. (1996). Integrating triggers and declarative constraints in SQL database systems. *Proceedings of the 22nd VLDB* (pp. 567-578). San Francisco: Morgan Kaufmann.

Codd, E. (1979). Extending the data base relational model to capture more meaning. *ACM Transactions on Database Systems, 4*(4), 397-434.

Date, C. (2000). *What, not how: The business rules approach to application development*. Addison-Wesley.

Date, C. & Darwen, H. (1997). *A guide to the SQL standard*. Addison-Wesley.

Decker, H. (1987). Integrity enforcement on deductive databases. In L. Kerschberg (Ed.), *Experts database systems* (pp. 381-195). Benjamin Cummings.

Decker, H. (1998). Some notes on knowledge assimilation in deductive databases. In B. Freitag, H. Decker, M. Kifer & A. Voronkov, A. (Eds.), *Transactions and change in* *logic databases* (pp. 249-286). Berlin: Springer-Verlag (LNCS 1472).

Decker, H. (2003). Translating advanced integrity checking technology to SQL. In J. Doorn & L. Rivero (Eds.), *Database integrity: Challenges and solutions* (pp. 203-249). Hershey, PA: Idea Group Publishing.

Demolombe, R. & Illarramendi, A. (1989). Heuristics for syntactical optimization of relational queries. *Information Processing Letters, 32*(6), 313-316.

Eswaran, K. & Chamberlin, D. (1975). Functional specifications of a subsystem for database integrity. *Proceedings of the 1st VLDB* (pp. 48-68). ACM Press.

Florentin, J. (1974). Consistency auditing of databases. *The Computer Journal, 17*(1), 52-58.

Fraser, A. (1969). Integrity of a mass storage filing system. *The Computer Journal, 12*(1), 1-5.

Godfrey, P., Gryz, J. & Zuzarte, C. (2001). Exploiting constraint-like data characterizations in query optimization. *Proceedings of SIGMoD'01*. Retrieved from *www.acm.org/sigs/sigmod/sigmod01/eproceedings/*.

Gorman, M. (2001). Is SQL really a standard anymore? *Whitemarsh Information Systems Corporation.* White Paper. Retrieved March 2004 from *www.ncb.ernet.in/education/modules/dbms/SQL99/issqlarealstandardanymore.pdf*.

Gupta, A., Sagiv, Y., Ullman, J. & Widom, J. (1994). Constraint checking with partial information. *Proceedings of the 13th PODS* (pp. 45-55). ACM Press.

Hammer, M. & McLeod, D. (1975). Semantic integrity in relational data base systems. *Proceedings of the 1st VLDB* (pp. 25-47). ACM Press.

Hammer, M. & Sarin, S. (1978). Efficient monitoring of database assertions (abstract). *Proceedings of SIGMOD'78* (p. 159). ACM Press.

Ibrahim, H. (2002). A strategy for semantic integrity checking in distributed databases. *Proceedings of the 9th ICPDS* (pp. 139-144). IEEE Computer Society.

McJones, P. (1997). The 1995 SQL reunion: People, projects, and politics. *SRC Technical Note 1997-018.* Retrieved from *www.mcjones.org/System_R/SQL_Reunion_95/sqlr95.html*.

Minsky, N. (1974). On interaction with data bases. *Proceedings of SIGMOD'74* (Volume 1, pp. 51-62). ACM Press.

Nicolas, J.-M. (1978). First order logic formalization for functional, multivalued and mutual dependencies. *Proceedings of SIGMOD'78* (pp. 40-46). ACM Press.

Nicolas, J.-M. (1982). Logic for improving integrity checking in relational data bases. *Acta Informatica, 18,* 227-253.

Sadri, F. & Toni, F. (2000). Abduction with negation as failure for active and reactive rules. *AI*IA 99: Advances in artificial intelligence* (pp. 49-60). Berlin: Springer-Verlag (LNCS 1792).

Stonebraker, M. (1975). Implementation of integrity constraints and views by query modification. *Proceedings of SIGMOD'75* (pp. 65-78). ACM Press.

Wilkes, M. (1972). On preserving the integrity of data bases. *The Computer Journal, 15*(3), 191-194.

KEY TERMS

Business Rule: Originally, a statement that defines or constrains the evolution of data pertaining to an enterprise's business. Business rules usually are implemented by integrity constraints or triggers or stored procedures.

Declarative vs. Procedural: Integrity constraints are declarative statements, expressed in languages such as predicate logic, datalog, or SQL. Since their evaluation can be very costly, the potentially troublesome hand-coding of procedural triggers and stored procedures is recommended by most database manuals. The main thrust of this chapter is about reducing the cost of using declarative integrity constraints and avoiding hand-coded implementation.

Integrity: In databases, 'integrity' is largely synonymous to 'semantic consistency', that is, the correctness of stored data with regard to their intended meaning. Integrity, as expressed by integrity constraints, should not be confused with a namesake issue often associated with data security.

Integrity Checking: Systematic tests, ensuring that integrity remains satisfied (see integrity satisfaction). If integrity is violated (see integrity violation), then the update causing inconsistency must be rejected or some other action must be taken to enforce a consistent state, i.e., one that satisfies integrity (see integrity enforcement).

Integrity Constraint: An integrity constraint is a requirement on the consistency of the information stored in a database. Integrity constraints are stated as yes/no conditions in the database schema. They are required to be satisfied in each database state. Thus, integrity constraints are invariants of the database which evolves via updates.

Integrity Enforcement: Actions taken to ensure that integrity remains satisfied across database updates (see integrity satisfaction). Conservative integrity enforcement rejects updates that would violate integrity (see integrity violation). Progressive integrity enforcement attempts to satisfy the update while maintaining integrity by further data modifications, possibly consulting the user.

Integrity Satisfaction: A given database state satisfies integrity if each integrity constraint in the database schema, posed as a query, returns the required yes/no answer. An ostensibly equivalent, but in fact slightly weaker definition says that integrity is satisfied if it is not violated (see integrity violation). The difference between both definitions manifests itself in incomplete databases.

Integrity Violation: In a given database state, integrity is violated if any one of the integrity constraints, posed as a query, returns the opposite of the required yes/no answer.

Predicate Logic: Well-established universal declarative data and knowledge representation language, with capacities of formalizing and automating deductive and other forms of logical reasoning. Its fundamental role for databases and particularly integrity constraints is undisputed and indispensable.

Semantic Integrity: The adjective in 'semantic integrity' is often added to distinguish a set of explicitly defined integrity constraints, either from structural constraints that are implicitly given by the used data model (e.g., data types in relational databases must be scalar) or from integrity as required for the sake of security.

Simplification of Integrity Constraints: Steps taken in order to reduce complexity and costs of integrity checking. Steps I-VI, as described in this article, are prime examples.

SQL: Widely used standard language for the declarative description, querying, and updating of stored data. SQL also allows expression of database schemata including integrity constraints. By an extension with procedural constructs, it also allows the expression and execution of triggers.

Static vs. Dynamic Integrity Constraints: By default, integrity constraints correspond to static properties that are independent of particular database states. As opposed to that, dynamic integrity constraints refer explicitly to several states, mostly consecutive ones, or to their transitions, involving temporal or other procedural integrity constraint constructs. A typical example of a genuine dynamic integrity constraint (otherwise not dealt with in this chapter) is 'salaries must never decrease'.

Trigger: An SQL statement that automatically generates an action (e.g., an update or a reset) upon a specified update request. Triggers are a procedural means to enforce and preserve integrity (see integrity enforcement).

Principles to Guide the Integration and Implementation of Educational Technology

Sara Dexter
University of Virginia, USA

INTRODUCTION

The Educational Technology Integration and Implementation Principles (eTIPs) are six statements that describe the K-12 classroom and school-level conditions under which the use of technology will be most effective. The eTIPs are an example of materials that can aid teachers in designing instruction and participating in creating supportive conditions for technology supported classroom instruction.

BACKGROUND

During the last decade, the call for teachers to be better prepared to teach with technology (CEO Forum, 1999, 2000; Office of Technology Assessment, 1995) has been repeated several times. In response, there are now standards in place to which new teachers are being held that explicitly describe the technology skills all teachers should have to be prepared to teach in a 21st Century school. These include the National Education Technology Standards for Teachers (ISTE, 2000), which were adopted by National Council for Accreditation of Teacher Education (NCATE) as a part of its accreditation requirements, and the Interstate New Teacher Assessment and Support Consortium standards (INTASC, 1992) used by many states as licensing requirements. In general, these standards call for teachers to be able to use technology in the classroom to plan and design learning environments and experiences, and support teaching, learning, and the curriculum. These standards, in turn, imply that teachers must make the consideration of technology use a routine part of their instructional decision making.

Teachers' decision making has been defined as the course of action during which teachers gather, organize, and interpret information, generate alternatives, select a specific course of action, and, after its implementation, consequently evaluate the effectiveness of the decision (Clark & Yinger, 1977; Lipham, 1974). The research literature emphasizes how critical teachers' planning and interactive decisions are in determining what they do, or do not do, in the classroom (e.g., Clark & Yinger, 1977; Jackson, 1968; Peterson & Clark, 1978; Shavelson, 1976). Shavelson and Stern (1981) posit that teachers' decision-making processes are influenced by schemata that are activated from memory.

MAIN THRUST OF CHAPTER

The Educational Technology Integration and Implementation Principles (or eTIPs) are one example of a set of statements that could serve as a schema, or the basis of a schema, for a teacher to organize his or her instructional decision making about the integration and implementation of technology. Principles can assist learners in recognizing and connecting ideas and in seeing how new and old ideas relate (Marzano, 2001), which are key tasks in developing the more elaborate schemas that are characteristic of expert teachers (Carter, 1990; Fogarty, Wang, & Creek, 1983; Kagan, 1992). The six eTIPs summarize what research suggests are the conditions that should be present in order for educational technology integration and implementation to be effective (Dexter, 2002), while offering the advantage of brevity over the 23 NETS-T standards and the five technology-specific statements in the INSTASC standards.

These eTIPs are organized into two dimensions: classroom and school-wide. The classroom principles expand upon the premise that effective technology integration requires the time and attention of teachers in the role of instructional designers, planning the use of the technology so it will support student learning. They assume that educational technology does not possess inherent instructional value but that a teacher must design into the instruction any value technology adds to the teaching and learning processes. Thus, the three classroom eTIPS prompt a teacher-designer to consider what he or she is teaching, what added value the technology might bring to the learning environment, and how technology can help to assess student learning.

Classroom-Level eTIPs

eTIP 1: Learning Outcomes Drive the Selection of Technology.

In order for learning outcomes to drive the selection of technology, teachers must first be clear about their lesson or unit's student-learning outcomes. This is an important first step in determining whether or not the educational technology available can be a support to teaching and learning. It will allow teachers to be more efficient as they search for available, appropriate technologies because they will quickly eliminate those that do not support their learning outcomes. Where technology does seem to support learning outcomes, teachers must also consider the cognitive demands made by the technology and if they are well-suited to the cognitive demands inherent in the learning outcomes. For example, if a learning outcome asks students to analyze or synthesize information, a drill and practice program or reference material on a CD-ROM probably isn't going to match as well as concept mapping or database software.

eTIP 2: Technology Use Provides Added Value to Teaching and Learning.

Using technology to add value—meaning to make possible something that otherwise would be impossible or less viable to do—might mean that it helps to individualize instruction or make it more responsive to a student's questions and interests or that it provides additional resources of information so instruction is more real-world, authentic, and current. Educational technology can also aid teachers in providing "scaffolds" that support learners as they move from what they already know and can do to what they are learning. Educational technology can also help teachers to create social arrangements that support collaborative as well as independent learning by facilitating communication and interaction patterns. Teachers can also use educational technology to support additional opportunities for learners to practice, get feedback, or allow for revision or reflection; thus, it supports knowledge acquisition and practice, so learners become more fluent in their knowledge.

Educational technology can aid students accessing information or representing it in new ways. It can increase access to people, perspectives, or resources and to more current information. Many times, software's interface design allows learner interaction or presents information in a multi-sensory format. Hyperlinks can allow learners to easily connect to related information. Built-in indexes and key word searching support learners by easing their search through a large amount of information to find what is relevant. These features all add value by increasing access to data or the users' control during that access. In terms of processing information, added value might mean that the educational technology supports students learning-by-doing or aids them in constructing mental models, or making meaning, by scaffolding their thinking. For example, a database can allow students to compare, contrast, and categorize information through query features. By asking students to create products with tool software, it requires them to think more deeply about the material in order to represent it with that tool (Jonassen, 2000). Educational technology can also add value to students' ability to show and articulate to others about what they have learned.

eTIP 3: Technology Assists in the Assessment of the Learning Outcomes.

At times, teachers will want to collect and return to students formative data, to let them know about their learning progress. Some software or hardware actually collects formative data during its use, and some technologies also provide help in the analysis of the information. Generally, these are software programs designed to assess student learning, such as tutorial or drill and practice software. Some of these programs, through screens or printouts of information, or other feedback mechanisms, support students' self-assessment of their learning.

Teachers will also want to collect summative information about students' achievement of the learning outcomes. Technology can assist teachers in collecting data that will help them understand how students are meeting or have met the learning outcomes for that lesson or unit. Products students produce through software, whether a database, "mind map," multimedia or word-processed report, or a Web site, demonstrate what they have learned about both the content of their product, the procedural knowledge required to produce it, and their ability to communicate. The capabilities a product might demonstrate include the skills of editing, analysis, group collaboration, or the operation of the software itself.

School-Level eTIPs

Part of what makes teachers' integration activities feasible or not is the level of technology support at a school. The three school-wide principles elaborate upon the premise that the school environment must support teachers in a role of instructional designer by providing adequate technology support. The presence of high-quality technology support programs are correlated to teachers' increased uses of educational technology (Dexter, Anderson & Ronnkvist, 2002). Thinking about the school-level principles while deciding whether or how to integrate technology can help a teacher to take the "technology ecology" of the setting into perspective during instructional design. Together they will help teachers to evaluate the level of access and support available to them in their integration work, which may help to determine whether or not, given their amount of planning time, a particular integration goal is realistic.

eTIP 4: Ready Access to Supported, Managed Hardware/Software Resources is Provided.

Teachers must have convenient and flexible access to and technical support for appropriate educational technology in order for them to utilize it in their classrooms. Perhaps of all the principles, this one is the most self-evident. Without available and working educational technology, it can hardly be utilized in a classroom. But, the two key words in this principle are ready and supported. Ready access means the technology should be close to where teachers need to use it and that it is scheduled flexibly, so that teachers have an opportunity to sign up for it when it is relevant for classroom work. Here, support specifically refers to technical support like troubleshooting help and scheduled maintenance. The idea of ready access should raise for the teacher questions about whether or not the students could be grouped together to work with the educational technology, if it could be a station through which students rotated, or if all students need to have simultaneous access to the educational technology. Ultimately, the access has to be practical. It must be ready enough that working through the logistics of providing students access to the technology does not outweigh the added value it provides.

eTIP 5: Professional Development is Targeted at Successful Technology Integration.

Technology professional development is key to teachers learning to integrate technology effectively into the classroom (CEO Forum, 1999). The learning needs can be thought of as being about (1) learning to operate the software and, (2) learning to use software as an integrated, instructional tool. Too often teachers' learning opportunities are just about the operation of the software. This is necessary, but teachers must also have learning opportunities that address more than these basic skills. Specifically, these learning opportunities should guide teachers in the instructional design I have laid out in the three classroom educational technology integration principles. By having sufficient time to explore educational technology and have their technological imagination sparked by examples of it in use, teachers can identify which materials match their learning outcomes (eTIP #1). Professional development sessions should also provide frameworks or criteria that can aid a teacher in determining whether or not an educational technology resource brings any added value to teaching or learning (eTIP #2). Likewise, through examples and discussion, teachers should have opportunity to consider how educational technology might aid the formative or summative assessment of students' learning (eTIP #3).

eTIP 6: Professional Community Enhances Technology Integration and Implementation.

This principle describes a professional collaborative environment for integrating and implementing technology. In such an environment, technology use would be more effective because the school organization would recognize the contribution individuals make to the collective knowledge of the school (Marks & Louis, 1999). And the entire staff would work toward consensus about the school's performance, in this case with technology, and how they could improve it (Marks & Louis, 1997). A collaborative professional community would serve as the vehicle for school-wide knowledge processing about technology integration and implementation, increasing the likelihood of reflective dialogue, sharing of instructional practices, and generally increasing collaboration on new practices.

FUTURE TRENDS

As educational technology and Internet access become ubiquitous in classrooms and new teachers headed into the classroom arrive from college already skilled in the operation of technology, it is likely that educational technologists will then be able to shift their research efforts to how best to develop teachers' instructional decision making about technology integration and implementation. This suggests that further research is needed about the schema of expert technology integrating teachers, and the key cognitive processes involved in designing and implementing effective technology integrated instruction. Future development efforts are needed in the area of developing instructional supports, such as cases and simulationsthat will aid novice integrators in developing the necessary knowledge and skills.

CONCLUSION

The research literature about teachers' instructional planning suggests that teacher educators working to develop K-12 educators' abilities to incorporate educational technology into the classroom should attend to the development of teachers' schema about technology integration and its implementation. By serving as a schema, or the basis of one, the Educational Technology Integration and Implementation Principles (eTIPs) can help teachers recognize and plan for the effective technology use that is represented in the NETS-T and INTASC standards. The eTIPs point out two key aspects of teachers designing effective integrated instruction: the technology use must match and support teaching and learning, and the larger school environment must provide support for the logistical and learning demands technology integration puts on teachers.

REFERENCES

Carter, K. (1990). Teachers' knowledge and learning to teach. In W. R. Houston (ed.), *Handbook of research on teacher education,* pp. 291-310. New York: Macmillan.

CEO Forum on Education and Technology (1999). Professional development: A link to better learning. Retrieved February 16, 2003, from: http://www.ceoforum.org/reports.cfm?RID=2.

CEO Forum on Education and Technology (2000). Teacher preparation STaR chart: A self-assessment tool for colleges of education. Retrieved February 16, 2003, from: http://www.ceoforum.org/reports.cfm?RID=3.

Clark, C. M. & Yinger, R. J. (1977). Research on teacher thinking. *Curriculum Inquiry,* 7(4), 279- 304.

Dexter, S. (2002). eTIPS-educational technology integration and implementation principles. In P. Rodgers (ed.), *Designing instruction for technology-enhanced learning,* pp.56-70. Hershey, PA: Idea Group Publishing.

Dexter, S., Anderson, R. E., & Ronnkvist, A. (2002). Quality technology support: What is it? Who has it? and What difference does it make*? Journal of Educational Computing Research*, *26* (3), *287-307*.

Fogarty, J. L., Wang, M. C., & Creek, R. (1983). A descriptive study of experienced and novice teachers' interactive instructional thoughts and actions. *Journal of Educational Research,* 77(1), 22-32.

International Society for Technology in Education [ISTE] (2000). National educational technology standards for teachers. Eugene, OR: International Society for Technology in Education.

Interstate New Teacher Assessment and Support Consortium [INTASC] (1992). Model standards for beginning teacher licensing and development: A resource for state dialogue. Retrieved September 27, 2004 from: *http://www.ccsso.org/content/pdfs/corestrd.pdf*

Jackson, P. W. (1968). *The way teaching is.* Washington, DC: National Education Association.

Jonassen, D.H. (2000). *Computers as mindtools for schools: Engaging critical thinking (2nd ed.).* Columbus, OH: Prentice-Hall.

Kagan, D. M. (1992). Professional growth among preservice and beginning teachers. *Review of Educational Research,* 62(2), 129-169.

Lipham, J. M. (1974). Making effective decisions. In J. A. Culbertson, C. Henson, & G. Morine-Deshimer (1978-1979). Planning and classroom reality: An in-depth look. *Educational Research Quarterly, 3*(4), 83-99.

Marks, H, M. & Louis, K.S. (1999). Teacher empowerment and the capacity for organizational learning. *Educational Administration Quarterly,* 5, 707-750.

Marks, H. M. & Louis, K.S. (1997). Does teacher empowerment affect the classroom? The implications of teacher empowerment for instructional practice and student academic performance. Educational Evaluation & Policy Analysis, 3, 245-275.

Marzano, R. J. (2001). *Designing a new taxonomy of educational objectives*. Thousand Oaks, CA: Corwin Press, Inc.

Office of Technology Assessment (1995). *Teachers and technology: Making the connection*, OTA-EHR-616, April. Washington, DC: U.S. Government Printing Office.

Peterson, P. L. & Clark, C. M. (1978). Teachers' reports of their cognitive processes during teaching. *American Educational Research Journal*, 15(4), 555-565.

Shavelson, R. J. (1976). Teachers' decision making. In N. L. Gage (ed.), *Yearbook of the National Society for the Study of Education: The psychology of teaching methods*. Chicago: University of Chicago Press.

Shavelson, R. J. & Stern, P. (1981). Research on teachers' pedagogical thoughts, judgments, decisions, and behavior. *Review of Educational Research*, 51, 455-498.

KEY TERMS

Added Value: traditional usage is as an indication that the particular packaging, delivery method, or combination of services in a product brings extra benefits than one would otherwise receive. Applied to educational technology, it communicates that the use of technology brings added value to the teaching or learning processes when it makes possible something that otherwise would be impossible or less viable to do.

Principle: ideas that can assist learners in recognizing and connecting ideas and in seeing how new and old ideas relate.

Professional Community: collaborative activities among a school's faculty members that focus on meaningful, shared issues in a school and also emphasize how each individual staff member can take responsibility for its achievement.

Professional Development: the National Staff Development Council defines this as the term that educators use to describe the continuing education of teachers, administrators, and other school employees.

Schema: mental constructs that aid learners in categorizing problems or situations and selecting appropriate courses of action for their effective resolution.

Technology Implementation: the putting into place at a system level of a school the conditions that support the integration of technology at the classroom level.

Technology Integration: the use of technology in teaching to support learning.

Privacy–Dangers and Protections

William H. Friedman
University of Central Arkansas, USA

INDIVIDUAL AND SOCIAL CONCERNS

It is no wonder that the average citizen is concerned about the difficulty of guarding one's privacy. Now, your own cell phone can reveal your ever-changing whereabouts by means of "location technology" (Lagesse, 2003). Chips that receive coordinates from global positioning satellites now make it possible to locate persons, cars, merchandise, in short, whatever we value. Like most new technology, it is easy to see advantages as well as drawbacks. Some positives of location technology are that ambulances, police and fire services can reach victims more quickly; driving suggestions can be delivered in real time to motorists (thus helping to avoid traffic tie-ups and prevent getting lost); advertisers can inform potential customers of the existence of a nearby hotel, store or restaurant; stores utilizing RFID (see the **KEY TERMS** section for explanations of possibly unfamiliar terms) can trace merchandise movement to reduce waste, replenish inventory, and stem shoplifting. Some negatives are that nefarious agents can also use location technology to track their prey; location-tracking history can be subpoenaed by one's legal adversaries; and it is inevitable that corporations and government will have an interest in conducting such monitoring (Griffin & Whitehead, 2001, 2002).

Privacy is an area that involves governmental, legal, social, managerial, and technical matters. Claims to privacy rights usually involve one or more senses of the notion of privacy. Table 1 gives some idea of the main connotations of "privacy" that are relevant to and will be applied to information technology; hence, they form the basis for the following discussion.

The first definition embodies respect for a computer user's wish to be in an insulated, protected space; no individual or program should interfere with the user's computer activity in a way that unexpectedly and undesirably disturbs that user. This presupposes freedom from concern that someone might hack in to the system and disrupt it, or barge into a chat room. Intrusive pop-up advertisements and junk e-mail can also contravene the first notion of privacy.

The second definition would suggest that any form of even unconcealed monitoring is a breach of privacy. An important exception in many people's minds is that government agencies might need to monitor computer activity when there is reason to suspect criminal or terrorist activity. A very similar exception arises when employers (Haag, Cummings & McCubbrey, 2004) feel that certain employees are abusing computer privileges, not working appropriately, or are otherwise jeopardizing the company's welfare. This can happen when the employee's computer activity leaves the company open to litigation or when the employee discloses company secrets (Brandt, 2001). Each of these exceptional cases, in order to be tolerated in an otherwise free society, presupposes some warranted authority for the monitoring. Less malicious, but perhaps equally bothersome is commercial monitoring to learn our buying habits by observing our Web searches and purchasing proclivities. Cookies to achieve this end are deposited and visible in a "Cookies" folder on our hard drives. Even when only aggregate, not identifiable individual data are harvested, most people would reject such a practice as beginning to slide down a slippery slope of more serious privacy invasion (Hamilton, 2001).

The third sense of privacy appeals to a desire to be free from being the object of concealed surveillance. Governmental and other spies have tried to plant programs or devices in suspect computers to log activities without the users knowing about it and report what they do to a distant receiving station.

The fourth sense expresses our desire not to suffer anyone's forcing his/her way into our space in plain sight in order to seize our computer or copy our files, in complete disregard of our wishes to keep these to ourselves. To have any property or information taken by stealth is offensive, but even more so for information that might presumably be detrimental to our best interests, such as credit card numbers.

The fifth connotation (as used particularly in database contexts) pertains to the natural expectation that our computer files be reserved exclusively for our own dissemination and access. To lose exclusive control over access and dissemination of our data files leaves us open to blackmail, financial fraud (as with discovery of our credit card numbers), malicious mischief, and even identity theft. While crimes may seem to be the modus operandi of ordinary crooks, we must contend with hackers and "script kiddies" who engage in this sort of pursuit for sport and, of course, sometimes also for profit.

The sixth meaning expresses our concern that when we purchase something or register at a Web site, unless we opt in, we do not ordinarily wish to have the information entrusted to the site shared with any other party. An opt-in permission granted to the site should imply that one is later able to cancel the service, or “opt-out”. A great deal of junk e-mail is delivered under the pretext that, at one point, one did opt-in for the program with a related third party, a statement that may or may not be correct.

Very few downloaders of software have the patience to inspect the license agreement that must receive assent before installation can proceed. Nearly everyone accepts it by checking “I agree” without actually scrolling and reading the extended window full of legalese. Unfortunately, the end-user license agreement may not indicate the full extent of what the software does. An honest EULA may state that the program performs anonymous profiling, an assurance that your computer activities are being documented for future use, but not with identifying characteristics of the user per se. This type of software is used to generate a marketing précis of one’s interests and attributes. Thus, if you visit Web sites that feature a certain service or product, you may be directed to other Web sites that feature the same thing. Vendors advertise with spyware companies, because of the selective targeting offered. By catering to consumers’ preferences as exhibited in Web browsing, they presumably also benefit because they are presented only with items that really interest them and are spared much spam. Despite this alleged advantage and with promises of not recording personal data, many consumers disapprove of this practice on general principles and try to delete the uninvited spyware, often a difficult if not impossible task.

There are many techniques to invade our privacy—in both hardware and software. Even the FBI is reputed to use the infamous but mysterious program, “Magic Lantern,” to spy on criminals attempting to hide their activities when using encryption for their e-mail. “When the user types the password, Magic Lantern software would capture it, giving the agency the ability to decrypt users’ communications” (Paulson, 2002).

A wide variety of such snares awaits law-abiding citizens as well, and can be conveniently lumped under the rubric of “spyware”. Table 2 summarizes some of the spyware that can be used to violate privacy.

A person’s first thought upon discovery of a parasite on his/her computer is how to remove it. Although parasites are not easily removed, anti-parasite programs can detect and remove them. In addition, several Web sites have parasite detection scripts that analyze your computer and provide you with removal instructions (for example, visit www.doxdesk.com/parasite). Parasites are often fellow travelers of demos and freeware that are offered on the Web. A surfer must take the trouble to read the small print that appears in the EULA and privacy policy at the bottom of most Web sites. Sometimes one can opt out of selecting any extra utility that asks be installed along with what you originally ordered.

Due care must be given to permission windows asking if you wish to run a certain program, the function of which is not totally clear. ActiveX controls on the Web cannot only install parasites that compromise our privacy but also install viruses. XSS involves everything: all sorts of nefarious activity such as account hijacking, changing of user settings, cookie theft, and fake advertising.

PRE-INTERNET LEGAL PROTECTION

While the U.S. Constitution does not explicitly mention a legal right to privacy, this right is traditionally considered to have been implied by the first, fourth, and fifth amendments to this constitution (Blumenfeld, 1998). The First Amendment does not really grant privacy in any of the senses mentioned earlier; instead, it guarantees free speech (which could pertain to the Internet, but not really to privacy). The fourth amendment prohibits unreasonable searches and seizures (applicable to data and computer equipment, of course). The Fifth Amendment states: “nor shall private property be taken for public use, without just

Table 1. The relevant senses of the concept of privacy

1. Freedom from being subjected to unwanted contact
2. Freedom from overt monitoring
3. Freedom from secret surveillance
4. Freedom from unauthorized intrusion
5. Confidentiality
6. Freedom from disclosure by others of personal data to unauthorized persons

Table 2. Techniques used to violate computer privacy

Web bugs
Parasites
Key loggers
Cookies
Port scanning
Password capture via algorithms
Cruising an office to detect computers left on
Cruising an office to find written records of passwords
ActiveX control
Cross-site scripting (XSS)

compensation;" but here the concern is with justice, not privacy. The third amendment is quite relevant to privacy, however. It assures that no soldier will be forcibly introduced into a citizen's household, thereby guaranteeing against privacy destroying intrusion by the quartered soldier (Friedman, 2002).

Louis D. Brandeis (1928) became a pioneer in the advocacy of a legal foundation for privacy when he wrote a dissenting opinion for a case about governmental wiretapping to entrap a bootlegger. It is almost prophetic in dealing with as yet unknown technologies:

"The progress of science in furnishing the government with means of espionage is not likely to stop with wire tapping. Ways may some day be developed by which the government, without removing papers from secret drawers, can reproduce them in court…"

The U.S. Privacy Act of 1974 has three important provisions: that there be no secret data banks, that effort must be made to see that the data are reliable, and that the information cannot be used for a purpose different from the original reason for collecting it (Friedman, 2002; Gurak, 2002; Helm, 2000).

POST-INTERNET LEGISLATIVE INITIATIVES TO THE RESCUE

At least a score of bills have been proposed in the U.S. Congress to protect us from Internet privacy abuse—all of the following principal bills and others are described along with their legislative status on this Web site of Thomas (2003):

- TITLE XIII—Children's Online Privacy Protection Act of 1998 (COPPA)
- Online Privacy Protection Act of 2003 (Introduced in the House)
- Consumer Privacy Protection Act of 2003 (Introduced in the House)
- Privacy Act of 2003 (Introduced in Senate)
- Privacy Act of 2003 (Introduced in Senate)
- Identity Theft Protection Act (Introduced in Senate)
- Social Security Number Misuse Prevention Act (Introduced in Senate)

Children under the age of 13 use the Internet for playing games, shopping, schoolwork, or idle surfing. COPPA endeavors to give parents more control by enlisting the aid of the Federal Trade Commission to establish new rules for Web site operators in order to safeguard our children's privacy while they are online. These rules are now part of the 1998 law, known as the Children's Online Privacy Protection Act (COPPA).

A similar act has been proposed to protect adults, namely, the Online Privacy Protection Act of 2003/H.R. 69 H. R. 69, which would require the Federal Trade Commission to prescribe regulations to protect the privacy of personal information collected from and about individuals (who are not covered by the Children's Online Privacy Protection Act of 1998) using the Internet.

An important aim of the Consumer Privacy Protection Act of 2003/H.R. 1636 (introduced in the U.S. House of Representatives) is to provide protection to all consumers by requiring notice of privacy policies and how data gathered from them on the Internet will be used.

Two U.S. Senate bills proposed to protect privacy and confidentiality are the Privacy Act of 2003/S.B. 745, which requires a consumer's prior agreement before personal data can be revealed to a third party, and the Identity Theft Protection Act/S.B. 223, which would require financial institutions to take strong proactive steps to secure the consumer's sensitive data (Cabral, 2001). Another protective bill, the Social Security Number Misuse Prevention Act/S.B. 228, governs the rules for sale and display of one's social security number by others as well as a vendor's requiring it for purchases.

CRITICAL ISSUES PERTAINING TO PRIVACY

Some difficult choices between conflicting values face our technological society (given in Table 3).

FUTURE TRENDS

As governments strive to combat terrorists (Ashcroft, 2004) as well as other suspected criminals by eavesdropping and entering their domiciles unannounced, and as parents and other authorities try to discover sexual activi-

P

Table 3. Conflicting values involving privacy

• Safeguarding one's personal data versus facilitating vendors' catering to our preferences, and as a result lowering costs (Cowley, 2002)
• Keeping one's personal computer activities secret versus the need for national security and fighting crime
• Monitoring employees at work versus infantilizing them psychologically (Friedman, 2002)

ties of children and pornographers, the privacy of even totally innocent parties will be increasingly diminished. Future research might concern the U.S. government's efforts (Center for Democracy and Technology, 2004) to thwart the "SAFE Act, a bipartisan bill that proposes modest fixes to the USA PATRIOT Act to ensure appropriate oversight of the Justice Department" (Center for Democracy and Technology, 2004). The purpose of the Patriot Act as stated in the first paragraph is "to deter and punish terrorist acts in the United States and around the world, to enhance law enforcement investigatory tools, and for other purposes"(USA Patriot Act, 2001).

Furthering this trend, technology is providing low-cost methods of eavesdroping and logging key strokes on a computer, so that idle but unsavorary persons will be able to discover confidental secrets.

CONCLUSION

There is much that an individual or organization can do to protect privacy, in all its connotations. Some of the many proactive and protective countermeasures one can take to thwart abuse are given in Table 4 (all suggestions help with commercial assaults on privacy [which correspond, in Table 1, with all listed senses of privacy]: and the last two might also aid in forestalling unwarranted governmental invasions [usually involving Table 1, senses: 2, 3, and 5]).

REFERENCES

Ashcroft, J. (2004, January 28). Letter to the Honorable Orrin Hatch, Chairman, Committee on the Judiciary, U. S. Senate. *http://www.cdt.org/security/usapatriot/20040128doj.pdf*

Brandeis, L.J. (1928). U.S. Supreme Court, Olmstead v. U.S., 277 U.S. 438. *http://caselaw.lp.findlaw.com/scripts/getcase.pl?court=US&vol=277&invol=438*

Brandt, A. (2001, December). Should your boss be allowed to search your hard drive? *PC World.*

Cabral, R. (2001, June 14-17). U.S. banking—financial privacy. In T. Gregson & D. Yang (Eds.), *2001 Hawaii Conference on Business*, Honolulu, Hawaii. CD-ROM Proceedings.

Table 4. Suggested ways we can act to protect our privacy

- Deal only with Web sites whose privacy policies are clear and to which you can assent
- Opt out of any suspect product or club registration
- Encrypt sensitive files on one's computer
- Utilize protective software (e.g., firewalls)
- Be careful not to reveal too much, knowingly or unknowingly (e.g., social security numbers or passwords)
- Use credit cards that allow one time use of numbers—each transaction uses a different temporary number linked to the real account number
- Work to secure protective legislation
- Join and cooperate with groups formed to fight privacy invasion

Center for Democracy and Technology. (2004, February 1). DOJ objects to SAFE Act. *http://www.cdt.org/*

Cowley, S. (2002, February 13). Chastened Comcast will stop tracking customer Web use. http://www.infoworld.com/articles/hn/xml/02/02/13/020213hncomcast.xml

Friedman, W. (2004, March). Perspectives on privacy. *Journal of American Academy of Business, 4*(1).

Griffin, K., & Whitehead, R. (2001, October). *Privacy in the age of technology. International Business and Economics Research Conf. Proceedings* [on CD-ROM]. (Also in *International Business and Economics Research Journal, 1*(4), April 2002, 29-38).

Gurak, L. (2002, March 15). Professor takes a critical look at online-privacy issues. *The Chronicle of Higher Education,* A38.

Haag, S., Cummings, M., & McCubbrey, D. (2004). *Management information systems for the information age.* New York: McGraw-Hill/Irwin.

Hamilton, R. (2001, June 14 -17,). Privacy Concerns?: Try enhancement data. *2001 Hawaii Conference on Business*, Honolulu, Hawaii.

Helm, A. (2000). *Everyday law.* Made E-Z Products, Inc., Deerfield Beach, FL.

Lagesse, D. (2003, September 8). They know where you are. *U.S. News and World Report Special Report, 135*(7), 32-38.

Paulson, L. (2002, March). Key snooping technology causes controversy. *Computer,* 27.

Thomas. (2003). Legislative information on the Internet.www.thomas.loc.gov

U.S.A. Patriot Act. (2001). http://www.epic.org/privacy/terrorism/hr3162.html

KEY TERMS

ActiveX Control: A Microsoft software module that enables another program to add functionality by calling ready-made components that blend in and appear as normal parts of the program.

Chat Room: Interactive "conversation" involving several people (via typing on a computer terminal or PC) usually centered on a single topic utilizing the Internet. Occasionally members of the chat room break off to form a smaller, more intimate conversation, to the exclusion of all others, thus invoking privacy privileges.

Cookie: Information, usually including a username, Internet address, and the current date and time, placed on the hard drive of a person using the World Wide Web by a Web site that one has visited. This information can be used to identify visitors who have registered or viewed the site, but also to report the visitor's unrelated Web activity, or worse, personal information stored by the user.

Cross-Site Scripting (XSS): An application to trick users into thinking they are dealing with a normal Internet situation, but the real purpose is to gather data from them.

Encryption: A transformation of data from an original readily understandable version (plaintext) to a difficult-to-interpret format (ciphertext) as a way of protecting confidentiality until changed back to the original. Images can also be encrypted to prevent recognition.

EULA: End user license agreement, which spells out the legal rights of both the purchaser and vendor of software.

Exploit: Weakness in a system through which hackers can gain entry into the system.Firewall: Hardware or software to prevent unauthorized users from gaining access to a computer or network.

GIF: Internet graphics image file format.

GPS: Global positioning system. It utilizes a wide-reaching radio broadcasting system, originally produced to aid navigation, consisting of a group of 24 satellites plus terrestrial receiving devices.

Identity Theft: The act of a person who successfully pretends to be another person for the purpose of committing fraud. An identity thief discovers some bit of your personal information and appropriates it for him/herself without your knowledge to acquire a privilege (e.g., driving license), some property, or some service for the victim is billed.

Key Logger, Keystroke Logger: Software that runs imperceptibly during computer use that records each keystroke, either saving the results to a log file or sending the results to a second party.

Magic Lantern: An e-mail containing an attachment that, if opened, would insert a "Trojan horse" that is activated upon launch of the popular encryption program, Pretty Good Privacy. Then it sends all the keystrokes the unwitting user types.

Opt-in: Consciously to accept some situation or condition in a contract. For example, to opt-in to an e-mail subscription means that you want to receive periodic messages, even from third parties (antonym: opt-out).

Parasites: Software that operates in the background and sends details of one's computer activity to an interested party. Such programs are often installed with demos and freeware obtained from the Web.

Pop-Up Advertisements: Windows that suddenly appear while surfing the Web, often involving splashy colors and loud sound effects.

Port Scanning: The practice of sending electronic queries to Internet sites in order to learn what level of security exists. Often the intent is to see if it can be compromised.

RFID: Radio frequency identification. Chips can broadcast information presently 3 to 10 meters away after being scanned by a special device using ordinary radio waves.

Script: A typically small program; also one that might do damage.

Script Kiddies: Derisive term for amateur hackers who illegally invade a system using the path of least resistance.

Spyware: Software that can be rapidly and unobtrusively installed on a computer and that subsequently passes information about one's Internet surfing habits to a designated Web site.

Trojan Horse: A program that appears to be legitimate but is designed to have destructive effects, as to data residing in the computer onto which the program was loaded.

Web Bug: A single pixel, transparent (or sometimes even visible) GIF image that can surreptitiously pass along information about a site user to second party.

.

Process–Aware Information Systems for Virtual Teamwork

Schahram Dustdar
Vienna University of Technology, Austria

INTRODUCTION

The question of the "right" organizational form and the appropriate information systems support remains of paramount importance and still constitutes a challenge for virtually all organizations, regardless of industrial background. Organizations distribute their required work activities among groups of people (teams), with teams constituting the main building block for implementing the work (tasks). In most cases, team members are organized as "virtual (project) teams." These teams are under heavy pressure to reduce time to market of their products and services and lower their coordination costs. Some characteristics of distributed virtual teams are that team (member) configurations change quite frequently and that team members report to different managers, maybe even in different organizations. From an information systems' point of view, distributed virtual teams often are self-configuring networks of mobile and "fixed" people, devices, as well as applications. A newly emerging requirement is to facilitate not just mobility of content (i.e., to support a multitude of devices and connectivity modes) to team members, but also to provide contextual information on work activities to all distributed virtual team members (Dustdar, 2002a, 2002b, 2002c). By context, we mean traceable and continuous views of associations (relationships) between artifacts (e.g., documents, database records), resources (e.g., people, roles, skills), and business processes. Context is composed of information on the "who, when, how, and why." The remainder of this chapter is organized as follows: The next section provides an overview of related work on classification systems of collaborative systems and provides an overview on evaluation aspects of current collaborative systems for virtual teamwork. Section 3 discusses some issues and problems related to the integration of artifacts, resources, and processes. Section 4 presents one proposed solution. Finally, Section 5 discusses some future trends and concludes the chapter.

FUNCTIONAL CLASSIFICATION OF COLLABORATIVE SYSTEMS

There has been a lot of work on classification models for collaborative systems. However, there is no one-and-agreed-upon taxonomy of analyzing and understanding collaborative systems. Academia and industry suggest various classification schemes. In industry, for example, people frequently use the term *e-mail* and *groupware* interchangeably. More generally, there is the tendency to classify categories of collaborative systems by naming a product (e.g., many use the terms *Lotus Notes* and *groupware* interchangeably). Academic research has suggested many different classification models. For a recent survey of collaborative application taxonomies, see Bafoutsou and Mentzas (2002). DeSanctis and Gallupe (1987), Ellis, Gibbs and Rein (1991), and Johansen (1988) suggest a two dimensional matrix based on time and place, where they differentiate between systems' usage at same place/same time (e.g., electronic meeting rooms), same place/different time (e.g., newsgroups), different place/different time (e.g., workflow, e-mail), different place/same time (e.g., audio/video conferencing, shared editors). This classification model helps one to easily analyze many tools on the market today; however, it fails to provide detailed insights on collaborative work activities themselves, as well as their relationship to business processes. Ellis (2000) provides a functionally oriented taxonomy of collaborative systems that helps one to understand the integration issues of workflow and groupware systems. The classification system of Ellis (2000) provides a framework in which to understand the characteristics of collaborative systems and their technical implementations.

The first category (Keepers) provides those functionalities related to storage and access to shared data (persistency). The metaphor used for systems based on this category is a "shared workspace." A shared workspace is basically a central repository where all team members put (upload) shared artifacts (in most cases, documents) and share those among the team members. Technical characteristics of “keepers” include database features, access control, versioning, and backup/recovery control. Examples of popular systems include *BSCW* (Bentley et al., 1997), IBM/Lotus *TeamRoom* (IBM, 2002), and the peer-to-peer workspace system *Groove* (Groove, 2002). The second category (Communicators) groups all functionality related to explicit communications among team members. This boils down to messaging systems (e-mail). Its fundamental nature is a point-to-point interaction model where team members are identified only by their name (e.g., e-mail address) and not by other means (e.g., skills, roles, or other constructs, as in some advanced workflow systems). The third category (Coordinators) is related to the ordering and synchronization of individual activities that make up a whole process. Examples of Coordinator systems include workflow management systems. Finally, the fourth category (Team-Agents), refers to semi-intelligent software components that perform domain-specific functions and thereby help the group dynamics. An example of this category is a meeting scheduler agent. Most systems in this category are not off-the-shelf standard software. Both evaluation models presented above provide guidance to virtual teams on how to evaluate products based on the frameworks. Current systems for virtual teamwork have their strength in one or two categories of Ellis’ framework. Most systems on the market today provide features for Keepers and Communicators support or are solely Coordinator systems (e.g., Workflow Management Systems) or Team-Agents. To the best of our knowledge, there is no system that integrates at least three of the above categories into one system. In the following section, we evaluate current collaborative systems categories for their usage in virtual teams and summarize their shortcomings with respect to the requirement for virtual teamwork.

Evaluation of Collaborative Systems for Virtual Teamwork

Cooperative tasks in virtual teams are increasing, and, as a consequence, the use of collaborative systems is becoming more pervasive. In recent years, it has increasingly become difficult to categorize systems according to the frameworks discussed previously, due to the increasing fuzziness of systems boundaries and to recent requirements for virtual teamwork. Traditional systems in the area of interest to virtual teamwork are groupware, project management (PM) and workflow management systems (WfMS). These system categories are based on different metaphors. Groupware systems mainly can be categorized along two lines (metaphors)—the *communications* or the *workspace* metaphor.

Communications-oriented groupware supports unstructured work activities using communications as the underlying interaction pattern. One very popular instance of communications-oriented groupware is e-mail. When e-mail is used as the main medium for virtual teams (as in most cases), data and associated information (e.g., attachments) remain on central mail servers and/or personal inboxes without any *context* information in which those e-mail communications were used (i.e., involved business processes, performed activities, created artifacts). Enterprise groupware systems generally focus on enterprise-wide messaging and discussion databases and do not support organizational components and structures, such as people and their associated roles, groups, tasks, and skills. This leads to “organizationally unaware” systems that treat all messages alike (semantically) and without any awareness of underlying business processes that are essential for efficient collaboration in project teams.

Workspace-oriented groupware, on the other hand, allows team members to upload or download artifacts using files and folders to organize their work. Groupware, as previously indicated, usually does not implement an underlying organizational model (i.e., providing information on the structure of a team, such as team members and their roles, skills, tasks, and responsibilities). The lack of explicit organizational structuring is a disadvantage and an advantage at the same time. It is disadvantageous because traditional groupware has no “hooks” for integrating business process information, which is important in order to integrate artifacts, resources, and processes. This will be discussed in more depth in the next section. The advantage of the lack of explicit organizational structure information is that these systems may be used in all organizational settings without much prior configuration efforts, and they lead to increased personal flexibility, as the proliferation of e-mail systems in teamwork demonstrates.

The second category, which we will briefly investigate in this section, is *project management systems*. As we have stated, virtual teamwork is, in most cases, organized as project work. Projects have well defined goals and are defined by their *begin* and *end* dates, as well as by the required resources and their tasks (work breakdown structure). It is interesting to note, however, that PM systems traditionally support the work of the project manager as the main (and sometimes the only) user of the PM system. They do not support dynamic interaction (instantiation) of processes. More recently, project management systems combine with information sharing tools (shared workspaces) to provide a persistent storage for artifacts. The enactment of the task by team members, as defined by the project manager, is not supported by PM systems. In other words, we can conclude that PM systems are not geared towards virtual teamwork, but focused more on the planning aspect. They provide “static” snapshots (usually in the form of GANNT charts) of projects and how they "should" be. There is no support for the work activities performed by the virtual team members.

The purpose of *workflow management systems* is to support the notion of processes within and, in some cases, between organizations (Aalst & Kumar, 2001; Bolcer, 2000; Bussler, 1999). However, WfMS' first requirement is to model a business process (build time) and then to enact this model (run time). This leads to substantial inflexibility for virtual teams (Ellis, 1995). In business, "exceptions are the rule;" therefore, modeling a process (project) is often not possible for creative, innovative virtual teams of knowledge workers such as in product development or consulting teams. A business process can be unstructured (ad hoc), semi-structured, or highly structured (modeled). For example, a business process such as "customer order entry" can be modeled using traditional WfMS. However, highly structured processes only can be enacted (instantiated) as they were designed. If an exception occurs, a workflow administrator needs to remodel the process before the execution can continue. This limits the usability of WfMS in a world where constant adaptation to new situations is necessary and where teams are increasingly mobile and distributed. An example of an *ad hoc* process is discussion of a project's design review using Groupware. A semi-structured process consists of groups of activities that are modelled; however, in contrast to a structured (modelled) process, it may also consist of activities that are not pre-defined. A process is semi-structured when there might be one or more activities between already modeled activities such as *assign process*, which are not known beforehand and therefore cannot be modeled in advance.

It is important to note that requirements for virtual teamwork do not follow the traditional boundaries of systems already presented. We differentiate between synchronous and asynchronous technologies for teamwork support. During our case study requirements analysis, we came to the conclusion that distributed product development in virtual communities requires a blend of synchronous and asynchronous systems support for communications, as well a basic support for asynchronous coordination of team members and their activities. In summary, the requirements for virtual teams cannot be met simply by using a combination of traditional synchronous and asynchronous systems, since the criteria for successful systems in this area differ substantially with traditional "enterprise information systems." We identified and implemented (see Section 4) four fundamental feature sets for our virtual team software—device independence; process-awareness; integration of artifacts, resources, and processes; and organizational awareness. Most systems on the market do not cater to the requirements of virtual teams; namely, *dynamic views of relationships* between artifacts, resources, and process awareness are vital to the work organization of virtual teamwork.

ON THE INTEGRATION OF ARTEFACTS, RESOURCES, AND PROCESSES

Organizations increasingly define the work activities to be fulfilled in virtual teams where team members from within the organization cooperate (i.e., communicate and coordinate work activities) with outside experts, and therefore form virtual teams, which in many cases operate as geographically dispersed teams. In fact, team members work on business processes; however, in many instances, team members view their own work as a project and not necessarily as part of a larger business process fulfilling a business goal in a larger context. The work of virtual team members often results in artifacts (e.g., documents) that need to be shared among virtual team members. The underlying assumption of this chapter is that *process-awareness* is increasingly important to virtual teams. Teamwork is a fundamental property of many business processes. Business processes have well defined inputs and outputs and serve a meaningful purpose, either within or between organizations. Business processes in general and their corresponding workflows, in particular, exist as logical models (e.g., weighted directed graphs). When business process models are executed, they have specific instances. A business process consists of a sequence of work activities. An activity is a distinct process step and may be performed either by a human agent or by a machine (software). A workflow management system enacts the real world business process for each process instance (Craven & Mahling, 1995; Dayal et al., 2001; Schal, 1996). Any activity may consist of one or more tasks. A set of tasks to be worked on by a user (human agent or machine) is called *work list*. The work list itself is managed by the WfMS. The WfMC (WfMC, 1995) calls the individual task on the work list a work item. Software systems for workflow management—Groupware—process modeling (Puustjärvi & Laine, 2001), and project management has been used to automate or to augment business processes in organizations (Casati et al., 2001; Hausleitner & Dustdar, 1999). Workflow management systems have been defined as "technology based systems that define, manage, and execute workflow processes through the execution of software whose order of execution is driven by a computer representation of the workflow process logic" (WfMC, 1995). Workflow systems generally aim at helping organizations' team members to communicate, coordinate, and collaborate effectively, as well as efficiently. Therefore, WfMS possess temporal aspects such as activity sequencing, deadlines, routing conditions, and schedules. WfMS are typically "organizationally aware" because they contain an explicit representation of organizational processes (process model). However, traditional WfMS present a rigid work environment consisting of *roles* and their associated *activities* and *applications*. In this context, they do not provide support for virtual teams such as frequently changing process participants, ad hoc formation of groups collaborating on a business process, and device independent support of group activities. Unfortunately, today's WfMS assume that each *work item* is executed by a *single* worker (Aalst & Kumar, 2001). Most WfMS focus on automating structured (modeled) intra-organizational business processes. Groupware, on the other hand, typically does not contain any knowledge or representation of the *goals* or underlying business *processes* of the group. We argue that, considering the top three problems occurring in virtual teamwork, increasing contextual information in the form of building relationships between artifacts, resources, and business processes solves the fundamental problems and, as an implication, the most dominant problems such as "difficulties in team communications" and "unclear work activities." Our approach for integration of artifacts, resources, and processes comprises a communications and coordination building block where team members exchange "enriched" messages. Workflow research has shown that modeling organizational structures has substantial benefits for business processes. Therefore, we allow modeling of organizational constructs such as groups, roles, skills, and organizational units. Each team member can be associated with those constructs, as shown in Section 4. Furthermore, an integrated database allows for attaching database objects to the communications and coordination activities of virtual team members, enabling integration of resources (organizational constructs) and artifacts. The process modeling component allows the creation of directed graphs consisting of tasks and their relationships with organizational constructs. The next section, therefore, discusses implementation issues on how to make context information (i.e., information about process instances), the team configuration (i.e., participants and their roles), their associated artifacts, and connectivity modes of group members (fixed, mobile, or ad hoc) accessible to all virtual team members.

THE CASE OF AN INTEGRATED INTERACTION MANAGEMENT SYSTEM FOR VIRTUAL TEAMS

In the following section, we will provide an overview of integration issues with which we are concerned, and design an integrated system for virtual teams called *Caramba* (Caramba Labs, 2002). An in-depth presentation of the conceptual foundations, the architecture, or the components is beyond the scope and focus of this chapter and can be found in Dustdar (2002b, 2002c, 2004). The Caramba software architecture is composed of multiple layers—middleware, client suite, and persistence store. Objects and services are accessed through the Transparent Access Layer (TAL) from the CarambaSpace platform (middleware). Depending on access mechanisms and the requested services (e.g., via Java client with RMI protocol or via Web browser with http), Caramba provides a unique way to handle requests using a metamodel framework to describe content and separate presentation, logic, and data. This model permits high flexibility and enables customization and extensions, as well as adopts new devices or technologies. The goal of this layer is to offer transparent access to a CarambaSpace. The TAL utilizes various services to transform, describe, manipulate, and observe objects. All objects managed through a CarambaSpace are described well using a metamodel description framework. Objects can be customized in their structure (e.g., adding columns to tables, adding relations to objects) and in their presentation by adopting their metamodel description. Any changes are dynamically reflected by client components. Based on the metamodel description framework Caramba enables various options to customize data and content and to integrate data from different resources (e.g., corporate databases). This layer also provides facilities for fine-grained object notification services and the implementation of customized services based on object observers. The middleware does not manage states and persistence of objects. Objects are stored, manipulated, and retrieved via the Persistence Layer (PEL). Caramba leverages and adopts standard Java-based technologies (e.g., JDBC, JNDI, HTTP, etc.) to access and integrate data.

An overall conceptual overview of how Caramba implements the requirements and how a work scenario of virtual teamwork may look is depicted in Figure 1. Virtual teams have one or more project managers and several resources (people) with various skill sets and professional backgrounds, as well as possibly different organizational affiliations. The daily teamwork consists of meetings, exchange of documents, and many communications (tasks, e-mails) being sent back and forth. For each project (business process), meetings, documents, and communications occur, and the trail of communications and interactions is depicted as lines between the team members. Without appropriate virtual team software, the relationship between artifacts, resources, and business processes is only available in the "heads" of the team members. For example, each team member has to remember *when* a particular document was sent to *whom* (e.g., a

Figure 1. Conceptual view on virtual team software support

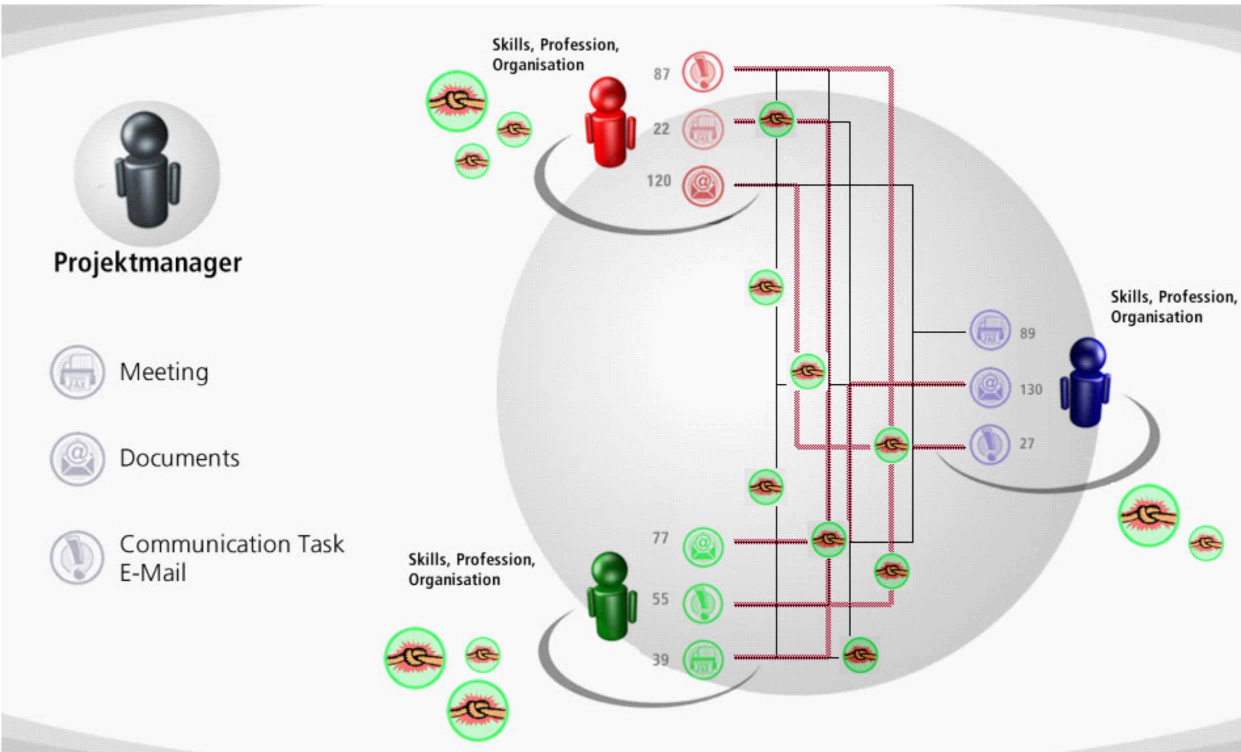
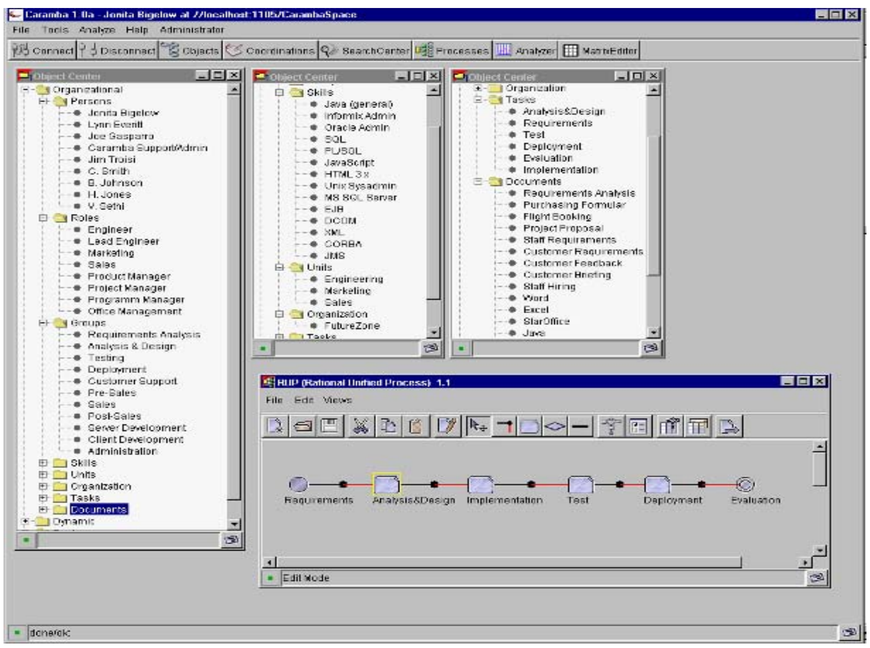


Figure 2. Modelling organizational resources and processes



customer) and *why* (i.e., as a part of a particular business process). The goal of virtual team software should be to explicitly provide this relationship information to all team members based on their views and interests.

In order to provide one example of what an implementation looks like, we present the Caramba components. The ObjectCenter component provides mechanisms to link activities with artifacts. Based on a metamodel, Caramba provides a set of organizational objects: Persons, Roles, Groups, Skills, Units, Organization, Tasks, and Documents (i.e., Templates). Utilizing these organizational constructs, an administrator can model any organizational structure, such as hierarchical, flat, or matrix. Each object class consists of attributes describing the object. The object class *Persons* contains attributes about the person, such as name, address, and so forth. The object class *Roles* allows definition of organizational roles such as "Head of IT." The object class *Groups* defines project settings such as "Product Team IT-Solutions." *Skills* enables the definition of required skill sets such as "Certified Java Developer." *Units* describes permanent departments such as "Marketing." The Object Center provides a means (by drag and drop) to link rows of object classes with each other, as depicted in Figure 2. It allows users to view relationships between who (organizational constructs) is performing which activities (Tasks) and using what (Documents). A business process modeller component enables a project manager to model a process template, which may be instantiated later using the built-in Workflow engine. Exceptions to the model are possible, without the need to remodel the process template, by choosing the communications (coordination) partner (from the ObjectCenter). The receiving end can read the appropriate message in his or her inbox.

CONCLUSION

During the last few years, virtually all business processes changed regarding their requirements for flexibility, interconnectivity, and coordination styles. Most business processes are based on teamwork. Most teams are organized as virtual teams, with team members coming from different organizations. In this chapter, we discussed the requirements of modern virtual teamwork and the problems associated with using traditional groupware, project, and workflow management systems for virtual teamwork. A fundamental need for distributed virtual teamwork is to have access to contextual information (i.e., to see a "knowledge trail" of who did what, when, how, and why). We presented the underlying conceptual issues and one implemented information system (Caramba) to support the integration of artifacts, resources, and business processes for virtual teams. Future virtual team systems should provide mechanisms for the integration of organizational models with artifacts and business processes in loosely coupled information systems. In our future work, we plan to design and implement support for definition, configuration, and composition of processes for virtual teams based on Web services. A Web service is an interface that describes a collection of operations that are network accessible through standardized XML messaging using Web servers or Application servers. A Web service is described using a standard, formal XML notion, called its *service description*. It can be published and found by other Web services. To summarize our recommendations and lessons learned, we think that for typical mid-size (e.g., 15-person) virtual teams (geographically dispersed), process-awareness, organizational awareness, and the integration of artifacts, resources, and processes is crucial. In most cases, we found that asynchronous systems support is of paramount importance when there are more team members in a virtual team and when more work occurs across different time zones.

ACKNOWLEDGEMENTS

The author thanks all team members of Caramba Labs Software AG for the fruitful and constructive discussions.

REFERENCES

Aalst, W.M.P., & Kumar, A. (2001). A reference model for team-enabled workflow management systems. *Data & Knowledge Engineering, 38*, 335-363.

Akademie für Führungskräfte (2002). Probleme bei der teamarbeit [Report]. Germany.

Bafoutsou, G., & Mentzsa, G. (2002). Review and functional classification of collaborative systems. *International Journal of Information Management, 22*, 281-305.

Bentley, R. et al. (1997). Basic support for cooperative work on the World Wide Web. *International Journal of Human-Computer Studies*, *46*, 827-846.

Bolcer, G.A. (2000, May and June). Magi: An architecture for mobile and disconnected workflow. *IEEE Internet Computing*, 46-54.

Bussler, C. (1999). Enterprise-wide workflow management. *IEEE Concurrency, 7*(3), 32-43.

Caramba Labs Software AG (2002). Retrieved January 15, 2002, from http://www.CarambaLabs.com

Casati, F. et al. (2001). Developing e-services for composing e-services. In *Proceedings CaiSE 2001*. Springer Verlag.

Craven, N. &Mahling, D.E. (1995). Goals and processes: A task basis for projects and workflows. In *Proceedings COOCS International Conference*. Milpitas, CA.

Dayal, U., Hsu M., & Ladin R. (2001). Business process coordination: State of the art, trends, and open issues. *Proceedings of the 27th VLDB Confererence*. Rome, Italy.

DeSanctis, G., & Gallupe, R.B. (1987). A foundation study of group decision support systems. *Management Science, 23*(5), 589-609.

Dustdar, S. (2002a). Mobility of context for project teams. *Proceedings of the International Workshop on Mobile Teamwork at the 22nd International Conference on Distributed Computing Systems*.

Dustdar, S. (2002b). Collaborative knowledge flow – Improving process-awareness and traceability of work activities. *4th International Conference on Practical Aspects of Knowledge Management*.

Dustdar, S. (2002c). Reconciling knowledge management and workflow management: The activity-based knowledge management approach. In H. Nemati, P. Palvia, & R. Ajami (Eds.), *Global Knowledge Management: challenges and opportunities*. Hershey, PA: Idea Group Publishing.

Dustdar, S. (2004, January). Caramba – A process-aware collaboration system supporting ad hoc and collaborative processes in virtual teams. *Distributed and Parallel Databases, 15*(1), 45-66.

Johansen, R. (1988). Groupware. Computer-support for business teams. *The Free Press*. New York.

Ellis, C.A. et al. (1995). Dynamic change within workflow systems. *Proceedings of COOCS International Conference*. Milpitas, CA.

Ellis, C.A. (2000). An evaluation framework for collaborative systems [Report]. University of Colorado at Boulder Technical Report CU-CS-9001-00.

Ellis, C.A., Gibbs, S.J., & Rein, G.L. (1991). Groupware: Some issues and experiences. *Communications of the ACM, 34*(1).

Groove (2002). http://www.groove.net

IBM (2002). http://www.ibm.com

Puustjärvi, J., & Laine, H. (2001). Supporting cooperative inter-organizational business transactions [Lecture]. *Proceedings of DEXA 2001*. Springer Verlag.

Schal, T. (1996). *Workflow management systems for process organizations*. New York: Springer.

Workflow management specification (1995). *Workflow Management Coalition*. Retrieved January 15, 2002, from http://www.wfmc.org/standards/docs/tc003v11.pdf

Workflow management specification glossary (1995). *Workflow Management Coalition (WfMC)*. Retrieved from http://www.wfmc.org

KEY TERMS

Knowledge Trail: Provides information on who did what, when, how, and why.

Process-awareness: (See knowledge trail)

Interaction Management System: An information system providing an environment for communications and coordination of work activities for virtual teams.

Workflow: Comprises cases, resources, and triggers that relate to a particular process.

Process: Indicates what tasks must be performed and in what order to complete a case.

Role: In order to perform tasks, skills are required. A role is a collection of complementary skills.

Task: An atomic process that is not divided further and is a logical unit of work.

Process-Based Data Mining

Karim K. Hirji
AGF Management Ltd, Canada

INTRODUCTION

In contrast to the Industrial Revolution, the Digital Revolution is happening much more quickly. For example, in 1946, the world's first programmable computer, the Electronic Numerical Integrator and Computer (ENIAC), stood 10 feet tall, stretched 150 feet wide, cost millions of dollars, and could execute up to 5,000 operations per second. Twenty-five years later, Intel packed 12 times ENIAC's processing power into a 12–square-millimeter chip. Today's personal computers with Pentium processors perform in excess of 400 million instructions per second. Database systems, a subfield of computer science, has also met with notable accelerated advances. A major strength of database systems is their ability to store volumes of complex, hierarchical, heterogeneous, and time-variant data and to provide rapid access to information while correctly capturing and reflecting database updates.

Together with the advances in database systems, our relationship with data has evolved from the prerelational and relational period to the data-warehouse period. Today, we are in the knowledge-discovery and data-mining (KDDM) period where the emphasis is not so much on identifying ways to store data or on consolidating and aggregating data to provide a single, unified perspective. Rather, the emphasis of KDDM is on sifting through large volumes of historical data for new and valuable information that will lead to competitive advantage. The evolution to KDDM is natural since our capabilities to produce, collect, and store information have grown exponentially. Debit cards, electronic banking, e-commerce transactions, the widespread introduction of bar codes for commercial products, and advances in both mobile technology and remote sensing data-capture devices have all contributed to the mountains of data stored in business, government, and academic databases. Traditional analytical techniques, especially standard query and reporting and online analytical processing, are ineffective in situations involving large amounts of data and where the exact nature of information one wishes to extract is uncertain.

Data mining has thus emerged as a class of analytical techniques that go beyond statistics and that aim at examining large quantities of data; data mining is clearly relevant for the current KDDM period. According to Hirji (2001), data mining is the analysis and nontrivial extraction of data from databases for the purpose of discovering new and valuable information, in the form of patterns and rules, from relationships between data elements. Data mining is receiving widespread attention in the academic and public press literature (Berry & Linoff, 2000; Fayyad, Piatetsky-Shapiro, & Smyth, 1996; Kohavi, Rothleder, & Simoudis, 2002; Newton, Kendziorski, Richmond, & Blattner, 2001; Venter, Adams, & Myers, 2001; Zhang, Wang, Ravindranathan, & Miles, 2002), and case studies and anecdotal evidence to date suggest that organizations are increasingly investigating the potential of data-mining technology to deliver competitive advantage.

As a multidisciplinary field, data mining draws from many diverse areas such as artificial intelligence, database theory, data visualization, marketing, mathematics, operations research, pattern recognition, and statistics. Research into data mining has thus far focused on developing new algorithms and tools (Dehaspe & Toivonen, 1999; Deutsch, 2003; Jiang, Pei, & Zhang, 2003; Lee, Stolfo, & Mok, 2000; Washio & Motoda, 2003) and on identifying future application areas (Alizadeh et al., 2000; Li, Li, Zhu, & Ogihara, 2002; Page & Craven, 2003; Spangler, May, & Vargas, 1999). As a relatively new field of study, it is not surprising that data-mining research is not equally well developed in all areas. To date, no theory-based process model of data mining has emerged. The lack of a formal process model to guide the data-mining effort as well as identification of relevant factors that contribute to effectiveness is becoming more critical as data-mining interest and deployment intensifies. The emphasis of this article is to present a process for executing data-mining projects.

BACKGROUND

The fields of machine learning, pattern recognition, and statistics have formed the basis for much of the developments in data-mining algorithms. The field of statistics is one of the oldest disciplines concerned with automatically finding structure in examples. Discriminant analysis (Fisher, 1936), for example, is the oldest mathematical classification technique used to separate data into classes by generating lines, planes, or hyperplanes. Through the pioneering work on classification and regression trees (CART) by Breiman, Friedman, Olshen, and Stone (1984), the statistical community has made an important contribution in legitimizing the use of decision trees, in data mining, for classification and regression. Pattern-recognition research emphasizes the creation of machines that can perform tasks more accurately, faster, and cheaper than humans (Fukunaga, 1972; Ripley, 1993), and has made an important contribution to data mining by popularizing the use of neural networks. A feed-forward neural network is a network in which the nodes (or processing units) are numbered so that all connections go from a node to one with a higher number. In practice, the nodes are arranged in layers with connections only to higher layers. Back propagation is an implementation for a feed-forward neural network in which error terms, from the output layer, are propagated back to the input layer so that the resulting connection weights at each node adjusted can be adjusted by means of an error-minimization method called gradient descent.

The multitude of data-mining algorithms can be linked to three main data-mining-problem approaches: clustering, association and sequential pattern discovery, and predictive modeling. Clustering (or segmentation) is concerned with partitioning data records into subsets. The *K*-means clustering algorithm is used for demographic clustering because categorical data are predominant. This algorithm, which is efficient for large databases, clusters a data set by determining the cluster to which a record fits best. Once clusters have been found in a data set, they can be used to classify new data. To uncover affinities among transaction records consisting of several variables, association algorithms are used. These algorithms are used to solve problems where it is important to understand the extent to which the presence of some variables implies the existence of other variables and the prevalence of this particular pattern across all data records. Sequential-pattern-discovery algorithms are related to association algorithms except that the related items are spread over time. Finally, the predictive-modeling data-mining-problem approach involves the use of a number of algorithms (e.g., binary decision tree, linear discriminant function analysis, radial basis function, back-propagation neural network, logistic regression, and standard linear regression) to classify data into one of several predefined categorical classes or to use selected fields from historical data to predict target fields.

The initial implementation of data-mining applications has been in the banking, consumer marketing, insurance, and telecommunications industries. Credit scoring, direct-mail target marketing, policy-holder risk assessment, and call graph analysis are but a few of the "killer" applications of data mining in these respective industries. As a result of some of the realized benefits of data mining, new applications are emerging in a number of areas including biomedicine where molecular data are combined with clinical medical data to achieve a deeper understanding of the causes for and treatment of disease, national security where unusual patterns and fraudulent behavior play a role in identifying and tracking activities that undermine security, pharmaceuticals where interest in understanding the 3D substructure of a molecule and how it interacts with the target is a crucial step in the design of new drug molecules, and ecology where large amounts of climate data, terrestrial observations, and ecosystem models offer an unprecedented opportunity for predicting and possibly preventing future ecological problems. Although the frontiers of data-mining applications continue to expand, focus on developing a data-mining process has not met with similar enthusiasm.

DATA-MINING PROCESS OVERVIEW

New product development (NPD) is a well-researched area (e.g., Hauptman & Hirji, 1999) and, thus, it is the foundation for the data-mining process model because NPD projects, by their very nature, are the most complex as they include systems, subsystems, components, and modules, as well as physical product and software aspects. Focusing on the NPD literature and synthesizing the elements of the various process models allows for the development of an information-centric process model for performing data-mining projects. Table 1 provides a baseline of what an inclusive process for performing data-mining projects might look like.

The phases in the baseline data-mining process include Phase 0, Phase 1, Phase 2, and Phase 3. Phase 0 is the *discovery* phase that supports the subsequent three phases. The set of proposed activities in this phase include (a) assessment of the organization's orientation toward data-centricity, (b) assessment of the capability of the organization to apply a portfolio of analytical techniques, and (c) strategy development for the use of analytics throughout the department or organization. Phase 1 is the *entry* phase. The underlying intent of this phase is to define the candidate business problem that is solvable and that can at least partially use existing data resident in the organization's databases. Prospecting and domain analysis, business problem generation and preliminary assessment, and data sensing are the proposed set of activities in this phase. Data sensing in particular is concerned with the representational faithfulness of the data set in question. Phase 2 is the *launch* phase. In this phase the data-mining project becomes a formal project with an associated capital and operational budget. The set of proposed activities in this phase include (a) secure project sponsorship, (b) project planning and project-team resourcing, (c) business problem refinement and validation of business assumptions and constraints, (d) development of the data strategy (i.e., explore the possible need for purchasing data from a third-party provider), and (e) formulation of the data-mining approach. Phase 3, the final phase, is the *execution and infusion* phase. The actual execution of data-mining algorithms takes place here as well as results analysis. The proposed set of activities in this final phase are (a) detailed business-problem definition, (b) data sourcing and enrichment, (c) execution of data-mining algorithms, (d) results interpretation, synthesis, and validation, and (e) information harvesting and business-strategy formulation. This final activity focuses on developing a strategy to tactically leverage the new insight and to communicate this on a department-, division-, or perhaps enterprise-wide basis.

APPLICATION IN BUSINESS

The 1990s are referred to as the "heyday" of the mutual-fund industry in North America. Falling yields from fixed-rate investments, new product creation, and the hope of higher returns drove investors to purchase mutual funds. It is not surprising that mutual-fund companies therefore experienced double-digit growth rates in the range of 25% to 65%. The picture today is vastly different as industry consolidation, a mature business cycle, and flat market growth are causing gross sales to increase at a decreasing rate. Retail investors are now more than ever before scrutinizing management expense ratios (MERs) and fund performance reporting and returns. Competition among the various fund companies is fierce and thus there is a renewed focus on both stealing market share and developing a compelling story about future sales growth rates across multiple product lines. Some of the tactics employed to achieve various business goals are product innovation and quality, superior fund performance, and exceptional sales, marketing, and client service.

The experiences to date of the application of the proposed data-mining process to an analytics project in the mutual-fund industry suggests that the model is playing a role in contributing to a successful outcome. The project is in the early aspects of Phase 3 and therefore because it is ongoing and no actual positive net present value has been realized, conclusions cannot be made at this time. However, some of the key accomplishments that can be attributed to the process model are as follow.

- Project has been formally approved as a result of the capital-budgeting process
- Project team and stakeholders are in place
- Business problem related to effective channel management is understood and defined
- Business entity and data models have been developed
- Strategies to close data gaps and augment data where necessary are in place
- Data-mining approach with an emphasis on clustering exists

Table 1. A summary of a data–mining process

Phase Name	Phase Identifier	Key Activity
Discovery	0	• Assess data centricity • Assess analytics capability • Develop analytics strategy
Entry	1	• Prospecting and domain analysis • Problem generation • Problem assessment • Data sensing
Launch	2	• Project sponsorship • Project planning and core project team • Problem refinement • Problem validation • Data strategy • Data-mining approach
Execution & Infusion	3	• Business-problem definition • Data sourcing and enrichment • Run data-mining algorithms • Results interpretation, synthesis, and validation • Information harvesting • Business-strategy formulation

FUTURE TRENDS

The need for near-real-time information to support business decision making is becoming more and more critical. Data are benign and therefore there is more focus on understanding the "story" behind the data and the relationships among them. From the vantage point of practitioners and managers, there appear to be new patterns emerging specifically in three areas. The software and enterprise-wide-package-implementation industry is experiencing major consolidation, and software companies are realigning their sales efforts to focus on the ever-growing mid-market segment. Once this shakedown is complete, many business intelligence software and solution providers will begin to bundle data-mining solutions with their existing offerings as a response to the already-established demand by customers in this market segment.

The data-mining process today is more akin to craft than interdisciplinary team-based problem solving. Once data mining gains acceptance and has a critical mass in organizations, there will undoubtedly be a shift to using this technology as part of a repertoire of tools to assist planners, product developers, managers, and others to develop strategies and tactical implementation plans for the organization. In this respect, improvement to existing tools in the areas of human interaction and visual presentation of information are expected. Finally, with the realities of virtual teams now solidified, the question of how virtual teams can effectively execute data-mining implementation projects will become relevant.

CONCLUSION

The paramount objective of publicly traded organizations is to focus on creating and maximizing shareholder wealth. However, recently well-publicized corporate shenanigans have shown that single-mindedly pursuing largesse at the expense of corporate social responsibility is not only fundamentally wrong, but also something that cannot be sustained indefinitely. Managers, C-suite executives, and corporate directors are now once again facing increasing scrutiny about revenue, profit, and earnings quality; adoption of accounting rules; and timely disclosures regarding the going concern of business entities. As managers craft, implement, and execute various growth, talent management, cost reduction, and competitive differentiation strategies, information technology will without a doubt continue to play an important role in enabling, supporting, leading, and transforming business, government, and not-for-profit organizations. Data mining as a technology is not an end, but rather a means to an end. Through the use of a disciplined and structured process, data-mining benefits can be obtained by the operationalization of data-mining results, via a business strategy, to achieve specific business-unit and enterprise-wide objectives.

REFERENCES

Alizadeh, A. A., et al. (2000). Distinct types of diffused large B-cell lymphoma identified by gene expression profiling. *Nature, 403*, 503-511.

Berry, M., & Linoff, G. (2000). *Mastering data mining*. New York: John Wiley & Sons, Inc.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. CA: Wadsworth & Brooks.

Dehaspe, L., & Toivonen, H. (1999). Discovery of frequent datalog patterns. *Data Mining and Knowledge Discovery, 3*(1), 7-36.

Deutsch, J. M. (2003). Evolutionary algorithms for finding optimal gene sets in microarray prediction. *Bioinformatics, 19*, 45-54.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM, 39*(11), 27-34.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics, 7*, 179-188.

Fukunaga, K. (1972). *Introduction to statistical pattern recognition*. Boston: Academic Press.

Hauptman, O., & Hirji, K. K. (1999). Managing integration and coordination in cross-functional teams: An international study of concurrent engineering product development. *R&D Management, 29*(2), 179-192.

Hirji, K. K. (2001). Exploring data mining implementation. *Communications of the ACM, 44*(7), 87-93.

Jiang, D., Pei, J., & Zhang, A. (2003). Toward interactive exploration of gene expression patterns. *SIGKDD Explorations, 5*(2), 79-90.

Kohavi, R., Rothleder, N. J., & Simoudis, E. (2002). Emerging trends in business analytics. *Communications of the ACM, 45*(8), 45-48.

Lee, W., Stolfo, S., & Mok, K. (2000). Adaptive intrusion detection. *Artificial Intelligence Review, 14*, 533-567.

Li, T., Li, Q., Zhu, S., & Ogihara, M. (2002). A survey of wavelet applications in data mining. *SIGKDD Explorations*, *4*(2), 49-68.

Newton, M., Kendziorski, C., Richmond, C., & Blattner, F. (2001). On differential variability of expression ratios: Improving statistical inference about gene expression changes from microarray data. *Journal of Computational Biology*, *8*, 37-52.

Page, D., & Craven, M. (2003). Biological applications of multi-relational data mining. *SIGKDD Explorations*, *5*(1), 69-79.

Ripley, B. D. (1993). Statistical aspects of neural networks. In O. E. Barndorff-Nielsen et al. (Eds.), *Networks and chaos: Statistical and probability aspects*. London: Chapman & Hall.

Spangler, W., May, J., & Vargas, L. (1999). Choosing data mining methods for multiple classification: Representational and performance measurement implications for decision support. *Journal of Management Information Systems*, *16*(1), 37-62.

Venter, J., Adams, M., & Myers, E. (2001). The sequence of the human genome. *Science, 291*, 1304-1351.

Washio, T., & Motoda, H. (2003). State of the art of graph-based data mining. *SIGKDD Explorations, 5*(1), 59-68.

Zhang, L., Wang, L., Ravindranathan, A., & Miles, M. (2002). A new algorithm for analysis of oligonucleotide arrays: Application to expression profiling in mouse brain regions. *Journal of Molecular Biology, 317*, 227-235.

KEY TERMS

Classification Trees: Type of decision tree that is used to predict categorical variables, whereas regression trees are decision trees used to predict continuous variables.

Cluster: Subset of data records; the goal of clustering is to partition a database into clusters of similar records such that records that share a number of properties are considered to be homogeneous.

Data Mart: Scaled-down version of an enterprise-wide data warehouse that is created for the purpose of supporting the analytical requirements of a specific business segment or department.

Data Mining: Analysis and nontrivial extraction of data from databases for the purpose of discovering new and valuable information, in the form of patterns and rules, from relationships between data elements.

Data Warehouse: A platform consisting of a repository of selected information drawn from remote databases or other information sources, which forms the infrastructural basis for supporting business decision making.

Information: Interpreted symbols and symbol structures that reduce both uncertainty and equivocality over a defined period of time.

Knowledge: Information combined with experience, context, interpretation, and reflection.

Operational Data Store: An integrated repository of transaction-processing systems that uses data-warehouse concepts to provide "clean" data in support of day-to-day operations of a business.

Programmed Instruction, Programmed Branching, and Learning Outcomes

Robert S. Owen
Texas A&M University-Texarkana, USA

Bosede Aworuwa
Texas A&M University-Texarkana, USA

INTRODUCTION

This chapter discusses the principles of two qualitatively different and somewhat competing instructional designs from the 1950s and '60s, *linear programmed instruction* and *programmed branching*. Our hope is that an understanding of these ideas could have a positive influence on current and future instructional designers who might adapt these techniques to new technologies and want to use these techniques effectively. Although these older ideas do still see occasional study and debate (e.g., Dihoff, Brosvic & Epstein, 2003, 2004), many current instructional designers are probably unaware of the learning principles associated with these (cf., Fernald & Jordan, 1991; Kritch & Bostow, 1998).

BACKGROUND

An important difference between these instructional designs is associated with the use of feedback to the learner. Although we could provide a student with a score after completing an online multiple-choice quiz, applications that provide more *immediate feedback* about correctness upon completion of each individual question might be better. Alternatively, we could provide *adaptive feedback* in which the application provides elaboration based upon qualities of a particular answer choice.

Below is a discussion of two qualitatively different instructional designs, one providing immediate feedback regarding the correctness of a student's answer, the other providing adaptive feedback based on the qualities of the student's answer. Suitability of one design or the other is a function of the type of learner and of the learning outcomes that are desired.

MAIN THRUST OF THE ARTICLE

Although the idea of non-human feedback would seem to imply a mechanical or electronic device, other methods could be used. Epstein and his colleagues, for example, have used a multiple-choice form with an opaque, waxy coating that covers the answer spaces in a series of studies (e.g., Epstein, Brosvic, Costner, Dihoff & Lazarus, 2003); when the learner scratches the opaque coating to select an answer choice, the presence of a star (or not) immediately reveals the correctness of an answer. Examples of the designs discussed below are based on paper books, but they are easily adaptable to technologies that use hyperlinks, drop-down menus, form buttons, and such.

Linear Programmed Instruction

The programmed psychology textbook of Holland and Skinner (1961) asked the student a question on one page (the quote below starts on page 2) and then asked the student to turn the page to find the answer and a new question:

"A doctor taps your knee (patellar tendon)
with a rubber hammer to test your __________."

The student thinks (or writes) the answer and turns the page to find the correct answer ("reflexes") and is then asked another question.

Questions or statements are arranged in sequentially ordered *frames* such as the single frame above. A frame is completed when the student provides a response to a stimulus and receives feedback. Skinner contended that this method caused learning through *operant conditioning*, provided through positive *reinforcement* for stimuli that are designed to elicit a correct answer (c.f., Cook, 1961; Skinner, 1954, 1958).

Skinner (and others who use his methods) referred to his method as *programmed instruction*, which incorporates at least the following principles (cf., Fernald & Jordan, 1991; Hedlund, 1967; Holland & Skinner, 1958, 1961; Whitlock, 1967):

- Clear learning objectives.
- Small steps; frames of information repeat the cycle of stimulus-response-reinforcement.
- Logical ordered sequence of frames.
- Active responding by a student who works at his/her own pace.
- Immediate feedback to the response in each frame with positive reinforcement for correct answers.

A technique in programmed instruction is to help the student a great deal at first, and then gradually reduce the cues in latter frames; this is called *fading* (Fernald & Jordan, 1991; Reiff, 1980). If correct responding suggests that a student is learning at a quick rate, *gating* can be used to skip over frames that repeat prior information (Vargus & Vargus, 1991). The programmer is expected to use information about student performance to make revisions; if the student is not succeeding, then it is due to a fault of the program, not to an inability of the student (Holland & Skinner, 1961; Vargus & Vargus, 1991).

Programmed Branching

Crowder (e.g., 1959, 1963) and others (e.g., Pressey, 1963) were critical of Skinner's approach, arguing that students not only learn from knowing a correct answer, but also learn by making mistakes. Crowder distinguished between his *automatic tutoring device* and the Skinner-type *teaching machine*, proposing that the automatic tutoring device is more flexible in allowing the student to receive an explanation when an error is made. Crowder (1959, pp. 110-111) provides an example of how this approach could be used in a programmed textbook:

> "In the multiplication of 3 x 4 = 12, the number 12 is called the *product* and the numbers 3 and 4 are called the
>
> Page 15 quotients.
> Page 29 factors.
> Page 43 powers."

In this *programmed branching* method of Crowder, the student is taken to one of several possible discussions depending on the qualities of the answer.

While Skinner's design would be expected to work only when stimuli elicit correct answers, Crowder's design allows for mistakes and must be designed to anticipate particular mistakes. Crowder believed that this method caused learning through *cognitive reasoning*. Whatever answer is chosen by the student, the programmed textbook (or machine) makes a *branch* to a discussion associated with issues relevant to the answer that was chosen. This is followed by a return to the same question if the student had made an incorrect choice, or a jump to new a *frame* containing the next question if the student had made a correct choice.

Learning Outcomes

Many issues have been raised about programmed instruction methods. Reiff (1980) discusses several criticisms:

- It does not take into consideration the sequence of development and readiness to learn (e.g., children of different ages or children vs. adults).
- It develops rote learning skills rather than critical thinking skills.
- Students can in some implementations cheat.
- The encouragement to respond quickly could develop bad reading habits.

Crowder's *programmed branching* design, which has received far less attention and study than Skinner's ideas, would seem to answer at least some of these criticisms. Crowder's design provides an explanation to both correct and incorrect answers, so the learner is not rewarded for cheating or working too quickly. Since the explanation is tied to the learner's thinking at the time a choice was made, Crowder's design would appear to be better to develop critical thinking skills, but might not be so good at developing rote learning skills. Crowder's design would appear to be better suited to students who have a greater readiness to learn, while perhaps not so well suited to a student who is at an earlier stage of learning a subject.

The above discussion suggests that each of these designs is useful, but that each is useful in different kinds of situations and that the *learning outcomes* of each approach might be different. Skinner's teaching machine, for example, might be more useful in situations where students are learning lists and definitions. The automatic tutoring device, on the other hand, might be more useful when the student is already at a higher level of understanding, whereby s/he can now use reasoning to derive an answer, or in situations where the student understands that there are degrees of right and wrong without concrete answers. The Skinner-type teaching machine might be better suited to "lower-order" levels of learning, while the Crowder-type automatic tutoring device might be better suited to "higher-order" levels of learning.

Although many ideas have been proposed with regard to a hierarchical perspective on "lower" and "higher" levels of learning, the most well-known, "Bloom's Taxonomy" (A Committee of College and University Examiners, 1956), originated in about the same timeframe as the ideas of Skinner and Crowder. "Bloom's Taxonomy" proposes that the objectives of learning lie on a hierarchical continuum: (1) knowledge of terminology and facts, (2) comprehension of translation and paraphrasing, (3) application, (4) analysis, (5) synthesis, and (6) evaluation.

"Bloom's Taxonomy" is actually only Part I of a two-part work. The first part above is known as the *cognitive domain*. Part II (Krathwohl, Bloom & Masia, 1964) focuses on the *affective domain*: (1) willingness to receive ideas, (2) commitment to a subject or idea, (3) feeling that an idea has worth, (4) seeing interrelationships among multiple ideas, and (5) the integration of ideas as one's own.

FUTURE TRENDS

Fernald and Jordan (1991) discuss several reasons as to why programmed instruction might have fallen out of use in more recent decades:

- It was seen to dehumanize the teaching process.
- Educators feared that it might be too effective and threaten their jobs.
- The importance of the learning principles weren't understood.
- Applications were often not effectively designed.

Technology, economics, and attitudes have since changed. As economics and student demand push us to use distance education methods, the first two arguments would seem to become more diminished in the future.

It is hoped that this chapter assists in diminishing the latter two arguments by introducing instructional designers to the principles discussed in this chapter and by encouraging instructional designers to create more effective designs with regard to appropriateness for a particular student audience, and with regard to the type and level of learning outcomes that are desired. By better understanding the past, we can better affect the future.

Curiously, there has been less attention devoted to Crowder's ideas of adaptive feedback than to Skinner's ideas of immediate feedback and reinforcement. We continue to see occasional research devoted to related issues, such as issues of immediate vs. delayed feedback (e.g., Dihoff et al., 2003, 2004; Kelly & Crosbie, 1997) or of allowing students to keep selecting answers from a multiple-choice set until the correct answer is finally discovered (Epstein et al., 2003). However, we still can only speculate with regard to conditions under which a Skinner-style of instructional design would be better and when a Crowder-style of design would be better. It is hoped that this chapter generates greater awareness of and use of these designs in new technologies, but also that greater interest in these ideas will stimulate more research into the learning mechanisms associated with them.

CONCLUSION

New technologies such as Web browsers now make it relatively easy for educators with the most modest of skills to present instructional frames in a linear sequential ordering or as branches that are dependent on the student's selection of answers from a list. In adapting some of these older ideas to newer technologies, we hope that instructional designers will be better equipped to select appropriate methods by considering:

- the student's level of readiness for learning,
- the basis for learning when different instructional designs are used, and
- the qualitatively different kinds of learning outcomes that are possible with different instructional designs.

REFERENCES

A Committee of College and University Examiners. (1956). *Taxonomy of educational objectives, the classification of educational goals, Handbook I: Cognitive domain*. New York: David McKay Company.

Cook, D.L. (1961). Teaching machine terms: A glossary. *Audiovisual Instruction*, *6*, 152-153.

Crowder, N.A. (1959). Automatic tutoring by means of intrinsic programming. In E. Glanter (Ed.), *Automatic teaching, the state of the art* (pp. 109-116). New York: John Wiley & Sons.

Crowder, N.A. (1963). On the differences between linear and intrinsic programming. *Phi Delta Kappan* (March). Reprinted in J.P. DeCecco (Ed.), *Educational technology: Readings in programmed instruction* (1964) (pp. 142-152). New York: Holt, Rinehart, and Wilson.

Dihoff, R.E., Brosvic, G.M. & Epstein, M.L. (2003). The role of feedback during academic testing: The delay retention effect revisited. *The Psychological Record*, *53*(4), 533-548.

Dihoff, R.E., Brosvic, G.M., Epstein, M.L. & Cook, M.J. (2004). Provision of feedback during preparation for academic testing: Learning is enhanced by immediate but not delayed feedback. *The Psychological Record*, *54*(2), 207-231.

Epstein, M..L., Brosvic, G.M., Costner, K.L., Dihoff, R.E. & Lazarus, A.D. (2003). Effectiveness of feedback during the testing of preschool children, elementary school children, and adolescents with developmental delays. *The Psychological Record, 53*(2), 177-195.

Fernald, P.S. & Jordan, E.A. (1991). Programmed instruction versus standard text in introductory psychology. *Teaching of Psychology, 18*(4), 205-211.

Hedlund, D.E. (1967). Programmed instruction: Guidelines for evaluation of published materials. *Training and Development Journal*, (February), 9-14.

Holland, J.G. & Skinner, B.F. (1961). *The analysis of behavior*. New York: McGraw-Hill.

Kelly, G. & Crosbie, J. (1997). Immediate and delayed effects of imposed feedback delays in computerized programmed instruction. *The Psychological Record*, *47*(4), 687-698.

Kritch, K.M. & Bostow, D.E. (1998) . Degree of constructed-response interaction in computer-based programmed instruction. *Journal of Applied Behavior Analysis*, *31*(3), 387-398.

Krathwohl, D.R., Bloom, B.S. & Masia, B. (1964). *Taxonomy of educational objectives, the classification of educational goals, handbook II: The affective domain*. New York: David McKay Company.

Pressey, S.L. (1963). Teaching machine (and learning theory) crisis. *Journal of Applied Psychology*, *47*, 1-6.

Reiff, J.C. (1980). Individualized learning through programmed materials. *Education*, *100*(3), 269-271.

Skinner, B.F. (1954). The science of learning and the art of teaching. *Harvard Educational Review*, *24*(2), 86-97.

Skinner, B.F. (1958). Teaching machines. *Science, 128*(October 24), 969-977.

Vargus, E.A. & Vargus, J.S. (1991). Programmed instruction: What it is and how to do it. *Journal of Behavioral Education*, *1*(2), 235-251.

Whitlock, G.H. (1967). Programmed learning: Some nonconfirming results. *Training and Development Journal*, *21*(6), 11-13.

KEY TERMS

Adaptive Feedback: Immediate feedback in the form of an explanation or discussion that is tailored to the qualities of the student's answer.

Automatic Tutoring Device: A device that uses programmed branching and adaptive feedback. Learning results from cognitive reasoning.

Cognitive Reasoning: Learning through the process of thinking about an issue; the student learns new ideas and relationships by relating an issue to previously learned material.

Frame: A small piece of information or a statement to which the student is exposed, such as a page with a single question. In linear programmed instruction, a frame includes a stimulus, a response, and reinforcement (positive feedback).

Hierarchy of Learning: The concept that learning can be sequentially ordered along a continuum from lower order to higher order. "Bloom's Taxonomy" is one of many that have been proposed.

Linear Programmed Instruction: A design whereby a series of frames are presented to the student in a specific sequential order. The student actively responds to stimuli in each frame and receives immediate feedback to that response. Learning results through operant conditioning.

Operant Conditioning: Learning through immediate positive feedback (reinforcement) regarding the correctness of an answer; the student learns to respond in a particular way to a particular question or issue (stimulus). **Fading** can be used by gradually reducing stimulus cues in subsequent frames when material is repeated.

Programmed Branching: A method whereby the student is taken to one of several possible explanations or discussions depending on the qualities of an answer that is given to a question. **Gating** is a simple skip of frames which repeat prior information when a student's answers suggest that the material has been adequately learned.

Teaching Machine: A device that uses linear programmed instruction, whereby frames present a question followed by feedback of the correct answer. Learning results from reinforcement of the student's correct answer.

Program Execution and Visualization on the Web

Cristóbal Pareja-Flores
Universidad Complutense de Madrid, Spain

J. Ángel Velázquez-Iturbide
Universidad Rey Juan Carlos, Spain

INTRODUCTION

Programming is a central activity in the computing profession. It is facilitated by different tools (editors, compilers, debuggers, etc), which are often integrated into programming environments. Programming also plays a central role in computer science education. For this purpose, a number of complementary tools were developed during the last decade: algorithm animators, program visualizers, problem generators, assignment graders, and so forth.

After the Web explosion, teachers of programming rapidly turned their attention to the Web. Although the Web has speed and power limitations, it also has several advantages that make it invaluable for educational purposes. Mainly, it provides universal accessibility and platform independence, and solves the distribution problem by always making available the last version of any tool.

This article describes Web-based tools for program execution and visualization (Pareja-Flores & Velázquez-Iturbide, 2002) in two sections. Each section gives a brief overview of their evolution, describes educational uses, and, in the case of visualization, includes some definitions and reports of lessons learned. Finally, we outline our view of future trends in the use of the Web for programming education and our personal conclusions.

BACKGROUND

The simplest use of the Web for programming education is as a public repository of high quality problems. Many collections have no structure or, at best, are lineally or hierarchically structured. We note several initiatives hosted by the ACM: the *Lab Repository* (Knox, 2002), *Computer Science Education Links* (McCauley, 2001), and the *ACM International Collegiate Programming Contest* (Skiena & Revilla, 2003). Other useful resources are also collected on Web sites, such as slides and audio lectures (Skiena & Revilla, 2003), algorithm animations (Brummond, 2001; Crescenzi et al., 2003), or programming tools (English, 2001). More advanced repositories provide a management system that, using (semi)structured mark-up languages, allows retrieving, maintaining and publishing. Good representatives are eXercita (Gregorio-Rodríguez et al., 2001, 2002) and SAIL (Kovourov, Tamassia, Bridgeman & Goodrich, 2000).

A more complex initiative consists in porting programming tools to be executed on the Web. We focus on two kinds of tools. The first kind is aimed at supporting program execution. Programming is a task that involves several activities: editing, compiling and testing, at least. These activities require running a number of applications: an editor, a compiler and the program itself. In addition to their conventional use in standalone computers, they can be used on the Web in a number of ways, increasing flexibility from an educational point of view.

A second kind of tools supports software visualization. This field studies the visual representation of software entities (Stasko, Domingue, Brown, & Price, 1998). Visualization requires an effort to abstract the target entity to visualize and to make a good graphical design that may yield many different representations: text vs. graphics, level of abstraction, static vs. dynamic visualizations, one or multiple views, behavior vs. performance, errors, and so forth. Algorithm animation is a subfield of software visualization aimed at the dynamic visualization a piece of software illustrating the main ideas or steps (i.e., its algorithmic behavior), but without a close relationship to the source code. The graphical nature of software visualization in general and algorithm animation in particular makes them very conducive to the hypermedia features of the Web.

Web-Based Program Execution

The simplest use of the Web for programming execution is as a medium to submit programs. A more advanced capability is the support of Web-based program edition, compilation and execution. This can be implemented in different ways, depending on how the client/server load is balanced. For instance, the system by Elenbogen, Maxim and McDonald (2000) includes a set of interactive Web exercises on C++ delivered by Java applets. A different option consists in using the Web as the medium to transfer programs and data to the server, which is then responsible for compiling and executing programs. In this approach, programs may be edited by the client and then remotely compiled and executed on a server (Hiltz & Kögeler, 1997). The server may also work as a black-box tester based on input-output pairs (Arnow & Barshay, 1999; Skiena & Revilla, 2003).

There are few systems with full programming capabilities on the server side, including debugging. Ibrahim (1994) developed a system that allowed the programmer to use the Web as a front-end to edit a program, send it to the Web server, and debug it by performing tracing actions on the program running at the server. Finally, other systems give support to the graphical visualization of execution (Domingue & Mulholland, 1997) or allow the user to experiment by controlling which parts must be executed and how (Berghammer & Milanese, 2001).

Algorithm Animation

Algorithm animation is a research field which is now 20 years old and still evolving. There is a consensus with respect to the videotape Sorting Out Sorting presented in 1981 by Baecker (1998) a landmark on animation, which included animations of nine sorting algorithms. Afterwards, some works established the main techniques for specifying and implementing algorithm animation: BALSA (Brown, 1988), Tango (Stasko, 1990), and Pavane (Roman, Cox, Wilcox & Plun, 1992). Systematic categorizations of software visualizations have been proposed since then (Price, Baecker & Small, 1998).

In the mid-nineties, many of the existing systems were ported to the Web, and many additional systems were specifically designed for the Web. A representative work from these years is that of Naps (1996), in which he carried out a study of the technical alternatives that could be used to make animations produced by previous systems available on the Web. Other systems are JCAT (Brown & Raisamo, 1997), Mocha (Baker, Cruz, Liotta & Tamassia, 1995), and JHAVÉ (Naps, Eagan & Norton, 2000). A second group from these years is formed by the systems that were not specifically designed for the Web, but based on stand-alone multimedia and hypermedia. Although they could not be used directly on the Web, they and Web-based systems had similar hypermedia features. A good example is the HalVis system (Hansen, Schrimpsher & Narayanan, 1999).

Another category of systems automatically produce program animations, enhanced with graphical representations. These can be considered as extensions of programming environments with advanced animation capabilities which are usable on the Web. Three examples are ISVL for Prolog (Domingue & Mulholland, 1997), KIEL for ML (Berghammer & Milanese, 2001), Jeliot for Java programming (Haajanen et al. 1997) and WinHIPE for functional programming (Naharro-Berrocal, Pareja-Flores, Velázquez-Iturbide & Martínez-Santamarta, 2001).

EDUCATIONAL USES OF WEB-BASED PROGRAM EXECUTION AND VISUALIZATION

Program Execution

Currently, the most common use of the Web for programming courses is as a communication medium, facilitating submission and administration of assignments and grades (Burd, 2000). As we have described previously, it may also be used as a medium to retrieve exercises and to send, run and even debug programs at the server.

Web-based program execution is also being used in more specific ways within courses. First, if the Web interface of the programming tool is carefully designed, only one aspect can be considered at a time (the condition of a conditional statement, the body of a loop, etc.), allowing novice students to concentrate on program fragments that illustrate certain syntactic or semantic elements. This feature is especially important at the beginning, when novices ignore program structure and other details. Second, program execution on the Web can be used as a testing tool. For instance, the teacher can use the system for student inquiry by proposing that the student predict the behavior expected for a given input, or vice versa, by guessing the input that yields a desired output (Elenbogen et al., 2000). Finally, by using the Web to drive visualizations and animations (as explained in the next section), information can be given graphically to the programmer about errors, about the progress of the computation, and so on.

Algorithm Animation

From the outset, the main use of algorithm animation was educational, rather that industrial (for instance, as a debugging tool). Algorithm animation systems have been used in several ways: as a complement to lectures on algorithms, for self-study, or within laboratories. A more demanding use of animation systems consists in requiring students to build their own animations.

The best documented experience ran for about 20 years at the Computer Science Department of Brown University (Bazik, Tamassia, Reiss & van Dam, 1998). All of the experiences reported in the available literature agree that students are highly motivated by algorithm animations. After various experiments and studies, the community has come to the conclusion that animations are effective only if students are actively engaged (Naps et al., 2003).

Lessons Learned from Using Animation Systems

It is difficult to provide canned recipes for the design of visualizations and animations. Gloor (1998) and Khuri (2001) summarize some general recommendations for their design. These are also commonly accepted suggestions for educational use:

- Make the meaning of the different graphical representations explicit, explained either by embedding them in the system or reinforced by allocating time to the subject during the course.
- Adapt to the knowledge level of the user, either novice or expert.
- Complement animations with explanations.
- Be interactive, allowing flexible control of animations (including movement in a backwards direction), customization, input data, or unfolding unexpected short questions.
- Provide multiple views, which must be consistently coordinated.
- Include performance information, for example, data collected during the execution of an algorithm, or animating several algorithms simultaneously.
- Include execution history.

FUTURE TRENDS

Program execution for programming courses is still an application field of the Web in which program educators have had little experience. From an educational point of view, inflexibility and delays are the most serious drawback of using the Web for program execution. Wider use of the Web will require careful implementation of comprehensive programming systems. The implementation of these systems is even more demanding than traditional ones, because a Web-based architecture and a different user interface must be designed. It will be some years before a satisfactory level of maturity is achieved by these kinds of Web applications.

In contrast, software visualization and animation achieved technical maturity in the early nineties. In the years thereafter, most efforts have focused on their educational adequacy. In this respect, some maturity is being achieved, but additional efforts are still necessary. We believe that we will see more efforts aimed at designing and using some of the elements constituting effective animations, and at facilitating their easy construction and customization by the user.

Finally, there are currently several research efforts aimed at designing and developing dynamic electronic books, with dynamic contents and virtual laboratories. We believe that these virtual laboratories for programming will incorporate interfacing with programming tools such as program compiling and execution, and will include automatic visualization and animation.

CONCLUSION

We have provided a panoramic view of the educational use of program execution and visualization on the Web. We have offered an overview of the main technical achievements and the educational uses of them. In the case of algorithm animations, we have also explained some of the lessons that have been learned. In the near future, we expect that there will be increased interest in Web-based tools to support these activities.

REFERENCES

Arnow, D., & Barshay, O. (1999). WebToTeach: An interactive focused programming exercise system. *Proceedings of the 29th ASEE/IEEE Frontiers in Education Conference* (pp.39-44) (session 12a9).

Baecker, R. (1998). “Sorting out sorting”: A case study of software visualization for teaching computer science. In J.T. Stasko, J. Domingue, M.H. Brown, & B.A. Price (Eds.), *Software visualization* (pp.369-381). Cambridge, MA: MIT Press.

Baker, J.E., Cruz, I.F., Liotta, G., & Tamassia, R. (1995). A model for algorithm animation over the WWW. *ACM Computing Surveys, 27,* 568-572.

Bazik, J., Tamassia, R., Reiss, S.P., & van Dam, A. (1998). Software visualization in teaching at Brown University. In J.T. Stasko, J. Domingue, M.H. Brown, & B.A. Price (Eds.), *Software visualization* (pp.383-398). Cambridge, MA: MIT Press.

Berghammer, R., & Milanese, U. (2001). KIEL – A computer system for visualizing the execution of functional programs. *Functional and (constraint) logic programming* (pp.365-368).

Brown, M.H. (1988). *Algorithm animation*. Cambridge, MA: MIT Press.

Brown, M.H., & Raisamo, R. (1997). JCAT: Collaborative active textbooks using Java. *Computer Networks and ISDN Systems*, *29*, 1577-1586.

Brummond, P. (2001). The complete collection of algorithm animations, *http://www.cs.hope.edu/~alganim/ccaa*.

Burd, D.D. (2000). Web based support of a programming class. In A. Aggarwal (Ed.), *Web-based learning and teaching technologies: Opportunities and challenges* (pp. 175-197). Hershey, PA: Idea-Group Publishing.

Crescenzi, P., Faltin, N., Fleischer, R., Hundhausen, C., Näher, S., Roessling, G., Stasko, J., & Sutinen, E. (2003). The algorithm animation repository, *http://algoanim.cs.ust.hk*.

Domingue, J., & Mulholland, P. (1997). Staging software visualizations on the Web. *Proceedings of the 1997 IEEE Symposium on Visual Languages*.

Elenbogen, B.S., Maxim, B.R., & McDonald, C. (2000). Yet, more Web-based exercises for learning C++. *Proceedings of the 31st SIGCSE Technical Symposium on Computer Science Education*, 290-294.

English, J. (2001). BURKS 6 Online, *http://burks.bton.ac.uk/*.

Gloor, P.A. (1998). User interface issues for algorithm animation. In J.T. Stasko, J. Domingue, M.H. Brown, & B.A. Price (Eds.), *Software visualization* (pp.145-152). Cambridge, MA: MIT Press.

Gregorio-Rodríguez, C., Llana-Díaz, L., Palao-Gostanza, P., Pareja-Flores, C., Martínez-Unanue, R., & Velázquez-Iturbide, J. Á. (2001). EXercita: Automatic Web publishing of programming exercises. *Proceedings of the 6th SIGCSE/SIGCUE Conference on Innovation and Technology in Computer Science Education* (pp.161-164).

Gregorio-Rodríguez, C., Llana-Díaz, L., Palao-Gostanza, P., Pareja-Flores, C., Martínez-Unanue, R., & Velázquez-Iturbide, J.Á. (2002). Exercita, *http://aljibe.sip.ucm.es.*

Haajanen, J., Pesonius, M., Sutinen, E., Tarhio, J., Teräsvirta, T., & Vanninen, P. (1997). Animation of user algorithms on the Web. *Proceedings of the 1997 IEEE Symposium on Visual Languages*.

Hansen, S., Schrimpsher, D., & Narayanan, N.H. (1999). From algorithm animations to animation-embedded hypermedia visualizations. *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications* (pp.1032-1037).

Hiltz, M., & Kögeler, S. (1997). Teaching C++ on the WWW. *Proceedings of the 2nd SIGCSE/SIGCUE Conference on Integrating Technology into Computer Science Education* (pp.11-13).

Ibrahim, B. (1994). World wide algorithm animation. *Proceedings of the 1st World-Wide Web Conference* (pp.305-316).

Khuri, S. (2001). Designing effective algorithm visualizations. In E. Sutinen (Ed.), *Proceedings of the First Program Visualization Workshop* (pp. 1-12). Joensuu, Finland: University of Joensuu.

Knox, D.L. (2002). The computer science teaching center, *http://www.cstc.org/*.

Kovourov, S., Tamassia, R., Bridgeman, S., & Goodrich, M. (2000). SAIL: A system for generating, archiving and retrieving specialized assignments using LaTeX. *Proceedings of the 31st SIGCSE Technical Symposium on Computer Science Education* (pp.300-304).

McCauley, R. (2001). Computer science education links *http://www.cacs.usl.edu/~mccauley/edlinks/*

Naharro-Berrocal, F., Pareja-Flores, C., Velázquez-Iturbide, J.Á., & Martínez-Santamarta, M. (2001). Automatic Web publishing of algorithm animations. *Upgrade, II*(2), 41-45.

Naps, T.L. (1996). Algorithm visualization served on the World Wide Web: Why and how. *Proceedings of the 1st SIGCSE/SIGCUE Conference on Integrating Technology into Computer Science Education* (pp.59-61).

Naps, T.L., Eagan, J.R., & Norton, L.L. (2000). JHAVÉ: An environment to actively engage students in Web-based algorithm visualizations. *Proceedings of the 31st SIGCSE Technical Symposium on Computer Science Education* (pp.109-113).

Naps, T.L., Roessling, G., Almstrum, V., Dann, W., Fleischer, R., Hundhausen, C., Korhonen, A., Malmi, L., McNally, M., Rodger, S., & Velázquez-Iturbide, J.Á. (2003). Exploring the role of visualization and engagement in computer science education. *ACM SIGCSE Bulletin*, *35*(3), 131-152.

Pareja-Flores, C., & Velázquez-Iturbide, J.Á. (2002). Program execution and visualization on the Web. In A. Aggarwal (Ed.), *Web-based learning and teaching technologies: Opportunities and challenges* (pp.236-259). Hershey, PA: Idea Group Publishing.

Price, B., Baecker, R., & Small, I. (1998). An introduction to software visualization. In J.T. Stasko, J. Domingue, M.H. Brown, & B.A. Price (Eds.), *Software visualization* (pp.3-27). Cambridge, MA: MIT Press.

Roman, C.-G., Cox, K., Wilcox, C., & Plun, J. (1992). Pavane: A system for declarative visualization of concurrent computations. *Journal of Visual Languages and Systems*, *3*, 161-193.

Skiena, S.S., & Revilla, M.A. (2003). *Programming challenges: The programming contest training manual*, Springer-Verlag.

Stasko, J.T. (1990). Tango: A framework and system for algorithm animation. *Computer*, *23*(9), 27-39.

Stasko, J.T., Domingue, J., Brown, M.H., & Price, B.A. (Eds.). (1998). *Software visualization*. Cambridge, MA: MIT Press.

KEY TERMS

Algorithm Animation: Visualization of a piece of software illustrating the main ideas or steps (i.e., its algorithmic behavior), but without a close relationship to the source code.

Information Visualization: Visualization of phenomena by means of appropriate representations. It is a field different from scientific visualization since information visualization emphasizes delivering visualizations that improve comprehension, whereas scientific visualization emphasizes delivering realistic visualizations.

Multiple Views: Visualization technique consisting in showing several aspects or views of a program, where any view typically shows few aspects and must be comprehensible by itself. Typical views are code vs. data, several levels of abstraction, logic vs. performance, history, and several simultaneous algorithms.

Pretty-Printing: Rudimentary form of program visualization whose goal is to make source code easier to read by means of spacing, indentation and layout. Pretty-printers are programs that systematically indent the source code of a source program according to its syntactic structure and given aesthetic rules. Sophisticated pretty-printers can also highlight different parts of a program and even generate a “program publication”.

Program Execution: To make a computer to follow the instructions stated in the program. It is also known as running a program.

Program Testing: To execute a program in order to check whether it is correct with respect to its specification. Testing requires a previous planning to design input data that cover all or almost all of the cases.

Program Tracing: To follow the execution of a program under control of the programmer. It is usually carried out with incorrect programs in order to debug and fix them. Typically, there are different operations that allow the programmer to control the pace of program execution and to watch the state of different structures.

Program Visualization: Visualization of a piece of software so that the representation has a close relationship to the source code.

Visualization: To make something visible by means of some representation. A visualization consists of two elements: a mental process and a graphical language. Note that “to visualize” is different from “to see”.

Web Repository: A Web site that contains a wide collection of items. It can be unstructured, semi-structured or structured.

Project Management Best Practices to Increase Success

Jeffrey L. Brewer
Purdue University, USA

INTRODUCTION

Utilizing good project management practices has become one of the key differentiators in delivering successful information technology projects. Kerzner (2001) defines project management as "the planning, organizing, directing, and controlling of company resources for a relatively short-term objective that has been established to complete specific goals and objectives" (p. 4). The field of project management has seen explosive growth in the amount of individuals holding a job title such as project manager, in the amount of research being conducted, and in the amount of books and articles being published. This paper explores the reasons for this growth, the reasons why project management has become so important to the on-going success of IT projects and thus the success of organizations and what future directions the field of project management will travel.

BACKGROUND

In 1995, a study entitled "CHAOS" was conducted by the Standish Group. The study surveyed 365 information technology (IT) executive managers in the United States who managed more than 8,000 IT application projects. The sample contained small, medium, and large companies across several industry segments including banking, securities, manufacturing, retail, wholesale, health care, insurance, and local, state, and federal organizations. The Standish Group also conducted focus group sessions and numerous personal interviews to provide a qualitative background for the survey results. The results of the study showed, as the title of the study indicates, that IT projects in the United States were in total disarray (see *Table 1*). "A huge portion of the more than $250 billion spent annually on IT application development is wasted because companies fail to utilize effective project management practices." Average cost overruns were 185%, average time overruns were 222%, only 16.2% of projects were counted as successful and the projects were only delivering 61% of the desired features. Successful projects were defined as meeting all project objectives on time and on budget. The study concluded that project management was one of the top catalysts to ameliorate these statistics.

Wilder & Davis (1998) agreed with the CHAOS study stating that poor project management is a major contributing factor leading to failed IT projects.

The Standish Group repeated the study in 2001 entitled "Extreme Chaos" and observed some noteworthy improvements (see *Table 1*). Successful projects had increased from 16.2% to 28%, and average time overruns had diminished from 222% to 63%; likewise average cost overruns went from 185% to 45% and delivery of required features rose from 61% of the total to 67%. The study listed the following items as contributors to the improvements in IT project results:

1. Improved project management
2. Better development tools
3. Reduction in hardware and software costs
4. Better management processes.

One of the major reasons for the improvements, mentioned in the CHAOS study, was attributed to better project management practices and better-trained project managers. When you look at how these distressing statistics were improved and read about some of the tremendous project disasters (Bailey, 1996; Gibbs, 1994; Lucas, 1995), they demonstrate how important project management has become. The importance of project management to today's organization continues to increase. Schwalbe

Table 1. Standish Group Study Results

	1995	2001
Successful IT Projects	16.2%	28%
Percent of projects cancelled	31%	23%
Average time overruns	222%	63%
Average Cost overruns	185%	45%
Delivery of required features	61%	67%

(2004) reports that the U.S. spends $2.3 trillion on projects every year, an amount equal to one-quarter of the nation's gross domestic product. All nations combined spend nearly $10 trillion of its $40.7 trillion gross product on projects of all kinds. More than half a million new IT application projects were started in 2001, up from 300,000 in 2000 (The Standish Group, 2001).

We can see from these statistics that project management is and will continue to be important to the success of today's organization. The next section outlines three key "best practices" that need to be adopted by organizations to allow project management to reach a higher level of success.

BEST PRACTICES

This section of the article lists and describes three project management best practices: a project management office, establishing a project management methodology, and finding or making good project managers. The top three best practices were chosen based on the literature review, personal interviews, and the author's 20 plus years of IT project management experience. The literature review consisted of journal articles, books, and case studies (Cai et al., 2004; Crawford, 2002; Johnson et al., 2001; Kerzner, 2003; McConnell, 1998; Murch, 2001; Perks, 2003; TechRepublic, 2001; The Standish Group, 2001; Visitacion, 2003).

Three large organizations with established project management practices were used to conduct the personal interviews. The interviews were done in face-to-face sessions held at the respective organization's facility. Two to three project managers with an average of 15 years of project management experience each were interviewed at each organization. The interviews were designed to serve two purposes: one, to substantiate the information that was gathered during the literature review and, secondly, to generate new ideas. The three organizations, two large pharmaceutical companies and a large cardiovascular medical product company asked that their names not be mentioned.

Establish a Project Management Office (PMO)

There are several variations that exist on what a PMO is; depending on what role a PMO plays in an organization and what level it operates at. A PMO is the "administrative mechanism by which a focal point is provided for organizational project management activities" (Rad, 2001). In some corporations, a PMO functions as a support organization that caters to multiple projects with administrative, time tracking, reporting, and scheduling services, while in some others it is merely responsible for business and technical management of a specific contract or program only. Depending on the maturity and capability of a PMO, it can serve different functions. Crawford (2002) discusses how PMOs can operate at three different levels. Level 1, or the individual project level, helps add value to individual projects by defining basic processes that can then be adopted by other projects. At Level 2, the PMO helps to diffuse the processes and uniform methodology to other projects and divisions. Level 3, the corporate level, has PMOs managing the entire collection of the organization's projects and reviewing their goals, history, and progress.

PMOs can help improve project success rates and establish standard project management practices throughout the organization (Kerzner, 2003). However, there is no uniform approach for success of a PMO. Each PMO has to conform to the specific company's culture. Robert Handler, vice-president of Meta Group's enterprise planning and architecture strategy service, feels that a PMO has to be "instituted in a way that doesn't fly in the face of the culture" (Santosus, 2003). If done correctly, a PMO can offer more accurate schedule estimates, improve stakeholder satisfaction levels and facilitate higher employee productivity rates. Even though many organizations have been moving from a functional organizational structure to a matrix or projectized structure in recent times, the PMO might represent a revolutionary change. Crawford (2002) states that, "reorganizing a company's work around projects is the equivalent of moving from a feudal system to participatory democracy."

The efficacy of a PMO has been questioned by several organizational decision-makers. As with any new technology or concept, there are proponents and detractors. There are those who dismiss the concept of a PMO as a fad and regard it with a high level of distrust. Tom Pohlman, an analyst at Forrester Research Group and author of the report *How Companies Govern Their IT Spending* feels that too many PMOs function as "process cops and report compilers for executive teams and often lose sight of what they are supposed to be doing — making sure projects are running effectively" (Hoffman, 2003). "People think about implementing a project office and they usually think bureaucracy paperwork and increased costs" (Bernstein, 2000).

The current concept of a PMO, which now has the responsibility for maintaining all project knowledge (Kerzner, 2003), evolved as recently as 2001 and hence it is still in its fetal stage. A study conducted by the Forrester group, based on telephone interviews with 704 North American IT decision-makers between late April and June of 2003 reported that 67% of the respondents said that their organizations have one or more PMOs, up from 53% the previous year. Gartner group has predicted that, "through 2004 companies that fail to establish a project office will experience twice as many major project delays, overruns, and cancellations as will companies with a project office in place."

Kerzner (2003) lists the following benefits of using a PMO:

- Standardization of operations
- Company rather than silo decision-making;
- Better capacity planning (i.e., resource allocations)
- Quicker access to higher quality information
- Elimination or reduction of company silos
- More efficient and effective operations
- Less need for restructuring
- Fewer meetings that rob executives of valuable time
- More realistic prioritization of work
- Development of future general managers

Adopt a Project Management Methodology

Achieving project management maturity and increased project success generally comes from having defined repeatable processes that are used on every project. These repetitive processes are referred to as the project management methodology (Kerzner, 2003). In a recent PM Network article, Jeff Sterba, Chairman, President and CEO, PNM Resources Inc., stated that they implemented project management methodologies in 2001 to manage enterprise project efforts. The tools and processes they developed allowed them to meet their goals on time by eliminating do-over work and controlling last-minute changes. "Now, we have trained more than 1,400 employees in project management skills, and these cross-functional team members help ensure success in all of our project management initiatives, saving us valuable resources and improving our productivity" (p. 31). In this same article; Tony Salvaggio, President, Computer Aid Inc., is quoted as saying "I saw that we could have a dramatic competitive advantage if we implemented advanced project management for all our activities…It keeps us from making monstrous mistakes and also keeps us from making the same mistakes repeatedly" (p. 32).

Www.dictionary.com defines a methodology as: a body of practices, procedures, and rules used by those who work in a discipline or engage in an inquiry; a set of working methods. Murch (2001) defines four key components:

- Guidelines – defined flexible steps necessary for successful application development
- Techniques – detailed process descriptions that support the activities throughout the product development lifecycle
- Tools – project management tools used in support of the methodology
- Templates – Reusable documents and checklists

The need for adopting a project management methodology is clearly established, what remains is which one and how.

To lower cost, reduce resource requirements for support, minimize paperwork, and eliminate duplicate effort an organization should maintain and support a single methodology across the organization (Kerzner, 2003). Many methodologies exist either commercially or as a by-product of hiring a consulting company that has its own. Murch (2001, p. 143) offers several examples: Process Engineer from Platinum, Inc., SUMMIT from Price Waterhouse Coopers, METHOD/1 from Andersen Consulting, and Architect from JMC, Inc. Many organizations have chosen to develop their own methodology starting with a defined, complete body of knowledge and adapting this to their organization.

All three of the organizations interviewed for this article had their own project management methodology, which was based on some established principles and methods. The path they each used to evolve from a standard body of knowledge to a tailored methodology was very similar:

1. Current assessment of projects, organizational culture and identification of control metrics
2. Obtain senior management commitment and executive champion
3. Training for the entire organization
4. Don't start from scratch, base it on some proven tools and techniques
5. Start with a light or less complex methodology and grow it as you learn
6. Integrate the project management methodology with other management processes.
7. Review lessons learned and adapt.

The Project Management Institute (PMI) describes their body of knowledge in a document called A Guide to the Project Management Body of Knowledge (PMBOK). The PMBOK defines 5 process groups (initiating, planning, executing, controlling, and closing), which overlap over time during the project, and nine knowledge areas (management of integration, scope, time cost, quality, human resources, procurement, risk, and communications). In the U.K. there is PRINCE2, a process-based approach for project management providing a customizable and scaleable method for the management of all types of projects. PRINCE2 stands for (PRojects IN Controlled Environments) and consists of seven major processes: starting up a project, initiating a project, managing stage boundaries, controlling a stage, managing product delivery, planning, and closing a project (Prince2, 2004). IEEE/EIA 12207 is another body of knowledge, which describes the major component processes of a complete software life cycle and the high-level relations that govern their interactions. This standard covers the life cycle of software from conceptualization of ideas through retirement. IEEE/EIA 12207 describes 17 processes that define the software development life cycle starting at initial concept, through requirements analysis, design, coding, testing, installation, operation, maintenance, and retirement (Gray, 2000).

Finding or Making Good Project Managers

As organizations are becoming more "project" based (Crawford, 2002), the current number of IT project management positions is creating demand for experienced competent project managers. It is important for these organizations to place individuals in a project manager role that have the skills to be successful. As demonstrated earlier in this article from the CHAOS studies done by The Standish Group, project management and better trained project managers are essential to help increase the success rates of IT projects. A project manager must have skills in addition to the basics of project management to succeed. According to David Foote, managing partner at Foote Partners LLC in New Canaan, Connecticut, "It requires all these soft skills that have to do with getting things that you want (and) adjudicating issues between people, managers, egos and agendas. It's how to get a job done without annoying people" (Melymuka, 2000). These soft skills can include basic leadership and team building abilities that are needed for the team to complete the designated project

Finding the right people to fill the roll of project manager has become a major problem for most organizations. Many of these "soft skills" are thought to be innate and may not be teachable. Organizations, in trying to find potential project managers from inside, have turned to personality tests, such as the Myers-Briggs Type Indicator or The Keirsey Temperament Sorter. "The MBTI descriptions of personality did what no other personality instrument has done before, be able to give most people some insight into themselves and others" (Viawest, 2003). Using these evaluations, organizations can examine a person based on the skills required to be a successful project manager and decide whether to put them in that role, whether they will need further training to be successful in a project manager role, or whether they would be best placed in a different position.

FUTURE TRENDS

Many key trends have emerged since the turn of the twenty-first century and will continue to guide project management in the near future. Listed below are several key trends that organizational leaders must pay close attention to.

- Strategic outsourcing – many firms are finding a positive ROI when looking outside of the United States to service many parts of their development efforts. These relationships will create a need for new project management skills in the areas of managing geographically remote teams and diverse culture issues.
- Product and service life-cycles becoming shorter – building solutions faster with higher quality for lower cost will lead to more complex projects to manage.
- Ever increasing rate of technology discovery – technology continues to mature at faster and faster rates adding more technology related risk to IT projects.
- The increasing role of the Internet – The Internet has softened the borders of our organizations, creating new and diverse projects and project teams, creating the need for project managers to master new tools and technique.
- Sophistication of end users – Users of today's technology are getting more sophisticated in their use of technology and more demanding forcing project leaders to be more technology knowledgeable.
- Increasing use of maturity models – Organizations are beginning to use maturity models to assess themselves on improvements in their project management practices. Models such as Capability Maturity Model Integrated (CMMI) from SEI or the new Organizational Project Management Maturity Model (OPM3) from PMI. Project leaders will be held accountable for more than just the successful completion of a single project - are they moving the organization up the maturity scale? (More information about these maturity models can be found at the relevant Internet sites: www.sei.com and www.pmi.com).

CONCLUSION

This article has established that better project management practices are essential to all organizations that wish to increase the success rate of current and future projects.

The number of IT projects is only increasing and requiring more and better-trained project managers. The success rate of IT projects has improved, although much more is needed. Establishing a PMO, although not a silver-bullet, holds great promise in improving repeatable successes. The tools and techniques needed to perform good project management practices have existed for some time; the issue has been and continues to be getting organizations to recognize the benefits to building and using a single sound methodology and training their people to become better project leaders.

REFERENCES

Bailey, R. (1996). Approving systems projects: Eight questions an executive should ask. *PM Network, 10*, 21-24.

Cai, Y., Ghali, S., Giannelia, M., Hughes, A., Johnson, A., & Khoo, T. (2004). Identifying best practices in information technology project management. Retrieved on April 2, 2004, from *www.pmforum.org/library/papers04/itpractices.htm*

Crawford, K. (2002). *The strategic project office*. New York, NY: Marcel Dekker, Inc.

Gibbs, W. (1994). Software's chronic crisis. *Scientific American, 271*, 86-95

Gray, L. (2000). *The guidebook to IEEE/EIA 12207: Standard for information technology software life cycle processes*. Fairfax, VA: Abelia Corporation.

Hoffman, H. (2003). Value of project management offices questioned. *Computerworld, 37* (29), 7.

Johnson, J., Boucher, K.D., Connors, K., & Robinson, J. (2001). Collaborating on project success. Retrieved on March 1, 2004, from *www.softwaremag.com/archive/2001feb/collaborativemgt.html*

Kerzner, H. (2003). *Project management: A systems approach to planning, scheduling, and controlling*. New Jersey: John Wiley & Sons, Inc.

Kerzner, H. (2003, June). Strategic planning for a project office. *Project Management Journal*, *34* (2), 13

Light, M. (2002, June 20). What are project office best practices and metrics? *Gartner*. Retrieved on February 26, 2004, from *www.itap.purdue.edu/itresources/gartner/research/107600/107652/107652.html*

Lucas, J.J. (1995). Work management: Why can't information managers manage? In *Proceedings of the Project Management Institute's 26th Annual Symposium* (pp. 304-310). New Orleans, LA. Upper Darby, PA: PMI.

McConnell, S. (1998). *Software project survival guide*. Redmond, WA: Microsoft Press.

Melymuka, K. (2000, March 27). Born to lead projects. *Computer World*. Retrieved October 31, 2003, from *www.computerworld.com/printthis/2000/0,4814,44218,00.html*

Murch, Richard. (2001). *Project management: Best practices for IT professionals*. New Jersey: Prentice Hall.

Our money's worth: Six executive officers recount how project management warranted their investment. (2004, January). *PM Network*.

Perks, M. (2003). *Best practices for software development projects*. Retrieved on March 2, 2004, from *www.computerworld.com/2003/0,4814,85198,00.html*

Prince2. (2004). What is PRINCE2? Retrieved on June 7, 2004, from *www.prince2.com/whatisp2.html*

Rad, P. (2001). Is your organization a candidate for PMO? *AACE International Transactions*, pg., PM71

Santosus, M. (2003, July 1). Office discipline: Why you need a project management office. *CIO.com*. Retrieved on March 1, 2004, from *www.cio.com/archive/070103/office.html*

Schwalbe, K. (2004). *Information technology project management*. Boston: Course Technology.

TechRepublic. (2001). *Project management best practices*. Retrieved February 24, 2004, from *www.techrepublic.com/2001-6240-0.html*

The Standish Group. (1995). *CHAOS*. Retrieved from *www.standishgroup.com*

The Standish Group. (2001). *CHAOS 2001: A recipe for success*. Retrieved from www.standishgroup.com.

Viawest. (2003). Keirsey temperament versus Myers Briggs types. . Retrieved October 29, 2003, from *http://users.viawest.net/~keirsey/difference.html*

Visitacion, Margo. (2003). *Project management best practices: Key processes and common sense*. Retrieved on March 5, 2004, from *www.forrester.com/*

Wilder, C., & Davis, B. (1998, November 30). False starts, strong finishes. *Information Week*.

KEY TERMS

Capability Maturity Model Integrated (CMMI): A five-level framework laying out a generic path to process improvement for software development in organizations.

Project: A temporary sequence of related activities that produces a unique product or service.

Project Management: The application of knowledge, skills, tools and techniques to project activities in order to meet or exceed stakeholder needs and expectations of a project.

Project Management Institute: International professional society promoting good project management practices.

Project Management Methodology: A defined repeatable process to aid in the successful completion of projects.

Project Management Maturity Model: A framework for helping organizations improve their practices and processes.

Project Management Office: A project office is a corporate management organization that evaluates measures and essentially enforces the performance and interaction of the implementation of IT project processes across a company's business units (Computerworld, 2003).

Project Management for IT Projects

Len Asprey
Practical Information Management Solutions Pty Ltd, Australia

INTRODUCTION

The provision of information systems, technology, and communications is a fundamental requisite to support competitive strategies and improve service levels in enterprises. Increasingly, with e-commerce developments using the Internet and World Wide Web (WWW), combined with increased compliance and regulatory requirements, medium to large businesses and governments are relying on IT to support operations and administration. Small to medium businesses also make use of Web, e-mail, and office productivity packages to gain commercial advantage.

Organizations establish projects to deliver business solutions using IT systems. Each project is unique and may be characterized by specific objectives, defined start and finish dates, a unique set of activities, resources (funds, people, and equipment), a discrete organization structure, and a requirement for management (Willson-Murray, 1997). Projects need to be managed, because set objectives have to be achieved within constraints such as scope, time, and budget.

Project management (PM) is not a discipline confined to the IT environment. PM is universal across industry sectors, both commercial and government sectors, and professions. The Project Management Institute (PMI), which had 70,000 members worldwide in 2001, included 15,000 members that had joined the Information Systems Specific Interest Group (Schwalbe, 2002).

International Standard (IEEE Std. 1490–1998) publishes a subset of the Project Management Institute's *A Guide to the Project Management Body of Knowledge*, embodying "generally accepted" knowledge and practices. The term "generally accepted" means that the content of the Standard applies to most projects, and that there is pervasive agreement about their suitability and value. However, it does not mean that the content applies consistently to all IT projects without taking into account appropriateness.

BACKGROUND

The framework of PM incorporates concepts of a project life cycle, thus demonstrating the nature of projects; essentially, projects are temporary endeavors with a defined start and end. Each project life cycle consists of phases. Each phase is a manageable portion of work, and the phases collectively make up the project life cycle. PM processes are applied to define, manage, and complete a project or phase of a project.

A Guide to the Project Management Body of Knowledge (Project Management Institute, 2002), commonly referred to as *PMBOK*, distinguishes five interrelated processes, as follows:

- *Initiating*: covers activities for initiating and authorizing a project and formulating a project charter
- *Planning*: involves defining objectives and selecting the best project approach
- *Executing*: relates to the coordination of people and resources to execute the plan
- *Controlling*: deals with monitoring, measuring progress, and managing variances to a plan
- *Closing*: deals with acceptance and closure of the project

However, there are alternative approaches, including the structured *PRINCE2* PM method. *PRINCE2* is a nonproprietary methodology developed by the U.K. Government (Office of Government Commerce, 2004). It features a startup component prior to project initiation and processes for directing a project, controlling a stage (or phase), managing stage boundaries, as well as closing a project. There are other approaches, such as four interrelated processes, being definition, planning, execution, and close-out (Verzuh, 1999).

It is important to understand the relationship between product life cycle methodologies and PM processes when initiating and managing an IT project. Both are interrelated, but the differentiation between the two is not always clear and may depend on the methodology adopted for the project.

Product life cycle methodologies provide guidelines for planning and implementing "products" such as IT systems. There is a diversity of life cycle methodologies in the commercial or public domains that apply to IT projects. Each methodology may offer unique features, but each usually consists of a series of phases, such as requirements, design, construction, and operation (Verzuh, 1999). They may present a product-centric life cycle approach with a more extensive series of phases (Willson-Murray, 1997).

PM processes may be applied to the project life cycle and the various phases within that cycle. Processes also intersect during a project or phase of a project. It is not unusual for product life cycles and PM processes to be combined for an IT project. Some PM methodologies integrate product-centric life cycles and processes, recognizing that IT projects deliver products, not activities (Bentley, 2000).

For example, a product life cycle might be applied to the deployment of a new IT system to cover the project from initiation to closure. The life cycle usually consists of a number of phases, one of which might be "procurement." PM processes might be applied to that specific phase to manage the procurement of hardware, software, communications, and professional services.

PM methodologies provide a generic interpretation of IT product life cycles and integrated PM processes. Specific methodologies may focus on the application of PM to types of IT products, covering a range of complex software implementation projects.

One example includes a project approach to supply chain management solutions (Ayers, 2003). Another example addresses the delivery of customer relationship management systems, to reduce the high failure rate in those types of projects (Gentle, 2002). Similar work has included the application of PM to integrated document and Web content management systems (Asprey & Middleton, 2003).

APPLICATION

PM processes are complemented in *A Guide to the Project Management Body of Knowledge* by nine knowledge application areas, covering the management of project integration, scope, time, cost, quality, human resources, communications, risk, and procurement (Project Management Institute, 2002).

Integration Management

Integration management relates to the coordination of elements such as developing and executing the project plan and integration of change control.

Plan development begins at initiation, once the aims of the project are mandated, initial constraints and assumptions are identified, and an approach is adopted. Planning draws upon other organizational strategies, plans, and knowledge resources.

Project managers utilize a work breakdown structure (WBS) technique to capture specific tasks in a logical assembly of related activities. The WBS is an input to a Gantt chart, and is often created in a Gantt chart tool, generally embodied within PM software. A Gantt chart is used to generate a project schedule, incorporating activities (or tasks), durations, start and finish dates, allocated resources, and linkages between activities.

The project plan is the key planning document for a project. Its structure varies depending on the PM methodology, but common elements are observable. A plan typically incorporates the project mandate (or charter), documents the project strategy or approach, and includes definition of scope inclusions and exclusions, and lists key stakeholders. It includes the description of the critical success criteria and performance measures. It defines the project organization, resourcing, budget and cost estimates, and key activity/deliverable and milestone dates. The plan also includes communication strategies and reporting structures, a risk/mitigation strategy, and is usually annexed with the relevant Gantt schedule.

When approved, the project plan is termed a "baseline" plan, being the original "approved" plan. Project execution is the process of putting the project plan into effect. The project manager is required to direct technical resources and coordinate organizational interfaces. The baseline project plan is monitored by the project manager, who takes corrective actions where activities or events impact the plan. Performance against the plan is evaluated by submission of status reports and planned meetings with a project sponsor or project board, who have governorship of the project.

Where variations to the scope of a project are required during the life cycle, these changes are subject to formal change control. Change control is a key component of integration management. It includes the management and processes for changes to project scope across all elements of PM. Integration management embraces the requirement to take into account the impacts of the change on elements such as schedule, costs, resources, risk, and quality management.

Change control usually involves submission of a change request form, which is logged in a change control register. The change request is assessed in terms of its impact on scope, costs, project timelines, and deliverables, and an evaluation of benefits and risk. The change request is then submitted for approval by the project board or other governing authority, and if approved, is then incorporated within the project scope.

Scope Management

Key to successful project delivery is the definition and management of scope, which defines the work to be performed by the project, and that work which is specifically excluded. Scope is typically defined during project initialization, encapsulated in a *project charter* or *project* *initiation document*, depending on the project methodology. The scope statement should include a detailed list of work that is within scope, and a definitive list of work that is not within scope. Where work activities are not determined as scope inclusions or exclusions, these need to be documented and resolved.

Complexities can arise with scope management, because scope may not be clearly understood by either customer or solution provider during the initiation of a project, or it may not be defined in sufficient detail to avoid ambiguities. The classic approach to scope management frequently involves a struggle between project manager and customer, particularly when requirements become clearer during the development of specifications.

An iterative approach to software development may be useful, and rapid application development (RAD) is one example of an iterative approach to requirements development, which helps to accelerate delivery of IT systems (Gantthead.com, 2004). However, RAD may not always be applicable, depending on a range of project characteristics, including scope, size, geographical dispersion, and other factors. RAD appears to be most useful where projects have a limited, well-defined scope, and a measurable outcome (Shelford & Remillard, 2003).

Time Management

Time management deals with processes aimed at the timely completion of projects. During the creation of a project plan, the specific activities that need to be performed are defined in the WBS, and these are related in a logical sequence. Resources are assigned to the activities and the level of effort required to perform each activity is determined.

Estimates for the duration of each activity are based on a number of factors. These include the nature of the activity and the capabilities of available resources. They also include the knowledge of the project manager or technical resources, and the application of quantitative measurements, where applicable. Consideration of duration needs to take into account the difference between "elapsed time" and "level of effort." For example, the elapsed time for end-user training might be two days, but the level of effort of a training manager during that activity might only be 0.5 days, as the resource activity might represent, for example, a coordination effort.

The durations assigned for activities are monitored during the execution of the project. If assigned activities take longer than planned, then those aspects need to be managed. Generally, contingencies are built into a project schedule to make provision for minor slippages in durations, but if a major slippage occurs, then change control processes may be invoked. If new activities emerge that were not anticipated during scope definition, change control may also be required.

Cost Management

Cost management provides practices for completion of projects within budget and embraces processes such as resource planning, estimating costs, preparing budgets, and controlling costs. Costs need to be determined during project planning and incorporated into a resource plan.

Project costing needs to include costs that will be directly incurred by executing the project (i.e., direct costs), for example, costs such as labor, materials, supplies and equipment, direct facility purchases, training, travel, and miscellaneous. When costing, provision must also be made for indirect costs, such as fringe benefits, facilities (e.g., rent, utility costs, and maintenance), and general and administrative support (Heerkens, 2002).

A common feature of IT projects is that they may be underfunded from the outset of the project. When cost overruns occur during project execution, it may indicate that inadequate cost analysis occurred during planning. Consequently, careful attention must be given to cost estimates when forecasting budget, as these can be subject to levels of inaccuracy.

Much depends on the domain experience of the project manager and technical resources. However, estimating difficulties can also have as their root cause business pressures to fit a project to an approved budget. Estimates may also be understated to provide an initial enticement for project approval. Scope creep can also impact costs, when project requirements are changed without change control (Frame, 2003).

Actual costs are captured during the project against an assigned project code that identifies the project as a cost center. These costs are typically managed using organizational financial management systems. These systems produce reports for analysis by the project manager and enable cost variances to be reported to the project sponsor or project board.

Quality Management

The international quality management standard (International Organization for Standardization, 2004) provides the framework for quality management requirements. Subsequent supporting standards have been developed, e.g., *HB 90.9-2000: Software Development—Guide to ISO 9001:2000*, which provides guidelines for software development (Standards Australia, 2004).

These types of standards are supported by nonproprietary approaches to quality management, such as total quality management and continuous improvement.

The essential elements for delivering IT systems and supporting effective PM are quality planning, quality control, and quality assurance. Quality planning involves identifying the quality standards and principles that are relevant to an IT project, and the development of a quality management plan for the project.

Quality control relates to the management of project deliverables, testing processes, and management of issues, risks, and change control, using monitoring on a regular basis. Quality assurance relates to verifying that the quality control processes are effective, eliminating unsatisfactory performance and instituting corrective actions.

Human Resource Management

The project plan identifies the resource types that are required to be assigned to project activities. It needs to be closely analyzed to determine project dependencies and establish whether there are any resource conflicts that would require schedule revision.

Once the project plan is thus optimized, available resources may be allocated to defined roles, and resource gaps can be identified for subsequent acquisition activities. Additional resources for the roles required may already exist within an organization, or they may have to be in-sourced under a contract arrangement or outsourced to a service provider.

A challenge facing IT project managers is that resources for the particular activities required may be scarce. Skilled resources may not be widely available for a new product or new version of a product, or resources that are skilled and experienced in the product might be in high demand, and unavailable. Planning for resources needs to occur early, and resource acquisition activities, where required, should commence soon thereafter.

Matrix management may be a complication, because resources assigned to a project may report to a line manager outside the project. This situation can give rise to resource management issues, not the least of which is conflicting priorities and directions, with consequential risk. The foundation for matrix management is a clear mandate communicated at project initiation, assisted by resolute project leadership and regular interaction between project and line managers.

It is a project manager's responsibility to provide leadership and direction, delegate actions and act as mentor, develop individual and team competencies, deal with conflict, drive mentor team building, and conduct performance appraisals. These responsibilities may be exacerbated when project members are not colocated.

Advances in technology and communications have also heralded the development of the "virtual team," The benefit of a virtual team is that it can be established quickly and can draw on experiences of resources that are not otherwise available in one location (Rad & Levin, 2003). Challenges can arise due to cultural and language diversity, and different time zones, such that the project manager needs to develop strategies to manage these aspects.

Communications Management

Communications management deals with timely and effective communications. The communication strategy should be developed and incorporated within the project plan. The strategy identifies the key stakeholders and their specific communication needs, such as the type of information they need and when and how it will be provided.

Information must be made available to stakeholders when required, and it must be relevant to the context represented by the stakeholder. For example, an executive manager is unlikely to have a requirement to review a detailed Gantt chart, and may simply require a high-level overview of key activities and milestones, and thus be informed of the manner in which the project is to be executed.

Project reporting also needs to be established such that the project board, project sponsor, project manager, and key stakeholders are aware of the reporting requirements, such as status reports, performance reporting, and budget forecasts.

At the conclusion of the project, there are closure activities for finalizing project information requirements and disseminating appropriate information to relevant stakeholders.

Risk Management

There are two fundamental techniques that support the endeavors of a project manager. These techniques are issue and risk management, each supported by relatively simple tools, i.e., an issues register and risk register.

An issue may be considered as any event or series of related events that have become an active problem that threatens the integrity of a project or related project (Young, 2003). An issue impacts upon a project's ability to satisfy objectives, time frames, or budgets. It needs to be dealt with as a matter of priority, due to the impacts on the project. The issue is documented in the issues register, and processes are invoked for assessing the issue, identifying the impact, and performing the action required, by whom and when, to resolve it.

A risk is a situation that may emerge and that may impact the project. The process of managing risk should begin with a review of the initial project assumptions (Kendrick, 2003). It involves identifying a risk, developing ways to reduce it, and determining how it is to be managed (Healy, 1997).

Each risk needs to be identified and evaluated in terms of its likelihood and potential impact on the project. The risk is documented in the risk register, together with mitigation actions, and a resource is assigned to the actions, with specific dates for resolution.

Procurement Management

Procurement management deals with the acquisition of products and services from parties external to the organization. The procurement process involves identifying what products and services need to be procured, when, and why. Consideration needs to be given to options such as build, lease, or buy, depending on the requirements of the project.

Solicitation of the product or service involves preparation of a detailed scope of work, encompassing requirements specifications, terms and conditions, and contractual conditions. The organization may use a request for information (RFI) to obtain an initial set of offers from suppliers, and then a request for proposal (RFP) to obtain firm proposals, potentially from a restricted set of suppliers. Alternatively, the organization may choose to use an RFP based on open or restricted procurement conditions.

The selection of a supplier involves evaluating proposals, assessing commercial compliance, product, or service capabilities, assessing delivery and support capabilities, and evaluating the risks associated with supplier performance and financial viability as a provider. The primary deliverable from the selection process is the award of contract.

Contract administration involves managing the relationship with the supplier, including contract change control, reporting, and payments. These activities might be supported by contract administration software and financial management systems for invoice processing and payments. Procurement management concludes with the completion of the contract, with formal acceptance and closure.

IMPORTANCE

The PM profession has emerged to meet a need within businesses and governments to deliver projects to scope, on time, and within budget. The requirement for PM has gained importance in the IT community because of the growing complexities in the development and integration of diverse business applications into unified enterprise applications architecture, and integration of the supporting hardware and communications infrastructure. The assignment of a project manager is aimed at assuring that one individual has the authority and responsibility for directing the project, to develop and execute the project

Table 1. A summary of critical issues for PM

Unclear project mandate/charter Project sponsor's inability to articulate aim and strategy of project may impact success	**Cost overruns** Inadequate cost estimating, scope creep, and time delays may impact the project budget
Inadequate planning Poor project planning may impair outcomes, with impacts on scope, time, and budget	**Poor quality** Deliverables may meet time frames but lack quality, creating a poor impression, and may require rework
Scope out of control Inadequate baseline definition of scope may impact project due to vagaries associated with scope definition; scope creep may impact budget, deliverables, and outcomes	**Lack of skilled/experienced resources** Competition for resources may present difficulties in securing skilled/experienced resources; resources may leave a project at a critical time
Time delays Inadequate management of time, i.e., resources focused on nonproject activities, may impact project schedule	**Lack of communication** Failure to communicate with all stakeholders may create tension and may impact the timely review/sign-off of deliverables
Issues/risks mismanaged Issues and risks that are not identified, qualified, quantified, and managed, may have a low, medium, or major impact on the project	**Procurement mismanaged** Inadequate specifications and procurement processes may impact project deliverables; inappropriate/inadequate technology may be selected

plan, monitor performance against the plan, and undertake corrective actions.

CRITICAL ISSUES

Some of the key issues facing project managers are presented in Table 1.

CONCLUSION

PM is vital to the successful implementation of IT projects, irrespective of whether the requirement is a multimillion dollar investment in software, infrastructure, and services for a global deployment, or the acquisition of a first computer, peripherals, software applications, and connectivity to meet the needs of a small business owner.

Knowledge of PM theory does not guarantee a successful project. The project manager has to apply PM principles using innate skills, acquired competencies, and learned experiences, typically relevant to the specific business or technology domain of the project. Professional certification is available to project managers through educational and industry groups, such as the Project Management Professional (PMP) certification offered by PMI.

REFERENCES

Asprey, L., & Middleton, M. (2003). *Integrative document and content management: Strategies for exploiting enterprise knowledge.* Hershey, PA: Idea Group.

Ayers, J. B. (2003). *Supply chain project management: A structured, collaborative and measurable approach*. Boca Raton, FL: Saint Lucie Press.

Bentley, C. (2000). *Managing projects the PRINCE2 way*. Hants, UK: Hampshire Training Consultants.

Frame, J. D. (2003). *Managing projects in organizations: How to make the best use of time, techniques and people* (3rd ed.). San Francisco: Jossey-Boss/Wiley.

Gantthead.com. (2004). Process/project RAD—RAD—Rapid application development process. Retrieved May 13, 2004, from http://www.gantthead.com/process/processMain.cfm?ID=2-19516-2

Gentle, M. (2002). *The CRM project handbook: Building realistic expectations and managing risk*. London; Sterling, VA: Kogan Page.

Healy, P. L. (1997). *Project management: Getting the job done on time and in budget*. Port Melbourne, Victoria, Australia: Butterworth Heinemann, Australia.

Heerkens, G. R. (2002). *Project management*. New York: McGraw-Hill.

International Organization for Standardization. (2004). *ISO 9000 Quality management standard*. Retrieved May 11, 2004, from http://www.iso.ch/iso/en/iso9000-14000/iso9000/qmp.html

Kendrick, T. (2003). *Identifying and managing project risk: Essential tools for future-proofing your project*. New York: AMACOM Books.

Office of Government Commerce. (2004). *Introduction to PRINCE2*. Retrieved May 10, 2004, from http://www.ogc.gov.uk/prince/about_p2/about_intro.htm

Project Management Institute. (2002). *A guide to the project management body of knowledge (PMBOK guide)* (2000 ed.). Newtown Square, PA: Project Management Institute.

Rad, P. F., & Levin, G. (2003). *Achieving project management success using virtual teams*. Boca Raton, FL: J. Ross Publishing.

Schwalbe, K. (2002). *Information technology project management* (2nd ed.). Canada: Course Technology.

Shelford, T. J., & Remillard, G. A. (2003). *Real web project management: Case studies and best practices from the trenches*. Boston, MA: Pearson Education.

Standards Australia. (2004). *HB 90.9-2000: Software development—Guide to ISO 9001:2000*. Retrieved May 11, 2004, from http://www.standards.com.au/catalogue/script/details.asp?DocN=AS539054873569

Willson-Murray, R. (1997). *Managing projects: A new approach*. Milton, Queensland: John Wiley & Sons.

Young, T. L. (2003). *The handbook of project management*. London; Sterling, VA: Kogan Page.

KEY TERMS

Gantt Chart: The Gantt chart displays schedule information in a graphical form. It may display tasks, duration, start/finish dates for activities and milestones, dependencies, and allocated resources.

PM: Project management.

PMBOK: *A Guide to the Project Management Body of Knowledge,* which is the term used to describe the collective knowledge within the project management profession. It is a book published by the Project Management Institute.

PMI: Project Management Institute: details for which can be obtained on the Internet at http://www.pmi.org

Rapid Application Development (RAD): A methodology that consists of techniques to accelerate application development. It features prototyping aimed at achieving demonstrable results quickly; iteration, an approach that commits to incremental development of the prototype; and a time-box concept that focuses on delivery, not scope.

Request for Information (RFI): An invitation sent to suppliers that seeks information about products or services and requests initial pricing estimates.

Request for Proposal (RFP): A request sent to suppliers that seeks proposals for products or services and requests detailed pricing information.

Work Breakdown Structure (WBS): A logical assembly of related activities that describes the totality of the work to be performed. Each activity may contain multiple subactivities or groups of subactivities in a descending order of detail.

Project Management and Graduate Education

Daniel Brandon
Christian Brothers University, USA

INTRODUCTION

Project Management is "the application of knowledge, skills, tools, and techniques to the project activities in order to meet or exceed stakeholder needs and expectations from a project" (Duncan, 1996). A project is defined as "a temporary endeavor undertaken to create a unique product or service" (Duncan, 1996). This article provides an overview of the coverage of the project management discipline in academic graduate education.

BACKGROUND

A number of professional organizations have developed around the world to address and foster this specific discipline. Most notable is the Project Management Institute (PMI, www.pmi.org) with about 100,000 members worldwide. Other major international organizations are the Association for Project Management (APM) and the International Project Management Association (IPMA) (Morris, 2001). These organizations have recognized there is a distinct skill set necessary for successful project managers, and the organizations are devoted to assisting their members develop, improve, and keep current these skills (Boyatzis, 1982; Caupin, Knopfel & Morris, 1998).

Several universities have also recognized the fact that project management involves distinct skills, and that the traditional degree programs and courses in both business schools and other schools do not adequately cover and/or integrate these skills. The *Chronicle of Higher Education* recently reported that seven Philadelphia-area corporations established ties with four universities in that region to improve the business skills of computer science and IT students; most of these key skills involved the project management skill sets, which are specifically identified later in this document (*Chronicles of Higher Education*, 2001).

Perhaps self-evident from the previous paragraph is the fact that the knowledge and training needed by project managers covers both traditional business disciplines and disciplines involved with building or making things. Often the skills involved with building or making things would be found in an engineering curriculum, and also in information technology or computer science curriculums.

Since the skill sets needed by project managers are extensive, and since these skills involve both business and engineering disciplines, and also since most candidate students are degreed working adults, most schools have developed their project management curriculums as graduate school programs. A number of universities also have a single "project management" course offered as a graduate or undergraduate course.

TYPES OF GRADUATE DEGREE PROGRAMS

An analysis of universities currently offering graduate project management programs indicates several types of programs being offered:

1. A master's level general degree program (such as an MBA) with a specialization in Project Management;
2. A full masters level (generally MS) program in project management; and
3. A "certification program" of several project management courses.

Some universities offer more than one of these program types. Also in some universities the program is offered in the School of Business (or Management) and in some schools the program is offered in the School of Engineering. In most universities, many of the courses appeared to be shared with other graduate degree programs; in other words, not all of the courses in the program are focused on project management.

PMI (and the other international project management organizations) have a certification program, and for PMI the designation is "Project Management Professional" (PMP). To obtain PMP certification, an individual must have 4,500 hours of documented project management experience over a period of six years, have a BS level college degree, and pass a rigorous 4-hour examination. The first PMP exam was given in 1984 to about 30 people, and today there are over 30,000 PMPs worldwide (Foti, 2001). Once the PMP status is obtained, an individual must earn 60 PDUs (Professional Development Units) every three years. Some universities offer a PMP Exam

Figure 1. Institutions offering graduate credit programs in project management

University	Organize	School	Certificate Program		MBA/MS Specialization		PM Masters Degree	
			# Courses	# PM	# Courses	# PM	# Courses	# PM
Amberton	KA	Business	4	4				
American Graduate Univ.	KA	Business					12	7
Boston University	KA	Business	8	8				
City University	KA	Business	6	6				
Colorado Technical University	Step	Both	6	6	13	6		
George Washington University	Step	Business					12	3
Int'l School of Info. Mgmt.	Step	Business	3	3	12	4		
Keller School of Management	KA	Business	6	4			14	6
Northwestern	Step	Engineering					12	4
Regis University	PG	Business			13	6		
Stevens Inst. Of Technology	Step	Business	4	4	12	4	12	6
U. of Management & Tech.	KA	Both	7	7				
U. of Wisconsin - Madison	KA	Business	6	6				
U. of Wisconsin - Platteville	Step	Business					12	5
University of Central Florida	Step	Engineering	5	1				
University of Maryland	Step + KA	Engineering					10	5
University of Texas - Dallas	Step	Business	6	1	10	1		
Western Carolina University	PG	Business					12	6
Wright State University	Step	Business			12	3		

Preparation course or cover exam prep material in one of their project management courses. However, most graduate programs do not cover exam prep; in fact, the graduate programs studied herein are more geared to providing the PDU credits for PMPs.

Figure 1 summarizes the program types for most of the U.S. universities offering project management programs "certified" by PMI. The list of such schools is on the PMI website (www.pmi.org). Out of the 19 schools listed, 11 offer a certificate program, six offer an MBA/MS specialization, and eight off a full Master's is project management. In 14 of the 19 schools, the program is entirely in the Business (or Management) school.

PROJECT MANAGEMENT KNOWLEDGE ORGANIZATION

PMI has developed an index of project management skills and knowledge called the "Project Management Body of Knowledge" (PMBOK). The PMBOK has been developed through several iterations over many years; the first version was developed in 1976 (Cook, 2004). The latest version (PMBOK, 2000) has been released (for certification testing beginning 1/2002) (PMI, 2000). It defines nine "knowledge area" which are organized into 37 "processes". The processes are grouped into five "process groups". This is illustrated in Figure 2 (for PMBOK, 1996) (Duncan, 1996).

Since so many resources have been put into the development and refinement of the PMBOK and it has been so well received by the project management community, it seemed prudent to us to organize our graduate program courses around the processes defined within PMBOK. The issue then became how do we "slice and dice" the processes as shown in Figure 2 into distinct (but integrated) courses. The PMBOK document itself organizes its write-up by knowledge area. However, most classic overall project management books and textbooks are organized by process groups (Badiru, 1989; Cleland & King, 1988; Hajek, 1984; Kerzner, 1980; Meredith & Mantel, 1989; Royce, 1988; Verzuh, 1999). There are however a number of books concerning particular parts of project management, and these cover particular knowledge areas, but they are not specifically written as "textbooks" (Fisher & Fisher, 2000; Fleming & Koppelman, 2000; Pinto & Trailer, 1999; Schuyer, 2001; Verma & Thamhain, 1996).

Looking at the universities currently offering degree programs to see how their curricula were organized, we defined three general types of organization:

1. "Step" – Courses are organized in the traditional manner from less depth to more depth over most of the knowledge areas. For example, the first course might be "Introduction to Project Management"; the next might be "Intermediate Project Management"; and the next would be "Advanced Project Management";
2. "KA" – Follows the PMBOK knowledge areas (Scope, Time, Cost, …); and
3. "PG" – Follows the PMBOK process groups (Initiation, Planning, …).

Figure 2. PMI process groups and knowledge areas

	Initiation	Planning	Executing	Controlling	Closing
Integration		Project Plan Development	Project Plan Execution	Overall Change Control	
Scope	Initiation	Scope Planning Scope Definition	Scope Verification	Scope Change Control	Scope Verification
Time		Activity Definition Activity Sequencing Activity Duration Estimation Schedule Development		Schedule Control	
Cost		Resource Planning Cost Estimating Cost Budgetting		Cost Control	
Quality		Quality Planning	Quality Assurance	Quality Control	
Human Resources		Organizational Planning	Staff Acquisition	Team Development	
Communications		Communications Planning	Information Distribution	Performance Reporting	Administrative Closure
Risk	Risk Identification	Risk Identification Risk Quantification Risk Response Development		Risk Response Control	
Procurement		Procurement Planning Soliciation Planning	Solicitation Source Selection Contract Administration	Contract Administration	Contract Closeout

Most programs do not fit entirely into one of these molds, but they were categorized according to the best fit. Overall, out of the 19 schools, 10 use primarily the Step method, six use primarily the KA method, and two use the PG area.

For schools offering certification, five use the Step method, six use the KA method, and none use the PG method. For schools offering the MBA/MS specialization, none use the KA method, one uses the PG method, and the rest use the Step method. For schools offering the full MS in Project Management, two use KA's, one uses PG's, and the rest (five) use the Step method.

The issue of course material organization is a difficult one for a university. As discussed earlier, universities offering these programs are taking different approaches in this area. We feel the "Step" approach is only useful for programs that have two or three project specific courses. The "KA" approach requires much more "course preparation" time, textbooks are limited, and instructors need depth in these skills. One possible curriculum design would be to use a combination of "PG" and "KA". For "PG", separation into two process "super-groups" may be appropriate: project planning and project control; both covering scope, time, and cost. Separate "KA" courses would likely involve: procurement, risk, quality, and human resources/communications.

PROJECT MANAGEMENT CONTENT IN PROGRAMS

As can be seen from Figure 1, not all of the courses in a project management program are project management specific courses. For most schools, the certification offering is made up of mostly project management specific courses (the #PM in Figure 1 is the number of project management specific courses). For the project management specialization, most schools use three to six project management specific courses. For the full MS Project Management degree, the number of project management specific courses is about one-third to one-half of the courses. These non-specific courses in the full MS degree program vary widely from school to school especially if the degree is in the Engineering school instead of the Business school. Some of these non-project management specific courses are typically: general management, organizational behavior, leadership, managerial accounting, information technology, finance, human resources, quantitative methods, quality assurance, procurement and contracting, and risk management.

DELIVERY

Some universities are offering some, all, or portions of their courses in the form of "distance learning". So the issue becomes: "where on the spectrum from 'bricks to clicks' should a program position itself." There are many pros and cons on both sides of this issue, and most of those pros and cons depend on exactly how a course is made available "online" and the university's overall vision, mission, and tradition. This issue encompasses most degree programs (not just project management), so we are not going to further debate it here, except to indicate it is highly dependent on a particular school's mission, tradition, and demographics. As discussed in the following section, the potential students for such a graduate program are working adults, so attention has to be given to the best delivery for that market. Many schools are holding classes on weekends or evenings to accommodate the adult audiences for these types of programs (*San Diego Business Journal*, 2001).

PROGRAM STAKEHOLDERS AND THEIR NEEDS

We have discussed and surveyed the needs of the stakeholders of a graduate program in our region. The external stakeholders we identified were those companies and those individuals who would benefit from such a program. The companies would benefit by the introduction or reinforcement of the specific project management methodologies into their organizations; this has both an educational and training perspective.

Our individual stakeholders are primarily degreed working professionals. This is similar to the market served by the other universities we investigated, since those other universities like ourselves are located in large metropolitan areas. These individuals benefit from a "continuing education" perspective that makes them individually more valuable. Those individuals having earned PMI PMP Certification would have another way to earn PDU credits (a credit course at a university earns 15 PDUs per semester credit hour). Currency of methods and tools is also quite important to both corporations and individuals.

FUTURE TRENDS

For future university programs in project management, four dimensions can be defined. The PMI PMBOK focuses on the dimension of breadth of the knowledge areas (and the 37 processes) but intentionally does not go into much depth. Going into depth gets into method and tool specifics. Thus, a future trend in university programs would be to address not only the breadth but also the depth of these key processes.

The next dimension identified is industry particulars. While there is much commonality to project management in all industries, there is also much that is specific to each area. For example, task estimation for an IT project is much different than task estimation in a construction project. So this should be considered as another added dimension to new programs, certainly not for all industries but for the major ones in a school's geographic region.

The next dimension we identified was that of time or "currency". This not only includes the use of current tools, but the practice of project management in the current business and technical environment. Issues such as "virtual teams", international coverage, and Web-based systems would be included in this dimension.

CONCLUSIONS

Herein, we have examined content, approach, and logistics issues in graduate education for project management. The university programs surveyed were all relatively new programs, so there is little or no data available for a statistical or comparative historical analysis at this time. In the future, one may be able to survey graduates from the different types of programs to determine the pros and cons of each type of program organization and the "best practices" for project management education.

REFERENCES

Badiru, A.B. (1989). *Project management in manufacturing and high technical operations*. Wiley Interscience.

Boyatzis, R. (1982). *The competent manager: A model for effective performance*. John Wiley & Sons.

Caupin, G., Knopfel, H., & Morris, P. (1998). *ICB IPMA competence baseline*. Zurich: International Project Management Association.

Chronicles of Higher Education. (2001, August 10). *47*(48), A45.

Cleland, D.I., & King, W.R. (1988). *Project management handbook*. Van Nostrand Reinhold.

Cook, D.L. (2004). Certification of project managers – Fantasy or reality. *Project Management Quarterly*, *8*(2), 32-34.

Duncan, W. (1996). *A guide to the project management body of knowledge*. Project Management Institute.

Fisher, K., & Fisher, M. (2000). *The distance manager: A hands on guide to managing off-site and virtual teams*. McGraw-Hill.

Fleming, Q., & Koppelman, J. (2000). *Earned value project management*. Project Management Institute.

Foti, R. (2001, September). The case for certification. *PM Network*.

Hajek, V.G. (1984). *Management of engineering projects*. McGraw-Hill.

Kerzner, H. (1980). *Project management. A systems approach to planning, scheduling, and controlling*. Van Nostrand.

Meredith, S.R., & Mantel, S.J. (1989). *Project management, A management approach*. John Wiley & Sons.

Morris, P. (2001, September). Updating the project management bodies of knowledge. *Project Management Journal*.

Pinto, J., & Trailer, J. (1999). *Essentials of project control*. Project Management Institute.

PMI. (2000). *A guide to the project management body of knowledge*. Project Management Institute.

Royce, W. (1988). *Software project management*. Addison-Wesley.

San Diego Business Journal. (2001, August 6). *22*(32), 23.

Schuyler, J. (2001). *Risk and decision analysis in projects*. Project Management Institute..

Verma, V., & Thamhain, H. (1996). *Human resource skills for the project manager*. Project Management Institute..

Verzuh, E. (1999). *Fast forward MBA in project management*. John Wiley & Sons.

KEY TERMS

APA: Association for Project Management.

IPMA: International Project Management Association.

Knowledge Area (KA): Project Management Knowledge Area; For PMI there are nine knowledge areas: Integration, Scope, Time, Cost, Quality, Human Resources, Communications, Risk, and Procurement.

Process Group (PG): A grouping of project management processes; for PMI, there are five process groups: Initiation, Planning, Executing, Controlling, Closing.

Professional Development Unit (PDU): A unit of continuing education for certified project managers.

Project Management: The application of knowledge, skills, tools, and techniques to project activities in order to meet or exceed stakeholder needs and expectations of that project.

Project Management Body of Knowledge (PMBOK): A consolidation and organization of information about project management including “best practices”.

Project Management Institute (PMI): The largest of the professional organizations which foster the discipline of project management.

Project Management Professional (PMP): The highest level of professional certification of project managers by PMI.

Project Management Models in IT

Marly Monteiro de Carvalho
University of São Paulo, Brazil

Fernando José Barbin Laurindo
University of São Paulo, Brazil

Marcelo Schneck de Paula Pessôa
University of São Paulo, Brazil

INTRODUCTION

The implementation of formal efficiency procedures is quite new in IT projects. There are different approaches regarding the best practices in the IT project management (Carvalho, Laurindo & Pessôa, 2003; Laurindo, Carvalho & Shimizu, 2003).

Humphrey (1989) identifies maturity levels in the IT project development process, based on the managerial behavior found in companies. The fundamental concepts of the process maturity derive from the belief that the development management process is evolutionary. Paulk, Weber, and Curtis (1995) identify the distinguishing characteristics between immature and mature organizations, as showed in Table 1.

CAPABILITY MATURITY MODEL (CMM)

The CMM (Humphrey, 1989; Paulk et al., 1995; Pessôa & Spinola, 1997) was developed by SEI—the Software Engineering Institute of Carnegie Mellon University—and presents five maturity levels, each corresponding to a set of structural requirements for key process areas (Figure 1).

Although each project is unique, it could be organized in a process to be applied in other projects. IT project managers used to apply a "methodology," that is, they established the steps to be followed in order to develop a system. Another singular characteristic is the dynamic technologies breakthrough that demands continuous improvements in the development methods and management of changing process, as described in CMM model, at Level 5, the highest level of maturity.

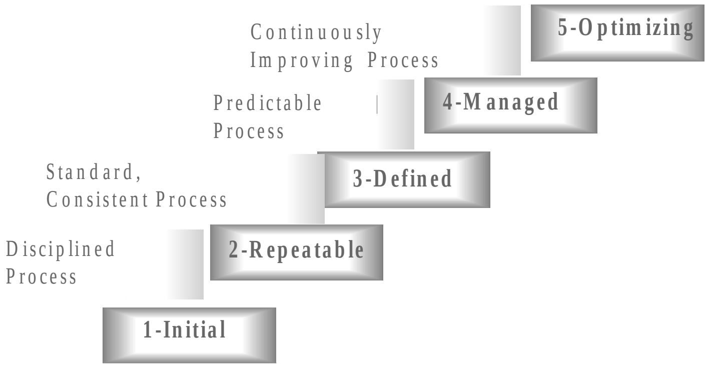
The CMM second level has a consistent project management structure, and the goal of this level is to deliver projects on time. To perform this, the model has several points that must be achieved, like effort and size estimation, strong process control (such as periodic meetings between technical people and managers), and several measures to show project status more clearly.

CMM is not an adequate reference for the assessment of internal methodologies, since it was not conceived to perform this kind of analysis. ISO 15504 (1998) proposed the standard project SPICE as a more appropriate model to evaluate maturity level of specific processes. While CMM level of maturity specifies a set of processes that have to be performed, ISO 15504 establishes maturity levels for each individual process: Level 0—Incomplete; Level 1—

Table 1. Immature organization x mature organization (Paulk et al., 1995)

IMMATURE ORGANIZATION	MATURE ORGANIZATION
• *Ad hoc*; improvised process by practitioners and managers • Not rigorously followed and not controlled • Highly dependent on personal knowledge • Little understanding of progress and quality • Compromising product functionality and quality to meet schedule • High risk when new technology is applied • High maintenance costs and unpredictable quality	• Coherent with action plans; the work is effectively achieved • Processes are documented and continuously improved • Perceptible top and middle management commitment • Well controlled assessment of the process • Product and process measures are used • Disciplined use of technology

Figure 1. Maturity levels (Paulk et al., 1995)



Performed; Level 2—Managed; Level 3—Established; Level 4—Predictable; Level 5—Optimizing. This is a different approach of CMM, since an organization does not perform a maturity level, but has a maturity profile: a maturity level is measured for each specific process. This new approach is very useful to the organization perspective because one can easily measure strong and weak points of their process, and plan improvement activities. Furthermore, from the companies' point of view, it is easier to understand staged levels, as the performed processes are already predefined.

The SPICE approach defined in standard ISO 15504 (1998) had firstly influenced *CMM for Systems Engineering*, published in 1995, and more recently influenced CMM I (CMM-I1; CMM-I2), just published in 2002. CMM-I, the integration model, was enhanced in two dimensions: *scope dimension* and *evaluation dimension*.

In the scope dimension, this new model incorporated other published models and covered all project activities, not only software, as the original software CMM did, but also other engineering fields. In the evaluation dimension, CMM-Il incorporated both approaches: the traditional (called staged CMM) and the maturity profile (called continuous CMM). Figure 2 shows the continuous CMM-I representation to be compatible with ISO/IEC 15504 standard.

CMM-I (and software CMM) considers that maturity level is an organizational characteristic and is independent of the professionals involved. Nowadays, there is a strong tendency towards the adoption of CMM-I models, which were sponsored by the Department of Defense (DoD); meanwhile ISO standards are less used.

PROJECT MANAGEMENT MODELS

Project management plays an important role in the competitive scenario and in the '90s achieved the status of methodology. The model proposed by the Project Management Institute—PMI (2000), called PMBoK, provides

Figure 2. Continuous maturity process representation in CMM-I (CMM-I1, 2002)

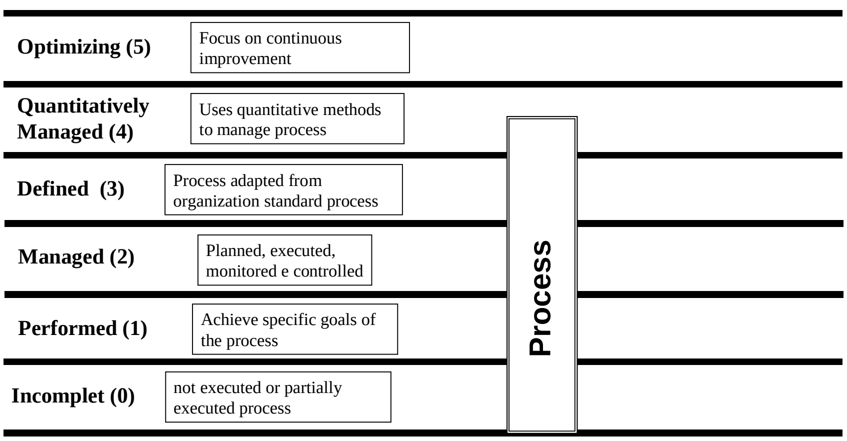
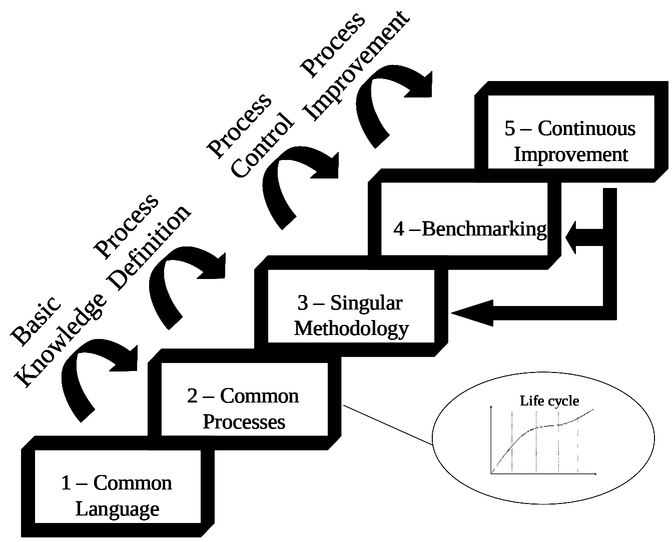


Figure 3. Project management maturity model (Adapted from Kerzner, 2001)



a framework to manage project efficiency, balancing scope expectations and the available resources in nine key areas (Rabechini & Carvalho, 1999).

Nevertheless, the PMBoK framework cannot provide a standard benchmark for project management capability as CMM to software engineering capabilities. In order to extend the Capability Maturity Model to project management, Kerzner (2000, 2001) proposes a Project Management Maturity Model (PMMM).

The PMMM differs in many aspects from the CMM, but this framework also introduces benchmarking instruments for measuring an organization's progress along the maturity model, detailing five levels of development for achieving maturity, as shown in Figure 3 (Carvalho et al., 2003).

It is important to highlight the differences in terminology between the CMM and PMMM (compare Figures 2 and 3) which could lead to misunderstanding when both models are being implemented in the IT domain of the same organization.

PMMM addresses the key knowledge areas across the project management process, in compliance with PMBoK, and integrates them with the Project Management Office (PMO) in the strategic level.

Kerzner (2000) identifies a life cycle in PMMM Level 2, Common Processes, which could be broken into five phases, as shown in Figure 4. It is important to note that some simultaneity among the phases can occur.

The embryonic phase means that the organization starts to recognize the benefits of project management (PM), usually by lower and middle levels of management. The two next phases are achieved when the PM concepts are accepted, and have visible support and commitment by executive and line management.

Kerzner (2001) emphasizes the growth phase as the most critical, because it is the beginning of the creation of the PM process and warns that different methodologies for each project should be avoided.

The last life cycle phase—maturity—is difficult to achieve due to several factors such as organizational resistance to project cost control and horizontal accounting.

The main characteristics of these life cycle phases emphasized by Kerzner (2001) are described in Table 2.

QUALITY SYSTEMS

It is important to note that the adoption of systems models, such as ISO 9000, focuses on the creation and maintenance of a quality system, applied to any process. The ISO 9001:2000 new version, published in the year 2000, was fully restructured to have a more clear process-focused approach. Other ISO standards offer an overview of these standards to the software field and contribute to deploying this approach to specific processes such as software products (ISO 9126-NBR 13596), quality requirements for software packages (ISO 12119), and the software life cycle process (ISO 12207).

ISO 9000-3 (2001) is a guide to help with ISO 9001 interpretation for the software field (Pessôa & Spinola,

Table 2. Life cycle phases' characteristics (Kerzner, 2001)

Phase	Characteristics
embryonic	• recognizing the need for PM • recognizing PM's potential benefits • applications of PM to the business • recognizing the changes necessary to implement PM
executive management acceptance	• visible executive support • executive understanding of PM • project sponsorship • willingness to change the way the company does business
line management acceptance	• visible line management support • line management commitment to PM • line management education • release of functional employees for PM training programs
growth	• development of company PM life cycles • development of a PM methodology • a commitment to effective planning • minimization of scope • selection of PM software to support methodology
maturity	• development of a management cost/schedule control system • integration of schedule and cost control • development of an educational curriculum to support PM

1997). The previous versions of this guide were developed by the ISO/TC/SC2 committee, the quality branch of ISO. Nowadays this ISO 9000-3 guide is being revised by ISO/IEC JTC1/SC7, the information technology branch. The ISO 9001:2000 new structure made this task easier than previous versions, and it is incorporating a map of the relationship between IT standards (ISO/IEC JTC1/SC7) and its respective quality systems described in ISO 9001 to this standard. For example, ISO 9001 specifies that the organizations have to identify their processes, and ISO 12207 proposes a set of processes involving software development, supply, acquisition, and maintenance. In addition to ISO 12207, other standards are referenced, such as ISO 9126 for software products, ISO 12119 for quality requirements for software packages, and ISO 15504 for software evaluation. This was considered an improvement of the standards structure that matches quality system standards with technical and specific standards.

ISO 9001:2000 standards have the purpose of certifying organizations whose quality systems comply with the standards, and provides a structure to manage quality independently of the organizational activity. This is not enough for specific fields of application, and for this reason there are complementary sets of standards in some areas, like QS-9000 for the automobile industry and TL-9000 for the telecommunications industry.

The standards from ISO/IEC JTC1/SC7 have the purpose of complementing the quality system for the IT-specific area, not focusing on applications as with the automobile or telecommunications industry, but considering the specifics of software and systems development.

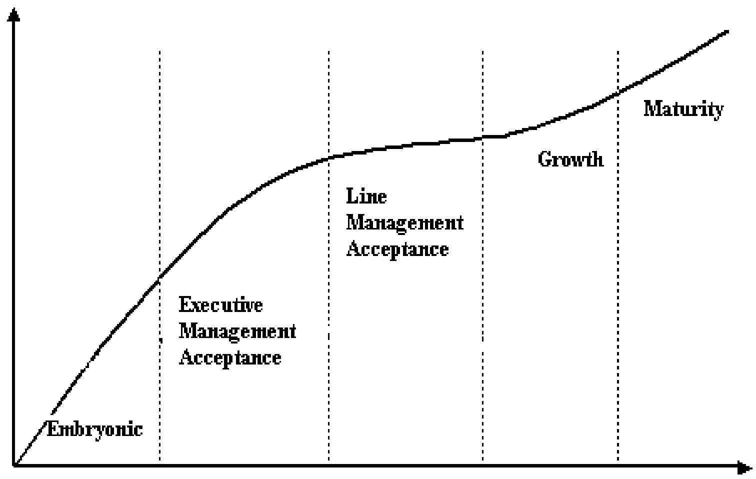
In general, ISO 9000 can be a good starting point for implementing a quality system. This system allows organizations to disseminate quality culture and create the initial structure to implement more specific quality systems.

In general, the quality system (ISO 9000), processes (CMM), and project (PMI) models have the possibility of mutual and complementary synergy, maintaining consistency with their fundamental points. On the other hand, there are important differences among these models, especially concerning the degree of abstraction. (Pessôa, Spinola & Volpe, 1997b; Tingey, 1997).

CONCLUSION

In spite of different approaches regarding the best practices in the IT project management, there is a general consensus about the importance of the following widely used efficiency models: the CMM (Humphrey, 1989; Paulk *et al.*, 1995), the Project Management Maturity Model—PMMM (Kerzner, 2000, 2001); the Project Management Body of Knowledge—PMBoK (PMI, 2000), and quality systems for software—ISO 9000-3 (2001) and ISO 12207 (1995). All of these are empirical and were developed based on best practices used in real projects. Although

Figure 4. Life cycle phases (Kerzner, 2000)



they have different approaches in their conception, they are rather more complementary than conflictive. These models consider that maturity level is an organizational characteristic and it is independent of the professionals involved.

REFERENCES

Carvalho, M.M., Laurindo, F.J.B. & Pessôa, M.S.P. (2003). Information technology project management to achieve efficiency in Brazilian Companies. In S. Kamel (Ed.), *Managing globally with information technology* (pp. 260-271). Hershey, PA: Idea Group Publishing.

CMM-I-1. (2002). *Capability Maturity Model Integration—version 1.1—for systems engineering and software engineering—continuous representation CMU/SEI/SW, V1.1—CMU/SEI –2002-TR01*. Retrieved February 2, 2002, from www.sei.cmu.edu.

CMM-I-2. (2002). *Capability Maturity Model Integration—version 1.1—for systems engineering and software engineering—staged representation CMU/SEI/SW, V1.1—CMU/SEI –2002-TR02*. Retrieved February 2, 2002, from www.sei.cmu.edu.

Humphrey, W.S. (1989). *Managing the software process*. Reading, MA: Addison-Wesley (SEI Series in Software Engineering).

ISO 12207. (1995). *ISO/IEC 12207:1995—information technology—software life cycle processes—ISO*.

ISO 9000-3. (2001, May). *Software engineering-guidelines for the application of ISO 9001:2000 to software*. Working draft WD4 ISO/IEC JTC-1 /SC7/WG18 N48.

ISO/IEC/TR15505-2—SPICE. (1998). *Technical report on information technology—software process assessment—part 2: A reference model for processes and process capability* (1st edition).

Kerzner, H. (2000). *Applied project management best practices on implementation*. New York: John Wiley & Sons.

Kerzner, H. (2001). *Strategic planning for project management—using a project management maturity model*. New York: John Wiley & Sons.

Laurindo, F.J.B., Carvalho, M.M. & Shimizu, T. (2003). Information technology strategy alignment: Brazilian cases. In K. Kangas (Ed.), *Business strategies for information technology management* (pp. 186-199). Hershey, PA: Idea Group Publishing.

Paulk, M.C., Weber, C.V., Curtis, B. & Chrissis, M.B. (1995). *The Capability Maturity Model: Guidelines for improving the software process/CMU/SEI*. Reading, MA: Addison-Wesley.

Pessôa, M.S.P., Spinola, M.M. & Volpe, R.L.D. (1997, October 14). Uma experiência na implantação do modelo CMM. *Proceedings of the Simpósio Brasileiro De Engenharia De Software, 11* (WQS´97—Workshop Qualidade De Software) (pp. 49-57), Fortaleza. São Paulo: Anais.

Pessoa, M.S.P. & Spinola, M.M. (1997, December 5). *Qualidade de processo de software: Um novo paradigma.* Iv Inftel—Congresso Petrobrás De Informática and Telecomunicações, São Paulo. São Paulo: Anais.

Pressman, R.S. (1987). Software engineering, a practitioner's approach (2nd edition). McGraw-Hill.

Project Management Institute. (2001). *A guide to the Project Management Body of Knowledge (PMBoK)*. MD: Project Management Institute.

Rabechini Jr., R. & Carvalho, M.M. (1999). Concepção de um programa de gerência de projetos em instituição de pesquisa. *Revista Valenciana Dèstudis Autonòmics*. Espanha: Valência.

Tingey, M.O. (1997). *Comparing ISO 9000, Malcolm Baldrige, and the SEI CMM for software: A reference and selection guide*. Englewood Cliffs, NJ: Prentice-Hall.

KEY TERMS

Capability Maturity Model (CMM): A framework to achieve maturity in project activities in the software field which presents five maturity levels, each corresponding to a set of structural requirements for key process areas.

CMM I (CMM-I1; CMM-I2): A model, which has enhanced in two dimensions: *scope dimension* and *evaluation dimension*. The CMM-Il incorporated both approaches: the traditional (called *staged CMM*) and the maturity profile (called *continuous CMM*).

ISO 9000-3: A guide to help ISO 9001 interpretation for the software field, i.e., the development, supply, acquisition, and maintenance of software.

ISO 15504: An international standard that proposes the standard project SPICE, which establishes maturity levels for each individual process: Level 0—Incomplete; Level 1—Performed; Level 2—Managed; Level 3—Established; Level 4—Predictable; Level 5—Optimizing.

Life Cycle: Common processes identified in PMMM Level 2 which could be broken into five phases: embryonic, executive management acceptance, line management acceptance, growth, and maturity.

Project Management Body of Knowledge (PMBoK): Provides a framework to manage project efficiency, balancing scope expectations and the available resources. This model proposes the following nine key areas: (i) integration, (ii) scope, (iii) time, (iv) cost, (v) quality, (vi) human resource, (vii) communication, (viii) risk, (ix) procurement.

Project Management Maturity Model (PMMM): A framework that introduces benchmarking instruments for measuring an organization's progress along the maturity model, detailing five levels of development for achieving maturity: Level 1—Common Language; Level 2—Common Processes; Level 3—Singular Methodology; Level 4—Benchmarking; Level 5—Continuous Improvement, as shown in Figure 3.

P

Promotion of E-Government in Japan and Its Operation

Ikuo Kitagaki
Hiroshima University, Japan

INTRODUCTION

In Japan, e-government has been considered since 2001 as one of the strategies of so called "e-Japan" (Ohyama, 2003). It had been decided that e-government shall be constructed within the fiscal year 2003. Preparation in terms of the legal system and technological developments made steady progress towards that goal. The construction of e-government should alleviate residents' burdens in terms of bureaucracy, enhance service quality rationalization, lean and transparent administrative agencies, countermeasures for natural calamities, more participation in policy making and administration by residents and so forth. Various tasks have been carried out at many places. For example, in autumn of 2002 a "basic residential register network" was established. Its initiation enjoyed broadly smooth operation. Residents had received administrative services only within certain jurisdiction limits until then. Now they are free to enjoy access to any administrative services from anywhere in Japan thanks to this e-system. Some local authorities introduced electronic tenders to enhance transparency of administration. Some local authorities adopted an electronic voting system in part of their areas. This paper explains the details of how the construction of the e-government came about and the status of its operation.

In constructing an e-government, basic researches in respect of relevant individual electronic chores are necessary. In reality, however, planning and drawing up an idea will often be brought about, depending on certain actual domestic social circumstances of the legal systems or certain consensus within and between relevant representative bodies of the government. Because of such circumstances, we have decided to list up general magazines easily available which report often on these themes, and the most up-to-date URLs of relevant organizations of the Japanese government (see Appendix).

BACKGROUND

Information processing on the part of the public administration has progressed in line with each development of computers (Makiuchi, 2003). It followed the progress of information processing in the private sector. The first steps included, as early as the 70s, information processing for specified jobs such as accounting and salary payments. It also controlled systems of government offices on the basis of one PC per person, as personal computers spread around 1990, internet technologies were introduced, and so forth. In particular, the Patent Office in Japan adopted an online system for patent application to speed up the processing of patent applications and it did substantially increase.

Nevertheless, given the progress and more use of the information processing for administration, the mentality that it only simply meant computers would replace mechanical style jobs lived a long time. In addition, it exerted a negative influence in that the conventionally vertical administration confined the progress of the information processing for administration to specific ministries and offices. And it made services difficult to access from the standpoint of the people.

The Mori Cabinet came into power in 2000. He decided to shift the nation to the use of state-of-the-art information under his "e-Japan strategies". He would emphasize the importance of information processing in Japanese society in order to get out of the economic slump. Objectives were placed in five fields; broadband infrastructure, education, public/administration services, e-commerce, and security. Among the items, administration services are selected so that e-government will be realized within the fiscal year 2003. Preparation in terms of legal systems and technological developments have made progress in that respect. Legal systems were reviewed. Now, there are the Online Transmission Regulations, Revised Basic Residential Register Law, Public Individual Certification Law etc. On the technological front, there were also some new introductions of electronic systems for administrative procedures, for example all-purpose reception system for applications, the introduction of a one-stop system for import/export harbor and taxation procedures as well as the electronic tender system. Electronic systems for revenue bookings allows payments per Internet and an online transfer system. The purposes of the introduction of information processing for administrative chore and business operations are; simplification of work for human resources, salaries, etc. and connection of the National Networks with LGWAN (Local Government Wide Area Network) owned by communalities.

Electronic procedures will be put into practice in various fields as described above. From their own perspectives, users consider more or less that the essential part of e-government is mainly the given opportunity of the electronic application and its ease.

E-government is one of the e-Japan Strategies in the Five Year Plan which started in 2000. Preparatory activities on legal systems and system construction are to be completed by March 2004. On the other hand, as regards set-up of broadband infrastructure, the objective was set that 30 million households would use it by the end of the fiscal year 2005. In reality, 50 million households (the actual number of subscribers is about 5.70 million) were able to use it as early as June 2002 due to the rapid progress of information telecommunication technologies. Partly as a result of these unexpected circumstances, e-Japan Strategies could enjoy another review only after one and half year of their inception.

E-Japan strategies seem to have advanced rapidly. An international comparison survey by a consultancy company called Accenture shows, however, that Japanese e-government ranked 17th.According to this study, the top three positions are Canada, Singapore and the USA. The same survey classifies and names these three nations as the "pioneering leaders". Australia, which clad the 4th, to the 13th positioning it calls the nations as "visionary challengers". It calls the 14th positioning to Japan, which is the 17th, as "emerging performers" and the 18th onwards as "platform builders". Whether or not they succeed in construction of e-government can be a good clue, a basis requirement for international competition for any country.

SYSTEM DEVELOPMENT CASES

Here are some cases of e-governments established on national scheme or at local authorities.

Basic Residential Register Network System

Resident registration cards, tax payment certificates etc. are normally applied for at the desk of an agency. The regional jurisdiction of a resident is responsible and issues them. E-government will enable any application at home or application and issuance from an office outside the resident district. Confirmation of the subject person beyond an administrative boundary and the necessary information exchange between administrative bodies in this respect becomes easier with the use of Basic Residential Register Network. The Basic Residential Register Law regulates undertakings for this system, how its operation should look and its practical use. Information contained in this system be the four items: name, birth date, gender, and address. It also contains resident card code numbers and information of any changes to them. The resident card code allows easy access to these four data items. It consists of eleven digits, selected at random, and will be issued only in Japan without duplication for any other person. This code can be changed at the request of the subject person. The system constellation is 3-fold in vertical; the levels consist of national, prefecture and municipalities. Depending on the distance between the "clients" making use of the common contents from each other, the option would be made which of the 3 levels of networks should be used. If two towns sit in the same prefecture, the communication server of one of the two towns and the dedicated line in connection with this server on the level of the prefecture would be used in order to communicate with the other town magistrate. If the communication should go to the level of prefecture-prefecture, the network of this constellation level would be working plus the national network involved. Communication servers have firewalls both inside and outside to prevent illegal access (Inoue, 2003; Yoshida, Mizokami & Watanabe, 2003)

ElectronicTender

Here is a case of Yokosuka City in Kanagawa Prefecture (Hirokawa, 2003). Up until the 1997, order placement of civil works in the city took place mainly by public tender in which 7-10 specified and pre-selected bidders could join the tender. However, it became known that there had been some collusions on the bidding. It was considered that reform is necessary. The tender system was changed in the fiscal year 1998 to a conditionally open tender where any business entity could participate, if inspection standards prescribed were satisfied. In the fiscal year 1999 information on order placements started to be shown on the agency's homepage.It became possible for companies wishing to participate to confirm general picture of an order. Where is the site? What is the nature of this work? Firstly, a company wishing to bid will transmit a tender application form to the contracting section by facsimile. If confirmed that this company is qualified to participate, it will purchase design documents from a designated printing company, and draw up an estimate. Then it sends the tender documents poste restante at the Yokosuka Post Office by registered mail with a delivery certificate by the deadline. On the date appointed, the contract section of the magistrate will open all the tender documents collected at that time and chose the successful bidder with representatives of the participating companies witnessing. All the tender results will be publicized on the homepage after that. The effect is that the number of tender companies increased by 2.5 times on average before and after the reform. At the same time the annual average success ratio (the ratio of a successful bid to the expected price) declined from 95.7% before the reform to 85.7%.

In the fiscal year 2001, the series of procedures above were to be carried out by the Internet. The purpose is to further widen the door to bidding companies and save on administrative work of the contract section. Companies put in a tender bid on the tender document transmission screen first and then transmit it by the Internet. The tender documents will at once be transmitted to the server of the authentication bureau, where the guarantee of originality and perception time stamps are added. Then it will be transmitted to the contract section. In other words, the function conventionally carried out at a post office has been transferred to an authentication bureau. The average success bid ratio for this fiscal year fell as low as 84.8%.

Noteworthy is the fact that the reform of the tender system was brought about only with the strong will on the part of the personnel, with the mayor on the top, down to all those who worked on the reform who did not give up facing various forms of resistance.

Electronic Voting

For electronic voting there are two options, voting made from a household PC and another made by ballot. Both are under examination now. The former is carried out where an electronic voting paper is obtained by a household from a voting server. Here, voting contents are filled in and transmitted to a sum-up server. However, a security problem has been pointed out in cases of electronic voting by household. Some people held the view that it should be limited to persons with physical disability. On the other hand, electronic voting at a ballot is carried out in that on voting contents decisions will be made via touch panels. Also, an electronic pen can be used at the voting terminal. This has been put into practice in Niimi City, Okayama Prefecture and part of the Hiroshima City. Various technologies for voting have been proposed, but none of them has been developed enough in any of the business in e-government. The reason appears to be that the frequency of an election itself is low; at most, it is once in a year. Therefore, cost efficiency of system development is not clear as a result. Furthermore, security must be perfect etc. A voting logic where contents and choice of an individual voter will not be known, not only to others but also to the system itself, is under study. The system must cope with bribery and corruption and misuse of strong power, and there remain difficult issues to be solved (Kitagaki, 2003; Shiraishi et al., 2003).

Various Resident Services

In Yokosuka City (Hirokawa, 2003) Geographical Information Service (GIS) was introduced to enhance and reach greater efficiency of administrative services in order to facilitate residents' participation in the regional community activities. Common use on the level of the municipal office was possible by integrating information owned by each division and bureau in the municipal office at one place, except for that information which would violate the protection of personal data. In the fiscal year 2002, Yokosuka City arranged so that citizens could obtain maps using Web-GIS from home. That was a result of an alliance with a map company in the private sector. On the Web site, information on medical institutions, bargain sales at shopping centers, sightseeing information, etc. became available. Many organizations and individuals such as the Chamber of Commerce and Industry, medical associations, SOHO workers etc. are engaged in preparation of these data and operation in cooperation.

In Kochi Prefecture (Ishikawa, 2003), the information super-highway has been in operation since the fiscal year 1997. Further improvement has been sought to accompany the preparation of the administrative networks and the changes in the information environment. In the fiscal year 2001, telecommunication services in the private sector started to be carried out as well, under a new basic conception. The main circuit of the information highway was changed from the previous 50Mbps to 2.4Gbps. Preservation control and stable operation throughout the year were guaranteed. In addition, main circuits were open to ISP (Internet service providers) and businesses in the private sector providing CATV services. They became compatible with IPv6, etc.

Sapporo City in Hokkaido showed its basic policy in the fiscal year 2002 amid the progress of various businesses with information technology in mind. It is considered a "city for cooperation between administration, citizens and enterprises". It should be an ideal state of city management, the city listens carefully to residents' voices. It started to operate a call center business to reflect this in administration. This center deals with various kinds of inquiries from citizens. The city has made experiments with citizens monitoring. Trial operation has further expanded since the fiscal year 2003. In addition, the city attempted to distribute IC cards. For example, it has prepared the environment to be able to use them for all the municipally operated subways.

Further, payment of administrative services with IC credit cards (Endo, 2003), provision of various electronic municipality software developed with the residents' standpoints (Shimada, 2003), as well as technological issues of its operation (Maeda, Okawa & Miyamoto, 2003), are under examination.

DISCUSSION AND FUTURE TRENDS

Smooth application of e-government in the near future depends on the following factors. First, is the convenience of the portal site, especially seen from the user's point of view. How quickly users can come across with the desired information at the portal site of e-government, is important. The electronic portals of the USA show construction standards for the portal of e-governments of each federal organization, state, etc. By doing so, each portal is linked organically to the whole. Compared to the case in the USA, some people point-out, the portal of the Japanese e-government does not give any thoughtful consideration to the integration of the whole. In fact, inconvenience may arise in that the same information must be input every time an electronic application is made to different representative bodies of the government without linkages to other representative bodies of the government.

The second factor is related to the present work contents and review of procedures. Basically, the information processing carried out by each representative body of individual governments has put in old procedures based on financial and accounting laws from the Meiji Era and shift them into electronic databases. Therefore, E-government now needs to reform the whole processing of work data as well as distribution structure, and change them to an effective and integrative work processing system.

The third is related to a greater efficiency of works and systems. In the future, review of excesses and deficiencies of procedures will be necessary. Experts will be required to add a new function conforming to the existing e-government. Experts sought must be fully familiar with both work contents and information technologies, and at the same time be able overview the whole landscape.

The fourth is related to the transfer of these works to the private sector. Outsourcing in the private sector will be necessary in order to raise the efficiency and rate of utilization of the introduced systems, and to reduce costs. At the same time, selection of outsourcing companies and contract contents are important in order for information of administration and individuals not to be leaked.

Lastly, all of the cases mentioned here are on-going projects, thus the relevant evaluation activity has not been yet done. In the near future, verification process ought to be planned.

REFERENCES

Endo, C. (2003). Credit smart card payment using NICSS token method in electronic government. *Information Processing Society of Japan*, *44*(5), 489-493.

Fujisaki E., Ohta, K. & Okamoto, T. (1993). Practical secret voting schemes using polling places. *Institute of Electronics, Information and Communication Engineers*, ISEC 93-24, 7-21.

Hirokawa, S. (2003). Approach to the local e-government. *Information Processing Society of Japan*, *44*(5), 461-467.

Inoue, M. (2003). The basic residential registers network system for e-government and e-local governments. *Information Processing Society of Japan*, *44*(5), 468-472.

Ishikawa, Y. (2003). Toward the implementation of local e-government-A case in Kochi prefectual government. *Information Processing Society of Japan*, *44*(5), 480-483.

Kitagaki. I. (2003). Full confirmation electronic voting model: Countermeasures to forcible or internal illicit improprieties. *Proceedings of the Information Resources Management Association*, 230-231.

Maeda, M., Okawa, Y. & Miyamoto, S. (2003). On the construction of electronic local governments from the standpoint of a technical vendor company. *Information Processing Society of Japan*, *44*(5), 499-502.

Makiuchi, K. (2003). E-Government (2nd version). *Information Processing Society of Japan*, *44*(5), 461-467.

Ohyama, N. (2003). Progress of e-government in Japan. *Information Processing Society of Japan*, *44*(5),455-460.

Segawa, M. (2003). IT turns to realization of e-collaboration city. *Information Processing Society of Japan*, *44*(5), 484-488.

Shimada, H. (2003). For the early realization of electronic local government. *Information Processing Society of Japan*, *44*(5), 494-498.

Yoshida, T., Mizokami, M. & Watanabe, T. (2003). Electronic filing to governments. *Information Processing Society of Japan*, *44*(5), 473-475.

APPENDIX

- The Strategic Headquarters for the Promotion of an Advanced Information and Telecommunications Network Society(Jan.22,2001) e-Japan strategies, [WWW document]. URL http://www.kantei.go.jp/jp/singi/it2/dai1/1siryou05_2.html
- The official residence of the Prime Minister (Feb.12,

2004). the Liaison Conference of Chief Information Officers (CIO) from Representative Bodies of the Government,[WWW document]. URL http://www.kantei.go.jp/jp/singi/it2/cio/index.html

- Administrative Management Bureau, Ministry of Public Management, Home Affairs, Posts and Telecommunications (Feb.16,2004).‰ the general window of e-Government, [WWW document]. URL http://www.e-gov.go.jp/
- Federal Citizen Information Center, Office of Citizen Services and Communications U.S. General Services Administration(Feb.16,2004). *The U.S. Governments Official Web Portal* [WWW document]. URL http://www.firstgov.gov

KEY TERMS

Basic Residential Register Network: A network to confirm the being of a person, a subject. Its use is common throughout Japan and jointly operated by local authorities.

E-Government: The government and local authorities which construct an electronic processing system for administrative procedures with full command of information telecommunication technologies. It provides various kinds of services for residents in line with these.

E-Japan Strategies: The national strategies decided in the first conference of the Strategic Headquarters for the Promotion of an Advanced Information and Telecommunications Network Society (the IT Strategic Headquarters) held on the 22nd January 2001. It was laid on the IT basic strategies decided at the Joint Conference of the IT Strategic Conference and the Strategic Headquarters for Information Telecommunications (IT) Technologies (27th November 2000).

Electronic Tender System: A system to carry out a series of works from notification of order placement information on the homepage etc., application for tender participation, sending a tender document, opening tender documents to the public announcement of the result.

Electronic Voting System: A system to vote electronically, using an information terminal of a personal computer etc. Various technologies have been created to prevent illegal actions such as alteration of voting contents or the use of abstainers' votes.

Online Communication Regulation Law: A law to enhance application in the administration, which is decided by laws exceeding 50,000, and cover all these laws for conducting application electronically. It is considered that this communication regulation law has completed the basic legal frameworks necessary for the e-government such as electronic signature law.

Public Individual Certification Law: A law to prepare a system to provide individual certification and authorization services in order to process on-line the public application and notification on-line.

Public Sector Case Study on the Benefits of IS/IT

Chad Lin
Edith Cowan University, Australia

Graham Pervan
Curtin University of Technology, Australia

INTRODUCTION

While organisations continue to invest heavily in IS/IT, research studies and practitioner surveys report contradictory findings on the effect of the expenditures on organisational productivity (Grover et al., 1998). Therefore, it is not surprising to see that the term "productivity paradox" is gaining increasing notoriety as several studies point toward fairly static productivity and rising IS/IT expenditure. Despite large investments in IS/IT over many years, it has been difficult to determine where the IS/IT benefits have actually occurred, if indeed there have been any (Willcocks & Lester, 1997). The main objective of this article is to generate a deeper understanding of issues surrounding the current practices and norms in managing IS/IT benefits and investments evaluation.

RESEARCH APPROACH

A case study was conducted to investigate the practices of IS/IT investment evaluation and benefits realisation in large Australian organisations. Semi-structured interviews were used to gain a deeper understanding of issues.

CASE DESCRIPTION

Seven interviews were conducted with participants from a Western Australian state government department and two major outsourcing contractors. The questions asked related to the formal benefits realisation methodology used, major outsourcing contracts, contractual relationship between department and contractors, and IS/IT investment evaluation methodology or technique deployed. Other data collected included contract documents, planning documents and minutes of relevant meetings. Around 80 pages of transcripts were coded and analysed. The analysis was conducted in a cyclical manner and followed guidelines for interpretive research set out by Klein and Myers (1999).

CASE STUDY RESULTS

A number of issues arose from the analysis of this text data and the key issues are presented next.

Issue 1: Lack of formal IS/IT investment evaluation methodology

Most of the participants claimed that some sort of formal methodology or process was put in place for evaluating these contracts. However, closer examination revealed that what was described did not constitute a formal IS/IT investment evaluation methodology. Participants wrongly considered various contract control mechanisms as a formal IS/IT investment evaluation methodology.

Issue 2: A formal IS/IT benefits realisation methodology was used

Every participant was aware that a formal IS/IT benefits realisation methodology was being used for outsourcing contracts and projects. A benefits realisation approach was used as an end-to-end process to assist in: (1) providing a rigorous process to select the right projects; (2) placing responsibility and accountability at the appropriate level; (3) driving process re-engineering through changes in the organisation; (4) ensuring benefits are realised; and (5) ensuring agreed re-investment of time savings applied as expected.

Issue 3: Lack of understanding of IS/IT investment evaluation methodology

The confusion indicated in Issue 1 about what constitutes a formal IS/IT investment evaluation methodology demonstrated a lack of understanding of such methodologies. This may be due to the fact that the department was unable to introduce a formal IS/IT investment evaluation methodology because it was required to follow the state government's outsourcing guidelines (MOPC, 2000; SSC, 1999b).

Issue 4: Existence of an informal IS/IT investment evaluation process

Despite the fact that no formal IS/IT investment evaluation methodology or process was used, contract control and evaluation mechanisms specified within the SLAs or government guidelines do represent an informal IS/IT investment evaluation process. Although these informal mechanisms may not totally replace a formal methodology (e.g., Kaplan and Norton's (1992) Balanced Scorecard), they were able to assist in evaluating the performance of the outsourcing contracts. These mechanisms were largely based on the standard state government contract process and purchasing guidelines (SSC, 1999a, 1999b).

Issue 5: Good understanding of benefits realisation practices

A benefits realisation methodology, called benefits realisation approach, was introduced before outsourcing IS/IT functions because there was a concern that IS/IT investments did not deliver value (which would jeopardize future government funding). To ensure that the IS/IT investments deliver the promised value and benefits as well as bring the focus back to the department's main business, a large internal change program was required.

Issue 6: Focus on quantitative IS/IT investment evaluation measures

Without employing more qualitative measures (e.g., relationship, culture and leadership) and a formal IS/IT investment evaluation methodology or process, the use of quantitative or accounting-based measures alone did not assist in full evaluation and monitoring of the performance.

Issue 7: Different motivations for outsourcing

Several reasons were put forward as the main motivation for IS/IT outsourcing. Only two of the four contractor representatives cited access to the required technical expertise as one of the department's reasons to outsource and two either did not know or did not respond to the question. However, all of the department's participants mentioned access to required technical expertise as a major reason to outsource some IS/IT functions. Therefore, the department's motivation for outsourcing was somewhat different from the contractors.

Issue 8: Success of the contracts perceived differently by stakeholders

Customer satisfaction, achieving the contractor's projected revenue, bringing value/benefits to the organisation, and meeting the SLA provisions were mentioned. Other criteria mentioned included technical competence to deliver what is required, risk factors, contractors' experience in a relevant area, and business continuity of the contractors. Both representatives from the first contractor mentioned achieving the projected revenue for themselves and satisfying customers as their only criteria for determining the success of their outsourcing contracts with the department. This may indicate that the first contractor's aim is to maximise the profit while maintaining a certain level of customer satisfaction. However, participants from the department seemed to have used different criteria for determining the success of the outsourcing contracts. Bringing value/benefits to the organisation, meeting the SLA provisions, and pricing/cost were mentioned by three out of four participants.

Issue 9: Better control over the IS/IT skill shortage within the department

Access to the required technical expertise was the most often cited reason for outsourcing. The department did not have the required IS/IT expertise to implement a major internal change program and outsource some of its IS/IT functions. In order to obtain the required technical expertise and skills from outside, the department had to transfer some of its IS/IT staff to the first contractor. According to most of the participants, those who went across to the first contractor were quite happy about the whole process. However, the department did not lose the capacity to manage and assess its own IS/IT needs after relinquishing the control of some of its IS/IT functions and staff, as has happened in many outsourcing organisations previously (Currie & Willcocks, 1998).

Issue 10: Embedded contract mentality

Staff of the department seemed to have a "contract mentality," as the operation of the contracts was based on the specifications set out in the SLAs within the outsourcing contracts. Several participants clearly indicated that there was a pre-agreed set of evaluation and control mechanisms in the SLAs within the outsourcing contracts such as metrics, monthly reports, reviews, and regular meetings. Moreover, half thought these contract control mechanisms were all part of the IS/IT investment evaluation methodology.

Issue 11: Lack of user involvement/participation in contract development

There appeared to be an organisational memory gap where units within the department possessed knowledge of different parts of the IS/IT systems development cycle. However, the knowledge did not seem to be shared by all units because different units participated in different stages. The department's whole outsourcing process may have been more successful if the participants were involved in both original tendering and outsourcing contracts negotiation as well as benefits realisation process.

Issue 12: Conflict between motivations and success criteria for outsourcing

It seemed that the participants had different expectations regarding the outsourcing as a whole and the outsourcing contracts. Alternatively, they may have felt that the outsourcing contracts had already brought in the required technical expertise and should not be used as a criterion for determining contract success

Issue 13: General lack of commitment by contractors

It was easy to see that department and contractors had different agendas in mind, despite the fact that the contract was a partnership-type arrangement. The first contractor's criteria for determining success of outsourcing contracts seemed to be maximisation of profit/revenue while keeping the customers satisfied. The contractors' lack of commitment could also be demonstrated by the fact that they either did not know why the department outsourced, or could not agree on a single reason for outsourcing. However, the department agreed that access to required technical expertise was a reason for outsourcing.

Issue 14: Restrictive government outsourcing contract guidelines

All state departments have to follow the contract tendering and outsourcing guidelines set out by the State Supply Commission (SSC, 1999a, 1999b). This was one reason why the department failed to adopt a formal IS/IT investment evaluation methodology. The state government guidelines specify numerous principles, steps, and procedures for state departments to follow when dealing with external contractors. However, they do not determine what formal IS/IT investment evaluation methodology a department should use.

Issue 15: Benefits realisation methodology provides rigorous processes

As mentioned in Issue 1, the benefits realisation approach provided the department with rigorous processes for achieving benefits delivery. These processes did not exist prior to its introduction. Instead, some crude methods had been used in an attempt to realise benefits within the organisation. These methods included budget cuts, projects statements presentations, and activity surveys. The introduction of the benefits realisation approach provided the department with rigorous processes.

Issue 16: Ability to manage the outsourcing contracts without external influence and assistance

To acquire technical expertise and skills, the department had to outsource some of its IS/IT functions and transferred some of its IS/IT staff to the first contractor. However, unlike other outsourcing organisations, the department appeared able to manage its outsourcing contracts internally without external influence or assistance. This may be because the department had transferred as few IS/IT staff to the first contractor as possible and, simultaneously, benefited from the second contractor's IS/IT technical expertise obtained under the contract.

CONCLUSION

This case study was conducted in a large government department with a mix of insourced and outsourced IS/IT activities. While the department appears to operate without any major problem, the mostly negative issues shown previously indicate weaknesses in the way it deals with the level of formality and integration in applying the methodologies. The problems mentioned in Issues 6 -14 were mostly caused by the lack of attention to IS/IT investment evaluation (Issues 1, 3, and 4). For example, if formal IS/IT investment evaluation was adopted by the department, more qualitative measures may been used to evaluate the outsourcing contracts (Issue 6).

So why didn't the department formally evaluate its IS/IT investments? One possible explanation was that the restrictive nature of the state government's outsourcing contract guidelines (Issue 14) made it difficult to implement a formal IS/IT investment evaluation methodology. Another explanation was that none of the IS/IT staff was familiar with the formal IS/IT investment evaluation process and hence possessed an "embedded contract mentality" (Issue 10) by simply following conditions set out within the SLAs. Seddon et al. (2001) suggest that under some circumstances, cost of formal IS/IT evaluations may seem likely to exceed benefits. However, several positive issues (Issues 9, 15, and 16) from this case study indicate that the use of a benefits realisation methodology enabled greater control and better management of outsourcing contracts.

Despite large investments in IS/IT over many years, it has been difficult for organisations to determine where benefits have occurred, if indeed there have been any. IS/IT investment evaluation practice remains a hotly debated topic in the IS literature. Little published work has been conducted in Australia and there is still a lot to be learned in the area of processes and practices of IS/IT investment evaluation and benefits management. We hope that more studies of the practice of IS/IT investment evaluation will benefit other researchers in this field and the business community as a whole. Through the research program introduced in this article it is hoped that better approaches may be developed for Australian organisations.

REFERENCES

Currie, W.L., & Willcocks, L.P. (1998). Analysing four types of IT sourcing decisions in the context of scale, client/supplier interdependency and risk mitigation. *Information Systems Journal, 8,* 119-143.

Grover, V., Teng, J., Segar, A.H., & Fiedler, K. (1998). The influence of information technology diffusion and business process change on perceived productivity: The IS executive's perspective. *Information and Management, 34,* 141-159.

Kaplan, R.S., & Norton, D.P. (1992, January-February). The balanced scorecard - Measures that drive performance. *Harvard Business Review,* 71-79.

Klein, H.K., & Myers, M.D. (1999). A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS Quarterly, 23*(1), 67-94.

MOPC. (2000, July). *Competitive tendering and contracting: Process guidelines for use by agencies* (3rd ed.). Ministry of Premier and Cabinet, Government of Western Australia.

Seddon, P., Graeser, V., & Willcocks, L. (2001, September 17-18). IT evaluation revisited: Plus change. *Eighth European Conference on Information Technology Evaluation*, Oriel College, Oxford, UK (pp. 1-18).

SSC (State Supply Commission). (1999a, April). *Policies and guidelines: A pocket guide for buying wisely*. Perth, Western Australia.

SSC (State Supply Commission). (1999b, April). *The contract process*. Perth, Western Australia. *http://www.ssc.wa.gov.au/pub/index.html*

Willcocks, L., & Lester, S. (1997). Assessing IT productivity: Any way out of the labyrinth? In L. Willcocks, D.F. Feeny & G. Islei (Eds.), *Managing IT as a strategic resource* (ch. 4, pp. 64-93). London: The McGraw-Hill Company.

KEY TERMS

Benefits Management: A managed and controlled process of checking, implementing and adjusting expected results and continuously adjusting the path leading from investments to expected business benefits.

Benefits Realisation: See Benefits Management.

IS/IT Benefits Realisation Methodologies: Approaches that are used to ensure that benefits expected in the IS/IT investments by organisations are realised or delivered.

IS/IT Investment Evaluation: This is the weighing up process to rationally assess the value of any acquisition of software or hardware that is expected to improve business value of an organisation's information systems.

IS/IT Investment Evaluation Methodologies: Approaches that are used to evaluate organisations' IS/IT investments.

IS/IT Outsourcing: The process of transferring IS/IT assets, staff, and management responsibility for delivery of services from internal IS/IT functions to external contractors.

Productivity Paradox: Despite large investments in IS/IT over many years, there have been conflicting reports as to whether or not the IS/IT benefits have actually occurred.

Public-Key Cryptography

Eberhard Stickel
Hochschule der Sparkassen - Finanzgruppe
University of Applied Sciences, Bonn GmbH, Germany

INTRODUCTION

Secure transmission of private information is a crucial issue in today's highly computerized world. Cryptographic algorithms are used to provide privacy of message transmission and to certify authenticity of sender and/or receiver of a message and message integrity as well as undeniability of transmitted messages (Menezes et al., 1996).

The message that is sent by the sender is called the plaintext, the secured message is called the ciphertext. To get the ciphertext, the plaintext is encoded by the sender. The receiver reconstructs the plaintext from the ciphertext by decoding. For encoding and decoding so-called keys are used (Koblitz, 1994, p. 55).

In the simplest setting, sender and receiver have agreed on a common private key, which is kept secret. This is called symmetric key cryptography. The secret private key is utilized for encoding and decoding messages sent between the two parties. For encoding the plaintext is XOR-ed with the secret key. The decoding is done in the same way using the ciphertext and the secret key on the receiver's side (Menezes et al., 1996, p. 15).

To be specific, if for example the plaintext message is given by the bit string 10011101 and the secret key is 11011100, the ciphertext is then given by 01000001. By using XOR-operations with ciphertext and key once more the plaintext is returned. Note that the XOR operation between two bit is defined as follows: $0 + 0 = 1 + 1 = 0$, $1 + 0 = 0 + 1 = 1$.

If key length and length of plaintext do not coincide, the plaintext may be blocked. This leads to the concept of block ciphers (Menezes et al., 1996, p. 223).

Symmetric key techniques generally can be implemented very efficiently. The corresponding algorithms are very fast. The problem, however, lies in the fact that two parties must have agreed on a common key before they can start to communicate and exchange messages. This would be highly impractical, for example, for transactions in electronic commerce (Menezes et al., 1996, p. 31).

BACKGROUND

A major breakthrough was the publication of the Diffie-Hellman key exchange scheme (Diffie & Hellman, 1976). The technique developed by Diffie & Hellman allows one to agree on a secret key through an insecure channel,for example, the Internet. The authors rely on a problem that is hard to solve, at least using today's knowledge and computing power: Let p be a large prime or the power of a large prime. Let g be a number with $1 < g < p$. g and p are publicly known. Given $g^a \bmod p$ and $g^b \bmod p$, compute $g^{ab} \bmod p$. This is the so-called Discrete Logarithm problem. Here mod refers to division with remainder. $a \bmod b$ is the remainder, if a is divided by b. Details may be found in Menezes et al. (1996, p. 515). The security of the method relies on the fact that it is impossible to solve the sketched problem in feasible time if the prime p is large enough. No efficient algorithms for solving the Discrete Logarithm problem are yet known.

Suppose that Alice and Bob want to agree on a secret key over an insecure channel. The Diffie-Hellman method runs as follows:

- Alice has published p, g and $g^a \bmod p$ as her public key. a is kept secret as private key.
- Bob chooses b which is kept secret, forms $g^b \bmod p$ and $(g^a)^b \bmod p$ and submits the latter number to Alice.
- Alice computes $(g^{ab})^{1/a} = g^b \bmod p$. This is the secret key exchanged.

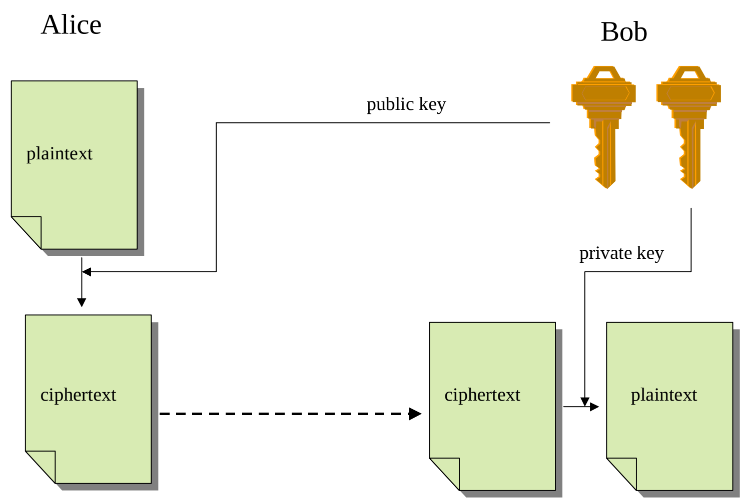
Note that various components of the algorithm are now publicly known. It is therefore called a public-key algorithm.

Various public-key algorithms are discussed in the literature and employed in practical applications. Probably the most famous algorithm is the RSA-algorithm, denoted by RSA after its founders Rivest, Shamir, & Adleman (Rivest et al., 1978). *Figure 1* shows the details.

RSA relies on the problem of factoring large numbers, for example, the product of two 1024-bit prime numbers. No algorithm exists so far that allows one to solve such a problem in feasible execution time. The basic idea of RSA is the following:

- Suppose Alice wants to transmit a secret message to Bob. Alice gets Bob's public key,for example, by accessing his Internet site.
- Alice encrypts the plaintext using the public key of Bob. Then she transmits the ciphertext to Bob.

Figure 1. Public-key cryptography



- Bob uses his secret private key to decrypt the ciphertext and recover the plaintext.

Generally public-key algorithms require a pair of keys. The public key is published, while the private key is kept secret. The public key is used for enciphering a message, while the private key is used to decipher the ciphertext.

Applying public-key methods for transmitting secret data is not efficient. The underlying algorithms are much slower and more complex than symmetric key methods (Menezes et al., 1996, p. 31). Typically a public-key algorithm is used to agree on a symmetric session key. During the session, data are protected by encrypting plaintext with the symmetric key obtained. This approach combines the advantages of symmetric and public-key technology.

Note, that there is still a problem that needs to be solved. It is necessary to assure the authenticity of the public key of a person, for example, of Bob's in the explanation of the RSA-algorithm. Otherwise, it is possible that a potential third party adversary pretends to be Bob (bogus identity). He would provide his public key to Alice, Alice would encrypt the plaintext using this key and the adversary would be able to read the secret message that was addressed to Bob. To solve this problem, trust-centers that certify the authenticity of public keys have been implemented (Menezes et al., 1996, pp. 491, 547, 581).

DIGITAL SIGNATURES

Public-key algorithms as discussed in the last section allow one to encipher secret messages. Thus, they provide privacy of information transmission. Especially in electronic commerce it is necessary to assure authenticity of messages as well as of communication partners. This is done by means of digital signatures (*Figure 2*).

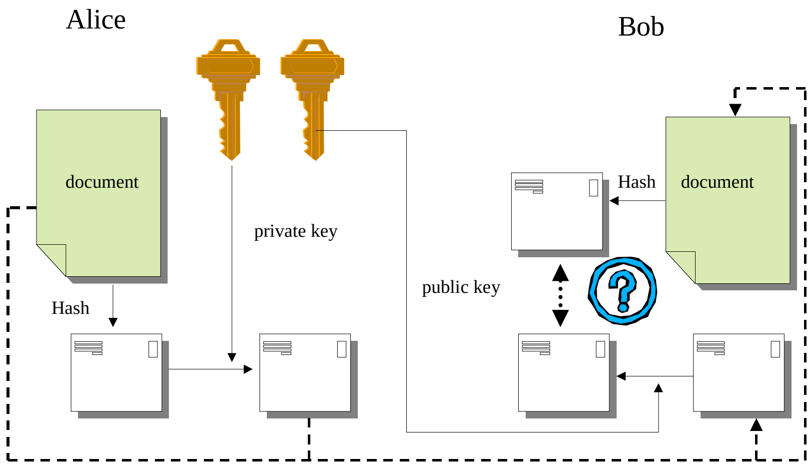
Fortunately, most of the well-known public-key techniques need only slight modifications in order to be used as digital signatures. The basic idea is as follows (Zhang & Wang, 2003, p. 24):

- Suppose Alice wants to sign a message *m* and forward it to Bob. First, Alice computes a so-called digital fingerprint *d* of the message *m*. This fingerprint usually consists of a bit string of a certain length that corresponds to the message *m*. Alice signs *d* with her private key. This means that Alice encrypts *d* using her secret private key.
- Bob fetches the public key of Alice and decrypts the signature. He thus holds *d* as well as the message *m*. A digital fingerprint *d'* is computed from the transmitted message. If $d = d'$, he accepts message and signature, otherwise the transaction is rejected.

Digital fingerprints are computed using so-called hash functions. A hash function compresses a message to a bit string of specified length. Hash functions should be collision resistant and not invertible. Collision resistance implies that two different (meaningful) messages should have different fingerprints (hash values). Lack of invertibility assures that it is infeasible to construct a meaningful message given a specified digital fingerprint. For details refer to Goldreich (2001, p.136).

Note, that in the process of signing the roles of private and public key are simply interchanged.

Figure 2. Digital signatures



If the two digital fingerprints computed do not coincide, the message has been corrupted or the sender did not provide a valid signature. In either case it is correct to reject the transaction.

Various signature schemes are discussed in the literature. Popular candidates are DSA (Digital Signature Algorithm), which is based on RSA (Menezes et al., 1996, p. 452) or ElGamal signatures (Menezes et al., 1996, p. 454).

Again, it is important to certify the correctness of the public key of the signer. As already indicated in the previous section this is generally done by trusted third parties. Such trust centers provide certificates signed by their digital signatures. Their signatures are certified in turn by trusted parties on a higher hierarchic level and so on (Li & Wang, 2003, p. 39).

Of course, it is possible to combine secure message transmission with digital signatures. In this case, sender and receiver need to possess a pair of private and public keys. The sender encrypts the message using the public key of the receiver. He also signs a fingerprint of the message using his private key. The receiver decrypts the message using his private key. He also decrypts the digital fingerprint using the public key of the sender. He then computes a digital fingerprint of the decrypted message and compares it to the decrypted fingerprint received. In case of equality the transaction is accepted, otherwise it is rejected.

FUTURE TRENDS

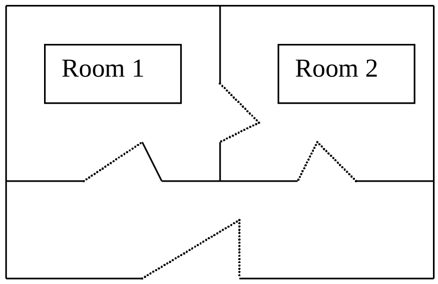
In many applications it is necessary to identify potential users to providers of resources (typically hosts or servers). If the identification is successful, access is granted. Passwords still are widely used for identification purposes. The use of passwords, however, has severe drawbacks (Menezes, 1996, p. 388). The password is usually transmitted. This allows password sniffing by potential adversaries. There is also potential hazard that users act carelessly with their passwords. One example is the use of meaningful, easy to guess passwords. The password is usually stored in a file located on the server. This password file may be attacked and corrupted. Various possibilities exist to improve on these potential weaknesses. In any case, however, meaningful data for identification purposes have to be transmitted. This is why passwords are considered as weak authentication techniques (Menezes et al., 1996, p. 388).

Challenge and response techniques provide a higher level of security. Server and client generally share a common secret key. The server randomly generates a bit string and sends it to the client (challenge). The client encrypts the bitstring using the shared secret key and resends the ciphertext to the server (response). The server checks by decrypting if the correct key was used (Menezes, 1996, p. 397).

Challenge and response identification does not require transmission of secret information (passwords). It still requires storage of a secret common key on the server, however. The corruption of the corresponding file is a potential hazard.

To summarize, it is desirable that secret information is neither transmitted nor stored on the server side. Zero-knowledge authentication schemes have been developed for this purpose (Goldreich, 1999, p. 12; Menezes et al., 1996, p. 405). A completely non-mathematical treatment of this rather quantitative subject can be found in Quisquater et al. (1990).

Figure 3. Illustration of zero-knowledge protocols



Zero-knowledge protocols allow one to prove to a third party that someone knows a secret without revealing this secret. The illustration in *Figure 3* will be used to motivate the key ideas behind this approach.

Figure 3 shows a house with an entrance hall and two rooms that may be accessed via the entrance hall. There is a door between the two rooms, which is generally locked. Only Alice has the secret key to open the door between these two rooms.

Suppose now that Alice wants to convince Bob that she has the key (i.e., that she knows a secret) without showing the key to Bob or letting Bob go into the two rooms (i.e., without revealing her secret).

At the beginning, Alice and Bob stand in front of the closed door of the entrance hall. Alice proceeds inside and closes the door. She then randomly goes into the left or the right room. Let us assume that in the first round she chooses the left room. Bob enters the hall and randomly chooses a bit b. If $b = 1$ he orders Alice to step out of the right door, otherwise he orders Alice to step out of the left door. In case $b = 0$, Alice may easily step out of the door without using her secret key. In case $b = 1$, she has to use the secret key enabling her to step out of the right door.

If Bob's choice is purely random, Alice has a chance of exactly 50% of stepping out of the correct door without having to reveal her secret. Thus, she may cheat with a probability of 50%. Conversely, Bob knows with a probability of 50% that Alice indeed possesses the secret key.

In order to reduce the probability of cheating this experiment has to be repeated for a suitable number of times. After 10 successful trials for example the probability of cheating is less than 0,1% (exactly $2^{-10} = 1/1024$). Thus, it is possible to reach arbitrary high levels of security that Alice knows the secret. Absolute security, however, can never be gained. This is why the approach sketched is called a probabilistic approach.

The idea sketched in *Figure 3* may be formalized in order to get an executable protocol. Usually the Fiat-Shamir protocol is applied (Fiat & Shamir, 1987). The security of this protocol relies on the difficulty of extracting discrete square roots. Let n be a composite number (e.g., the product of two distinct large prime numbers). The problem of computing x if $x^2 \bmod n$ is known is a computationally hard problem (comparable to the Discrete Logarithm problem and the problem of factoring large integers). The secret used in this protocol is the knowledge of a square root of a known integer $y \bmod n$. When executing the protocol a third party should be convinced that such a square root is known without revealing the square root.

CONCLUSION

Public-key cryptography is widely used in today's computerized world. Public-key algorithms allow to secure the privacy of communication and to assure the integrity of exchanged messages and the identity of the sender of a message (digital signatures). They may also be used for solving the identification problem between clients and servers offering valuable resources. Public-key algorithms rely on the existence of computationally hard problems such as the Discrete Logarithm problem. If the underlying problem size (key size) is large enough, it is infeasible to break these algorithms using today's knowledge and computing power. However, knowledge and computing power are permanently improving so that consequently the key size needs to be increased continuously. Public-key techniques are also applied in the area of electronic payment systems (Kou, 2003).

An excellent more detailed summary of current techniques and research in this area may be found in Delfs & Knebl (2002) or Biham (2003).

REFERENCES

Biham, E. (2003). *Advances in cryptology*. EUROCRYPT 2003. LNCS 2656. Berlin: Springer.

Delfs, H., & Knebl, H. (2002). *Introduction to cryptography*. Berlin: Springer.

Diffie, W., & Hellman, M.E. (1976). New directions in cryptography. *IEEE Transactions on Information Theory*, *22*, 644-654.

Fiat, A., & Shamir, A. (1987). How to prove yourself: Practical solutions to identification and signature problems. In *Advances in Cryptology: Crypto 86*. Lecture Notes of Computer Science 263 (pp. 186-194). Berlin: Springer.

Goldreich, O. (1999). *Modern cryptography, probabilistic proofs and pseudorandomness*. Berlin: Springer.

Goldreich, O. (2001). *Foundations of cryptography*. Basic Tools. Cambridge: Cambridge University Press.

Koblitz, N. (1994). *A course in number theory and cryptography* (2nd ed.). Berlin: Springer.

Kou, W. (2003). *Payment technologies for e-commerce*. Berlin: Springer.

Li, H., & Wang, Y. (2003). Public-key infrastructure. In W. Kou (Ed.), *Payment technologies for e-commerce* (pp. 39-70). New York: Springer.

Menezes, A.J., van Oorschot, P.C., & Vanstone, S.A. (1996). *Handbook of applied cryptography*. New York: CRC Press.

Quisquater, J.J., Guillou, L., & Berson, T. (1990). How to explain zero-knowledge protocols to your children. In *Advances in Cryptology-Crypto 89*. Lecture Notes of Computer Science 434 (pp. 102-109).

Rivest, R.L., Shamir, A., & Adleman, L.M. (1978). A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM, 21*, 120-126.

Zhang, F., & Wang, Y. (2003). Security fundamentals. In W. Kou (Ed.), *Payment Technologies for E-Commerce* (pp. 7-38). New York: Springer.

KEY TERMS

Symmetric-Key Cryptography: Sender and receiver agree on a common secret key. This key is used to exchange encrypted messages.

Public-Key Cryptography: Public-key cryptography allows the exchange of encrypted messages without prior key agreement. The receiver of an encrypted message has a secret private key and a published public key. The public key is used by anyone who wishes to send a message to the receiver. Only by means of the private key of the receiver, can the original message be retrieved by decoding.

Trust Center: It is important to assure the identity of the owner of a public key. Otherwise potential adversaries may pretend to have another identity by using their own public keys in confidential transactions. Trust centers certify the identity of owners of public keys.

Digital Signatures: Digital signatures may be used to electronically sign documents. The signature can be checked by anyone and if properly used assures integrity of the signed message, as well as the identity of the signer.

Digital Fingerprints: Larger documents are not signed in its entirety. Generally a digital fingerprint of the larger document is computed and signed. For this purpose hash functions are used. They should be collision resistant and not invertible. Then two different meaningful messages have different fingerprints and it is not possible to construct a meaningful message, given a fingerprint.

Challenge and Response: This is a cryptographic technique used to identify a client to a server. Server and client share a common secret key. The server sends bit strings (challenges) to the client who encrypts these strings and resends them to the server (response). The server may then check the correctness by applying the secret key and comparing the result to the original bit string.

Zero-Knowledge Protocol: This is another cryptographic technique to provide identification to a third party. For this purpose and in contrast to other techniques no secret information is transmitted between the communicating partners. Nor is any secret information stored on the third parties server. Zero-Knowledge protocols try to prove that the client knows a secret without revealing this secret to a third party.

Computationally Hard Problems: A mathematical problem is considered computationally hard if a slight increase in problem size dramatically increases the time for solving the problem. Typically, runtime is an exponential function of problem size for computationally hard problems. A lot of cryptographic applications are based on computationally hard problems like the Discrete Logarithm problem, the Discrete Square Root problem and the problem of factoring large integers.

Q

QoS-Oriented MAC Protocols for Future Mobile Applications

Alexander Markhasin
Siberian Telecommunications University, Russia

Stephan Olariu
Old Dominion University, USA

Petia Todorova
Fraunhofer FOKUS, Germany

INTRODUCTION

Common intended access to a physically unshareable resource requires arbitration. Since in wireless communications the transmission medium cannot be shared without loss of information, *medium access control* (MAC) plays the role of an arbiter stipulating, for each of the intended users, the conditions and parameters of exclusive access rights. The major challenge in designing a good MAC protocol is to provide each user with the negotiated quality-of-service (QoS) parameters while, at the same time, utilizing the network resources as efficiently as possible. Thus, an ideal MAC protocol must strike a balance between the conflicting goals of QoS provisioning, efficiency, simplicity, and service controllability. The MAC protocol improvements ensure a dominant effect on the development of the QoS and other future breakthrough wireless technologies.

BACKGROUND

Future generation of global wireless and satellite multimedia communications for mobile multi-service applications requires dynamic (soft) control and stringent QoS guarantees. The MAC technology employed depends on the bandwidth allocation strategy, the scheduler algorithms used, and on the MAC protocol itself. As showed in Markhasin, Olariu, and Todorova (2004), QoS-oriented MAC technology, along with the all-IP/ATM architecture, is expected to provide cost-efficient broadband multi-service to mobile and geographically distributed users anytime and anywhere.

Assume that at some time t the system contains N_t mobile users belonging to k data priority service classes, with input multimedia traffic intensities G_{ikt}, output traffic intensities S_{ikt}, and total input intensity $G_t = \sum_{i,k} G_{ikt} \le C_{MAC}$, where C_{MAC} is the MAC protocol efficiency, and $C_{MAC} = \max_{\{G_t,t\}} \sum_{i,k} S_{ikt}(G_{ikt})$. A *QoS-oriented MAC* protocol must guarantee for each point in time t belonging to some interval $[t_1, t_2)$ the following required values:

a) QoS characteristics: average delay, loss probability, throughput;
b) traffic parameters: peak and average rates; and
c) bandwidth: the amount of bandwidth necessary to support the negotiated QoS.

Perhaps the most natural and straightforward arbitration strategies used in MAC is token based. Specifically, an artifact known as the *token* regulates access to the transmission medium. The arbiter allocates the token to users according to a scheduling strategy. The user that possesses the token has access to the resource; once the token is relinquished, the access right to the resource is lost. Many multimedia applications require effective control of the *token requests generation rate* variation, including token bank fair queuing (TBFQ) proposed in Wong, Zhu, and Leung (2003), recurrent M-sequences RS-token broadcast reservation (TBR-RS) proposed by Markhasin (1996), and also scheduler algorithm *priority parameters* regulation, local queue *buffer size*, and *ARQ quantity* of automatic request for repeat adjusting, as also shown in Markhasin (1996).

One can define static (*hard) QoS provisioning* as a session with static control parameters satisfying predetermined traffic profiles $G_{ikt} = const(t)$. Bandwidth resource $Y_{it} = const(t)$, ...$t \in [t_1, t_2)$ and other control parameters are regulated at long-term intervals only. Therefore hard QoS out-of-profile traffic is not guaranteed, so that performance degradation is possible (Wong et al., 2003).

Similarly, one can define *soft QoS provisioning* as a session with dynamic controlled/adapted QoS characteristics according to traffic profiles variation $G_{ikt} = \mathrm{var}(t)$. Bandwidth resource $Y_{it} = \mathrm{var}(t)$, $\ldots t \in [t_1, t_2)$ and other control parameters are regulated at short-term intervals (Markhasin, 1988).

According to Fattah and Leung (2002) and Wong et al. (2003), existing scheduler algorithms can be classified as *work-conserving* with non-priority service disciplines and *non-work-conserving* using priority service disciplines.

In order to develop QoS-oriented MAC protocols for all-IP/ATM mobile applications, Markhasin (2001) shows that one has to overcome three main impediments:

- the time barrier: concerning the degradation of long-delay MAC efficiency when the round-trip time is significantly increased;
- the dynamical barrier: concerning the instability of the dynamic on-the-fly control of QoS provisioning; and
- the economic barrier: concerning the unacceptably large cost of the wireless broadband ATM networks with centralized architecture.

A SURVEY OF MAC PROTOCOLS

The main goal of this section is to provide an overview of various proposed MAC protocols and to compare them in terms of their ability to provide mobile users with QoS guarantees. Historically, *ALOHA* was the first MAC protocol offered for packet radio networks proposed (Abramson, 1970). It is classical *free (random) access* protocol. Unfortunately, the maximum throughput of this protocol is only 1/2e. Some two years later, the well-known *slotted ALOHA*, which doubles the efficiency of ALOHA, was offered (Roberts, 1972). The next fundamental contribution was the development of several *carrier-sense* protocols, known generically as CSMA (Kleinrock & Tobagi, 1975). Some multi-access methods, later referred to as slotted and non-slotted ALOHA and CSMA, were proposed and analyzed (Markhasin, 1969). There also the *slotting effect* about protocol efficiency increasing was opened. Rubin (1979), Tobagi (1980) and others show that the majority of other classical *free*, *controlled*, and *hybrid* access methods were also proposed in the 1970s-1980s.

The first generation of MAC protocols was not QoS-aware. Some adjusting mechanisms were used for adapting the time intervals division mainly to burst data traffic. Differentiation and service quality control problems were raised in the 1980s for integrated packet radio and satellite networks (Rubin, 1979; Li, 1984; Markhasin, 1985). The development of asynchronous transfer mode (ATM), along with multimedia information technologies and *m-applications*, triggered an interest in guaranteed QoS control. Wong et al. (2003) noted that much research has been focused on hard QoS provisioning of integrated services. For *short-delay wireless media*, *hard QoS-aware* demand assignment protocols SFRs type, including DQRUMA (Karol, Liu & Eng, 1995) and others (for an overview, see Gumalla & Limb, 2000). The TBQF scheduling scheme for wireless packet-switched networks, which provides *soft QoS* guarantees for heterogeneous traffic, was also proposed (Wong et al., 2003).

The soft QoS-oriented multifunctional medium access control (MFMAC) technology for *long-delay satellite mediums* was proposed by Markhasin (2001), who also introduced MFMAC technology based on soft QoS-aware RS-token broadcast reservation (TBR-RS) MAC protocol (Markhasin, 1996). This protocol uses the recur-

Table 1

a) MAC classification matrix with protocol examples

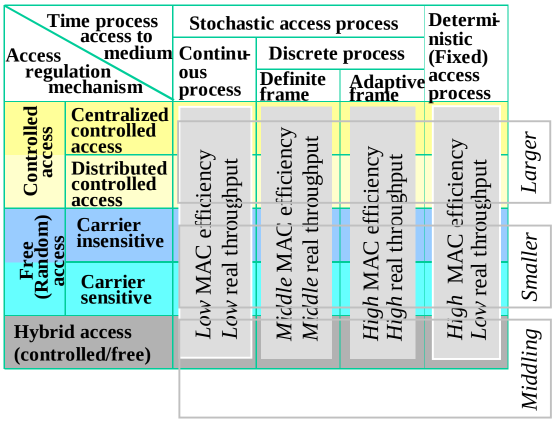
Access regulation mechanism \ Time process access to medium		Stochastic access process: Continuous process	Stochastic access process: Discrete process – Definite frame	Stochastic access process: Discrete process – Adaptive frame	Deterministic (Fixed) access process
Controlled access	Centralized controlled access	Non-slotted Polling	DAMA, FPODA, ARDA, SFR	Slotted Polling	S-TDMA, PAMA
	Distributed controlled access	Non-slotted Token-passing	Superframe Reservation (SFR)	RS-Token Broadcast Reservation, MLMA	JTIDS
Free (Random) access	Carrier insensitive	Pure ALOHA	Slotted ALOHA		
	Carrier sensitive	Non-slotted CSMA	Slotted CSMA (v_0 =1)	Slotted CSMA, Ethernet	
Hybrid access (controlled/free)		ALOHA/ Polling	DSA_{++},GPRS DQRUMA, TBFQ	Reserved ALOHA, MASCARA	

b) The examples of MAC protocol short-delay efficiencies

Access regulation mechanism \ Time process access to medium		Stochastic access process: Continuous process	Stochastic access process: Discrete process – Definite frame	Stochastic access process: Discrete process – Adaptive frame	Deterministic (Fixed) access process
Controlled access	Centralized controlled access	$1/(1+v_{large})$	$1/(1+v_s N/J)$	$1/(1+v_{middle})$	$1/(1+v_{small})$
	Distributed controlled access	$1/(1+v_{middle})$	$1/(1+v_s N/J)$	$1/(1+v_{small})$	$1/(1+v_{small})$
Free (Random) access	Carrier insensitive	$1/2e$	$1/e$		
	Carrier sensitive	$1/(1+a_{cf} w_{small})$ $a_{cf} > 1$		$1/(1+a_{af} w_{small})$ $a_{cf} > a_{af} > 1$	
Hybrid access (controlled/free)		$\frac{1}{1+a_{ch} w_{large}}$ $a_{ch} > 1$	$\frac{1}{1+a_{dh} w_m N/J}$ $a_{dh} > 1$	$\frac{1}{1+a_{ah} w_{middle}}$ $a_{ah} > 1$	

Table 2

2a) Short-delay efficiency



2b) Long-delay efficiency (*Time barrier* overcoming)

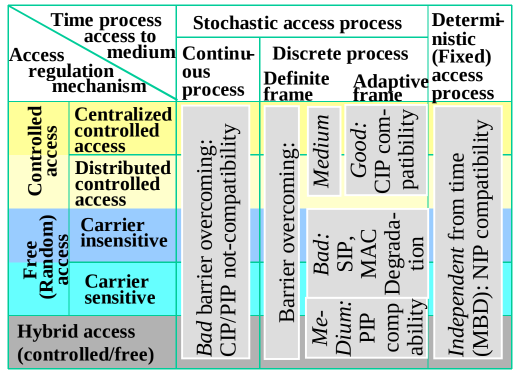
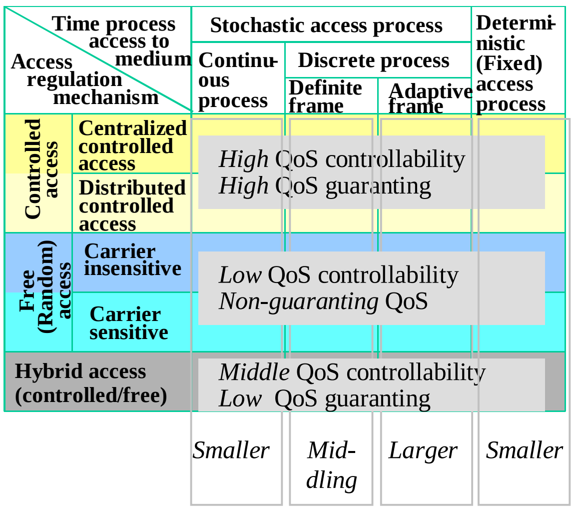
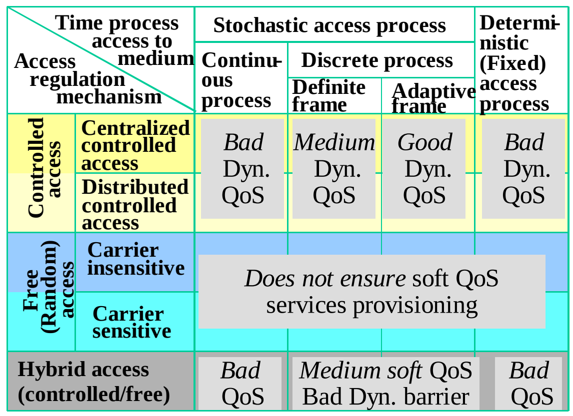


Table 3

3a) Static QoS Control



3b) Soft QoS Control (dynamical barrier overcoming)



rent M-sequences (RS) MAC-addressing opportunities to organize highly effective multiple access to long-delay space mediums. Based on TBR-RS MAC protocol, the multifunctional broadband ATM Spacemedium Hyperbus Architecture for satellite multimedia system was developed and presented (Brand et al., 2001). A satellite communication *scheduler problem* consisting of reservation and allocation of network resources to individual connections before data transmission, based on the information available, was clearly stated and investigated (Todorova & Markhasin, 2003). A novel paradigm for future m-commerce applications was proposed and implemented based on a completely distributed all-IP/ATM architecture for global satellite multimedia communications and soft QoS-oriented MAC technology (Markhasin, Olariu & Todorova, 2004).

For comparison of QoS-oriented MAC protocols, matrix systematization is well suited (Markhasin, 1984). Matrix systematization is based on two fundamental characteristics of MAC protocols: the matrix rows represent the *media access regulation mechanisms*, and the matrix columns represent the *media access time processes* (Table 1).

Table 1b shows an example of MAC protocols short-delay (medium bit delay MBD ≤ 10 bit) efficiencies. One can see that the matrix *ij*-th elements of MAC protocol characteristics are correlated via matrix *i*-th strings and *j*-th columns, respectively. As shown by Markhasin (1984), the highest short-delay efficiency ensures the adaptive protocols with controlled access mechanisms, adaptive time process, and smallest MAC expenses (see Table 1b).

Table 2 illustrates a comparison of MAC protocols in terms of short-delay and long-delay (MBD much larger than 10 bits) efficiency and real maximum throughput criterions. The MAC protocols with *adaptive frame* time processes (third column) ensure a highest short-delay efficiency and real maximum throughput. The same column protocols with controlled access and parallel-conveyer MAC instruction processing ensure the best time barrier overcoming abilities (Markhasin, 1996).

Table 3 compares MAC protocols in terms of soft QoS-provisioning capabilities. The following criteria were assumed: a) for static (hard) control—as its characteristics of the *controllability*, *differentiation*, and *guarantee* by the QoS and traffic parameters control, and bandwidth resources assignment; b) for dynamic (soft) control—as, added to criteria defined above, the *dynamical efficiency* and *stability* by the dynamical QoS and traffic parameters control, and bandwidth resources assignment.

The best soft QoS capabilities ensure MAC protocols with *adaptive frame* time access, *controlled (reservation)* access mechanisms, and *parallel-conveyer processing* of MAC instructions. One of them is the MAC protocol TBR-RS, which provides stable and automatic "on-the-fly" adaptation of the M-periodical hyperframe to traffic intensity (Markhasin, 1988, 1996). At the same time, this protocol allows one to realize universal RS-token MAC tools *"all-by-one"* for soft (up to real-time) control of QoS, traffic parameters, and bandwidth resources.

FUTURE TRENDS

Now that mobile networks are being called upon to support real-time interactive multimedia traffic, such as video tele-conferencing, these networks must be able to provide their users with quality-of-service (QoS) guarantees. Although the QoS provisioning problem arises in wireline networks as well, mobility of hosts and scarcity of bandwidth makes QoS provisioning a challenging task in wireless networks. It has been noticed that multimedia applications can tolerate and gracefully adapt to transient fluctuations in the QoS that they receive from the network. The management of such adaptive multimedia applications is becoming a new research area in wireless networks. As it turns out, the additional flexibility afforded by the ability of multimedia applications to tolerate and to adapt to transient changes in the QoS parameters can be exploited by protocol designers to significantly improve the overall performance of wireless systems. A major role in the task of QoS provisioning for multimedia application will be played by the MAC technology.

CONCLUSION

This article introduces an advanced QoS-aware MAC technology and a fully distributed architecture for future global wireless multi-service platforms 3G+/4G, which can support an integrated set of mobile multimedia applications, such as m-commerce, m-enterprise, m-business, m-positioning, e-government, e-work, e-education, e-health, e-culture, and so forth. This QoS-oriented MAC technology provides advanced broadband communication multi-services for a large population of geographically dispersed users, in support of anytime-anywhere communications.

ACKNOWLEDGEMENT

This work was supported, in part, by NATO collaborative grant PST.CLG979033.

REFERENCES

Abramson, N. (1970). The Aloha system—Another alternative for computer communications. *AFIPS Conference Proceedings*, *37*, 281-285.

Brandt, H., Todorova, P., Lorenz, P., Markhasin, A., Milne, P. & Ristol, S. (2001, April), Multifunctional distributed broadband ATM with dynamic control of QoS Hyperbus over satellite. *Proceedings of the 19th AIAA International Communication Satellite System Conference* (pp. 1-7), Toulouse, France.

Fattah, H. & Leung C. (2002). An overview of scheduling algorithms in wireless multimedia networks. *IEEE Wireless Communications*, *9*(5), 76-83.

Frigon, J.-F., Chan, H.C.B. & Leung, V.C.M. (2001). Dynamic reservation TDMA protocol for wireless ATM networks. *IEEE JSAC*, *19*(2), 370-383.

Gummalla, A.C.V. & Limb, J.O. (2000).Wireless medium access control protocols. *IEEE Communications Surveys*, (Second Quarter). Retrieved from www.comsoc.org/pubs/surveys.

Karol, M.J., Liu, Z. & Eng, K.Y. (1995). An efficient demand-assignment multiple access protocol for wireless packet (ATM) networks. *Wireless Networks*, *1*(3), 267–279.

Kleinrock, L. & Tobagi, F.A. (1975). Packet switching in radio channels: Part 1. Carrier sense multiple-access methods and their throughput-delay characteristics. *IEEE Transaction on Communications*, *COM-23*(12), 1400-1416.

Li, W.O.K. & Tsun-Yee, Y. (1984). An integrated voice and data multiple access scheme for a land-mobile satellite system. *IEEE Proceedings*, *72*(11), 199-209.

Markhasin, A. (1969, November). About the packet loss problem by bursty data transmission in the enterprise control systems. *Reports of the Scientific-Technical Conference on Design and Implementation of the Electronic Control Systems for Discrete Technology Enterprises, Section 8: Technical Facilities* (pp. 27-31).

Markhasin, A. (1984). *Architecture of data radio networks*. Novosibirsk, Russia: Science Publishing House.

Markhasin, A. (1985). Architecture of an integrated wide-area radio data network. *Automatic Control & Computer Sciences*, *19*(5), 11-18.

Markhasin, A. (1988). Flexible integrated service radio networks: Architecture and dynamic control of structure, functions and delivery. *Automatic Control & Computer Sciences*, *22*(3) 12-19.

Markhasin, A. (1996). Multi-access with dynamic control of the traffic and service quality in broadband ATM networks. *Optoelectronics, Instrumentation and Data Processing*, *3*, 93-99.

Markhasin, A. (2001, June). Advanced cost-effective long-delay broadband ATM medium access control technology and multifunctional architecture. *Proceedings of the IEEE International Communication Conference* (ICC'2001) (pp. 1914-1918).

Markhasin, A., Olariu, S. & Todorova, P. (2004). QoS-oriented medium access control for all IP/ATM mobile commerce applications. In N.S. Shi (Ed.), *Mobile commerce Applications* (Chapter 14, pp. 303-331). Hershey, PA: Idea Group Publishing.

Roberts, L.G. (1972). Aloha packet system with and without slots and capture. *ACM SIGCOMM Computer Communication Review*, 28-42.

Rubun, I. (1979). Access-control disciplines for multi-access communication channels: Reservation and TDMA schemes. *IEEE Transactions on Information Theory*, IT-25, 516-536.

Tobagi, F.A. (1980). Multi-access protocols in packet communications systems. *IEEE Transaction on Communications*, *COM-28*(4), 468-488.

Todorova, P. & Markhasin, A. (2003). Quality-of-service-oriented media access control for advanced mobile multimedia satellite systems. *Proceedings of the 36th Annual Hawaii International Conference on System Sciences* (HICSS 35, Big Island of Hawaii) (pp. 1-8).

Wong, W.K., Zhu, H. & Leung, V.C.M. (2003). Soft QoS provisioning using the token bank fair queues scheduling algorithm. *IEEE Wireless Communications*, 10(3), 8-16.

KEY TERMS

Collision: A usually destructive event resulting from the simultaneous access to the same bandwidth resource by multiple users.

Multiple-Access (MA): Procedures regulating the simultaneous use of a common transmission medium by multiple users.

Controlled MA: A class of coordinated centralized or distributed collision-free *multiple- access* protocols.

Free (random) MA: A class of non-coordinated distributed *multiple-access* protocols where *collisions* may occur.

Hybrid MA: A class of *multiple-access* consisting of a combination of *controlled multiple-access* and *free multiple-access*.

Medium Access Control (MAC): According to the OSI terminology, an interface between the physical layer and LLC sub-layers.

Medium Access Control (MAC) Protocol: A set of rules, algorithms, and interfaces that define and control access to a common *medium*.

Quality of Service (QoS): A collection of performance parameters for network service including bandwidth, average delay, jitter, packet loss probability, among many others.

Qualitative Methods in IS Research

Eileen M. Trauth
The Pennsylvania State University, USA

INTRODUCTION

As information technologies have evolved, so too has our understanding of the information systems that employ them. A significant part of this evolving understanding is the role of the human contexts within which information systems are situated. This, in turn, has led to the need for appropriate methods of studying information systems in their context of use. Just as decisions about information systems need to be considered within their contexts of use, so also do choices about the appropriate research methodologies to employ for studying them. Increasingly, qualitative methods are chosen as an appropriate method for studying contextual aspects of information systems development, use and impact.

Qualitative research refers to research methods that engage in the interpretation of words and images rather than the calculation of numbers. These methods include: ethnography, case study, action research, interviews, and text analysis (i.e., conversation analysis, discourse analysis, and hermeneutics). Qualitative research can be theory-driven in much the same way as quantitative analysis. However, it can also employ grounded theory techniques in order to develop theory (Glaser & Strauss, 1967).

Following some early uses of qualitative methods in the 1980s (e.g., Benbasat et al., 1987; Kaplan & Duchon, 1988; Lee, 1989; Mumford et al., 1985), there has been a significant growth in the use of qualitative methods for information systems research since the 1990s (e.g., *Journal of Information Technology*, 1998; Lee et al., 1997; *MIS Quarterly*, 1999, 2000; Nissen et al., 1991; Trauth, 2001).

Accompanying the increased use of qualitative methods for IS research has been a discussion of various methodological issues. Among the key aspects of this dialogue are discussions about the suitability of qualitative methods for various types of research and issues arising from a particular type of qualitative methods: interpretive methods. This article presents a reflection on some these discussions in the form of a consideration of five factors that can influence the choice of qualitative (particularly interpretive) methods for information systems research.

FACTORS INFLUENCING THE DECISION

The Research Problem

The research problem, *what* one wants to learn, should determine *how* one goes about learning it. Heaton (1998) chose observation, interview and document analysis to examine the social construction of computer-supported cooperative work in two different cultures in order to learn how the meaning of "culture" was reflected in the design of systems. Trauth (2000) used ethnographic methods to explore the influence of socio-cultural factors on the development of a nation's information economy. Bentley et al.'s (1992) ethnographic study of the work practices of air traffic controllers informed their design of an interface to an air traffic control database. Walsham and Sahay (1999) conducted extensive interviews to gain an in-depth understanding of the implementation of geographical information systems for administrative applications in India. Phillips (1998) employed public discourse analysis to reveal the way in which concerns about anonymity, surveillance, and privacy are integrated into public understanding of a consumer payment system.

The Epistemological Lens

Orlikowski and Baroudi (1991) considered the influence of the epistemological lens – positivist, interpretive or critical – on the conduct of IS research. While there is some positivist, qualitative IS research (e.g., Lee 1989), most qualitative IS research is either interpretive or critical because of the assumption that "our knowledge of reality is a social construction by human actors" that precludes obtaining objective, value-free data (Walsham, 1995, p. 376). The interpretive epistemology has also spawned IS research employing hermeneutic methods (e.g., Boland, 1985, 1991, and Trauth & Jessup, 2000). Ngwenyama and Lee (1997) used the critical lens to examine information richness theory.

The Uncertainty Surrounding the Phenomenon

According to Galliers and Land (1987), the added complexity from including relations with people and organizations in a view of information systems introduces greater imprecision and the potential for multiple interpretations of the same phenomenon. Hence, alternatives to quantitative measurement are needed. Others argue that the less that is known about a phenomenon the more difficult it is to measure it. Benbasat et al. (1987) explained that the case study approach is appropriate for IS research areas in which few previous studies have been carried out. Paré and Elam (1997) built theories of IT implementation using case study methods. Orlikowski's (1993) rationale for choosing qualitative methods and grounded theory to study the adoption of CASE tools rested on the absence of systematic examination of the organizational changes accompanying the introduction of CASE.

The Researcher's Skills

The absence of formal study of qualitative methods may serve as a barrier to choosing these methods. Orlikowski (1991) suggested that institutional conditions have inhibited the teaching of qualitative methods because of the functionalist/positivist perspective of American business schools where IS is typically taught. These institutional conditions, within which doctoral studies are conducted and dissertations are written, have inhibited the use of alternative research paradigms and methodologies with long-term implications for the choice of methods used in IS research. Schultze's (2001) reflection on her decision to choose interpretive methods for her dissertation illustrates the importance of institutional influence. Exposure to advisors with expertise in interpretive methods gave her methodological opportunities not available to other students.

The Academic Politics

The choice of research methods is influenced by the country in which one works, whether one has completed the PhD, whether one has a tenured position, one's academic rank, and the particular inclinations of the university at which one works. The norms and values of the field are reinforced during one's education and beyond. What is taught in research methods seminars sets the standard for "acceptable" research. Advice to junior faculty, peer review of journal articles and the tenure review process all reinforce those norms and values. Fitzgerald and Howcroft (1998) described the polarization of positions into "hard" and "soft" perspectives. Klein and Myers (1999) contributed to closing this methodological divide by developing a set of principles for conducting and evaluating interpretive field studies.

FUTURE TRENDS

As our understanding of the context of information systems grows, our desire to understand and explain contextual factors will motivate researchers to explore new ways to employ qualitative methods. Therefore, we can expect greater use of the critical epistemological lens in the use of qualitative methods in IS research. We can also expect to see the increased use of *virtual* qualitative research methods. That is, the traditional face-to-face methods of data generation used in qualitative research will find increasing analogues in the virtual world. We can expect to see, for example, "virtual ethnographies," "virtual participant observation" and "online interviews".

CONCLUSION

The primary advantage of using qualitative, particularly interpretive, methods is the flexibility it affords the researcher during data generation and analysis. The main disadvantage of qualitative, particularly interpretive, methods is overcoming concerns about validity and generalization of findings. The concepts of both statistical validity and statistical generalization need to be redefined for qualitative research. The *MIS Quarterly* "Special Issue on Intensive Research" has addressed the validity issue by publishing exemplar research papers that provide evaluative criteria for other researchers to use in establishing validity. The generalizability issue is being addressed in thoughtful pieces such as the recent article by Lee and Baskerville (2003).

Despite these issues, the acceptance of qualitative methods for IS research is evidence of a growing consensus that these methods make a valuable contribution to the study of information systems in context. In making the decision to use qualitative methods a number of factors must be taking into consideration. These factors relate to the characteristics of the research problem, the researcher and the research environment.

REFERENCES

Benbesat, I., Goldstein, D.K., & Mead, M. (1987). The case research strategy in studies of information systems. *MIS Quarterly, 11*(3), 369-386.

Bentley, R., Hughes, J.A., Randall, D., Rodden, T., Sawyer, P., Shapiro, D., & Sommeville, I. (1992). Ethnographically-informed systems design for air traffic control. In J. Turner & R. Kraut (Eds.), *Sharing perspectives: Proceedings of ACM Conference on Computer-Supported Cooperative Work* (pp.123-129). New York: ACM Press.

Boland, R.J. (1985). Phenomenology: A preferred approach to research on information systems. In E. Mumford, R.A. Hirschheim, G. Fitzgerald & T. WoodHarper (Eds.), *Research methods in information systems* (pp.193-201). Amsterdam: NorthHolland.

Doolin, B. (1998). Information technology as disciplinary technology: Being critical in interpretive research on information systems. *Journal of Information Technology, 13*(4), 301-312.

Fitzgerald, B., & Howcroft, D. (1998). Towards dissolution of the IS research debate: From polarization to polarity. *Journal of Information Technology, 13*(4), 313-326.

Galliers, R.D., & Land, F.F. (1987). Choosing appropriate information systems research strategies. *Communications of the ACM, 30*(11), 900-902.

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory*. Chicago: Aldine Publishing Co.

Heaton, L. (1998). Talking heads vs. virtual workspaces: A comparison of design across cultures. *Journal of Information Technology, 13*(4), 259-272.

Journal of Information Technology. (1998). *Special Issue on Interpretive Research in Information Systems*, *13*(4).

Kaplan, B., & Duchon, D. (1988). Combining qualitative and quantitative methods in information systems research: A case study. *MIS Quarterly, 12*(4), 571-586.

Klein, H.K., & Myers, M.D. (1999). A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS Quarterly*, *23*(1), 67-93.

Lee, A.S. (1989). A scientific methodology for MIS case studies. *MIS Quarterly, 13*(1), 33-50.

Lee, A.S., & Baskerville, R.L. (2003). Generalizing generalizability in IS research. *Information Systems Research*, *14*(3), 221-243.

Lee, A.S., Liebenau, J., & DeGross, J.I. (Eds.). (1997). *Information systems and qualitative research.* London: Chapman & Hall.

MIS Quarterly. (1999). *Special Issue on Intensive Research, 23*(1).

MIS Quarterly. (2000). *Special Issue on Intensive Research, 24*(1).

Mumford, E., Hirschheim, R.A., Fitzgerald, G., & WoodHarper, T. (Eds.). (1985). *Research methods in information systems.* Amsterdam: NorthHolland.

Ngwenyama, O.K., & Lee, A.S. (1997). Communication richness in electronic mail: Critical social theory and the contextuality of meaning. *MIS Quarterly*, *21*(2), 145-167.

Nissen, H.-E., Klein, H.K., & Hirschheim, R. (Eds.). (1991). *Information systems research: Contemporary approaches and emergent traditions.* Amsterdam: North-Holland.

Orlikowski, W.J. (1991). *Relevance versus rigor in information systems research: An issue of quality — the role of institutions in creating research norms.* Panel Presentation at the IFIP 8.2 Working Conference on the Information Systems Research Challenges, Copenhagen, Denmark.

Orlikowski, W.J. (1993). CASE tools as organizational change: Investigating incremental and radical changes in systems development. *MIS Quarterly*, *17*(3), 309-340.

Orlikowski, W.J., & Baroudi, J.J. (1991). Studying information technology in organizations: Research approaches and assumptions. *Information Systems Research*, *2*(1), 1-28.

Paré, G., & Elam, J.J. (1997). Using case study research to build theories of IT implementation. In A.S. Lee, J. Liebenau & J.I. DeGross (Eds.), *Information systems and qualitative research* (pp. 542-568). London: Chapman & Hall.

Phillips, D. (1998). The social construction of a secure, anonymous electronic payment system: Frame alignment and mobilization around ecash. *Journal of Information Technology, 13*(4), 273-284.

Schultze, U. (2001). Reflexive ethnography in information systems research. In E.M. Trauth (Ed.), *Qualitative research in IS: Issues and trends* (pp. 78-103). Hershey, PA: Idea Group Publishing.

Trauth, E.M. (2000). *The culture of an information economy: Influences and impacts in the Republic of Ireland.* Dordrecht, the Netherlands: Kluwer Academic Publishers.

Trauth, E.M. (2001). *Qualitative research in IS: Issues and trends.* Hershey, PA: Idea Group Publishing.

Trauth, E.M., & Jessup, L. (2000). Understanding computer-mediated discussions: Positivist and interpretive analyses of group support system use. *MIS Quarterly*, *24*(1), 43-79.

Walsham, G. (1995). The emergence of interpretivism in IS research. *Information Systems Research, 6*(4), 376-394.

Walsham, G., & Sahay, S. (1999). GIS for district-level administration in India: Problems and opportunities. *MIS Quarterly*, *23*(1), 39-65.

KEY TERMS

Case Study: An examination of a phenomenon in its natural setting using fixed boundaries such as time.

Critical Research: Critique of the status quo through the exposure of what are believed to be deep-seated, structural contradictions within social systems.

Ethnography: Research characterized by an extended period in the field and which involves the researcher being immersed in the community being studied.

Grounded Theory: A method used to systematically derive theories of human behavior from empirical data.

Interpretive Research: Exploring the deeper structure of a phenomenon within its cultural context by examining the subjective meanings that people create.

Positivist Research: Research premised on the existence of a priori fixed relationships among phenomena.

Qualitative Methods: Methods for data collection and analysis that focus on understanding of text and observation rather than numeric calculation.

Quality Assurance Issues for Online Universities

Floriana Grasso
Liverpool University, UK

Paul Leng
Liverpool University, UK

INTRODUCTION

Online delivery of degree-level programmes is an attractive option, especially for working professionals and others who are unable to contemplate full-time residential university attendance. If such programmes are to be accepted, however, it is essential that they attain the same standards and quality as conventionally delivered degrees. The key challenge is to find ways to ensure that the qualities that make university education attractive are preserved in the context of a new and quite different model of delivery.

Many systems have been developed to support online learning (see, e.g., Anderson & Kanuka, 1997; Davies, 1998; Persico & Manca, 2000; Suthers & Jones, 1997; Yaskin & Everhart, 2002). These systems may or may not mimic conventional lecture-room teaching, but will necessarily involve major differences in the ways in which teaching and student support are organised. Furthermore, the Internet lends itself naturally to an internationalisation of education delivery, but this also poses challenges for universities that have developed their structures within the framework of national education systems. To address these issues, it may be desirable for the university to work in partnership with other agencies, for example to provide local support services for students. This too, however, may introduce new problems of quality control and management. We will discuss here what structures are required to ensure the quality of the education provided and the standards of the degrees offered in this context.

BACKGROUND

The emergence of the Internet as a way of delivering higher education has led to examinations of its implications for education policy in many national and international contexts. A set of benchmarks for quality of online distance education was developed by the (U.S.-based) Institute for Higher Education Policy (2000). This identified a total of 24 benchmarks, in seven categories. In the UK, the Quality Assurance Agency for Higher Education has issued guidelines on the Quality Assurance of Distance Learning (QAA, 2000a), with a similar scope to those of the IHEP. A comparison of the main headings of the two frameworks is illustrated in Table 1. Also relevant, when the delivery model involves partnership with external agencies, is the QAA Code of Practice in relation to Collaborative Provision (QAA, 2000b). Similar issues are examined in an Australian context by Oliver (2001), and from Hong Kong by Yeung (2002). Yorke (1999) discusses quality assurance issues in relation to globalised education, touching especially on collaborative provision. Other perspectives are offered by Pond (2002), Little and Banega (1999), and Davies et al. (2001).

Much of the work in this field reflects "an implicit anxiety that the 'values' of traditional teaching may some-

Table 1. Comparison of U.S. and UK QA frameworks

<table>
<tr><th>IHEP (USA)</th><th>QAA (UK)</th></tr>
<tr><td>Institutional Support</td><td>System Design (i.e., institutional issues)</td></tr>
<tr><td>Course Development</td><td rowspan="2">Programme Design (course development and structure)</td></tr>
<tr><td>Course Structure</td></tr>
<tr><td>Teaching and Learning</td><td>Programme Delivery</td></tr>
<tr><td>Student Support</td><td>Student Support</td></tr>
<tr><td>Faculty Support</td><td></td></tr>
<tr><td>Evaluation and Assessment</td><td>Student Assessment</td></tr>
<tr><td></td><td>Student Communication and Representation</td></tr>
</table>

how be eroded" (Curran, 2001). There is consequently, in most prescriptions, a strong emphasis on replicating in an online context the characteristics of quality that we might expect to (but do not always) find in conventional teaching. Thus, one of the precepts of (QAA, 2000a) calls for "…..managing the delivery of each distance learning programme of study in a manner that safeguards the academic standards of the award"; and one of the benchmarks of the IHEP specifies that "Feedback in student assignments is provided in a timely manner". Unexceptionable as they are, these requirements are not peculiar to online distance learning. The key issue is not, therefore, one of defining quality assurance criteria, but rather that of providing structures to ensure their implementation.

QUALITY ASSURANCE FOR ONLINE DEGREES

Pedagogic Issues

Before examining quality assurance as such, we will first consider questions relating directly to the pedagogic approach used in online learning. In this respect, the premise that quality in online learning involves only a replication of on-campus characteristics is, we believe, limiting. We start, instead, from the standpoint that lecture-based teaching, whatever its merits, is not necessarily an ideal which online teaching must emulate. Students all too frequently attend lectures in an entirely passive mode, expecting to listen and receive the information they require while making no positive contribution themselves. Interaction between lecturer and students, and within groups of students is low, especially in the large classes that are typical of modern universities.

Conversely, online teaching makes it possible to recreate, through the medium of moderated online discussion, an atmosphere that is closer to that of a small-group on-campus seminar, and, paradoxically, can be far more involving and interactive than is typically the case in on-campus teaching. Two broad principles inform the approach: *constructivism* (Wilson, 1996), and *collaborative enquiry*. Collaborative enquiry via Internet-mediated communication provides a framework for this mode of learning (Stacey, 1998). The aim is to use the medium to foster the creation of a *learning community* (Hiltz & Wellman, 1997) that will enable dialogue between participants, sharing of information, and collaborative project work.

Moderated discussion (Collins & Berge, 1997) is a key feature of the teaching paradigm here, and serves a number of purposes that are relevant to the question of quality. Most obviously, it provides the means by which students may share information and experience, comment on the course materials and assignments, raise questions, and bring to the class knowledge and expertise that is outside the experience of the course teacher. To a significant extent, the students thus participate actively in the teaching process, augmenting the overall learning experience. Less obviously, there are other issues of quality in which classroom discussion can have a role; we will discuss these next.

Quality Assurance Issues

Key issues of quality assurance in an online degree programme include:

- Academic control
- Academic standards
- Staff appointment and training
- Monitoring of programme delivery
- Assessment procedures
- Student identity and plagiarism
- Student progression and support

Our review of these issues, next, draws on our experience with the online degree programmes at the University of Liverpool (Gruengard, Kalman & Leng, 2000).

Academic Control

The primary requirement of the frameworks defined by the QAA and other bodies is that all academic aspects of an online degree programme should remain the responsibility of the parent university, which should have structures and procedures that are effective in discharging this responsibility. The issue here is that the academic standards and quality of the programme may be threatened by differences between the parties involved in its delivery, especially when there is only an indirect relationship between the university and some of the people involved (for example, regional partner organisations, or locally-based tutors).

In principle, these problems can be resolved by placing online degree programmes firmly within the framework defined by the university for approving and managing its courses. To oversee this, we have at Liverpool established a dedicated organisational unit within the university, the e-Learning Unit.

Academic Standards

A corollary this is that, wherever possible, the quality management of an online programme should follow procedures that are comparable to those established for other degrees of the university, especially in respect to those procedures that define and maintain the academic standards of the degree. These will include course and module approval and review procedures, assessment criteria, and so forth. In most cases, it should be possible to exactly replicate the procedures that apply on campus.

Staff Appointment and Training

In the Liverpool University online programmes, teaching is principally carried out by part-time *instructors* who are based throughout the world. All appointments, however are subject to University approval using the same procedures and criteria that apply for the approval of non-established staff to teach on internal programmes. All instructors are required to first undertake an (online) training programme, over a period of 6 weeks, in which they are instructed on the use of the software platform and the methodology and pedagogic approach used in the programme. Further to this, in the first module they teach, the instructor is overseen by an academic *mentor* whose role is to advise and guide the novice.

Monitoring of Programme Delivery

A key aspect of quality assurance, especially when instructors as well as students are not campus-based, is the monitoring of module delivery. Interestingly, it is easier to do this effectively than is the case for on-campus teaching. In the Liverpool University model, *all* significant communications between staff and students are made electronically, as postings in or through a "virtual classroom". Thus, all are subject to scrutiny by "lurking" in the classroom while the module is being delivered, and/or by examining the recorded history of the class subsequently. In this way the academic staff of the e-Learning Unit monitor the delivery of each module to ensure that the appropriate standards and quality criteria are met. The module monitor is required to complete a report that also incorporates a review of feedback from students and the comments of the instructor. These reports are reviewed routinely by a Board of Studies.

Assessment Procedures

Assessment in this model is readily subject to moderation by the module monitor, who has access to all the relevant information contained in the record of the virtual classroom. Review of assessment outcomes is an explicit part of the end-of-module procedures, and these are finally subject to confirmation by the Board of Examiners, which in the UK framework includes an external examiner from another university. The external examiner also has full access to the virtual classroom and so, in practice, has the opportunity for a more detailed examination of assessment processes and standards than is usually possible in on-ground teaching.

Strict management of the assessment process is in our experience essential when instructors are drawn from a wide variety of cultures. In this case consistency of assessment can only be achieved by defining very clear grading descriptors and criteria, and by insisting that the interpretation of these is moderated firmly within the university's assessment model.

Student Identity and Plagiarism

One of the questions that most exercises organisations considering online learning is that of how to confirm the identity of participants, together with the related question of protecting against plagiarism. In this respect a key role is played by discussion in the virtual classroom. Participation in discussion provides a means of monitoring the effective involvement of each student, and assists in preventing impersonation and plagiarism. Research has shown (Klemm & Snell, 1996; Lai, 1997) that involvement in online discussion is rarely wholly effective unless moderated by external facilitators, and we believe that it is important that it be made a requirement, equivalent to the attendance requirements of on-campus degrees.

The fact that all communications take place online, and are recorded and preserved indefinitely, provides further protection against plagiarism. It is easy to apply programs that perform comparisons of work submitted in the virtual classroom, or use services that perform checks against plagiarism throughout the Web.

Student Progression and Support

The requirement to participate in online discussion also provides a means of monitoring student progress. If a student is failing to keep up with the requirements of the programme, this rapidly becomes apparent as his/her contributions to the discussion falter. At this point the instructor can intervene to investigate and take action if required.

FUTURE TRENDS

A key precept for quality assurance of online degree programmes is that, wherever possible, procedures and structures that are thought to be necessary on campus should have an online equivalent. As the previous discussion reveals, creating equivalent processes is usually

Table 2. Comparison of on-campus and online characteristics

	On Campus	**Online**
Teaching/learning mode	Predominantly lecture-based	Predominantly seminar-based
Interpersonal interactions	Low: via classroom discussions/questions	High: via moderated e-mail dialogue
Verification of student identity	Personal appearance	Textual/linguistic characteristics
Student support	Face-to-face meetings with tutors	E-mail interactions with tutors
Review of standards of attainment	Scrutiny of examination scripts	Inspection of work in online classrooms
Mentoring/monitoring of staff performance	Inspection of selected lectures/classes	"Lurking" in ongoing online classes

possible, but some significant differences emerge, highlighted in Table 2. The comparison of conventional lecture-based teaching with our model of online learning is not, we believe, to the disadvantage of the latter.

Although online degree programmes will not replace campus-based education, we believe there will be a strong future trend towards programmes that will meet particular areas of demand, especially the needs of working adults. Examples of successful online higher education programmes are, so far, relatively few, especially in Europe. We believe firmly that the successful programmes of the future will be those that focus on pedagogy, and give precedence to academic standards and quality assurance, rather than those that emphasise technological aspects or focus on low-cost delivery.

CONCLUSION

Quality in online degree programmes is often perceived to imply a replication of on-campus characteristics. We believe, conversely, that in some respects online delivery provides an opportunity to enhance the quality of learning opportunities for students. Especially this is so when the learning paradigm encourages a high degree of discussion and interaction between staff and students in the virtual classroom.

In other respects, it is indeed necessary to maintain comparability with on-campus degrees, especially in relation to academic standards. Here, however, the key issues do not relate to the definition of standards and quality criteria, but to the creation of mechanisms to uphold them. We believe it is necessary to establish well-defined and rigorous monitoring procedures for this purpose. Again, a learning environment that emphasises classroom discussion is a help in many aspects of quality management.

REFERENCES

Anderson, T., & Kanuka, H. (1997). On-line forums [1]: New platforms for professional development and group collaboration. *Journal of Computer-Mediated Communication, 3*(3).

Collins, M.P., & Berge, Z.L. (1997). *Moderating online electronic discussion groups*. Paper presented at the American Educational Research Association, Chicago. Retrieved on September 20, 2004 from *http://www.emoderators.com/moderators/sur_aera97.html*

Curran, C. (2001). The phenomenon of on-line learning. *European Journal of Education, 36*(2), 113-132.

Davies, G. (Ed.). (1998). Teleteaching '98: Distance learning, training and education. *Proc XV IFIP World Computer Congress*, Vienna/Budapest.

Davies, G., Doube, W., Lawrence-Fowler, W., & Shaffer, D. (2001). Quality in distance education. *Proc 32nd SIGCSE Technical Symposium on Computer Science Education*, ACM (pp. 392-393).

Gruengard, E., Kalman, Y.M., & Leng, P (2000). University degrees via the Internet: A new paradigm for public-private partnership. *Innovations Though Electronic Commerce* (Proc IEC2000), Manchester (pp. 46-53).

Hiltz, S.R., & Wellman, B. (1997). Asynchronous learning networks as a virtual classroom. *Comm ACM, 40*(9), 44-49.

Institute for Higher Education Policy. (2000). *Quality on the line: Benchmarks for success in Internet-based distance education.*

Klemm, W.R., & Snell, J.R. (1996). Enriching computer-mediated group learning by coupling constructivism with collaborative learning. *Journal of Instructional Science and Technology, 1*(2).

Lai, K-W. (1997). Computer-mediated communication for teenage students: A content analysis of a student messaging system. *Education and Information Technologies*, *2*, 31-45.

Little, D.L., & Banega, B.H. (1999). Development of standards or criteria for effective online courses. *Educational Technology and Society, 2*(3), 4-15.

Oliver, R. (2001). Assuring the quality of online learning in Australian higher education. *Proceedings of Moving Online Conference II.*

Persico, D., & Manca, S. (2000). Use of FirstClass as a collaborative learning environment. *Innovations in Education and Training International, 37*(1), 34-41.

Pond, W.K. (2002). Distributed education in the 21st century: Implications for quality assurance. *Online Journal of Distance Learning Administration*, *5*(2).

QAA. (2000a). (Quality Assurance Agency for Higher Education): *Distance learning guidelines.* Retrieved on September 20, 2004 from *http://www.qaa.ac.uk*

QAA. (2000b). (Quality Assurance Agency for Higher Education): *Code of practice for the assurance of academic quality and standards in higher education: Collaborative provision.* Retrieved on September 20, 2004 from *http://www.qaa.ac.uk*

Stacey, E. (1998). Learning collaboratively in a CMC environment. In G. Davies (Ed.), *Teleteaching '98, Proceedings of XV IFIP World Computer Congress,* Vienna/Budapest, Austrian Computer Society (pp. 951-960).

Suthers, D., & Jones, D. (1997). An architecture for intelligent collaborative educational systems. *Proceedings of AI&ED'97*, Kobe, Japan.

Wilson, B.G. (1996). *Constructivist learning environments: Case studies in instructional design.* Educational Technology Publications.

Yaskin, D., & Everhart, D. (2002). Blackboard Learning System (Release 6): Product overview. Retrieved on September 20, 2004 from *http://www.blackboard.com/docs/wp/LSR6WP.pdf*

Yeung, D. (2002). Towards an effective quality assurance model of Web-based learning: The perspective of academic staff. *Online Journal of Distance Learning Administration*, 5(2).

Yorke, M. (1999). Assuring quality and standards in globalised higher education. *Quality Assurance in Education, 7*(1), 14-24.

KEY TERMS

Constructivism: A form of learning in which students construct their own unique understanding of a subject through a process that includes social interaction, so that the learner can explain understandings, receive feedback from teachers and other students, clarify meanings, and reach a group consensus.

Globalised Education: Educational programmes in which both students and educators may be globally distributed.

Higher Education Programme: The processes, learning materials, and associated procedures and facilities that lead to the completion of a degree or related qualification.

Moderated Discussion: Discussion that is supervised, partly directed, and evaluated by a programme tutor.

Moderation of Assessment: External oversight of an assessment process to ensure that appropriate standards are maintained.

Online Learning: A programme of education that is carried out wholly or primarily through the medium of the Internet.

Quality of Education: Refers to the effectiveness of the programme in promoting student learning and achievement, in relation to its expected outcomes and standards.

Standards (of Education): Refer to the level of attainment achieved by students, in relation to the qualification awarded.

Virtual Classroom: An internet-mediated forum for distribution of learning materials, classroom discussion, and collaborative working.

Q

Quality of UML

John Krogstie
NTNU and SINTEF, Norway

INTRODUCTION

According to Booch, Rumbaugh, and Jacobson (1999), developing a model for an industrial strength software system before its construction is regarded increasingly as a necessary activity in information systems development. The use of object-oriented modeling in analysis and design started to become popular in the late eighties, producing a large number of different languages and approaches. Lately, UML (2004) has taken a leading position in this area.

In this article, we give an overview assessment of UML using a generic evaluation framework. We will first present the evaluation framework. We will then evaluate the language quality of UML before pointing to the future direction and potential of UML.

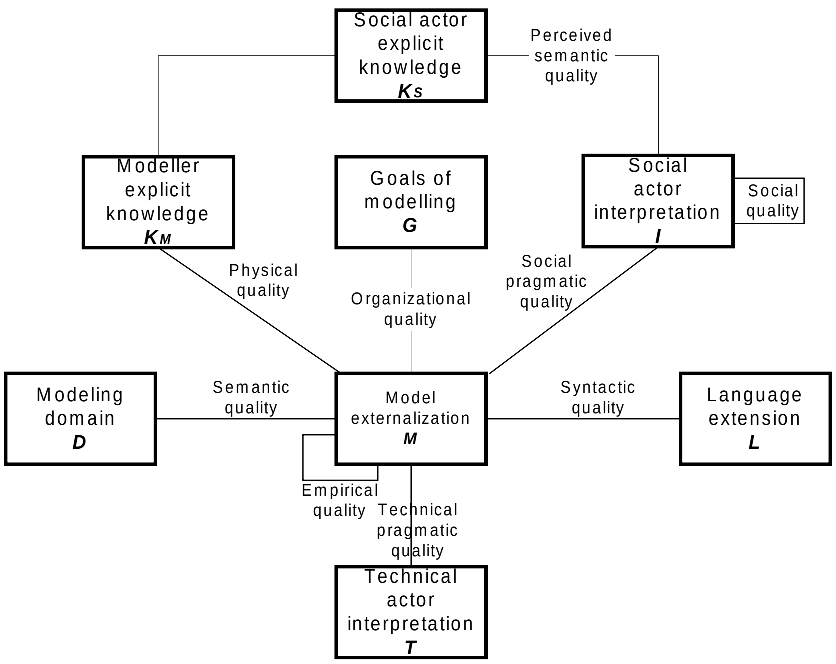
BACKGROUND

Krogstie, Sindre and Lindland (1995) and Krogstie and Sølvberg (2003) have developed a framework for quality of models and modeling languages.

The main concepts of the framework and their relationships are shown in Figure 1 and are explained in the following. Quality has been defined referring to the correspondence between statements belonging to the following sets:

- G, the goals of the modeling task.
- L, the language extension, that is, the set of all statements that are possible to make according to the graphemes, vocabulary, and syntax of the modeling languages used.
- D, the domain, that is, the set of all statements that can be stated about the situation at hand.
- M, the externalized model itself.
- K_s, the relevant explicit knowledge of those being involved in modeling. A subset of these is actively involved in modeling, and their knowledge is indicated by K_M.
- I, the social actor interpretation, that is, the set of all statements that the audience thinks that an externalized model consists of.
- T, the technical actor interpretation, that is, the statements in the model as "interpreted" by modeling tools.

Figure 1. Framework for discussing the quality of models



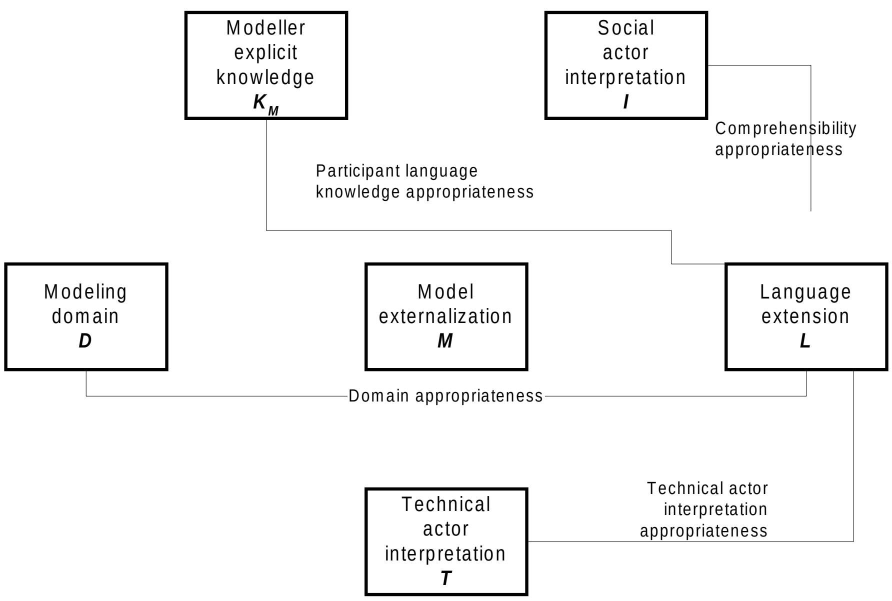
The main quality types are indicated by solid lines between the sets, and are described briefly in the following:

- Physical quality: The basic quality goals on the physical level are externalization, that the knowledge K of the domain D has been externalized, and internalizeability, that the externalized model M is available.
- Empirical quality deals with predictable error frequencies when a model is read or written by different users, coding (e.g., shapes of boxes) and HCI-ergonomics for documentation and modeling-tools. For instance, graph layout to avoid crossing lines in a model is a mean to address the empirical quality of a model.
- Syntactic quality is the correspondence between the model M and the language extension L.
- Semantic quality is the correspondence between the model M and the domain D. This includes validity and completeness.
- Perceived semantic quality is the similar correspondence between the audience interpretation I of a model M and his or hers current knowledge K of the domain D.
- Pragmatic quality is the correspondence between the model M and the audience's interpretation of it (I). We differentiate between social pragmatic quality (to what extent people understand the models) and technical pragmatic quality (to what extent tools can be made that interpret the models).
- The goal defined for social quality is agreement among audience members' interpretations I.
- The organizational quality of the model relates to that all statements in the model contribute to fulfilling the goals of modeling (organizational goal validity), and that all the goals of modeling are addressed through the model (organizational goal completeness).

Language quality relates the modeling languages used to the other sets. Four quality areas for language quality are identified, with aspects related both to the meta-model and the notation as illustrated in Figure 2.

1. Domain appropriateness. This relates to the language and the domain. Ideally, the conceptual basis must be powerful enough to express anything in the domain, not having what (Wand & Weber, 1993) terms construct deficit. On the other hand, you should not be able to express things that are not in the domain, that is, what is termed construct excess (Wand & Weber, 1993). Domain appropriateness is primarily a mean to achieve physical quality, and through this, to achieve semantic quality.
2. Participant language knowledge appropriateness relates the participant knowledge to the language. Participant language knowledge appropriateness is primarily a mean to achieve physical and pragmatic quality.
3. Participant comprehensibility appropriateness relates the language to the social actor interpretation. The goal is that the participants in the modeling effort using the language understand all of the possible statements of the language. Comprehensi-

Figure 2. Language quality in the quality framework



bility appropriateness is primarily a mean to achieve empirical and pragmatic quality.

4. Technical actor interpretation appropriateness relates the language to the technical audience interpretations. For tools interpretation, it is especially important that the language lend itself to automatic reasoning. This requires formality (i.e., both formal syntax and semantics being operational and/or logical), but formality is not necessarily enough, since the reasoning must also be efficient to be of practical use. This is covered by what we term analyzability (to exploit any mathematical semantics) and executability (to exploit any operational semantics). Different aspects of technical actor interpretation appropriateness are a mean to achieve syntactic, semantic and pragmatic quality (through formal syntax, mathematical semantics, and operational semantics).

OVERVIEW OF EVALUATION

The evaluation is structured according to the four areas of language quality presented previously. Before presenting the evaluation, we will position UML in relation to the sets of the quality framework.

Domain: According to OMG (2004), UML is a language for specifying, visualizing, constructing, and documenting the artifacts of software systems, as well as for business modeling and other non-software systems. In other words, UML is meant to be used in analysis of business and information, requirement specification and design. UML is meant to support the modeling of (object-oriented) transaction systems, real-time and safety critical systems. As illustrated in Favre (2003), UML is used and adapted for a number of different specific areas.

Language: We Have Based the Evaluation on UML (version 2.0).

The sets "Knowledge", "Model", and "Interpretation" must be judged from case to case in the practical application of a modeling language and tools. Also, when it comes to weighting the different criteria against each other, this must be done in the light of the specific modeling task, such as has been done, for example, by Arnesen and Krogstie (2002).

Due to the limitation on the length of an article of this kind and the breadth of the evaluation, we will only have room for presenting some of the major results. See Krogstie (2003) for a more detailed description of using the framework for evaluating UML.

Domain Appropriateness

Looking briefly on the coverage of the seven main modeling-perspectives in information systems modeling (Krogstie & Sølvberg, 2003), we find:

- UML has a very good support for modeling according to an object-oriented perspective, especially for design.
- The structural perspective is also well-supported, although not as well as in languages made specifically for this purpose.
- The behavioral perspective is supported using statecharts.
- The functional (process) perspective is supported on a high level through use case modeling, a language that has been highly criticized for not being well-defined (Hitz & Kappel, 1998). Whereas use cases are meant for requirements modeling, activity diagrams can be used for simple procedural descriptions by showing control flow and the production of data or objects in a process flow. This is useful for design. Many have also attempted using activity diagrams for business models. Improvements of activity diagrams is introduced in UML 2.0 in this regard. The lack of traditional dataflow in activity diagrams has been noted as a problem.
- The actor-role perspective is partly covered using the collaboration diagrams. Using roles in sequence diagrams or "swimlanes" in activity diagrams, we also get a role-oriented view, but there is no intuitive way to represent organizational and group-structures and relationships in UML.
- There are some general problems with constraints expressed in an OO modeling framework (Høydalsvik & Sindre, 1993). Temporal and deontic constraints are hard to express. The same problem applies to non-functional requirements. There is no support for modeling of goal-hierarchies (Mylopoulos, Chung, & Tu, 1999).
- The language-action perspective, which is most useful for the analysis of businesses, is not supported.

A meta-model of UML is defined (using UML), and there exist extension mechanisms to make the language more applicable in specific domains.

Most of UML is first useful during design. These parts of the language should not be used in analysis and requirements specification, even in areas where the transition from analysis to design is "seamless". (There is much evidence that, especially for business systems, this transition is far from seamless even when using object-oriented modeling in all domains (Davis, 1995; Høydalsvik & Sindre, 1993).) Proper guidelines for avoiding this are not consistently provided, and there is no support for avoiding using analysis and design concepts in the same model.

There are also mismatches between underlying basis and external representation. In sequence diagrams, for instance, a number of concepts are semantically vacant (Morris & Spanoudakis, 2001). Some of the problems are addressed in UML 2.0. We also find examples of concrete constructs in the meta-model, with no representation in the notation (e.g., namespace and model).

Participant Language Knowledge Appropriateness

It can be argued that for those being familiar with the main OO-modeling concepts and main OO modeling-languages, the core of UML should not represent a too steep learning curve. Almost all CS and IS-degrees now include courses where UML is lectured and used. On the other hand, we have noted the complexity of the language previously described. The large number of constructs in UML is partly due to its diverse diagramming techniques (Siau & Cao, 2001). Constructs in use case diagrams are very different from constructs in class diagrams. Class diagrams are also very different from activity diagrams or statecharts. This diversity causes problems. First, there are more constructs to learn. Second, the knowledge and experience gained in using one diagramming technique cannot be transferred to another diagramming technique in UML.

Participant Comprehensibility Appropriateness

Some main observations on this area are:

UML is argued to be overly complex, with 233 different concepts (in UML 1.4) (Castellani, 1999). In Siau and Cao (2001), a detailed comparison is presented. Although UML has more diagramming techniques when compared to object-oriented methods such as OMT, OOAD and Shlaer/Mellor, each of the diagramming techniques in isolation is no more complex than techniques found in these methods. On the overall method level on the other hand, UML stands out noticeably, being most complex according to most of the metrics.

With so many concepts, it is not surprising that some redundancy and overlap is witnessed.

Examples of lacking symbol differentiation is found, for example,.that both classes and objects are shown using rectangles. On the other hand, since the text-style of the object gives an indication of this, this is a minor problem.

UML contains many possibilities of adding (often small) adornments to the models, which often are difficult to see and comprehend.

A uniform use of symbols is not adhered to. An example is that different symbols are used for a role if it is external to the system (pin-man) or internal (rectangle).

Technical Actor Interpretation Appropriateness

The UML-syntax is rigorously defined, and the language is described through a meta-model made in the structural model of UML with accompanying natural language descriptions of the semantics. Using UML to model UML means that some of the definitions are circular, and this leaves UML (formally speaking) undefined. This would be unproblematic if most practitioners already understood the meaning of the concepts (classes, inheritance, and associations) that are involved. In UML 2.0, a formal (operational) action language has been included to support a wider repertoire of modeling techniques.

FUTURE TRENDS

Modeling as a general technique within information systems development has received increasing interest over the last years, and will continue to be of high importance, for example, in connection to the OMG MDA approach and business process support. Due to its strong support, UML is probably the best general modeling language to adopt as a basis for object-oriented development if one is not already using another language with good tool support that one is satisfied with. Most of the accidental problems such as inconsistencies in the language-descriptions found in earlier version of UML seem to be addressed in UML 2.0, but there are still concerns, and UML 3.0 is already planned, but it is difficult to judge what this will look like.

CONCLUSION

UML has been developed and refined over a number of years, but there are still some major weaknesses with the approach. Even if it has not been possible to agree on a standard process, outline process guidelines need to be included – even if the best that can be done is to describe a number of alternatives. Particularly problematic is the logical/physical confusion in the UML-definition. As discussed by Davis (1995), there are fundamental differences between the models related to analysis, design, and requirement specification. What our investigation has also illustrated is that although there is a perceived need to extend the expressiveness and formality of the language, the language has several weaknesses regarding comprehensibility appropriateness, and is already viewed as difficult to comprehend, with a steep learning curve. The distinction between infrastructure and superstructure provided for UML 2.0 tries to address this, but it is too early to judge to what extent this is successful. Looking at the accepted 800-page proposal for UML 2.0 does not give us much hope that UML will be particularly much easier to learn and use, especially because of the strong need for backward consistency.

REFERENCES

Arnesen, S., & Krogstie, J. (2002). Assessing enterprise modeling languages using a generic quality framework. *Proceedings of EMMSAD'02*, Toronto, Canada May.

Booch, G., Rumbaugh, J., & Jacobson, I. (1999). *The unified modeling language: User guide*. Addison-Wesley.

Castellani, X. (1999). Overview of models defined with charts of concepts. In E. Falkenberg, K. Lyytinen, & A. Verrijn-Stuart (Eds.), *Proceedings of the IFIP8.1 Working Conference on Information Systems Concepts (ISCO4); An Integrated Discipline Emerging*, September 20-22, Leiden, The Netherlands, (pp. 235-256).

Davis, A. (1995). Object-oriented requirements to object-oriented design: An easy transition? *Journal of Systems and Software, 30*(1/2), July/August, 151-159.

Favre, L. (Ed.) (2003). *UML and the unified process*. Hershey, PA: IRM Press.

Hitz, M., & Kappel, G. (1998). Developing with UML – Some pitfalls and workarounds. In J. Bézivin, & P.-A. Muller (Eds.), *UML'98 - Beyond the notation*, June 3-4, Mulhouse, France (pp.9-20). Springer-Verlag.

Høydalsvik, G.M., & Sindre, G. (1993). On the purpose of object-oriented analysis. In A. Paepcke (Ed.), *Proceedings of the Conference on Object-Oriented Programming Systems, Languages, and Applications (OOPSLA'93)*, September (pp. 240-255). ACM Press.

Krogstie, J. (2003). Evaluating UML using a generic quality framework. In L. Favre (Ed.), *UML and the unified process* (pp.1-22). Hershey, PA: IRM Press.

Krogstie, J., & Sølvberg, A. (2003). *Information systems engineering – Conceptual modeling in a quality perspective*, Trondheim, Norway: Kompendiumforlaget.

Morris, S., & Spanoudakis, G. (2001). UML: An evaluation of the visual syntax of the language. *Proceedings of HICSS 34*.

Mylopoulos, J., Chung, L., & Tu, E. (1999). From object-oriented to goal-oriented requirements analysis. *Communications of the ACM, 42*(1), January, 31-37.

OMG (2004). Unified Modeling Language v 2.0 OMG Web site *http://www.omg.org*

Siau, K. & Cao, Q. (2001). Unified modeling language (UML) – A complexity analysis. *Journal of Database Management*, January-March, 26-34.

Wand, Y., & Weber, R. (1993). On the ontological expressiveness of information systems analysis and design grammars. *Journal of Information Systems, 3*(4), 217-237.

KEY TERMS

Analysis Model: A model developed to learn all aspects of a problem domain to determine the best way to solve a specific set of user needs.

Design Model: A model developed to represent the optimal technical solution of a specified user need (as represented in a requirements model).

Model: An abstraction represented in a modeling language.

Modeling Language: A language (i.e., a set of symbols, and rules for how to combine these symbols) to represent knowledge relevant in information systems development.

Requirements Model: A model to represent the external requirement to a system without taking into account how the system looks inside.

Unified Modeling Language (UML): A general-purpose visual modeling language that is used to specify, visualize, construct and document the artifacts of a software system.

Visual Modeling Language: A diagrammatic modeling language.

Querying Multidimensional Data

Leonardo Tininini
Istituto di Analisi dei Sistemi e Informatica "Antonio Ruberti", Rome, Italy

INTRODUCTION

A powerful, easy-to-use querying environment is without doubt one of the most important components in a multidimensional database. Its effectiveness is influenced by many aspects, both logical (data model, integration, policy of view materialization, etc.) and physical (multidimensional or relational storage, indexes, etc.). Multidimensional querying is often based on the core concepts of multidimensional data modeling, namely the metaphor of the data cube and the concepts of facts, measures and dimensions (Agrawal, Gupta, & Sarawagi, 1997; Gyssens & Lakshmanan, 1997). In contrast to conventional transactional environments, multidimensional querying is often an exploratory process, performed by navigating along dimensions and measures, increasing/decreasing the level of detail and focusing on specific subparts of the cube that appear "promising" for the required information.

BACKGROUND

Multidimensional data are obtained by applying aggregations and statistical functions to elementary data, or more precisely to data groups, each containing a subset of the data and homogeneous with respect to a given set of attributes. For example, the data "Average duration of calls in 2003 by region and call plan" is obtained from the so-called fact table, which is usually the product of complex activities of source integration (Lenzerini, 2002) on the raw data corresponding to each phone call in that year. Several groups are defined, each consisting of calls made in the same region and with the same call plan, and finally applying the average aggregation function on the duration attribute of the data in each group. The pair of values (region, call plan) is used to identify each group and is associated with the corresponding average duration value. In multidimensional databases, the attributes used to group data define the dimensions, whereas the aggregate values define the measures of data.

The term multidimensional data comes from the well known metaphor of the data cube (Gray, Bosworth, Layman, & Pirahesh, 1996). For each of the n attributes, which are used to identify a single measure, a dimension of an n-dimensional space is considered. The possible values of the identifying attributes are mapped to points on the dimension's axis, and each point of this n-dimensional space is thus mapped to a single combination of the identifying attribute values and hence to a single aggregate value. The collection of all these points, along with all possible projections in lower dimensional spaces, constitutes the so-called data cube. In most cases, dimensions are structured in hierarchies, representing several granularity levels of the corresponding measures (Jagadish, Lakshmanan, & Srivastava, 1999). Hence, a time dimension can be organized into days, months and years; a territorial dimension into towns, regions and countries; a product dimension into brands, families and types. When querying multidimensional data, the user specifies the measures of interest and the level of detail of the information required by indicating the desired hierarchy level for each dimension. In a multidimensional environment, querying is often an exploratory process, where the user "moves" along the dimension hierarchies by increasing or reducing the granularity of displayed data. The operation of drill-down corresponds to an increase in detail, for example, by requesting the number of calls by region and month, starting from data on the number of calls by region or by region and year. Conversely, roll-up allows the user to view data at a coarser level of granularity (Cabibbo & Torlone, 1997).

OLTP VS. OLAP QUERIES

Multidimensional querying systems are commonly known as On-Line Analytical Processing (OLAP) systems (Li & Wang, 1996), in contrast to conventional On-Line Transactional Processing (OLTP) systems. The two types have several contrasting features, although sharing the same requirements on fast "on-line" response times.

- *Number of records involved.* One of the key differences between OLTP and multidimensional queries is the number of records required to calculate the answer. OLTP queries typically involve a rather limited number of records, accessed through primary key or other specific indexes, which need to be processed for short, isolated transactions or to be issued on a user interface. In contrast, multidimensional queries usually require the classification and aggregation of a huge amount of data (Gupta, Harinarayan, & Quass, 1995).

- *Indexing techniques.* Transaction processing is mainly based on the access of a few records through primary key or other indexes on highly selective attribute combinations. Efficient access is easily achieved by well-known and established indexes, particularly B+-tree indexes. In contrast, multidimensional queries require a more articulated approach, as different techniques are required, and each index performs well only for some categories of queries (Chan & Ioannidis, 1998; Jürgens & Lenz, 1999).
- *Current state vs. historical DB's.* OLTP operations require up-to-date data. Simultaneous information access/update is a critical issue, and the database usually represents only the current state of the system. In OLAP systems, the data does not need to be the most recent available and should, in fact, be time-stamped, thus enabling the user to perform historical analyses with trend forecasts. However, the presence of this temporal dimension may cause problems in query formulation and processing, as schemes may evolve over time and conventional query languages are not adequate to cope with them (Vaisman & Mendelzon, 2001).
- *Target users.* Typical OLTP system users are clerks, and the types of query are rather limited and predictable. In contrast, multidimensional databases are usually the core of decision support systems, targeted at management level. Query types are only partly predictable and often require highly expressive (and complex) query language. However, the user usually has little experience even in "easy" query languages like basic SQL: the typical interaction paradigm is a spreadsheet-like environment based on iconic interfaces and the graphical metaphor of the multidimensional cube (Cabibbo & Torlone, 1998).
- *Dimensions and measures.* Early statistical database research has already shown (Shoshani & Wong, 1985) that the standard relational model and operators (commonly used to represent and query transactional databases) are inadequate for effective representation and querying of multidimensional data. This led to the distinction between category attributes (the dimensions) and summary attributes (the measures). The distinction between dimensions and measures is also at the basis of most models for OLAP systems. However, as noted by several authors, this distinction has some drawbacks, mainly because some operations easily expressible in relational algebra become cumbersome in multidimensional models. Some authors have proposed multidimensional models with a symmetrical treatment of measures and dimensions to cope with this problem (Agrawal, Gupta, & Sarawagi, 1997; Cabibbo & Torlone, 1997; Gyssens & Lakshmanan, 1997).

EXPRESSING MULTIDIMENSIONAL QUERIES

As noted earlier, the metaphor of the data cube and the concepts of facts, measures and dimensions are fundamental to both multidimensional data modeling and querying. In particular, techniques proposed in the literature and/or implemented in commercial systems to retrieve such data are based on the idea of determining the cube of interest and then navigating along the dimensions, increasing or decreasing the level of detail through roll-up and drill-down or selecting specific subparts of the cube through the operation of slice and dice.

The query languages for multidimensional data support both these standard operations and additional ones for performance of more sophisticated calculations. A first broad distinction can be made among:

- Languages based on an algebra (usually an extension of the relational algebra), where queries are expressed by using operators representing facts, measures and dimensions. Examples of these languages are the grouping algebra proposed by Li and Wang (1996) and the algebra for "symmetrical" cubes (Agrawal, Gupta, & Sarawagi, 1997).
- Languages based on a calculus (usually an extension of the relational calculus), where queries are expressed in a more declarative way. An example is MD-CAL, a multidimensional calculus for fact tables (Cabibbo & Torlone, 1997).
- Visual languages, usually relying on an underlying algebra, and based on a more interactive and iconic querying paradigm: This is the approach of most commercial OLAP products. A visual query language for statistical aggregate data was proposed by Rafanelli, Bezenchek, and Tininini (1996) and for the MD model by Cabibbo and Torlone (1998).

Multidimensional query languages can also be classified by the type of model used to represent the data:

- Query languages based on a relational representation of multidimensional data, hence based on extensions of the relational algebra and calculus.

ROLAP (Relational OLAP) systems are based on such languages. Examples are that described by Gyssens and Lakshmanan (1997), and SQL(H) (Jagadish, Lakshmanan, & Srivastava, 1999), an extension of SQL to embed hierarchies in the relational model.
- Query languages based on specifically designed multidimensional models, usually based on an abstraction of cubes or fact tables, to which the algebra operators are applied. Many commercial, so-called MOLAP (Multidimensional OLAP) systems are based on these languages.

Some further issues must be considered to express multidimensional queries in a statistical environment, mainly because: (i) it is crucial to preserve privacy in statistical databases and excessively detailed aggregate data can disclose confidential information about individual persons; and (ii) much of the data in statistical databases originates from sample surveys, and corresponding aggregate data are not significant at the finest levels of detail.

As a consequence, an interface enabling free navigation of dimension hierarchies would often lead the user to express a query referring to non-permissible data. Specific techniques are therefore required to establish a trade-off between the characteristic freedom of multidimensional navigation and the constraints of significance and disclosure prevention imposed by statistical databases.

EVALUATING MULTIDIMENSIONAL QUERIES

Query evaluation is the process of translating the query expressed in the chosen language into an (efficient) evaluation plan which retrieves the necessary information and calculates the required results. Note that the adoption of a particular query language may have a significant impact not only on how easily the desired query can be expressed, but also on how efficiently it can be evaluated.

A fundamental requirement of OLAP systems is the ability to perform multidimensional analyses in on-line response times. As multidimensional queries usually involve a huge amount of data to be aggregated, the only way to achieve this is by pre-computing some queries, storing the answers permanently in the database and reusing them (almost exclusively) in the query evaluation process. These pre-calculated queries are commonly referred to as materialized views. From an abstract point of view, the determination of the evaluation plan can be seen as the process of transforming the source query into an equivalent target query, referring (possibly exclusively) to the materialized views. This is known in literature as the (aggregate) query rewriting problem. Aggregate query rewriting techniques are usually based on:

- A collection of rewriting (transformation) rules usually applicable to queries of specific forms. The rules are iteratively applied to the original query to decompose it into more elementary sub-queries, until all (or most of) the sub-queries obtained are available as materialized views. This is the approach in, for instance, Gupta, Harinarayan, and Quass (1995).
- The detection of homomorphisms from the views to the query (as for non-aggregate queries). This enables the query optimizer to distinguish queries usable for rewriting from the useless ones. Usable queries are then combined, usually in conjunctive (AND) form. Rewriting techniques of this type are presented in Srivastava, Dar, Jagadish, and Levy (1996) and Grumbach, Rafanelli, and Tininini (1999).
- Query equivalence criteria, as in Cohen, Nutt, and Serebrenik (1999). The original query, the usable views and the rewritten query are all expressed through an extension of Datalog with aggregate functions (again COUNT, SUM, MIN and MAX) as the query language. Several candidate rewritten queries are considered, and the views in each body are unfolded (i.e., replaced by their body in the view definition). Finally, the unfolded candidate is compared with the original query to verify equivalence through use of known equivalence criteria for aggregate queries.

All rewriting algorithms proposed in the literature are based on the attempt to obtain a rewritten query with a particular form by using (possibly only) the available views. An interesting question is: “Can I rewrite more by considering rewritten queries of more complex form?” and the even more ambitious one “Given a collection of materialized views, is the information they provide sufficient to rewrite a query?” In Grumbach and Tininini (2000), the problem is investigated in a general framework based on the concept of query subsumption.

An interesting issue related to multidimensional query evaluation is the efficient computation of the materialized aggregate views. As for OLTP queries, the use of indexes can have dramatic consequences on the overall process efficiency. However, unlike OLTP queries where B+-tree based indexes (Comer, 1979) achieve excellent performances in almost all practical situations, there is no “universally good” index for OLAP queries, but rather a variety of techniques, each of which may perform well for specific data types and query forms but be inappropriate for others (Chan & Ioannidis, 1998; Jürgens & Lenz, 1999).

FUTURE TRENDS AND CONCLUSIONS

In this article, we have discussed the main issues related to the querying of multidimensional data. Querying is strictly connected to the way data is modeled, and particularly to the concepts of data cube, fact table, measure and dimension. Multidimensional queries have several features which distinguish them from conventional queries on transactional systems. Aggregation functions and dimension hierarchies play a major role in this distinction. Although there is a general consensus on the key concepts, a common unified framework for multidimensional data querying is still lacking, particularly a standardized query language independent from the specific storage technique (e.g., MOLAP and ROLAP).

We have presented the main characteristics of multidimensional query languages proposed in the literature and the main issues related to the evaluation of such queries, in particular the query rewriting problem, that is, how an original query can be transformed into an equivalent one expressed in terms of some materialized views and consequently more efficiently computable. A substantial effort is still needed in this field, and important contributions may stem from results in other research areas like logic programming and automated reasoning.

REFERENCES

Agrawal, R., Gupta, A., & Sarawagi, S. (1997). Modeling multidimensional databases. In International Conference on Data Engineering (ICDE'97) (pp. 232-243).

Cabibbo, L., & Torlone, R. (1997). Querying multidimensional databases. In Int. Workshop on Database Programming Languages (DBPL'97) (pp. 319-335).

Cabibbo, L., & Torlone, R. (1998). From a procedural to a visual query language for OLAP. In Int. Conference on Scientific and Statistical Database Management (SSDBM'98) (pp. 74-83).

Chan, C.Y., & Ioannidis, Y.E. (1998). Bitmap index design and evaluation. In ACM International Conference on Management of Data (SIGMOD'98) (pp. 355-366).

Cohen, S., Nutt, W., & Serebrenik, A. (1999). Rewriting aggregate queries using views. In ACM Symposium on Principles of Database Systems (PODS'99) (pp. 155-166).

Comer, D. (1979). The ubiquitous b-tree. ACM Computing Surveys, 11(2), 121-137.

Gray, J., Bosworth, A., Layman, A., & Pirahesh, H. (1996). Data cube: A relational aggregation operator generalizing group-by, cross-tab, and sub-total. In International Conference on Data Engineering (ICDE'96) (pp. 152-159).

Grumbach, S., Rafanelli, M., & Tininini, L. (1999). Querying aggregate data. In ACM Symposium on Principles of Database Systems (PODS'99) (pp. 174-184).

Grumbach, S., & Tininini, L. (2000). On the content of materialized aggregate views. In ACM Symposium on Principles of Database Systems (PODS'00) (pp. 47-57).

Gupta, A., Harinarayan, V., & Quass, D. (1995). Aggregate-query processing in data warehousing environments. In International Conference on Very Large Data Bases (VLDB'95) (pp. 358-369).

Gyssens, M., & Lakshmanan, L.V.S. (1997). A foundation for multi-dimensional databases. In International Conference on Very Large Data Bases (VLDB'97) (pp. 106-115).

Jagadish, H.V., Lakshmanan, L.V.S., & Srivastava, D. (1999). What can hierarchies do for data warehouses? In International Conference on Very Large Data Bases (VLDB'99) (pp. 530-541).

Jürgens, M., & Lenz, H.J. (1999). Tree based indexes vs. bitmap indexes " a performance study. In Intl. Workshop on Design and Management of Data Warehouses (DMDW'99).

Lenzerini, M. (2002). Data integration: A theoretical perspective. In ACM Symposium on Principles of Database Systems (PODS'02) (pp. 233-246).

Li, C., & Wang, X.S. (1996). A data model for supporting on-line analytical processing. In Conference on Information and Knowledge Management (CIKM'96) (pp. 81-88).

Rafanelli, M., Bezenchek, A., & Tininini, L. (1996). The aggregate data problem: A system for their definition and management. SIGMOD Record, 25(4), 8-13.

Shoshani, A., & Wong, H.K.T. (1985). Statistical and scientific database issues. IEEE Transactions on Software Engineering, 11(10), 1040-1047.

Srivastava, D., Dar, S., Jagadish, H.V., & Levy, A.Y. (1996). Answering queries with aggregation using views. In International Conference on Very Large Data Bases (VLDB'96) (pp. 318-329).

Vaisman, A.A., & Mendelzon, A.O. (2001). A temporal query language for OLAP: Implementation and a case study. In 8th International Workshop on Database Programming Languages (DBPL 2001) (pp. 78-96).

KEY TERMS

Data Cube: A collection of aggregate values classified according to several properties of interest (dimensions). Combinations of dimension values are used to identify the single aggregate values in the cube.

Dimension: A property of the data used to classify it and to navigate on the corresponding data cube. In multidimensional databases, dimensions are often organized into several hierarchical levels, for example, a time dimension can be organized into days, months and years.

Drill-down (/roll-up): Typical OLAP operation, by which aggregate data are visualized at a finer (/coarser) level of detail along one or more analysis dimensions.

Fact: A single elementary datum in an OLAP system, the properties of which correspond to dimensions and measures.

Fact Table: A table of (integrated) elementary data that are grouped and aggregated in the process of multidimensional querying.

Measure: A numeric value obtained by applying an aggregate function (like count, sum, min, max or average) to groups of data in a fact table.

Materialized View: A particular form of query whose answer is stored in the database to speed up the evaluation of further queries.

Query Rewriting: Process by which a source query is transformed into an equivalent one referring (almost exclusively) to a collection of materialized views. In multidimensional databases, query rewriting is fundamental to obtain acceptable (online) response times.

Slice and Dice: Typical OLAP operation by which specific subparts (“slices”) of the data cube are selected and arranged for visualization.

R

Real Options Analysis in Strategic Information Technology Adoption

Xiaotong Li
University of Alabama in Huntsville, USA

INTRODUCTION

Many information resource managers have learned to be proactive in today's highly competitive business environment. However, limited financial resources and many uncertainties require them to maximize their shareholders' equity while controlling the risks incurred at an acceptable level. As the unprecedented development in information technology continuously produces great opportunities that are usually associated with significant uncertainties, technology adoption and planning become more and more crucial to companies in the information era. In this study, we attempt to evaluate IT investment opportunities from a new perspective, namely, the real options theory. Its advantage over other capital budgeting methods like static discounted cash flow analysis has been widely recognized in analyzing the strategic investment decision under uncertainties (Amram & Kulatilaka, 1999; Luehrman, 1998a, 1998b). Smith and McCardle (1998, 1999) further show that option pricing approach can be integrated into standard decision analysis framework to get the best of the both worlds. In fact, some previous IS researches have recognized the fact that many IT investment projects in the uncertain world possess some option-like characteristics (Clemsons, 1991; Dos Santos, 1991; Kumar, 1996). Recently, Benaroth and Kauffman (1999) and Taudes, Feurstein and Mild (2000) have applied the real options theory to real-world business cases and evaluated this approach's merits as a tool for IT investment planning.

As all real options models inevitably depend on some specific assumptions, their appropriateness should be scrutinized under different scenarios. This study aims to provide a framework that will help IS researchers to better understand the real options models and to apply them more rigorously in IT investment evaluation. As the technology changes, the basic economic principles underlying the real options theory do not change. We do need to integrate the IT dimension into the real options based investment decision-making process. Using electronic brokerage's investment decision in wireless technology as a real-world example, we show the importance of adopting appropriate real options models in IT investment planning. By specifically focusing on the uncertainties caused by IT innovation and competition, our study also gives some intriguing results about the dynamics between IT adoption and the technology standard setting process.

REAL OPTIONS THEORY

It is generally believed that the real options approach will play a more important role in the highly uncertain and technology driven digital economy. Before reviewing the real options literature body that is growing very rapidly, we use an example to give readers an intuitive illustration of the values of real options and their significance in financial capital budgeting.

Pioneer Venture: The Value of a Growth Option

In this example, the management of a large pharmaceutical company wants to decide whether to acquire a young biomedical lab. If they decide to acquire it, they should provide $100,000 funding to cover the initial costs for the pioneer venture. Five years after the initial funding, the management will decide whether to stop the pioneer venture or to expand it significantly according to the market situation at that time. If they choose to expand it, additional $1,000,000 is needed. The cost of capital is assumed to be 15%. Five years after acquisition of the lab, the management will face two scenarios. The good scenario will occur with 60% likelihood, while the bad one will have 40% likelihood of happening. All expected future cash flows during the next 10 years are given in Table 1. Using standard capital budgeting method, we can find that the NPV for the pioneer venture is -$15,215. For the period of large-scale production, the NPV is -$71,873. As the NPVs for both periods are negative, it seems that the management should give up the acquisition. However, the acquisition will be a good investment if we consider the growth option associated with it. By acquiring the lab, the company also buys a growth option that enables it to

Table 1. Projected cash flows in the example of pioneer venture project

Year	Pioneer Stage	Larger Scale Stage	Total Cash Flows	Discount Rate
0	-$100,000		-$100,000	15%
1	$10,000		$10,000	
2	$10,000		$10,000	
3	$50,000		$50,000	
4	$50,000		$50,000	
5	$20,000	-$1,000,000	-$980,000	
6		$100,000	$100,000	
7		$100,000	$100,000	
8		$500,000	$500,000	
9		$500,000	$500,000	
10		$200,000	$200,000	
	Large Scale Stage	Good Scenario	Bad Scenario	Prob (good)
5	-$1,000,000	-$1,000,000	-$1,000,000	0.6
6	$100,000	$130,000	$55,000	
7	$100,000	$130,000	$55,000	
8	$500,000	$650,000	$275,000	
9	$500,000	$650,000	$275,000	
10	$200,000	$260,000	$110,000	
	NPV Pioneer Stage	-$15,215.42		
	NPV Large Scale Stage	-$71,872.54		
	NPV with Growth Option	$13,749.98		
	Value of the Option	$28,965.40		

expand the lab when the conditions are favorable 5 years later. In this case, the good scenario will occur with 60% likelihood. After simple calculation, it is easy to find that the growth option has a value of $28,965. Combining its value with the negative NPV during the pioneer venture period, the adjusted NPV of the acquisition is $13,750, which means this investment is strategically plausible.

Many researchers recognized the potential of this options pricing theory in capital budgeting because traditional DCF (discounted cash flows) technique has its inherent limitation in valuing investments with strategic options and many uncertainties. Table 2 gives a comparison between an American call option on a stock and a real option on an investment project. Despite the close analogy, some people may still question the applicability of option pricing theory on real options that are usually not traded in a market. However, Cox, Ingersoll and Ross (1985) and McDonald and Siegel (1984) suggest that a contingent claim on a non-traded asset can be priced by subtracting a dividend like risk premium from its growth rate.

Recent development in real option theory focuses on the valuation of more complicated real options like shared options, compounded options and strategic growth options. Dixit and Pindyck (1994) examine the dynamic equilibrium in a competitive industry. Their model suggests that a firm's option to wait is valuable when uncertainty is firm-specific. For industry-wide uncertainty, there is no value to wait because of the asymmetric effects of uncertainty.

FOUR CATEGORIES OF IT INVESTMENT OPPORTUNITIES

As shown in Figure 1, we have four types of IT investment opportunities based on the two criteria: (i). Shared opportunities with high IT switching costs; (ii). Shared opportunities with low IT switching costs; (iii). Proprietary opportunities with low IT switching costs; (iv). Proprietary opportunities with high IT switching costs. It is worth noting that each category has distinctive requirements on the application of real options models. We use the continuous-time model developed in McDonald and Siegel (1986) as a benchmark to show why we differentiate IT investment opportunities based on the two criteria. It basically suggests that the option to defer uncertain investment is very valuable and should be taken into account when a company makes investment decisions. A

Table 2. Comparison between an American call option and a real option on a project

AMERICAN CALL OPTION ON STOCK	REAL OPTION ON A PROJECT
Current Stock Price	Present Value of Expected Cash Flows
Option Exercise Price	Investment Cost of a Project
Right to Exercise the Option Earlier	Right to Invest in the Project at any time before the Opportunity Disappears
Stock Price Uncertainty	Project Value Uncertainties
Option Price	Value of Managerial Flexibility Associated with the Project
Expiration Time	Time Window of the Investment Opportunity
Traded in Financial Market	Usually not Traded
Easy to Find a Replicating Portfolio	Hard to Find a Replicating Portfolio

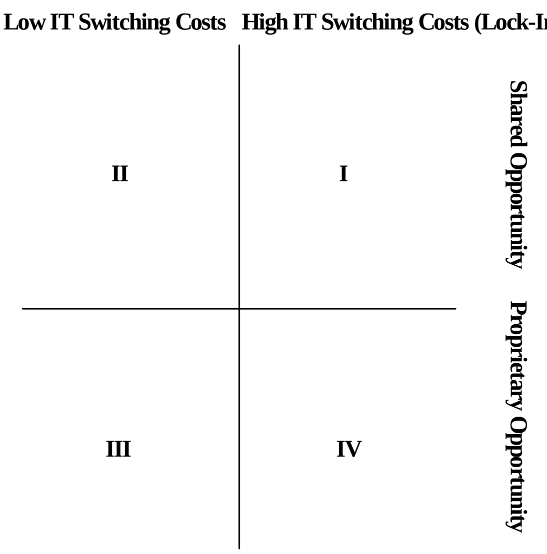
major assumption of this model is that there is no competitive erosion; in other words, the investment project is a proprietary opportunity. Without this assumption, the value of the project should not follow the symmetric geometric Brownian motion described in their model. The reason is simple: the existence of potential competition makes the distribution of future project value asymmetric, with high project value less likely to occur. It is worth noting that the well-known Black-Scholes option pricing formula is also based on the assumption that the underlying asset price follows the geometric Brownian motion. In the real business world, most investment opportunities are shared or at least partially shared. Especially in the IT business sector where intensive competition is pervasive, those real options models assuming symmetric uncertainty in investment opportunity value are generally inappropriate. Intuitively, competition pressure will decrease the value of the option to defer an investment. There are usually two approaches to deal with this issue. One approach is to model the competitive entries as exogenous shocks. For examples, Dixit and Pindyck (1994) and Trigeorgis (1991) use a Poisson Jump process to describe the competitive arrivals. Their studies show that the effect of the competitive erosion can be expressed as the following equation

Strategic NPV= NPV+ (Value of Option to Wait - Competitive Loss).

In other words, strong competition will restrict managerial flexibility if the investment opportunity is shared. We need to evaluate the investment opportunity in the context of the real options theory by considering the growth option and the waiting option simultaneously. Alternatively, we can incorporate the preemptive effect into the standard real options models. For example, Li (2001) proposes a real options model with strategic consideration based on the model in McDonald and Siegel (1986).

The other criterion we used to categorize different IT investment opportunities is the IT switching cost. We all know that future uncertainty makes the options embedded in an investment opportunity valuable. Theoretically, there is no need to single out technology uncertainty from

Figure 1. Four categories of IT investment opportunities



all other uncertainties in the real options model. All these uncertainties have the same effect: they make the future payoff of an investment project less predictable. However, we will concentrate on the technology uncertainty in this study because it plays a pivotal role in affecting IT investment payoff. Perhaps the most important question that management faces before committing an IT investment is whether the technology adopted is the right choice. More specifically, will the adopted technology be the best solution to maximize the expected investment payoff? Clearly there is not a simple answer to this question because there are so many uncertainties involved. Some very promising or popular IT solutions may become obsolete in a few years. In some other cases, some neglected IT solutions may evolve to be the standard solution. Nevertheless, most technology uncertainties can be resolved as the process of technology competition goes forward. A typical process of technology competition includes:

- Problem identification: An important problem is identified and new technology is sought to solve it.
- Technology solutions proposition: Several technology developers/vendors propose different solutions to solve the problem.
- Solution testing and comparison: Different technology solutions are competing in the market and their effectiveness is tested and compared.
- Technology standardization: The best solution will flourish over time. Based on it, the technology to solve the problem will be standardized.

For many IT investment projects, decision makers face an uncertain technology environment where several IT solutions are competing in the market. Obviously, the future successes of these projects will to some extent depend on whether the IT solutions adopted will win the technology competition. Consequently, decision makers do have an incentive to use the deferring option to let more technology uncertainties be resolved. Under this scenario, many option-to-wait models can be easily extended to find the optimal investment strategy. However, to apply these real options models we must presume that there are significant technology switching costs once an IT solution is adopted. Otherwise, the uncertainties in technology competition will not make the option to wait valuable because the decision makers can easily switch to other IT solutions after they implement the investment project. As pointed out by Shapiro and Varian (1998), the IT switching costs are very significant in many cases. They use the term "technology lock-in" to describe the situation where management has little flexibility to switch to other technology solutions once they have adopted one IT solution.

Now it should be clear why we use IT switching cost as the second criterion to classify different IT investment opportunities. When the IT switching cost is significant (technology lock-in), the option to wait is valuable. Therefore, real options analysis should concentrate on the managerial flexibility in deferring an IT investment to let more technology uncertainties be resolved. When the switching cost is low, high IT uncertainties cannot be used to justify the wait-and-see policy. On the contrary, we should use real options analysis to quantify the value of the option to switch that usually makes an investment opportunity more appealing to the management.

To summarize our discussion, let us look at the four categories of IT investment opportunities based on the two criteria.

- Category I: Shared investment opportunity with high IT switching cost. For this type of IT investment opportunity, we must consider both the strategic benefit of early preemptive investment and the valuable option to wait. Potential competitive pressure forces investors to be proactive. However, preemptive investment will incur the loss of the valuable option to wait. So for this type of IT investment opportunity, the key in the real options analysis is to consider the strategic growth option and the option to wait at the same time. By balancing the two contradictory effects, we can find the optimal investment point at which the expected investment payoff will be maximized.
- Category II: Shared investment opportunity with low IT switching cost. For this type of IT investment opportunity, early preemptive investment is usually the best strategy. As we discussed before, it is beneficial to invest early to preempt potential competitors. Moreover, IT uncertainties will not make the wait-and-see strategy more appealing because the IT switching cost is low. Therefore, real options models should be used to quantify the values of the growth option and the switching option embedded in the IT investment opportunity.
- Category III: Proprietary investment opportunity with low IT switching cost. It is worth noting that the option to wait is a very valuable component of a proprietary investment opportunity. However, technology uncertainty will not contribute a lot to the value of the option to wait because the IT switching cost is low for investment opportunities in this category. So in the real options analysis we should pay attention to other business uncertainties that may increase the value of the option to wait.
- Category IV: Proprietary investment opportunity with high IT switching cost. Wait-and-see is the dominant strategy for this type of IT investment

opportunity. So real options analysis should concentrate on the option to defer an investment. With the presence of technology lock-in, decision makers should be more patient before they commit a proprietary investment.

In the real business world, an IT investment opportunity may dynamically evolve from one category to other ones. So decision makers should be very cautious when they conduct real options analysis. In the next section, we use a real-world case to show the importance of adopting appropriate real options models as the IT investment opportunity evolves.

CONCLUSION

Although some recent studies recognized the potential of real options theory in evaluating strategic IT investment opportunities, we believe that the applicability of various real options models should be scrutinized under different scenarios. Standard real options models assuming symmetric uncertainty in future investment payoffs cannot be directly applied to the shared opportunities because of the competitive erosion. With the presence of potential competitive entry, real options analysis should balance the strategic benefit of preemptive investment and the value of the option to wait. IT switching cost is another important factor we must consider when we conduct real option analysis. As high IT switching cost or technology lock-in is very common in the digital economy, decision-makers should pay more attention to the technology uncertainties before committing early investment to preempt their competitors.

REFERENCES

Amram, M., & Kulatilaka, N. (1999). *Real options, managing strategic investment in an uncertain world*. Boston: Harvard Business School Press.

Benaroth, M., & Kauffman, R.J. (1999). A case for using real options pricing analysis to evaluate information technology project investments. *Information Systems Research, 10*(1), 70-88.

Clemons, E.K. (1991). Evaluating strategic investments in information systems. *Communications of the ACM, 34*(1), 22-36.

Cox, J., Ingersoll, J., & Ross, S. (1985). An intertemporal general equilibrium model of asset prices. *Econometrica, 53,* 363-84.

Dixit, A., & Pindyck, R. (1994). *Investment under uncertainty*. Princeton University Press.

Dos Santos, B.L. (1991). Justifying investment in new information technologies. *Journal of Management Information Systems, 7*(4), 71-89.

Kumar, R. (1996). A note on project risk and option values of investments in information technologies. *Journal of Management Information Systems, 13*(1), 187-93.

Li, X. (2001). *Optimal timing for brokerage to go wireless-a real options approach*. Unpublished PhD dissertation. The University of Mississippi.

Luehrman, T. (1998a, July-August). Investment opportunities as real options: Getting started with the numbers. *Harvard Business Review*, 51-64.

Luehrman, T. (1998b, September/October). Strategy as a portfolio of real options. *Harvard Business Review*, 89-99.

McDonald, R., & Siegel, D. (1984). Option pricing when the underlying asset earns a below-equilibrium rate of return: A note. *Journal of Finance, 39*(1), 261-265.

Shapiro, C., & Varian, H. (1998). *Information rules: A strategic guide to network economy*. Harvard Business School Press.

Smith, J., & McCardle, K. (1998). Valuing oil properties: Integrating option pricing and decision analysis approach. *Operations Research, 46*(2), 198-218.

Smith, J., &. McCardle, K. (1999). Options in the real world: Some lessons learned in evaluating oil and gas investments. *Operations Research, 47*(1), 1-15.

Taudes, A., Feurstein, M., & Mild, A. (2000). Options analysis of software platform decisions: A case study. *MIS Quarterly, 24*(2), 227-43.

Trigeorgis, L. (1991). Anticipated competitive entry and early preemptive investment in deferrable projects. *Journal of Economics and Business, 43*(2), 143-145.

KEY TERMS

Black-Scholes Option Pricing Model: A model that is used to calculate the value of an option by taking into account the stock price, strike price and expiration date, the risk-free return, and the standard deviation of the stock's return.

Deferred Option: Option to defer a project or an investment gives a firm an opportunity to make an investment at a later point in time.

Managerial Investment Flexibility: Flexibility in the timing and the scale of an investment provided by a real investment option.

Net Present Value (NPV): The present value of an investment's future net cash flows minus the initial investment.

Option: The right, but not the obligation, to buy or sell an asset by a pre-specified price on before a specified date.

Real Options Theory: Financial valuation tool that helps in calculating the value of managerial flexibility under uncertainties.

Switching Costs: Switching costs refer to the hidden costs consumers face when switching from one product or technology to another in the marketplace.

Reasoning About User Preferences

Tamara Babaian
Bentley College, USA

INTRODUCTION

Personalization has been identified as a key task to the success of many modern systems. As Riecken (2000, p. 28) writes in the editorial of the special issue of the *Communications of the ACM* devoted to this subject, "personalization means something different to everyone." Personalization can take various forms; however, it can be broadly described as the set of mechanisms by which an application is tailored to a particular end user and his or her goal.

Modern systems typically have a large set of features designed to carry out a multitude of tasks, or operate using an enormous wealth of information available on the Internet. The effectiveness with which a system is able to help its user achieve the desired effect, as well as the user's satisfaction from interacting with the system, depend critically, among other factors, on the user's ability to identify and use relevant, customizable options. Configuring the system for optimal performance requires that the user specifies his or her individual preferences related to many different tasks. However, the user's ability to provide this kind of personal information is often greatly impaired by the following drawbacks in the way personalization is implemented.

- Customization is carried out as a separate process that is taken out of context of the task in which such personal information is used, thus obscuring from the user the purpose and advantages of supplying such information.
- The amount of potentially useful personal information is sometimes overwhelming, thus the systems are installed with a set of settings that are considered typical. Further customization has to be initiated by the user. However, inexperienced users rarely take advantage of customization even if they are aware of potential benefits due to the lack of information on available options. As a result, experience demonstrates (Manber, Patel, & Robison, 2000) that many users shy away from customization when they can benefit from it a great deal.

The items above characterize the shortcomings of a typical process of configuring customizable features. On the other hand, there are problems that the developers of software face in *designing* for personalization. As Pednault (2000) points out, the underlying representation of "the human-side and the technology-side" is the key. However, representations currently in use, at times, lack flexibility to be easily adjustable and reusable. This is largely a consequence of the absence of a rigorous model of what constitutes personalization. The lack of such model results in ad hoc representations used by most systems.

The approach to personalization that we present here is inspired by the view of interfaces as the means of collaboration between a human user and a computer system, articulated by Shieber (1996), in which the collaborative model of human-computer interaction is contrasted to the view of a system as a mere set of tools available to the user. As a theoretical framework, collaboration theory and its existing philosophical and formal mathematical accounts (Bratman, 1992; Cohen & Levesque, 1991; Grosz & Kraus, 1996) can inform both design and usability analysis of systems, and highlight problems that need to be addressed to make interfaces better collaborative partners. Examples of interfaces that have been created following this view have already been built and are described in Babaian, Grosz, & Shieber (2002), Ortiz and Grosz (in press), Rich, Sidner, and Lesh (2001), and Ryall, Marks, and Shieber (1997).

Theories of collaboration postulate:

1. commitment of the parties to a shared goal, and
2. sharing of knowledge and communication in an effort to establish agreement and mutual knowledge of the recipe for completing the task

as the key features of a collaborative activity. Stemming directly from this view, in our approach the collaborator system has the ability to elicit personal information at the time when the user is most motivated to provide it, that is, when the system is processing a task for which such information is critical. The novelty of our approach and its implementation also lies in defining the task of collecting personal information declaratively via informational goals and preconditions on the actions that the system takes in response to a user's request. This is enabled by

(a) the use of a knowledge base that stores all gathered preference information, and

(b) an automated reasoning and planning system that can reason autonomously about knowledge, lack of knowledge, and actions that the system may take to acquire the necessary but missing information.

The system can perform information gathering autonomously (by inspecting available personal information, such as, for example, a person's Internet bookmarks) as well as by direct user querying. This approach to personalization ensures gradual adaptation of the system to the user's preferences. At the same time, the declarative definition of personal information and its relationship to system actions make it easy to fine-tune personalization options, resulting in a more easily adjustable and extendable design.

BACKGROUND

The problem of end-user tailoring, also known as customization of software, is not new (see, for example, Morch, 1997). Recent explosion of the Internet and its ubiquity in our everyday life have created new challenges and opportunities for the advancement of research on this subject, in particular, in the area of customizing information-access interfaces. Numerous works have addressed the issue of information overload and the resulting need for effective information retrieval and presentation of the results that is tailored to the needs of each individual visitor. A thorough review of these works is beyond the scope of this chapter; however, we briefly describe the leading approaches and provide the reader with a set of further references. Availability of logs of Web site usage has provided an excellent opportunity and an exciting domain for technologies such as machine learning and data mining (for a review see Anderson, 2002; Pierrakos, Paliouras, Papatheodorou, & Spyropoulos, 2003). In the artificial-intelligence community, two approaches to automated personalization on the Web have been explored and used most successfully: adaptive Web sites and collaborative filtering. Adaptive Web sites and Web site agents (e.g., Pazzani & Billsus, 1999; Perkowitz & Etzioni, 2000) attempt to dynamically tailor the layout and the contents of a Web site or suggest a navigation path for each individual user by observing the user's initial interaction with the Web site and matching it to the previously observed behaviors of others. Likewise, collaborative filtering (Amazon.com is probably the most familiar example) is a technique that creates recommendations that are based on the choices of previous users with similar interests or requests.

Recently, many traditional concepts and techniques of artificial intelligence have been applied in the area of intelligent user interfaces, in particular, to interface personalization (we refer the reader to the *Proceedings of the International Conference on Intelligent User Interfaces*). An article by Weld et al. (2003) surveys a set of research projects aimed at developing representations and methods for user-directed customization and automatic system adaptation of various kinds of application interfaces.

Many applications of machine-learning and data-mining technologies to Web-based computing have been enabled by the availability of logs recording various details of the interaction of millions of users with the Web sites. At the same time, non-Web-based systems (e.g., common desktop editors, spreadsheets, etc.) have also benefited from the emerging culture of personalization and now commonly incorporate a few personalizable features. Nevertheless, the advancement of research in personalization of common desktop applications has been lagging behind, partly due to the absence of detailed data on their actual usage. The method of software customization presented in this article is applicable to a broad set of software tools and not limited to just Web-based systems.

FUTURE TRENDS

Goal-Directed Personalization in Writer's Aid

Writer's Aid (Babaian et al., 2002) is a system that works simultaneously with an author writing a document, helping him or her with identifying and inserting citation keys, and autonomously finding and caching papers and associated bibliographic information from various online sources.

At its core, Writer's Aid contains

(a) a knowledge base that contains a system's knowledge about the state of the world, and
(b) an automated planning system.

The planner is given a description of the actions that Writer's Aid can execute and works to automatically combine these actions into a plan that will achieve a specified goal. Each action is described via preconditions that must be true prior to executing the action, and effects that the action brings about. Plan generation is accomplished by representing both goals and actions using a logic-based language and using a specialized reasoning engine that can infer what is true after performing a sequence of actions. For an example, consider the following action of searching a user's personal directories for bibliographic collections.

Action 1: FindLocalBibliographies

Preconditions: none
Effects: knowing locations of all bibliographic collections of a user

Personalization in Writer's Aid consists of the initial tune-up of the system to the user's parameters, and the dynamic personalization that occurs while Writer's Aid works on accomplishing a user goal.

Initial Tune-up

Initial tune-up occurs at the time of the installation. The goal of the initial tune-up is to identify and enter certain user-specific parameters, such as, for example, the user's own locally stored bibliographic collections, his or her preferred online bibliographies, and so forth. To direct the system to collect data about the location of local bibliographies, it is sufficient to post the following goal on the list of goals to be accomplished during the tune-up.

Personalization Goal 1 = knowing the locations of all of the user's bibliographic collections

In response, Writer's Aid will generate a plan (in this case consisting of a single Action 1) described above, which accomplishes Personalization Goal 1, and thus provides Writer's Aid with access to the user's personal bibliographies.

This approach to the initial customization, directed by a set of declarative goals that the planner automatically finds a way to achieve, greatly simplifies the personalization process. Personalization goals are now separated from the rest of the system code, and the overall personalization design is much more flexible then the typical hard-coded procedure.

Dynamic Personalization

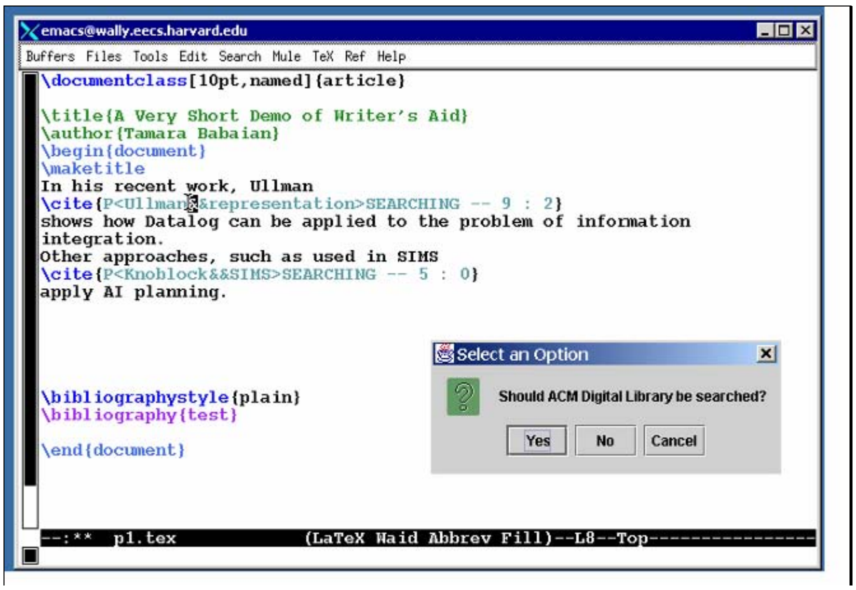
Imagine the following scenario. Writer's Aid is working to locate a viewable version of a paper that the user has requested. The plan for locating the paper includes an action of querying a known paper collection, namely, the ACM digital library. In order to avoid wasting time on searching collections of papers on subjects unrelated to the user's research field, this action contains a precondition that the paper collection be one of the user's *preferred* collections.

Action 2: QuerySourceForPaper (source, paper)

Precondition: source must be user's preferred source
Effects: knowing whether source contains viewable version of paper.

Writer's Aid does not know if ACM digital library is the user's preferred bibliography, so it cannot establish the precondition unless it executes an action (namely,

Figure 1. In the editor window, the user has entered two citation requests (identified by the blue /cite labels). While working on a request for papers related to keywords "Ullman" and "representation," Writer's Aid is asking the user whether ACM Digital Library is one of his or her preferred sources.



Action 3 described in Figure 1) of asking the user himself or herself to obtain the necessary information.

Action 3: AskUserAboutSource (source)

Precondition: user permits system to post questions
Effects: knowing whether source is a user-preferred source

The user's response determines whether the ACM digital library will be queried; it is also recorded in the Writer's Aid knowledge base for future use.

Dynamic personalization occurs gradually, always within a context of a particular task, thus eliciting the user's input at the time it is used and providing the user with the understanding of how the personal information is being used by the system.

DISCUSSION

We have presented a novel approach to personalization, which involves mixed-initiative interactions between the user and the computer system. This approach uses an explicit representation of the knowledge about the domain and the user, and relies on automated reasoning and planning techniques for

(a) identifying the necessary yet unknown personal-preference information, and
(b) constructing and executing a plan to obtain such information from various sources, including the user.

Note that the classical planning framework, in which the planning agent is assumed to have a *complete* specification of the world state, cannot be used in the system interface for the task of representing and acquiring relevant domain and user-preference information because, initially, the system has only incomplete knowledge about the user. Thus, the formalism used for the knowledge base and the planning system should satisfy the following needs.

1. It should be able to effectively represent and reason with only a *partial*-state specification. Indeed, an interface that adapts to a particular environment and user should be able to start with little or no knowledge of the domain and the user, dynamically identify the need for such knowledge, and collect it using information-gathering actions.
2. The planner should perform such information gathering in a nonredundant way and be able to draw all relevant and valid conclusions from the information it has collected, especially when the user is one of the sources of information. If the system ever repeated a question for which it had already obtained the user's answer, the user would likely be annoyed and stay away from taking time to provide any additional personalization information.

The planners that do not rely on the assumption of complete information (also known as a closed-world assumption, or CWA) are said to operate in an open world. Clearly, a planner that is used for collecting customization information instead of relying on the predefined default must be an open-world planner since it must assume that the information is initially unknown and plan to collect it.

The planning formalism used in Writer's Aid, PSIPLAN (Babaian, 2000; Babaian & Schmolze, 2000), satisfies these requirements: It effectively represents and reasons about the known and the unknown given a partial specification of the world. Furthermore, it includes an algorithm for planning in an open world that operates to ensure nonredundancy of information gathering.

Personalization via knowledge preconditions remedies commonly occurring problems with customization, outlined in the introduction, by adopting a mixed-initiative approach to customization. However, special attention should be given to aspects of mixed-initiative interfaces to assure the system acts in a manner that does not greatly disrupt the user's ongoing activity. On the other hand, the user must have access to the same customization data as the system and be able and aware of the way of modifying those settings at any time.

Deployment and experimental evaluation will doubtlessly identify ways of further improvement of dynamic personalization. A set of principles of mixed-initiative user interfaces introduced by Horvitz (1999), and the recent study of instant-messaging interruption on the user's performance in ongoing computing activity (Cutrell, Czerwinski, & Horvitz, 2001) can provide a starting point for further investigations.

CONCLUSION

We presented a method of software customization that is based on an explicit declarative representation of (a) user preferences and (b) knowledge preconditions of system actions. Automated reasoning about actions and change is used to relate particular user preferences to system actions. The advantages of using this framework include the ability of the system to recognize the lack or availability of user-preference information that is relevant to an action immediately prior to the system performing the action via a precondition satisfaction check. Furthermore, by employing automated planning, the system is capable of creating and executing a plan aimed at collecting the missing preference information.

Representing a personalization task via a set of information goals addresses the problems with the way personalization is approached in most modern systems, outlined in the beginning of this paper, in the following ways.

- It leads to preference elicitation that occurs within the context of the particular task that requires personal information, thus informing the user of his or her choices, motivating the response, and ensuring its accuracy.
- Personalization occurs gradually at the times when the personal information is critical to the satisfaction of a user's goal and is initiated by the computer system, thus relieving the user from the potentially time-consuming task of specifying all preferences at once.
- Personalization defined declaratively via information goals separates customization of the interface from the overall system architecture, making the interface more easily adjustable and extendable.

REFERENCES

Anderson, C. R. (2002). *A machine learning approach to Web personalization*. PhD thesis, University of Washington.

Babaian, T. (2000). *Knowledge representation and open world planning using psi-forms*. PhD thesis, Tufts University.

Babaian, T., Grosz, B. J., & Shieber, S. M. (2002). A writer's collaborative assistant. *Proceedings of Intelligent User Interfaces'02*, 7-14.

Babaian, T., & Schmolze, J. G. (2000). PSIPLAN: Open world planning with psi-forms. *Proceedings of the Fifth International Conference on Artificial Intelligence Planning and Scheduling, AIPS'00.*

Bratman, M. E. (1992). Shared cooperative activity. *The Philosophical Review, 101*(2), 327-341.

Cohen, P., & Levesque, H. (1991). Teamwork. *Nôus, 25*, 487-512.

Cutrell, E., Czerwinski, M., & Horvitz, E. (2001). Notification, disruption, and memory: Effects of messaging interruptions on memory and performance. *Proceedings of Human-Computer Interaction: INTERACT '01*, 263-269.

Grosz, B. J., & Kraus, S. (1996). Collaborative plans for complex group action. *Artificial Intelligence, 86*(2), 269-357.

Horvitz, E. (1999). Principles of mixed-initiative user interfaces. *Proceedings of CHI'99*, 159-166.

Manber, U., Patel, A., & Robison, J. (2000). The business of personalization: Experience with personalization of Yahoo! *Communications of the ACM, 43*(8), 35-39.

Morch, A. (1997). Three levels of end-user tailoring: Customization, integration, and extension. In M. Kyng & L. Mathiassen (Eds.), *Computers and design in context* (pp. 51-76). The MIT Press.

Ortiz, C., & Grosz, B. (2002). Interpreting information requests in context: A collaborative Web interface for distance learning. *Autonomous Agents and Multi-Agent Systems*, 5, 429-465.

Pazzani, M. J., & Billsus, D. (1999). Adaptive Web site agents. In O. Etzioni, J. P. Müller, & J. M. Bradshaw (Eds.), *Proceedings of the third international conference on autonomous agents (Agents'99)* (pp. 394-395). Seattle, WA: ACM Press.

Pednault, E. P. D. (2000). Representation is everything. *Communications of the ACM, 43*(8), 80-83.

Perkowitz, M., & Etzioni, O. (2000). Adaptive Web sites. *Communications of the ACM, 43*(8), 152-158.

Pierrakos, D., Paliouras, G., Papatheodorou, C., & Spyropoulos, C. D. (2003). Web usage mining as a tool for personalization: A survey. *User Modeling and User-Adapted Interaction, 13*(4), 311-372.

Rich, C., Sidner, C., & Lesh, N. (2001). Collagen: Applying collaborative discourse theory to human-computer interaction. *AI Magazine, 22*(4), 15-25.

Riecken, D. (2000). Personalized views of personalization. *Communications of the ACM, 43*(8), 26-28.

Ryall, K., Marks, J., & Shieber, S. (1997). An interactive constraint-based system for drawing graphs. *Proceedings of UIST*, 97-104.

Shieber, S. (1996). A call for collaborative interfaces. *ACM Computing Surveys, 28A.*

Weld, D., Anderson, C., Domingos, P., Etzioni, O., Gajos, K., Lau, T., et al. (2003). Automatically personalizing user interfaces. *Proceedings of IJCAI-2003*, 1613-1619.

KEY TERMS

Automated Planning: An area of artificial intelligence concerned with solving the problem of finding a sequence of actions that transforms a specified initial state into a specified goal state.

Closed-World Assumption (CWA): An assumption made in classical planning that all facts that are not implied by the knowledge base are *false*. CWA rests on the assumption of *completeness* of knowledge.

Collaborative Interface: Interface in which the interaction between the user and the system is guided by the principles of collaborative behavior.

Dynamic Personalization: End-user tailoring that occurs during the regular use of the system within a context of a user task as opposed to a specific context-free activity directed at the selection of customization options.

Knowledge-Based Personalization: Personalization that is based on (a) explicit representation of knowledge about the user and the domain of operation and (b) automatic inference from that knowledge.

Open-World Assumptioin (OWA): An assumption made while reasoning with *incomplete* knowledge: all facts that are not implied by the knowledge base are assumed to be *unknown*.

Personalization: A set of mechanisms built into a system in order to tailor it to a particular user and his or her goal .

R

Recursive Nature of the Market for Enterprise Applications

Fergal Carton
University College Cork, Cork, Ireland

Frédéric Adam
University College Cork, Cork, Ireland

INTRODUCTION

The paper explores the recursive elements of the market for enterprise systems by examining the evolution of the sales discourse from vendors of enterprise applications. Enterprise systems are continually sold and implemented, on the basis that greater integration of the modules supporting business functions is a good thing. In this paper we question this assumption, principally based on the redundancy of much of the information that is produced.

Following the development of computer applications from early Material Resources Planning (MRP) days through to today's latest offerings in the form of Enterprise Resource Planning II (ERP II), we try to understand the circumstances which have generated the requirement (needs discourse). In parallel, it is posited that the sales discourse is characterised by a continual extension and re-packaging of existing solutions, responding to the business users' evolving requirements with ever greater integration between the operational modules. This tendency to over-integrate exacerbates the problem of information overload that is experienced by managers trying to monitor the organisation's performance, efficiency and effectiveness.

MRP was once implemented to gain better visibility and control of inventory, because it was understood that this was the most costly element of the cost of goods sold. Reducing inventory levels is a well understood management goal in most manufacturing organisations. On the other hand, the abundance of information that has accompanied the gradual computerisation of business functions doesn't seem to elicit a similarly "economical" attitude towards information. Instead of encouraging information excess, we argue in favour of a "Just-in-Time" approach to information provision, where appropriate information is delivered where and when it is needed, rather than exhaustive information being available to all. Going as far back as the fundamental design issues of enterprise applications, we question whether business value can be gained from continually integrating business functions into a single data structure.

BACKGROUND

The focus of industry in the years following the Industrial Revolution was on providing as much output as possible, as opposed to controlling inventory (Mokyr, 2001). With this change came the increasing need for systems to support the increasingly complex nature of mass production facilities and activities (O'Gorman, 2004).

Research had shown, however, that the main problem of managers was not a lack of relevant information, rather an overabundance of irrelevant information (Ackoff, 1967). In that era, the constraining factor on the level of computerisation was cost. The gradual commoditisation of technology has meant that storage hardware is (relatively) cheap, therefore there is no culture of economy with its use (or abuse).

This "changing cost balance" has been attributed to the original growth in the uptake of computer applications to support MRP (Miller & Sprague, 1975). Although the MRP logic was already available and widely applied, the use of systems had been hitherto prohibitively expensive for most businesses. Computation costs were falling as inventory costs were rising. The rapid update capability of computers, coupled with the MRP logic and appropriate data, made it possible for managers to cope intelligently with the thousands of changes that inevitably occur between the planning and execution of primary tasks.

The natural propensity of computer manufacturers is to sell new systems that use lots of computer time (Miller & Sprague, 1975). The same increase in price performance ratio prompted the adoption Enterprise Resource Planning (ERP) systems in the 1990's, integrated systems capable of uniting and correlating the basic units of the business transaction (from sales order to finished goods, from demand forecast to master production schedule).

In order to achieve this integration, ERP systems rely on large central relational databases. The amount of storage and memory required to manipulate and operate these databases grew in tandem with the improvement in cost/performance of the hardware. Furthermore, software houses gradually moved away from the client/server model to the "thin client", capable of running on any PC with a browser. For the first time it was feasible for an organisation to operate its entire transaction processing infrastructure from a remote centralised server, using the internet to deliver functionality to the desktop.

Sammon et al. (2003) describes these 2 components of ERP systems as the solution to "operational" integration problems and "informational" requirements of managers. These are the same concepts expressed by Zuboff (1988) in describing the use of technology not only to automate manual tasks, but also to "informate" management tasks, such that "events, objects and processes become visible, knowable and shareable in a new way".

ERP systems are therefore expected to deliver the following benefits: (1) reduce costs by improving efficiencies through computerization; and (2) enhance decision-making by providing accurate and timely enterprise-wide information (Poston and Grabski, 2001).

Whether these centralized information systems really are capable of delivering both types of benefit has been a topic of debate for some time. "The notion that a company can and ought to have an expert (or a group of experts) create for it a single, completely integrated super-system – an MIS – to help it govern every aspect of its activity is absurd", according to Dearden (1972).

The Trend Towards Greater Integration

In a traditional manufacturing organisation, materials accounted for 75-80% of the total cost of provision of the cost or service (O'Gorman, 2004). The attitude of planners in the 70's was therefore to develop methods that minimised inventory excess (in materials, WIP or finished goods).

Conversely, the focus of today's ERP vendors as they strive for ever greater integration has been to provide as much information as possible (analogous to a "build to stock" model in manufacturing terms) rather than trying to control it.

ERP systems, with their focus on the integration of processes and their dependence on the integrity of data at the point of entry, can be compared to virtual assembly lines, where each stage in the business process is optimised for the throughput of high volumes of transactions.

A major downside to this level of integration of business processes is that informational "stock-outs" can occur (one small piece of relatively unimportant information missing can block a business critical transaction). A classic example would be an exchange rate missing blocking an invoice from printing.

One of the benefits of employing what ERP vendors call "best practice" is that all transactions must fit in the same system model, regardless of the relative importance of the transactions. Thisignores the 80:20 rule as elaborated by Orlicky (1975), in what is probably the definitive book on MRP, according to Browne, Harhen & Shivnan (1996). If 20% of the components account for 80% of the cost, why apply the same rigour to recording transactional movements of inventory across 100% of components?

Sammon & Adam (2004) describe how businesses can succumb to the "ERP steamroller" of integration in the area of procurement. The integration of procurement into one single instance ERP system implies a rationalisation of local suppliers and purchasing patterns and the elimination of redundant suppliers. This can result in the organisation losing its ability to vary the source of supply. It can also have the effect of "steamrolling" local differences in the supply base, for example, locally sourced components not having exactly the same specification as counterparts in other countries. As with all elements of master data (suppliers, parts, customers etc.), integrated systems covering global operations are intolerant of local nuances in data structure.

One downside to the large scale integration of business processes as exemplified in ERP systems is the onus it puts on data capture: the more integrated the system, the more data is required at the point of entry in order that flags and triggers encountered during subsequent steps in the process are populated. Broadly speaking, ERP systems push the onus of data quality back to the point of entry, decentralising responsibility for data quality back to the rightful owners of that data.

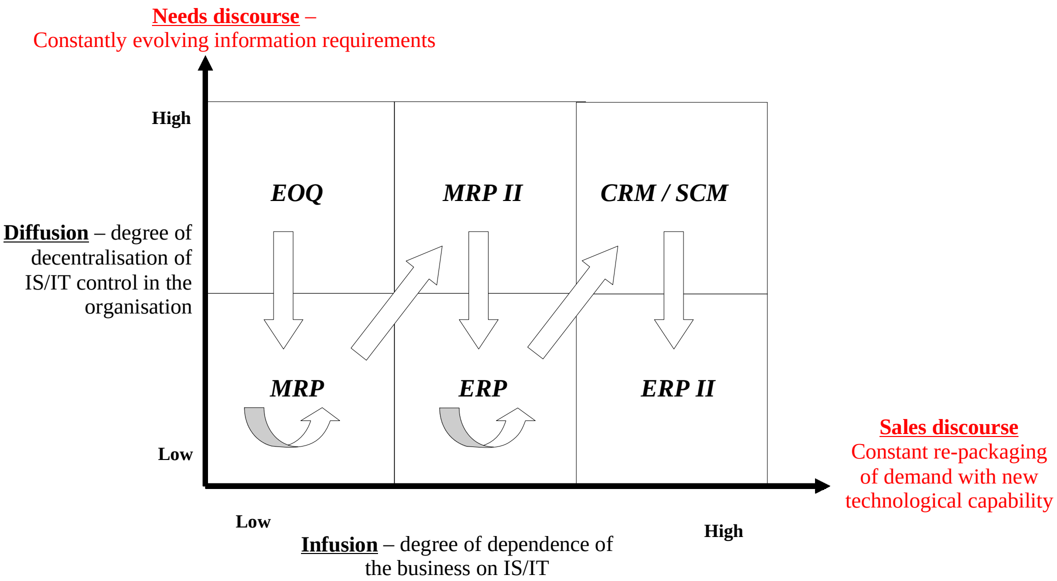
The Cyclical Nature of the ERP Market

The table in Figure 1, adapted from the IT planning matrix developed by Sullivan (1985), depicts the different stages in the evolution of planning requirements and corresponding management approaches. This evolution is plotted against the 2 main forces of change in IT: dependence of the business on IT (Infusion), and the degree of decentralisation of IT planning and control (Diffusion).

In simple manufacturing processes with few dependencies, material acquisition could be based on a principle of Economic Order Quantity (EOQ), whereby re-ordering of stock items was triggered automatically based on a minimum stock level. MRP originated in the early 60's as a computerised approach for the planning of materials acquisition and production for more complex manufacturing processes where interdependencies between components existed. Orlicky (1975) realised that a computer enabled the detailed application of the technique, making it effective in managing manufacturing inventories.

Based around the Bill of Materials (BOM), early applications exploded a production plan for a top level parent item into a plan of production and purchasing for compo-

Figure 1: The evolution of planning requirements and corresponding management approaches.



nent items. These systems were implemented on large mainframe computers run in centralised material departments for large companies.

A strong influence in the promotion of MRP was the campaign of communication launched by the 11,000 member American Production & Inventory Control Society in the early 70's. The computer industry soon followed with a range of applications. Almost all the major computer manufacturers had developed and were pushing software applications to support MRP, and virtually every major industrial consulting firm was advising on their implementation (Miller & Sprague, 1975).

Starting, therefore, under the "traditional" influence of IT, the use of these applications became more widespread, leading to the extension of the software to support various related operational functions. In particular, the combination of the planning and execution modules, with the potential for feedback from the execution cycle to the planning cycle, was termed closed-loop MRP.

So closed-loop MRP, together with some financial modules, developed into an integrated approach to the management of manufacturing resources. This approach became known as Manufacturing Resource Planning or MRP II. From the 1980's onwards, MRP II applications became available at lower cost on minicomputers and then microcomputers.

The attraction of the move to MRP II in the 80's lays not only in its role as decision making support, linking as it did, the planning and execution modules of manufacturing; but more importantly, in its integrative role within the manufacturing organisation (Browne, Harhen, Shivnan, 1996). Similarly, the attraction of moving to ERP in the 90's lay not only in its integrative role within the organisation, but also in its integrative role within the corporation, forcing geographically disparate entities to toe a common line with respect to operational procedures.

FUTURE TRENDS

The growth in popularity of ERP systems can be linked to an increasing business trend towards globalization, mergers and acquisitions. To be successful, a global company must be able to control and coordinate their various remote operating units. Accurate, real-time information provided by an ERP system has the ability to integrate the more remote subsidiaries into corporate practice because an ERP system allows the sharing of information in standard format across departments, currencies, languages and national borders. Thus, ERP systems can be used to provide a "common language" between units (Bingi *et al.*).

On the other hand, many post-ERP companies are realising that despite the huge implementation expense of the project, the business is no better off in terms of having access to summarised information which might assist managers in making decisions. CIO's might well demand, after the extravagant sales discourse and staggering budgets commanded by ERP vendors, what is the cost of having stockpiles of detailed transactional information?

As we look to the future, computation costs are still falling, while pressure on cost reduction has been growing since the worldwide economic recession in the new millennium. Vendors, desperate to make up for the dearth in large IT investments, attributed to Y2K fears, never saw the purse strings loosened as IT budgets were reduced in line with all corporate spending cuts. What is the solution? The natural propensity of computer manufacturers is to tout new systems that use lots of computer time (Miller & Sprague, 1975).

Cue the emergence of the Enterprise Resource Planning II market, billed by some vendors as the means to realising the benefit originally planned for ERP, and by others as the continuation of the "integration" of business processes. This continuing integration is focused on the area of Product Lifecycle Management, whereby engineers and product designers are "integrated" into the sales process at one end of the cycle, to allow better translation of customer requirements, and to manufacturing at the other end of the cycle, to facilitate the development of efficient production processes.

CONCLUSIONS

In a survey of 20 US companies, Lederer and Mendelow (1987) found that top managers did not view information as a business resource to be managed for long-term benefit. They only appreciated its criticality when they could not get what they needed.

Business resources (for example, assets, inventory, employees) are managed carefully because managers understand that making efficient use of these resources is key to the survival of the business. Investment in these areas is always evaluated carefully with respect to what contribution they will bring to the business.

It has been notoriously difficult in the past to cost-justify investments in information systems, and therefore vendors have opportunistically hitched their offerings to requirements that have an in-built business case (reducing inventory costs, for example).

A fundamental question worthy of further research is: to what extent do vendors instigate new demand for products simply by integrating application modules that share a common underlying data structure?

In their discussion of management fads and fashions in the ERP community, Sammon and Adam (2004) suggest that there is a mutually beneficial undercurrent to the constant re-packaging of the expert discourse: for every fad there will be a corresponding "panacea". In dealing with the complex problems facing organisations today, the language promoted by vendors in the ERP community tends to help formulate old problems in new ways, such that clients can elaborate the needs discourse in a way that is new, stimulating and irrefutable.

However, with each further step into the spiders web of integration, and letting what "can" be done take precedence over what "needs" to be done, the organisation is perhaps forgetting that: a) this apparent progress rarely gives tangible benefits, and b) flexibility is often more important than integration.

REFERENCES

Ackoff, R. (1967, December). Management misinformation systems. *Management Science*, *14*(4).

Berger, P. (1998) *PGI: les services valent cher*. *Le Monde Informatique*. 25th Sept 1998, #779

Bingi *et al.*, (1999, Summer). Critical issues affecting an ERP implementation. *Information Systems Management*, 7-14.

Browne, J., Harhen J., & Shivnan, J. (1996). *Production management systems*. Addison-Wesley.

Dearden, J. (1972, January/February). MIS is a mirage. *Harvard Business Review*, *50*(1), p. 90.

Horwitt, E. (1998, March). Enduring a global roll-out – and living to tell about it. *Computerworld.*

Inmon, W.H. (2000). ERP and data warehouse: Reading the tea leaves. Retrieved October 18, 2001, from *www.billinmon.com/library/articles/arterpfu.asp*

Kirwin, B. (1995). *Total cost of ownership: A powerful management tool*. Gartner Group Inc.

Lederer A.L. & Mendelow A.L. (1987). Information resource planning: overcoming difficulties in identifying top management's objectives. *MIS Quarterly*, *11*(3), pp 389-399

Miller, J.G., & Sprague L.G. (1975, Sept-Oct). Behind the growth in materials requirements planning. *Harvard Business Review*, 84-89.

Mokyr, J. (2001). The rise and fall of the factory system: Technology, firms and households since the Industrial Revolution. *Carnegie Rochester Conference Series on Public Policy*. *55*(1), 1-45.

O'Gorman, 2004. *The enterprise resource planning era: Lessons learned and issues for the future.* Hershey, PA: Idea Group Inc.

Orlicky, J. (1975). *Materials requirements planning: The new way of life in production and inventory management.* New York: McGraw Hill.

Sammon, D. & Adam, F. (Eds.). (2004). Towards a model for investigating non-decision making in ERP Communities. In *The Enterprise Resource Planning Decade: Lessons Learned and Issues for the Future*. Hershey, PA: Idea Group Inc.

Sammon, D., Adam, F. & Carton, F. (2003, December). The realities of benefit realisation in the monolithic enterprise systems era- Considerations for the future. *Electronic Journal of Information Systems Evaluation.*

Sullivan, C.H . (1985, Winter). Systems planning in the information age. *Sloan Management Review.*

Ward, J., Peppard J. (2002*). Strategic planning for information systems* (3rd ed.). John Wiley & Sons.

Zuboff, S. (1988). *In the age of the smart machine: The future of work and power*. Oxford: Heinemann Professional Publishing.

KEY TERMS

Bill of Materials (BOM): A Bill of Materials is a hierarchical product structure, showing the sub-components and interdependencies of any given finished good. Akin to a recipe, it is the underlying link between end product demand and material requirements. It also facilitates production costing, as each component in the hierarchy can be costed.

Closed loop MRP: The combination of MRP functionality with planning and production execution modules, with the potential for feedback from the execution cycle to the planning cycle, is termed "closed loop MRP".

Economic Order Quantity (EOQ): An approach to defining the lot size for purchasing raw materials, the EOQ is a mathematical expression of the trade-off between ordering costs (for purchased items) or set-up costs (for manufactured items) and the cost of storing material as inventory. If set-up or ordering costs are high, it may make sense to deal in larger batches, with the inherent knock-on effect of increasing inventory costs.

Enterprise Resource Planning (ERP): ERP systems are integrated applications that satisfy the transaction processing requirements for a wide range of business activities, including purchasing, production planning, warehouse management, inventory control, sales order processing, distribution, finance and human resources.

Enterprise Resource Planning II (ERP II): Enterprise Resource Planning II (ERP II) is a term that has been coined to denote the applications aimed at satisfying organisations who have already implemented ERP. This appears to include the realisation of efficiency gains originally planned for ERP, the implementation of ERP solutions to more vertical market segments and the further integration of key business processes (for example, to include PLM)

Manufacturing Resource Planning (MRP II): MRP II is closed loop MRP, enhanced with some financial modules for production costing and creditor management, usually running on a single integrated technical platform. MRP II was the immediate pre-cursor of Enterprise Resource Planning systems.

Material Requirements Planning (MRP): MRP, originating in the early 60's, was a computerised approach to the planning of materials acquisition and production for more complex manufacturing processes where interdependencies between components exist. The application of computers to MRP was a significant initial step in the evolution of ERP systems.

Product Lifecycle Management (PLM): Product Lifecycle Management is an approach whereby engineers and product designers are "integrated" into the sales process at one end of the cycle, to allow better translation of customer requirements, and to manufacturing at the other end of the cycle, to facilitate the development of efficient production processes. The "lifecycle" is derived from the notion of managing product design from inception through to execution and eventually retirement.

Relating Cognitive Problem–Solving Style to User Resistance

Michael J. Mullany
Northland Polytechnic, New Zealand

INTRODUCTION

This chapter explores cognitive problem-solving style and its impact on user resistance, based on the premise that the greater the *cognitive difference* (*cognitive gap*) between users and developers, the greater the user resistance is likely to be. Mullany (1989, 2003) conducted an empirical study demonstrating this. This study contradicts the findings of Huber (1983) and supports Carey (1991) in her conclusion that cognitive style theory, as applied to IS, should not be abandoned. Mullany's findings, in fact, are the opposite. Kirton (1999, 2004) supported Mullany's results. In particular, Mullany made use of Kirton's (2004) adaption–innovation theory. The emergent instrument, called the Kirton adaption–innovation inventory (KAI; Kirton, 1999, 2004), was used by Mullany as his measure of cognitive style.

Mullany's study also investigated the relationship between user resistance and user ages and lengths of service in the organisation. It failed to show any relationship between these factors and user resistance. This countermands the findings of Bruwer (1984) and dismisses any intimation that older or longer-serving employees are necessarily more resistant to change as myths.

BACKGROUND

Ever since the early 1980s, experts have identified user resistance to new systems as an expensive time overhead (see studies by Hirschheim & Newman, 1988, and Markus, 1983). Some authors suggest the greater importance of age and length of service. Bruwer (1984), for instance, claimed to have demonstrated that the older or longer-serving an employee, the more resistant he or she is likely to be to a new computer system. Clarification of issues surrounding user resistance has also highlighted *cognitive style theory* as potentially important, but to date, its impacts have only been sparsely researched in relation to user resistance, many of the prior studies being open to question. This research, on the other hand, proposes that a system will fail when the developer and user differ significantly in their problem-solving approaches. To reduce user resistance, it thus makes sense to recommend system designs that suit the user's approach to problem solving.

This issue appears only to have been studied empirically by Mullany (1989, 2003). He formulated the research question, "Is there a relationship between user resistance to a given information system and the difference in cognitive style between the user and the developer?" With the aid of his own instrument for measuring user resistance and the Kirton adaption–innovation instrument (Kirton, 1999) to measure the cognitive styles of users and associated system developers, he found a highly significant relationship between developer–user cognitive style differences and the level of user resistance to systems.

Why no other studies along similar lines have been reported in credible current research is difficult to explain. One possibility is that the literature contains speculative studies, such as that by Huber (1983), that discredit cognitive-style theory as a tool in understanding system success. Other studies, such as that by Carey (1991), while encouraging the continued use of cognitive-style theory in studying system phenomena, do not demonstrate its predictive success in information systems (IS). The remainder of this chapter thus examines the meaning and measure of cognitive style, the measure of user resistance, the specific findings of Mullany (1989, 2003), and outlooks for the future in this area of research.

THE MEANING AND MEASURE OF COGNITIVE PROBLEM-SOLVING STYLE

Liu and Ginther (1999) defined *cognitive style* as, "An individual's consistent and characteristic predispositions of perceiving, remembering, organizing, processing, thinking and problem-solving." Schroder, Driver, and Streufert (1967), in a discussion of human information processing, suggested that organisms "either inherit or develop characteristic modes of thinking, adapting or responding and go on to focus upon adaptation in terms of information processing." In short, an individual exhibits characteristic ways of processing information (and, hence, solving problems), known as his or her "cognitive style." Table 1 gives an historic summary of key experts over the years

Table 1. Cognitive-style constructs: Key studies.

Reference	Cognitive-Style Construct	Instrument
Kelly (1955)	Cognitive complexity or simplicity	**RepGrid** (Repertory grid)
Jung (1960)	Jungian typology	**MBTI** (Myers–Briggs type indicator)
Witkin et al. (1967)	Field dependence or independence	**EFT** (Embedded figures test)
Hudson (1966)	Converger or diverger	None
Schroder et al. (1967)	Cognitive complexity	**DDSE** (Driver's decision-style exercise)
Ornstein (1973)	Hemispherical lateralisation	Brain scan
Kirton (1976)	Adaptor–innovator continuum	**KAI** (Kirton adaption–innovation inventory)
Taggart (1988)	Whole-brain human information processing	**HIP** (Human information-processing instrument)

who have endeavoured to name and measure the construct of cognitive style. Of these, the MBTI (Myers–Briggs type indicator) is the most used in current, credible research literature, followed by the KAI (Kirton, 1976, 1984). As previously stated, the only evident effort made to relate cognitive style to user resistance was carried out by Mullany (1989) using the KAI. The reason for his preferred use of the KAI stemmed from its ability to provide a near-continuous, bipolar scale, convenient for finding correlations and associations. The MBTI, by contrast, yields only certain cognitive classifications, where no mutual order is evident. The correlation with other factors would then have been more difficult to show statistically.

Turning to the theory behind the KAI, Kirton (1999) identified two extremes of cognitive style; namely, the *adaptor* and the *innovator*. The adaptor tends to follow traditional methods of problem solving, while the innovator seeks new, often unexpected, and frequently less-accepted methods. The adaptor tends to "do well" within a given paradigm, where the innovator tends to "do differently," thus transcending accepted paradigms. The adapter is prepared to be wedded to systems, solving problems "in the right way," but is often seen as "stuck in a groove." The innovator has little regard for traditions, is often seen as creating dissonance, and elicits comments such as, "He wants to do it his own way, not the 'right' way." All humans, Kirton proposed, can be located on a continuum between the extremes of these two cognitive styles.

Both cognitive extremes can be highly creative, can resist change, and can act as agents for change. Adaptors support changes to the conservative, back to the "good old ways," and resist changes to novel methodologies. Innovators support changes toward unprecedented systems and technologies and resist changes to the traditional.

Kirton's instrument, the KAI, has been widely demonstrated to be a successful measure of his construct of cognitive problem-solving style. The instrument takes the form of a questionnaire, on which the respondent has to rate himself or herself against 33 character traits. KAI scores can range from 32 to 160, with a mean of 96 and a standard deviation of about 16. A person scoring above the mean of 96 is considered to be an innovator; conversely, a person scoring below 96 is rated as an adaptor. However, in the range of 80 to 112 (that is, within one standard deviation of the mean), a third cognitive style can be identified—the mid-scorer. Such persons tend to have human rather than technical problem-solving preferences and can relate better to the extreme scorers than either can to the other.

A DESCRIPTION AND MEASURE OF USER RESISTANCE

Mullany (1989) measured user resistance at personal interviews with the key user of each system selected for investigation. The user was asked to list the problems that he or she recalled had occurred during the system's development and implementation. They were asked, in effect, to make complaints, in confidence, against the system and its manner of implementation. Then they were requested to rate the severity of each complaint on a seven-point scale (with seven representing the most severe weighting). The sum of severities of all the complaints measured the respondent's *resistance score* or *R-score*. Obvious criticisms of the R-score method are as follows:

1. It may be highly influenced by the cognitive style of the interviewer.
2. At an interview, the user might forget certain crucial problems that had been experienced.

Mullany refuted (1) on the grounds that the same person (himself) did all the interviewing in his study. He assumed (2) to be of limited impact, because the object of the R-score method is to observe the user in the process of complaining. Consequently, the resistant user is capable of exaggerating or even inventing complaints, making the issue of those forgetten less relevant. However, he conceded the limitation that there are covert forms of resistance, such as absenteeism and withdrawal, that are not necessarily related to overt complaints.

To investigate a relationship between cognitive-style differences and user resistance, Mullany (1989) set out to collect data from a suitable sample of computer system developers and users. Bivariate data were to be collected, namely, the analyst–user KAI difference and the R-score for each user. The association between these was then to be measured. For his association measure, he used both the Kendall-*t* and the more traditional Spearman-*r*, which are equally reliable for significance testing (Liebetrau & Kendall, 1970). According to Kendall (1970; who invented the Kendall-*t* measure of association) and as confirmed by Liebetrau (1983), sample sizes of 10 to 20 are sufficient for such tests. The author thus selected a much larger sample size of 34 systems in 10 South African organizations. However, the following further criticisms were identified and addressed:

1. The sample size is small compared with some other studies in IS.
2. A user who champions a system may point out deficiencies in the hopes of improving that system.

Referring to the first of these criticisms, one should be alerted to the fact that sample representivity is more important than size in obtaining reliable results. In fact, the larger the sample, the less the researcher is likely to be able to guarantee a lack of significant bias. For example, suppose that with the aid of a suitable instrument, one sets out to measure the diligence of some human population. If a large sample size is sought through a postal, Web-based, or e-mail survey (the only practicable methods for really large samples), only the most diligent respondents are likely to respond, giving a bias to the more diligent and, thus, casting serious doubt on the results. To reduce this effect, Mullany (1989) collected all the data at personal interviews with the analysts and users. Furthermore, organisations he approached were requested to provide a fair spread of systems in use. He thus used legitimate power lent to him by the organisations to interview those as he required.

The second criticism was addressed by obtaining approval to keep all employees' responses confidential and to make this clear at each interview. This meant that a user would be unlikely to complain to Mullany in the hopes of achieving some system improvement, as he or she knew that no information would be relayed to the rest of the organisation. Every effort was made to preserve standard interviewing conditions: these being freedom from pressure or interruption and complete assurance of confidentiality. In short, interviewing conditions similar to those of face-to-face counselling were achieved. Furthermore this technique of measuring user resistance has been confirmed by respected researchers. First, both Markus (1983) and Hersheim et al. (1988) identified complaint as an overt symptom of, if not even a form of, resistance. Kirton (2004), in a discussion of Mullany's study, confirmed the technique as valid.

THE RELATIONSHIP BETWEEN USER RESISTANCE AND THE DIFFERENCES IN COGNITIVE STYLES BETWEEN THE USER AND THE DEVELOPER

The key developer and key user of each were interviewed. In each case, measures were obtained for the developer KAI score, user KAI score, and user R-score. At the same time, demographic data were collected; most particularly, the ages and lengths of service of the respondents, in order to test the findings of Bruwer (1984). A relationship as an association was found for the user R-scores versus the absolute differences between developer and user KAI scores. The association (with $p < 0.005$) proved to be strong, suggesting that user resistance can be minimized by matching a user with a developer of similar cognitive style. However, no significant associations were found between the ages and lengths of service of users and their R-scores, in contradiction of Bruwer's (1984) results. Rosen and Jerdee's (1976) study, which sought a similar result for occupational groups in general, agrees with Mullany's findings in this respect.

An interpretation of the R-score was demonstrated based on a near-perfect direct proportion that proved to exist between the weighted and nonweighted numbers of the users' complaints. In this relationship, the constant of proportionality was found to be 3.913 (that is, nearly 4). The R-score can thus be described as approximately four times the number of complaints a user will make retrospectively, in private, concerning a system and its manner of implementation.

FUTURE TRENDS

This study reignites the issue of cognitive style as an important issue in IS and completely countermands the conclusions of Huber (1983). It substantially strengthens the case made by Carey (1991) that cognitive-style issues in IS research should not be abandoned. Further, it suggests that user resistance and the related constructs of user dissatisfaction and system success can be predicted from cognitive-style measures (that is, KAI scores) prior to system development.

Areas for further research centre upon the main limitation of this study. For instance, there is little known regarding how the developer-user cognitive gap influences the system development life cycle (SDLC) over a significant passage of time, and neither this study nor any other found in the literature has achieved this. In fact, the literature is devoid of any attempts to conduct such research. A longitudinal study where SDLC curves are compared with the developer-user cognitive gap would be of immense importance and interest. New rules for system development based on cognitive-style testing would be expected to emerge.

CONCLUSION

It is clear that cognitive problem-solving style, as defined by Kirton and measured using the KAI, impacts user resistance. The greater the cognitive gap between users and developers, the greater the user resistance is likely to be. This contradicts the findings of Huber (1983) and supports Carey (1991) in her conclusion that cognitive-style theory, as applied to IS, should not be abandoned. Mullany's findings, in fact, are the opposite.

The failure to show any relationship between users' ages and lengths of service, and their resistance ratings, countermand the findings of Bruwer (1984) and suggest that organisations should be alerted to the danger of discriminating against older or longer-serving users or dispensing with their services on such grounds.

As mentioned above, areas for further research centre upon the main limitation of this study. A longitudinal study where SDLC curves are compared with the developer–user cognitive gap would be of great importance and interest.

REFERENCES

Bruwer, P. J. S. (1984). A descriptive model of success for computer-based information systems. *Information and Management, 7*, 63-67.

Carey, J. M. (1991). The issue of cognitive style in MIS/DSS research. In J. Carey (Ed.), *Human factors in information systems. An organizational perspective* (pp. 337-348). Norwood, NJ: Ablex Publishing Corp.

Hirschheim, R., & Newman, M. (1988). Information systems and user resistance, theory and practice. *The Computer Journal, 31*(5), 398-408.

Huber, G. P. (1983). Cognitive style as a basis for MIS designs: Much ado about nothing? *Management Science, 29*(5), 567-579.

Hudson, L. (1966). *Contrary imaginations*. United Kingdom: Methuen.

Jung, C. G. (1960). *The basic writings of CG Jung*. New York: Pantheon.

Kelly, G. A. (1955). *The psychology of personal constructs*. New York: Norton.

Kendall, M. G. (1970). *Rank correlation methods*. London: Charles Griffin & Co.

Kirton, M. (1976). Adaptors and innovators: A description and measure. *Journal of Applied Psychology, 61*(5), 622-629.

Kirton, M. (1984). Adaptors and innovators—Why new initiatives get blocked. *Long Range Planning, 17*(2), 137-143.

Kirton, M. (1999). *KAI manual* (3rd ed.). Berkhamstead: UK: Occupational Research Centre.

Kirton, M. (2004). *KAI Certification Course*. Birkhamstead, Hertfordshire: Occcupational Research Centre.

Liebetrau, A. M. (1983). *Measures of association*. Beverly Hills, CA: Sage Publications.

Liu, Y., & Ginther, D. (1999). Cognitive styles and distance education (invited submission). *Online Journal of Distance Learning Administration, 2*(3), 118.

Markus, M. L. (1983). Power, politics, and MIS implementation. *Communications of the ACM, 26*(6), 430-444.

Mullany, M. J. (1989). An analysis of the relationship between analyst–user cognitive style differences and user resistance to information systems. Master's thesis. Cape Town, South Africa: University of Cape Town.

Mullany, M. J. (2003). Forecasting user satisfaction. *The International Principal, 7*(1), 10-12.

Ornstein, R. E. (1973). *The nature of human consciousness*. San Francisco, CA: Viking Press.

Rosen, B., & Jerdee, T. H. (1976). The nature of job-related age stereo-types. *Journal of Applied Psychology, 61*(2), 180-183.

Schroder, H. M., Driver, M. J., & Streufert, S. (1967). *Human information processing. Individuals and groups functioning in complex social situations*. New York: Holt, Rinehart and Winston.

Taggart, W. M. (1988). A human information processing model of the managerial mind: Some MIS implications. In J. Carey (Ed.), *Human factors in information systems. An organizational perspective* (pp. 253-268). Norwood, NJ: Ablex Publishing Corp.

Witkin, H. A., Moore, C. A., Goodenough, D. R., & Cox, P. W. (1977). Field dependent and field independent cognitive styles and their educational implications. *Review of Educational Research, 47,* 1-64.

KEY TERMS

Adaptor: An adaptor tends to follow traditional methods of problem solving, tending to "do well." He or she is often seen as "stuck in a groove" (Kirton, 1999).

Association: A relationship between two statistical variables. Unlike a *correlation*, an association does not yield a quantitative result but is contingent upon the ranking of the bivariate data values only.

Cognitive Gap: The difference in cognitive problem-solving style between two people, especially two people who are obliged to interact as members of a group or team.

Cognitive Problem-Solving Style: The position an individual occupies between two extremes of cognitive problem-solving style personality; namely, *the adaptor* and *the innovator*.

Cognitive Style: An individual exhibits characteristic ways of processing information and, hence, solving problems, known as his or her "cognitive style."

Innovator: The innovator seeks new, often unexpected, and frequently less acceptable methods. He or she has little regard for traditions, is often seen as creating dissonance, and elicits comments such as, "He wants to do it his own way, not the 'right' way" (Kirton, 1999).

KAI (Kirton Adaption-Innovation Inventory): An instrument that measures cognitive problem-solving style. It takes the form of a questionnaire, on which the respondent is asked to rate himself or herself against 32 character traits.

R-Score (Resistance Score): A method of measuring user resistance where, at personal interviews with the key user of a given system, the user is asked to list system problems and then to rate the severity of each on a seven-point scale. The sum of severities of all the complaints measures his or her *R-score* (Mullany, 1989).

User Resistance: Any user behaviour, action, or lack of action that inhibits the development, installation, or use of an information system.

Relationship Cardinality Constraints in Relational Database Design

Dolores Cuadra
Universidad Carlos III de Madrid, Spain

Paloma Martinez
Universidad Carlos III de Madrid, Spain

Elena Castro
Universidad Carlos III de Madrid, Spain

INTRODUCTION

Database modelling is a complex task that involves conceiving, understanding, structuring and describing real universes of discourse (UD) through the definition of schemata using abstraction processes and data models. Traditionally, three phases are identified in database design: conceptual, logical and physical design. The conceptual modelling phase represents the most abstract level since it is independent of any database management system (DBMS) and, consequently, it is very close to the user and allows him/her to collect almost completely the semantics of the real world to be modelled.

A conceptual schema, independent of the data formalism used, plays two main roles in the conceptual design phase: a *semantic* role, in which user requirements are gathered together and entities and relationships in a UD are documented, and a *representational* role that provides a framework that allows a mapping to the logical design of database development. Three topics are involved in the database conceptual modelling process: data modelling formalism, methodological approach, and CASE tool support. One of the most extended data modelling formalisms, extended entity relationship (EER) model, has proved to be a precise and comprehensive tool for representing data requirements in information systems development, mainly due to an adequate degree of abstraction of constructs that it includes. Although the original ER model was proposed by Chen (1976), many extensions and variations as well as different diagrammatic styles have been defined (Hull & King, 1987; McAllister, 1998; Peckhan & Maryanski, 1988).

In database conceptual analysis, one of the most difficult concepts to be modelled is the relationship concept, especially higher order relationships, as well as its associated cardinalities. Some textbooks (Boman et al., 1997; Ullman & Widom, 2001) assume that any conceptual design can be addressed by considering only binary relationships since its aim is to create a computer-oriented model. We understand the advantages of this approach although we believe that it may produce certain loss of semantics (some biases are introduced in user requirements), and it forces to represent information in rather artificial and sometimes unnatural ways.

Concerning the logical design, the transformation process of conceptual schemata into relational schemata should be performed trying to preserve the semantics included in the conceptual schema completely; the final objective is to keep such semantics in the database itself and not in those applications that access to the database. Nevertheless, sometimes a certain loss of semantics is produced, for instance, *foreign key* and *not null* options in the relational model are not sufficient to control relationship cardinality constraints.

CONCEPTUAL MODEL REVISED

Central concepts of the ER conceptual model are entities and relationships; these constructs were introduced by Chen (1976) and have been incorporated in other conceptual models although with different names: *class, type*, and so forth, for entities and *associations* for relationships. Nevertheless, those concepts do not have a precise semantics, and, consequently, it is necessary to fix their meaning.

Although the entity concept is widely used and accepted, there is no agreement on one definition; for instance, Thalheim (2000) collects 12 different entity denotations. Although experts are not able to give a unique definition, the underlying concept is coincidental in all of them and its usage as design element does not suppose great disadvantages. An entity definition is not given here, just to highlight, according to Thalheim (2000), an entity is a *representation* abstraction with modeling purposes. Date (2004) adds that the represented concept is a distinguishable object, but we do not consider this feature as essential because it depends on the designer's point of view.

The relationship concept is more confused; it is defined as an *association* among entities. This definition offers many interpretations; for instance, in several design methods there are some differences depending on the number of relations can participate in other relationships as in HERM (Thalheim, 2000) by means of association entities as in UML, OMG (2000), or by grouping as clusters a set of entities and relationships (Teorey, 1999). These differences occur because a relationship combines *association* features with *representation* features, and therefore, it might be considered a relationship (if association aspects are highlighted) or an entity (if representation aspects are emphasized).

Cardinality constraint is one of the most important restrictions that can be established in a relationship, and in general, in a conceptual schema. Its functionality is to limit the number of entity occurrences associated in a relationship. Even though it is a simple concept, the definition of this constraint admits several variants.

Two main approaches are discussed: the Chen's constraint that is an extension of the mapping constraint (a special case of cardinality constraint that considers only the maximum cardinality and that for binary relationships can be 1:1, 1:N or N:M) (Chen, 1976); the Chen's constraint has been adopted or extended in different data models and methodologies. On the other side, the MERISE approach (Tardieu, Rochfeld, & Coletti, 1983) incorporates the participation semantics. These two approaches meet each other when cardinality constraints for binary relationships are defined (excepting the natural differences in graphical notations). Both of them represent the same semantics in binary relationships although the way of expressing it is different. Table 1 summarises the semantics associated to cardinality constraint for the occurrences of *A* in the binary relationships *R*.

In n-ary relationships, the two approaches Chen and Merise, previously commented, do not represent the same semantics. Table 2 summarises the semantics associated to cardinality constraint for *A* in the n-ary relationship *R* depending of the approach.

TRANSFORMATION OF CONCEPTUAL SCHEMATA INTO RELATIONAL SCHEMATA

The main difficulty when transforming a conceptual schema into a relational schema is information preservation. Generally, to achieve a complete mapping between both models and keeping their inherent and semantic restrictions from conceptual model to relational model is quite complicated. Usually, restrictions that cannot be applied in the relational model must be reflected in the application in some different way, that is, outside the DBMS. In this way, there are several extensions to the relational model proposed by Codd (1970), Codd, (1979), Date, (2004) and Teorey (1999) that provide a more semantic model.

The principal transformation rules are described in most database textbooks (Date, 2004; Elmasri & Navathe, 2003; Ramakrishnan & Gehrke, 2002), but these rules do not reflect the cardinality constraints transformation. In general, these rules can be classified for binary relationships transformation depending on the relations number that generated (Table 3). Option 1 makes a relation for each constructor, therefore the R relationship is transformed in a relation. Option 2 uses *key propagation* that is applied to a relationship which mapping constraint is 1:N or 1:1. Last, Option 3 only considers just one relation, to store the information of A, B and R constructors. This option

Table 1. Cardinality constraint in binary relationship summary

	A — R — B	
Cardinality constraint for **A**	Minimum	Maximum
0	**Optional**: there are occurrences of entity A that not participates in the relationship R	Inapplicable
1	**Mandatory**: all occurrences of entity A participates in the relationship R	**Uniqueness**: there is at most one occurrence of the entity B related to an occurrence of the entity A
k>1	**K-Mandatory**: each occurrence of the A participates at least K times in the relationship	**K-Limit**: There are at most K occurrences of the B related to each occurrence of the A
N	Inapplicable	Without limit of maximum participation

Table 2. Cardinality constraint in n-ary relationship summary

	Chen Approach		Merise Approach	
Cardinality constraint for **A**	Minimum	Maximum	Minimum	Maximum
0	Presence of unknown information	Inapplicable	Optional participation of the A ocurrences	Inapplicable
1	No constraint	For each (n-1) record there is one unique ocurrence related with one ocurrence of the entity A in the R	Mandatory participation of the A	For each A ocurrence there are at most one ocurrences related in the R
k>1	For each (n-1) record there are at least more than one ocurrences of the entity A related in the R	For each (n-1) record there is at most K ocurrences related with one ocurrence of the entity A in the R	For each A ocurrence there are at least more than K ocurrences related in the R	For each A ocurrence there are at most K ocurrences related in the R
N	Inapplicable	Without limit of maximum participation	Inapplicable	Without limit of maximum participation

is more restrictive than the others and only can be applied in a relationship which mapping constraint is 1:1.

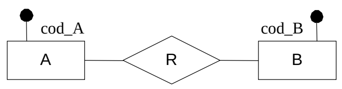
Table 4 explains when the different options explained in the Table 3 can be used depending on cardinality constraints (max. and min.) associated with the entities A and B and summarizes main transformation rules, indicated by options and relational constraints that must be included to add semantics. Relational constraints considered are: primary key (PK), unique key (UK), foreign key and the reference to relation X (FK_X), and optional attributes (N).

As we can see in Table 4, in most cases the minimum cardinality, the three first rows in the table, is not considered so its associated semantics is lost at the transformation.

When we transform n-ary relationships, we must take into account:

- the cardinality constraint approach used in the conceptual model;
- whether a relationship should be decomposed in minor grade relationships; and
- if the study cardinality constraint validation is needed.

Table 3. Binary relationship transformation options



Option 1	**Option 2**	**Option 3**
A (cod_A,...)	A (cod_A,...,cod_B)	R (cod_A, cod_B,...)
B (cod_B,...)	B (cod_B,...)	
R (cod_A, cod_B,..)		

The relationship decomposition and cardinality constraint validation study is a complex task that needs the interaction with domain experts to be solved. Therefore, transformations rules, on the whole, presented easy rules.

An example of different transformation rules with the Merise and Chen approaches for a ternary relationship is presented in Table 5. As it is shown in the table, rules only depend on mapping constraint.

In addition, the principal transformation rules for a ternary relationship, adopted for most design methodologies, appear independent of minimum cardinality constraints. This situation appears also in higher relationships.

Therefore, we can summarize that the cardinality constraint is one of the most important restrictions that can be established in a relationship, and in general, the semantics associated is loss in the transformation process to relational model.

CRITICAL ISSUES OF CARDINALITY CONSTRAINTS

After reviewing all conceptual constructs, it is needed to achieve a conceptual model as a tool that is able to reflect the situations that frequently appear in data modelling scenarios where it is required. The redefinition of the conceptual model, taking into account previous aspects as well as the development of a wider notation, are tasks to be dealt with.

The detection and specification of abstractions in an UD that lead to a correct and complete schema are critical problems that combine psychological and methodological aspects. There are many other aspects associated with the specification of constraints. Their identification and validation require more formal treatments. Identification can be faced up to with a lexical analysis of a problem

Table 4. Transformation rules summarised

A and B cardinalities	Option	Constraint
(0/1,n) (0/1,n)	1	UK on R: (cod_A, cod_B) FK_A on R: cod_A FK_B on R: cod_B
(0,1) (0/1,n)	1	PK on R: cod_A NN on R: cod_B FK_A on R: cod_A FK_B on R: cod_B
	2	FK_B on A: cod_B N on A: cod_B
(1,1) (0/1,n)	2	FK_B on A: cod_B NN on A: cod_B
	1	PK on R: cod_A NN on R: cod_B FK_A on R: cod_A FK_B on R: cod_B
(0,1) (0,1)	1	CC on R: cod_A o cod_B NN on R: cod_A, cod_B FK_A on R: cod_A FK_B on R: cod_B
	2	FK_B on A: cod_B NN on A: cod_B o FK_A on B: cod_A NN on B: cod_A

A and B cardinalities	Option	Constraint
(0,1) (1,1)	2	FK_A en B: cod_A NN on B: cod_A
	1	UK on R: cod_A, cod_B NN on R: cod_A, cod_B FK_A on R: cod_A FK_B on R: cod_B
(1,1) (1,1)	2	FK_B on A: cod_B NN on A: cod_B **or** FK_A on B: cod_A NN on A: cod_B
	1	CC on R: cod_A, cod_B NN on R: cod_A, cod_B FK_A on R: cod_A FK_B on R: cod_B
	3	PK on R: cod_A NN on R: cod_B

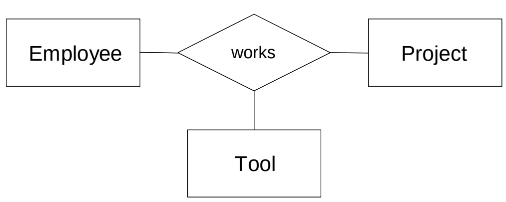
UK : Unique Key
PK: Primary Key
FK_A: Foreign Key reference to A
N: Column Null
NN: Column Not Null

description. Syntactic and semantic validation of relationships is a critical aspect of internal coherence with the UD. To obtain these properties, it is very important to:

- clarify the relationship definition;
- define formally relationship cardinality constraints and consider the two main approaches unification to higher order relationships;
- classify several approaches and provide conceptual model mappings; and
- provide a set of rules to validate relationships.

On the other hand, another problem is how to preserve the cardinality constraint semantics in n-ary relationships for their implementation in a DBMS. There are some approaches, for example, in Balaban and Shoval (2002), an extension of the extended entity relationship (EER) model is made by including methods to control the semantics associated with Chen's cardinalities. This approach adds a dynamic component to the model, but any database management system supports it. Camps (2002) makes a complete analysis of the transformation of maximum cardinality constraints into the relational model, considering

Table 5. Transformation rules to apply an example



Employee (cod_e,...)
Project (cod_p,...)
Tool (cod_t,...)
works (cod_e, cod_p, cod_t,...)

Chen Approach

Rule 1: **N:M:P**
Candidate keys = {(cod_emp, cod_proy, cod_tec)}
Rule 2: **1:N:M**
Candidate keys = {(cod_proy, cod_tec)}
Rule 3: **1:1:N**
Candidate keys = {(cod_proy, cod_tec), (cod_emp, cod_tec)}
Rule 4: **1:1:1**
Candidate keys = {(cod_proy, cod_tec), (cod_emp, cod_proy), (cod_emp, cod_tec)}

Merise Approach

Rule 1: **N:M:P**
Candidate keys = {(cod_emp, cod_proy, cod_tec)}
Rule 2: **1:N:M**
Candidate keys = {(cod_emp)}
Rule 3: **1:1:N**
Candidate keys = {(cod_emp), (cod_proy)}
Rule 4: **1:1:1**
Candidate keys = {(cod_emp), (cod_proy), (cod_tec)}

Chen and Merise cardinalities for ternary relationships. They use equivalencies between cardinality constraints and functional dependencies, but this work does not concern to the transformation of minimum cardinalities problem. Finally, Cuadra et al. (2002) present the transformation of EER schemata into relational schemata using an extended relational model with active capabilities giving solutions to this problem.

CONCLUSIONS

Conceptual models are well-known tools to achieve a good design of information systems. Nevertheless, the understanding and use of all constructs and constraints which are presented in these models are not an easy task, and sometimes it causes loss of motivation. Disagreements between main conceptual models, such as the confusion in the use of some of their constructs and some open problems in these models, are shown.

Another important topic treated in this chapter is the conceptual schema transformation into logical schema. Some solutions are presented in order to clarify the relationship construct and to extend the cardinality constraint concept, as well as several approaches to preserve the cardinality constraint semantics in n-ary relationships.

REFERENCES

Balaban, M., & Shoval, P. (2002). MEER- An EER model enhanced with structure methods. *Information Systems, 27*, 245-275.

Boman, M. et al. (1997). *Conceptual modelling*. Prentice Hall, Series in Computer Science.

Camps, R. (2002, June). From ternary relationship to relational tables: a case against common beliefs. *ACM SIGMOD Record, 31*(2), 46-49.

Chen, P.P. (1976). The entity-relationship model: Toward a unified view of data. *ACM Transactions on Database Systems, 1*(1), 9-36.

Codd, E.F. (1970, June). A relational model of data for large shared data banks. *CACM 13*(6).

Codd, E.F. (1979, December). Extending the database relational model to capture more meaning. *ACM Transactions on Database Systems, 4*(4), 397-434.

Cuadra et al. (2002). Preserving relationship cardinality constraints in relational schemata. In *Database Integrity: Challenges and Solutions*. Hershey, PA: Idea Group Publishing.

Date, C.J. (2004). *An introduction to database systems* (8th ed.) Reading, Mass.: Addison-Wesley.

Elmasri, R., & Navathe, S.B. (2003). *Fundamentals of database systems* (4th ed.). Addison-Wesley.

Hull, R., & King, R. (1987). Semantic database modelling: Survey, application, and research issues. *ACM Computing Surveys, 19*(3), 201-260.

McAllister, A. (1998). Complete rules for n-ary relationship cardinality constraints. *Data & Knowledge Engineering, 27*, 255-288.

OMG (2000). Unified modelling language specification, version 1-3. Object Management Group.*ACM Computing Surveys, 31*(1), 63-103.

Peckham, J., & Maryanski, F. (1988). Semantic data models. *ACM Computing Surveys, 20*(3), 153-189.

Ramakrishnan, R., & Gehrke, J. (2002). *Database management systems* (3rd ed.). MacGraw-Hill International Editions.

Tardieu, H., Rochfeld, A., & Coletti, R. (1983). *La Méthode MERISE. Tome 1: Principles et Outils*. Paris : Les Editions d'Organisation.

Teorey, T.J. (1999). Database modeling and design: The entity-relationship approach (3rd ed.). San Mateo: Morgan Kaufmann.

Thalheim, B. (2000). *Entity-relationship modelling. Foundations of database technology*. Springer-Verlag.

Ullman, J.D., & Widom, J. (2001). *A first course in database systems*. Prentice-Hall International.

KEY TERMS

Cardinality Constraints: One constraint established in a relationship. It limits the number of entity occurrences that are associated in a relationship.

Chen Approach: One way to calculate cardinality constraints. It limits the participation of a combination of the other entity(ies) with an entity in the relationship.

Conceptual Models: Tools to achieve a good design of information systems. These tools are used to express information system requirements specification, and their principal characteristic is easy and intuitive use.

Entity: The basic element in a conceptual model that represents an abstraction with modelling purposes.

Logical Models: In a design methodology, tools to transform conceptual schemata into a schemata near to implementation. A well known principal model is the relational model.

Merise Approach: Another way to calculate cardinality constraints. It limits the participation of an entity in the relationship.

Relationship or Association: The basic element in one conceptual model. It is defined as an association among entities.

Semantics: Property that collect UD specifications and that should be preserved in the transformation schemata for all methodology phases.

ENDOTES

[1] From now on, we will use the original names (entity and relationship).

Reuse of Formal Specifications

Laura Felice
Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina

Daniel Riesco
Universidad Nacional de San Luis, Argentina

INTRODUCTION

During the Rigorous Approach to Industrial Software Engineering (RAISE) specification development process, a variety of components and infrastructures are built. All of these components are not independent, but related to one another, especially when we specify different systems into the same infrastructure. The RAISE method (Bjorner, 2000) is based on the idea that software development is a stepwise, evolutionary process of applying semantics-preserving transitions. Thus, the reuse process is crucial in all of the stages of the development, but there is no explicit reference to the specification reusability in this development process.

Software components are typically very rich in information, making the task of characterizing them and capturing their relevant properties difficult. However, this is not the only reason that makes software reuse difficult. Krueger (1992) provides a brief general survey and very clear view of different approaches for software reuse.

Information retrieval methods based on analyses of natural-language documentation have been proposed to construct software libraries (Helm & Maarek, 1991; Maarek, Berry & Kaiser, 1991). Software components represented by natural language can make the retrieval process a task with ambiguity, incompleteness and inconsistency. Using a rigorous method to the retrieval of a component can minimize all of these problems.

Based on these observations, we introduce a Reusable Component (RC) model for the definition of the reusable component structure into RAISE. Related to this method, it is important to emphasize the work of Beltaifa and Moore (2001). They propose an infrastructure to support reuse improving the efficiency of reusing software components.

RC model integrates RAISE Specification Language (RSL) specifications (George et al., 1992) and object-oriented code. RC model describes object-oriented classes at different levels of abstraction:

- Specialization: hierarchies of RSL implicit specifications related by formal specialization relationship.
- Realization: hierarchies of RSL complete algebraic specifications related by realization relationship.
- Code: hierarchies of imperative RSL schemes related by implementation relationship and linked to object-oriented code.

Also, a rigorous process for reusability of RC components is defined. Its manipulation, by means of specification building operators (Rename, Extend, Combine and Hide), is the basis for the reusability.

Our approach allows that the properties of components formally specified can be characterized by giving a functional (RSL specification) description. Therefore, they may be useful to someone searching for a particular component.

Different possible classes of existing RC components may be retrieved using a formal reasoning technique: an exact match to the query specification, a component more general than the query, or a component more specific than the query.

BACKGROUND

Different approaches to specify reusable components functionalities have been proposed. The way in which the components can be used with others can play a critical role in the reuse implementation.

Related with the RAISE method, we emphasize the work of Beltaifa. They propose an infrastructure to support reuse which improve both the ease and efficiency of reusing software components. The main difference with our work is the integrated process defined for all stages of the development method.

As a typical related work, we can mention Hennicker and Wirsing (1992) who present a model for reusable component definition. A reusable component is defined as an unordered tree of specifications where any two consecutive nodes are related by the implementation relation and the leaves are different implementations of the root. The work of Chen and Cheng (1997) is another approach that provides a formalism to register components properties to reuse them based on the architecture and integration of the system. They are related to LOTOS tools to facilitate the retrieval of the reusable component.

On the other hand, the work of Zaremski and Wing (1997) is related to the specification matching. It is very important to emphasize this proposal has been referenced by a lot of authors.

There are two main activities in the RAISE method: writing an initial specification, and developing it towards something that can be implemented in a programming language (George, 2002). Writing the initial specification is the most critical task in software development. If it is wrong, that is, if it fails to meet the requirements, the following work will be largely wasted. It is well known that mistakes made in the life-cycle are considerably more expensive to fix than those made later.

What kinds of errors are made at the beginning? The main problem is that we may not understand the requirements. The requirements are written in a natural language, and, as a result, likely to be ambiguous. The aim of the initial specification is to capture the requirements in a formal and precise way.

MAIN THRUST OF RAISE AND REUSE

The aim of the project RAISE was to develop a language, techniques and tools that would enable industrial usage of "formal methods" in the construction of software systems. The results of this project include the RSL language, which allows us to write formal specifications; a method to carry out developments based on such specifications, and a set of tools to assist in edition, checking, transforming and reasoning about specifications.

RSL is a "wide spectrum" language that can be applied at different levels of abstraction as well as stages of development. It includes several definition styles such as model-based or property-based, applicative or imperative, sequential or concurrent.

A development in RAISE begins with an abstract specification and gradually evolves to concrete implementations. The first specification is usually an abstract applicative one, for example, functional or algebraic. A first algebraic specification should have:

- A hierarchy of modules whose root is the system module;
- A module containing types and attributes for the non-dynamic identified entities; and
- The signatures of the necessary functions associated with types. These functions should be categorized as generators (if the associated type or a type dependent on it appears in their result types) and as observers. Besides, preconditions should be formulated for partial functions. These preconditions are expressed by means of functions called "guards".

The specification may contain invariants expressed as functions.

RC Model Description

RC describes object classes at three different conceptual levels: specialization, realization and code. These names refer to the relations used to integrate specifications in the three levels. A more detailed description can be found in Felice, Leonardi, Favre, and Mauco (2001).

RC Components

The specialization level describes a hierarchy of incomplete RSL specifications as an acyclic graph. The nodes are related by the specialization relationship. In this context, it must be verified that if P(x) is a property provable about objects x of type T, then P(y) must be verified for every object y of type S, where S is a specialization of T.

Specialization level reconciles the need for precision and completeness in abstract specifications with the desire to avoid over-specification.

Every leaf in the specialization level is associated with a sub-component at the realization level. A realization sub-component is a tree of complete specifications in RSL:

- The root is the most abstract definition.
- The internal nodes correspond to different realizations of the root.
- Leaves correspond to sub-components at the implementation level.

If E1 and E2 are specifications E1 can be realized by E2 if E1 and E2 have the same signature and every model of E2 is a model of E1 (Hennicker & Wirsing, 1992).

Adaptation of reusable components, which consumes a large portion of software cost, is penalized by overdependency of components on the physical structure of data.

The realization level allows us to distinguish decisions linked with the choice of data structure. In RAISE, there are four main specification style options. They are applicative sequential, imperative sequential, applicative concurrent and imperative concurrent (George, Haxthausen, Hughes, Milne, Prehn, & Pedersen, 1995). Associated with them, there are two styles: abstract and concrete. Imperative and concrete styles use variables, assignments, loops, channels (in concurrent specifications), and so forth, that are related to design decisions about data structures. Every specification at the realization level is linked to sub-components at the code level.

The code level groups a set of schemes in RSL associated with code. The RAISE method provides translation processes, which start with a final RSL specification and produce a program in some executable language, for example, C++ using the translation tool component of the RAISE toolset (George, 2001).

RC Relationships

It is worth considering that the three relations (Specialization, Realization and Code) form the "RAISE implementation relation" (George et al., 1992). Any formal system that aims to provide a means of specification and development must provide a notion of implementation. That is, if specification E1 is related to specification E2 in the model, we need to know if E1 and E2 are in the "RAISE implementation relation". The following properties must be satisfied:

"Properties preservation: All properties that can be proved about E1 can also be proved for E2 (but not vice versa in general).
Substitutivity: An instance of E1 in a specification can be replaced by an instance E2, and the resulting new specification should implement the earlier specification".

RAISE DEVELOPMENT PLAN APPLYING A REUSE MODEL

Engineers usually proceed from applicative to imperative specifications. The proposal is to introduce the RC model for the definition of the reusable component structure into RAISE method. Where does one introduce this model? RAISE developers start with the module scheme specification, define an abstract applicative module, and develop a sequence of concrete applicative modules and the corresponding set of imperative modules from the final applicative modules. Figure 1 summarizes the overview of the method.

During the first stages of development, when the interaction with the stakeholders is crucial, the use of client-oriented requirements engineering techniques seem to be necessary in order to enhance the communication between the stakeholders and the software engineers. It has been proposed that a systematic reuse approach integrates natural language requirement specifications with formal specifications in RSL. Some heuristics are described to develop a formal specification in RSL starting from models which belong to the requirements baseline (Mauco & George, 2000).

The goal is that engineers can make reuse in all development stages. Therefore, by introducing the RC model in all the development steps, it will be possible to make use of abstraction, selection, specialization, and integration of software artifacts in the RAISE method.

THE REUSE PROCESS

Formal specifications are used to model the problem requirements and the function of the library components. The specifications are written in RSL language. The classification scheme consists of a collection of formal definitions representing possible component features in the domain. The formalization of the scheme permits automated classification of the specifications. The retrieval mechanism is based on syntactic comparison of features sets. The components returned by the retrieval mechanism are passed on to a more detailed evaluation that uses specification matching to determine reusability.

The work of Penix (1998) is an important approximation, because he proposes automated component retrieval and adaptation using a heuristic based on specification semantics to indicate component reusability.

Figure 1. Overview of the RAISE method

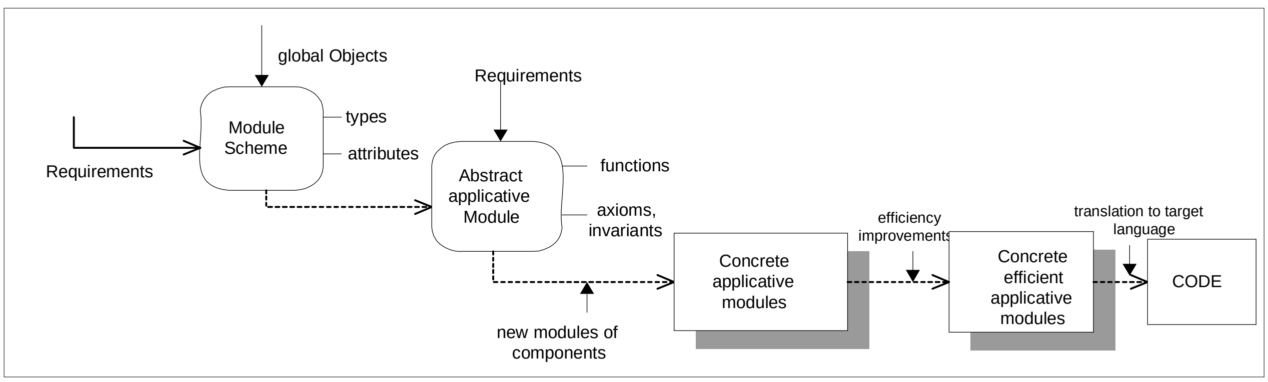
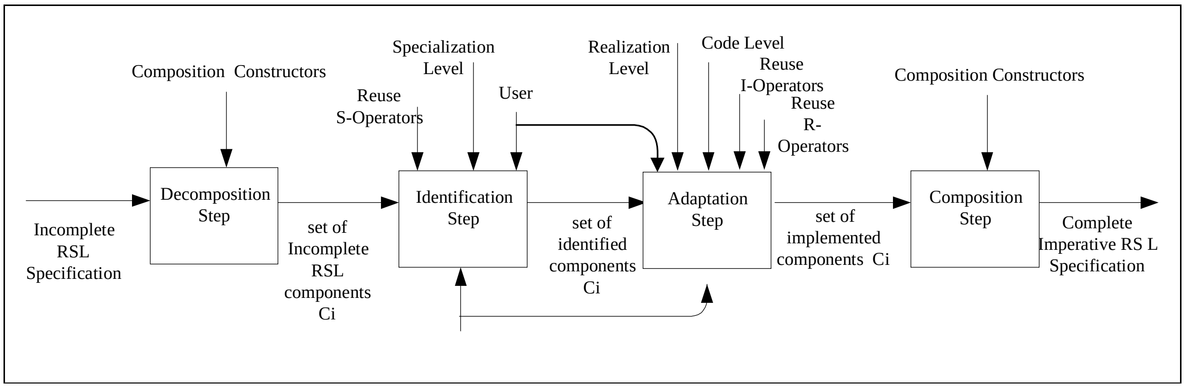


Figure 2. The method



The results of specification matching determine the relationship that exists between the retrieved components and the requirements specification. The adaptation phase allows in determining whether a mechanism exists to adapt or combine the retrieved components to solve the problem.

The main idea is to transform the incomplete RSL specification into complete imperative specification by reusing existing components. The method has the following steps: decomposition, identification, adaptation and composition depicted in Figure 2.

- Decomposition step: the decomposition of a goal specification Eg into sub-specifications E1, E2, ... En is formalized.
- Identification step: for each specification Ei a component Ci (in the specialization level) and a sequence s1,s2,...,sn of RSL specifications must be identified, verifying the implementation relation. A node in Ci must be selected as a candidate to be transformed. The identification of a component is correct if it can be modified by rename, hide and extend operators to match the query Ei.
- Adaptation step: not only a leaf in the sub-component associated in the realization level but also a sequence of operators used in the previous steps are applied. Then, a scheme in the code level is selected and the same operators in the selected leaf are applied.
- Composition step: the sub-specifications Ei and their implementations are composed. Next, the Identification step is introduced.

RC Identification

The component classification is described by a set of features. Each feature represents a property or an attribute of the component. They can be refined to give them a more precise meaning. The objective is to support effective retrieval of component. So, in order to be useful in practice, the component specifications should be independent of the kind of component, allowing the storage of different kinds of them, for example, requirements definitions, designs, documentation, and so forth. A very important approach that provides a formalism to register components properties is the work of Chen and Cheng (1997). It is based on the architecture and integration of the system facilitating the retrieval of a reusable component.

Being a RSL specification is a collection of modules, and a module basically a named collection of declarations as either a scheme or an object -objects and schemes defined using class expressions-, we propose a classification based on a feature-oriented model. Each feature describes a property of the component, and is characterized by:

(a) *kind of component*: describing the function of different kinds of components like: requirements specifications, designs specifications, systems specifications, documentation, and so forth;
(b) *operations* performed;
(c) *component structure*: modules involved (schemes and objects):
(c.1) *granularity of each module*: list of objects related with the class;
(d) *relationships* to another component ('implements' relations and composition of components); and
(e) *component specification style* (applicative sequential, imperative sequential, imperative sequential, applicative concurrent or imperative concurrent and abstract and concrete styles).

By localizing the possible components, the goal is to compare a component with the query. This process has two essential steps: signature matching and semantic matching. The signature matching enables a syntactic comparison of a query specification with specifications

Signature–Matching:

Query-Signature X RC-Library X Match-Predicate $\rightarrow$Set-of RC-Components

Signature-Matching(Q,C,P) = { c $\in$ C/ P(Q,C)}

existing in RC reusable components. The semantic matching compares the specifications dynamic behavior. The bases of the signature matching come from Zaremski and Wing (1997), even though they were adapted to the identification of RC components.

The signature of a specification consists of a set of sorts and a set of operations, each operation being equipped with a particular functionality.

Let L=<SL, FL> be the signature of a library specification and Q=<SQ,FQ> the signature of a query specification where SL and SQ are sets of sorts and FL and FQ are sets of operation symbols, the signature matching is defined in the figure above.

This means that given a query specification Q, a RC library C and a predicate P, it gives back the RC components that satisfy P. The signature matching is based on operations matching. Different kinds of operation matching can be applied. They are the exact, generalized and specialized matching of operations. This matching requires a specifications signature matching (sorts and operations) and the axioms proofs between pairs of operations.

Searching for a Reusable Candidate

The RC library is a fundamental piece to the engineering because it contains existing components that will be reused later in the large systems constructions. The appropriate use of it requires:

- the organization of the library, in order to facilitate the search of the components and to improve the efficiency retrieval, and
- the comparison between the description of the library component with the description of the new description.

The identification consists: first, in a syntactic comparison between a library component in RSL language and the user query one, and then, in a semantic comparison of the library component behavior with the query specification behavior both expressed by axioms. The former description corresponds to the syntactic matching, and the latter corresponds to the semantic matching. A very important factor to consider is the granularity levels of components. They are the list of objects that belong to a RSL module. The component can modify its size, from construct operations of the language to modules for big software systems. To describe and to reuse components they must be encapsulated.

FUTURE TRENDS

Our goal is to solve a problem which is a weakness of the RAISE formal method. Therefore, the proposal of our future work is to apply not only the software components reuse but also a domain specifications reuse, that is, in the confines of the domain engineering.

CONCLUSION

An overview of the strategy to classify and select a RAISE reusable component is presented. The RC model serves for the description of reusable components and a transformational process with reuse from RSL specifications to code is presented. This model aims to solve a problem that is a weakness of RAISE formal method.

REFERENCES

Beltaifa, R. & Moore, R: (2001). A software reuse infrastructure for an efficient reuse practice. (Technical Report No. 230). Retrieved from: *http://www.iist.unu.edu*

Bjorner, D. (2000). Software engineering: A new approach; Lecture Notes, Technical University of Denmark.

Chen, Y. & Cheng, B (1997). Formally specifying and analyzing architectural and functional properties of components for reuse. *Proceedings of 8th Annual Workshop on Software Reuse*, Columbus, OH.

Felice, L., Leonardi, C., Favre, L., & Mauco, V (2001). Enhancing a rigorous reuse process with natural language requirement specifications. *Proceedings of '2001 Information Resources Management Association International Conference'* (pp. 271-275) Toronto, Canada.

George, C. (2001). RAISE tools user guide (Technical report No.227). Retrieved from *http:// www.iist.unu.edu*

George, C. (2002). Introduction to RAISE (Technical report No. 249). Retrieved from *http:// www.iist.unu.edu*

George, C., Haff, P., Havelund, K., Haxthausen, A., Milne, R., Nielsen, C., Prehn, S. & Ritter, K. (1992). *The RAISE specification language*. Prentice Hall.

George, C., Haxthausen, A., Hughes, S., Milne, R., Prehn, S., & Pedersen, J. (1995). *The RAISE development method*. BCS Practicioner Series. Denmark. Prentice Hall.

Helm, R. & Maarek, Y. (1991). Integrating information retrieval and domain specific approach for browsing and retrieval in object-oriented class libraries. *Proceedings of OOSPLA91* (pp.47-61).

Hennicker, R. & Wirsing, M. (1992). A formal method for the systematic reuse of specifications components. Lecture Notes in Computer Science 544, Springer-Verlag.

Krueger, C. (1992). Software reuse. *ACM Computer Surveys*, *24*(2).

Maarek, Y., Berry, D., & Kaiser, G. (1991). An information retrieval approach for automatic constructing software libraries. *IEEE Trans. Software Engineering, 17*(8), 800-813.

Mauco, V. & George, C. (2000). Using requirements engineering to derive a formal specification (Technical Report No. 223). Retrieved from *http:// www.iist.unu.edu*

Penix, J. (1998). Unpublished doctoral dissertation, University of Cincinnati, Cincinnati, OH.

Zaremski, A. & Wing, J. (1997). Specification matching of software components. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, *6*(4), 333-369.

KEY TERMS

Algebraic Specification: A technique whereby an object is specified in terms of the relationships between the operations that act on that object. A specification is presented in four parts:

1. Introduction part, where the sort of the entity being specified is introduced and the name of any other specifications that are required are set out;
2. Informal description of the sort and its operations;
3. Signature, where the names of the operations on that object and the sorts of their parameters are defined; and
4. Axioms, where the relationships between the sort operations are defined.

Reusable Component: The aim of reusable components is to provide services to other components and to require services from others. There are three concepts associated with them: their interfaces, their architectures and their realizations.

Reuse Process: A method having the following steps: the decomposition of the goal specification, the identification for each specifications, the adaptation step using the appropriate operators and the composition of the sub-specifications.

Rigorous Process: A mathematically-based technique for describing and reasoning about system properties. It provides frameworks within which people specify, develop, and verify systems in a systematic manner.

Semantic Matching: Semantic matching compares the specifications dynamic behavior.

Signature Matching: Signature matching enables a syntactic comparison of a query specification with specifications existing in a model.

Specification Matching: The process to determine if two software components are related. It allows us to distinguish when two component specifications match, we know that certain guarantees about the behavior of the system will hold if we substitute one component for the other.

Risk Management in the Digital Economy

Bob Ritchie
University of Central Lancashire, UK

Clare Brindley
Manchester Metropolitan University, UK

INTRODUCTION

Digital Economy and Risk: A Two-Edged Sword

The digital economy has been generated by radical changes to every aspect of business and commerce in the last two decades. These changes are far more significant than the developments in Information and Communication Technologies (ICTs) that have largely facilitated the digital economy itself. Every business sector has witnessed changes in the competitive structure of the marketplace, consumer preferences, buying habits, marketing and promotional strategies, production operations, internal administration systems, supply chain arrangements, and the opening up of the global economy. Managers would concede that the uncertainties and risks in running their businesses, as a result of such changes, are not only much greater than previously but increasing. However, the digital economy is a two-edged sword in the sense that the ICTs generating the additional uncertainties and risks also provide the means to enable decision makers to manage them more effectively. The key to survival in the digital economy rests with the abilities of the managers to utilize ICTs effectively to manage uncertainties and risks.

ICTs have largely been seen as helping to enhance database access, analytical powers, and the communications capacity of managers. The justification for these efforts has been based on the premise that more and better quality information will result in reduced uncertainty and improved risk perceptions in decision situations. In short, the outcome would be reflected in "better quality" decisions in terms of risk assessment and resolution.

BACKGROUND

The Digital Economy

The term "digital economy" reflected in the title of this paper may be viewed from a variety of perspectives:

1. Technology developments, especially those relating to the digital communication of data and information, are usually considered the primary driver in the creation of the digital economy. Increases in speed, improvements in capacity, accuracy, reliability, general quality, and ease of use are all features that have enabled the widespread adoption and development of digital technologies. The developments in audiovisual communication technologies and wireless technologies are opening up further opportunities for the transmission and exchange of business intelligence.
2. Socio-economic changes have been equally influential in the rapid adoption of the new technologies. Individuals of all ages, educational and social backgrounds are prepared to regularly use mobile communications, access the Internet and engage in interactive video communications, often with friends and family in different parts of the globe. The impact that these changes in individual and group social behaviors have had on the rate of adoption of new technologies should be fully appreciated. The reasons underlying such changes are multifaceted and complex and beyond the scope of our present discussion, although they have broadly been influenced by individual social and economic needs.
3. Micro-economic factors at the level of the individual organization have been responsible for "pulling" and "pushing" organizations and their management towards increased attention and adoption of the digital economy. Significant "pull" factors include demands from end users of the product or service (e.g., requests for more detailed information on the product/service prior to and subsequent to the purchase, in terms of performance, maintenance, modifications, upgrades). The "push" factors are typically associated with the business organization seeking to maintain its competitive position by offering services equivalent to those of its main competitors, especially if these may be viewed as providing a distinctive competitive advantage (e.g., providing detailed product information via the Web and enabling customers to order directly). Some of

Table 1: Key topic areas presented in this article.

Key topic areas relating to digital economy and risk presented in this article include:
• primary elements of the digital economy
• overview of risk and risk management
• risk and uncertainty
• individual/organizational response to resolving risk
• role of information search and corporate intelligence
• contribution of the digital economy to risk resolution
• individual characteristics and risk perceptiveness
• management of risks
• risk perception
• information processing and risk resolution
• risk management within the digital economy

the issues involved will be discussed further in later sections of the paper.

4. Macro-economic factors are particularly significant in enabling the development of the digital economy, although they are often less evident when exploring individual product/market developments. Changes in legislation affecting consumer rights, guarantees of financial transactions, security of information held on computer systems, and commercial contracts negotiated via the Internet are all examples of the changes in the macro-economic environment needed to facilitate and support the development of the digital economy. Without such changes, individual organizations and customers might consider the risks of such commercial transactions to be too high. In essence, the responsiveness of governments and other similar institutions have lagged behind many of the demands placed on them by the rate of change in the digital economy.

It may be argued that defining the term "the digital economy" remains problematic due to the number of perspectives from which this term may be viewed and due to the number and interactive nature of the variables involved.

OVERVIEW OF RISK AND RISK MANAGEMENT

The key dimensions of risk and its management are represented in Table 2.

Risk and Uncertainty

This seemingly simple term "risk" has proved somewhat problematic in arriving at an agreed definition. Most academic fields and researchers (e.g., Dowling & Staelin, 1994; Knight, 1921) provide variations on the theme, though most would agree that risk relates to two dimensions: the *Likelihood* of a particular event occurring (i.e., probability), and the *Consequences* should this event occur.

In the case of the consequences, it has been common to assume that these are generally undesirable, e.g., financial loss or even loss of life. Sitkin and Pablo (1992, p.9) define risk as "the extent to which there is uncertainty about whether potentially significant and/or disappointing outcomes of decisions will be realised." Similarly, MacCrimmon and Wehrung (1986) identified three components of risk: the magnitude of loss, the chance of loss, and the potential exposure to loss. However, it is important to recognize that there would be no point in taking risks unless there were some benefits to compensate for the possible negative outcomes (Blume, 1971). An associated feature is that of differing risk perceptions. Different individuals, groups of individuals, and organisations may view or perceive the risks (i.e., the likelihood of occurrence, nature and scale of negative consequences, and the potential rewards) differently (e.g., Forlani & Mullins, 2000; March & Shapira, 1987).

The term "uncertainty" typically reflects the ambiguity surrounding the decision situation in terms of the precise nature of the situation, its causes, possible solutions, and the reaction of others to possible actions taken. Rowe (1977) has defined uncertainty as the absence of information concerning the decision situation and the need to exercise judgment in determining or evaluating the situation.

Table 2. Key dimensions of risk and its management

• risk and uncertainty
• risk solution
• role of information search and corporate intelligence
• contribution of the digital economy
• individual characteristics
• management of risk
• risk perception

Risk Resolution

A natural reaction by decision makers facing uncertainty and risk is to seek to resolve the uncertainty and risk inherent in the decision situation. There are several actions that may be taken:

- seek to understand the parameters of the "problem" or situation;
- assess the predictability of the parameters;
- consider the possibility of eliminating or ameliorating the risks; and
- assess the attitudes of the decision takers towards the risks.

These elements represent the process that individuals and organizations develop to manage the risk, to position the organization, and to develop appropriate strategies to manage the impact of such events (Bettis, 1982). The increasing complexity and multitude of relationships consequent to the digital economy pose new challenges for the management of risk at the organizational level.

Role of Information Search and Corporate Intelligence

The decision maker confronted with a risky decision situation naturally follows a process of gathering more information, processing this in different ways, and evaluating this in relation to the risks faced. This information is then used—maybe consciously or unconsciously—by the decision maker to assess to what extent the risk has been removed, reduced, or resolved. If the decision maker is not satisfied that s(he) has achieved sufficient resolution of the risks involved, then further information searches, analysis, and processing will occur. However, this intelligence-gathering process need not necessarily result in improved understanding and risk resolution. Uncovering more information may reveal more influencing variables causing perceptions of greater uncertainty, both as a consequence of uncovering new influencing variables and an increasing sense of complexity. Ritchie and Marshall (1993) argued that the quality and sufficiency of the information available will influence the perception of risk by those involved in the decision-making process. Ritchie and Brindley (1999) argued that risk perception is both the consequence of information search and analysis as well as its determinant.

Contribution of the Digital Economy

Information can be considered as a risk-reduction or risk-insulating tool, on the basis that more and better information would result in improved decision making and more effective risk management. The rapid pace of ICT developments (Kalokota & Robinson, 2000; Rycroft & Kash, 1999) and the emerging digital economy (Brindley & Ritchie, 2001) demonstrate that both individuals and organizations have easier access to more information that is more immediate and arguably more relevant. The decision makers can therefore access through their own desktops, internal data, external market reports, competitor information, etc. Swash (1998) recognized that it is essential to overcome the problem of information overload by ensuring the Internet is used frequently, thus improving the users' surfing proficiency and their knowledge of "useful" sites. Zsidisin and Smith (2004), in examining the impact of the Internet on supply chain decisions, suggested that it is natural that information exchange improves. Improvements in the voluntary exchange of information may produce better knowledge of the situations surrounding the dynamics of a business or commercial relationship. This provides greater potential for detecting, averting, and managing risks.

Individual Characteristics and Risk

It is generally accepted that there are differences in individual risk perception, information-processing styles, and decision-making attributes (e.g., Chung, 1998). There is perhaps less agreement on the nature and consequences of such individual differences. For example, differences in gender have been posited as the reason for women being more meticulous in information search, responsive to decision cues, and being more risk averse than men (e.g., Chung, 1998). Others have failed to establish significant differences (e.g., Masters & Meier, 1988), suggesting that other contextual variables (e.g., social, educational, and experiential background) may be more relevant to any differences in behavior. For example, the decision to start a business may generate differences in risk-taking behavior that is not gender-, culture- or age-related (Busenitz, 1999; Shapira, 1995). A number of authors (see Forlani & Mullins, 2000; Ghosh & Ray, 1997; Kahneman & Lovallo, 1993) have suggested that the provision of structured decision approaches and information search frameworks can provide the appropriate mechanism to overcome any biases and improve the quality of decisions.

Management of Risk

The strands of risk together with business intelligence in the digital economy may be captured in the three dimensions of individual/group/organizational behavior when confronted with risk in given decision situations:

- Willingness to seek further resolution of the decision situation faced, both in terms of the context and the decision-specific variables.
- Desire to identify and measure the risks in some way, either objectively or subjectively.
- Information search and processing to support the decision.

These three dimensions are closely inter-related and are employed to modify the risk perceptions of the decision makers. In addition to these decision-making activities is the range of activities associated with risk management. Some risk management activities may be undertaken prior to the decision itself (e.g., insuring against certain risks) or after the decision (e.g., effective management of relationships with customers to reduce the likely incidence of disputes). While activities of this type may be employed fairly readily with local markets, many managers may find it more difficult to avoid the risks resulting from increased global competition in their home or local markets, consequential of the digital economy.

Risk Perception

Throughout the decision process, an individual is seeking information to remove the uncertainties and to resolve the risks involved in the decision.The factors that contribute to the perceptions of uncertainty and risk experienced by the individual may be represented in table 3.

The integration and interaction of many of these components increases the complexity of understanding both risk perception and decision behavior. For example, although funding may be available to undertake intensive research, the pressures of time to respond to a given situation may preclude such activities even though they may be considered desirable.

Information Processing and Risk Resolution

The information-processing behavior of the individual may be presented in terms of the process described in Diagram 1. The decision maker seeking to resolve the risks perceived will search for appropriate information, process, analyze, evaluate, and synthesise this with other information relating to the situation. The contribution of the digital economy is evident not only in providing the initial intelligence but in assisting the decision maker in the processing and manipulation of the data prior to the decision-making stage. It is suggested that the decision maker either consciously or sub-consciously assesses whether the risks have been sufficiently resolved or not. In many respects, this may be posing the question concerning one's own degree of confidence in proceeding with the decision, though it may also be influenced by the factors identified.

An assessment that the risks had not been adequately resolved would normally result in further information search, analysis, synthesis, etc., repeating this cycle until the individual believed the risks had been sufficiently resolved to proceed. It is unlikely that the decision maker would achieve the situation where s(he) was fully confident concerning risk resolution. The probable outcome, even after seeking further risk resolution, is that of having to make the decision even though the risks are not fully resolved, as a consequence of pressure of time, lack of resources to pursue further information, or the feeling that further search is unlikely to prove cost-effective. This is probably a very common outcome reflected in the statements of many decision makers that "there are many other angles that we ought to explore but we no longer have the time to do so."

In situations where the decision maker considers that the risks have been adequately resolved, one might conclude that the normal procedure would be to proceed with the decision. Another possible outcome could be suggested where the decision maker decides to delay the decision. This may be simply a reluctance to commit oneself even if one is fairly clear and confident of the decision to be made. Alternatively, it may reflect a more deliberate tactic of delay to await changes in the decision situation, to assess the likely responses of other participants in the competitive situation, or perhaps in the hope that someone else may take the responsibility. The making of the decision is no longer viewed as the concluding stage in the process. Attention is increasingly being directed to the post-decision activities that seek to ensure

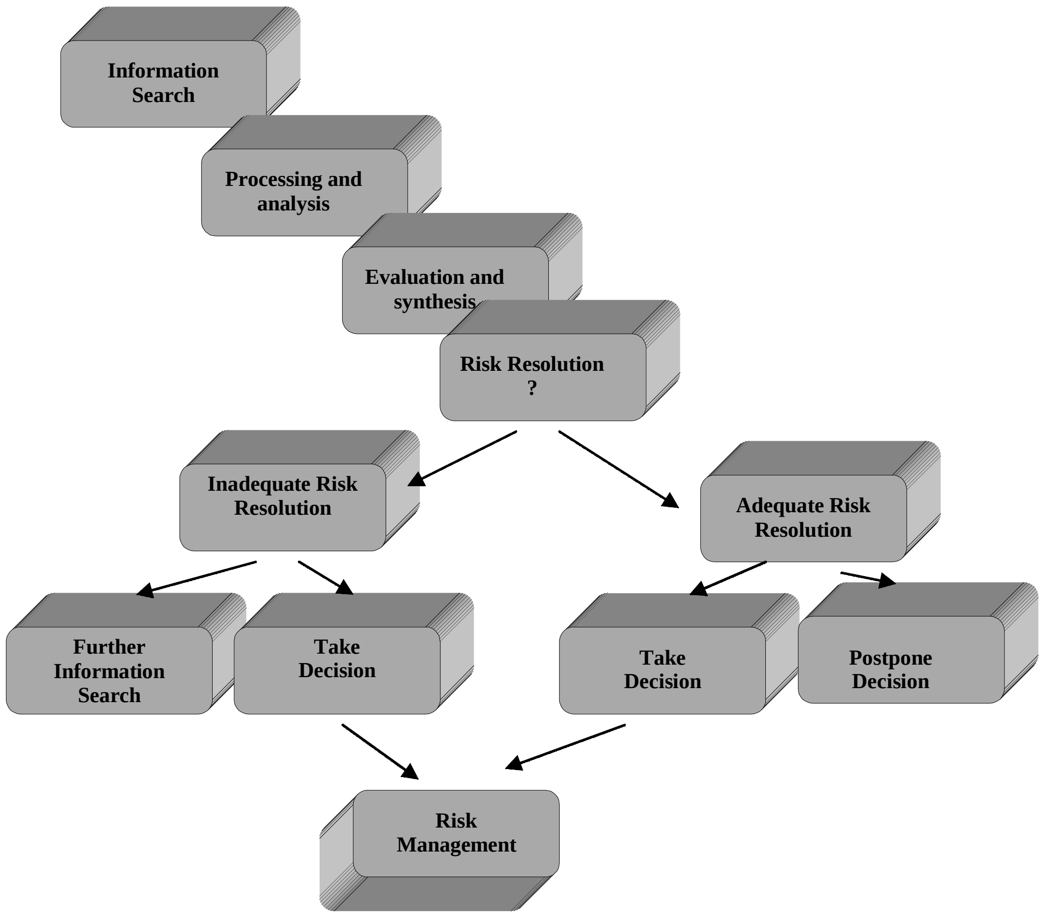
Table 3. Main groups of factors influencing individual risk perceptions

• direct implications in terms of costs and benefits and the scale of these
• other decision outcomes, known with less certainty (e..g. reactions of peers and colleagues)
• time available in which to take the decision, typically constrained
• funding available to undertake the research, analysis and evaluation processes involved
• achievement of personal goals that decision maker(s) seeks to satisfy
• data quality which may significantly influence the nature of the risks perceived

Source: Brindley and Ritchie (2001)

R

Diagram 1: Information processing and risk resolution. Adapted from Ritchie and Marshall (1993).



that the opportunity is taken to avoid potential risks anticipated in the decision processing to avoid risks seen as less likely to impact on the situation, and to minimize the likelihood and consequences of those risks seen as more likely to occur. A generic term for these activities is *Risk Management*. Examples of Risk Management activities are presented in Table 4.

FUTURE TRENDS

Risk Management within the Digital Economy

The digital economy will impact on organizations irrespective of their size, geographic location, or sector. While such developments will engender greater uncertainty and risks, they will also facilitate solutions and opportunities for organizations to resolve and manage the new risks. Specific challenges for the organizations in the digital economy are:

1. Fundamental changes in the marketplace as a consequence of providing more direct communications with the consumer, greater heterogeneity in the market in terms of consumer needs and behavior, and a movement in the balance of power towards the consumer as opposed to the manufacturer or service provider.
2. Individuals and organizations are unlikely to behave in a rational and structured manner to resolve risks. The previous emphasis on predictable patterns of information search, corporate business intelligence, and evaluation of information is likely to be replaced by less predictable demands as the nature of the problems encountered are less predictable. This will have significant implications for the development and design of business intelligence systems.
3. Predictions that the digital economy—by providing improved access to information and processing capabilities—will lead to improved decisions is unlikely to be sustainable.

Table 4. Risk Management Activities

• insuring against the financial consequences of particular outcomes
• developing formal contractual arrangements with trading partners to limit scope for non-prediction behaviour
• training and staff development
• communication of intentions within the appropriate internal and external networks to ensure involvement and commitment (e.g. relationship marketing)
• detailed planning, monitoring and control at all stages
• building close relationships with key partners to generate a sense of common purpose and trust (e.g. partnering)
• developing more effective Risk Management strategies through more effective communications at all levels both within the organisation and with external partners

4. The modus operandi of competitive and business relationships within the digital economy will evolve as the nature and extent of risks faced change, both in terms of novel risk situations encountered and the rate at which these occur.
5. The development and training of the individual in terms of the appropriate knowledge, skills, and behavior patterns provides an important requirement to utilize the information available effectively.
6. The nature of what constitutes "effective" risk management may well change in the digital economy (Ritchie & Brindley, 1999).

CONCLUSIONS

The digital Economy has produced a fundamental change in the nature of the risks faced and increased the likelihood that the marketplace will remain turbulent, unstable, and risk prone. Managers now need to have the capability—through improved knowledge, skills, and understanding—to identify, analyze, and manage these competitive developments and the associated risks. Associated with the improved capability to manage the risks is the ability to implement a wider range of risk management strategies to ameliorate the consequences of the incidence of risks and their consequences. The digital economy and the associated ICTs improve the opportunities to ensure effective risk management.

REFERENCES

Bettis, R.A. (1982). Risk considerations in modeling corporate strategy. *Academy of Management Proceedings*, 22-25.

Blume, M.E. (1971). On the assessment of risk. *Journal of Finance*, 26(1), 1-10.

Brindley, C.S. & Ritchie, R.L. (2001). The information-risk conundrum. *Marketing Intelligence and Planning*, 19(1), 29-37.

Busenitz, L.W. (1999). Entrepreneurial risk and strategic decision making: It's a matter of perspective. *The Journal of Applied Behavioral Science*, 35(3), 325-340.

Chung, J.T. (1998). Risk reduction in public accounting firms: Are women more effective? *International Review of Women and Leadership*. 4(1),39-45.

Dowling, R.G. & Staelin, R. (1994).A model of perceived risk and intended risk-handling activity. *Journal of Consumer Research*, 21(1),119-25.

Forlani, D. & Mullins, J.W. (2000). Perceived risks and choices in entrepreneurs' new venture decisions. *Journal of Business Venturing*, 15 (4),305-322.

Ghosh, D. & Ray, M.R. (1997). Risk, ambiguity, and decision choice: Some additional evidence. *Decision Sciences*, 28 (1),81-104.

Kahnemann, D. & Lovallo, D. (1993) Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management Science*, 39(1),17-31.

Kalakota, R. & Robinson, M. (2000). *e-Business*. Reading, MA: Addison-Wesley Longman.

Knight, F.H. (1921). *Risk, uncertainty and profit*. Boston and New York: Houghton Mifflin Company.

MacCrimmon, K.R. & Wehrung, D.A. (1986). *Taking risks: The management of uncertainty*. New York: Free Press.

March, J.G. & Shapira, Z. (1987).Managerial perspectives on risk and risk taking. *Management Science*, 33 (11),1404-1418.

Masters, R. & Meier, R. (1988). Sex differences and risk-taking propensity of entrepreneurs. *Journal of Small Business Management*, 26(1), 31-35.

Ritchie, R.L. & Brindley, C.S. (1999). Relationship marketing as an effective approach to organisational risk management strategies. *Proceedings of the Fourth International Conference on the Dynamics of Strategy*, 22-23 April 1999, Surrey, UK, 1, 313-323.

Ritchie, R.L. & Marshall, D.V. (1993). *Business risk management*. London: Chapman and Hall.

Rowe, W.D. (1977). *Anatomy of risk*. New York: Wiley.

Rycroft, R.W. & Kash, D.E. (1999). *The complexity challenge*. London: Pinter Publishers.

Shapira, Z. (1995). *Risk taking: A managerial perspective*. New York: Russell Sage.

Sitkin, S.B. & Pablo, A.L. (1992). Reconceptualizing the determinants of risk behaviour. *Academy of Management Review*. 17(1),9-38.

Swash, G. (1998). UK business information on the Internet. *New Library World*. 99(1144), 238-242.

Zsidisin, G.A. & Smith, M.E. (2004). Early supplier involvement as a tool for reducing supply risk in supply chain risk. In C. Brindley (Ed.), UK: Ashgate Publishing, 117-129.

KEY TERMS

Decision support: the tools, techniques, and information resources that can provide support to the decision maker in improving the efficiency and effectiveness of his/her decisions. Many of these decision support tools may employ ICTs and be part of the Management Information System itself.

Digital economy: accepts as its foundation, the ICT developments and represents the impact that these have had on the conduct of business and commercial activities. Changes in markets and supply chains as well as increasing global competition all represent what is encapsulated within the term the digital economy.

Information and Communication Technologies (ICTs): a generic term used to encapsulate the diverse range of technological developments (e.g., computer storage and retrieval, computing capacity, wired communications, wireless communications, portable technologies) that have enhanced the internal and external activities of organizations. Especially important is the manner in which these strands of technological development have been integrated to provide greater synergy.

Management Information: a term that covers a wide variety of sources and types of information that may prove valuable to the decision making, management, and control of an organization. This term would include quantitative and qualitative information types, internal and externally sourced information, as well as classifying the information in terms of its quality (e.g., accuracy, detail, relevance, timeliness).

Risk: in a limited manner, the decision situation in which the full range of possible outcomes are known with certainty and the probability of their occurrence can be assessed accurately, usually by some objective means (e.g., rolling the dice is a classic risk decisions situation). More usually, the probabilities must be assessed subjectively, often based on previous experiences or intuition, and the outcomes themselves may not be fully identifiable. The term "risk" is used commonly to generally define decision situations that are really a combination of classical risk and uncertainty, i.e., the more normal decision situation in organizations.

Risk Management: the range of activities that may be taken to avoid the occurrence of an undesirable event or to modify, minimize, or eliminate the consequences should the event occur (e.g., an insurance policy against particular risks would not prevent the occurrence, but would compensate for the financial and other consequences of the outcome).

Risk Perception: the term used to express how a situation is viewed or seen by the decision maker(s). Individual characteristics, experiences, and beliefs may influence the way in which we might view a given situation as being either more or less risky. Usually this is measured on a subjective and relative scale (i.e., Situation A is perceived as riskier than B) rather than on an objectively measurable scale.

Uncertainty: the situation where less than perfect knowledge exists about a particular problem or decision requirement. There exists a wide variation in terms of degrees of uncertainty from extreme uncertainty (i.e., very limited knowledge of outcomes or likelihood of their occurrence) to near certainty (i.e., almost complete knowledge of the outcomes and the likelihood of occurrence). Generally, an uncertain decision situation refers to one containing ambiguity about part or all of the decisions parameters.

Rotating Banner Advertisements on the World Wide Web

Subhasish Dasgupta
The George Washington University, USA

Rajesh Chandrashekaran
Fairleigh Dickinson University, USA

INTRODUCTION

Advertising spending on the Internet has soared. Indeed, by some accounts, Internet advertising is projected to reach $23.5 billion by 2005 (eMarketer, 2002). Although there are several ways of advertising on the Internet, for example, buttons, banners, paid links, superstitials, and so forth, banner advertising is the most common form of advertising on the Internet (Meland, 2000). Advertising using banners (usually placed near the top of a page) is currently the most popular form of online advertising. Banners may be static (stationary) or dynamic (rotating). In the case of static banners, all visitors to a particular site are exposed to the same banner. In contrast, dynamic banners describe cases where ad servers to a particular site deliver different banners to different clients/visitors. This approach presents the possibility of time/space sharing among different advertisers.

This article discusses one particular type of dynamic/ rotating banner advertising. Specifically, we present a model to deliver and track the effectiveness of dynamic rotating banner advertisements. The advertising is dynamic in that different banners may be delivered to different users, and it is rotating in that each user (depending on the length of time spent at that site/page) will be exposed to multiple advertisements. These banners may be from a single advertiser (different executions of an ad) or from different advertisers. The latter allows for better space/time sharing and cost effectiveness.

Rotating ads provide the ability to deliver multiple ads to users by associating many ads with a single Web page. Most Web sites have popular pages that users visit often. Rotating banners allow companies to deliver more than one advertisement for these pages, thereby increasing their yield from the page.

Measuring click-through, that is, clicking on an ad banner, has been and remains important in assessing the effectiveness of online advertisements. Research has shown that there are many factors that influence peoples' click- through behaviors (Cho, 2003a). For example, many banner-related factors, for example, the size of the banner, location of the banner, and dynamic animation (Razzouk & Seitz, 2003) and other individual factors, for example, person's involvement with the product/service being advertised (Cho, 2003b; Yoon, 2003) determine the effectiveness of banner ads on the Internet. For example, Cho (2003b) confirmed that while people whose involvement is low are generally less likely to seek additional information, they are also more likely to be influenced by ad-related factors.

However, of late, click-through rates have been declining (Dreze & Hussherr, 2003). This disturbing trend has caused advertisers to think about ways in which click-through rates may be improved. Recent findings suggest that advertisers may be able to reverse the trend by using rotating and dynamic banner advertisements. For example, Chatterjee, Hoffman and Novak (2003) found that there is significant heterogeneity in consumers' click-proneness to banner advertisements and that there are significant gains from repeated exposures to banner ads – but mainly for consumers who are not otherwise click-prone. In addition, consumers who are more involved with the product are more likely to click than those who are not (Cho, 2003b). Such findings might suggest that rotating (rather than static) banner ads, which allow for space/time sharing enabling multiple exposures to messages from multiple advertisers, may be an effective way to improve click-through rates and effectiveness of banner advertisements. Recent experimental evidence also confirms that the level of attention and message recall or association of banner advertisements is also a function of position (i.e., top versus bottom) and graphics (Razzouk & Seitz, 2003).

BACKGROUND

Review of Web Advertisement Models

The process of advertising on the Web consists of two parts: delivery of advertisement to a client computer using the Internet, and tracking effectiveness of the delivered advertisement. There are two basic types of advertisement delivery models: delivery-focused and measurement-focused. Delivery-focused models emphasize the advertisement delivery method and do not have the ability to track effectiveness of the advertisement. Measurement-focused models use modified advertisement delivery mechanisms that help in measuring effectiveness.

We consider two delivery-focused models: the static ad model and the dynamic ad model. In the static ad model, when a user requests a page, the Web server responds with the requested page. The page has content and HTML code to insert a banner image into the Web page. This banner image is the advertisement. A server finds the requested banner image for the Web page. The page content and banner image are then transmitted to the user's computer, that is, the client, over the Internet. A Web server log records transfer of the content page as a "hit" in the server's log. When an ad file is sent out to the client, the server also records that the user's browser successfully downloaded the advertisement. In this model each page has only one ad associated with it and this ad is changed in batches, either once per day or once per week.

The dynamic ad model is very similar to the static ad model described previously. In addition to the Web server, there is a separate server called an ad server that stores all the banner advertisements that are delivered to the client. The ad server also has special software that makes a decision regarding which ads should be served to the client or user. A summary of delivery-focused models is provided in Table 1.

We consider two measurement-focused models. The first, cache-measurement, "allows for the appropriate measurements of ads stored and provided from cache by a proxy server or browser, as well as those ads actually provided by the ad server" (Bennett 1998). The second, browser measurement, allows for the recording of ad-related activity using software (e.g., Java) that runs on the requesting browser. A summary of the measurement-focused models is provided in Table 2.

Static and Rotating Banner Advertisements

Static banner advertisements are defined as those in which the banner space is utilized by a single advertiser, whose banner appears along with a Web page through the duration of client's visit. Each such ad provides a link with the advertiser's home page via a "button". We should note that the term *static* simply refers to the fact that the ad belongs to a single advertiser. However, each ad may include moving or animated elements (text, graphics, or both) and other features that attempt to draw attention to the ad.

In contrast, *rotating banner advertisements* refer to ads belonging to different advertisers that can share the same banner space for the duration of the page visit. That is, two or more advertisements appear in succession on the user's screen in the same banner space. Each ad appears on the user's screen for a predetermined duration, and is then replaced by another ad belonging to a different advertiser. This "rotation" continues as long as the page is displayed on the user's screen. Furthermore, similar to static ads, each ad in the rotation provides a link to the advertiser's page for the duration of display.

FUTURE TRENDS

A Framework for Delivery and Tracking of Rotating Banners

In this section we develop a theoretical framework to examine how rotating banner ads may be delivered and evaluated. Figure 1 shows the proposed framework for the delivery and tracking of rotating banners.

The proposed system consists of two components: ad delivery and ad tracking. The following sections describe each in turn.

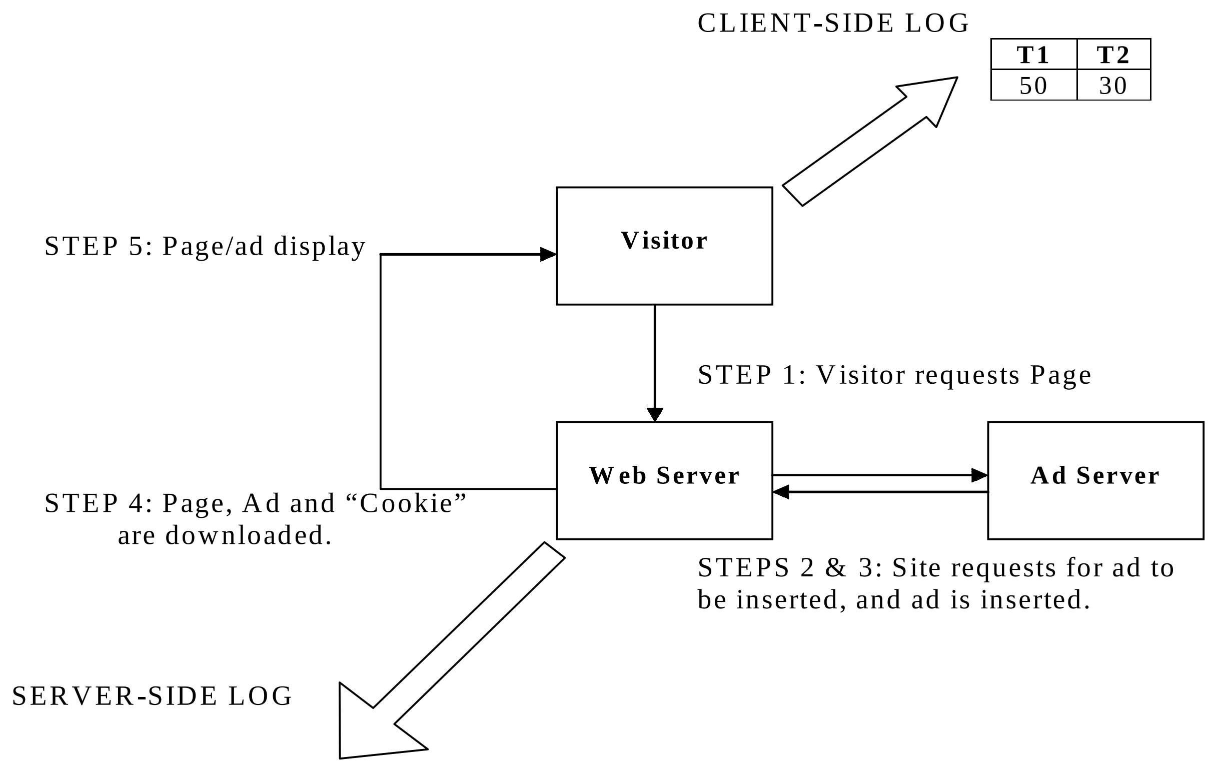
Table 1. Delivery focused models

Model	Content Delivered by	Ad Delivered by	Delivery Mechanism
Static Ad Model	Web server	Web Server	One ad associated with each page. These ads are changed in daily or weekly intervals.
Dynamic Ad Model	Web server	Ad server	Different ads are dynamically associated with each page. This can be done based on the user.

Table 2. Measurement focused models

Model	Content Delivered by	Ad Delivered by	Measurement Mechanism
Cache-measured	Web server	Ad server	Emphasis is on counting ads provided from cache by a proxy server.
Browser-measured	Web server	Ad server	A browser run program keeps track of Ads downloaded.

Figure 1. Ad delivery and tracking system



Client	Time (T)	S1	S2	C1	C2
1	0	1	0	0	0
	30	1	1	0	0
	60	2	1	0	0
	80	2	1	1	0

Note:
T = Time elapsed since page and ads were downloaded
S1 and S2 = Number of opportunities to see ads 1 & 2 respectively
C1 and C2 = Click status of ads 1 & 2 respectively
T1 and T2 = Total time of display of ads 1 & 2 respectively

Ad Delivery

When a Web page is downloaded from the server it downloads the contents of the page (containing the Java Script for the rotation of the ad images), the ad images, and a program to keep track of time spent on the page (T_p) and click status of each ad (C_i). Once the page and the ads have been loaded on the client's browser, the Java Script program that is built into the Web page is executed using the client-computer's resources and provides the ad rotation. This program is crucial to the implementation of the rotating banner system.

Ad Tracking

The tracking system consists of two sub-systems. The one that resides on the Web (or ad) server is called the server side sub-system, and the other, which utilizes the user's computing resources, is called the client side sub-system. Each sub-system is further described next. For explanation purposes we consider a system that consists of two rotating ads: Ad 1 and Ad 2.

Server Side Sub-System

The server-side sub-system is responsible for maintaining a log of the time (T) elapsed since the page was sent (start of the download by client). In addition, it maintains counts (S_1 and S_2) of the number of opportunities that the user has had to see each ad (ads 1 and 2 respectively). S_1 and S_2 are incremented based on the frequency of rotation, that is, how many seconds (t) each ad is displayed for. The log file also maintains a record of the "click status" of each ad using counters C_1 and C_2 for ad 1 and ad 2 respectively. Initially, all counters are set to zero. When the client clicks on a particular ad (say Ad 1), the respective counter (here, C_1) is incremented by one, but C_2 remains unchanged (refer to server side log in Figure 1).

Using this program, the Web server can not only measure the click-through rate of each ad, but can also measure the number of opportunities the client has had to see each ad. The latter measure can be used to determine reach, gross impressions, and other variables that may be used in sophisticated pricing models. This additional information also allows the server to estimate an ad's effectiveness on dimensions other than those that are behavioral in nature. Existing models concern themselves only with click-through rates (see IAB and MB Interactive study, 1997).

Client-Side Sub-system

The client-side sub-system is responsible for maintaining a log on the client's side. This log contains such information as the client code and duration of display for each ad (T1 and T2). The sub-system uses a program that is downloaded to the client's computer along with the Web page requested. The code in the program creates a time stamp when the ad was received using the client's own computer clock and stores it in the hard drive. This mechanism operates like a "cookie," and can be implemented using programs like Java applets.

Consider the scenario when the client clicks on one of the rotating banners. At this time, the data in the cookie (i.e., the entire log from the client's side) are transmitted back to the server log, which then updates the appropriate fields. In this case, the time is noted based on when the click-through was made. Based on the start time and end time transmitted to the server log, the actual time duration (T_1 and T_2) for which each ad was displayed on the client's browser can be calculated and stored in the server log.

Consider another scenario in which the client does not click on any of the ads on the page, but moves to a different page on a different server. In this case, as soon as the user moves to a different page the end time is noted from the computer's internal clock and stored on the hard disk as a "cookie" by the client–side program. The next time the same client connects to the Web site (server), the data in the cookie are transmitted back to the server log, and the actual ad display times are recorded in the log.

It is important to note that there may be differences between the server-side and client-side logs because the server-side log does not consider the time taken to transmit data and images from the server to the client and back. The client program will be more accurate since its time logs are not affected by the time taken to transmit data and images. To reconcile these differences, the client side data are used to calculate the final values of S1, S2, T1 and T2.

CONCLUSIONS

In this article we have presented an overview of models used for delivery and tracking banner advertising on the Web. We introduced the concept of rotating banner advertising and presented a model that can be used for delivery and measurement of such advertising. For advantages and limitations of this model see Dasgupta and Chandrashekaran (2001).

We would like our readers to note that technology in this field is changing rapidly and newer methods of Web advertising are being introduced all the time. But we still believe that this encyclopedia article provides the right overview of technologies issues, marketing concepts, and terms and definitions related to the field of Web-based advertising.

REFERENCES

Bennett, R. (1997). How Internet ads are delivered and measurement information. ABC Interactive, White Paper. Retrieved August 2001 from *http://www.accessabvs.com/webaudit/admeasurement.html*

Chatterjee, P., Hoffman, D.L., & Novak, T.P. (2003). Modeling the clickstream: Implications for Web-based advertising efforts. *Marketing Science, 22*(4), 520-541.

Cho, C.H. (2003a, summer). Factors influencing clicking of banner ads on the WWW. *CyberPsychology & Behavior, 6*, 201-215.

ho, C.H. (2003b). The effectiveness of banner advertisements: Involvement and click-through. *Journalism and Mass Communication Quarterly, 80*(3), 632-645.

Dasgupta, S., & Chandrashekaran, R. (2001). Delivery and tracking of rotating banner advertisements on the World Wide Web: An information systems model. In A. Gangopadhyay (Ed.), *Managing business with electronic commerce*. Hershey, PA: Idea Group Publishing.

Dreze, X., & Husser, F. (2003). Internet advertising: Is anybody watching. *Journal of Interactive Marketing, 17*(4), 8-23.

EMarketer. (2002). Online advertising perspective. Retrieved August 2001 from */www.emarketer.com/analysis/eadvertising/20010412_ead.html*

IAB Online Advertising Effectiveness Study. (1997). Retrieved on August 27, 2004 from *http://www.mbinteractive.com*. San Francisco, CA: Millard Brown Interactive.

Kurose, J.F., & Ross, K.W. (2000). *Computer networking*. NY: Addison-Wesley Longman Inc.

Meland, M. (2000, February 13). Banner ads get sexy. *Forbes*, 28-29.

Razzouk, N., & Seitz, V.A. (2003). Banner advertising and consumer recall: An empirical study. *Journal of Promotion Management, 9*(1, 2), 71-80.

Yoon, S.-J. (2003). An experimental approach to understanding banner adverts' effectiveness. *Journal of Targeting, Measurement and Analysis for Marketing, 11*(3), 255-272.

KEY TERMS

Applet: An applet is a program written in the Java™ programming language that can be included in an HTML page, much in the same way an image is included. When you use a Java technology-enabled browser to view a page that contains an applet, the applet's code is transferred to your system and executed by the browser's Java Virtual Machine (JVM) (*http://java.sun.com/applets/*).

Browser: A computer software program that requests Web pages and other associated applications over the Internet and that can display these files using the right format

Browser Log: Browser log is a computer file (program) running on the client's browser that lists all requests for individual files and ads.

Cache: Cache is a storage area on the user computer's hard disk where recently viewed Web pages are stored.

Client: A client refers to a computer that requests and receives data and services from servers on a computer network. Computer users work with clients to access information on the Internet and World Wide Web.

Cookies: A collection of information, usually including a username and the current date and time, stored on the local computer of a person using the World Wide Web, used chiefly by Web sites to identify users who have previously registered or visited the site (*http://dictionary.reference.com*).

HTML: HyperText Markup Language is the language in which most pages on the World Wide Web are written. These pages can be read using a browser.

Java: Java is an object-oriented language that is widely used for Internet or Web-based applications. It was designed specifically for distributed environments.

Javascript: Javascript is Netscape's simple, cross-platform, World Wide Web scripting language. Javascript runs in only three environments – as a server-side scripting language, as an embedded language in server-parsed HTML, and as an embedded language run in Web browsers, where it is the most important part of DHTML (*http://dictionary.reference.com*).

Proxy Server (or Web Cache): This is a server that lies in between the organizational network and the Internet. It has its own disk storage and stores copies of recently requested objects (Kurose & Ross, 2000).

Web Server: A Web server is a computer that is addressable by a URL and that houses objects. Objects include Web pages (HTML files), JPEG images, and other applications or programs.

Web Server Log: Also called an Access log, this is a list of all requests for individual files and ads that users (or clients) have made from a Web site (Whatis.com 2000).

Scenarios for Web-Enhanced Learning

Jane E. Klobas
Bocconi University, Italy and the University of Western Australia, Australia

Stefano Renzi
Bocconi University, Italy

INTRODUCTION

After several years of Internet use in traditional universities, some patterns have emerged, both in the nature of use, and in understanding the conditions associated with successful adoption and application of Web-enhanced learning (WEL). This article summarizes, in the form of nine scenarios, the ways in which the Internet is being used to enhance learning in traditional universities. It also discusses the changes needed if universities are to benefit more widely from WEL.

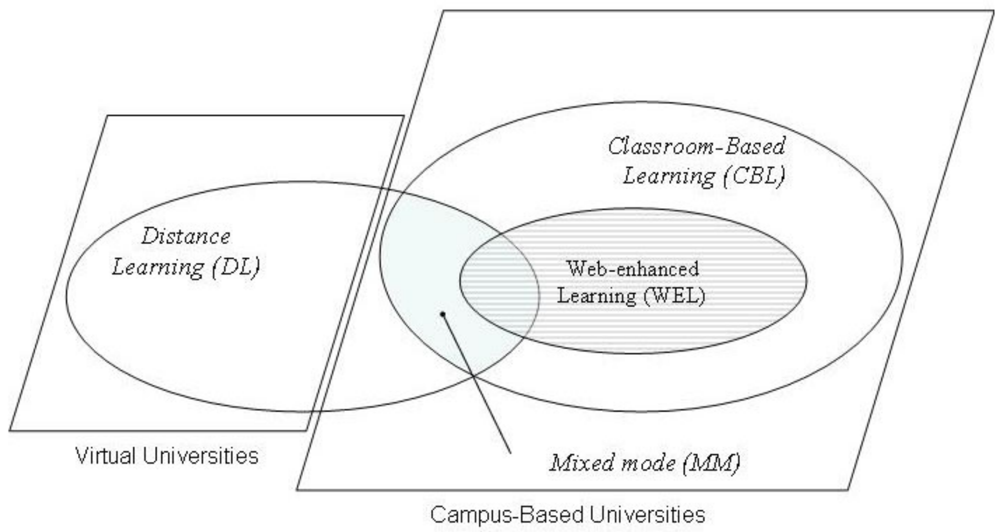
BACKGROUND

The World Wide Web is used by universities to make courses available to students who are distant from campus (*distance learning*, DL) and to enhance learning by students who attend courses on campus (*Web-enhanced learning*, WEL). Universities may be classified on the basis of the modes of learning that they offer. *Virtual universities* offer access to courses by DL only. Traditional, or *campus-based universities*, offer courses that are based on formal lessons held in classrooms or laboratories (*classroom-based learning*, CBL), but may also offer courses by DL or *flexible learning* (FL), which is a combination of DL and CBL.

WEL is the use of the Web to enhance CBL in traditional universities. WEL provides students studying in the classroom with access to electronic resources and learning activities that would not be available to them in traditional classroom-based study. Simple forms of WEL provide access to the Web from within the classroom, using the Web as a platform for real-time demonstration or as a digital library. More sophisticated forms of WEL blend activities in the classroom with Web-enabled learning activities that promote collaborative learning among students, even when they are distant from the classroom.

Figure 1 illustrates the relationship between the modes of learning offered by universities. WEL is represented as that portion of CBL that uses the Web to enhance learning. When it is used to blend in-classroom and out-of-

Figure 1. Relationship between Web-enhanced learning (WEL) and other modes



classroom activities, WEL shares the characteristics of DL and FL.

WEL differs from flexible learning in that the focus of the lesson remains the traditional classroom. With FL, classroom-based learning is mixed with learning at a distance. In the most common form of FL, *distributed learning* (also known as *blended learning* or *mixed mode learning*), students participate in formal lessons both in the classroom and at a distance, according to a schedule prepared by the instructor. Some flexible learning may be enhanced by use of the Web, for example, to provide discussion forums in which students studying at a distance and in the classroom may participate together, but use of the Web is not necessary for flexible learning.

This article is concerned with integration of online learning and classroom-based learning to achieve effective and manageable WEL for campus-based students. The focus is on change across a university system rather than in an individual classroom. We argue that WEL adds most value when it is used to enable new forms of learning, and in particular, online collaborative learning by students working at a distance from the classroom, as well as within it (Rudestam & Schoenholtz-Read, 2002). This value can only be obtained through attention at the institutional level to the organizational transformation required to implement, support, and sustain WEL (Bates, 2000).

WEL SCENARIOS

Nine distinct scenarios for use of WEL can be identified (Table 1, based on Klobas & Renzi, 2003). They can be divided into four groups: *information provision* scenarios, in which the Web is used to provide information to students and others outside the classroom; *classroom*

Table 1. A hierarchy of WEL use scenarios

Scenario	Label	Use
INFORMATION PROVISION SCENARIOS		
1	Catalog	Provision of static, and primarily logistic, information about the course.
2	Notice Board	Distribution of course materials in electronic form.
3	Class Resource	Provision of additional materials and references in response to student and teacher experience in the course as it progresses.
CLASSROOM RESOURCE SCENARIOS		
4	Classroom Resource	Use of the Web for demonstration or as a digital library during classroom sessions.
5	Streaming Video	Broadcast of classroom sessions.
INTERACTIVE LEARNING SCENARIOS		
6	Virtual Classroom	Synchronous interactive classroom sessions that include video and voice communication among instructors and students.
7	Interactive Web	An interactive environment outside the classroom.
8	CSCL	Computer-Supported Collaborative Learning.
EXPERIMENTAL SCENARIO		
9	Experimental	An experimental environment for innovative use of the Web.

resource scenarios, in which the Web is used to extend the classroom, either by providing access to resources in the classroom or by enabling lessons to be broadcast outside the classroom; *interactive learning* scenarios, which range from interactive virtual classrooms to the use of the Web to support collaborative learning among students working at a distance; and an *experimental* scenario, in which the Web is used to experiment with technology and pedagogy in ways not envisaged by the preceding scenarios. Any or all of the scenarios may be used alone or in combination, in a single course or by a whole university.

Information Provision

The first group of scenarios (1 to 3) represent incremental changes to traditional classroom-based learning. In these scenarios, the Web is used as an information delivery mechanism that provides students with some flexibility in the time and place with which they access some of the information required for the course. The scenarios range from simple publication of course catalog information to use of streaming video to permit students to 'attend' classes outside the classroom. The information and communications technology (ICT) infrastructure, training, and skills required for successful use of each scenario is quite simple.

The simplest and most common of WEL scenarios consists of provision of basic *catalog* information about a course: course description, list of textbooks, name of teacher(s), class schedule, allocated classroom(s), and examination procedures. Most university Web sites contain at least a subset of this information. This is a simple scenario to manage. The university needs a Web server and the staff to maintain it. The information posted in such catalogs is often pooled or available from a single source. Because it is static and seldom needs to be updated more than once a semester, the effort involved in maintaining currency is quite low.

In Scenario 2, *notice board*, teachers use the Web to distribute course materials in electronic form. Such material might include: educational material used by teachers in the classroom (slides, case studies, newspaper articles, site URLs related to the course content), past exam scripts and solutions, and official University communication. The content may be made available all at once before the course begins in the online equivalent of a course book, or from time to time during the life of the course (for example, lesson slides may be put online after each lesson).

Use of the Web in Scenario 3, *class resource*, is more dynamic than in Scenario 2. The teacher selects new material to add to the course Web site during the course, in response to questions asked, interest expressed, and other experiences of how the students are responding to the course as it is delivered. In courses with multiple classes (sections), each class may share the material available on the common course notice board, but may have quite different additional resources.

Effective adoption of Scenario 2 and Scenario 3 requires more extensive ICT infrastructure than Scenario 1 to permit students to access the course Web sites on a regular basis. At Scenario 2, some universities make staff available to load materials to the site on behalf of the teacher, but at Scenario 3, the teachers need the IT and the training to be able to load their own materials. This level therefore marks a significant shift in the resources required to support WEL. At the same time, it marks a significant shift in the value added by WEL; at this level, WEL provides the opportunity to quickly and easily provide students with access to current material of direct relevance to their learning and experience in the course as it unfolds.

The Web as a Classroom Resource

In Scenario 4, *classroom resource*, the teacher may use the Web to access reference materials, presentations, and demonstrations from sources outside the classroom. In this scenario, the Web provides convenient access to resources that might previously have been drawn from other sources or accessed in other ways by the teacher. While relatively simple for the teacher, this scenario requires provision of secure access to appropriate IT and Internet infrastructure from within the classroom.

Scenario 5, *streaming video*, requires more substantial investment in technology, including quality recording equipment in classrooms and the staff to operate and maintain it, high-end servers, and high-speed networks to all locations where the video may be watched. For effective use, teachers need to learn how to structure and present visual aids that will be suitable both in the classroom and for presentation by video. The primary uses of streaming video, to allow students the option of 'attending classes' from outside the classroom and to review the teacher's presentation of different material (Creighton & Buchanan, 2001), represent only an incremental change in the nature of education.

Learning Through Interaction

Use of the Web is more dynamic in the interactive learning scenarios, which involve interaction between the teacher and students, and among the students themselves.

In the WEL *virtual classroom* scenario (Scenario 6), the Web is used to transmit complete classroom lessons using synchronous video, voice, whiteboard, and other utilities. Teachers and students from different locations may share lesson components, ask questions, respond, and interact with one another in a variety of ways. For campus-based students, a virtual classroom provides the opportunity to share classroom experiences with teachers and students in classrooms located on other campuses (Hiltz & Wellman, 1997). Universities considering this option for WEL should weigh the considerable costs of investment in ICT infrastructure, training, and course redesign against the return to students.

In Scenario 7, *interactive Web*, the interactions are somewhat passive, based mainly on the use of course forums, resource contributions, self-evaluation tests, delivery of assignments, and secure online exams. This is the most common application of online learning platforms. Teachers require considerable training and support to adopt this scenario effectively. Students also require preparation, both in use of the functions of the technology, and in how to use the provided functions to improve the quality of their course experience and learning.

A more complex interactive scenario is *CSCL* (computer-supported collaborative learning, Scenario 8), an environment where at least the online component of teaching and learning is based primarily on interactions among students working in groups. This scenario includes collaborative group learning activities that go beyond those possible with simple course forums. Such activities may include group projects that involve sharing materials or preparation of joint documents. This scenario offers greater potential for improving the quality of learning at the university than any of the preceding scenarios.

Indeed, the power of WEL to change the quality of education is based on its potential to make collaborative learning possible. WEL makes a difference when it is used to enable students to learn collaboratively (Friedlander, 2002; Klobas & Renzi, 2003; Lammintakanen & Rissanen, 2003; Rudestam & Schoenholtz-Read, 2002). Students, themselves, describe the value of participation in learning communities of peers (Hamilton & Zimmerman, 2002), while educators claim that participation in collaborative learning not only results in better quality learning of course subject matter, but also in changes in the world view of the students and their capacity for lifelong learning and contribution to society (Klobas, Renzi, Francescato & Renzi, 2002; Rudestam & Schoenholtz-Read, 2002). Furthermore, collaborative learning that makes a difference does not need expensive technologies. CSCL does not require the investment in ICT infrastructure of Scenario 6, and can be implemented with simple asynchronous conferencing software (Hazemi & Hailes, 2002; Hiltz & Turoff, 2002).

Experimental Scenario

The final scenario, *experimental* (Scenario 9), provides an environment for teachers to experiment with new applications of the Web in the classroom, new Web-based technologies, new educational strategies, and the interaction between these innovations. While current thinking focuses on CSCL as the most advanced form of WEL, the existence of an experimental scenario reminds us to be open to further changes in learning theory and technology. It also reminds us of the need to evaluate new approaches to learning associated with WEL.

FUTURE TRENDS

Universities across the globe, in developed and developing countries, have been quick to adopt technologies to support WEL, including the infrastructure to support widespread use of the Web as an information resource, and university-wide platforms to support online learning. But this rapid adoption of technology has had relatively little impact on the education of campus-based students (Middlehurst, 2003, as cited in Collis & Van der Wende, 2002; Observatory of Borderless Education, 2002). Those changes that have occurred have been incremental rather than transformational. The Web is most frequently used to provide access to resources—as a substitute for, or complement to, notice boards, distribution of handouts, and use of the library—rather than to provide access to new forms of learning. Thus, the Web is being used to automate rather than to transform university education. Few attempts to go beyond this simple automation have been successful (Pollock & Cornford, 2000). Those universities that are making greater use of the Web are not distinguished from the others by virtue of their information technology infrastructure, but in terms of their focus on students, markets, and policy. Those looking to 'stretch the mould' of the future university emphasize flexibility in the location of learning, and have policies in place for quality, future markets, and costs and efficiency (Collis & Van der Wende, 2002).

Radical changes in our approach to university education are therefore needed if universities are to benefit from WEL. Bates (2000) claims:

"If universities and colleges are successfully to adopt the use of technologies for teaching and learning, much more than minor adjustments in current practice will be required. Indeed, the effective use of technology requires a revolution in thinking about teaching and learning." (p. v)

That revolution, according to Rudestam and Schoenholtz-Read (2002), "demands a reexamination of our core beliefs about pedagogy and how students learn" (p. 4), based on theories of constructivism and collaborative learning (Leidner & Jarvenpaa, 1995). Change on this scale requires vision and leadership, as well as appropriate resources.

Bates (2000) calls for "fundamental change in the way our higher education institutions are organized and managed" (p. 5), because "history dictates that the introduction of new technology is usually accompanied by major changes in the organization of work" (p. 1). This requires leadership, vision, policies, planning, and evaluation that emphasize the educational goals of WEL rather than just its technological characteristics (Bates, 2000; Collis & Van der Wende, 2002; Friedlander, 2002; Klobas & Renzi, 2003; Pollock & Cornford, 2000; Surry, 2002).

While they have put the necessary ICT infrastructure in place, universities have paid relatively little attention to development of the human resources needed for successful adoption of WEL (Collis & Van der Wende, 2002). Financial investment in training and skill development for teachers, students, and technical staff is required. It is not surprising, given arguments for attention to pedagogy rather than technology, that successful adoption of WEL is associated with training to develop teachers' ability to design courses that use WEL to improve pedagogy rather than training that emphasizes the features of specific software (Klobas & Renzi, 2003).

Human resource issues associated with optimization and management of the ICT infrastructure also need to be addressed. This requires attention to engagement with the university's partners in the supply and maintenance of the ICT infrastructure for WEL (Pollock & Cornford, 2000). The ICT infrastructure for WEL involves several layers of technology (internal and external networks, servers, and applications), and successful WEL requires skillful coordination of suppliers (Klobas & Renzi, 2003).

Changes in the reward systems for university staff are necessary to support the changes in work demanded by WEL (Bates, 2000; Collis & Van der Wende, 2002; Klobas & Renzi, 2000, 2003; Surry, 2002). Such changes in reward systems require short-term financial investment, but in the longer term are associated with changes in the nature and structure of work in the university.

CONCLUSION

WEL provides traditional universities with opportunities to enhance the quality of the education they provide to students on campus. While most current approaches to WEL involve incremental changes to classroom teaching, greater value is obtained through application of WEL to improve opportunities for collaborative learning among students. Success therefore requires attention—at the most senior levels—to educational values, financial and human resources, and transformation of educational processes and organizational structure, as well as to technology. WEL will become a natural part of the educational model of those universities with the management commitment and skill to implement and sustain the transformation required.

REFERENCES

Bates, A.W.T. (2000). *Managing technological change: Strategies for college and university leaders.* San Francisco: Jossey Bass.

Bento, R. & Schuster, C. (2003). Participation: The online challenge. In A.K. Aggarwal (Ed.), *Web-based education: Learning from experience* (pp. 156-164). Hershey, PA: Information Science Publishing.

Brown, J.S., Collins, A. & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-42.

Clulow, V. & Brace-Govan, J. (2003). Web-based learning: Experience-based research. In A.K. Aggarwal (Ed.), *Web-based education: Learning from experience* (pp. 49-70). Hershey, PA: Information Science Publishing.

Collis, B. & van der Wende, M. (2002). *Models of technology and change in higher education: An international comparative survey on the current and future use of ICT in higher education* (Report). Enschede, The Netherlands: University of Twente, Center for Higher Education Policy Studies.

Creighton, J.V. & Buchanan, P. (2001). Toward the e-campus: Using the Internet to strengthen, rather than replace, the campus experience. *EduCause Review, 36*(2), 12-13.

Friedlander, L. (2002, May 8-11). Next-generation distant learning. In F. Fluckiger, C. Jutz, P. Schulz & L. Cantoni (Eds.), *Proceedings of the 4th International Conference on New Educational Environments* (pp. 3-6), Lugano, Switzerland. Lugano: University of Applied Sciences Southern Switzerland and University of Southern Switzerland, Berne: net4net.

Hamilton, S. & Zimmerman, J. (2002). Breaking through zero-sum academics. In K.E. Rudestam & J. Schoenholtz-Read (Eds.), *Handbook of online learning: Innovations in higher education and corporate training* (pp. 257-276). Thousand Oaks, CA: Sage Publications.

Hazemi, R. & Hailes, S. (2002). Introduction. In R. Hazemi & S. Hailes (Eds.), *The digital university: Building a learning community*. London: Springer-Verlag.

Hiltz, S.R. & Turoff, M. (2002). What makes learning networks effective? *Communications of the ACM*, *45*(4), 56-59.

Hiltz, S.R. & Wellman, B. (1997). Asynchronous learning networks as a virtual classroom. *Communications of the ACM*, *40*(9), 44-49.

Klobas, J.E. & Renzi, S. (2000). Selecting software and services for Web-based teaching and learning. In A.K. Aggarwal (Ed.), *Web-based learning & teaching technologies: Opportunities and challenges* (pp. 43-59). Hershey, PA: Idea Group Publishing.

Klobas, J.E. & Renzi, S. (2003). Integrating online educational activities in traditional courses: University-wide lessons after three years. In A.K. Aggarwal (Ed.), *Web-based education: Learning from experience* (pp. 415-439). Hershey, PA: Information Science Publishing.

Klobas, J.E. Renzi, S., Francescato, D. & Renzi, P. (2002). Meta-risposte all'apprendimento online (Meta-response to online learning). *Ricerche di Psicologia, 25*(1), 239-259.

Lammintakanen, J. & Rissanen, S. (2003). An evaluation of Web-based education at a Finnish university. In A.K. Aggarwal (Ed.), *Web-based education: Learning from experience* (pp. 440-453). Hershey, PA: Information Science Publishing.

Leidner, D.E. & Jarvenpaa, S.L. (1995). The use of information technology to enhance management school education: A theoretical view. *MIS Quarterly*, *19*(3), 265-291.

Middlehurst, R. (2003). Competition, collaboration and ICT: Challenges and choices for higher education institutions. In M. van der Wende & M. van der Ven (Eds.), *The use of ICT in higher education: A mirror of Europe.* Utrecht: Lemma.

Observatory of Borderless Education. (2002, August). Online learning in Commonwealth universities: Selected data from the 2002 Observatory Survy, Part 1. *Briefing Note,* (7). Retrieved September 12, 2004 from *http://www.obhe.ac.uk/products/briefings/publicaccesspdfs/OnlineLearningCUpartone.pdf*

Palloff, R.M. & Pratt, K. (2002). Beyond the looking glass: What faculty and students need to be successful online. In K.E. Rudestam & J. Schoenholtz-Read (Eds.), *Handbook of online learning: innovations in higher education and corporate training* (pp. 171-184). Thousand Oaks, CA: Sage Publications.

Pollock, N. & Cornford, J. (2000). Theory and practice of the virtual university. *Ariadne,* (24). Retrieved September 12, 2004 from *web.ariadne.ac.uk/issue24/virtual-universities*

Rudestam, K.E. & Schoenholtz-Read, J. (2002). The coming of age of adult online education. In K.E. Rudestam & J. Schoenholtz-Read (Eds.), *Handbook of online learning: Innovations in higher education and corporate training* (pp. 3-28). Thousand Oaks, CA: Sage Publications.

Sauter, V.L. (2003). Web design studio: A preliminary experiment in facilitating faculty use of the Web. In A.K. Aggarwal (Ed.), *Web-based education: Learning from experience* (pp. 131-154). Hershey, PA: Information Science Publishing.

Shapiro, J.J. & Hughes, S.K. (2002). The case of the inflammatory e-mail: Building culture and community in online academic environments. In K.E. Rudestam & J. Schoenholtz-Read (Eds.), *Handbook of online learning: Innovations in higher education and corporate training* (pp. 91-124). Thousand Oaks, CA: Sage Publications.

Surry, D.W. (2002, April). A model for integrating instructional technology into higher education. *Proceedings of the Annual Meeting of the American Educational Research Association,* New Orleans, LA. AERA.

KEY TERMS

Blended Learning: *See Mixed Mode Learning*.

Collaborative Learning: Learning that occurs through the exchange of knowledge among learners. Collaborative learning is a form of social learning.

Computer-Supported Collaborative Learning (CSCL): Collaborative learning that occurs via the medium of computer-based communication networks such as the Internet.

Distributed Learning: *See Mixed Mode Learning*.

Flexible Learning: Systems in which students may choose to complete some of their learning on campus and some of their learning off campus.

Mixed Mode Learning: Study that combines traditional face-to-face learning with learning at a distance in a structured program. The Web may be used to enhance learning during study by one or both of these modes. Mixed Mode is also known as Blended Learning and Distributed Learning.

Online Learning Activities: Learning activities in which students interact with resources, or other students, or both, using the capabilities of the Internet or other computer-based communication networks.

Social Learning: Learning through social interaction with other people.

Web-Enhanced Learning (WEL): Use of the World Wide Web (Web) to provide students studying in the classroom with access to electronic resources and learning activities that would not be available to them in traditional classroom-based study. The simplest forms of WEL provides information about a course on the Web and access to the Web from within the classroom. More sophisticated forms of WEL blend activities in the classroom with Web-enabled online learning activities, which promote collaborative learning among students even when they are distant from the classroom.

Security and Trust of Online Auction Systems

Pouwan Lei
University of Sussex, UK

Chris Chatwin
University of Sussex, UK

Rupert Young
University of Sussex, UK

INTRODUCTION

An auction is a market with an explicit set of rules determining resource allocation and prices on the basis of bids from market participants (McAfee & McMillan, 1987). Generally speaking, an auction is the standard means for performing an aggregation of supply and demand in the marketplace to effectively establish a price for a product or service. It establishes prices according to participants' bids for buying and selling commodities, and the commodities are sold to the highest bidder. Simply stated, an auction is a method for allocating scarce goods, a method that is based upon competition between the participants. It is the purest of markets: a seller wishes to obtain as much money as possible for the commodity offered, and a buyer wants to pay as little as necessary for the same commodity. Traditionally, there are three protagonists in the auction: sellers, buyers, and auctioneers. An auction offers the advantage of simplicity in determining market-based prices. It is efficient in the sense that an auction usually ensures that resources accrue to those who value them most highly and ensures also that sellers receive the collective assessment of the value. Indeed, auctions are conducted in accordance with formal rules for governing market access, trade interaction, price determination, and trade generation (Friedman, 1993). In the case of a traditional physical auction, a seller will choose an auction house based on the service: the form of licensing, the availability of suitable insurance, suitable descriptions and access to the commodities, payment terms, and security of goods before and during the auction process. The buyer or seller needs to come to the market or sends his/her representative.

Online auction systems provide immediate access advantages compared with their physical auction systems counterpart. Participants may join an online auction system, effectively placing bids using a computer on an anywhere-anytime basis. The access is not only limited to desktop computers, but also handheld devices such as mobile phones. In online auctions, transactions take place based on information (product descriptions), and the products move from seller directly to buyers only after online transactions are completed. They facilitate buyers and sellers in: meeting, the listing of items for sale independent of physical location, exchanging information, interacting with each other, and ultimately completing transactions. They offer significant convenience, allowing trading at all hours, and provides continually updated information; and they allow buyers and sellers to trade directly, by bypassing traditional intermediaries and lowering costs for both parties. Online auctions are global in reach, offering buyers a significantly broader selection of goods to purchase, and provide sellers with the opportunity to sell their goods efficiently to a broader base of buyers. Factors that make online auctions attractive may also present disadvantages. Many online auctions simply list the items for sale. No attempt is made to verify and check that the merchandise actually exists or that the description is accurate. Transaction trustworthiness and security are the two most significant problems.

EXISTING PROBLEMS

Online auctions have become very popular. One of the most successful online auctions, eBay (www.ebay.com) purports that the number of active users worldwide has increased from 27.7 million in 2002 to 41.2 million in 2003, and the number of listings was about 292 million in 2003 (eBay, 2004). Nielsen/Netratings, the global standard for Internet audience measurement and analysis, also reports that eBay was one of the top five sites in Germany and the United Kingdom in February 2004 (Nielsen/Netratings, 2004). The research firm, the Aberdeen Group, found that 94% of net market transactions were through auctions, with just 6% through catalog sales (Pritchard, 2002). Most auctions are open to the public. Whatever you want, you can find. Given the rapid success of the virtual market, no de facto standards exist as to the bidding rules and policies governing the online auction business. Although online auctions have been developing for many years, there are still two major problems: trustworthy transaction and security. Regarding the first problem, trustworthy transactions, many auction sites describe themselves merely as meeting places for buyers and sellers. They simply allow sellers to list merchandise offered for trade and do not verify that the merchandise actually exists or is accurately described. They only use an email address to identify the traders—buyers and sellers. After the auction is over, it is the seller's responsibility to deal directly with the buyer concerning payment and delivery. The auction companies do not hold any responsibility for the transaction. Auction fraud is therefore an increasingly difficult problem in the virtual market. The common types of auction fraud are as follows (National Consumer League, 2001):

i) Failure to deliver: Buyers pay for an item, which is never received.
ii) Misrepresentation: Items received do not match up to the original description.
iii) Shill bidding: A seller, or an associate, places a fake bid intended to drive up prices.
iv) Selling black-market goods: The goods are typically delivered without authentic merchandize, warranty, or instructions.

Among the complaints that the Federal Trade Commission (FTC) received about auction fraud, the two most frequent are 'failure to deliver' and 'misrepresentation'. However, in the last few years there is a new trend of increased 'shill bidding'. These problems effectively prevent some Internet users from participating in Internet auctions. According to FTC's Auction Fraud Report, Internet auction fraud entails 64% of all Internet fraud that is reported (Enos, 2001). The total dollar loss for all complaints exceeded US$12.3 million. Internet auction fraud has become a significant problem.

The second problem, security, is naturally a big concern for any business on the Internet. Since data is being transported over public networks, this makes it possible for third parties to snoop and derive critical information. Security and safety is an important topic in conducting business on the Internet; online auctions are no exception. During the auction, buyers and sellers have to submit their personal information to the system, as well as provide electronic payment for their goods. Hundreds and perhaps thousands of credit card numbers, home addresses, and phone numbers were exposed for months through a security hole on many Internet auction sites. Few auction sites provide security features such as SSL and VeriSign security. In the survey of protections on smaller auction sites, there is less than 20% implementing security technology (Selis, Ramasastry & Wright, 2001). Furthermore, most online auctions do not enforce strong authentication, relying instead on a user ID and password, or may be an email account, to establish the validity of a client. Once this minimal information is supplied, people are free to enter into the online auction system and participate in bidding. Moreover, no minimally acceptable standard exists for ensuring that auctioneers protect users against the loss of personal information by the auctioneer. There are no established minimum-security standards or licensing bodies to protect the privacy rights of customers. People are risking their personal information. Ensuring security and trust in electronic communication is a principle requirement for achieving the trust necessary in gaining widespread acceptance of Internet auction systems as a medium for commerce.

SECURITY & CONFIDENTIALITY

The challenge in security is to build an online auction system (OAS) with safe communication and collaboration between legitimate users. The following sums up the fundamental security needs for OAS:

i) The need to identify and authenticate legitimate users, thus identifying and granting access to bid information, content, and supporting services.
ii) Provision of a security system with fine-grained access control that will allow, on the one hand, legitimate users access to resources, while on the other, protecting sensitive information from hackers and unauthorized users (i.e., all other users).
iii) OAS should ensure that private, tamperproof communication channels for auction participants exist. Hence, processing of their transaction is secure.
iv) OAS should provide auditing and logging facilities to track site security and misuse.
v) OAS should provide secure data transactions from sellers to OAS and from OAS to buyers.
vi) Database system security is another consideration in OAS. In order to make sure that no unauthorized or authorized user can access any data in the database system, OAS should clearly identify data held, conditions for release of information, and the duration for which information is held.

Authentication is often considered the single most important technology for OAS. It should be computationally intractable for a person to pretend to be someone else when logging into an OAS. It should be virtually impossible for a third party to alter e-mail addresses, digital signatures, or the content of any document without detection. In addition, it should be equally

Figure 1. E-payment systems

Proprietary Payment Scheme	Third-Party Payment System: P2P	Escrow Service
Buyers pay money to payment company	Seller signs up for P2P Web site	Auction closes
Seller send out goods to buyer	Seller sends money request to buyer	Buyers pays I-escrow
Payment company sends payment to seller	P2P Web site notifies buyer by e-mail	Sellers ships goods
	Buyer signs up	Buyer approves goods
	Buyer sends money to seller	I-escrow pays sellers
	P2P Web site notifies seller by e-mail	
	Seller ships	

difficult for someone to mimic the Internet address of a computer when connecting to the OAS. Various authentication technologies are available for determining and validating the authenticity of users, network nodes, files, and messages; several levels of authentication must be considered. Validating the identity of users during the login process to the system is supported by encryption technologies to support authentication. Encryption is the fundamental technology that protects information as it travels over the Internet. Four properties are used to describe the majority of encryption functions of interest to OAS. These are: *confidentiality, authentication, integrity*, and *non-repudiation*. A cryptosystem comes with two procedures, one for encryption and one for decryption (Garfinkel, 1995). Cryptographic systems include Secure Socket Layer (SSL), Public Key Infrastructure (PKI), and Secure Electronic Transaction (SET).

A consumer-to-consumer (C2C) online auction system is designed for sellers and buyers; the online auction site acts as an intermediary. Sellers and buyers will interact with each other for their payment transaction. In order to prevent trustworthy transaction problems, OAS should provide a mechanism for trustworthiness such that the identity of the parties is established/verified. Almost every online business makes use of registration to identify and classify their customers. However, the difficulty lies in identifying information, which can be readily verified, which is also unique, difficult to fabricate, and does not reduce the potential customer base. Most systems therefore are relatively weak at ensuring the validity of information offered to identify registrants. The limits for ensuring trustworthy online registration are principally set by the availability of online verification services. The OAS may be able to do data field type checking (validate post codes or names). The one verifiable piece of information under current systems might be the customer e-mail address. In practice the only sure way of ensuring customer trustworthiness might be to limit the customer base to a set of certified users.

However, it is not the case in business-to-business (B2B). B2B marketplaces involve the transactions from one business to another via the Internet. No customer is involved in the transaction. A strict and legal entity is required between businesses. Businesses can access sites for procurement and services to obtain multiple bids, issue purchase orders, make payments, and so forth. The bidders are now the purchasing managers of other businesses. All sellers are registered and cleared as a certified business or commercial identity. The problems with B2B mainly arise if the transaction involves multiple countries. 'Cross-border' transactions involve taxes, duties, customs procedures, and satisfaction of legal requirements. For instance, online payments are becoming popular in Europe. However different countries have had different rules on an electronic signature and electronic invoices. At present, the European Committee for Standardization is resolving many issues related to online payments among the members (George, 2004). In Asia, the legal system is weak, inefficient, and corruptive in some countries. Hence it will become difficult to conduct commerce with companies in other countries if the country has no such legislation. Supplier-buyer enablement (B2B) is easy to support in Singapore and Hong Kong, but it is still in its infancy in the Philippines, Indonesia, India, and China (Choy, 2000). The legal framework will need a much longer time to become established in these countries.

Banking plays a critical role in commerce and therefore auction systems, as it typically represents the authority responsible for policing the final settlement of payment. In e-commerce as a whole, however, banks often lag behind the rate of technological change in other sectors of commerce. Buyers may have several payment options, including credit card, debit card, personal check, cashier's check, money order, cash on delivery, and escrow services. Credit cards offer buyers the most protection, including the right to seek a refund from the credit card issuer if the product is not delivered or if the product received is not the product ordered. Many sellers in C2C auctions do not accept credit cards. There are several reasons for this. From the seller's perspective, there will be a charge on them, and the average value of most purchases was US$100 or less (National Consumer League, 2001). The use of a credit card for payment will add cost to the sellers. From the buyer's perspective, it is very dangerous to disclose the credit card information to a person they have never met before. They may use your credit card information for mischief. Payment by check, cashier's check, or money order directly to the seller accounts for 69% of payment methods. However, those methods have no protection for the buyers. Money can be electronically transferred between buyers and sellers in a fast and low-cost way. The e-payment methods are shown in Figure 1. A proprietary payment system is a system in which the buyer pays a payment company rather than the seller, and the payments company pays the sellers. Escrow services allow buyers to deposit money in trust with a company, which will not release the funds to a seller until certain conditions are met or verified. In person-to-person (P2P) payment, it is first necessary for the payer to register with a P2P Web site, giving the payment provider authorization to debit a personal bank or credit card account.

CONCLUSION

Except for some notable large auction systems, most small online auction systems do not implement any security technology, which is the foundation for trusted transactions. Should international legislation be drafted for law enforcement of Internet auctions? It may be likened to legislation for road safety: it is illegal for drivers and passengers to ride in a car without a seat belt. In other words, the online auction systems should only be operated with essential security features such as SSL and privacy policy. Credit cards give the best protection to the customers, however the risk is high as the buyer's information about the seller is limited to an e-mail address. For high-value transactions, bringing in the rules of the traditional auction house may be a trend to maintain the confidence of both buyers and sellers. Consumers have various ways to protect themselves from auction fraud. It is important to educate them about the choices of payment methods related to the degree of protection available. There is always a trade-off between cost and risk. Another major problem facing both C2C and B2B online auction systems is the legal framework under which they operate, since it is not limited to one nation, but is 'cross-border'. In C2C, a perpetrator of fraudulent transactions may be from another country. It may thus be difficult to take legal action against him/her. Meanwhile in B2B, besides the issues of taxation and currency exchange, there are difficult issues relating to legal authority. Who will arbitrate or convene legal hearings in B2B? Online auction systems play a dominant role in e-commerce; it is therefore an important channel for trading. In order to make it a secure and a trusted marketplace, there is an urgent requirement for international management and control.

REFERENCES

Choy, J. (2000). Asian e-marketplaces face challenges. *Asia Computer Weekly*, (December 11-17).

EBay. (2004). EBay Inc.'s fourth quarter and full year 2003 financial results. Retrieved March 22, 2004, from *investor.ebay.com*

Enos, L. (2001). Online auctions top FBI Net fraud list. *E-Commerce Times*, (March 7). Retrieved October 27, 2003, from *www.ecommercetimes.com/perl/story/7986.html*

Friedman, D. (1993). The double auction market institution: A survey. In D. Friedman and J. Rust (Eds.), *The double auction market institutions, theories and evidence* (pp. 3-26). Santa Fe Institute Studies in the Science of Complexity. Addison-Wesley.

Garfinkel, S. (1995). *PGP: Pretty good privacy*. O'Reilly & Associates.

George, N. (2004). Cheque is not in the post. *Financial Times* (UK), (May 26).

McAfee, R.P. & McMillan, J. (1987). Auctions and bidding. *Journal of Economic Literature*, (June), 699-738.

National Consumer League. (2001, January 31). Online auction survey summary. Retrieved from *www.nclnet.org/onlineauctins/auctionsurvey2001.htm*

Nielsen/Netratings. (2004). Nielsen/Netratings' top rank. Retrieved March 22, 2204, from *nielsen-netratings.com*

Pritchard, S. (2002). Online auctions: Streamlined benefits for buyers and sellers. *Financial Times* (UK), (March 13).

Selis, P., Ramasastry, A. & Wright, C.S. (2001, April 17). Bidder beware: Toward a fraud-free marketplace—best practices for the online auction industry. Retrieved October 27, 2003, from the Washington State Attorney General's Web site, *www.atg.wa.gov/consumer/auctions/home.htm*

KEY TERMS

Digital Certificates: Used to authenticate both parties. Certificate authorities (CAs) must issue these certificates. These are trusted third parties that have carried out identity checks on their certificate holders and are prepared to accept a degree of liability for any losses due to fraud. CAs also issue the public and private keys.

Dutch Auction: A popular kind of auction at many sites, commonly used when a seller has a number of the same items to sell, for example, 10 posters. The auctioneer starts with a high asking price. The seller then gradually decreases the offer price, and the first person to bid is the winner.

English Auction: By far the most popular auction method. Bidding takes the form of an ascending price auction where a bid must be higher in price than an existing bid in order to win the auction.

Public Key Infrastructure (PKI): An Internet trust model based on public key cryptography (encryption is conducted with a dual key system: a public key known to everyone and a private key known only to the recipient of the message). PKI offers the advantages of authentication and non-repudiation, which SSL lacks.

Reserve Auction: In this case the seller sets a reserve price, or the lowest price the seller is willing to transact on.

Secure Electronic Transaction (SET): Despite SSL's popularity, MasterCard, Visa, and several other companies developed SET. Released in 1997, SET v1.0 established a standard specifically for handling electronic payments, describing field formats, message types, protocol handshaking, and encryption mechanisms. The key difference between SET and SSL is that SET has digital certificates for all involved parties as an integral part of its design.

Secure Socket Layer (SSL): SSL protocol provides secure links over the Internet between a Web browser and a server. SSL was developed by Netscape Communications in 1995 and is embedded in Web browsers. Its adoption has been widespread as it is relatively inexpensive.

Security Issues in Distributed Transaction Processing Systems

R. A. Haraty
Lebanese American University, Lebanon

INTRODUCTION

Transaction-processing systems (TPS) are becoming increasingly more available as commercial products. However, the approaches to the issues associated with using TPS in multilevel secure environments are still in the research stage. In this article, we address the issues of multilevel security in distributed transaction-processing systems. A distributed transaction-processing system (DTPS) is a collection of a finite number of centralized transaction-processing systems connected by a computer network. Each of these transaction-processing systems is controlled by a software layer and can be accessed both remotely and locally. Properties of a DTPS, such as data replication, may have a substantial effect on the security of the system. The security policies and integrity constraints adopted at each site may result in global security having inconsistent states. We address the issues of achieving a multilevel secure DTPS, and discuss the security constraints and data replication.

In this work, we address the issues of achieving a multilevel secure DTPSs system and discuss the security constraints and the replication of data items. The next section provides some background. Then, next, an overview of a distributed transaction-processing system is presented. In the fourth section, security-related issues are discussed. In the fifth section, a multilevel secure distributed transaction-processing system is presented. Then, in the next section, future trends are presented. The final section concludes the article.

BACKGROUND

Several commercial and military applications require a multilevel secure transaction-processing system (MLS/TPS). In an MLS/TPS, users are assigned classification levels that we denote by "clearances," and data items are assigned sensitivity levels. There are three interesting architectures that have been used to build MLS/TPSs from untrusted ones. These architectures are known as the integrity lock architecture, the kernelized architecture, and the data distribution architecture (Air Force Studies Board, 1983). While most of the techniques for TPS security are developed for traditional centralized TPSs, more TPS researchers are making substantial contributions to the development of a distributed TPS (Getta, 2003; Haraty, 1999; Haraty & Rahal, 2002; O'Connor & Gray, 1988).

A DTPS is a collection of a finite number of TPSs connected by a computer network (Ozsu & Valduriez, 1999). Each of these TPSs is controlled by a transaction management software layer and can be accessed both remotely and locally. A DTPS integrates information from the local TPS and presents remote users with transparent methods to use the total information in the system. An effective TPS system serves to maintain the ACIDity properties (i.e., atomicity, consistency, isolation, and durability) of transactions and must be superimposed on the preexisting local TPSs (Gray & Reuter, 1993).

One proposed architecture for MLS/TPS is the replicated architecture. This approach is being explored in several ongoing research efforts, including the Naval Research Laboratory Secure Information through replicated architecture (SINTRA) project (Thuraisingham, 1987). Data replication in DTPS has several implications for the security of the system. Replication allows data items in different local TPSs to be identified as logically belonging to the same entity. The security policies adopted by each site may result in global security having inconsistent states, because of the difference of local representation and management.

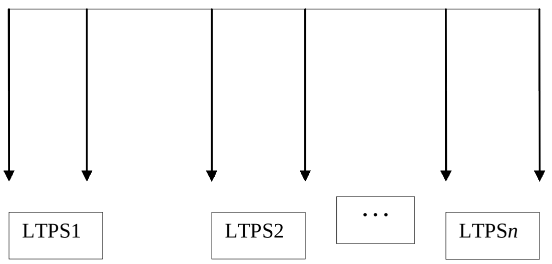
OVERVIEW OF DISTRIBUTED TRANSACTION-PROCESSING SYSTEMS

A DTPS consists of a set of preexisting local TPSs {LTPS*i* | 1 d•*i* d" *m*}, distributed among several interconnected sites. Each LTPS*i* is a software layer on a set of data items D*i*. Figure 1 depicts the architecture of a DTPS.

SECURITY ISSUES

Processes that execute on behalf of users are referred to as subjects. Objects, on the other hand, correspond to a

Figure 1. Distributed transaction-processing system



data item. Objects can be files, records, or even fields. In this section, we present the notion of object classification with emphasis on the problem of conflicting security constraints due to replication.

A security classification is a function that associates each subject and each object with a given level of security. Many classifications, such as the security lattice, exist (Denning, 1976). However, a well-known classification is four-value function (DOD paradigm) that classifies objects into unclassified (U), confidential (C), secret (S), and adopt top secret (TS). A simple policy that can be established using a classification function SL is as follows:

Subject X can access (read) Object Y iff SL (Y) $\leq$ SL(X)

A security constraint consists of a data specification and a security value. The data specification defines any subset of the TPS. The security values can be given by a classification function. Specific values are unclassified, confidential, secret, and top-secret. Thuraisingham (1987) defined two types of security constraints—internal constraints and external constraints:

1. Internal constraints classify the entire TPS as well as relations, attributes, and tuples within a relation. These constraints can be applied to data, as they are actually stored in the TPS.
2. External constraints classify relationships between data and the results obtained by applying operations on the stored data, such as sum, average, and count. Among these constraints are the functional constraints and the dynamic constraints.

These security constraints are subject to inconsistency and conflicting local security constraints. A good global security approach should reject inconsistent security constraints and inconsistent clearance of users. Examples of the inconsistencies encountered include:

- **Conflicting security constraints:** Such constraints classify the same facts into different categories.
- **Overlapped security constraints:** These constraints cover overlapped data domains.
- **Inconsistent security level of replicated data:** Cases where different copies of replicated data may belong to different security cases.
- **Access privileges of users to replicated data:** Instances where a user may have different access rights on replicated data at different sites.

Several solutions have been proposed to solve these inconsistencies and define a global security policy that respects the local ones (Pfleeger, 1989; Thuraisingham, 1987).

There are several ways to combine local policies. The optimal combination should give a policy that defines all component policies and is still secure.

MULTILEVEL SECURE DISTRIBUTED TRANSACTION-PROCESSING SYSTEMS

There are two strategies for building MLS/DTPS from DTPS. These strategies include data replication and per-level-based distribution. The scope of this article does not include the issues associated with network security; but, it is particularly important to have the various local TPSs. Instead, we will assume that interconnection between the various local TPSs is secure and focus attention on security that has to be provided due to replication and other properties specific to the TPS.

The data distribution approach physically replicates low-level data at all higher-level TPSs. The advantage of the replicated architecture is that is fairly secure (McDermott & Sandhu, 1991). No performance overhead is associated with multilevel queries, because they are locally executed. On the other hand, because data is replicated, there is overhead associated with broadcasting updates of lower-level data to higher-level TPSs in a correct and secure manner. This broadcasting mechanism is known as “data synchronization” (Air Force Studies Board, 1983).

In the per-level-based approach, data are physically stored in separate local TPSs according to sensitivity level. Early examples of this approach were presented by Hinke and Schaefer (1975). The advantage of this approach is that updating transactions does not produce inconsistencies. Performance overhead associated with multilevel queries is a major disadvantage.

Global Commitment in Secure Environment

An important aspect of a correct TPS is atomic commitment (Bernstein et al., 1987). Unfortunately, the local TPS in a MLS/DTPS system cannot support atomic commitment, so the two-phase commit (2PC) protocol (Bernstein et al., 1987) cannot be implemented. 2PC is known to introduce covert channels. In order to establish a covert channel, there must be two cooperating agents/subjects in the system and an encoding scheme. There are two main types of covert channels: covert storage channels and covert timing channels.

Covert storage channels disclose information from high to low subjects by manipulating a physical object that can or cannot be seen by the low subjects. For example, suppose there are two subjects of different security levels. Suppose also that these processes share a common resource—the available disk space. The secret subject creates a secret file that takes all of the available disk space to store the file. When the low subject attempts to create a file and store it onto the common disk, its request is denied. Through this denial, the high subject can signal information to the low subject. These signals are in terms of 0 and 1 bits that the low subject has to decode and turn into useful messages.

Covert timing channels can covertly send information by modulating observable delays of a common resource. This delay must be measured by low subjects cleanly; otherwise, the channel becomes noisy. For example, suppose we have two subjects again operating the low and high levels. The high subject can modulate the disk access time of the low subject by issuing numerous disk requests (thus transmitting a bit of 1) or zero disk requests (thus transmitting a zero). A system that is free from any type of covert channel is called covert channel secure.

Several distributed commitment protocols have been defined. A scheduler in MLS/DTPS that produces committable execution guarantees that a distributed transaction (a unit of work with execution sites: TPS1, TPS2, …, TPS*n*) becomes committed after it has been locally committed. The commitment of a distributed transaction means all of its subtransactions are committed. In this article, we follow the definition proposed by Bernstein et al. (1987):

If one subtransaction commits, then all other subtransactions will eventually commit.

We assume, in this article, that each subtransaction of a distributed transaction is designed to be executed in only one container. One can then say that a subtransaction has a security level.

FUTURE TRENDS

Future work will involve taking a closer look at MLS/DTPS and defining new and better ways of handling transaction management as well as query processing. Future work will also involve extending security issues to temporal and multimedia databases.

CONCLUSION

The security issues presented in this article highlight the intricacies required to architect a MLS/DTPS. We hope to address these issues further and to identify potential prototypes and engineering solutions that meet the requirements of MLS for DTPS.

REFERENCES

Air Force Studies Board, Committee on Multilevel Data Management. (1983). *Multilevel data management*. National Research Council.

Bernstein, P. A., Hadzilacos, V., & Goodman, N. (1987). *Concurrency control and recovery in database systems*. Reading, MA: Addison-Wesley.

Denning, D. (1976). Secure information flow in computer systems. Ph.D. dissertation. Purdue University.

Getta, J. R. (2003). Hybrid concurrency control in multilevel secure database systems. In *Proceedings of the IASTED International Conference—Applied Informatics*. Innsbruck, Austria.

Gray, J., & Reuter, A. (1993). Transaction processing: Concepts and techniques. San Francisco, CA: Morgan Kaufmann.

Haraty, R. A. (1999). A security policy manager for multilevel secure object oriented database management systems. In *Proceedings of the International Conference on Applied Modelling and Simulation*, Cairns, Queensland, Australia.

Haraty, R. A., & Rahal, I. (2002). A bit vectors algorithm for accelerating queries in multilevel secure databases. In *Proceedings of the International Conference on Computer Science, Software Engineering, Information Technology, e-Business, and Applications (CSITeA'02)*, Foz do Iguazu, Brazil.

Hinke, T., & Schaefer, M. (1975). Secure database management system, RADC-TR-75-266.

McDermott, J. P., & Sandhu, R. S. (1991). A single-level scheduler for the replicated architecture for multi-secure database. In *Proceedings of the Seventh Annual Computer Security Applications Conferences*.

O'Connor, J. P., & Gray, J. W. (1988). A distributed architecture for multilevel database security. In *Proceedings of the Security Conference*.

Ozsu, M. T., & Valduriez, P. (1999). *Principles of distributed database systems*. Upper Saddle River, NJ: Prentice Hall.

Pfleeger, C. P. (1989). *Security in computing*. Upper Saddle River, NJ: Prentice Hall.

Thuraisingham, M. B. (1987). Security of database systems. *Computer and Security*, *6*(6).

KEY TERMS

Covert Channel: This is a channel that is not meant to route information, but nevertheless does.

Multilevel Secure Transaction-Processing System: This is a system whereby database users are assigned classification levels, and data items are assigned sensitivity levels.

Security Lattice: This is a partial (or total) order of security classes, where there is a least upper bound that dominates all the other security classes and a greatest lower bound that is dominated by all security classes.

Subject: This corresponds to a user or, more correctly, to a process that is running on behalf of a user.

Two-Phase Commit (2PC): This is an atomic commitment protocol that behaves as follows: The coordinator asks the participants to vote on commitment; if any votes No, the coordinator informs all participants to Abort; if all participants voted Yes, then the coordinator informs all participants to Commit.

Self-Organizing Networked Learning Environments

Jon Dron
University of Brighton, UK

INTRODUCTION

The Internet has long been touted as an answer to the needs of adult learners, providing a wealth of resources and the means to communicate in many ways with many people. This promise rarely has been fulfilled, and often, when it is, by mimicking traditional instructor-led processes of education. As a large network, the Internet has characteristics that differentiate it from other learning environments. As Kelly (1998) puts it, "the sum of the value of a network increases as the square of the number of members." Because these interactions are mediated through computers and may be with many users at once, this is a notable underestimate.

Churchill said, "We shape our dwellings and afterwards our dwellings shape us" (Churchill, 1943, p.23). If this is true of buildings, then it is even more applicable of the fluid and ever-changing virtual environments made possible by the Internet. Our dwellings are no longer fixed, but may be molded by the people that inhabit them. This article discusses a range of approaches that makes use of this facility to provide environments that support groups of adult learners by adapting to their learning needs using nothing more than their interactions to provide structure and shape to their learning.

BACKGROUND

Jonathan Darby (2003) identifies three generations of networked learning environments used in adult education. First generation systems are direct analogues of traditional courses, simply translating existing structures and course materials. Like their traditionally delivered forebears, they are dependent on individual authors. Second generation systems tend to be team-built and designed for the medium from pedagogical first principles, but still within a traditional course-based format. Third generation systems break away from such course-led conventions and provide such things as just-in-time learning, guided paths through knowledge management systems, and personalized curricula. This chapter is concerned primarily with third generation environments.

Saba's interpretation of Moore's theory of transactional distance predicts that in an educational system, as structure increases, dialogue decreases and vice versa (Moore & Kearsley, 1996; Saba & Shearer, 1994). What is significant in differentiating learning experiences is not the *physical* distance between learners and teachers, but the *transactional* distance, measured by the degree of interaction between them. Highly structured educational activities have a high transactional distance, while those involving much discussion have a lower transactional distance. This chapter is concerned with the structured variety of educational activity.

In a traditional learning environment, the structure of the experience is provided by the teacher or instructional designer. One of the few justifications for the lecture form is that it provides a framework for understanding a topic, emphasizing the important points and ordering them in a manner that is intended to be beneficial to learners. However, learners will not benefit equally from any given structure, as different learners learn differently (Kolb, 1984). It would be better if learners could select appropriate approaches for their needs. Without a teacher, help with this might be provided by the opinions of other learners. However, eliciting those opinions, assessing their reliability or relevance, actually finding the resources in the first place, and, once found, fitting them into a structured learning experience is difficult. Several approaches to these problems are available, but first it is necessary to introduce a few concepts of self-organization.

Self-Organizing Processes

Self-organization of the kind we are discussing is an emergent process in which the interactions of the participants and the software lead to organization that does not arise as an intentional, programmed, or planned process, but which arises from the nature of the system itself. Such processes are very common in nature and human social systems. Two in particular are of interest here—*evolution* and *stigmergy*.

Based primarily on work following that of Darwin (1872), evolution is one of the most powerful self-organizing principles whereby a process of replication with variation combined with natural selection (survival of the fittest) leads to a finely balanced self-adjusting system. It is important to note that "fittest" does not mean "best" by any other measure than the ability to survive in a given environment.

Stigmergy, a form of indirect communication through signs left in the environment (Grasse, 1959), leads to self-organized behavior, examples of which include everything from ant trails and termite mounds to forest footpaths, money markets, and bank-runs. For example, ants wander randomly until they find food, after which they return to the nest leaving a trail of pheromones. Other ants are more likely to wander where such pheromone trails mark the route. When they, too, find food, they leave a trail of pheromones. The stronger the trail, the more other ants are drawn to it. This positive feedback loop continues until the food runs out, after which the trail slowly dies away.

A full discussion of the many factors that result in a successful self-organizing system is beyond the scope of this chapter. However, the following brief discussion provides a flavor of what is involved.

Self-organizing processes occur through local interactions. For systems to develop any sort of complexity, it is necessary for these interactions to occur at a number of scales. For instance, the interactions of bacteria in an ant's gut affect the ant; groups of ants can affect tree growth; tree growth can affect climate. Local interactions should form small clusters that in turn interact with each other, leading to ever increasing scales of self-organization. A hierarchical scale where the slow moving features of a system shape the faster changing leaf nodes is a common feature of self-organizing systems, from forests to cities (Brand, 1997). Parcellation is also an important feature of such systems (Calvin, 1997). As Darwin found in the Galapagos Islands, isolated populations tend to develop differently and more rapidly than their mainland counterparts. Any self-organizing system relies on interactions between more or less autonomous units. The precise level of interactivity varies, but it is interesting to note that for a system which teeters at the edge of chaos, neither too stable to change nor too changeable to develop, the average number of connections between interacting agents tends to stabilize around just over two (Kauffman, 1995). Systems must be capable of change, being in a more or less permanently unfinished state. Perfect systems cannot evolve (Shirky, 1996). Equally, systems in perpetual flux can never achieve the stability to achieve self-organization.

Some Examples of Self-Organised Learning in Practice

For many seekers of knowledge today, the starting point is often Google (http://www.google.com). Google's PageRank algorithms use as their basis a principle described by Kleinberg (1998) as Latent Human Annotation (LHA). The principle behind LHA is that most Web pages provide links to other sites if those sites are considered in some way "good." A simplified formulation of this is that the more backlinks (links pointing into a site) that point to a given site, the higher its approval rating. Combined with a content-based search for keywords, documents returned often have a high degree of relevance and reliability. This approach is self-organized, incorporating evolution (unlinked sites "die") and stigmergy (more-visited sites get more links pointing to them). It does not rely on a central controlling authority to provide decisions on a resource's usefulness or give a structure to the content that is returned. However, the large number of results returned, problems with term ambiguity, and the lack of a means of identifying relevant resources for specific learner needs (beyond simple content-based searching) makes Google a relatively poor tool for finding resources from which to learn.

Wiki Wiki allows anyone in the whole world to edit any Wiki page. The potential for chaos is enormous, and yet Wikipedia (http://www.wikipedia.org), an encyclopedia generated by thousands of volunteers with no central authority, is not only possible, but hugely successful. In Wikipedia, self-organization largely occurs through the goodwill of the participants. Although anyone may vandalize a page, the community that creates it quickly removes such defacements, leading to a highly reliable and comprehensive source of knowledge. The success of Wikipedia may be ascribed to many factors, not least of which are its strong structure and simple policies. Interestingly, it makes use of a meta-wikipedia where Wikipedians may discuss issues relating to articles to be posted. This parcellation contributes greatly to the evolution of ideas.

Weblog communities form through links provided from one weblog (a kind of easy-to-edit interactive online diary) to another. Links pointing to a weblog may be discovered through Google (using the related: keyword) or explicitly through referrer logs and backlink tools. As links between weblogs grow, they start to form small, stigmergic clusters (Gregorio, 2003). This is in accordance with the principles of small world networks (Watts & Strogatz, 1998) whereby connections between a small number of nodes within a much larger network form natural clusters. As long as an appropriate cluster can be found, such networks provide a powerful means of finding structure in a subject. One way to find such networks is through a recommender system.

PHOAKS (http://www.phoaks.com) is a recommender system. Like Google, it makes use of LHA. It searches for references to Web sites in newsgroup postings and, having filtered out those that are obviously advertisements, matches themes of the newsgroup with the search term entered and provides rankings based on the results (Terveen et al., 1997).

Collaborative filters are recommender systems that make use of implicit and/or explicit preferences to provide recommendations. For instance, if you and I share a liking for nine books, and if you like a further book that I have not read, then it is likely that I will also like that book. Collaborative filters are very useful for matching users' preferences, but tend to be less successful when seeking learning resources because to learn is to change. My previous preferences for particular resources will be a less reliable guide to my future needs than preferences for books, movies, or Web sites, because the act of using a learning resource will (if successful) change me in ways that are likely to differ from how they have changed you. Nonetheless, systems such as Altered Vista (Recker et al., 2000) have achieved some success in this area.

Explicitly Emergent Systems

There are a few systems that explicitly exploit self-organizing principles to the benefit of learners. The selection presented here is a relatively small subset that indicates how this area is developing.

An exception to the rule that collaborative filters cannot cope well with changing needs is the CoFIND system (Dron et al., 2003), which combines both evolutionary and stigmergic principles to provide both social navigation and collaborative filtering. Rather than simple good-bad ratings, it employs a multi-dimensional matrix of "qualities" that loosely translate into those things that learners find useful about a resource (e.g., if it is good for beginners, detailed, exciting, etc.). Because qualities are created by learners and used by other learners to provide explicit ratings, they provide a kind of footprint of the learning needs that led to a particular resource being recommended. This remains even after the learner has moved on. By basing its recommendations on an explicit metadata model rather than a user model, it thus overcomes the problem of changing user models. In keeping with evolutionary principles, not just the resources in the system but the metadata that describe them are in competition with each other. Combined with positive feedback loops generated by social navigation using stigmergy, this means that each CoFIND system develops into a unique ecosystem composed of smaller, interacting ecosystems.

The SEO system also explicitly employs stigmergy, using agents to gather information from experts, which it combines in what is described as a "Kempelen box," named after the Mechanical Turk, an 18th century chess-playing machine that, in reality, contained a human player. These agents leave information trails that provide self-reinforcing paths like ant trails to useful knowledge (Small, 2001). The box provides answers to questions through the use of clusters of agents formed through a combination of stigmergy and evolutionary processes that weed out less useful clusters. Successful agent clusters are those that are most queried. Thus, the structure of the system adapts to the kind of questions being asked of it.

Jaakko Kurhila's (2002) EDUCO system makes use of stigmergy in the form of social navigation features to provide not only visual indicators of the relative popularity of documents within the system, but also real-time indicators of who is currently viewing them. This is combined with a chat system that enables interactions between users, providing a powerful incentive to visit pages that are currently being viewed. A similar principle is employed in Dron's (2003) the *Pavement*, which also makes use of stigmergic self-organizing principles suggested by Jacobs (1961) in "The Death and Life of Great American Cities."

FUTURE TRENDS

Darby's (2003) third generation of learning environments already exists, but by far the majority of effort is still being expended in the construction of first- and (occasionally) second-generation resources. Meanwhile, the huge popularity of systems such as Google demonstrates that there is a demand for discovering information, often in the service of learning. With the ever increasing trend towards meaningful metadata being appended to resources using open, XML-based standards, particularly Resource Description Framework (RDF, the foundation of the Semantic Web), the accuracy of such searches will be immeasurably improved over the coming years. However, to turn such information into useful knowledge and learning, a more sophisticated set of tools is needed. Such tools will combine the knowledge of many people, effectively amplifying intelligence and operating in some senses as a kind of group mind.

CONCLUSION

The World Wide Web is more than a collection of documents. Increasingly, the face presented to the end user is the result of a large amount of processing. The dynamic generation of resources to learn from is moving slowly away from rule-based machine- or human-governed information to a more symbiotic relationship in which the strengths of machines and the strengths of people are combined and, in the process, mutually enhanced. In the process, the high transactional distance of resource-based learning is reduced by glimpses of the footprints of others, a kind of social translucence (IBM_Social_Computing_Group, 2003). In a world where our roles are changing faster than the traditional course-based approaches to the delivery of learning can address, the resulting group mind can adapt itself more readily and effectively to the needs of individual learners than can any single human being.

REFERENCES

Brand, S. (1997). *How buildings learn*. London: Phoenix Illustrated.

Calvin, W.H. (1997). The six essentials? Minimal requirements for the Darwinian bootstrapping of Quality. *Journal of Memetics, 1*.

Churchill, W. (1943). HC Deb 28 October 1943 c403.

Darby, J. (2003). *UK eUniversities worldwide: Who we are and what we want from standards*. Retrieved December 14, 2003, from *http://www.imsglobal.org/otf/IMS-Darby.pdf*

Darwin, C. (1872). *The origin of species* (6th ed.).

Dron, J. (2003). *Sidewalks on the information superhighway* [Report]. *Proceedings of E-Learn 2003*, Phoenix, AZ.

Dron, J., Mitchell, R., & Boyne, C.W. (2003). Evolving learning in the stuff swamp. In N. Patel (Ed.), *Adaptive evolutionary information systems*. Hershey, PA: Idea Group Publishing.

Grasse, P.P. (1959). La reconstruction du nid et les coordinations inter-individuelles chez Bellicoitermes natalenis et Cubitermes sp. La theorie de la stigmergie: Essai d'interpretation des termites constructeurs. *Insect Societies, 6*, 41-83.

Gregorio, J. (2003). *Stigmergy and the World-Wide Web*. Retrieved December 13, 2003, from *http://bitworking.org/news/Stigmergy/*

IBM_Social_Computing_Group (2003). *Social computing group FAQ*. Retrieved April 28, 2003, from *http://www.research.ibm.com/SocialComputing/SCGFAQs.htm#WhatIsSocialTranslucence*

Jacobs, J. (1961). *The death and life of great American cities*. London: Pimlico.

Kauffman, S. (1995). *At home in the universe: The search for laws of complexity*. London: OUP.

Kelly, K. (1998). *New rules for the new economy*. New York: Penguin Group.

Kleinberg, J.M. (1998). *Authoritative sources in a hyperlinked environment* [Report]. *Proceedings of the 9th ACM-SIAM Symposium on Discrete Algorithms*.

Kolb, D.A. (1984). *Experiential learning*. Englewood Cliffs, NJ: Prentice Hall.

Kurhila, J., Miettinen, M., Nokelainen, Petri, & Tirri, H. (2002). *Use of social navigation features in collaborative e-learning* [Report]. *Proceedings of E-Learn 2002*, Montreal, Canada.

Moore, M.G., & Kearsley, G. (1996). *Distance education: A systems view*. Belmont, WA: Wadsworth.

Recker, M.M., Walker, A., & Wiley, D. (2000). *An interface for collaborative filtering of educational resources* [Report]. *Proceedings of the 2000 International Conference on Artificial Intelligence*, Las Vegas, NV.

Saba, F., & Shearer, R.L. (1994). Verifying key theoretical concepts in a dynamic model of distance education. *The American Journal of Distance Education, 8*(1), 36-59.

Shirky, C. (1996). In praise of evolvable systems. *ACM Net_Worker*.

Small, P. (2001). *A self-organizing, living database for volatile data*. Retrieved from http://www.stigmergicsystems.com/stig_v1/papers/page1.html

Terveen, L., Hill, W., Amento, B., McDonald, D., & Creter, J. (1997). PHOAKS: A system for sharing recommendations. *Communications of the ACM, 40*(3), 59-62.

Watts, D., & Strogatz, S. (1998). Collective dynamics of "small-world" networks. *Nature, 393*, 440-442.

KEY TERMS

Collaborative filter: A form of recommender system that uses implicit or explicit recommendations of others to provide advice (see Recommender System).

Recommender system: A computer program that recommends some sort of resource based on algorithms that rely on some sort of user model, some sort of content model, and some means of matching the two.

Social navigation: The transformation of an interface (usually Web-based) by using the actions of visitors.

Social translucence: An indirect or oblique indication of the presence of others visiting a Web site.

Stigmergy: A form of indirect communication whereby signs left in the environment influence the behavior of others who follow.

Transactional distance: A measure of the relative amounts of dialogue and structure in an educational activity. Of necessity, as one increases, the other decreases, and vice versa. More autonomous learners require less dialogue than more dependent learners.

Latent Human Annotation (LHA): The unintentional communication of a recommendation or other information as a by-product of another process (e.g., the provision of hyperlinks on a Web page, which are then used by search engines to provide rankings of the linked pages.

Emergent behavior: Behavior that arises out of the interactions between parts of a system and cannot be predicted easily or extrapolated from the behavior of those individual parts.

Semantic Web Fundamentals

Grigoris Antoniou
Institute of Computer Science, FORTH, Greece

Vassilis Christophides
Institute of Computer Science, FORTH, Greece

Dimitris Plexousakis
Institute of Computer Science, FORTH, Greece

Martin Doerr
Institute of Computer Science, FORTH, Greece

INTRODUCTION

World Wide Web: A Critical View

The World Wide Web (Berners-Lee, Cailliau, & Groff, 1992; Berners-Lee, 1999) has changed the way people communicate with each other and the way business is conducted. It lies at the heart of a revolution that is currently transforming the developed world toward a knowledge economy (Neef, 1997), and more broadly speaking, to a knowledge society.

Most of today's Web content is suitable for human consumption. Even Web content that is generated automatically from databases is usually presented without the original structural information found in databases. Typical uses of the Web today involve humans seeking and consuming information, searching and getting in touch with other humans, and reviewing the catalogs of online stores and ordering products by filling out forms.

These activities are not particularly well supported by software tools. Apart from the existence of links that establish connections between documents, the main valuable, indeed indispensable, tools are search engines. Keyword-based search engines, such as AltaVista (www.altavista.com), Yahoo (www.yahoo.com), and Google (www.google.com; Page & Brin, 1998), are the main tool for using today's Web. It is clear that the Web would not have been the huge success it was, were it not for search engines. However, there are serious problems associated with their use; the most important ones are listed in Table 1.

BACKGROUND

The Semantic Web Vision

The Semantic Web (Berners-Lee, Hendler, & Lassila, 2001; Davis, Fensel, & van Harmelen, 2002; Fensel et al., 2002; Antoniou & van Harmelen, 2004) proposes to overcome the difficulties listed above by making Web content machine processable. The key point is that the semantics (meaning) of Web content must be explicitly represented and processed. This aim will be achieved by combining the following technologies:

Table 1. A list of problems associated with keyword-based search engines

High recall, low precision Too many, mostly irrelevant pages are retrieved
Low or no recall Key relevant pages are not retrieved
Sensitivity to chosen vocabulary Slight changes in vocabulary may cause significant changes in results
Results are single Web pages Information may be spread across various pages
Human involvement is necessary To interpret retrieved pages, and to combine information

- *Explicit Metadata*: Web content will carry its meaning "on its sleeve" through appropriate semantic markup.
- *Ontologies*: They will describe semantic relationships between terms and will serve as the foundation for establishing shared understanding between applications.
- *Logical Reasoning*: Automated reasoning-enabled tools will make use of the information provided by metadata and ontologies.

More on ontologies and reasoning are found in the following sections. As a simple example, suppose that you are searching for photos of an orange ape in an annotated online collection of digital photos. Suppose that picture 1 is annotated as "playing orangutan." Then this picture can be retrieved on the Semantic Web, although its annotation does not contain the words "orange" or "ape" (so, a keyword-based search would fail). This can be achieved through interplay of (a) the annotation, (b) information contained in an ontology about animals that states that orangutans are apes and are orange, and (c) reasoning that combines the above information to conclude that the picture is relevant to the user's query.

The Semantic Web vision was created by Tim Berners-Lee, the person who created the WWW. The Semantic Web activities are coordinated by the World Wide Web Consortium (http://www.w3.org/2001/sw/). Table 2 collects a few critical issues of Web technologies addressed by the Semantic Web initiative.

FUTURE TRENDS

Ontologies

An ontology is a formalization of a shared conceptualization of a particular domain. It supports interoperability between applications at the semantic level (the meaning of information) and also allows people to talk about objects of common interest. Typically, it describes the objects of the domain and the relationships that hold between them.

Objects of the same kind are organized in so-called classes that are collections of objects sharing certain characteristics. For example, in a university domain, classes may be professors, students, administrative and technical staff, courses, lecture theaters, etc. Individual objects (called *resources* in Web terminology) can be declared to be instances of a certain class.

Once we have defined classes, we can also define relationships between them. One particular kind of relationship is a class hierarchy. A class A is a subclass of a class B (and B is a superclass of A) if every instance of A is also an instance of B. For example, the class of professors is a subclass of the class of all university employees. Figure 1 shows a sample class hierarchy for the university domain.

The subclass relationship is a general-purpose relationship between classes. Further relationships, called *properties*, can be defined by the user. In the university domain, such properties might be that a course is taught by a particular professor, that a student takes a course, and that a professor is head of a department.

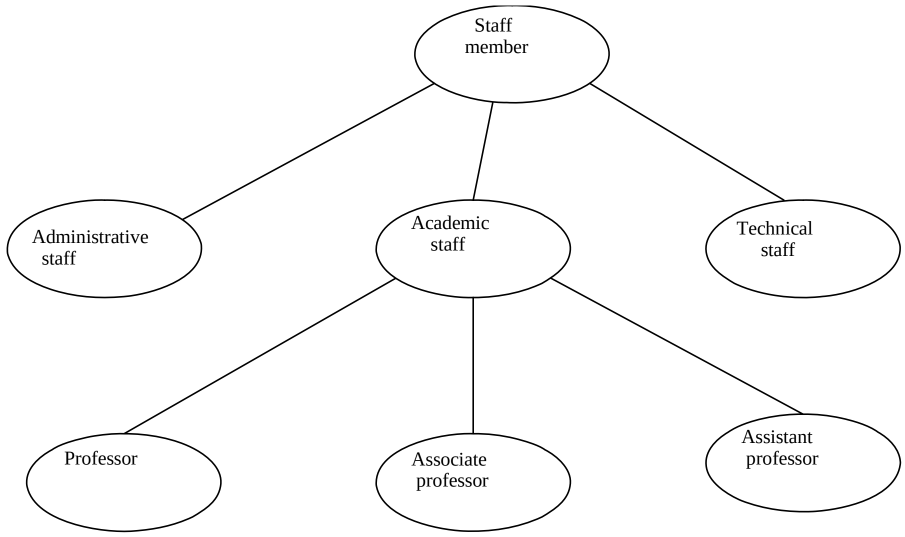
The interplay of classes and properties opens interesting modeling possibilities. For one, it is possible to define domain and range restrictions. For example, one can specify that a course can only be taught by an academic staff member (*range restriction*) and that only a course, and not, say, a lecture theater, can be taught (*domain restriction*).

Moreover, the valuable concept of *inheritance* can be utilized. Suppose that person *X* is declared to be an associate professor. Then *X* is allowed to teach a course, because he inherits this possibility from its superclass academic staff member. This way, we can avoid adding superfluous information to the ontology (*X* is an academic staff member; *X* is a staff member).

Table 2. Critial issues of Web technologies (Khosrow-Pour, 2004) addressed by the Semantic Web

Cyberloafing	Surfing the Internet, wasting time, and accessing inappropriate materials
Flooding of the Web with content	Including information that is not helpful
Having inadequate search facilities on the WWW	Lack of high-level query language search engines for locating, filtering, and presenting information
Maintaining integrity of data	Maintaining up-to-date and accurate information on the site for viewers to use

Figure 1. Class hierarchy in the university domain



Information such as the above can be expressed in an ontology language. Ontology languages can be informal, e.g. thesauri, or formal, when they rely on some kind of logical formalism.

Ontologies have become valuable tools in various domains, for example, in biology (Bodenreider, Mitchell, & McCray, 2003), medicine (Unified Medical Language System, UMLS), and culture (CIDOC/CRM).

Reasoning

When an ontology language is formal, it is possible to reason about the knowledge expressed in a particular ontology. For example, we may reason about the following:

- *Class Membership*: If *x* is an instance of class C, and C is a subclass of D, then we can infer that *x* is an instance of D.
- *Equivalence of Classes*: If class A is equivalent to class B, and B is equivalent to class C, then we can infer that A is equivalent to C.
- *Consistency*: If we declared that classes C and D are disjoint, and *x* is an instance of both C and D, then there is an error.
- *Classification*: If we declared that certain property–value pairs are sufficient conditions for membership in class A, then if an individual *x* satisfies such conditions, we can conclude that *x* must be an instance of A.

Derivation such as the preceding can be made mechanically instead of being made by hand. Such reasoning support is important, because it allows one to do the following:

- Check the consistency of an ontology and the knowledge
- Check for unintended between classes
- Automatically classify instances of classes

Automated reasoning support allows one to check many more classes than could be checked manually. Checks like the preceding ones are valuable for designing large ontologies, where multiple authors are involved, and for integrating and sharing ontologies from various sources.

Relevance to Sample Application Fields

Tables 3 and 4 illustrate the relevance of the Semantic Web vision to the application areas of knowledge management and business-to-consumer (B2C) e-commerce (Fensel, 2001). Other areas that would greatly benefit from the realization of the Semantic Web vision include business-to-business (B2B) e-commerce, e-learning (Nilsson, Palmer, & Naeve, 2002), e-government (Klischewski, 2003), Web services (McIlraith, Son, & Zeng, 2001), and peer-to-peer systems (Nejdl et al., 2002).

S

Table 3. Semantic Web and knowledge management

• Knowledge will be organised in conceptual spaces, that is, according to concepts and their interrelations, not according to keywords or other ad hoc means.
• Automated tools will support maintenance by checking for inconsistencies and extracting new knowledge, hidden in the stored knowledge.
• Keyword-based searching will be replaced by query answering based on what needs to be retrieved (semantic query answering).
• Defining views on certain parts of information (even parts of documents) will be possible. This way, certain pieces of knowledge will be hidden from certain groups of persons, while they will be accessible to others.

Table 4. Semantic Web and B2C e-commerce

• Pricing and product information will be extracted correctly, and delivery and privacy policies will be interpreted and compared to user requirements.
• Additional information about the reputation of online shops will be retrieved from other sources, for example, independent rating agencies or consumer bodies.
• The low-level programming of wrappers, necessary for today's shopbots, will become obsolete.
• More sophisticated shopping agents will be able to conduct automated negotiations, on the buyer's behalf, with shop agents.

CONCLUSION

The Semantic Web is the next step in the development of the WWW. It proposes to dramatically improve the Web's functionality and usability by making the meaning of Web content machine processable.

At present, most of the effort has been placed on the design of formal languages for describing Web resources and for writing Web ontologies. The most important languages are the language of resource description framework (RDF; Lassila & Swick, 1999), which allows one to state simple facts, and the ontology languages RDF Schema (Brickley & Guha, 2003) and OWL (Dean & Schreiber, 2003). The success of the Semantic Web will critically rely on two factors:

1. The development of professional tools for annotating pages and linking them to existing ontologies, for developing ontologies, and for reasoning with them—RDF Suite (Karvounarakis et al., 2002) is a good collection of tools supporting RDF and RDF Schema
2. The uptake by users—initial successes can be expected either in environments with central authority (e.g., in the knowledge management of large organizations) or in the emergence of virtual communities

REFERENCES

Antoniou, G., & van Harmelen, F. (2004). *A Semantic Web primer*. Cambridge, MA: The MIT Press.

Berners-Lee, T. (1999). *Weaving the Web*. New York: Harper.

Berners-Lee, T. J., Calilliau, & Groff, J. F. (1992). The World Wide Web. *Computer Networks & ISDN Systems, 25*(4), 454–459.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American, 284*(5), 34–43.

Bodenreider, O., Mitchell, J. A., & McCray, A. T. (2003). Biomedical ontologies. In *Proceedings 2003 Pacific Symposium on Biocomputing* (pp. 562–564). Singapore: World Scientific.

Brickley, D., & Guha, R. V. (2003). *RDF vocabulary description language 1.0: RDF Schema*. Retrieved from *http://www.w3.org/TR/rdf-schema/*

CDOC/CRM. (No Date). CIDOC conceptual reference model. Retrieved from *http://cidoc.ics.forth.gr*

Davis, J., Fensel, D., & van Harmelen, F. (2002). *Towards the Semantic Web: Ontology-driven knowledge management*. New York: Wiley.

Dean, M., & Schreiber, G. (2003). OWL Web ontology language reference. Retrieved from *http://www.w3.org/TR/owl-ref/*

Fensel, D. (2001). *Ontologies: Silver bullet for knowledge management and electronic commerce*. Berlin: Springer-Verlag.

Fensel, D., Hendler, J., Lieberman, H., & Wahlster, W. (2003). *Spinning the Semantic Web*. Cambridge, MA: The MIT Press.

Karvounarakis, G., Alexaki, S., Christophides, V., Plexousakis, D., & Scholl, M. (2002). RQL: A declarative query language for RDF. In *Proceedings of the 11th International World Wide Web Conferenc (WWW'02)*.

Khosrow-Pour, M. (2004). An overview of Web-enabled technologies assessment and management: Critical issues. Retrieved from *http://www.idea-group.com/custom/encyclopedia/sampleManuscript.doc*

Klischewski, R. (2003). Semantic Web for e-government. In R. Traunmüller (Ed.), *Proceedings of EGOV 2003* (pp. 288–295). Berlin: Springer, LNCS 2739.

Lassila, O., & Swick, R. R. (1999). *Resource description framework (RDF) model and syntax specification*. Retrieved from *http://www.w3.org/TR/REC-rdf-syntax/*

McIlraith, S. A., Son, T. C., & Zeng, H. (2001). Semantic Web Services. *IEEE Intelligent Systems, 16*(2), 46–53.

Neef, D. (1997). *The knowledge economy*. Oxford: Butterworth-Heinemann.

Nejdl, W., Wolf, B., Qu, C., Decker, S., Sintek, M., Naeve, A., Nilsson, M., Palmér, M., & Risch, T. (2002). EDUTELLA: A P2P networking infrastructure based on RDF. In *Proceedings of the 11th International World Wide Web Conference (WWW'2002)* (pp. 604–615).

Nilsson, M., Palmér, M., & Naeve, A. (2002). Semantic Web meta-data for e-learning—Some architectural guidelines. In *Proceedings of the 11th World Wide Web Conference (WWW2002)*, Hawaii, USA.

Page, L., & Brin, S. (1998). The anatomy of a search engine. In *Proceedings of the Seventh International WWW Conference (WWW'98)*. Brisbane, Australia, April 14–18.

UMLS. (No date). *Unified Medical Language System*. Retrieved from *http://www.nlm.nih.gov/research/umls/umlsmain.html*

KEY TERMS

Class: A collection of objects sharing certain characteristics (such as producer, customer, product, company, etc.) Individual objects are called instances of the class. The classes of an ontology describe the important concepts of a particular domain being modeled.

Class Hierarchy: Classes are usually organized in a conceptual space along a generalization/specialization axis. A class A is more general (superclass) than a class B when each instance of B is also an instance of A.

Ontology: A consensual and formal description of shared concepts in a domain. Typically, it organizes the objects of a domain in classes and includes a hierarchy of classes (e.g., printers are hardware devices). Ontologies are used to aid human communication and shared understanding and also communication among software applications.

OWL: The current W3C standard for defining Web ontologies.

Property: Properties are used to establish links between classes. For example, books are published by publishers. The property "publishes" relates a publisher with a particular book.

Property Hierarchy: As with classes, properties can be organized in generalization/specialization taxonomies. For example, the property "is CEO of" relating persons to companies is a more specific property than (subproperty of) "works for."

RDF (Resource Description Framework): This is the basic language of the Semantic Web. It is used for describing Web resources.

RDF (Resource Descrioption Framework) Schema: This is a primitive ontology language that is integrated with RDF. Its basic operations are the definition of classes and properties, their organization in hierarchies, and domain and range restrictions on properties.

XML (Extensible Markup Language): An application-independent meta-language for defining markup languages. It serves as the basis for syntactic interoperability among information systems.

Shaping the Evolution of Mobile Commerce

Andreas Rülke
PRTM, UK

Anand Iyer
PRTM, UK

Greg Chiasson
PRTM, UK

INTRODUCTION

How does a breakthrough technology, or a breakthrough combination of formerly separate technologies, become a viable business? What are the necessary conditions, competencies, and organizing mechanisms? Which enterprises are in the best positions to provide the various competencies and to organize the new business? How will the new business unfold?

Such are the questions posed by mobile commerce (m-commerce) or wireless electronic commerce, the business result of the Internet's convergence with broadband wireless communications. M-commerce's value chain model can help shed light on how this new business will develop and grow in the years to come.

Generally speaking, "value chain" can be defined as a map of the entire set of competencies, investments, and activities required to create, produce, deliver, maintain, and reap the proceeds from a product or service, and the relationships among those investments and activities. The profits and competitive advantages of participation in a given value chain reside dynamically within the chain, accumulating at the positions of greatest value. The enterprises that hold these positions have a great deal of control over how the chain operates and how the benefits are distributed.

The wireless telecommunications industry has changed dramatically since the late 1970s and early 1980s, when the first commercial handsets were launched. Since that time, mobile cellular communications systems have completed two entire generations of development, as have their respective value chains. The first mobile generation, or '1G,' used analog technology for transmitting voice calls, and had a correspondingly simple value chain (see Figure 1). Its value chain was very similar to that of fixed lines.

The second wireless generation ('2G') used digital technology and limited bandwidth. It provided Short Message Service (SMS) text messaging in addition to voice service.

An intermediary step between the second and third generations is 2.5G, which has higher bandwidth than 2G and "always on" wireless access.

BACKGROUND

We are now beginning to see the results of the third generation (3G) of wireless development. The 3G networks operate at a much faster speed than those of its predecessors and feature much greater capabilities: the Multimedia Messaging Service (MMS), which supports simultaneous transmission of various media—images, text, video, and audio clips—along with voice (Rowello, 2001). The 3G-value chain is also considerably more complex, as Figure 3 illustrates.

Figure 1. First-generation value chain: cellular voice service

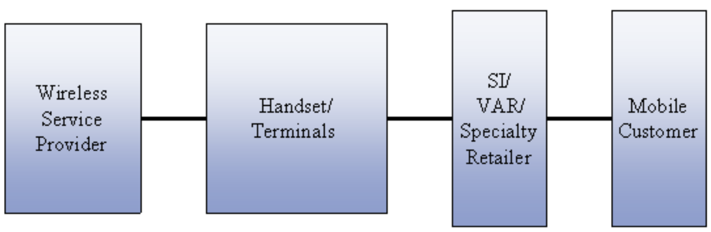


Figure 2. Second-generation value chain: digital voice and data

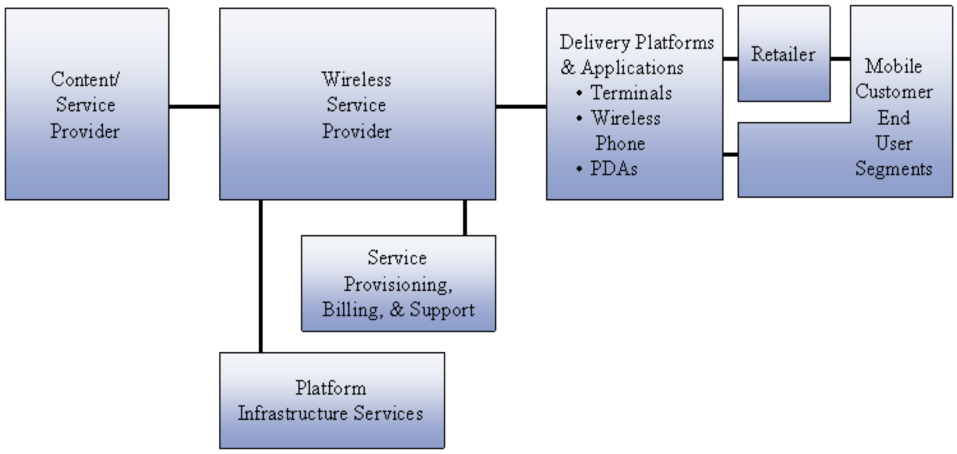
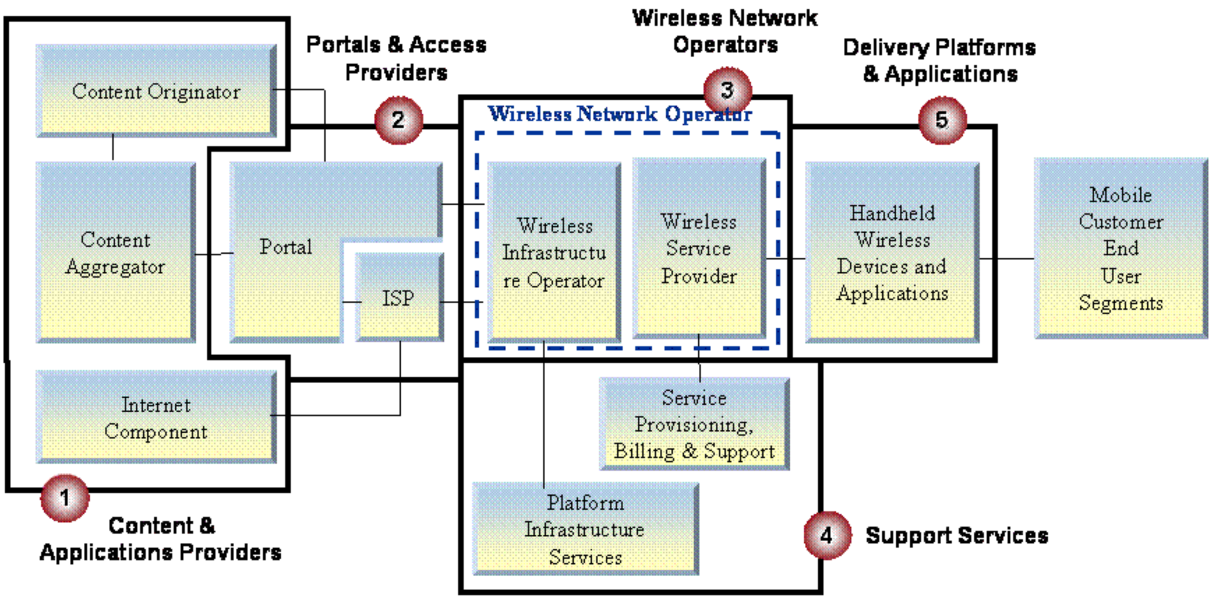


Figure 3. Next-generation value chain: the wireless Internet



In an article published in 2000, we presented a model of the unfolding '3G' mobile commerce value chain, which groups the participants into five major elements (Rulke & Chiasson, 2000):

- Element 1—Content and Application Providers: e.g., Bloomberg and MapQuest
- Element 2—Portal and Access Providers: e.g., Yahoo! and AvantGo
- Element 3—Wireless Network Operators: e.g., Sprint PCS and Vodafone
- Element 4—Support Services: e.g., SpectraSite and Convergys
- Element 5—Delivery Platforms and Applications: e.g., Nokia and palmOne

This value chain is highly horizontal, and reflects the paths and supporting capabilities required to consummate mobile commerce: to create, aggregate, sell, and deliver content.

The time for 3G is imminent. The network technology is finally sound enough to be introduced to the general public. Traditional defined m-commerce and location-based services were slower to roll out than planned (Carbone, 2003) but are ready now, while 3G content, such as ringtones and music, are becoming increasingly important in the industry. Well-performing handsets are still in short supply (Tanner, 2004) but should be available for a rollout in 2005. For these reasons, this is an apt time to reexamine our 3G value chain and consider its relevance for strategic planning efforts.

MAIN THRUST OF THE CHAPTER

Value chain analysis is predicated on the fact that some forms of participation in horizontally distributed businesses are more advantageous than others. Because wireless network operators are in a position to leverage their existing relationships with their mobile phone customers into m-commerce relationships, they currently occupy one of the strongest positions within the mobile commerce value chain. Moreover, incumbent network operators have cleared the very high capital-investment hurdle facing would-be competitors.

Market history underscores the importance of seizing these two advantages. Many providers of wireline Internet services were slow to press their advantage in developing Internet offerings, only to see their customers migrate to new entities such as Yahoo! and AOL.

Wireless operators' third current advantage within the m-commerce value chain is their information on their subscribers' whereabouts. In the U.S., this information is becoming fairly accurate with the advent of "E911" emergency location-determination requirements. Knowledge of subscribers' locations is invaluable in directing contextual advertising messages, and in providing services such as directions to and phone numbers of nearby businesses. This information provides more than a source of advantage over other m-commerce participants; it affords a significant advantage over wireline ISPs, which do not typically track users' whereabouts.

Wireless operators have other advantages as well. They have extensive billing systems in place, which are generally flexible enough to capture m-commerce as well as access charges. This is of particular advantage in areas of the world where credit cards are less common, or where there is greater reluctance to use them for online transactions.

Value chain analysis also reveals how the positions of advantage within a chain may shift. The key challengers to wireless operators will be the Internet and "dot.com" companies that are part of the World Wide Web. These companies include many thousands of ISPs, business portals, content providers, and other software companies. These entities tend to be extremely quick to react to—or even to create—market changes. They're in the business of being first to market with a product that works. They're also in the business of continuous product improvement, iteratively building customer solutions that are very much on target.

FUTURE TRENDS

As traditional and nontraditional wireless enterprises converge to form content/delivery partnerships in the mobile commerce space, the positions of advantage within the industry's value chain will shift according to emerging patterns of consumer demand and preference. How will the distribution of revenues and profits change over time? What types of service offerings and business structures will be required to manage these changes? How can companies best implement the necessary changes?

In 2001, PRTM conducted a survey in order to get a sense of the wireless industry's expectations associated with the rollout of 3G wireless technologies over the next five years. We used our value chain model as the basis for the survey. A total of 91 respondents participated. Wireless infrastructure manufacturers, terminal and handset makers, wireless operators, content and applications providers, portal companies, and providers of a variety of specialized wireless-related services from more than 80 countries were all represented. A brief overview of the findings was published in the October 15, 2001, issue of *Telephony* magazine (Chiasson & Guyton, 2001).

The responses were quite consistent with the expectations stemming from our value chain model and analysis. Given this consistency and rollout delays, the survey results continue to be a relevant source of information on how the 3G value chain will impact m-commerce.

Companies are looking to capitalize on or compensate for shifts in revenues and profits among the various elements of the wireless value chain through integrated next-generation offerings. Respondents expected a gradual revenue shift away from wireless network operators and toward content and applications providers as the locus of value shifts from transport toward content.

New location-based wireless services, coupled with an improved ability to charge for content (through micro-billing, ASP models, etc.), are the key factors behind the anticipated reapportionment of revenues within the wireless industry. At the same time, wireless operators face both mounting competition and the commoditization of their offerings, much as we saw with land-line long-distance operators.

While the largest share of revenue will continue to accrue to wireless operators over the next five years, the distribution of revenue across the value chain elements will become more equitable. Respondents expected content and applications providers' annual revenues to grow the fastest, doubling from 11% to 22% of the total. This still makes sense, given the coming growth in data-based content and the growing consumer willingness to pay for it.

Today, most of the revenue is generated through voice, and that revenue belongs to the wireless operators. In the future, portals will give customers access to graphics-rich Web sites via mobile handset screens. To limit their decline in revenue share, wireless operators need to avoid becoming mere fungible pipelines.

Respondents expect the wireless industry as a whole to become more profitable over the next five years. Net profit margins are expected to grow, with the largest increases in profitability accruing to content and application providers that are able to take advantage of economies of scale by spreading their fixed costs over wider customer bases. Wireless operators are expected to hold net profit margins constant by focusing on their most profitable customers, and through increasing efficiencies.

For the wireless industry's expected gains to materialize, companies will need to develop offerings that span multiple value chain elements. The large majority of respondents (78%) report that their next-generation business plans focus on multiple value chain elements. Furthermore, companies with superior relative rates of revenue growth attach the greatest importance to integrated offerings. Traditional wireless companies will look to share in the growth and profitability of the new content- and applications-based developments, while companies new to wireless will need to offer solutions that combine delivery with content. Recent telematics ventures such as Wingcast and OnStar are excellent examples of nontraditional players (automotive OEMs in this case) partnering with existing wireless value chain participants to bring new integrated offerings to end consumers.

Multi-element participation in the value chain, both direct and indirect, is expected to increase over the next five years. The largest shift is expected from "one element" to "two element" companies. The number of companies participating in only one value chain element is expected to decrease from 39% in 2001 to 28% by the end of 2005. Given the delayed roll-out of the 3G systems by early 2004, this percentage might have been delayed, but we still expect the same basic trend to take place.

Partnerships will be the preferred means of integrating across the 3-G wireless value chain. The desire to provide integrated offerings, and the expectation of participating in multiple value chain elements, both indicate the transition away from transaction-oriented interactions among value chain participants in favor of more closely coupled business structures. Although the development of cross-chain capabilities in-house is the ultimate in integration, companies generally prefer to partner with holders of existing capabilities. The content and applications providers element, which is expected to enjoy the greatest revenue and profitability growth, is expected to see the most partnering activity. Wireless operators expect to form, on average, 21 partnerships with content and application providers by 2003. Conversely, content and application providers expect to form, on average, two partnerships with wireless operators and three partnerships with delivery platforms and applications by 2003.

PRTM surveyed respondents about a range of anticipated partnership types, including alliances, minority investments, joint ventures, and acquisitions, with varying degrees of coupling and tightness of integration. While the types of partnerships preferred depend on respondents' particular locations within the value chain, strategic/product alliances are the most likely approach to creating partnerships in all value chain elements.

Overall, respondents expect to form "tighter" partnerships (i.e., acquisitions, joint ventures, minority investments) with wireless operators and delivery platforms/applications providers. "Looser" partnerships (i.e., non-exclusive strategy/product alliances and marketing alliances) are favored with content and applications providers, and with portals and access providers. Verizon Wireless, AT&T Wireless, Sprint PCS, and Palm all have alliances with Yahoo! Mobile, for instance, and are listed as official partners on Yahoo!'s Web site.

The two most important reasons cited for forming partnerships are to exploit capabilities not available in-house, and to gain time-to-market advantage. Contributing reasons for partnering are to obtain a cost advantage over in-house capabilities, and to leverage the brands or customer bases of potential partners. Brand and customer leverage were particularly cited by content and applications providers as reasons for partnering.

The highly horizontal, distributed nature of the mobile-commerce value chain makes one fact plain: no single enterprise has the wherewithal to provide a true end-to-end solution that simultaneously optimizes business, shareholder, and customer objectives. But if partnership is vital, what is the right partnership model? Value chain analysis can guide you to the answer.

Each value chain element can be broken down into its constituent links. For instance, delivery platform applications consist of the key hardware, software, operating systems, standards, and so forth that comprise the technological heart of the solution. Support services can be broken down into installation, provisioning, customer care, and billing. Each of those links can, in turn, be subdivided into layers. For example, the hardware link of delivery platform applications can be decomposed into cellular receivers, GPS receivers, processors, logic, and the like.

Once the value chain has been broken down into its underlying elements, it is possible to associate elements with types or classes of candidate alliances. This point raises an important principle: An alliance should be mutually exclusive, yet collectively exhaustive. In other words, it should consist of the minimum number of players (element suppliers) required to deliver the solution, with no overlaps and no gaps. And, of course, the fewer the players, the better: less administrative and governance complexity, and more margin to go around.

We'll use "N" to represent this minimum number of mutually exclusive/collectively exhaustive players. Once you've determined your candidate list of N candidates, the options in terms of business arrangement can be depicted along a simple spectrum. At the far left of the spectrum is the consortium. At the far right is the formation of a new company ("newco"). Along the way are intermediate forms of business arrangement.

While there are no explicit variables that are intended to depict this continuum, one can envision that factors such as degree of management control, investment required, and governance complexity may increase from left to right.

In the context of mobile commerce, it has been our experience that the partnership forms on the right side of the spectrum have worked better than those on the left side.

A set of strategic and operational criteria can be applied to determine the best form for the business alliance from the options along the spectrum. The criteria include:

- Overall ability of the solution to meet customer goals
- Degree of up-front investment required
- Time to market requirement
- Degree of legal/regulatory compliance difficulty
- Degree of public acceptance of the solution offering
- Ability of the partnership form to absorb start-up costs
- Degree of brand leverage potential

In addition to leveraging multiple brands, the newco approach also leverages multiple core competencies, minimizing the new business's learning curve. The approach is well suited to combining best-in-class capabilities to produce a best-in-class offering. If a heavy infusion of capital into the partnership is required, the newco approach has the advantage of being able to draw capital from multiple sources. This approach can also be fastest to market, and has the greatest ability to absorb start-up costs. In addition, this approach has a major advantage when the time comes for market launch. If the newco brings together five companies, then it also brings together five sets of customers, creating a large and well-primed trial market.

CONCLUSION

To describe mobile commerce as an emerging business opportunity, or even as a 'business of businesses," would grossly fail to capture its extraordinary dynamism. M-commerce is a genesis in progress; a new and growing source of value at the confluence of two technological revolutions. The metaphor of a Web is useful in that it conveys the idea of an intricate and purposeful pattern, but the pattern of m-commerce is anything but fixed or final. Its patterns are being woven today, at the interconnections of the many diverse technologies, competencies, investments, and business models that comprise the mobile commerce value chain. M-commerce participation opportunities are value-chain (or "value-Web") participation opportunities, and the key to participation is partnership.

The value chain model presented here, along with the guidance offered on selecting the most appropriate form of partnership, will help companies set appropriate participation strategies, execute their strategies effectively and efficiently, and revise their strategies as circumstances change and opportunities arise. We have identified some clear trends in revenue distribution within the m-commerce value chain: for example, the trend toward a more equitable distribution of revenue across the elements of the chain over the next five years, notably favoring content and applications providers.

The emergence of mobile commerce, in all its complexity and flux, calls some basic business assumptions into question. For instance, to the old question, "Who owns the customer?", m-commerce value chain strategy poses the question, "Who *are* my customers in this Web of entities and partnerships, and to what degree can I own the relationships with value chain partners and end users?" The possibilities for positioning and branding within—and across—the value chain appear boundless. Mobile commerce will become even more complex in the future. The companies that begin proactively carving out their positions now will be best able not just to cope with the value chain's mounting complexities, but to shape the chain's evolution to their advantage.

ACKNOWLEDGEMENT

Parts of this article are excerpted from *Mobile Commerce: Technology, Theory, and Applications* (2003), edited by Brian E. Mennecke and Troy J. Stradler, and published by Idea Group Inc. (www.idea-group.com)

REFERENCES

Carbone, J. (2003). Watch wireless grow. *Purchasing, Chemicals Ed.*, *132*(15), 30.

Chiasson, G. & Gutyon, J. (2001). Survey draws landscape of next generation wireless value chain. *Telephony*, (October 15).

Rowello, R. (2001). The Internet unplugged: Consumer electronics in a 3G world. *Insight, 13*(1), 46-49.

Rulke, A. & Chiasson, G. (2000). Prospering in the age of mobile commerce. *Insight, 12*(3), 58-61.

Tanner, J.C. (2004). Hotspots go wide. *Wireless Asia, 7*(1), 22-25.

KEY TERMS

ASP: Application Service Provider.

Content Aggregator: Businesses that transform individuated content into specific and customer-tailored forms.

Content Originators: Businesses that create the highly specific types of content that is enhanced, combined, packaged, transmitted, and sold to customers.

Delivery Platforms and Applications: Handheld wireless device makers, handheld computing devices, personal digital assistants, and mobile handsets.

Internet Service Providers (ISPs): Provide the hardware that connects customers to content and applications providers, usually for a monthly fee.

Mobile Commerce: Wireless electronic commerce.

MMS: Multimedia Messaging Service.

Network Infrastructure Operators: The network-facing elements of wireless networks which provide the software and hardware that enable online communications.

OEM: Original Equipment Manufacturer.

Platform Infrastructure Services: Entities that own portfolios of telecommunication towers in various countries and rent space on the towers to network infrastructure operators.

Portals: A single, convenient point of access to all the products and services produced by the content originators/aggregators.

Service Provisioning, Billing, and Support: Various individual elements of customer service which may be outsourced by wireless service providers.

SMS: Short Message Service.

Wireless Service Providers: The customer-facing elements of wireless networks; the services whose quality is perceived by customers; speed of connection, clarity, and so on.

Sharing Organizational Knowledge through Knowledge Repositories

M. Asim Qayyum
University of Toronto, Canada

INTRODUCTION

Knowledge repositories are increasingly being viewed as a special form of knowledge management in organizational memory information systems (OMISs). Presented in this paper are the design concepts and guidelines for building a knowledge repository, and its practical implementation in the form of a prototype. The needs and the organizational and technical challenges associated with the undertaking of such a project are identified, and recommendations and strategies for overcoming the restrictions are discussed. In light of these repository-building guiding principles, the prototype of such a knowledge management system is envisaged to be a Web-based electronic repository of online pedagogical resources, built to help foster a learning organization that works together to gather and share knowledge. As per the design guidelines, information resources within this prototype are combined with user insights and experience in the form of associated annotations and then categorized within a subject tree to appear as knowledge to the repository users.

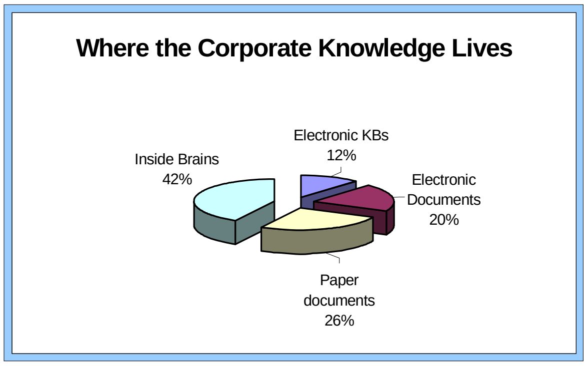
BACKGROUND

"Only 2 percent of information gets written down, the rest is in people's mind," says David Owens, vice president of knowledge management at Unisys Corporation and vice chair of the Conference Board's Learning and Knowledge Management Council (Hickins, 1999). This comment clearly illustrates that knowledge management (KM) places equal emphasis on capturing the tacit knowledge that is in people's heads, rather than targeting just the explicit knowledge that can be stored in a more shareable format. By managing its knowledge an organization would know more, and the more it knows, the more successful it will be. And this comes only after an understanding of the kind of information that is available to the members of an organization, where it is and how it can be accessed (Hackbarth & Grover, 1999).

While the figure quoted by David Owens is probably derived from personal experiences, a study of the empirical division of knowledge in an organization was carried out by the Delphi Group when they looked at KM practices in approximately 700 U.S. companies (Hickins, 1999). The results, presented in Figure 1, illustrate the fact that only a portion of the corporate knowledge is in shareable format while the majority (42%) of any one kind of knowledge resides inside people's heads. However, people leave organizations, taking away the knowledge that is stored in their heads. Therefore, organizations must build knowledge management (KM) systems, such as knowledge repositories, to retain the maximum possible tacit knowledge and make it available to the people who need it. This simple need for KM systems is supported by results from another survey, again by Delphi group, of 370 business professionals which showed that 28% had already begun or completed KM projects, while 93% said that they would undertake such projects by 2000 (Anthes, 1998).

Data and information need to be integrated to arrive at knowledge, and what is data to some may be information for others. Knowledge is, however, information that has been edited and analyzed in such a manner to make it useful. It has the greatest human contribution, stems from people, is the most difficult to manage, and is mostly context-specific (Grover & Davenport, 2001). And when knowledge from the past is brought to bear on present activities and thus affects the level of organizational effectiveness, then it is called organizational memory, or OM (Stein, 1992). A knowledge management system should then manage and expand this organizational memory while retaining a strong organizational learning founda-

Figure 1. Breakdown of knowledge areas within an organization



tion so as to successfully increase an organization's potential for taking effective action (Alavi, 2000; Davenport & Prusak, 1998). Such a knowledge management system would act as an organizational memory information system (OMIS), which provides a coherent integration of dispersed know-how, or OM, from all over the organization (Stein). An important aspect of OMIS is that the projects requiring its use do not involve a simple development of the system but require the incorporation of a concept of organizational development with a focus on enterprise-wide knowledge sharing and learning (Lehner, Mair, & Klosa, 1998). Thus, while an advanced database system will view information as a resource, an OMIS will focus on managing knowledge and contribute to learning ability, flexibility, and mastering of organizational change.

Development of an OMIS is not technology-driven but people-driven (Hickins, 1999), and these systems can firmly be placed as organizational knowledge management systems as they adapt to the social dynamics of the workplace. These social dynamics may include factors such as work habits, perceived benefits, and knowledge sharing. A knowledge repository can then be classified as a special form of OMIS because it embodies the phases of acquisition, retention, maintenance, and retrieval within its knowledge management framework (Hackbarth & Grover, 1999; Maier & Klosa, 1999). Processing of knowledge can be added to this list, and this may involve sorting, filtering, organizing, analyzing, comparing, correlating, mining, or a simple labelling of knowledge so that others can find it (Seng, Zannes, & Pace, 2002). These processes are necessary so that the objective of a knowledge repository can be achieved, which is to capture knowledge for later access by organizational members, and common repository techniques using these concepts include Lotus Notes, Web-based intranets, and document management tools (Grover & Davenport, 2001).

FISKR AS A KNOWLEDGE REPOSITORY

Sharing of tacit knowledge is aimed at gathering knowledge locked in people's heads, their notebooks, or desktops or simply lying in the filing drawer, things that were envisaged to be achieved by this repository. Thus, this was the rationale behind the design of the Faculty of Information Studies Knowledge Repository (FISKR) at FIS in University of Toronto. This repository is capable of acquiring and retaining "structured information," such as resources or pointers to resources of many different kinds, including, but not limited to, electronic or printed material, slide presentations, multimedia files, student papers, theses, reviews, or any other resources that are relevant to the subject matter. Retained in addition to structured information was "informal knowledge," which took the form of user-contributed annotations associated with each resource and provided added value in terms of conveying knowledge. Support was also made available to the faculty members for the creation of course pages and reading lists.

What differentiated FISKR from any other database was the fact that tacit information, which is ordinarily not shared and is deemed useful for the learning community, was made available in a shareable format to its members. Thus, the "implicit" data and information added to this repository consist of not just the resources but the associated contributor's experience, insight, and use context, which was made available to the members of a community in an "explicit" and organized form, or was now knowledge for its users. Such a design was also demonstrated in the conceptual design of a learning organization's KM system (Hall, Paradice, & Courtney, 2003). Nonaka and Takeuchi (1995) addressed this key issue of organizational knowledge creation and labeled it as "knowledge conversion," where human knowledge is created and expanded through social interaction between tacit knowledge and explicit knowledge.

The actual design of repository-based KM projects can be classified in two ways. The top-down approach looks at the knowledge present within a department and seeks to bring that together within a KM system by utilizing the services of a group of specialized personnel. Bottom-up approach on the other hand would tend to identify user needs and create a repository first and then encourage the users to add their knowledge to it. The latter category was chosen for implementing FISKR as it suited an educational environment.

A notable example of building a bottom-up knowledge repository as an OMIS was the Eureka project at Xerox (Hickins, 1999). It was noted at that time that the database housed more than 5,000 user "contributions," which were available to the Xerox employees via the World Wide Web right on their laptops. Docushare was another such Xerox project. Similarly, Hewlett-Packard used Lotus Notes to capture tips. Reviewing FISKR's design in context of the knowledge sharing at these companies, it was noted that success came out of the overall cooperative culture developed there as the management believed that technology must support the distributed sharing of knowledge because the work had become very cooperative in nature. This was in line with recommendations put forward by Davenport, DeLong, and Beers (1998), who suggested that a key success factor for knowledge management project was that knowledge repository creation should be accompanied by encouraging and facilitating communication among organizational members, thereby improving knowledge access and enhancing the knowledge environment. And once this effective communication and learning start occurring within an organization, then they often result in "intelligence" or "knowledge" (Seng et al., 2002). Thus, the system development has to work in parallel with the development of an organization-wide sharing culture. The successful deployment of these two factors is ultimately expected to lead to accelerated learning and innovative sharing environment to an academic setting, such as the one in which this project was implemented.

The other type of repository creation approach, the top-down approach, was demonstrated in the Optical Recorded Information Online Network (ORION) project, which was developed at the Federal Express Corporation. The main objective was to cope with the massive employee documentation requirements of its ever-increasing worldwide workforce, estimated to be at over 90,000 employees (Candler, Palvia, Thompson, & Zeltman, 1996).

PROTOTYPE DEVELOPMENT GUIDELINES

In a learning organization such as FIS, the critical issue was how the individual learning could be transferred to the organization, and this objective was implemented through this project. So when the knowledge transfer is to take place by constructing an OMIS, then a key challenge is to harness as much organizational memory into the OMIS as is feasible and appropriate (Hackbarth & Grover, 1999). To achieve this and to ensure the success of FISKR, it was essential that this repository blended in with the culture of FIS and that the students and faculty used the system without considering it as an additional burden imposed on top of their already overworked schedules. It was realized that if this KM system was viewed as extra work, then it may end up as a straight failure, as people tend to ignore systems which require them to input data (Hickins, 1999). Thus, the need was to develop a system that would be useful to the users in terms of supporting their class work and that they would accept as part of their routines.

In view of the above-mentioned constraints, some guidelines (King, Marks, & McCoy, 2002; Tobin, 1998) adapted during the development of this repository were as follows:

- To act as a useful learning tool, it had to be well organized, kept up-to-date, and be easy to access and search. The users will tend to look for other sources if they find the repository complicated to use.
- The information in the repository had to be kept accurate. If not monitored properly, data that has once been entered could stay there forever, even when it had been outdated or replaced by information that is more accurate. To implement this feature, this repository drew on the services of subject specialists, who classified and verified all user inputs and monitored the quality of contents.
- All members of the organization were given direct access to the repository. While certain portions of this repository were password protected, no other approval was necessary for majority of the users.

The bottom line here was that the information in the repository should be the information that the users need and that it may not be limited to information generated from within the organization. In addition, as the repository grew through user submissions, it amassed the information that users required, and this information came from within or external to the organization. It was only after implementing these conditions that it could be said that the information contained within this repository conveyed knowledge to the community members. Thus, as demonstrated by the project objectives and design and the actual prototype (that will be described in the implementation sections), the standards of a knowledge repository were adequately met in the FISKR implementation as it employed the phases of acquisition, retention, maintenance, processing, and retrieval.

Lastly, the prototype of FISKR has been introduced to students in at least one course and during a couple of presentations and seminars. The interest has been encouraging so far, and over time, the intention is to move out of the prototyping phase to the actual system implementation phase.

The Architectural Representations

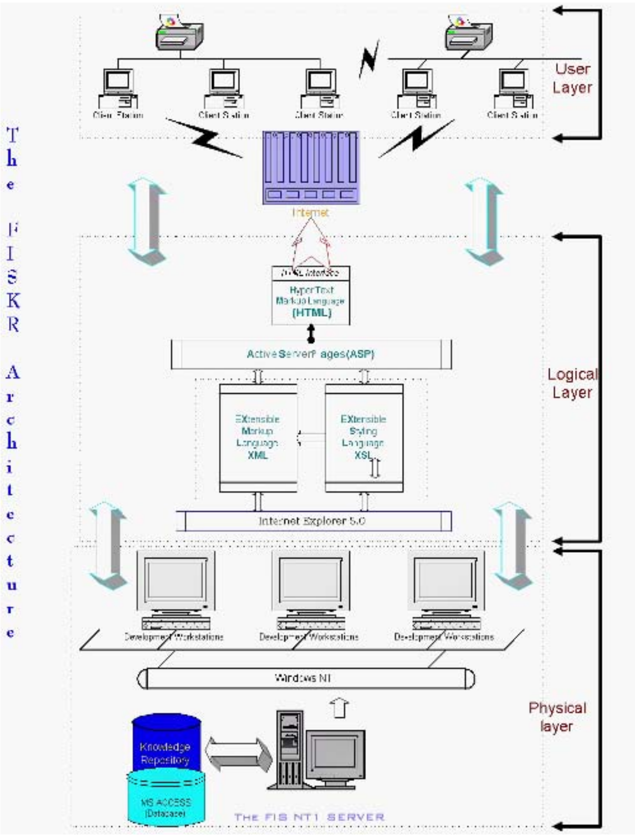
The three-tier physical architecture of the Faculty of Information Studies Knowledge Repository (FISKR) is demonstrated in Figure 2, which shows the various hardware and software components that have been utilized in this prototype. Briefly, the system was Web-based, built upon a Windows NT platform using the test server available at FIS. Microsoft's (MS) Active Server Pages (ASP) was used as the main script language with MS Access acting as the knowledge base at the back-end. Extensible Markup Language (XML) was used to structure the data.

Operational Features

The prototype of FISKR currently provides three options to the user:

1) "Search FISKR" allows an individual to search resources housed in FISKR by keyword, author

Figure 2. FISKR physical architecture



name, title, or type of resource, that is, whether it is a book, journal article, conference paper, research article, etc.

2) "Browse FISKR" allows an individual to browse resources housed in FISKR by various predetermined classification schemes.

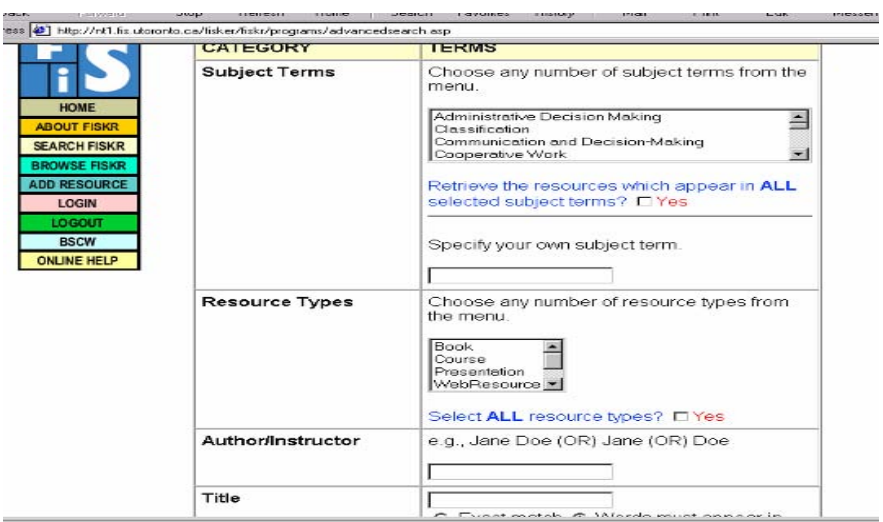
3) "Add a Resource" allows an individual to add resources to FISKR as well as annotations. FIS faculty members are given the additional option of entering a "Course" as a resource.

A typical search screen is shown in Figure 3. This is an advanced search screen, where the user is presented with a basic keyword search option also. A search can also be carried out by date or author information.

When browsing resources according to some subject terms, a typical query will result in a legend at the top, which is present in all search result screens to guide the user in identifying the type of the resource. The user can then drill down to retrieve further information about the resource. When adding a resource, the user is presented with an entry screen, which lists all possible resources that can be entered into the system.

Initial efforts focused on identifying many possible resources within the faculty. However, because of the modular nature of this system, further types of resources can easily be added to the repository without incurring any major design changes. As it is, the resources can be divided into two major virtual categories. The first would be termed as that of containing explicit resources or pointers to them, such as books and other published material. The second category contains the implicit resources, which are comprised of unpublished and little-known personal works, such as student papers, theses, and multimedia files, e.g., the audio and video of some lecture or presentation. The objective was to process categorized information to create knowledge for the repository users (Seng et al., 2002).

Figure 3. FISKR search options



S

FUTURE TRENDS: FISKR-II

FISKR has significant potential for the development and sharing of pedagogical resources within and between schools and departments in the University of Toronto, as well as organizations outside the university. This is mainly because of the structure provided to the data using XML, which will help facilitate data interchange. Commercially, the FISKR framework, as well as its tools and applications would be of great interest to organizations designing and developing their own internal knowledge repositories. The use of XML and object-oriented methodologies to mark up and represent the content of resources permits great breadth and flexibility in the kinds of resources that FISKR can house.

Some future enhancements to this system could be as follows:

- E-mails could be a part of the repository as they play an important role in capturing the elusive tacit knowledge flowing within the organization.
- For retrieving this (implicit) knowledge contained in embedded annotations from the knowledge base, the user can have the option of looking up annotations by made faculty members or renowned researchers, which would lead to availability of more credible information about the resource.
- The use of ontologies will bring about more organized storage within the knowledge base, which can then be exploited by an automated acquisition of explicit resources using Web wrappers.
- A Java-based tree-shaped structure of all the subject taxonomies present in the repository could help the users add nodes to a tree within the subject taxonomies, which have already been created by subject specialists.

CONCLUSION

Presented were the design concepts of a knowledge repository and one completed example. The experiences with this development process helped sum up the repository success factors as being: A repository should be envisioned to be a collection of shared resources, and the greater the information-handling capacity made shareable through the repository, the more valuable the system will be for its community of users. Moreover, the organization implementing such a project should plan to establish a long-standing focus on managing information content and providing "smart, friendly" access to information content. The prototype FISKR is a direct application of these principles to the organizational knowledge base, and benefits from all the areas of expertise that the faculty can bring to bear on it. This helped the project in meeting its ultimate goal of supporting a learning organization by capturing and sharing knowledge.

FISKR prototypes can be viewed at:

- **Original version:** http://www.fis.utoronto.ca/research/programs/fiskr
- **Customized version:** http://www.fis.utoronto.ca/fiskr/splash.htm

REFERENCES

Alavi, M. (2000). Systems for managing organizational knowledge. In M. F. Price (Ed.), *Framing the domains of IT management: Projecting the future through the past* (pp. 15-28). Cincinnati, OH: Pinnflex Educational Resources.

Anthes, G. (1998). Defending knowledge. *Computerworld, 32*(7).

Candler, J. W., Palvia, P. C., Thompson, J. D., & Zeltman, S. M. (1996). The Orion project staged business process reengineering at FedEx. *Communications of the ACM, 39*(2), 99-107.

Davenport, T. H., DeLong, D. W., & Beers, M. C. (1998, Winter). Successful knowledge management projects. *Sloan Management Review,* 43-57.

Davenport, T. H., & Prusak, L. (1998). *Working knowledge: How organizations manage what they know*. Boston: Harvard Business School Press.

Dieng, R., Corby, O., Giboin, A., & Ribiere, M. (1999). Methods and tools for corporate knowledge management. *International Journal of Human-Computer Studies, 51*, 567-598.

Grover, V., & Davenport, T. H. (2001, Summer). General perspectives on knowledge management: Fostering a research agenda. *Journal of Management Information Systems,* 5-21.

Hackbarth, G., & Grover, V. (1999). The knowledge repository: Organizational memory information systems. *Information Systems Management, 16*(3), 21-30.

Hall, D., Paradice, D., & Courtney, J. F. (2003). Building a theoretical foundation for a learning-oriented knowledge management system. *JITTA: Journal of Information Technology Theory & Application, 52*(2), 63-89.

Hickins, M. (1999). Xerox shares its knowledge. *Management Review, 88*(2), 40-45.

King, W. R., Marks, P. V. J., & McCoy, S. (2002). The most important issues in knowledge management. *Communications of the ACM, 45*(9).

Lehner, F., Mair, R. K., & Klosa, O. W. (1998). *Organizational memory systems—Applications of advanced database and network technologies* (Research Paper No. 19). University of Regensburg, Department of Business Informatics.

Maier, R. K., & Klosa, O. W. (1999, April). *Organizational memory systems to support organizational information processing.* Paper presented at the Proceedings of the 1999 ACM SIGCPR Conference on Computer Personnel Research, New Orleans, LA.

Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. Oxford University Press.

Seng, C. V., Zannes, E., & Pace, R. W. (2002). The contributions of knowledge management to workplace learning. *Journal of Workplace Learning Volume, 14*(4), 138-147.

Stein, E. W. (1992). A method to identify candidates for knowledge acquisition. *Journal of Information systems, 9*(2), 161-178.

Tobin, D. R. (1998). Networking your knowledge. *Management Review, 87*(4), 46-48.

KEY TERMS

Annotations: Comments, typographical corrections, hypotheses, or ratings given by a reader playing the role of an author, which makes a statement about the document or some part of it at a certain time.

Knowledge: The most valuable form of content in a continuum, starting at data, encompassing information, and ending at knowledge.

Explicit Knowledge: Knowledge that can be codified in words, numbers, or rules and can take the shape of documents, or perhaps result in the production of some type of equipment.

Implicit Knowledge: Knowledge that has been acquired through experience and task execution, where it helps individuals in performing their work.

Knowledge Management: Identification and analysis of available and required knowledge assets and related processes, and the subsequent planning and control of actions to use these assets and processes to achieve organizational objectives.

Knowledge Repository: A collection of meaningful data and information, including implicit and explicit data, which serves a particular community of practice.

Organizational Memory: When knowledge from the past is brought to bear on present activities, thus affecting the level of organizational effectiveness.

Organizational Memory Information Systems: A coherent integration of dispersed know-how from all over the organization.

ENDNOTE

1. Further information on this topic can be found in Dieng, Corby, Giboin and Ribiere's (1999) article, which provides a useful discussion and survey of methods, techniques, and tools aimed at managing corporate knowledge from a memory designer's perspective.

S

Signature-Based Indexing Techniques for Web Access Logs

Yannis Manolopoulos
Aristotle University of Thessaloniki, Greece

Alexandros Nanopoulos
Aristotle University of Thessaloniki, Greece

Mikolaj Morzy
Poznan University of Technology, Poland

Tadeusz Morzy
Poznan University of Technology, Poland

Marek Wojciechowski
Poznan University of Technology, Poland

Maciej Zakrzewicz
Poznan University of Technology, Poland

INTRODUCTION AND BACKGROUND

Web servers have recently become the main source of information on the Internet. Every Web server uses a Web log to automatically record access of its users. Each Web-log entry represents a single user's access to a Web resource (e.g., HTML document) and contains the client's IP address, the timestamp, the URL address of the requested resource, and some additional information. An example log file is depicted in Figure 1. Each row contains the IP address of the requesting client, the timestamp of the request, the name of the method used with the URL of the resource, the return code issued by the server, and the size of the requested object.

Conceptually, Web-log data can be regarded as a collection of clients' access-sequences, where each sequence is a list of Web pages accessed by a single user in a single session. Extraction of user access-sequences is a non-trivial task that involves data cleaning and user session tracking (Cooley, Mobasher & Srivastava, 1999). Figure 2 presents an example client sequence derived from the Web log from Figure 1.

Web logs are typically used to perform Web usage analysis, i.e., to identify and investigate user access-patterns. Such patterns can be discovered by means of popular data-mining algorithms for sequential pattern discovery (Chen, Park & Yu, 1998; Pei, Han, Mortazavi-Asl & Zhu, 2000). Each access pattern is a sequence of Web pages which occurs frequently in a collection of user access-sequences. Sequential access-patterns provide information about typical browsing strategies of users visiting a given Web site, for example, "10% of users

Figure 1. Example of a Web log

```
150.254.31.173  -- [21/Jan/2003:15:48:52 +0100] "GET /mmorzy " 301 328
150.254.31.173  -- [21/Jan/2003:15:48:52 +0100] "GET /mmorzy/index.html " 200 9023
150.254.31.173  -- [21/Jan/2003:15:48:52 +0100] "GET /mmorzy/acrobat.gif " 304
144.122.228.120 -- [21/Jan/2003:15:48:56 +0100] "GET /imgs/pp1.gif " 200 2433
150.254.31.173  -- [21/Jan/2003:15:48:58 +0100] "GET /mmorzy/research.html " 200 8635
60.54.23.11     -- [21/Jan/2003:15:48:59 +0100] "GET /mmorzy/db/slide3.htm " 200 24808
150.254.31.173  -- [21/Jan/2003:15:49:03 +0100] "GET /mmorzy/students.html " 200 7517
150.254.31.173  -- [21/Jan/2003:15:49:08 +0100] "GET /mmorzy/db_course.html " 200 9849
144.122.228.120 -- [21/Jan/2003:15:49:16 +0100] "GET /reports/repE.html " 200 76685
150.254.31.173  -- [21/Jan/2003:15:49:22 +0100] "GET /mmorzy/html.gif " 200 1038
150.254.31.173  -- [21/Jan/2003:15:49:22 +0100] "GET /mmorzy/zip.gif " 200 1031
144.122.228.120 -- [21/Jan/2003:15:50:03 +0100] "GET /imgs/polish.gif " 200 109
```

Figure 2. Example of a client's access sequence

```
/mmorzy/index.html → /mmorzy/research.html → /mmorzy/students.html →
/mmorzy/db_course.html
```

visited the page X, and later the page Y." After some frequently occurring sequences have been discovered, the analyst should be able to search for user access-sequences that contain the patterns. Such queries, called pattern queries, have numerous applications, for example, searching for typical/atypical user access-sequences.

Since Web logs tend to be large, a natural solution to support processing of pattern queries would be indexing Web access-sequences. Unfortunately, traditional indexing techniques for single-valued attributes, such as B-trees (Comer, 1979) or bitmap indexes (O'Neil, 1987), and multidimensional indexes, such as R-trees (Guttman, 1984), are inefficient or even not applicable, as they do not take ordering into consideration. These techniques can be applied merely to locate sequences built from the same set of elements as the query sequence, thus they are likely to introduce many false drops if the actual task is subsequence search. Therefore, novel dedicated indexing methods for Web-log data should be researched.

SEQUENTIAL DATA INDEXING METHODS FOR PATTERN QUERIES

Let I be a set of *items* (e.g., distinct URLs). A *data sequence* X is defined as an ordered list of items (e.g., client's access sequence). Thus, $X = \langle x_1 x_2 \dots x_n \rangle$, where each $x_i \in I$ (x_i is called an *element* of X). We say that a data sequence $X = \langle x_1 x_2 \dots x_n \rangle$ *is contained* in another data sequence $Y = \langle y_1 y_2 \dots y_n \rangle$ if there exist integers $i_1 < i_2 < \dots < i_n$ such that $x_1 = y_{i1}, x_2 = y_{i2}, \dots, x_n = y_{in}$. Given a database D of data sequences (e.g., Web-log file) and a data sequence Q, a *pattern query* consists in finding in D all data sequences that contain Q. In other words, a pattern query formulates a problem of finding all data sequences containing a set of user-defined elements in a specified order.

An *item mapping function* $f_i(x)$, where $x \in I$, is a function that transforms a literal into an integer value. Henceforth it is assumed that literals are mapped to consecutive positive integers starting from 1. An *order mapping function* $f_o(x,y)$, where $x, y \in I$ and $x \neq y \Rightarrow f_o(x,y) \neq f_o(y,x)$, is a function which transforms an item sequence $\langle x\, y \rangle$ into an integer value. It has to be noticed that the intuition for the use of $f_o(x, y)$ is that it takes ordering into account. Henceforth we consider order mapping functions of the form $f_o(x, y) = a * f_i(x) + f_i(y)$, where a is greater than all f_i values. Consequently, f_o values are always greater than f_i values, and the property $x \neq y \Rightarrow f_o(x,y) \neq f_o(y,x)$ is satisfied since $f_o(x,y) = f_o(w,z) \Leftrightarrow x = w \wedge y = z$. Given a sequence $X = \langle x_1 x_2 \dots x_n \rangle$, the *equivalent set* E of X is defined as:

$$E = \left(\bigcup_{x_i \in X} \{ f_i(x_i) \} \right) \cup \left(\bigcup_{x_i, x_j \in X,\, i<j} \{ f_o(x_i, x_j) \} \right)$$

where f_i is an item mapping function and f_o is an order mapping function.

According to the above formula, an equivalent set is the union of two sets: one resulting from considering each element separately, and one resulting from considering pairs of elements. $S(E)$ denotes the former set, consisting of values of f_i, and $P(E)$ denotes the latter set, consisting of values of f_o. A significant property of equivalent sets is that for two sequences Q, P and the corresponding equivalent sets E_Q and E_P, if Q is contained by P, then $E_Q \subseteq E_P$. Therefore, equivalent sets allow us to express a pattern query problem as the problem of finding all sets of items that contain a given subset.

Example 1. For instance, for $I = \{A, B, C, D, E\}$ we have $f_i(A) = 1, f_i(B) = 2, f_i(C) = 3, f_i(D) = 4, f_i(E) = 5$, and $f_o(x, y) = 6 * f_i(x) + f_i(y)$ (e.g., $f_o(A,B) = 8$). Given a sequence $X = \langle A, C, D \rangle$, using the mapping functions that were described above, we get: $E = (\{f_i(A)\} \cup \{f_i(C)\} \cup \{f_i(D)\}) \cup (\{f_o(A, C)\} \cup \{f_o(A, D)\} \cup \{f_o(C, D)\}) = \{1, 3, 4, 9, 10, 22\}$.

Equivalent sets can be efficiently represented with *superimposed signatures*. A signature is a bit string of L bits (L is called signature *length*) and is used to indicate the presence of elements in a set. Each element of a set can be encoded using a hash function into a signature that has exactly m out of L bits equal to '1' and all other bits equal to '0'. The signature of the whole set is defined as the result of the superimposition of all element signatures. In this approach it is assumed that m is equal to one, and the signature of the element x is the binary representation of the number $2^{x \bmod L}$. Given two equivalent sets E_1, E_2 and their signatures $sig(E_1)$, $sig(E_2)$, it holds that $E_1 \subseteq E_2 \Rightarrow (sig(E_1) \,\&\, sig(E_2)) = sig(E_1)$, where & is a bit-wise operator.

Signatures provide a quick filter for testing the subset relationship between sets. Therefore, if there exist any bits of $sig(E_1)$ that are equal to '1' and the corresponding bits of $sig(E_2)$ are not also equal to '1', then E_1 is not a subset of E_2. The inverse of the latter statement, however, does not hold in general. Evidently, *false drops* may result from collisions due to the superimposition. To verify a drop, we have to examine the corresponding sequences with the containment criterion.

Below we describe two selected signature-based sequential data indexing methods to optimize pattern queries: SEQ(C), which uses complete signatures of equivalent sets, and SEQ(A), which uses signatures of approximated equivalent sets.

A Simple Sequential Index (SEQ(C) – "Complete")

Let D be the database of sequences to be indexed. A simple data structure for indexing elements of D is based on the paradigm of the signature file (Faloutsos & Christodoulakis, 1984) and is called SEQ(C). It corresponds to the direct use of signatures of equivalent sets. The construction algorithm for SEQ(C) is given in Figure 3a.

Example 2. Let us consider the following example of SEQ(C) index entry construction. Assume the data sequence to be indexed is $X = \langle A, C, D, E\rangle$. Assume the set I of items and item mapping functions and order mapping functions from Example 1, and L=10.

The equivalent set for the data sequence X is the following: $E = \{f_i(A), f_i(C), f_i(D), f_i(E), f_o(A,C), f_o(A,D), f_o(A,E), f_o(C,D), f_o(C,E), f_o(D,E)\} = \{1, 3, 4, 5, 9, 10, 11, 22, 23, 29\}$.

Therefore, the SEQ(C) index entry will consist of the following signature (starting with the least significant bit): $sig(E) = 1111110001$.

The algorithm for querying the structure for a given sequence Q, is given in Figure 3b. Initially (Step 1), the equivalent set, E_Q, of Q is calculated. Then, each signature in the structure is examined against the signature $sig(E_Q)$ (Step 4). The verification of each drop is applied at Steps 5-7. The result, consisting of the sequences from D that satisfy query Q, is returned in set R.

Using Approximations of Equivalent Sets (SEQ(A) – "Approximate")

In Nanopoulos, Zakrzewicz, Morzy, and Manolopoulos (2003), a different method for organizing equivalent sets is proposed which significantly outperforms SEQ(C) thanks to reduction of the index size. It is based on the observation that the distribution of elements within sequential patterns is skewed, since the items that correspond to frequent subsequences have larger appearance frequency. Therefore, the pairs of elements that are considered during the determination of an equivalent set are not equiprobable.

Due to the above, some pairs have much higher co-occurrence probability than others. The sizes of equivalent sets can be reduced by taking into account only the pairs with high co-occurrence probability. This leads to approximations of equivalent sets, and the resulting method is denoted as SEQ(A). The objective of SEQ(A) is the reduction of the sizes of equivalent sets with a reduction of the lengths of the corresponding signatures.

The construction algorithm for SEQ(A) is depicted in Figure 4, where $supp_D(x, y)$ denotes the support of an ordered pair (x, y) in D (i.e., the normalized frequency of sequence $\langle x\ y\rangle$ (Agrawal & Srikant, 1995)), where $x, y \in I$ and the pair $(x, y) \in P(E)$. SEQ(A) reduces the

Figure 3. SEQ(C) method: (a) construction algorithm (b) search algorithm

```
1. F = Ø
2. forall  P ∈ D
3.    E = Equivalent_Set(P)
4.    F = F ∪ {<sig(E), pointer(P)>}
5. endfor
```

(a)

```
1. E_Q = Equivalent_Set(Q)
2. R = Ø
3. forall  <s, pointer(P)> ∈ F
4.    if (s & sig(E_Q)) = sig(E_Q)
5.              retrieve P from D
6.              if Q is contained in P
7.                        R = R ∪ {P}
8.              endif
9.    endif
10. endfor
```

(b)

sizes of equivalent sets by considering for each element $i \in I$, only k most frequent ordered pairs with i as the first element. In Step 2 of the algorithm, those frequent pairs are discovered and represented in the form of *NN* sets containing k most frequent successors for each element. In Steps 8-10, the *NN* sets are used to filter out infrequent pairs.

Example 3. Let us consider the following example of SEQ(A) index entry construction. Assume the data sequence to be indexed is $X = <A, C, D, E>$. Assume the set I of items and item mapping functions and order mapping functions from Example 1, k=1 and L=10. Given are the following support values for the element pairs: $supp_D(D,E) = 0.280$, $supp_D(E,B) = 0.220$, $supp_D(A,C) = 0.120$, $supp_D(C,E) = 0.101$, $supp_D(A,D) = 0.075$, $supp_D(C,D) = 0.053$, $supp_D(A,E) = 0.040$, $supp_D(B,A) = 0.037$, and for the other pairs, that is, (A,B), (B,C), (B,D), (B,E), (C,A), (C,B), (D,A), (D,B), (D,C), (E,A), (E,C), (E,D), the support values are less than 0.037.

Based on the support values of the element pairs, we construct the *NN* sets for each item. Each *NN* set contains only 1 item (k=1) which forms the strongest pair:

$NN(A) = \{C\}$, $NN(B) = \{A\}$, $NN(C) = \{E\}$, $NN(D) = \{E\}$, $NN(E) = \{B\}$.

The equivalent set will not represent the pairs that are not represented in the *NN* sets:

$E = \{f_i(A), f_i(C), f_i(D), f_i(E), f_o(A,C), f_o(C,E), f_o(D,E)\} = \{1, 3, 4, 5, 9, 23, 29\}$.

Therefore, the SEQ(A) index entry will consist of the following signature:

$sig(E) = 0101110001$.

The search algorithm of SEQ(A) is analogous to that of SEQ(C). However, Step 1 of the algorithm depicted in Figure 3b has to be modified accordingly (identical approximation has to be followed for the equivalent set of a query pattern):

1. E_Q = Equivalent_Set(Q)
1a. **forall** $(x, y) \in P(E_Q)$
1b. **if** *y* Ï*NN*(*x*)
1c. remove pair (x, y) from E_Q
1d. **endif**
1e. **endfor**

During the generation of the approximation of the query's equivalent set, the *NN* sets are used, which implies that they have to be kept in memory. However, this presents a negligible space overhead.

FUTURE TRENDS

Future research on sequence indexing should focus on extending the most promising technique introduced so far—the method based on approximations of equivalent sets. It may be interesting to consider alternative approximation schemes like: (1) global frequency threshold for

Figure 4. SEQ(A) method: construction algorithm

1. **forall** $i \in I$
2. find $NN(i) = \{i_j \mid i_j \in I, 1 \le j \le k, i_j \ne i, \forall\, l \notin NN(i)\ \ supp_D(i, i_j) \ge supp_D(i, l)\}$
3. **endfor**
4. $F = \emptyset$
5. **forall** $P \in D$
6. E = Equivalent_Set(P)
7. **forall** $(x, y) \in P(E)$
8. **if** $y \notin NN(x)$
9. remove pair (x, y) from E
10. **endif**
11. **endfor**
12. $F = F \cup \{<sig(E), \text{pointer}(P)>\}$
13. **endfor**

ordered pairs of elements within a sequence (using most frequent pairs instead of most frequent successors for each item), and (2) information-content measures (considering only pairs that carry more information than others).

CONCLUSION

In this article, the problem of efficient indexing of large Web access logs for pattern queries has been considered. We have described signature-encoding schemes, which are based on equivalent sets. Equivalent sets capture the ordering among the elements of access sequences, which is very important in the case of Web access logs. We have presented selected signature-based indexing methods build upon the concept of equivalent sets. Advanced topics on Web-log indexing, including application of signature trees to organize SEQ(C) and SEQ(A) indexes, can be found in Manolopoulos et al. (2004).

REFERENCES

Agrawal, R. & Srikant, R. (1995). Mining sequential patterns. In P.S. Yu & A.L.P. Chen (Eds.), *Proceedings of the 11th International Conference on Data Engineering* (pp. 3-14). Taipei, Taiwan: IEEE Computer Society.

Chen, M.S., Park, J.S. & Yu, P.S. (1998). Efficient data mining for path traversal patterns. *IEEE Transactions on Knowledge and Data Engineering*, *10*(2), 209-221.

Comer, D. (1979). The ubiquitous B-tree. *ACM Computing Surveys*, *11*(2), 121-137.

Cooley, R., Mobasher, B. & Srivastava, J. (1999). Data preparation for mining World Wide Web browsing patterns. *Knowledge and Information Systems*, *1*(1), 5-32.

Faloutsos, C. & Christodoulakis, S. (1984). Signature files: An access method for documents and its analytical performance evaluation. *ACM Transactions on Office Information Systems*, *2*(4), 267-288.

Guttman, A. (1984). R-trees: A dynamic index structure for spatial searching. In B. Yormark (Ed.), *SIGMOD'84, Proceedings of Annual Meeting* (pp. 47-57). Boston: ACM Press.

Manolopoulos, Y., Morzy, M., Morzy, T., Nanopoulos, A., Wojciechowski, M. & Zakrzewicz, M. (2004). Indexing techniques for Web access logs. In D. Taniar & J.W. Rahayu (Eds.), *Web information systems* (pp. 309-338). Hershey, PA: Idea Group Publishing.

Nanopoulos, A., Zakrzewicz, M., Morzy, T. & Manolopoulos, Y. (2003). Efficient storage and querying of sequential patterns in database systems. *Information and Software Technology*, *45*(1), 23-34.

O'Neil, P.E. (1987, September 28-30). Model 204 architecture and performance. In D. Gawlick, M.N. Haynie & A. Reuter (Eds.), *Proceedings of High Performance Transaction Systems, 2nd International Workshop* (HPTS 1987) (pp. 40-59). Asilomar Conference Center, Pacific Grove, California.

Pei, J., Han, J., Mortazavi-Asl, B. & Zhu, H. (2000, April 18-20). Mining access patterns efficiently from Web logs. In T. Terano, H. Liu & A.L.P. Chen (Eds.), *Proceedings of Knowledge Discovery and Data Mining, Current Issues and New Applications, 4th Pacific-Asia Conference* (PAKDD 2000) (pp. 396-407), Kyoto, Japan.

KEY TERMS

Data Mining: A database research area that aims at automated discovery of non-trivial, previously unknown, and interesting regularities, trends, and patterns in large data sets.

Database Index: An auxiliary physical database structure that is used to speed up the retrieval of data objects from the database in response to certain search conditions. Typically, indexes are based on ordered files or tree data structures.

Sequential Pattern: Frequently occurring subsequence in a collection of sequences.

Signature Index: A database index that uses signatures to represent values of the indexed data objects. Signatures are typically encoded as bit strings.

URL: Uniform Resource Locator, the global address of a Web page or other resource on the Internet. A URL has three basic parts: the protocol identifier, IP address or domain name of the server, and the relative path to the resource on the server.

Web Access Log: A text file into which a Web server saves information about every request that it handles.

Web Server: A computer program running on a machine permanently connected to the Internet, dedicated to serving requested Web pages to the user's browser (client).

Similarity Web Pages Retrieval Technologies on the Internet

Rung Ching Chen
Chaoyang University of Technology, Taiwan

Ming Yung Tsai
Chaoyang University of Technology, Taiwan

Chung Hsun Hsieh
Chaoyang University of Technology, Taiwan

INTRODUCTION

In recent years, due to the fast growth of the Internet, the services and information it provides are constantly expanding. Madria and Bhowmick (1999) and Baeza-Yates (2003) indicated that most large search engines need to comply to, on average, at least millions of hits daily in order to satisfy the users' needs for information. Each search engine has its own sorting policy and the keyword format for the query term, but there are some critical problems. The searches may get more or less information. In the former, the user always gets buried in the information. Requiring only a little information, they always select some former items from the large amount of returned information. In the latter, the user always re-queries using another searching keyword to do searching work. The re-query operation also leads to retrieving information in a great amount, which leads to having a large amount of useless information. That is a bad cycle of information retrieval. The similarity Web page retrieval can help avoid browsing the useless information. The similarity Web page retrieval indicates a Web page, and then compares the page with the other Web pages from the searching results of search engines. The similarity Web page retrieval will allow users to save time by not browsing unrelated Web pages and reject non-similar Web pages, rank the similarity order of Web pages and cluster the similarity Web pages into the same classification.

In this article, we will introduce the technology of similarity Web page retrieval which includes searching Web pages and classifying similarity Web pages. For searching Web pages, we will specify the types of search engines and the policy of ranking search results first. And then, two algorithms used for finding related Web pages, the Cocitation algorithm and Latent Linkage Information (LLI) algorithm, will be introduced. The classification method can make the Web pages with similarity vector attributed will be clustered in the same class, and the user retrieves the related Web pages more effectively.

BACKGROUND

The Web Pages Searching

With the growing information of the Internet, search information efficiently is more important than ever to do. Search engines are used to find the Web pages with the same keyword on the Internet (Chakrabarti, Joshi, & Tawde, 2001; Wen-Syan & Candan, 2002). Only by keypressing a keyword, search engines can return information quickly and easily. Jenkins classifies the search engines into three categories (Barroso, Dean, & Holzle, 2003; Jenkins, Kackson, Burden, & Wallis, 1998).The categories include the classified directories search engines, automated search engines, and metasearch engines. The classified directories search the engines' classification of Web data by human. The automated search engines collect the Web resources from the WWW automatically. The metasearch engines provide interface that receives the user query and merges the searching results from various search engines. For the three types of searching, engines will be described briefly as follows (Madria & Bhowmick, 1999). The classified directories' search engines collect the Web data from the WWW by the editorial staff and describe the Web site with text mention. The editors must classify the Web pages to suitable category. Therefore, a user can find the useful information by classification directory easily (Baeza-Yates, 2003). The automated search engines employ the retrieval programming (called Robot or Spider) to collect the Web data from WWW voluntarily, construct the URL (Uniform Resource Locator) index and title, and provide that to the user (Wen-Syan & Candan, 2002). A metasearch engine uses a concept of a computer agent, between the users and many search engines. Its responsibility is to make the users search more conveniently for various aspects of the search engines (Tirri, 2003). A metasearch engine receives user's query and passes the query to various searching engines, if reformatting the query is needed, and then collects and rearranges the results from the search engines (Madri & Bhowmick, 1999; Baeza-Yates, 2003; Wen-Syan & Candan, 2002). Henzinger, Motwani, and Silverstein indicate a number of reasons why the metasearch engines are needed.

The Web Mining

The mining technology is different from general search work. The mining method can find the data in horizontal relation, and the relation between data and data, that general searching work can not do. When mining technologies are applied to finding Web pages, we called it Web mining. The Web mining framework includes three types which are: Web content mining, Web structure mining, and Web usage mining (shown in Figure1) (Flake, Lawrence, Giles, & Coetzee, 2002; Kao, & Lee, 2000; Madria & Bhowmick, 1999; Sundaresan, & Yi, 2000; Zhang, & Dong, 2002).

In traditional search technologies, the user sends a keyword to a search engine, and then the search engine uses the keyword to find related pages, but it spends too much time and returns too many irrelevant pages. Therefore, the hyperlink analysis and anchor text analysis are proposed to find the related pages. Using the hyperlink structure to find the relative pages can find the related pages quickly and easily, but hyperlink structures have no content information. Hence, there are two assumes in hyperlink study. When a page A links to another page B, indicating page B is recommended by author page A. If two pages have the same links, they might be on the same topic (Henzinger, 2000; Qiu, Hemmje, & Neuhold, 2000; Zhengyu, Qingsheng, & Yukun, 2001). Hence, the hyperlink analysis can be used to ranking pages, Web pages community construction, Web searching improvement, Web clustering and visualization, as well as finding the related pages. Tirri thinks the earlier Web search technologies are not good enough by keyword searching, and he used ontology and hyperlink analysis knowledge to do the searching. He considers the next generation of Web search must have the personal search results, easily find the information that users want and make synonym by itself (Hodgson, 2001; Singh, 2002; Tirri, 2003). We described the similarity Web pages searching technology in the following methods.

(1) Cocitation Method

Dean and Henzinge proposed the Cocitation algorithm that uses the hyperlink structure to find the related pages (Dean,& Henzinger, 1999; Jingyu, & Yanchun, 2003). The algorithm is based on two definitions: (A) The two pages are co-cited if they have a common parent page; (B) The number of their common parent pages, named degree of cocitation. In this algorithm, a user can use a query term for a search engine to find a require page, and then can use a URL of the require page to construct a vicinity graph by hyperlink, and then can analyze the graph to find the related pages of the required page. Hence, the algorithm employed the hyperlink of Web pages to find the related page. It is simple and computes easily, but when two pages have the same degree of cocitation, the system cannot decide which Web pages are actually related pages. Also, it may have a topic drift problem if the anchor of a Web page constructs malice hyperlink purposely. The similarity measures between two pages are shown in Figure 2. The page “u” and page “a” have the same parent page A. They are cocited, and the degree of the cocitation is 1. In the same way, the in degree of Page B is 2. Hence, the degree of all pages that related to pages u can be computed.

(2) Latent Linkage Information (LLI) Algorithm

Although, the Cocitation algorithm can evaluate the similarity of Web pages, it cannot determine which pages are better when the pages have the same degree. Therefore, Hou and Zhang utilize the linkage matrix to express the relevance of the Web page topology. The matrix can reveal the depth relationships of the pages, and then find the

Figure 1.Web mining classification

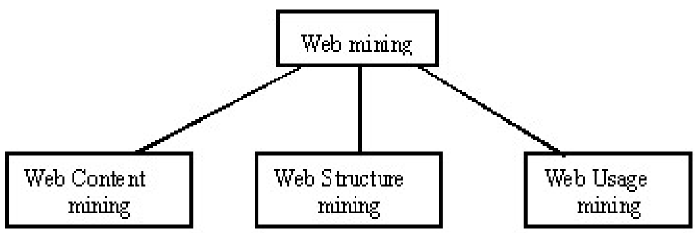
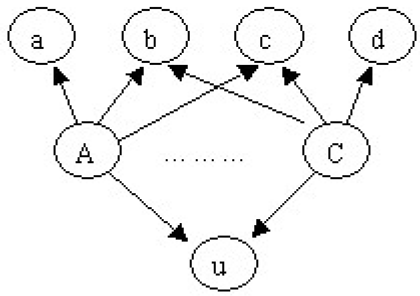


Figure 2. In degree of b is two



relative pages easily. In order to reveal the deep relationships, the singular value decomposition (SVD) concept is used in linear algebra. In the same way, it extends the Cocitation algorithm to construct the direction graph with parent pages and children pages (FS and BS). The topological relationships between BS and P_u are expressed at a linkage matrix A. The topological relationships between FS and C_u are expressed at a linkage matrix B as follows:

$$A = (a_{ij})_{mxn},$$

$$a_{ij} = \begin{cases} 1, when.page.i.is.a.child.of.page.j,\ page.i \in BS,\ page.j \in P_u \\ 0, otherwise \end{cases}$$

$$B = (b_{ij})_{pxq},$$

$$b_{ij} = \begin{cases} 1, when.page.i.is.a.parent.of.page.j,\ page.i \in FS,\ page.j \in C_u \\ 0, otherwise \end{cases}$$

Hence, the ith row of matrix A can be considered as the coordinate vector of page i (BS) in an n dimensional space; the jth column of matrix A can be considered as the coordinate vector of page j (P_u) in the m dimensional space. In the same way, the ith row of matrix B can be considered as the coordinate vector of page i (FS) in a q dimensional space; the jth column of matrix B can be considered as the coordinate vector of page j (C_u) in a p dimensional space. In other words, it can transfer the matrix A and B to find the topological relationships in different multi-dimensional spaces.

(3) PageRank Agorithm

Brin and Page proposed the PageRank algorithm that uses the in-link structure to rank the results for users (Brin & Page, 1998). When a page B links to page A, it indicates page B is important to page A. For example, if a paper is cited by many other papers, the paper is more important. The PageRank algorithm extends this idea to use the in-link structure to rank the results of search engines. The PageRank is defined as follows:

$$PR(i) = (1-d) + d(PR(T_1)/C(T_1) + ... + PR(T_n)/C(T_n)),$$

where *PR(Ti)* is the page rank of pages *Ti* which links to page A; $C(T_n)$ is the number of the out-link of page T_n and function d is a damping factor which can be set between 0 and 1.

The $PR(i)$ is computed by Web structure, and it can be used to rank the results of Web search. The PageRank algorithm is used to monitor a model of user's behavior. There is a "random surfer" who browses the Web random and goes to another page by links until he gets bored and starts on another page again. The value d is the probability at each page that the "random surfer" will get bored and go to another random page. Therefore, the PageRank can easily compute and rank the result in search engine by hyperlinks (Satronov & Parashar, 2003).

Web Pages Classification

Users usually input a keyword to a search engine to find the related Web pages and often return the massive of Web pages. Users hope to get requiring pages from the returned results, but they contain too many irrelative Web pages; that makes the problem worse. To decide which Web pages are required wastes much time. The data classification method is applied to many fields, it can help user quickly get related information. Previously, WWW was not popular, the Web pages information does not propagate fast, so the system can use artificial method to classify Web pages (Asirvatham & Ravi, 2001). Now, the Web pages expand quickly, so using artificial method to classify Web pages has no efficiency. Therefore, using information technology to automatic classification Web pages is needed. There are many methods using Web pages classification such as Naïve Bayes, k-NN approach, inductive rule learning, neural network, decision tree (Chan, Tokuda, & Akira, 2003; Kwon & Lee, 2003; McCallum, & Nigam, 1998) and support vector machine (Dumais & Chen, 2000). Before the classification operation, the features extraction is necessary. In general, there is the Boolean model, the vector space model, and the probability model (Yates & Ribeiro-Neto, 1999) for features extraction methods. We describe the information retrieval model first.

The information features extraction has three common models: Boolean model, vector space model, and probability model. We will introduce the three models and then weight the keywords through TF-IDF. On the other hand, the method of Web page classification will be a brief specification, too.

(1) Boolean Model

Boolean model is based on fuzzy set theory and Boolean algebra classification (Yates& Ribeiro-Neto, 1999). The weight of a keyword is represented as binary codes.

$W \in \{0,1\}$, 0 stands for keyword W not appeared;
1 stands for keyword W appeared.

A query document q, after the transform Boolean function calculates similarity between classified document d and q, has two results:

1: stands for query document q related to document d.
0: stands for query document q not related to document d.

(2) Vector Space Model

This model uses keywords vector to estimate the angle between query document and classified documents and to determine the query document belonging to which category.

A query vector: $q = (W_{1q}, W_{2q}, .., W_{tq})$
q:query document
t:total of keywords in query document q
Classified document vector: $d = (W_{1d}, W_{2d}, .., W_{td})$
d: already classified document vector.

The similar measure between d and q can be estimated by the cosine angle.

(3) Probability Model

The probability model uses Bayes theorem to determine the documents' category. A query document q is an estimated probability values between document and document. The form is as follows:

$$\frac{P(d\ relate\ to\ q)}{p(d\ non-relate\ to\ q)}$$

For example, if document D have n keywords W_{i1}, W_{i2}, W_{in}, the probability of D belong to C_I is:

$$P(C_I \mid W_{i1}, W_{i2}, .., W_{in})$$

According to Bayesian theorem:

$$P(C_I \mid W_{i1}, W_{i2}, .., W_{in}) = \frac{P(C_I) \times P(W_{i1}, W_{i2}, .., W_{in} \mid C_I)}{P(W_{i1}, W_{i2}, .., W_{in})}.$$

It can calculate the probability of D belonging to each category $C_{I.}$ The category that has maximum of probability value belongs to query document.

(4) TF-IDF

In documents, the more keywords appear, the more important the keywords are. Generally, the importance of a keyword in a document considers two factors. One is the frequency of the keyword in the document, named Term Frequency (TF), the other is the numbers of this keyword appearing in all documents, called Document Frequency (DF). The high value of TF corresponds with the high importance of the keyword in the document. The product of TF and Inverse Documents Frequency (IDF) is used to represent the importance of the keyword in a document, named TF-IDF.

$W_{i,j} = TF_{i.j} \times IDF_i$
$W_{i,j}$: it is the importance of the keyword i in document j.

$$TF_{ij} = \frac{n_j}{n_{all}},$$

where n_j is the times of the keyword *j* appear in thesis *i* and n_{all} is the total TF that is meaningful in thesis *i*. TF_{ij} is the frequency of a keyword *j* in thesis *i*.

$IDF_i = \log\left(\frac{N}{n_i}\right)$, where N is the number of documents and n_i is the frequency of keyword i in all documents.

(5) Web Page Classification Methods

Presently, there are many Web page classification approaches, such as k-NN, neural network method. The k-NN method (Kwon & Lee, 2003) calculates the likelihood between the category and relevant Web page. In order to improve efficiency, have joined the feature selection, used HTML tags and new similarity degree to evaluate the Web page classification. The neural network method (Liu & Zhang, 2001; Selamat & Omatu, 2004) do the preprocessing of the Web page, include stemming and moved the stop words. Then a portion is to reduce the dimension of feature vector (Lam & Lee, 1999), another portion is to each category extraction the keyword and assign the weight. Combine the two types of feature vectors and input to neural network for training. According the neural training results, the system can classify the Web pages into suitable classes.

FUTURE TRENDS

Using the content and the link information features is one of a tendency for finding the related Web pages in the future researches. The content information includes the, tag, and real word and meta information. In WWW, Web pages usually use HTML language to compose the Web pages content which contains tags, out links, and descript terms metadata. The HTML language uses twin tags to represent the characters. However, the Web pages represent the same content with traditional documents when the tags are removed. For HTML tags, each tag has various functions. For examples, the tags pair <title> and </title> stands for the words of title between the tags, as well as the tags pair <table> and </table> stands for a table form between the tags. If these tags are analyzed, we also can get much useful information. The out-links and in-links also contain much information. In HTML file, the metadata described terms that would not appear in Web pages, but it can help the content of the Web pages understand. For example, the metadata of code type description can allow the system to understand which language is used in the Web page. So, how to effectively assemble these features is an interesting topic for the similarity Web pages retrieval.

On the other hand, precisely classifying the Web pages to the correct category also makes the similarity Web pages retrieval effectively. How to correct classification of the Web pages automates by the feature of Web content is also a hot research topic. When we input a keyword in a search engine to do a search, we can get a lot of Web pages. However, the search engine cannot translate the keyword into various languages which make the search results not to stretch out various nation information. Appling the keyword translation technology to a search engine is also a public research direction.

In addition, when two pages have the same present pages, they might be the related pages; we can extend this idea to the Web pages having the same child pages—they might be the related pages. Moreover, the topic discovery is a new technology that finds the related pages. It will automatically cluster the related Web pages together by the representative topics. From representative topics, we can easily determine the related Web pages. Therefore, many of the searching results will be transferred to various classes, and then we find the related pages are easy and do not worry about the return of too much information. Finally, the similarity Web searching will be extended to the functions of multimedia content similarity Web pages retrieval. For example, to do the similarity images retrieves on the Internet.

CONCLUSION

The information on the Internet is growing fast. How to efficiently search is an important research theme. The hyperlink analysis can find the relative pages easily, but it is not precise. Hence, combining the Web content search and the Web structure search finding the relative pages is the trend of study. Many researchers use the hyperlink graph of pages and consider the content of Web pages to find the related pages. In a different way, using the HTML tags to find the related Web pages is also a worthy research direction. The Web pages classification automatically is also a direction to find the similarity Web pages. Many methods are mentioned in the article. The similarity Web pages will be an interesting research topic for scholars to do the research.

REFERENCES

Asirvatham, A. & Ravi, K. (2001). Web page classification based on document structure. *International Institute of Information Technology Hyderabad*, India 500019.

Baeza-Yates, R. (2003). Information retrieval in the Web beyond: Beyond current search engines. *Approximate Reasoning*, *34*, 97-104.

Barroso, D. & Holzle. (2003). Web search for a planet: The Google cluster architecture. *Micro IEEE*, *23*(2), 22-28.

Brin, S. & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems, 30*(1-7),107-117.

Chakrabarti, S., Joshi, M. & Tawde, V. (2001). Enhanced topic distillation using text, markup tags, and hyperlinks. Proceedings of the 24th Annual International ACM SIGIRC Conference on Research and Development in information Retrieval *New Orleans, Louisiana, USA*, 208-216.

Chan, L., Tokuda, N. & Akira, N. (2003). A new differential LSI space-based probabilistic document classifier. *Information Processing Letters*, *88*, 203-212.

Dean, J. & Henzinger, M. (1999). Finding related pages in the World Wide Web. *IEEE Computer Networks*, *31*(11-16), 1467-1479.

Dumais, S. & Chen, H. (2000). Hierachical classification of Web content. *Proceedings of the 23rd annual international conference on research and development in information retrieval (SIGIR2000)*, 256-263.

Flake, G., Lawrence, S., Giles C. & Coetzee, F. (2002). Self-organization and identification of Web communities. *IEEE Computer*, *5*(3), 70-99.

Henzinger, M.R. (2001). Hyperlink analysis for the Web. *IEEE Internet Computing*, *5*(1), 45-50.

Henzinger, R., Motwani, R. & Silverstein G.. (2002). Challenges in Web search engines. *SIGIR Forum*, *36*(2).

Hodgson, J. (2001). Do HTML tags flag semantic content? *IEEE Internet Computing*, *5*(1), 20-25.

Jenkins, C., Kackson, M., Burden, P. & Wallis, J. (1998). Searching the World Wide Web: An evaluation of available tools and methodologies. *Journal on Information and Software Technology*, 39, 985-994.

Jingyu, H. & Yanchun, Z. (2003). Finding relevant Web pages from linkage information. *IEEE Computer Society*, 15(4), 940-951.

Kao, B. & Lee, J. (2000). Anchor point indexing in Web document retrieval. *IEEE Transactions*, *30*(3), 364-373.

Kwon, O., Lee, J. (2003). Text categorization based on k-nearest neighbor approach for Website classification. *Information Processing and Management*, 25–44.

Lam, S. & Lee, D. (1999, April). Feature reduction for neural network based text categorization. *Proceedings of the 6th International Conference on Database Systems for Advanced Applications, Hsinchu, Taiwan.*

Liu, Z. & Zhang, Y. (2001). A competitive neural network approach to Web page categorization. *International Journal of Uncertainty, Fuzziness and Knowledge Based System*, *6*, 731-741.

Madria, S. & Bhowmick, S. (1999). Research issues in Web data mining. *Proceedings of Data Warehousing and Knowledge discovery. First International Conference on Research Issues in Web Data Mining, 39* (pp. 303-312).

McCallum, A. & Nigam, K. (1998). A comparison of event models for Naïve Bayes text classification. *AAAI-98 Workshop on Learning for Text Categorization.*

Qiu, Z., Hemmje, M. & Neuhold, E. (2000). Using link-based domain models in Web searching**.** *Digital Libraries: Research and Practice*. 13(16),152-159.

Satronov, V. & Parashar, M.(2003). Optimizing Web servers using page rank prefetching for clustered accesses. *Information Sciences*, *150*, 165-176

Selamat, A. & Omatu, S. (2004). Web page feature selection and classification using neural networks. *Information Sciences*, *158*, 69-88.

Singh, M.P. (2002). Deep Web structure. *IEEE Internet Computing*, *6*(5), 4-5.

Sundaresan, N. & Yi, J. (2000). Mining the Web for relations. *Computer Networks*, *33*, 699-711.

Tirri, H. (2003). Search in vain: Challenges for Internet search computer. *Computer*, *36*(1), 115-116.

Wen-Syan, L. & Candan, K. (2002). Query relaxation by structure and semantics for retrieval of logical Web document. *IEEE Knowledge and Data Eng.*, *14*(4), 768-791.

Zhang, D. & Dong, Y. (2002). A novel Web usage mining approach for search engines. *Computer Networks. 39*, 303-310.

Zhengyu, Z., Qingsheng, Z. & Yukun, C. (2001). A new semantic expression method for Web pages based on OCDB. *Computer Networks and Mobile Computing*, 16(19), 8-12.

Yates, R. & Ribeiro-Neto, B. (Eds.). (1999). *Modern information retrieval*. Addision Wesley Longman Limited.

KEY TERMS

Classification: A technique dividing a dataset into mutually exclusive groups. Unlike clustering, classification relies on predefined classes.

Clustering: A technique that uses features to find the linking pages of each other automatically. Usually, the Web pages with near features will be clustered together.

Cocited: The Web pages have the same parent page or have the same child pages called the Web pages are cocited.

In Degree: The number of Web pages that link to current Web page.

Information Retrieval: A technology which can retrieve useful information effectively for some themes from the Internet by artificial or automatic method.

k-Nearest Neighbor: A technique that classifies each Web page in a set based on a combination of the classes of the k records most similar to it.

Machine Learning: The purpose of machine learning is how to make computers learning, can let computer to do something. The goal is to make computers improve the performance.

Out Degree: The number of Web pages that current Web page point out.

Simple Methods for Design of Narrowband High-Pass FIR Filters

Gordana Jovanovic-Dolecek
National Institute for Astrophysics, Optics, and Electronics, Mexico

INTRODUCTION

Digital signal processing (DSP) is an area of engineering that "has seen explosive growth during the past three decades" (Mitra, 2001). Its rapid development is a result of significant advances in digital computer technology and integrated circuit fabrication (Smith, 2002; Jovanovic-Dolecek, 2002). Diniz, Silva, and Netto (2002) state that "the main advantages of digital systems relative to analog systems are high reliability, suitability for modifying the system's characteristics, and low cost."

The main DSP operation is digital signal filtering, that is, the change of the characteristics of an input digital signal into an output digital signal with more desirable properties. The systems that perform this task are called digital filters. The applications of digital filters include the removal of the noise or interference, passing of certain frequency components and rejection of others, shaping of the signal spectrum, and so forth (White, 2000; Ifeachor & Jervis, 2001).

Digital filters are divided into finite impulse response (FIR) and infinite impulse response (IIR) filters. FIR digital filters are often preferred over IIR filters because of their attractive properties, such as linear phase, stability, and the absence of the limit cycle (Mitra, 2001; Diniz, Silva & Netto, 2002). The main disadvantage of FIR filters is that they involve a higher degree of computational complexity compared to IIR filters with equivalent magnitude response (Mitra, 2001; Stein, 2000). In past years, many design methods have been proposed to reduce the complexity of FIR filters (Jou, Hsieh & Kuo, 1997; Kumar & Kumar, 1999; Webb & Munson, 1996; Lian & Lim, 1998; Bartolo & Clymer, 1996; Kuo, Chien & Lin, 2000; Abeysekera & Padhi, 2000; Yli-Kaakinen & Saramaki, 2001; Coleman, 2002; Jovanovic-Dolecek, 2003).

We consider high-pass (HP) linear-phase narrowband filters. It is well known that one of the most difficult problems in digital filtering is the implementation of narrowband filters. The difficulty lies in the fact that such filters require a high-order design with a large amount of computation, making them difficult to implement (Mitra, 2001; Grover & Deller, 1999; Stein, 2000).

We propose an efficient design of high-pass linear-phase narrowband digital filters based on the corresponding low-pass (LP) filter. In this design we use the interpolated FIR (IFIR) structure and the sharpening recursive running sum (RRS) filter (Jovanovic-Dolecek, 2003). In the next section we describe the transformation of a low-pass into a high-pass filter, followed by descriptions of an IFIR structure, RRS filter, and the sharpening technique. Finally, we present the design procedure, along with an example.

BACKGROUND

Transformation of LP into HP Filter

Instead of designing a high-pass filter by brute force, we can transform it into a low-pass filter. We replace the desired cutoff frequencies of the high-pass filter ω_p and ω_s, by the corresponding low-pass specifications as follows:

$$\begin{aligned} \omega_p^{'} &= \pi - \omega_p \\ \omega_s^{'} &= \pi - \omega_s \end{aligned} . \quad (1)$$

Given these specifications, a low-pass FIR filter can be designed. From this auxiliary low-pass filter, the desired high-pass filter can be computed by simply changing the sign of every other impulse response coefficient. This is compactly described as:

$$h_{HP}(n) = (-1)^n h_{LP}(n), \quad (2)$$

where $h_{HP}(n)$ and $h_{LP}(n)$ are the impulse responses of the high-pass and the low-pass filters, respectively. In that way the IFIR structure proposed for design of LP filters can also be used for HP filters.

IFIR Structure

The interpolated finite impulse response structure, proposed by Neuevo, Dong, and Mitra (1984), is an efficient realization of a high-order liner-phase LP FIR filter. Instead of designing one high-order liner-phase filter $H(z)$, two lower order liner-phase LP filters are computed. One of them is called the shaping or model filter $G(z)$, and the other one is the interpolator filter $I(z)$.

Suppose that the specifications of the original filter $H(z)$ are: pass-band edge ω_p, stop-band edge ω_s, pass-band ripple R_p, and minimum stop-band attenuation A_s. The specification of the LP filter $G(z)$ can then be expressed as follows:

$$\begin{aligned} \omega_p^G &= M\omega_p \\ \omega_s^G &= M\omega_s \\ R_p^G &= R_p / 2\,, \\ A_s^G &= A_s \end{aligned} \tag{3}$$

where M is the interpolation factor, and the upper index G stands for the filter $G(z)$. The expanded filter $G(z^M)$ is obtained by replacing each delay z^{-1} in the filter $G(z)$ with z^{-M} . In the time domain this is equivalent to inserting M-1 zeros between two consecutive samples of the impulse response of $G(z)$. The expansion of the filter $G(z)$ introduces M-1 images in the range $[0, 2\pi]$, which have to be eliminated. This is why the filter interpolator $I(z)$ is needed. The general low-pass IFIR structure is given in Figure 1.

The high-pass filter design depends on the parity of the interpolation factor. For M even, there is an image at high frequency. If all other images along with the original spectrum are eliminated, the high-pass filter results. Therefore, the interpolator filter is a high-pass filter with the following specifications (Jovanovic-Dolecek, 2003):

$$\begin{aligned} \omega_p^I &= \pi - \omega_p^{'} \\ \omega_s^I &= \frac{2\pi}{M}\frac{M-2}{2} + \omega_s^{'} = \frac{\pi(M-2)}{M} + \omega_s^{'} \\ R_p^I &= R_p / 2 \qquad\qquad , \\ A_s^I &= A_s \end{aligned} \tag{4}$$

where upper index I stands for the interpolator $I(z)$. Now, the resulting IFIR filter is the desired high-pass filter.

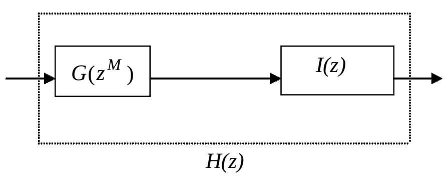
For M odd, however, there are no images at high frequency. In order to obtain the low-pass filter, all images have to be eliminated using the interpolator filter, so that only the original spectrum remains. In this case the interpolator $I(z)$ has the following specifications (Jovanovic-Dolecek, 2003):

$$\begin{aligned} \omega_p^I &= \pi - \omega_p^{'} \\ \omega_s^I &= \frac{2\pi}{M} - \omega_s^{'} \\ R_p^I &= R_p / 2 \qquad . \\ A_s^I &= A_s . \end{aligned} \tag{5}$$

Since the resulting IFIR filter is a low-pass filter, we apply transformation (2) to achieve the desired high-pass filter.

To further simplify the proposed design, a simple RRS filter, described in the next section, is used as the interpolator in the IFIR structure.

Figure 1. IFIR structure



Recursive Running-Sum Filter

The simplest low-pass FIR filter is the moving-average (MA) filter. Its impulse response $g(n)$ is given by:

$$g(n) = \frac{1}{M}\sum_{k=0}^{M-1} g(n-k) . \tag{6}$$

All impulse response coefficients are equal to 1, and so the filter requires no multiplications.

Its transfer function is given by:

$$G(z) = \frac{1}{M}[1 + z^{-1} + ... + z^{-(M-1)}] = \frac{1}{M}\sum_{k=0}^{M-1} z^{-k} \quad . \tag{7}$$

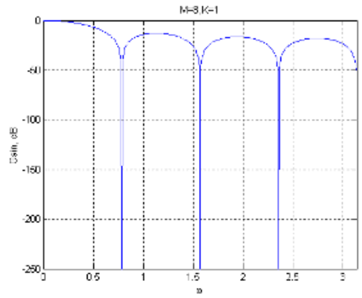
A more convenient form of the above transfer function for realization purposes is given by:

$$H_{RRS}(z) = [G(z)]^K \;= \left[\frac{1}{M}\frac{1-z^{-M}}{1-z^{-1}}\right]^K , \tag{8}$$

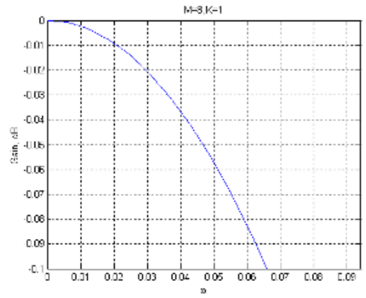
which is also known as a recursive running-sum filter (RRS) (Mitra, 2001), or a boxcar filter. The scaling factor $1/M$ is needed to provide a dc gain of 0 dB, and K is the number of the cascaded sections of the filter. The magnitude response of the filter can be expressed as:

$$\left|H_{RRS}(e^{j\omega})\right| = \left| \; \left[\frac{\sin(\omega M/2)}{M\sin(\omega/2)}\right]^K \right| . \tag{9}$$

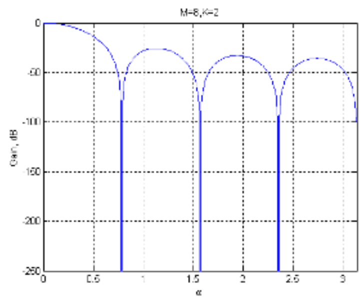
Figure 2. RRS filters for M=8 and K=1, 2, and 4



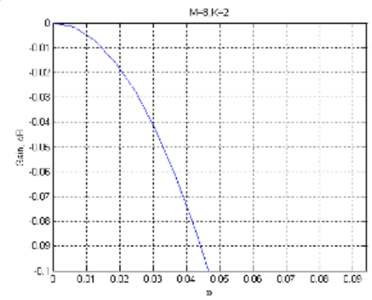
a. Overall magnitude response



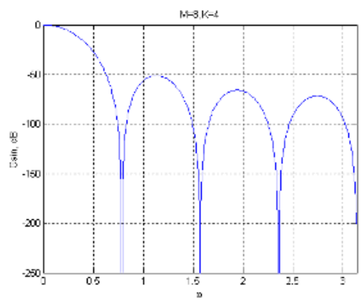
b. Pass-band zoom



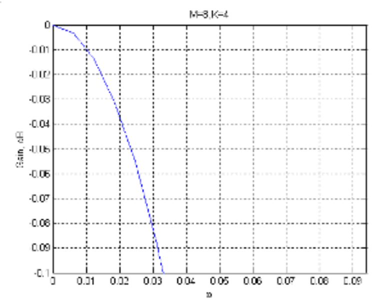
c. Overall magnitude response



d. Pass-band zoom



e. Overall magnitude response



f. Pass-band zoom

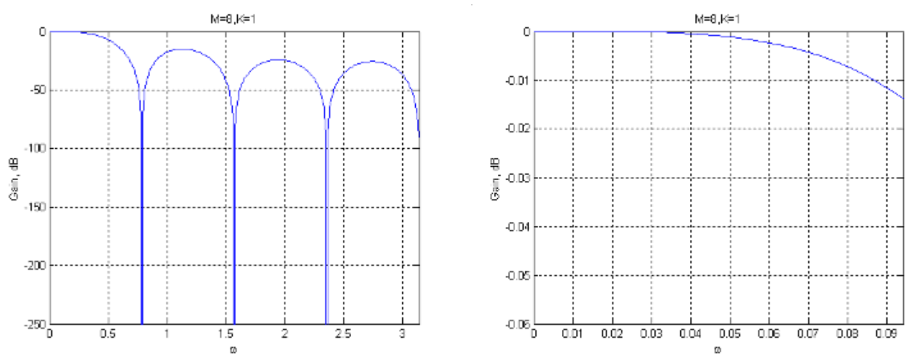
The magnitude response has nulls at integer multiples of ($2\pi/M$). This provides natural alias-rejections introduced by expanding the shaping filter $G(z)$ by factor M. The stop-band attenuation is improved as a result of the increased number of stages K. However, since the pass-band is downgraded, the RRS filter—when used as the interpolator in an IFIR structure—has limited application. Figure 2.a. illustrates the overall magnitude responses and the corresponding pass-band details of the RRS filter for M=8 and K=1, 2, 3, and 4.

In the next section we show how to improve the characteristics of the RRS filter using the sharpening technique.

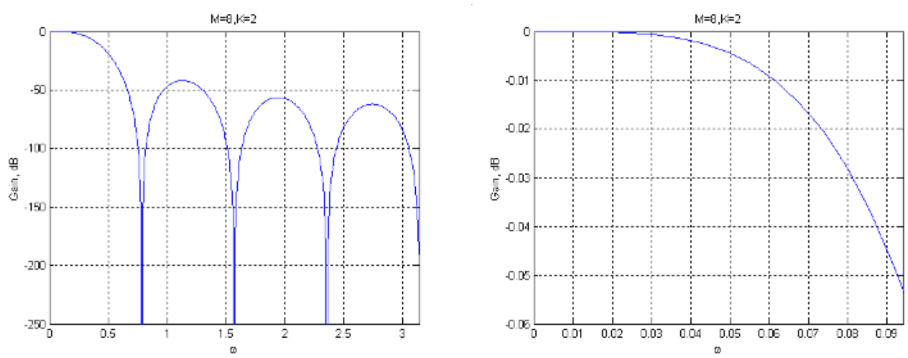
Sharpening RRS Filter

The improved RRS characteristics are based on a technique for sharpening the response of a filter proposed by Kaiser and Hamming (1977) and Kwentus, Jiang, and Willson (1997). This technique improves both the pass-band and stop-band of an RRS filter by using multiple copies of the original filter. The simplest sharpened RRS filter magnitude characteristic $|H_{shRRS}(\omega)|$ is related to the original RRS filter as:

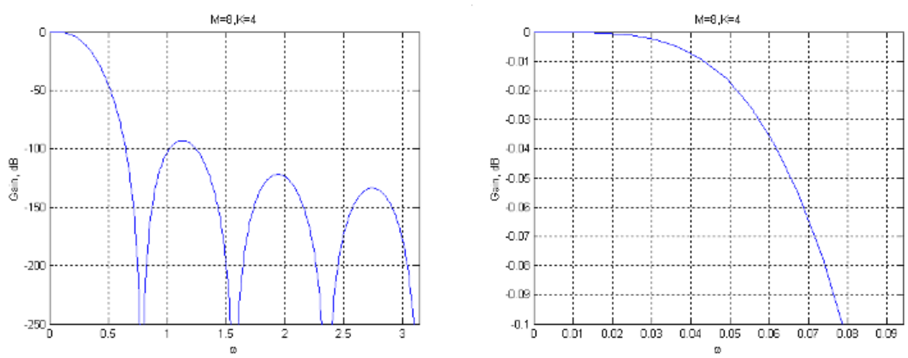
Figure 3. Sharpened RRS filter



a. Overall magnitude response *b. Pass-band zoom*

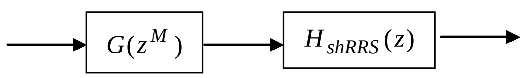


c. Overall magnitude response *d. Pass-band zoom*



e. Overall magnitude response *f. Pass-band zoom*

Figure 4. Sharpening IFIR structure



$$\left| H_{shRRS}(e^{j\omega}) \right| = \left| 3H_{RRS}^2(e^{j\omega}) - 2H_{RRS}^3(e^{j\omega}) \right| , \qquad (10)$$

where the RRS filter magnitude response $H_{RRS}(e^{j\omega})$ is given in Equation (9).

The sharpening RRS filter has improved both pass-band and stop-band characteristics over the ones of the RRS filter, thereby making it a good candidate for the interpolator in an IFIR structure, as illustrated in Figure 3. The corresponding block diagram is shown in Figure 4.

In the following section we outline the design procedure based on the structure in Figure 4 and illustrate it with an example.

Filter Design Procedure

We consider two cases, depending on whether M is even or odd.

The following seven-step procedure may be used to design a sharpening HP IFIR filter when M is an even number:

1. Transform the high-pass specifications into the low-pass ones.
2. Choose the value M.
3. Design the low-pass filter $G(z)$.
4. Expand the filter $G(z)$ M times.
5. Choose the value K and design the low-pass RRS filter. Transform the low-pass RRS filter into a high-pass filter, and design the corresponding sharpening HP RRS filter.
6. Cascade the expanded filter $G(z^M)$ and the sharpening high-pass RRS filter.
7. Check if the specification of the resulting filter is satisfied. If not, choose different values for M and K, and repeat the design.

If the interpolation factor M is an odd number, both filters in the IFIR structure—the shaping filter and the interpolator filter—need to be transformed into high-pass filters. Detailed explanation of the design is presented in the following steps:

1. Transform the high-pass specifications into the low-pass ones.
2. Choose the value M.
3. Design the low-pass filter $G(z)$.
4. Expand the filter $G(z)$ M times.
5. Transform the expanded filter into the high-pass one.
6. Choose the value K and design the LP RRS filter. Transform it into an HP RRS filter and a design HP sharpening RRS filter.
7. Cascade the expanded filter $G(z^M)$ and the sharpening high-pass RRS filter.
8. Check if the specification of the resulting filter is satisfied. If not, choose different values for M and K, and repeat the design.

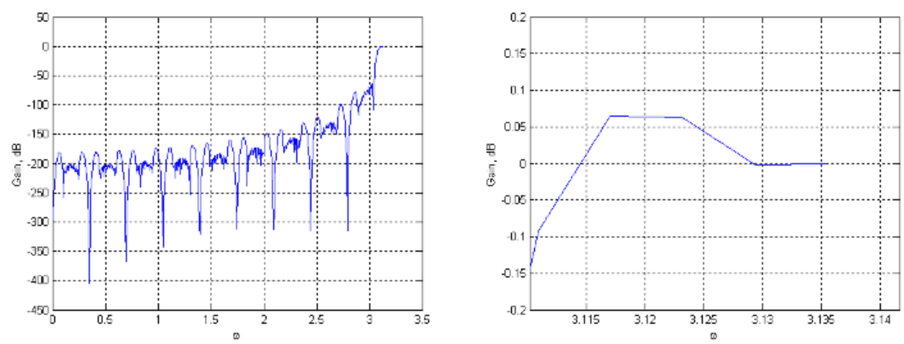
We illustrate the above design in the following example.

Example 1.

Consider a high-pass filter with pass-band edges at $\omega_p = 0.99\pi$ and $\omega_s = 0.97\pi$, pass-band ripple $R_p = 0.2$ dB, and the minimum stop-band attenuation $A_S = -60$ dB.

The high-pass filter that satisfies the specification, when designed using Parks-McClellan algorithm, has an order $N_H = 254$. By choosing $M = 18$ and $K = 3$, the order

Figure 5. Designed filter



a. Overall magnitude response *b. Pass-band zoom* of the shaping filter $G(z)$ is only 16. The magnitude response of the designed filter is shown in Figure 5.

FUTURE TRENDS

Presented here is the low complexity design of narrowband high-pass filters based on the IFIR structure and the sharpening technique. The basic idea behind the frequency masking technique, proposed to reduce the complexity of FIR filters (Lim & Lian, 1993), is to design the overall filter using several less complex sub-filters, including the model filter and its complement, and the corresponding interpolation filters. This technique is the generalization of the IFIR technique and can be also used for the wide-band FIR filters with the narrow transition bands. Indeed, both techniques can be combined if appropriate. Less-complex sub-filters can be designed using the simple filters like cosine and RRS filters. Future work must be dedicated to the problem of the optimization of the design, and the trade-off between the complexity of the sub-filters and the complexity of the sharpening polynomial.

CONCLUSION

The design of low-complexity high-pass narrow-band FIR filters has been presented. The method is based on the use of the IFIR structure where the interpolator is the sharpening RRS filter. The RRS filter requires no multipliers and no storage for the interpolation coefficients. The sharpening technique improves the pass-band and the stop-band characteristics of an RRS filter. The overall result is the lower computational complexity of the resulting designed filter. The method is useful for narrowband HP filter design.

REFERENCES

Abeysekera, S.S. & Padhi, K.P. (2000). Design of multiplier-free FIR filters using a LADF sigma-delta modulator. *Proceedings of the 2000 IEEE International Symposium on Circuits and Systems. Emerging Technologies for the 21st Century, 2,* 65-68.

Bartolo, A. & Clymer, B.D. (1996). An efficient method of FIR filtering based on impulse response rounding, *IEEE Transactions on Signal Processing*, 46(August 8), 2243-2248.

Coleman, J.O. (2002). Cascaded coefficient number systems lead to FIR filters of striking computational efficiency. *Proceedings of the 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1,* 513-516.

Diniz, P.S.R., da Silva, E.A.B. & Netto, S.L. (2002). *Digital signal processing, system analysis and design*. Cambridge: Cambridge University Press.

Grover, D. & Deller, J.R. (1999). *Digital signal processing and the microcontroller*. Englewood Cliffs, NJ: Prentice-Hall.

Ifeachor, E.C. & Jervis, B.E. (2001). *Digital signal processing: A practical approach* (2nd edition). Englewood Cliffs, NJ: Prentice-Hall.

Jou, Y.D., Hsieh, C.H. & Kuo, C.M. (1997, August 4). Efficient weighted least-square algorithm for the design of FIR filters. *IEE Proceedings of Visual Image Processing, 144,* 244-248.

Jovanovic-Dolecek, G. (Ed.). (2002). *Multirate systems: Design and applications*. Hershey, PA: Idea Group Publishing.

Jovanovic-Dolecek, G. (2003). Design of narrowband highpass FIR filters using sharpening RRS filter and IFIR structure. In J. Peckham & S.J. Lloyd (Eds.), *Practicing software engineering in the 21st century* (pp. 272-294). Hershey, PA: IRM Press.

Kaiser, J. & Hamming, R. (1977). Sharpening the response of a symmetric nonrecursive filter by multiple use of the same filter. *IEEE Transactions on Acoustics, Speech, and Signal Processing,* ASSP-25(October), 415-422.

Kumar, B. & Kumar, A. (1999). Design of efficient FIR filters for amplitude response: By using universal weights. *IEEE Transactions on Signal Processing, 47*(February 2), 559-563.

Kuo, C.J., Chien, H.C. & Lin, W.H. (2000). Neighboring full-search algorithm for multiplierless FIR filter design. *IEICE Transactions on Fundamentals of Electronics Communications & Computer Sciences,* Inst. Electron. Inf & Commun. Eng, Japan, *11,* 2379-2381.

Kwentus, A.Y., Jiang, Z. & Willson, A.N. (1997). Application of filter sharpening to cascaded integrator-comb decimation filter. *IEEE Transactions on Signal Processing, 45*(February 2), 457-467.

Lian, Y. & Lim, Y.C. (1998). Structure for narrow and moderate transition band FIR filter design. *Electronic Letters, 34*(1), 49-51.

Lim, Y.C. & Lian, Y. (1993). The optimum design of one- and two-dimensional FIR filters using the frequency response masking technique. *IEEE Transactions on Circuits and Systems-II, 40*(February), 88-95.

Mitra, S.K. (2001). *Digital signal processing: A computer-based approach*. New York: McGraw-Hill.

Neuevo, Y., Dong, C-Y. & Mitra, S.K. (1984). Interpolated impulse response filters. *IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-23*(June), 301-309.

Oppenheim, A.V. & Schafer, R.W. (1999). *Discrete-time signal processing* (2nd edition). Englewood Cliffs, NJ: Prentice-Hall.

Smith, S. (2002). *Digital signal processing: A practical guide for engineers and scientists*. New York: Newnes.

Stein, J. (2000). *Digital signal processing: A computer science perspective*. New York: Wiley-Interscience.

Webb, J.L. & Munson, D.C. (1996). A new approach to designing computationally efficient interpolated FIR filters. *IEEE Transactions on Signal Processing, 44*(August 8), 1923-1931.

White S. (2000). *Digital signal processing: A filtering approach*. Delmar Learning.

Yli-Kaakinen & Saramaki, T. (2001). A systematic algorithm for the design of multiplierless FIR filters. *Proceedings of the 2001 IEEE International Symposium on Circuits and Systems, 2*, 185-188.

KEY TERMS

Cutoff Frequencies: The frequencies that determine the pass-band (the frequencies that are passed without attenuation) and the stop-band (the frequencies that are highly attenuated).

Frequency Response $H(e^{j\omega})$: The discrete-time Fourier transform of the impulse response of the system; it provides a frequency-domain description of the system. In general, it has a complex value.

High-Pass Digital Filter: Digital filter that passes only high frequencies defined by the pass-band cutoff frequency and attenuates all frequencies from 0 to cutoff stop-band frequency.

IFIR Structure: The cascade of the expanded shaping or model filter $G(z^M)$ and the interpolator $I(z)$.

Impulse Response $h(n)$: The response of a digital filter to a unit sample sequence, which consists of a single sample at index $n = 0$ with unit amplitude.

Low-Pass Digital Filter: Digital filter that passes only low frequencies defined by the pass-band cutoff frequency and attenuates all high frequencies from the cutoff stop-band frequency to p.

Magnitude Response $|H(e^{j\omega})|$**:** Absolute value of the complex frequency response.

Phase Response: Phase of the complex frequency response.

RRS Filter: Recursive running-sum filter is an FIR filter with all coefficients equal to 1 presented in a recursive form.

Sharpening: The procedure for improving both the pass-band and the stop-band of a symmetrical FIR filter by repetition of the original filter according to a chosen polynomial.

Simulation and Gaming in IT Education

Norman Pendegraft
University of Idaho, USA

INTRODUCTION

The rapid change in information technology presents several serious problems to IS educators. In particular, the number of basic ideas that must be mastered is constantly increasing while the time available is not. This makes it essential to use class time efficiently as well as effectively. Simulations and simulation games provide an interesting and useful tool to help in this effort.

BACKGROUND

Simulation, case studies, role playing, and gaming are related teaching methods based on experiential learning. They permit experience or experimentation with a situation modeling the real world (Senge, 1990). On a deeper level, simulation is claimed by some to be a fundamentally new way of studying the world (Pagels, 1998). The idea that students learn better by doing goes back at least to Dewey (1938). The key idea underscoring this approach is that people learn better from experience than from reading or listening (Corbeil, Laveault & Saint-Germain, 1989). By compressing time, the simulation allows the students to experience the consequences of their own actions or to see how a system operates.

Simulation as a teaching tool suggests several approaches. Perhaps most obvious in an information systems curriculum is computer simulation, a tool well known in the information systems community. Using this technique, a computer program is written which exhibits behavior that models the behavior of the system under study. Butterfield and Pendegraft (1998) described a spreadsheet simulation of a Fourier Series, adding sine wave to construct a square wave, thus demonstrating how bandwidth limits data rates. Simulation can be easily used in the information systems classroom by having students operate or create simulations of relevant technology. Campbell (1996) created a simulation of a computer and had his students write assembly language programs to execute on the simulation. Englander (2003) uses Little Man Computer, a simple paper simulation of a CPU, as an example to explain basic CPU architecture, CPU operation, and machine language. In an extension of that idea, Pendegraft and Stone (2003) had their students develop a Visual Basic simulation of a Little Man central processing unit on which they ran programs mandated by the instructor. In addition to having to execute simple programs written in Little Man's machine language, their simulation had to deal with other architectural issues like input and output.

Case studies are a time-honored approach of instruction in strategy courses (see for example, Burgelman, Maidique & Wheelwright, 2001). Barker (2002) suggests that they can also be very valuable for teaching technical skills such as software development. In some sense, a case study is a role play, with the student acting the part of an analyst examining the case situation.

Role playing and simulation gaming are similar approaches in that they use simulated worlds, but instead of creating or observing or analyzing that world, students are immersed in it. Role playing is a method in which students are presented a scenario simulating some real situation, and assigned roles in that scenario. The scenario can be based on real or simulated situations (Barker, 2003). Participants then assume the roles of relevant persons in the scenario and act out the situation to see what happens. Role playing is a commonly and successfully used tool in IS education (for example, Christozov, 2003).

According to Greenblat (1988), simulation gaming includes role playing as an element, Whereas role playing allows participants to play the roles as they please, simulation gaming emphasizes the interactions of the roles and constraints of various types on the players. In some sense, a simulation game strives to teach about a specific situation, while a role play or game may have a more general lesson. The additional structure allows us to focus the students' attention on key issues. Dennis (2002) and Pendegraft (2002, 2003) both developed classroom gaming exercises to help students understand TCP/IP from the inside. Their exercises will be examined in more detail later in this article.

After many years of using such exercises at all levels (undergraduate, graduate, and executive), it is the author's opinion that they are very useful and that major benefit accrues to the instructor in preparing the simulation as well as to the students when they play the game. Simulation and gaming are student-centered learning, that is, the student is actively involved in the learning rather than passively observing the instructor (Greenblat, 1998). The student does the work, makes decisions, and sees the impact of the decisions. Role playing and simulation gaming attempt to take advantage of this by creating a situation in which a student may "play a game" in which time is compressed and attention can be focused on a few key ideas. Finally, these kinds of exercises are fun. The class gets to move around, talk, and frequently laugh. Simulations and games epitomize the idea that learning should be fun.

USING A GAME TO TEACH TCP/IP

TCP/IP is a layered protocol for controlling data flow in a packet switched network. For a more complete discussion of TCP/IP see, for example, Hunt (1998). Figure 1 lists the layers. TCP/IP uses packet switching, which means that each message is broken into pieces (called packets), each of which contains part of the message. Each packet is augmented with a header that contains (among other things) the addresses of the source and destination machines. At each step along the way, a router looks at the header, determines the address of the next host along the way, and forwards the packet to that next host.

Understanding the addressing presents a major hurdle to understanding TCP/IP. There are two types of addresses in each packet. The IP header includes the IP addresses of the source and destination machines. These do not change as the packet moves through the network. DLL layer header includes the MAC addresses of the source and destination machines for the current hop. These change at each hop. Since there are two destination addresses and two source addresses, some students have difficulty understanding how the addressing works. One way to help explain the addresses is to have the students play a game.

As examples of this approach, consider two similar games, one designed by Dennis (2002) and one by Pendegraft (2002, 2003), to help teach how TCP/IP works. Both are published elsewhere and so will not be described in detail here. Both games are designed to be run in one class session in a course on telecommunications. Both seem suitable for undergraduates or graduate students.

Figure 1. TCP/IP layers

Layer		Address
Application		
Transport Control	TCP	Port
Internet Protocol	IP	IP
Data Link Layer	DLL	MAC
Physical Layer		
TCP/IP Layers		

Both could be adapted to class sizes ranging from a dozen to more than 40.

The games have similar structure. The class is divided into teams, each team representing a host. Each player represents one layer on that host. A network map and instructions for each layer are given to the players. Figure 2 shows a sample network, and Figure 3 gives an excerpt from Dennis's instructions for the Data Link Layer.

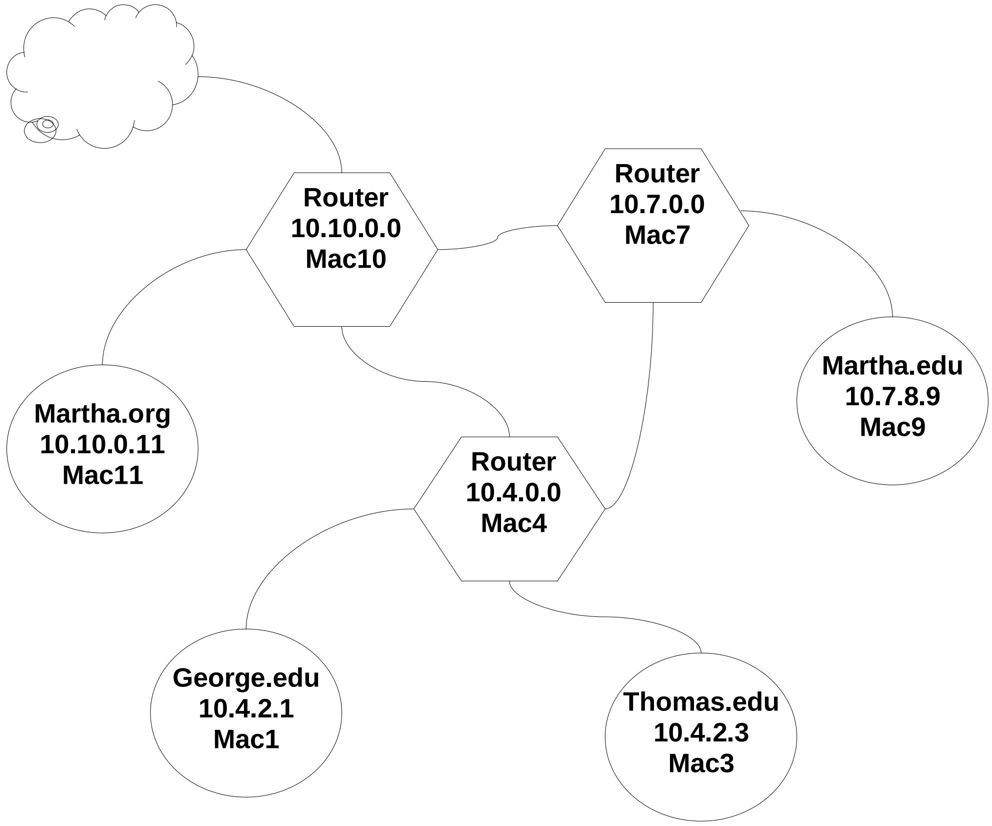
In the play of the simulation, an application layer player writes a message to another application layer player on a paper form. The form is then handed to the TCP player of the sending host. The TCP player adds the TCP header and hands the packet to the IP player. The IP player adds the IP header and hands the packet to the DLL player. The message is then passed to another host where each player strips off the header for that layer and hands the message upward, or in the case of IP forwards it as necessary.

There are several structural differences. Pendegraft gives the IP layer a routing table (see Figure 4 for an example), while Dennis has the IP layer use the network map to determine first hop IP addresses. Pendegraft uses a separate envelope for each layer and its header, while Dennis has different forms for each layer's header data which are taped to the message form. Another difference is that in Pendegraft's game, there is no physical layer player. Instead, the DLL players hold the ends of pieces of wire representing the physical layer. When a DLL player finishes adding the MAC addresses to the packet, he strings the packet on the wire and by raising his hand, sends the packet to the next machine. Such silliness introduces some humor into an otherwise dry subject, and helps keep the students interested and involved.

The games offer different points of view. Dennis's game allows many messages on the network at one time; consequently, it offers a more complete (and hence more complex) model of TCP/IP. In Pendegraft's game only one message is sent at a time and the entire class follows it along the way, immediately discussing problems that may occur such as a player incorrectly addressing a packet. In Dennis's game that sort of error is handled in discussion between the affected players. One result of these two points of view is that in Dennis's game, each player has a different experience that can be shared in the post-game discussion, while in Pendegraft's game, there is a more shared experience.

Both games simplify TCP/IP, ignoring some issues like handshaking or error detection. This is not to say that these issues are unimportant, but that these games focus attention on a limited set of issues of paramount importance, primarily addressing. Pendegraft's game is designed to allow including such complexities "notionally." That is, the instructor can intervene with an external complexity. For example, after a couple of messages have

Figure 2. Network map



been successfully transmitted, the instructor can announce that a packet was garbled in transmission and let the DLL player handle the problem. Or he can announce that at particular data link or router is temporarily out of service, obligating the IP player to reroute traffic.

FUTURE TRENDS

Among the trends in IT that seem relevant is the trend toward increased system complexity, paralleled with efforts to hide that complexity from the user. Examples include third-generation cell phone services, distributed databases, and network-centric systems like Planetnet. These systems raise questions about how to prepare students to deal with a world of increasing complexity. It may be that simulations and games offer a way to let students experience complex systems without having to delve into all the details.

CONCLUSION

Despite their differences, both games have been successfully used to help teach a complex subject. As the two games show, it is possible to design a game to accomplish a specific purpose in a class or to fit the instructor's style. Student response to such exercises is generally favorable for both their educational and entertainment value.

Understanding IT basics will remain an essential part of the education of IT professionals. As technology

Figure 3. DLL instructions (Dennis, 2002, p. 374)

1. Accept outgoing messages from the network layer, format them, add error-control information, and pass them to the physical layer.
2. Accept incoming messages from the physical layer. If they contain no errors, pass them to the network layer. If they contain an error, destroy the message.

Figure 4. Routing table for host 10.4.0.0

Destination	First Hop
10.4.2.1	10.4.2.1
10.4.2.3	10.4.2.3
10.7.*.*	10.7.0.0
*	10.10.0.0

evolves, it will continue to be a challenge to help students understand the basics and prepare themselves to keep learning. Simulation and gaming offer effective and fun tools to help students learn about new technology. They do so in a way that reinforces the need to learn how to learn and to continue learning.

REFERENCES

Barker, S. (2002). Training business students to be end-used [sic] developers: Are case studies the best option? *Proceedings of the IRMA International Conference,* Seattle, WA.

Barker, S. (2003). Business students as end-use developers: Simulating "real-life" situation through case study approach. In T. McGill (Ed.), *Current issues in IT management* (pp. 305-312). Hershey, PA: IRM Press.

Burgelman, R.A., Maidique, M.A. & Wheelwright, S.C. (2001). *Strategic management of technology and innovation* (3rd edition). Boston: McGraw-Hill.

Butterfield, J. & Pendegraft, N. (1998). Fourier analysis: Creating a virtual laboratory using computer simulation. *Informing Science, 1*(3).

Campbell, R.A. (1996). Introducing computer concepts by simulating a simple computer. *SIGCSE Bulletin*, *28*(3).

Christozov, D. (2003). Real live cases in training management of information resources during the transition to market economy. In T. McGill (Ed.), *Current issues in IT management* (pp. 297-303). Hershey, PA: IRM Press.

Corbeil, P., Laveault, D. & Saint-Germain, M. (1989). *Games and simulation activities: Tools for international development education.* Quebec: Canadian International Development Agency.

Dennis, A. (2002). *Networking in the Internet age.* New York: John Wiley & Sons.

Dewey, J. (1938). *Experience in education.* New York: Collier.

Englander, I. (2003). *The architecture of computer hardware and systems software* (3rd edition). New York: John Wiley & Sons.

Greenblat, C.S. (1988). *Designing games and simulations: An illustrated handbook.* Newbury Park, CA: Sage Publications.

Hunt, C. (1998). *TCP/IP network administration* (2nd edition). Sebastapol: O'Reilly.

Pagels, H. (1998). *Dreams of reason.* New York: Simon and Schuster.

Pendegraft, N. (2002). The Internet game. *Proceedings of the International Conference of the Information Resources Management Association,* Seattle, WA.

Pendegraft, N. (2003). The TCP/IP game. In T. McGill (Ed.), *Current issues in IT management* (pp. 117-124). Hershey, PA: IRM Press.

Pendegraft, N. & Stone, R. (2003, December 12-14). Using a simulation assignment to teach CPU operations. *International Conference on Informatics & Research*.

Senge, P. (1990). *The fifth discipline*. New York: Doubleday.

KEY TERMS

Case Study: An instruction tool containing a detailed description of a real-world situation.

Computer Simulation: A simulation built using a computer language.

Experiential Learning: Learning based on experiences rather than listening or reading.

Game: A simulation in which people are part of the model and their decisions partially determine the outcome.

IP (Internet Protocol): The protocol governing the operation of the network layer of the Internet.

Role Playing: An element in gaming which requires the student to assume the role of another.

Simulation: A dynamic model of a system focusing on selected aspects of that system.

TCP (Transport Control Protocol): Protocol governing the operation of the transport layer of the Internet.

S

Simulation for Business Engineering of Electronic Markets

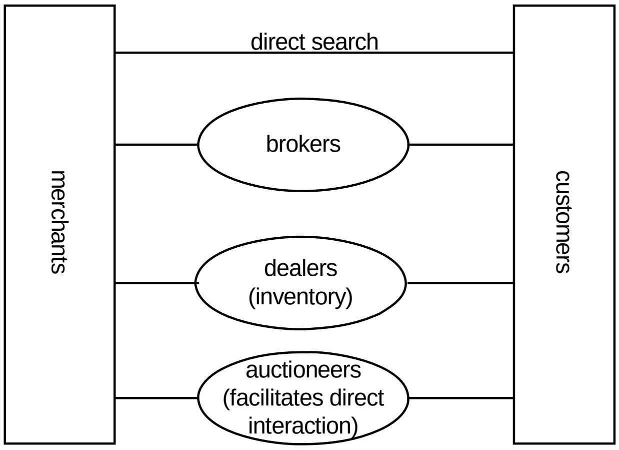
Marijn Janssen
Delft University of Technology, The Netherlands

INTRODUCTION

Information and communication technology (ICT) is applied to support the creation of electronic markets (e-markets). Electronic data interchange (EDI) is the exchange of structured data between information systems of different organizations. EDI is often considered as the first step towards e-markets. In the course of time EDI was extended and interorganizational information systems (IOS) were used to create electronic marketplaces. IOS serve as intermediaries between buyers and sellers. Probably the most well known examples of IOSs are airline reservation systems, which are closed systems having typically a limited number of participants.

With the advent of the Internet e-markets, targeting a global audience at low costs has emerged as a new way of doing business. An e-market system is an interorganizational information system that allows participating buyers and sellers to exchange information about prices and products (Bakos, 1991). Today's e-markets can be used for selling all kinds of products and even services and are typically built around publicly accessible networks, where entrance can be limited based on some kind of conditions. E-markets offer basic market functions and the current base is beginning to emphasize capabilities that aim to satisfy management information, risk management needs and enable integration of information systems (Dai & Kauffman, 2002). E-markets in business-to-business situations are witnessing an evolution towards newer market structures (Tomak & Xia, 2002). It is still unclear what the exact effects of electronic market are, how they emerge, and which markets will eventually turn out to be the most successful in the industry (Fairchild et al., 2004).

Figure 1. Examples of market structures



The value propositions of e-markets often center on two key dimensions (Le, 2002). The first dimensions deals with demand and/or supply aggregation to overcome market fragmentation (through search cost efficiency, price transparency and product-cost savings, market liquidity, network externalities and customer lock-in). The second dimension addresses inter-firm collaboration for greater supply chain performance (through streamlined workflows and process-cost savings, customer lock-on, and business process integration). Design decisions are critical, as they determine the market microstructure influencing liquidity, volatility and business propositions to stakeholders. The development and the introduction of e-markets requires the developer to carefully identify, evaluate, and understand the possible impact of the various design alternatives because they determine the market microstructures (Levecq & Weber, 2002). A business engineering methodology can be of help to design and develop e-markets by providing insight into current market and potential e-markets and by evaluating the implications of potential markets. Simulation can be used to compare the performance of the current and possible "to be" situations in a business engineering methodology. The *objective* of this article is to discuss research issues concerning the simulation of e-markets for business engineering.

BACKGROUND

E-markets can use a large variety of coordination mechanisms and market structures to coordinate demand and supply. A taxonomy of market structures based on how traders search out their counterparts is shown in Figure 1 (Garbade, 1982). In practice, all kinds of market structures and matching mechanisms can be used in accordance with traders' interest (Dai & Kauffman, 2002).

The core of e-markets is the *coordination* of the various interdependent activities performed by autonomous organizations during the information, agreement and settlement phases (Lindemann & Schmid, 1999). There are two opposing views on coordination. In a *coordination of tasks* approach the design of processes is dependent on the coordination mechanisms that manage the dependencies between tasks (Malone & Crowston, 1994). The *coordination of commitments* approach emphasizes networks of commitments that organizations establish through intentional acts of speech (Winograd & Flores, 1987). This coordination approach emphasizes the fulfillment of human commitments and describes activities in terms of contracts and promises. The combination of these two coordination perspectives leads to the definition of an *e-market* as the coordination of interdependent activities performed by autonomous organizations by exchanging data between information systems of buying, selling and facilitating organizations, allowing them to agree on and fulfill commitments.

Requirements of traders on e-markets are not easily elicited and can demand innovative mechanisms or deliberate trade-offs. The timely sharing of information among traders is often a major issue (Christopher, 2002). Information sharing is necessary for efficient coordination of the processes of traders, whereas for example sellers want to avoid information being used to negotiate lower prices or shared with competitors. Another typical issue in the business engineering of e-markets is the selection of matching mechanisms, as buyers and sellers can have different and even opposing requirements. For example one trader might want to minimize trading time, while another might want to maximize the reliability of a chosen trading partner. The most conspicuous opposing requirement is that buyers want to have the lowest price at the best possible trading conditions while sellers want to have the highest possible price to maximize revenue.

In short, a large number of other trade-offs and decisions need to be made before an efficient and effective e-market can be established. A list of business engineering issues is shown in Table 1. Some limitations are coming from the state-of-the-art of the technology and from market and/or product characteristics; others are coming from the opposing requirements and needs of the parties involved. Simulation for business engineering of e-markets can help decision makers to gain insight into these issues. This should support them to make deliberate choices, without having to experiment in real life, which could be costly and even result in a loss of customers.

SIMULATION FOR BUSINESS ENGINEERING

E-markets are by nature complex and analytic methods that can only be applied in a limited way. Analytical techniques analyze market mechanisms in a state of equilibrium and often focus on fair and efficient market mechanism (Mas-Colell, Whinston & Green, 1995). Although these approaches contribute to insight into and design of matching mechanisms, they do not help decision makers to evaluate the impact of e-markets for a practical situation. An e-market should be evaluated prior to implementation on criteria such as costs, utilization, trading time, delivery time, number of bids, matching chance, and so on. Analytical approaches do also not grasp the time-dependent dynamics resulting from the interplay between actors executing business processes.

Simulation of business processes constitutes one of the most widely used applications of operational research, as it allows us to understand the essence of business systems, to identify opportunities for change, and to evaluate the effect of proposed changes on key performance indicators (Law & Kelton, 1991). The philosophy behind a business engineering approach is to develop a simulation model of the current market, experiment with this model, and experiment with alternative market situations (Sol, 1982). An analysis of e-markets needs to begin with understanding of traditional market processes and should investigate how conventional transaction methods are changed as a result of e-market adoption (Lee & Clark, 1997).

Animation is often a standard feature of simulation. An animation model is a graphical representation of a problem situation and includes visualization of the time-ordered dynamics of objects, a static background, an overview of performance indicators and a user-interface (Vreede & Verbraeck, 1996). The purposes of animation are to facilitate decision makers to acquire insight into the dynamic interactions of the modeled system, the performance of the “as is” and “to be” situation and to facilitate communication between parties involved in a dynamic modeling study.

FUTURE TRENDS

Business engineering methodology provides context to the simulation technique (Greasley, 2003). Business engineering approaches proposed by Streng (1994) and Giagles, Paul and Doukidis (1999) tackled the identification of the added value of EDI by making use of discrete-event simulation. Nikolaidou and Anagnostopoulos (2003) use a simulation approach for modeling distributed systems. Janssen (2001) developed a business engineering methodology, including a simulation approach for electronic intermediaries. Business engineering using simulation should be based on the distinguished characteristics of e-markets and provide insight into current and potential markets and should evaluate the added value. Dependent on the characteristics of the market under study, certain issues shown in Table 1 dominate. Business engineering ensures the incorporation of the relevant business engineering issues and simulation can be used to aid conscious decision-making. A business engineering approach should help to focus on those issues most relevant to the particular situation and help to find solutions to those issues (Janssen, 2001). An interesting research direction is about which elements should be included in the business engineering for which type of markets.

A large number of independent organizations with their own strategies and sometimes even opposing aims carry out business with each other. The relations between organizations can change during the trading process, as organizations can enter or leave the playing field. The most powerful abstractions are the ones that minimize the semantic gap between the units of analysis that are intuitively used to conceptualize the problem and the constructs present in the modeling approach. Ideally, organization should be simulated as distributed systems, where each system is represented by an autonomous entity.

Software agents are autonomous entities that can be used for simulating organizations. The so-called agents are autonomous entities with their own interests and goals so they can decide to enter or leave a trading situation. *Agent-based simulation* has appeared for modeling organizations within e-markets (Janssen, 2001; Ramat & Preux, 2003).

In the past, electronic matching mechanisms were primarily used for commodity products (Lee & Lee, 1998). Over time, matching mechanisms became more sophisticated and were able to deal with product characteristics, delivery terms and preferences of buyers and sellers. Current research on mechanisms focuses on supporting semi-competitive matching approaches where agents strive to reach a fair and reasonable agreement for both parties, which nevertheless maximizes their own payoff (Luo et al., 2003). To evaluate the implications of these mechanisms, simulation of e-markets should focus on *emulation*, which means that actual software is written to execute matching mechanisms.

Communication over the Internet using a *Web-based* simulation is preferable for supporting communication between the researcher and persons involved in the design process. Another research issue is to use *distributed* simulation so multiple participants could interact with a simulation environment at the same time. In this way the participants can gain experiences with a hypothetical market situation. This might lead to an increase in insight into the problem situation by participants and might help designers of e-markets to make better decisions.

CONCLUSION

The quest towards innovative forms of e-markets has just started. During the business engineering of e-markets numerous trade-offs and decisions have to be made influencing the performance and possible adoption of an e-market. Business engineering using simulation should help decision-makers to focus on the most relevant issues and to make appropriate decisions and trade-offs without having to experiment in real-life situations. It is still unclear which the most relevant issues are to focus on when engineering e-markets.

Ideally, e-markets should be modeled with the help of a distributed, agent-based simulation environment that makes use of emulated mechanisms. Decision-makers should be able to view the animation over the Internet, manipulate parameters and view the consequences of their actions. The accomplishment of this ideal needs ample research attention in the domain of distributed simulation, agent-based simulation, Web-based animation and emulation of mechanism.

REFERENCES

Bakos, Y. (1991). A strategic analysis of electronic marketplaces. *MIS Quarterly*, *15*(3), 295-310.

Christopher, J. (2003). The effect of delays in information exchange in electronic markets**.** *Journal of Organizational Computing and Electronic Commerce, 12*(2), 121-131

Dai, Q., & Kauffman R.J. (2002). Business models for Internet-based B2B electronic markets. *International Journal of Electronic commerce, 6*(4), 41-72.

Fairchild, A.M., Ribbers, P.M.A., & Nooteboom, A.O. (2004). A success factor model for electronic markets: Defining outcomes based on stakeholder context and business process. *Business Process Management Journal, 10*(1), 63-79.

Garbade, K. (1982). *Securities markets.* New York: McGraw-Hill.

Giaglis, G.M., Paul, R.J., & Doukidis, G.I. (1999). Dynamic modeling to assess the business value of electronic commerce. *International Journal of Electronic Commerce, 3*(3), 35-52.

Greasley, A. (2003). Using business-process simulation within a business-process reengineering approach. *Business Process Management Journal, 9*(4), 408-420.

Janssen, M. (2001). *Designing electronic intermediaries.* Doctoral dissertation. Delft University of Technology, Delft, the Netherlands.

Law, A.M., & Kelton, D.W. (1991). *Simulation modeling and analysis.* New York: McGraw-Hill.

Le, T.T. (2002). Pathways to leadership for business-to-business electronic marketplaces. *Electronic Markets, 12*(2), 112-119.

Lee, H.G., & Clark, T.H. (1997). Market process reengineering through electronic market systems: Opportunities and challenges. *Journal of Management Information Systems, 3*, 113-136.

Lee, H.G., & Lee, R.M. (1998). Electronic call market for commodity transactions: Design of computer-mediated order matching system. *Journal of Organizational Computing and Electronic Commerce, 8*(4), 307-334.

Levecq, H., & Weber, B.W. (2002). Electronic trading systems: Strategic implications of market design choices. *Journal of Organizational Computing and Electronic Commerce, 12*(1), 85-103.

Lindemann, M.A., & Schmid, B.F. (1998). Framework for specifying, building, and operating electronic markets. *International Journal of Electronic Commerce, 3*(2), 7-21.

Luo, Z., Jennings, N.R., Shadbolt, N., Leung, H., & Lee, J.H. (2003). A fuzzy constraint based model for bilateral, multi-issue negotiations in semi-competitive environments. *Artificial Intelligence, 148,* 53-102.

Malone, T.W., & Crowston, K. (1994). The interdisciplinary study of coordination. *ACM Computing Surveys, 26,* 87-119.

Mas-Colell, A., Whinston, M.D., & Green, J.R. (1995). *Microeconomic theory*. New York: Oxford University Press.

Nikolaidou, M., & Anagnostopoulos, D. (2003). A distributed system simulation modeling approach. *Simulation Modeling Practice and Theory, 11,* 251-267.

Ramat, E., & Preux, P. (2003). Virtual laboratory environment (VLE): A software environment oriented agent and object for modeling and simulation of complex systems. *Simulation Modelling Practice and Theory, 11,* 25-55.

Sol, H.G. (1982). *Simulation in information systems development.* Doctoral dissertation. University of Groningen, Groningen, the Netherlands.

Streng, R.J. (1994). *Dynamic modelling to assess the value of electronic data interchange: A study in the Rotterdam port community*. Doctoral dissertation. Delft University of Technology, Delft, the Netherlands.

Teich, J., Wallenius, H., & Wallenius, J. (1999). Multiple-issue auction and market algorithms for the World Wide Web. *Decision Support Systems, 26*(1), 49-66.

Tomak, K., & Xia, M. (2002). Evolution of b2b market-places. *Electronic Markets, 12*(2), 1-8.

Vreede, G.J. de, & Verbraeck, A. (1996). Animating organizational processes: Insight eases change. *Simulation Practice and Theory, 4*(4), 245-263.

Winograd, T., & Flores, F. (1987). *Understanding computers and cognition. A new foundation for design*. Reading, MA: Addison-Wesley.

KEY TERMS

Agent-Based Simulation: Simulation of organizations as interacting autonomous entities with their own interests and goals and discrete processes.

Animation Model: An animation model is a graphical representation of a problem situation that can consist of a visualization of the time-ordered dynamics of objects, a static background, an overview of performance indicators and a user interface.

Business Engineering: The integral design of both organizational structures and information systems.

Coordination of Commitments: The actions by humans leading to the completion of work. Coordination is described in terms of contracts and promises consisting of recurring loops of requesting, making and fulfilling commitments.

Coordination of Tasks: The management of dependencies between tasks.

Discrete-Event Simulation: Discrete-event simulation models a system by changing the systems state at discrete points in time.

Electronic Markets: The coordination of interdependent activities performed by autonomous organizations by exchanging data between information systems of buying, selling and facilitating organizations allowing them to agree on and fulfill commitments.

Electronic Market System: An interorganizational information system that allows participating buyers and sellers to exchange information about prices and products.

Emulation: Actual software is written to execute something, instead of simulating it.

Matching Mechanisms: Well-defined set of rules for determining the terms of an exchange of something for money.

Spot Sourcing: Buying based on incidental and last-minute opportunities. Spot sourcing is transaction oriented and often involves short-term and incidental relationships.

Systematic Sourcing: Buying using pre-negotiated contracts with qualified sellers. These contracts are often long-term in nature and tend to have a cooperative nature.

Simulation in Information Systems Research

Theresa M. Vitolo
Gannon University, USA

Chris Coulston
Penn State Erie, USA

INTRODUCTION

Simulation is a standard research technique of the natural sciences, social sciences, and engineering disciplines. The paradigm of simulation provides an accepted mode of development, validation, and verification by which complex, highly dynamic interactions can be probed and analyzed. The approach enables researchers to phrase experiments in a controlled environment where the concepts, variables, and relationships of the domain can be manipulated. Then, employing standard experimental design techniques, the simulation represents the behavior of the underlying system. The behavior can be analyzed statistically for its regularity and critical values where the dynamics may become unstable. Further, once the simulation model is developed, sensitivity analysis can be done to probe the interdependencies of the elements of the underlying system upon one another.

Without the use of computers, the concept of simulation uses physical models to provide the experimental framework examining complex systems. These physical models incorporate dimensions of a real system, abstracted into the model. If one wants to explore a pilot's ability to manage a cockpit's physical controls under different degrees of information loading, one constructs a mock cockpit and offers the pilot different amounts of information to handle. The experiment of the physical simulation relies on the modeler validly abstracting relevant aspects of the real cockpit into the mock cockpit.

Computer-realizable simulation has become an option with the development of computers and the representation of models within a computing environment. Programming known mathematical representations of physical behaviors became a standard use of computers with programming languages such as FORTRAN, designed to numerically analyze mathematical functions within acceptable tolerances. In fact, the disciplines of management science and operations research are defined by their various modeling approaches to represent the behavior of physical systems into mathematical and computational frameworks - simulation being one of the frameworks. In fact, the initiation of these fields are usually traced to their successes in the 1940s with military-related research of dynamic systems using computers and simulation.

Today, simulation continues to be a tool for understanding complex, dynamic systems—but not for proposing optimal solutions to their interactions. Optimization cannot be a mathematical analysis goal of simulation; other mathematical approaches exist for this analysis goal. Typically, the dynamics and interactions of a simulated system are too complex to be captured into a mathematical structure, lending itself to optimization algorithms. Simulation provides a methodical way to describe and to examine the dynamic space of the system's behavior.

The usage of simulation to understand complex systems is not done solely in the natural and physical sciences. Social sciences such as economics and psychology have relied on the concept as a technique to test theories of behavior. Even performing well-defined, "what-if?" analysis in spreadsheet applications to assess supply-and-demand interactions constitutes simulation. Social psychologists might use a simulation framework by developing a research model based on probabilistic structures. The probabilities describe the tendency of elements in a social system such as a crowd to react and change into a mob. In the vernacular of the field, simulation examines the modeled elements, tendencies, and interactions, by manipulating the "stat rats" in order to probe the behavior.

Numerous computational approaches have been developed to handle simulations. General-purpose programming languages such as FORTRAN or C++ can represent the model and realize its execution. General-purpose simulation packages such as SIMAN (Pegden, Shannon & Sadowski, 1990), ARENA (Kelton, Sadowski & Sturrock, 2004), GPSS (Schriber, 1974), or MATLAB (Palm, 2001) have been classically used to represent the model and to facilitate the representation of time for the execution of the model. In addition, special-purpose simulation environments have been developed for specific domains ranging from inventory logistics, chemical manufacturing, to air-traffic control.

Modeling and statistical processes are foundations to simulation. Hannon and Ruth (2001) emphasize dynamic modeling perspectives; Shanbhag and Rao (2003) target stochastic processes as an aspect to simulation; Severance (2001) focuses on the systems-thinking perspective guiding simulation development; Powell and Baker (2004) show how simulation can be realized in easily accessible spreadsheet software; and Chung (2004) captures the general manner of the basics and techniques in a handbook for simulation modeling.

Simulation's advantage stems from the safe, abstract environment invented by the model. Often the underlying system to be understood may be (1) dangerous to manipulate, (2) unavailable for controlled manipulations, or (3) non-existent at the current period of time. The simulation of a proposed new rail-transport system based on a current city's environment captures all three of these conditions.

Simulation also has its down side. A good simulation depends upon a good model of the underlying system. Creating a good model requires analytical, mathematical, and logic skills. Executing the simulation requires computational skills. Understanding the results of the simulation requires knowledge about the underlying system and about the mapping of it into the model. These skills require training and development which are not readily available in the average technical professional.

BACKGROUND

Simulation incorporates experimental design as its underlying paradigm for understanding complex, dynamic systems. By manipulating the input variables and system parameters within acceptable and legal value ranges, the behavior of the system can be mapped as a function of the changes defined by the experimental design. More than one experimental design¾more than one set of experiments, more than set of variable perturbations¾may exist for any modeled system. The modeler determines what aspects of the system need to be captured in the model in order to be probed by the experiments. The process begins with a conceptual model of the system.

A prerequisite to constructing a simulation is to understand the behavior of the system and to represent it sufficiently in a valid model, which can be realized (executed) in a computational format. If the behavior is not well understood, the abstraction into the model may not accurately represent the relationship between the input variables and the output variables. The process of validation examines the correspondence between the modeled input and output pairings, and the actual input and output pairings. This correspondence should show statistical goodness-of-fit. If it does not, the conceptual model needs to be revised. Until this correspondence is valid, proceeding with the experimental design would not be rational.

Models, long a standard of research and engineering, are an abstraction of the significant features and relationships of a system. This abstraction enables the system to be studied for the purpose of prediction or comparison. The experimenter knows the system intimately and defines a level of analysis appropriate in order to enable the prediction of a system's behavior or to compare the effects of change upon the system's behavior.

Generally, models may be represented as linguistic, iconic, or physical representations. The choice depends upon how the modeler expresses the fundamental relationships of the target behavior. Examples of each of these are as follows:

1. A mathematical model of a business's order fulfillment interactions requires equations¾a model based on the language of mathematics.
2. A process model of an organization's enterprise-wide data distribution plan requires diagrams of nodes and connections - a model capturing behavior possibilities by the use of icons.
3. A construction, 3D model of a company's proposed new production plant requires an architectural layout of objects of buildings and land¾a model showing components' interactions based on physical scale.

Although a variety of texts present simulation, classic coverage of the simulation development process from model conceptualization to simulation execution can be found in Banks and Carson (1984), Law and Kelton (2000), Winston (1987), and Zeigler (1984, 1976). Coverage of a variety of classic implementation platforms for simulation can be found in Watson and Blackstone (1989). Aburdene (1988) presents both discrete-event simulation and continuous simulation, and provides examples across a variety of disciplines and applications from information diffusion to modeling the arms race. Pegden et al. (1990, pp. 12-13) identify the simulation process as a set of 12 steps. These steps cover the project management aspects of the simulation study as well as the conceptual management of the experiment.

1. Problem definition: Identifying the goal and purpose to the simulation
2. Project planning: Identifying and coordinating the hardware, software, staff, and management resources required
3. System definition: Identifying the system to be studied—in classical general systems analysis terms
4. Conceptual model formulation: Extending the system definition to incorporate formal variables, relationships, and interaction logic
5. Preliminary experimental design: Identifying an experimental framework by which to assess the study,

that is, which control variables, output variables, factors to vary

6. Input data preparation: Identifying and collecting data to configure parameters and probability distributions
7. Model translation: Representing the system and experiment in the language of a software application designed for simulation
8. Verification and validation: Confirming the accuracy, integrity, and credibility of the model's parameters, distributions, and output
9. Final experimental design: Finalizing the structure of the experiment to test a specific, stated hypothesis
10. Experimentation: Executing the simulation according to the experimental design
11. Analysis and interpretation: Examining the output of the simulation study in light of the hypothesis of the experiment
12. Implementation and documentation: Using the results of the analysis to drive action as a consequence of study

These 12 steps give a researcher an environment to manipulate, repeatedly, in order to explore any relationships appropriately captured by the model of the system. Since the simulation can be used to explore conditions that have not yet occurred, but which are plausible within the definition of the model, the simulation enables a researcher to pose problems and to experience solutions to the problems before they are needed. The reliability and credibility of the simulation are powerful reasoning tools for research.

The value of Step 8 is directly based on the goal of Step 11, interpretation. Simulation requires the model to be executed under statistical conditions, mimicking the concept of "in the long run." One run of a simulation cannot give statistically significant results and cannot prove the validity of the model. Simulation requires the model to have its statistical elements manipulated appropriately so that the tendency to respond to an input is correctly captured by the parameters used for the statistical elements. For instance, if a researcher has analyzed a system and has assessed that according to a normal distribution with a mean of σ and a standard deviation of μ, users will request services from an information help desk, and if the simulation is being constructed to assess the need for expanded staffing of the help desk, then the validity—the truthfulness of description—of the parameters to the normal distribution describing users' requests-for-services will influence the final interpretation of the staffing question. If the parameters are wrong—or if the assessment of a normal distribution is wrong—then the validity of the interpretation and the value of the simulation is lost.

THE CURRENT ROLE FOR SIMULATION IN INFORMATION SYSTEMS RESEARCH

While simulation as a technique for examining complex and dynamic systems is widely accepted in the physical and social sciences, the approach is not used, understood, or taught as a tool for information systems (IS) research. As such, its use as a tool for information systems management by IS professionals is limited. The fundamental gist of creating good simulations is to know the underlying system being modeled. IS professionals known their systems intimately, and yet, are not using a general-systems theoretic approach such as simulation for assessing their research or their problems.

IS research has not taken advantage of simulation to examine complex situations as other domains have. A variety of reasons exist for the limited use of simulation in IS research.

- The field is relatively young compared to the other domains using simulation.
- The formalization of the problems and systems within the IS field has not reached the level of maturity and understanding existing in other domains. Much IS research has focused on capturing and describing the behavior of IS systems, not on predicting the expression of the behavior.
- The technical competencies to develop a valid and verified simulation are not standard components of IS curriculums. IS curriculums do not provide the mathematical and analytical emphasis required to construct a robust simulation. IS curriculums stress the technical competencies to plan, to build, and to deploy systems, but not to predict the ramifications of incorporating new processes into an organization from a scientific perspective.

IS professionals who understand and build decision support systems for the traditional functional areas of industry have incorporated simulation as a modeling environment for assessing a variety of management decisions. Yet, the field has not embraced the same modeling approach to support management concerns of an IS's nature. Ironically, IS professionals are asked to build simulation environments and applications, but may not be trained in the basics of such tools.

However, as the field of IS matures, IS professionals will want better understanding and control of the systems they build, manage, and modify. Also, the organizations relying on the systems will require the IS professionals to do these tasks better. The use of simulation by IS researchers can be used to address the need for accountability and prediction. The problem domains for the application of simulation will come from the industrial settings of IS professionals. The understanding of the components of the problem domain and how they could be expressed by simulation are tasks for IS researchers.

From risk management in project development to security assessment in IS operations, IS professionals require quantitative appraisals that are reliable indicators of performance and expectations. While the language of "indicators of performance and expectations" is common to marketing, production, and human resources efforts, the incorporation of these indicators into a model is not as common. If an IS researcher wants to assess how effective different change management techniques would be for a development project, the elements of the development project, the character of the change management techniques, and the different response tendencies of the users and project coordinators could be modeled and simulated. The simulation could be probed, suggesting an effective technique and its timing, reflecting the needs of the organization and of the project. Currently, IS professionals might use a technique at hand which need not necessarily fit the project.

As IS executives embrace the concept of portfolio management of IS efforts in an organization, the concept requires quantitative grounding. If a CIO is to define the appropriate mix of on-going projects and strategic initiatives, some rational basis for promoting one mix over another is required. By building a simulation of the IS efforts of an organization, a CIO could examine how different project mixes could influence the work and future of the organization. IS research could support the portfolio management by defining suitable mappings of organizations and projects into model components to be incorporated by the CIOs into a specific model. Specific models would then be simulated.

IS tactical management wants to add knowledge management to its repository of intellectual assets. While an organization has multiple sources of knowledge, IS management needs to define a rationale to identify which sources should be captured, how to capture the source(s), and what aspects of the source(s) to capture. Without an environment to assess these decisions, knowledge management efforts could simply become a burden to the IS professionals and useless to the organization. Simulation can provide the environment, once again, to quantitatively view the value of the knowledge asset in light of its potential to the organization. IS research could examine the questions of knowledge management

IS operations - like any other department of operations - can benefit from simulation, as it has been used historically in manufacturing, logistics, or facilities layout. This aspect is an accepted concept of information resources management. However, the use of classical operations management techniques to IS operations has been less common, but is changing. The assessment of network capabilities provides situations for simulation (for usage of simulation to review network capabilities, see Bagrodia, Gerla, Kleinrock, Short & Tsai, 1995; Lee, Gerla & Toh, 1999). Requirements assessment as part of the development of information systems has also proposed the use of simulation (see Rozenblit & Liu, 1990). The trend towards incorporating total cost of ownership (TCO) assessments into information systems management is another example area of operations where simulation would be a viable tool.

FUTURE TRENDS

Traditionally, IS research gives the IS profession insight into methodologies, techniques, and technologies. IS professionals can benefit from an analytical framework for examining the use, need, and character of IS at the strategic, tactical, and operational level of analysis. Simulation could provide this analytical framework for reviewing the efficiency and effectiveness of proposed and operational systems. The simulation results would provide comparisons between proposals or explanations of problematic events.

The reasons for simulation not being more common in IS research can be attributed to four points: the structure of IS curriculums, the modeling approaches driving current research, the separation of IS project development from review after deployment, and the newness of the IS field. Each of these reasons impose limitations upon IS research and its potential use of simulation. The conditions of these reasons, however, are changing. As the IS field matures and moves away from ad hoc techniques, and moves towards repeatable, sustainable processes, more scientific techniques will be sought to assess situations. This trend is occurring now in the profession. The trend will also influence IS curriculums to change in order to promote the scientific exploration of the situations.

CONCLUSION

Researchers in the natural sciences, social sciences, and engineering have relied on the methodology of simulation in order to represent, assess, and predict behavior in systems. Simulation requires a variety of skills from systems analysis to mathematical modeling to logical construction in order to create, verify, validate, and execute a simulation model. All of these skills are not common to all IS professionals, and the methodology of simulation is not widely known or taught to them either. However, the problems IS professionals and researchers address can benefit from the use of simulation. If IS curriculums include the skill sets needed for simulation and include simulation in their learning objectives, then simulation can be used as one of the quantitative techniques for addressing IS problems and research questions.

REFERENCES

Aburdene, M.F. (1988). *Computer simulation of dynamic systems*. Dubuque, IA: William C. Brown.

Bagrodia, R., Gerla, M., Kleinrock, L., Short, J. & Tsai, T.-C. (1995). A hierarchical simulation environment for mobile wireless networks. *Proceedings of the 27th Conference on Winter Simulation* (pp. 563-570).

Banks, J. & Carson, J. (1984). *Discrete-event system simulation*. Englewood Cliffs, NJ: Prentice-Hall.

Chung, C.A. (2004). *Simulation modeling handbook: A practical approach*. Boca Raton, FL: CRC Press.

Kelton, W.D., Sadowski, R.P. & Sturrock, D.T. (2004). *Simulation with ARENA* (3rd Edition). Boston: McGraw-Hill Higher Education.

Law, A.M. & Kelton, W.D. (2000). *Simulation modeling and analysis* (3rd Edition). New York: McGraw-Hill.

Lee, S., Gerla, M. & Toh, C.-K. (1999). A simulation study of table-driven and on-demand routing protocols for mobile ad hoc networks. *IEEE Network, 13*, 48-54.

Liu, K.C. & Rozenblit, J.W. (1990). Applying knowledge-based system design and simulation in information system requirements determination. *Proceedings of the 22nd Conference on Winter Simulation* (pp. 407-411).

Palm, W.J. III. (2001). *Introduction to MATLAB 6 for engineers*. Boston: McGraw-Hill.

Pegden, C.D., Shannon, R.E. & Sadowski, R.P. (1990). *Introduction to simulation using SIMAN*. New York: McGraw-Hill.

Powell, S.G. & Baker, K.R. (2004). *The art of modeling with spreadsheets: Management science, spreadsheet engineering, and modeling craft*. Hoboken, NJ: John Wiley & Sons.

Schriber, T. (1974). *Simulation using GPSS*. New York: John Wiley & Sons.

Severance, F.L. (2001). *System modeling and simulation: An introduction*. New York: John Wiley & Sons.

Shanbhag, D.N. & Rao, C.R. (Eds.). (2003). *Stochastic processes: Modelling and simulation*. Amsterdam & Boston: Elsevier.

Watson, H.J. & Blackstone, H.J. Jr. (1989). *Computer simulation*. New York: John Wiley & Sons.

Winston, W. (1987). *Operations research: Applications and algorithms*. Boston: PWS-KENT.

Zeigler, B.P. (1984). *Multi-facetted modeling and discrete event simulation*. San Diego, CA: Academic Press Professional.

Zeigler, B.P. (1976). *Theory of modeling and simulation*. New York: John Wiley & Sons.

KEY TERMS

Conceptual Model: The abstraction of relevant aspects of a target—usually physical—system into a representation system consisting of symbols, logic, and mathematical constructs.

Continuous System: A system whose state variables show its behavior change continuously with respect to time. Hence, the behavior of the system is seen as changing continuously over time. Such systems are usually modeled using differential equations.

Discrete System: A system where the state variables showing its behavior change only at isolated, discernible points of time. Hence, the behavior of the system is seen as changing in distinct, separate moments of time. Such systems can be modeled using difference equations or incremental event analysis.

Experimental Design: The process of defining dependent and independent variables of the underlying model and acceptable ranges for these variables. In terms of input, output, and model, the experimental design describes which input variables and parameters to manipulate and according to what range of changes. As these elements are manipulated, specific output variables will be assessed. The connection between the input- and output-set pairings is embodied in the description of the model. Identifying what items to manipulate and which correspondence should reveal the effect of the manipulation describes the plan of the experiment.

Simulation: (1) An analysis technique capturing the behavior of a system as mathematical and logical algorithms into a model. The model is represented and realized in an environment where the resultant behavior of the model can be captured, analyzed, and used to further the understanding of the system. Simulation is not an optimization technique, but can be used with sensitivity analysis to probe the parameter space of the underlying system. (2) The computational execution of a simulation model in a simulation environment according to an experimental design.

Simulation Environment: A computing environment for realizing and executing a simulation model, giving output values corresponding to states of behavior in the physical, real system.

Simulation Model: The representation of a system in terms of the fundamental components, relationships, and interactions of its behavior, captured as mathematical and logical algorithms.

Validation: The process of showing the correspondence between the physical system and the simulated system is within acceptable tolerances of output performance. Validation relies on statistical confirmation of the simulated output to fit the behavior exhibited by target, physical system.

Verification: The process of determining the correctness of the programmed components in terms of typographical errors, numerical accuracy, algorithm completeness, and logical unity. Verification covers all aspects of logical, syntactic, and semantic debugging of computer programs.

Small Business Transformation Through Knowledge Management

Nory B. Jones
University of Maine Business School, USA

Jatinder N. D. Gupta
University of Alabama in Huntsville, USA

INTRODUCTION

In the last decade, the importance of knowledge as a source of sustainable competitive advantage has gained widespread acceptance. Business practitioners and academics alike recognize that what is "between the ears" (Tiwana, 2000) of their employees represents the source of creativity and innovation that nourishes and sustains the organization. Furthermore, the ability to harness the intellectual capital in an organization probably represents the most important aspect relating to the creation of an intelligent enterprise.

However, most research on the topic of knowledge management (KM) and intellectual capital has focused on larger organizations. Because small businesses account for a major portion of the total number of businesses, jobs, and growth in many world economies (Wong & Radcliff, 2000), it is important to understand the impact of knowledge management on small businesses as well. We need to understand the correlation between knowledge management practices, the ability of a small business to transform itself into an intelligent enterprise and any resulting performance or competitive improvements KM may provide.

Because a "build it and they will come" approach to knowledge management usually does not work, this article discusses and integrates the concepts of adoption and diffusion of innovations with knowledge management theories to help transform a small business into an intelligent enterprise. The ultimate goal of this chapter is to provide small businesses with some consistent theories and practices that may help improve their competitiveness in a turbulent world.

KNOWLEDGE MANAGEMENT IN SMALL BUSINESS

What is Knowledge Management?

Karl Wigg is credited with coining the term "knowledge management" (KM) at a 1986 Swiss Conference sponsored by the United Nations International Labor Organization. He defined KM as "the systematic, explicit, and deliberate building, renewal, and application of knowledge to maximize an enterprise's knowledge-related effectiveness and returns from its knowledge assets" (Wigg, 1999). Thus, KM represents an organization's ability to capture, organize, and disseminate knowledge to help create and maintain competitive advantage. It is becoming widely accepted as a key part of the strategy to use expertise to create a sustainable competitive advantage in today's business environment. It enables the organization to maintain or improve organizational performance based on experience and knowledge. It also makes this knowledge available to organizational decision-makers and for organizational activities (Beckman, 1999; Pan & Scarbrough, 1999). Therefore, we can assert that knowledge management represents a key strategy in creating and sustaining an intelligent enterprise, capable of outperforming its competitors.

Using KM in Small Business

Why is it important for small businesses to use knowledge management to become "learning organizations" or "intelligent enterprises"? According to Wong et al. (1997), "Many of the factors which have promoted the growth of SMEs also require their managers to acquire new skills. In fast-growing small firms, the management team will be constantly developing and the skills needed will change as both cause and effect of the development of the firm itself." The bottom line is that for a small business to succeed and thrive in a changing world, it must continually learn and adapt better and faster than its competitors. Knowledge management provides the tools and strategies to achieve this (Anderson & Boocock, 2002).

Guimaraes (2000) further suggests that small businesses face greater pressures from chains owned by large corporations, increased regulations and politics, and greater competition due to increasing business globalization. He asserts that innovation, facilitated by knowledge management, may be the key to their survival and success in difficult times. Chaston et al. (2001) support this view in their statement: "Organizational learning [knowledge management] is increasingly being mentioned in the literature as a mechanism for assisting small firm survival". It is "the most effective and practical way through which to increase Small and Medium-Size Enterprise (SME) sector survival rates during the early years of the new millennium". They contend that by assisting employees and facilitating their learning and knowledge sharing, they can creatively develop new products, better and more efficient processes, and identify new ways of building better relationships with customers. Thus, it appears that knowledge management techniques of acquiring, sharing and effectively using knowledge may represent a crucial means of transforming a small business into an intelligent enterprise, resulting in improved performance by facilitating innovation, idea creation, and operating efficiencies.

Influence of Adoption and Diffusion

How do adoption and diffusion factors influence KM in small businesses and their goal of becoming an intelligent enterprise? By understanding factors that facilitate the adoption and diffusion of innovations, small businesses may improve their chances of success in a knowledge management initiative. Based on many years of adoption and diffusion of innovations research, Rogers (1995) developed a model often considered the foundation for the adoption and diffusion of innovations. We include strategies and processes in the definition of "innovations" in the context of this chapter. This model proposes three main elements influencing the adoption and diffusion of innovations, including: the innovation, communication channels, and social systems.

Based on the research of Rogers and others, the factors that appear to significantly influence adoption and diffusion include the relative advantage of an innovation (the degree to which the innovation is perceived as better than what it supercedes), which is positively related to its rate of adoption and continued and effective use and the influence of culture and social systems. The social systems similarly influence people's attitudes and willingness to adopt new processes or technologies. Finally, communication channels such as mass communication and interpersonal channels have been found to be effective promotional avenues to facilitate the successful awareness and use of a new process or technologies. Other studies propose that technological change within organizations represents a cumulative learning process where firms will seek to improve and diversify their technology in areas that enable them to build upon their current strategies in technology (Alange et al., 1998). Thus, prior experience appears to influence the willingness to adopt and the rate of adoption and diffusion in addition to Rogers' variables. The concept of absorptive capacity (Cohen & Levinthal, 1990) further suggests that an organization's ability to absorb new knowledge or for an innovation to diffuse throughout is based on its prior experience with this knowledge or innovation. In terms of creating a learning organization or an intelligent enterprise, this theory says that the greater the absorptive capacity of the organization, the greater is the ability of its employees to absorb and use new knowledge effectively.

APPLICATIONS TO SMALL BUSINESS

How can these theories be applied to a small business for transformation into an intelligent enterprise? First, relying on Rogers' classic theories, a small business can easily communicate the advantages of these new technologies and practices using mass media channels, such as company newsletters, e-mail, or company meetings. After making people aware of the new KM system, the business can use interpersonal channels to effectively persuade people to try them and continue using them. By using homophilous colleagues (individuals with similar attributes such as common beliefs, education, social status and values), a small business can more effectively persuade people to adopt and then use the KM systems. These "knowledge champions" represent very important motivators and influencers because they are trusted and respected by their peers (Jones et al., 2003). This is very important because the literature is replete with cases of companies investing in new technologies and new management strategies, which are simply viewed with skepticism as the "latest fad". However, by carefully selecting peers who are trusted and respected, the adoption and diffusion process can be greatly facilitated.

Similarly, the literature describes the huge influence of culture on the effectiveness of KM systems and the development of a learning organization (intelligent enterprise). Rogers' theories can be effectively applied to the cultural aspect as well. The effect of norms, opinion leaders, and change agents can exert a profound influence on the adoption and diffusion of an innovation throughout a social system. This is because norms (culture) can exert a powerful influence on people's willingness to accept or reject an innovation depending on whether it is compatible with their existing values and norms. Therefore, by using change agents as cultural influencers, small businesses can greatly enhance the transition to becoming a learning organization/ intelligent enterprise while using knowledge management systems to provide competitive advantages. Finally, in an organizational social system, such as within a small business, a powerful individual within the organization, such as the business

Table 1. Contributing factors to the adoption & diffusion of innovations

Success Factor	Characteristic
Innovation:	
o Relative advantage	o The perceived benefit from the innovation that makes it superior to existing tools or processes.
o Compatibility	o How well the innovation is perceived as consistent with the existing values, past experiences, and needs of potential adopters.
o Complexity	o How difficult it is to understand and use.
o Triability	o Ability to experiment with the innovation.
o Observability	o Ability to see tangible results from using it.
Communication Channels:	
o Type of communication channel	o **Communication Type:** Mass media: used for initial awareness of innovation versus interpersonal: used to persuade potential people about benefits.
o Type of individual	o **Individual Type:** Homophilous; opinion leader, knowledge champion.
Social Systems (Culture):	
o Organizational Leaders (Change Agents)	o People in a position of authority who can make the adoption decision.
o Opinion Leaders (Peers)	o Those individuals who can exert a positive influence on others to adopt and use an innovation; individuals who have similar attributes such as common beliefs, education, social status and values, who can persuade their peers to adopt the new system and change values and norms within the culture.
o Prior knowledge/ experience/attitude towards innovation	o Attitudes and prior experience and knowledge may influence the adoption and diffusion of an innovation. The theory of absorptive capacity tells small businesses that the more knowledge they provide their employees, the greater their capacity to absorb more and become a more proactive learning organization/intelligent enterprise.

owner, can also exert a strong influence on the adoption and diffusion of a KM system as well as facilitating a cultural shift to becoming a learning organization. Table 1 summarizes these factors.

FUTURE TRENDS

As mentioned, small businesses often face major challenges of coping with resource deprivation. This includes not only fewer technological resources, but also less focused expertise to enable them to make critical decisions or improve products or processes. They need to learn how to learn, change, and adapt better and faster than their competitors and better meet the needs of their customers. By understanding how they can facilitate a knowledge sharing initiative, a small business may be able to make better use of scarce resources and become more intelligent than their competitors; become better at learning, innovation and creativity. For a small business to increase the speed and effectiveness of the diffusion process for a knowledge management initiative or other processes/technologies, it may be important for them to understand the components of an innovation, communication channels, and/or social systems within the context of a small business environment.

Therefore, incorporating a knowledge management strategy holds some promise as a mechanism for small businesses to improve performance. However, the “build it and they will come” philosophy may be simplistic and an investment in knowledge sharing technologies may be wasteful if the systems are not used effectively. Therefore, small businesses should be aware of the issues associated with adoption and diffusion of knowledge management practices.

Similarly, small businesses should understand that creating a learning organization/ intelligent enterprise potentially involves huge cultural changes. Thus, a major challenge is whether a small business owner is interested in fundamentally changing the culture of the organization, including the reward and recognition structures. For example, Pan and Scarbrough (1999, 2000) found that the major challenge to implementing a knowledge management initiative was successfully overcome when the CEO initiated cultural changes that actively promoted and rewarded knowledge acquisition, knowledge sharing, and knowledge use. This led to significant performance improvements such as considerably reduced time in process for product development, reduced costs and improved customer satisfaction. However, this involved a complete commitment to cultural change by a new CEO, who made dramatic changes in the culture and the reward/recognition structures and company policies.

CONCLUSION

With respect to opportunities and challenges for SMEs, knowledge management appears to represent a mechanism for survival, growth and prosperity. As knowledge management continues to diffuse throughout the global business world, the competitive pressures may necessitate some form of KM as a pre-requisite for survival. Therefore, implementing these diffusion methods should prove beneficial to small business owners and operators.

Small businesses can benefit from understanding the factors associated with diffusion, including Roger's five major elements. Another important consideration for small businesses is the cultural issues surrounding their organizations and resources issues that will prove important to employees. By understanding these issues, small businesses may be able to tailor the adoption and diffusion of technologies or a knowledge sharing system to their unique situation, thus improving their chances for success.

While the small business studied represents only one situation, it is interesting that several consistencies with the literature emerged, lending some possible connections and conclusions between the effective diffusion of knowledge management systems and a causal relationship with improved performance. The implication is that knowledge management can facilitate the creation of a learning organization or an intelligent enterprise. This improves creativity and innovation that further enhances productivity, problem solving and customer satisfaction. By promoting a culture that continually learns and improves, organizations can develop a sustainable competitive advantage that is not easily imitated by competitors.

REFERENCES

Advocacy, O. (2001). *Small business economic indicators for 1999: Executive summary*. Washington, DC: U.S. Small Business Administration.

Alange, S., Jacobsson, S., & Jarnehammar, A. (1998). Some aspects of an analytical framework for studying the diffusion of organizational innovations. *Technology Analysis and Strategic Management, 10*(1), 3-21.

Anderson, V., & Boocock, G (2002). Small firms and internationalisation: Learning to manage and managing to learn. *Human Resource Management Journal, 12*(3), 5-24.

Beckman, J. (1999). The current state of knowledge management. In J. Liebowitz (Ed.), *Knowledge management handbook.* Boca Raton: CRC Press.

Chaston, I., Badger, B., & Sadler-Smith, E. (2001). Organizational learning: An empirical assessment of process in small U.K. manufacturing. *Journal of Small Business Management, 39*(2), 139-151.

Cohen, W.M., & Levinthal, D.A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly, 35*(1), 128-153.

Daugherty, P.J., Germain, R., & Droge, C. (1995). Predicting EDI technology adoption in logistics management: The influence of context and structure. *Logistics and Transportation Review, 331*(4), 309-325.

Davenport, T.H., & Prusak, L. (2000). *Working knowledge: How organizations manage what they know.* Boston: Harvard Business School Press.

Guimaraes, T. (2000). The impact of competitive intelligence and IS support in changing small business organizations. *Logistics Information Management, 13*(3), 117-130.

Holland, R. (1998). *Planning against a business failure.* Agricultural Extension Service, University of Tennessee, ADC Info #24.

Jones, N., Herschel, R., & Moesel, D. (2003) Using "knowledge champions" to facilitate knowledge management. *Journal of Knowledge Management, 7*(1), 49-63.

Karahanna, E., Straub, D.W., & Chervany, N.L. (1999). Information technology adoption over time: A cross-sectional comparison of pre-adoption and post-adoption beliefs. *MIS Quarterly, 23*(2), 183-213.

Kolle, C. (2001). SMEs are the 'backbone of the economy'. *Asian Business, 37*(3), 68-69.

Kwon, B. (2001). The post-WTC economy: How long a slump? *Fortune Small Business, 11*(9), 18-24.

Mcgee, M.K. (1999). Lessons from a cultural revolution-Proctor & Gable is looking to IT to change its entrenched culture- and vice versa. *Information Week,* 46-50.

McGregor, J., & Gomes, C. (1999). Technology uptake in small and medium-sized enterprises: Some evidence from New Zealand. *Journal of Small Business Management, 37*(3), 94-102.

Pan, S.L., & Scarbrough, H. (1998). A socio-technical view of knowledge-sharing at Buckman Laboratories. *Journal of Knowledge Management, 2*(1), 55-66.

Pan, S.L., & Scarbrough, H. (1999). Knowledge management in practice: An exploratory case study. *Technology Analysis & Strategic Management, 11*(3), 359-374.

Puccinelli, B. (1998). Overcoming resistance to change. *Inform, 12*(8), 40-41.

Reisenberger, J.R. (1998). Executive insights: Knowledge— the source of sustainable competitive advantage. *Journal of International Marketing, 6*(3), 94-107.

Rogers, E.M. (1995). *Diffusion of innovations* (4th ed.). New York: The Free Press.

Scheraga, D. (1998). Knowledge management competitive advantages become a key issue. *Chemical Market Reporter, 254*(17), 3-6.

Thong, J.Y.L. (1999). An integrated model of information systems adoption in small businesses. *Journal of Management Information Systems, 15*(4), 187-214.

Tiwana, A. (2000). *The knowledge management toolkit.* Upper Saddle River, NJ: Prentice Hall PTR.

Wigg, K.M. (1999). Introducing knowledge management into the enterprise. In J. Liebowitz (Ed.), *Knowledge management handbook.* Boca Raton: CRC Press.

Wong, C., Marshall, N., Alderman, N., & Thwaites, A. (1997). Management training in small and medium sized enterprises: Methodological and conceptual issues. *International Journal of Human Resource Management, 8*(1), 44-65.

Wong, W.L.P., & Radcliffe, D.F. (2000). The tacit nature of design knowledge. *Technology Analysis & Strategic Management, 12*(4), 493-512.

KEY TERMS

Absorptive Capacity: An organization's ability to absorb new knowledge; often based on the prior experience and knowledge base of its employees. The greater the absorptive capacity of an organization, the greater its ability to learn and adapt to changing market forces.

Adoption: Changes in employee attitudes, perceptions, and actions that lead them to try new practices, activities, or innovations that are different from their normal routines or behaviors.

Diffusion: The process by which new behaviors, innovations, or practices spread through the organization as people learn about them from other employees and try them out.

Intelligent Enterprise: An organization that is able to understand changing environments, and adapt with agility and resourcefulness in order to outperform competitors.

Knowledge Management: An organization's ability to capture, organize and disseminate knowledge to help create and maintain competitive advantage.

Learning Organization: An organization that continually strives to create a culture and environment that promotes learning, exploration, and innovation and that continuously transforms itself to achieve superior competitive performance.

Organizational Culture: The set of shared values and norms within an organization that shape its behaviors, attitudes, and actions.

Relative Advantage: The degree to which an innovation or process is perceived as being superior to that which preceded it.

SMEs Amidst Global Technological Changes

Nabeel A. Y. Al-Qirim
United Arab Emirates University, UAE

BACKGROUND AND IMPLICATIONS

In small countries such as New Zealand, small to medium-sized enterprises (SMEs) are defined as enterprises employing 19 or fewer employees. Small enterprises are defined as those employing zero to five full-time employees (FTEs) (often called microbusinesses), and medium-sized enterprises as those employing six to nineteen FTEs. Other countries, such as the United States and European countries, define their SMEs as having a much larger number of employees (200–500 or fewer).

SMEs contribute significantly to the economies and to the employment levels of different countries in the world. For example, SMEs constitute around 95 percent of enterprises and account for 60–70 percent of employment within the countries of the Organisation for Economic Cooperation and Development (OECD, 1997) and other countries across the globe, including the United States. Not to forget that SMEs are usually the source of most of the profound inventions and innovations (Iacovou, Benbasat, & Dexter, 1995).

Historically, SMEs have been accused of being uncritical about the strategic importance of IT and its use in their businesses. This laggardness in adopting or using IT in business was attributed to various organisational, technological, and environmental deficiencies in SMEs. The recent emergence of the Internet, in general, and the Web, in particular, revolutionises business activities (Abell & Lim, 1996) and promises to provide unprecedented opportunities to SMEs to expand in scope and in market reach.

However, despite the apparent media hype (Premkumar & Roberts, 1999) and the enthusiasm among academicians (Adam & Deans, 2000; Abell & Lim, 1996; *Infotech Weekly*, 1997; Poon & Swatman, 1999a) and professionals (Deloitte, 2000; IDC, 1998; PWHC, 1999) about electronic commerce (EC), the published EC research portrayed a gloomy picture about EC uptake and use by SMEs. Thus, investigating reasons behind such laggardness in adopting and in using EC effectively is essential. This research attempts to highlight some of the important issues that could assist in bridging the existing divide between SMEs and EC. These issues could be of interest to SMEs and to other stakeholders interested in SMEs and EC.

ELECTRONIC COMMERCE SUCCESS IN SMES

In the SMEs scenario, different research emphasised the different EC advantages to SMEs (Abell & Black, 1997; Abell & Lim, 1996; Adam & Deans, 2000; Deloitte, 2000; Poon & Swatman, 1997, 1998, 1999a,b; PWHC, 1999):

1. The Internet is an efficient communication medium and a vast resource for information. The SMEs could use e-mail technology to communicate efficiently with their buyers and suppliers, reducing communication costs, including the buying of expensive equipment (e.g., fax/telex).
2. The Internet provides added-value services to customers/partners/suppliers by providing different primary/supplementary information about the organisation's industry, products, and services on their Web sites. This could result in increasing the loyalty and the stickiness of their customers (customer resource management; CRM). The preceding tangible and intangible tactics are of strategic importance in retaining and increasing customer bases by increasing switching costs.
3. The Internet would provide new opportunities to SMEs, otherwise not possible before the introduction of the Internet, such as the ability to reach global markets and the ability to mass-customise products and services to appeal to the different tastes of global consumers.
4. SMEs would adopt EC for image-enhancement purposes. Having an Internet account (URL, dot-com, Web page) and printing an e-mail address on business cards and letterhead were reported as major drivers as well. Whether the SMEs were able to elevate from such initial depiction to a more strategic posture in adopting more strategic EC initiatives, such as selling and buying online, is worth further investigation from the perspective of the different countries.

On the other hand, the EC research highlighted the following impediments:

1. Technological impediments: e.g., security (privacy concerns, viruses, e-payments), legalities (enforceability of contracts, confirmations of receipt, prosecutions), policies (lack of global or unified standards), telecommunication services [bandwidth, convergence, reliability and quality of services (QoS)]
2. Organisational impediments: cost, busy nature, small size and limited resources, lack of knowledge/expertise about EC
3. Environmental impediments: Relating to the lack of regulatory frameworks pertinent to the above technological impediments highlighted in (1), above, either at the one-country level or even at the global level.

In the light of the above advantages and impediments, most of the existing EC research found most of the SMEs not witnessing real benefits (direct sales and tangible profits) in the short term due to difficulties in selling products over the Internet (Adam & Deans, 2000; Poon & Swatman, 1998, 1999a). Face-to-face interactions with customers and buyers proved to be more dominating than electronic interfaces (Ba et al., 1999; Poon & Swatman, 1997). They found the key motives for SMEs to adopt EC are the long-term indirect benefits, e.g., ongoing business transformation and new business initiatives (new opportunities), which could resemble a preparatory stage (infrastructure) for the long-run direct benefits stage (secure returning customers and form long-term business partnerships) (Poon & Swatman, 1998, 1999a). However, the biggest challenge for the SMEs here is to succeed in moving from such simple and preparatory EC initiatives (driven mostly by hype from the media, professionals, and researchers) to more sophisticated and strategic initiatives (e.g., efficiency à effectiveness à strategic advantage).

On the one hand, having EC requires an apparent investment in different areas: technological infrastructure upgrades or replacement, EC integration with existing IT systems, EC consultants, investments in bandwidth and applications (Web site, intranet, extranet, etc.). However, this considerable investment in the EC infrastructure is necessary to make it possible to process information efficiently, handle heavy traffic, and deliver satisfactory performance. SMEs would perceive this to be an expensive endeavour and, hence, represent a barrier to EC adoption (MOED, 2000; PWHC, 1999). It is worth mentioning here that unlike the investment in information science/information technology (IS/IT), which requires high initial investment and smaller ongoing maintenance and support costs, EC would require considerable continued investments in upgrading, overhauling, and replacing the whole EC system with an innovation or new designs, etc. Most probably, the investment in EC would materialise in the long term only as highlighted earlier (Poon & Swatman, 1998, 1999). However, this depends on different factors, such as the ability to develop economies of scale (Ba et al., 1999; Poon, 2000), e.g., having a well-established online customer base and ongoing business that enables the firm to sell massively and cheaply at the same time.

With the introduction of new EC technology like the intranet, Internet electronic data interchange (EDI), extranet, Web site, etc., there would be some fundamental changes in work processes and current practices (Alexander, 1999; Behrendorff & Rahman, 1999). EC is not only a new way of selling and marketing, but also a new way of thinking, which requires a change of mindset. Teo, Tan, and Buk (1998) pointed to the fact that organisations attempting to adopt the Internet should expect a possible change in communication and culture patterns. EC is changing the way business is conducted, even with individual customers. Firms that are able to streamline their products or processes or delivery agents on the Internet will be able to shift entirely to the pure EC arena (Choi et al., 1997). The success stories of small businesses using the Internet are apparent and are publicised and reported by the media. However, most of the businesses existing on the Internet are not necessarily transacting information-based products only, but rather complementing the sale and the delivery of a physical product with such things as publishing information about the usability of a physical product (e.g., user manuals), tracking the shipment, etc. (Teo et al., 1998).

Most of the IS literature on SMEs (Blili & Raymond, 1993; Cragg & King, 1992, 1993; Harrison et al., 1997; Jarvenpaa & Ives, 1991; Thong, 1999; Thong & Yap, 1995, 1996) and EC in SMEs (Poon & Swatman, 1998, 1997, 1999a, 1999b) emphasises the role and the characteristics of the manager (usually the owner) as a product champion. Poon and Swatman (1998, 1999a) pointed to the manager's role in their EC study, where they found direct management involvement was the norm in the different cases. Although the managers of small business lack formal IT qualifications and training, they were champions in adopting EC, specifically in microbusinesses, where the sole decision maker was the director of the business.

Due to the recent nature of EC, it is expected that the adoption decision for EC would include some sort of high-risk elements. Hence, the adoption decision for EC would require a risk-taking manager. Poon and Swatman (1997, 1998, 1999a) found that the entrepreneurial perspective differed between the different firms in their study. Managers/owners embraced EC technology and attempted to exploit it to the maximum. The managers who championed Internet adoption in their organisations demonstrated an innovative and risk-taking attitude toward EC, despite lacking formal IT training.

Adam and Deans (2000) and Poon and Swatman (1998) pointed to the market scope of small business, where SMEs transacting with international markets would perceive many advantages from the Internet, such as cost savings and market communication in comparison with other SMEs operating in local markets. In this scenario, EC is perceived to increase global competition and provide different opportunities to SMEs. Poon and Swatman (1999a) asserted that if a small business retained a high percentage of customers and competitors online, this would increase the chances of adopting EC.

The field of EC is relatively new, and the actual functioning and utilisation of EC technologies are still unknown to most organisations (Teo et al., 1998), including SMEs. Therefore, it is expected that SMEs planning to adopt EC would seek assistance from consultants and vendors in the industry in different areas, such as planning and strategy, training, development, and implementation (Deloitte, 2000). Determining how efficient the technology vendors are in providing feasible and well-integrated EC products and services to SMEs is worth investigating in different countries.

CONCLUSION

SMEs contribute significantly to the national economies and to the employment levels of different countries and represent a viable source for inventions and innovations. The emergence of EC in the early 1990s could provide different opportunities to the small business sector to overcome its inadequacies. However, in review of the electronic commerce/business (EC) literature in organisations, in general, and in SMEs, specifically, it was observed that the available research portrayed a gloomy picture about EC uptake and use by SMEs. Therefore, this research attempted, by reviewing recent EC research, to depict an agenda for EC success in SMEs. By following the suggested guidelines in this research, SMEs could be in a better position to assess the viability of the new EC phenomenon to their survivability in the long term. Specifically, these points are addressed to the managers/owners of the SMEs in identifying the different perspectives surrounding the new innovations. These factors are of high importance to researchers, SMEs, professionals (including educational institutions), and policy makers in driving SMEs and EC forward.

REFERENCES

Abell, W., & Black, S. (1997). Business use of the Internet in New Zealand: A follow-up study. Retrieved August 8, 2000, from *http://www.scu.edu.au/ausweb96/business/abell/paper.htm*

Abell, W., & Lim, L. (1996). Business use of the Internet in New Zealand: An exploratory study. Retrieved August 8, 2000, from *http://www.scu.edu.au/ausweb96/business/abell/paper.htm*

Adam, S., & Deans, K. (2000). Online business in Australia and New Zealand: Crossing a chasm AusWeb2k—The Sixth Australian World Wide Web conference,

Alexander, A. (1999, December). Tuning small business for e-Commerce: Consultants say business consulting is essential, even in e-commerce. *Accounting Technology*, *15*(11), 48–53.

Ba, S., Whinston, A., & Zhang, H. (1999, December). Small business in the electronic marketplace: A blue print for survival. *Texas Business Review*. University of Texas, Austin.

Behrendorff, G., & Rahman, S. (1999). Adoption of electronic commerce by small to medium enterprises in Australia. In F. Tan, P. Corbett, & Y. Wong (Eds.), *Information technology diffusion in the Asia Pacific: Perspective on policy, electronic commerce and education* (pp. 130–147). Hershey, PA; London: Idea Group Publishing.

Blili, S., & Raymond, L. (1993). Information technology: Threats and opportunities for small and medium-sized enterprises. *International Journal of Information Management, 13*, 439–448.

Choi, S., Stahl, D., & Whinston, A. (1997). *The economic of electronic commerce*. New York: Macmillan Technical Publishing.

Cragg, P., & King, M. (1992). Information systems sophistication and financial performance of small engineering firms. *European Journal of Information Systems, 1*(6), 417–426.

Cragg, P., & King, M. (1993). Small-firm computing: Motivators and inhibitors. *MIS Quarterly*, March.

Deloitte Touche Tohmatsu. (2000). Deloitte e-Business survey: Insights and issues facing New Zealand business. Retrieved August 8, 2000, from *http://www.deloitte.co.nz/images/acrobat/survey.pdf*

Iacovou, C., Benbasat, I., & Dexter, A. (1995, December). Electronic data interchange and small organisations: Adoption and impact of technology. *MIS Quarterly*, 465–485.

Infotech Weekly. (1997, April 1). New Zealand Internet use. Retrieved May 15, 2000, from *http://www.nua.net/surveys/index.cgi?f=VS&art_id=863080905&rel=true*

International Data Corporation (IDC). (1998). Ecommerce booming in New Zealand. Nua Internet Services: Retrieved April 30, 1998, from http://www.nua.ie/surveys/index.cgi?f=VS&art_id=905354498&rel=true. Retrieved May 15, 2000, from *http://www.nua.ie/surveys/index.cgi?f=VS&art_id=905354498&rel=true*

Jarvenpaa, L., & Ives, B. (1991, June). Executive involvement and participation in the management of information technology. *MIS Quarterly, 15*(2), 205–227.

OECD. (1997). Small business, job creation and growth: Facts, obstacles and best practices.

Poon, S. (2000). Business environment and Internet commerce benefits—A small business perspective. *European Journal of Information Systems*, *9*, 72–81.

Poon, S., & Swatman, P. (1997). Internet-based small business communication. *International Journal of Electronic Commerce*, *7*(2), 5–21.

Poon, S., & Swatman, P. (1998). A combined-method study of small business Internet commerce. *International Journal of Electronic Commerce*, *2*(3), 31–46.

Poon, S., & Swatman, P. (1999a). An exploratory study of small business Internet commerce issues. *Information & Management*, *35*, 9–18.

Poon, S., & Swatman, P. (1999b). A longitudinal study of expectations in small business Internet commerce. *International Journal of Electronic Commerce, 3*(3), 21–33.

Premkumar, G., & Roberts, M. (1999). Adoption of new information technologies in rural small businesses. *The International Journal of Management Science (OMEGA*), *27*, 467–484.

PWHC (Pricewaterhousecoopers). (1999, September 24). SME electronic commerce study (TEL05/97T). Retrieved April 10, 2000, from *http://apec.pwcglobal.com/sme.html*

Rihga Colonial Club Resort, Cairns, June 12–17. Retrieved August 8, 2000, from *http://ausweb.scu.edu.au/aw2k/papers/adam/paper.html*

Teo, T., Tan, M., & Buk, W. (1998). A contingency model of Internet adoption in Singapore. *International Journal of Electronic Commerce*, *2*(2), 95–118.

Thong, J. (1999). An integrated model of information systems adoption in small business. *Journal of Management Information Systems, 15*(4), 187–214.

Thong, J., & Yap, C. (1995). CEO characteristics, organisational characteristics and information technology adoption in small business. *Omega, International Journal of Management Sciences*, *23*(4), 429–442.

Thong, J., & Yap, C. (1996). Information technology adoption by small business: An empirical study. In K. Kautz & J. Pries-Heje (Eds.), *Diffusion and adoption of information technology* (pp. 160–175). London: Chapman & Hall.

KEY TERMS

Economical Importance of SMEs: SMEs contribute significantly to the economies and to the employment level of different countries in the world. For example, SMEs constitute around 95% of enterprises and account for 60% to 70% of employment within the countries of the OECD (OECD, 1997) and other countries across the global including the U.S. Not to forget the SMEs are usually the source for most of the profound inventions and innovations (Iacovou, Benbasat, & Dexter, 1995).

IT/E-Commerce Adoption and Use in SMEs: Historically, SMEs have always been accused of being uncritical about the strategic importance of IT and its use in their businesses. This laggardness in adopting or using IT in business was attributed to various organizational, managerial, technological and environmental deficiencies in SMEs. The recent emergence of the Internet in general and the Web in particular revolutionizes business activities (Abell & Lim, 1996) and promises to provide unprecedented opportunities to SMEs to expand in scope and in market reach.

Small Business Internet Commerce: the use of Internet technology and applications to support business activities of a small firm (Poon, 1999).

Small to Medium-Sized Enterprises (SMEs): In small countries such as New Zealand, SMEs are defined as enterprises employing 19 or fewer employees. Small enterprises are defined as those employing zero to five full-time employees (FTEs) (often called microbusinesses) and medium-sized enterprises as those employing six to nineteen FTEs. Other countries, such as the United States and European countries, define their SMEs as having a much larger number of employees (200–500 or fewer).

SMEs in Knowledge-Based Economies

S

Sushil K. Sharma
Ball State University, USA

Nilmini Wickramasinghe
Clevelend State University, USA

Jatinder N.D. Gupta
University of Alabama in Huntsville, USA

INTRODUCTION

Today's Information Age is having a dramatic effect on businesses as well as on the life styles of people. Globalization, rapid technological change and the importance of knowledge in gaining and sustaining competitive advantage characterize this information age. Traditionally, economists have seen capital, labor, and natural resources as the essential ingredients for economic enterprise. The new economy of the 21st century is increasingly based on knowledge with information, innovation, creativity and intellectual capitalism as its essential ingredients (Persaud, 2001). Today's modern economy then, is based more on intangibles, information, innovation, and creativity, and their abilities to facilitate expanding economic potential (Persaud, 2001) and the exploitation of ideas rather than material things. Many new terms have been coined for this new economy such as "knowledge-based economy", "borderless economy", "weightless economy". and "digital economy" to name a few (Woodall, 2000). This new economy seems to defy the basic economic law of scarcity which means, if a physical object is sold, the seller ceases to own it. In this new economy, however, when an idea is sold, the seller still possesses it and can sell it over and over again (Woodall, 2000). Traditional economic theory assumes that most industries run into "diminishing returns" at some point because unit costs start to rise, so no one firm can corner the market. In the new economy, knowledge-based products and services have "increasing returns" because knowledge-based products are expensive to produce for the first time, but cheap to reproduce. High fixed costs and negligible variable costs give these industries vast potential for economies of scale (Sharma, Wickramasinghe & Gupta, 2003; Woodall, 2000).

The shift to a knowledge-based economy results largely from developments in information and communications technologies. A company's knowledge assets are inherent in the creativity of its knowledge workers combined with technological and market know-how (Halliday, 2001). Information can now be delivered with such speed that companies must develop their knowledge assets to solve competitive problems. Knowledge-based economies offer huge opportunities for small and medium-sized enterprises (SMEs) to develop entirely new high-value products and services, add value to existing products and services, reduce costs, develop new export markets, and add value to existing activities. Implicit promises include access to world markets, low cost entry into new markets, and the ability to gain efficiencies in business processes. However, these promises may be illusory for most SMEs. Technological, organizational, and marketing hurdles are also making it more difficult for SMEs to succeed in knowledge-based economies (Acs, Carlsson, & Karlsson, 1999; Persaud, 2001). This article identifies those major factors that are hindering the success of SMEs in knowledge-based economies.

THE KNOWLEDGE-BASED ECONOMY

In a knowledge-based economy, knowledge drives the profits of the organizations for gaining and sustaining competitive advantage. Intellectual capital, that is, employees, their knowledge of products and services, and their creativity and innovity, is a crucial source of knowledge assets. The knowledge-based economy is all about adding ideas to products and turning new ideas into new products. Thus, the knowledge-based economy is all about adding ideas to products and turning new ideas into new products. Realizing the importance of knowledge assets, many companies have changed their traditional organization's structures. The traditional command and control model of management is rapidly being replaced by de-centralized teams of individuals motivated by their ownership in the companies (McGarvey, 2001; Sharma et al., 2003).

The new structure of the economy is emerging from the convergence of computing, communications and content. Products are becoming digital and markets are be-

Table 1. Key factors for SMEs

Factors	Description
Lack of Sustained Technology Developments	While borderless electronic trade and online trading hubs are opening up, many more opportunities for small- and medium-sized enterprises (Jordan, 2000), is proving to be a difficult task for SMEs in ever increasing technological environments. Only 35% of small businesses have an Internet Web site, and of those only 2% have sites with e-commerce transactions (Kleindl, 2000). Asian SMEs are even further behind the times due to poor penetration of Internet technologies and slow growth of electronic commerce infrastructure (Jordan, 2000).
Lack of Proper Information	SMEs require proper guidance in making the right choice of technology suited to their needs. Many SMEs do not have the ability, time or energy to shift to new technology either due to lack of expertise at their own level or absence of proper guidance, advice and support from big companies. Not only do they lack information about the availability and sources of the new technology, but they also lack a resource base for searching for partners (Jordan, 2000).
Inertia to Change	Many small businesses are too content to change, despite the often-surprising rewards of taking that step into the future. For example, SMEs can benefit from powerful enterprise resource planning (ERP) tools and customer relationship management (CRM) software technology solutions without investing in them, because they can rely upon rental-based enterprise solutions such as application service providers (ASPs) to provide these solutions.
Slow to Adapt E-commerce	The Internet and the Web environments are bringing fundamental changes in traditional business models. SMEs often lack the resources of larger firms and may not have the brand name recognition of Internet first movers.
Encroachment by Big Corporations into Niche Areas	The Internet and the online environments are having a major impact on how businesses operate. The Web environments allow customers to easily search and find competitive information and new sources of supply (Kleindl, 2000). SMEs have traditionally been able to gain the advantage over larger competitors by developing personalized relationships with customers, customizing their offerings, and efficiently targeting niche markets (Kleindl, 2000).
Commoditization of Suppliers	SMEs face additional pressure due to the commoditization of suppliers. This occurs when the customer can find a large number of suppliers with relatively similar offerings and then have them bid for the sale.
Competition from First Movers	SMEs face competition with online competitors who have first mover advantages. Firms that are Web pioneers already have cost advantages as they move along experience curves that relate to personnel training and management in Web practices.
Weak Supply Chains	Online communication and developing links with suppliers are impacting almost all businesses including SMEs.
Lack of Expertise	The Internet allows SMEs to communicate and transact business at any time. SMEs can use the Internet and information technologies to create unique products and services for their customers that differ from the competition.

Table 1. Key factors for SMEs (continued)

Barriers to International Trade	Export and internationalisation are important to the survival of many SMEs because of their tremendous potential for enhancing sales growth, increasing efficiency and improving quality (Masurel, 2001). SMEs find it difficult to form international partnerships, or may lack managerial experience and competence, and have difficulty in gathering information.
Lack of Awareness of Management Processes	As our economy becomes more connected to the global marketplace, SMEs need awareness of management processes and tools to create competitive advantage (Monk, 2000).
Long Lead Times	The rapid economic and technological developments in the globally-oriented business world make the strategic use of information technology (IT) essential yet SMEs are somewhat behind regarding the application of IT (Knol & Stroeken, 2001).
Poor Management and Accounting Practices	The role of finance has been viewed as a critical element for the development and participation of small and medium-sized enterprises (Cook, 2001). Levy has highlighted the limited access to financial resources available to smaller enterprises compared to larger organizations and the consequences for their growth and development (Levy, 1993).
Intense Market Competition in International Trade	Many SMEs are relatively inexperienced operators in international trade. SMEs face the challenge of intense market competition for trade promotion. Adapting to a new global trading environment presents a challenge to SMEs.
Inefficient Value Chains	The rapid development of the World Wide Web as a communication and marketing medium offers small and medium-sized enterprises tremendous opportunities of access to world markets, low-cost entry into new markets, and the ability to gain efficiencies in business processes (Kleindl, 2000). However, SMEs require restructuring of value chains because business models of a knowledge-based economy are very different than the earlier traditional businesses.

coming electronic. In the old economy, information flow was physical: cash, checks, invoices, bills, reports, face-to-face meetings, analog telephone calls, radio or television transmissions, and so forth. Relationships with trading partners, customers and suppliers, distribution networks, intellectual property, patents, image, and so forth, are all elements of a knowledge economy. These elements represent intellectual capital. The knowledge-based economy is based on the application of human know-how to everything we produce; and hence, in this new economy, human expertise and ideas create more and more of the economy's added value. The knowledge content of products and services is growing significantly, as consumer ideas, information, and technology become part of products. In the new economy, the key assets of the organization are intellectual assets in the form of knowledge (Anonymous, 2001; Cowey, 2000; Gold, Malhotra, & Segars, 2001; Granstrand, 2000; McGarvey, 2001; Sharma et al., 2003).

KEY ISSUES AFFECTING PARTICIPATION AND SUCCESS OF SMES IN THIS ECONOMY

It is not enough just to highlight the critical elements of today's knowledge-based economy; we must now understand the various factors affecting the participation and success of SMEs in this economy. We identify 15 important factors which are summarized in Table 1.

CONCLUSION

We believe that SMEs should find many more new opportunities in today's knowledge-based economy, but at the same time, may be threatened in competitive environments that can force SMEs to modify or completely abandon many current business practices. SMEs must consider new business models that take advantage of existing and emerging Internet-based technologies in order to stay competitive. SMEs have to introduce new technologies to enter into global markets and find ways to attract customers if they want to compete in global scenarios. As larger organizations are entering into Internet-enabled businesses, SMEs have to upgrade their tools and techniques to equip themselves with the abilities to compete with large organizations. In addition, more has to be done to encourage stability and growth in the small to medium-sized business sector, and to establish globally competitive firms. Many small manufacturers do not have the resources required to transform their businesses in response to the digital economy's rapid evolution. It may be necessary for SMEs to redesign many of their traditional business processes to compete in the new economy. SMEs face the challenges of fast changing technology and business scenarios in the knowledge-based economy. Combining the right business strategy along with the most effective technology tools would improve the chances of SMEs participation in knowledge-based economy.

REFERENCES

Acs, Z.J., Carlsson, B., & Karlsson, C. (1999). *The linkages among entrepreneurship, SMEs and the macroeconomy*, Cambridge: Cambridge University Press.

Anonymous (2000). Web returns for SMEs. *Management Accounting*, 78(5), 14.

Anonymous (2001). What knowledge economy? *Information World Review*, *170*, 1.

Boulton, R.E.S., Libert, B.D.. & Samek, S.M. (2000) A business model for the new economy. *The Journal of Business Strategy, 21*(4), 29-35.

Budhwani, K. (2001). Becoming part of the e-generation, *CMA Management*, *75*(4), 24-27.

Clieaf, M.V. (2001). Leading & creating value in the knowledge economy. *Ivey Business Journal*, *65*(5), 54-59.

Cook, P. (2001). Finance and small and medium-sized enterprise in developing countries. *Journal of Developmental Entrepreneurship*, *6*(1), 17-40.

Cooper, J. (2001). Enhancing the competitive success of Canadian SMEs. *CMA Management*, *75*(5), 16-21.

Cowey, M. (2000). Knowledge economy—fact or fad? *New Zealand Management, 47*(4), 54-55.

Cowey, M. (2001). Managing in the new economy. *New Zealand Management*, *48*(3), 66-67.

Duhan, S., Levy, M., & Powell, P. (2001). Information systems strategies in knowledge-based SMEs: The role of core competencies. *European Journal of Information Systems*, *10*(1), 25-40.

Fjeldstad, O.D. & Haanaes, K. (2001). Strategy tradeoffs in the knowledge and network economy. *Business Strategy Review*, *12*(1), 1-10.

Fontes, M. & Coombs, R. (2001). Contribution of new technology-based firms to the strenghtening of technological capabilities in intermediate economies. *Research Policy*, *30*(1), 79-97.

Gold, A.H., Malhotra, A., & Segars, A.H. (2001). Knowledge management: An organizational capabilities perspective. *Journal of Management Information Systems*, *18*(1), 185-214.

Granstrand, O. (2000). The shift towards intellectual capitalism - The role of infocom technologies. *Research Policy*, *29*(9), 1061-1080.

Gulisano, V. (2001). Succeeding in the digital logistics age. *World Trade*, *14*(5), 42-43.

Gupta, A.K. & Govindarajan, V. (2001). Converting global presence into global competitive advantage. *The Academy of Management Executive*, *15*(2), 45-58.

Halliday, L. (2001). An unprecedented opportunity. *Information World Review*, *167*, 18-19.

Heneman, R.L., Tansky, J.W., & Michael Camp, S. (2000). Human resource management practices in small and medium-sized enterprises: Unanswered questions and future research perspectives. *Entrepreneurship Theory and Practice*, *25*(1), 11-26.

Jordan, T. (2000). Transform your business. *Asian Business*, *36*(5), 30-32.

Kandell, J. (2000). CRM, ERM, one-to-one-decoding relationship management theory and technology. *Trusts & Estates*, *139*(4), 49-53.

Karkoviata, L. (2001). Making the customer king. *Asian Business*, *37*(2), 47-48.

Klaas, B.S., McClendon, J., & Gainey, T.W. (2000). Managing HR in the small and medium enterprise: The impact of professional employer organizations. *Entrepreneurship Theory and Practice*, *25*(1), 107-124.

Kleindl, B. (2000). Competitive dynamics and new business models for SMEs in the virtual marketplace. *Journal of Developmental Entrepreneurship*, *5*(1), 73-85.

Knol, W.H.C. & Stroeken, J.H.M. (2001). The diffusion and adoption of information technology in small- and medium-sized enterprises though IT scenarios. *Technology Analysis & Strategic Management*, *13*(2), 227-246.

Kupiec, E. (2000) Shifting strategies: Challenging the traditional business knowledge. *CMA Management, 74*(4), 15-17.

Levy, B. (1993). Obstacles to developing indigenous small and medium enterprises: An empirical assessment. *The World Bank Economic Review, 7*(1), 65-83.

Lu, J.W. & Beamish, P.W. (2001). The internationalization and performance of SMEs. *Strategic Management Journal*, *22*(6/7), 565-586.

Mars, R.D.D. (2000). SMEs flourish. *Asian Business*, *36*(5), 53-54.

Masurel, E. (2001). Export behaviour of service sector SMEs. *International Small Business Journal*, *19*(2), 80-84.

McAdam, R. (2000). The implementation of reengineering in SMEs: A grounded study. *International Small Business Journal*, *18*(4), 29-45.

McGarvey, R. (2001). New corporate ethics for the new economy. *World Trade*, *14*(3), 43.

McMahon, R.G.P. (2001). Business growth and performance and the financial reporting practices of Australian manufacturing SMEs. *Journal of Small Business Management*, *39*(2), 152-164.

Monk, R. (2000). Why small businesses fail, CMA Management; Hamilton.

Persaud, A. (2001). The knowledge gap. *Foreign Affairs, 80*(2), 107-117.

Peursem, K.A.V. & Wells, P.K. (2000). Contracting practices in professional accounting SMEs: An analysis of New Zealand firms. *International Small Business Journal*, 19(1), 68-82.

Said, A.J. (2000). Helping small firms trade effectively with the Internet. *International Trade Forum*, 3, 16-19.

Sharma, S.K., Wickramasinghe, N., & Gupta, J.N.D. (2003). What should SMEs do to succeed in today's knowledge-based economy. In N.A.Y. Al-Qirim (Ed.), *Electronic commerce in small to medium-sized enterprises: Frameworks, issues and implications* (pp.289-303). Hershey, PA: Idea Group Publishing.

Stewart, T.A. (2000). Three rules for managing in the real-time economy. *Fortune, 141*(9), 332-334.

Tapscott, D. (1996). *Digital economy – Promise and peril in the age of networked intelligence.* McGraw-Hill.

Warren, L. & Hutchinson, W.E. (2000). Success factors for high-technology SMEs: A case study from Australia. *Journal of Small Business Management*, *38*(3), 86-91.

Woodall, P. (2000). Survey: The new economy: Knowledge is power. *The Economist*, *356*(8189), 27-32.

Wickramasinghe, N. (2002). Do we practise what we preach: Are knowledge management systems in practice truly reflective of knowledge management systems in theory? *Business Process Management Journal*, forthcoming.

KEY TERMS

Customer Relationship Management (CRM): Simply stated, customer relationship management (CRM) is about finding, getting, and retaining customers. CRM are those aspects of a business strategy which relate to techniques and methods for attracting and retaining customers.

E-Commerce: The conduct of buying and selling of products and services by businesses and consumers over the internet. E simply means anything done electronically, usually via the Internet. E-commerce is the means of selling goods on the Internet, using Web pages.

Knowledge Economy: The knowledge-based economy is all about adding ideas to products and turning new ideas into new products. Relationships with trading partners, customers and suppliers, distribution networks, intellectual property, patents, image, and so forth, are all elements of a knowledge economy. These elements represent intellectual capital.

Learning Organizations: Those that have in place systems, mechanisms and processes, that are used to continually enhance their capabilities and those who work with it or for it, to achieve sustainable objectives " for themselves and the communities in which they participate.

Small and Medium-sized Enterprises (SMEs): As per European Commission definition, small and medium-sized enterprises, referred as "SMEs", are defined as enterprises which: have fewer than 250 employees; and have either an annual turnover not exceeding 40 million EURO, or an annual balance-sheet total not exceeding 27 million EURO as shown in Table 2.

Web Portal: Commonly referred to as simply a *portal*. Web portal is a Web site or service that offers a broad array of resources and services, such as e-mail, forums, search engines, and online shopping malls. The first Web portals were online services, such as AOL, that provided access to the Web, but now most of the traditional search engines have transformed themselves into Web portals to attract and keep a larger audience.

Table 2. SMEs

	Medium-sized	Small	Micro-enterprise
Max. number of employees	Max 250	Max. 50	Max. 10
Max. turnover (in million ECU)	40	7	-
Max. balance-sheet total (in million ECU)	27	5	-

Social Responsibility and the Technology Paradigm in Latin America

Herberto Ochoa-Morales
University of New Mexico, USA

INTRODUCTION

In Latin America, regional as well as multilateral integration schemes have a predominant role within integration agreements. A representation of this includes MERCOSUR: Brazil, Argentina, Uruguay and Paraguay. The Andean Community of Nations (CAN) is composed of Bolivia, Ecuador, Colombia, Peru, and Venezuela and the Group of Three (G3): Colombia, Mexico, and Venezuela. Another regional pact is the CARICOM, composed of English speaking countries within the Caribbean Basin. These organizations have the intent to establish, among other components, free trade areas, customs unions, common markets, and economic unions; all covenants that, in the future, may evolve into a political union (S.C.A. et al., 1998).

Under the scheme of regional integration, a high flow of goods, services, and investments between countries will be originated primarily under the format of foreign direct investments (FDIs). From the economic perspective, the outcome is trade and, therefore, stimulus to economic growth. By the year 2000, Latin America's regional agreements CAN and MERCOSUR, without considering other regional pacts with Chile, had a potential market of 310 million consumers (UN-CEPAL, 1999a). Chile's contribution alone is 15.2 million potential customers. It should be emphasized that the CAN countries will have, by the period 2000-2005, an average increase in population rate of 17.98 per thousand, while MERCOSUR will have 13.96 and Chile 11.8 per thousand increases respectively (UN-CEPAL, 1999b).

The research literature concurs on the importance of technology as main factor embedded in the productivity equation. The "digital gap" between developed countries (DC) and the less developed ones (LDC) is greater than the one built by economic indicators such as productivity; and socio-economic ones like "standard of living". In March 2000, the number of users on the Internet was approximately 304 million. The United States of America and Canada have 45%, Europe 27%, the Asia-Pacific region 23%, and Africa and the Middle East 1.5%. Latin America and the Caribbean hold 8% of the world population, but only 3.5% of Internet users and less than 1% of the global e-commerce, but at the same time the growth rate has been the highest in the world, and the number of users was 14-fold within the 1995 to 1999 period (UIT, 2000).

The literature emphasized the growth of e-commerce in the decade of the 90s that has occurred by improvement of computer-based power and convergence taking place with telecommunications. Nevertheless, there are other factors associated with the developments that include, but are not limited to, the role and social responsibilities of the public and private sectors in driving and sustaining infrastructure development. E-commerce has shown a rapid development in Latin America. Brazil reached 4 million users in 1999. This represents 50% of the interconnected population - Argentina 12%, and Chile 4%. It is necessary to emphasize the fact that 80% of electronic commerce is realized within six realms: supermarkets, books, hardware and software, electronic equipment, music, and financial services (UN-CEPAL, 2000b).

Another "gap" present in Latin America that has a great repercussion on the digital economy is the one that could be defined by its components: socio-economics and technology. Further, there is an uneven distribution of wealth between countries, and within them. A large price differential regarding telecommunications cost and coverture exists. The governments of the region have accomplished basic strides so the mass population will have access to the Internet. Peru has created The Peruvian Scientific Network, known by its Spanish abbreviation, RCP. The network is composed of 1,000 public centers that provide service to 40% of the network. In Argentina, the program argentina@internet.todos has approximately 1,000 tele-centers located in low income and remote areas (UIT, 2000).

The position of the Latin American countries within the framework of a knowledge-based society could be described as a transitional one. To acquire the objective of one, state intervention through laws and regulations, and furthermore, private and public actions will be essential in view of the peculiar contrasts in the region. It will be wearisome to expect the market forces alone to furnish the needed mechanisms. Also, the implementation of adequate legal framework that determines the rules and regulations, not only for the suppliers of services, but to compensate power concentration generated by the technology in the hands of the industrialized countries and multinational enterprise (MNE) are imperative (Katz & Ventura-Dias, 2000).

The literature concurs that computer information systems is a function of various parameters, and among them, the ones that could be identified as related to communication and diffusion are: cost of telephone service, and the structure and behavior of the market that compose the Internet services. There are least five relevant parameters in the market of information transmission that will be identifiable, which contribute to shape the Internet: 1) the carriers, 2) the access providers, 3) the service providers, 4) the content providers and 5) the end users. These schemes generate conflict and competition. International firms cover the first two levels; meanwhile, the rest are national enterprises within country members of regional or multilateral agreements (UN-CEPAL, 2000b).

BACKGROUND

In the last decade, the telecommunication sector in Latin America has grown enormously. Privatization and the development of new technologies have performed a critical role in this process. During the decade of the 90s, 2/3 of the countries of the region totally or partially privatized the telecommunication domain. At the same time, the arrival of new technologies such as cellular telephones and cable television has generated substantial changes in the sector. The case of Venezuela and Paraguay deserves special attention due to the fact that there are more cell phones than conventional ones (UIT, 2000).

In Latin America only 1/3 of all homes have telephone service. The growth and covertures of the telecommunication sector are functions of the regulatory framework in which they are developed. In many cases, monopolies have been created. Other good indicators of improvement in the sector are the digitalization of the telephone systems, an increase in the number of public telephones, and the improvement of repair time (UIT, 2000). Social factors have to be taken in consideration regarding the infrastructure of telecommunications; 25% of the region population lives on an income of $1 a day. The access to the Internet services in absolute terms is less than the U.S., although it is prohibited to the great majority of the population due to poverty. Government involvement could provide a solution to the problem of subsidizing services and the necessary hardware and software (UN-CEPAL, 2000b). A notion exists that the Internet revolution would narrow the gap between the world's rich and poor. The evidence shows that the opposite could happen and many developing countries are located on the wrong side of a widening knowledge gap if they do not act almost immediately. The DCs use different approaches to deal with the digital divide. The U.S. approach comes out of the "trade vector" because U.S. firms have large investments in the New Economy. They aggressively seek market expansions. The "state vector" represented by European firms emphasizes the state responsibility and sustainable development. The northern hemisphere countries', "donor states," approach is to find the best way to penetrate foreign markets and use their investments in the most cost-effective mode. The southern hemisphere countries, "host states," are concerned with how to attract investments that could generate growth and therefore jobs and wealth that will assist in the struggle to reduce social inequities (Conhaim, 2001).

Across the literature, authors agree on a definition of social responsibility as moral obligation of business organizations to seek goals that will provide common good for the communities that are beyond those required by business itself. According to Adams (2000), the business' main responsibility is the shareholders concern; therefore, society should not expect large investment or involvement within the communities where they function. Also, the author stated that business organizations look forward for the government to provide and maintain the necessary infrastructure for them to operate. Epstein (1998) confirmed that corporations are institutions that exist to fulfill societal purpose and the common good of the community is the underlying principle for their continued existence. The author defined corporate social policy process (CSPP) as a concept to provide the tools and operational framework to assist business managers in the consideration of social policies in the decision-making process. The foundation of CSPP is the internalization within business organizations' key elements from business ethics, corporate social responsibility and corporate social responsiveness (Epstein, 1998). The latter eliminates a vacuum in the decision-making process that could have as a result the instance that the literature recognized as "bounded rationality," which is the "good enough solution" for a particular problem.

Brazil has the privilege to count within its information technology assets The Committee for Democracy in Information Technology (CDI). This not-for-profit organization has as its goal to reduce the digital gap affecting individuals of low-income communities not only in the country but also throughout the world. In addition to bringing information technology (IT) to the less privileged, CDI promotes notions of human rights, literacy, ecology, health, and non-violence, among other important social teachings that will help to cover the social responsibility vacuum created by the lack of involvement of domestic businesses, MNEs, and public and private institutions. Currently, Brazil has 19 regional CDI centers with 311 schools of information technology and citizenship. Internationally, CDI operates 25 centers located in

Chile, Colombia, Japan, Mexico and Uruguay, for a grand total of 336 schools. The Japan Center located in Tokyo is mainly used to collect hardware that later is sent to LDCs. CDI was created in 1993 by a young professor of information technology, Rodrigo Baggio, with an initial slogan of "Computers for everyone" (*htpp://cdi.org.br*).

United Nations' Economic and Social Council, based on the decision 1999/281, resolved that the high level segment of the agenda for the year 2000 would be dedicated to "The Development and International Cooperation in the XXI Century: The Function of Information Technology in the Context of a World Economy Based in Knowledge". Therefore, representatives of Latin American and Caribbean Countries met in the town of Florianopolis, Santa Catarina, Brazil on June 20 to 21, 2000 to issue the guidelines to design and implement the necessary mechanism to move these countries into the "knowledgeable society" (UN-CEPAL, 2000a).

The mechanism of social responsibility will need to be based on an agenda that will contain several public policies to increase the efficiency and equity during the transition to a knowledge-based society. These include, but are not limited to, cost of telecommunication services, access to the digital network, and cost and accessibility to the computational structure. Education of the users at any level is necessary and access must be provided to the mass population with scarce financial resources for the information society. To reach the latter, the establishment of terminals in public places and community centers is necessary. All of these have to be performed within a legal framework that provides the essential elements to guarantee electronic transactions and therefore the ability to generate a large volume of trade using these media.

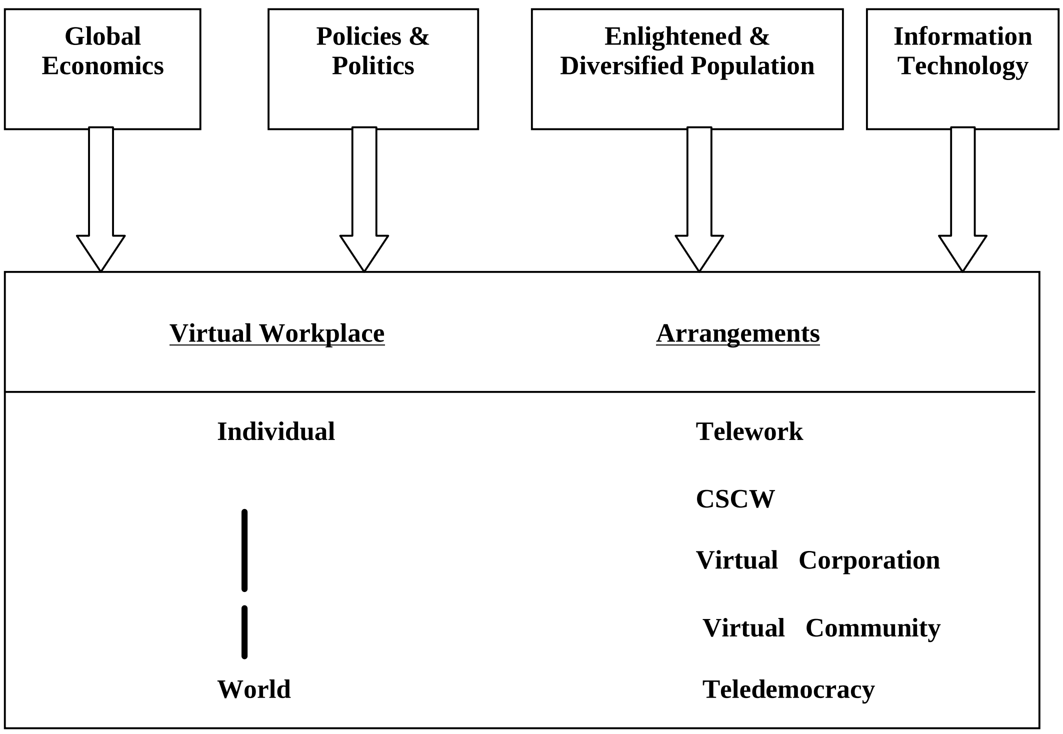
FUTURE TRENDS

The synergism generated by countries, members of regional and/or multilateral agreements allows them to obtain the necessary endowment to develop the infrastructure required for a virtual society. Also, this synergism contributes to dissipate the concentration of power originated by the technology in the hands of the DC and MNEs, and provides the necessary significance to establish policies that eliminated the vacuum created by the lack of involvement of business organizations regarding the social implications caused by the virtual society.

CONCLUSION

Driving Forces Model (Figure 1) depicts the framework in which the virtual society evolved (Igbaria, 1999). The

Figure 1. Driving Forces, 1999 ACM, Inc. Reprinted by permission



regional trading block already in place in Latin America will generate an expansion of business into global markets, creating a global economy in which new standards for trade will be present that include, but are not limited to, electronic payments like e-cash, and electronic data interchange (EDI) among businesses located elsewhere. Due to economic and social factors in Latin America, the role of regional governments to provide oversight for this new arrangement is not only preponderant but could be controversial. Elements such as the amount of control to be adopted by the government, regulations, and privatization will be relevant to the development of the knowledge-based society within the legal framework - same as the education and exposure of the population to the new virtual society (Ochoa-Morales, 2001). The Peruvian case portrays the importance of privatization versus a state monopoly in the telecommunication industry. It shows that privatization not only increases the amount of investments and the growth of many parameters within the telecommunication area, but also allows the mass population to have access to the technology (Ochoa-Morales, 2002). Any new policies should contain essential provisions addressing the importance of social responsibility and accountability. This will aid in reaching the goals of helping in the realm of social inequalities. Pertaining to the technology sector, a component that will exert large influence is the volatility of the communication segment due to the availability of new technology and changes thereof. The workplace arrangements such as tele-work, computer-supported cooperative work (CSCW), among others described in the model (Figure 1), are elements that will change the way business is done conventionally and will cause a great impact in social and cultural values within the context of society.

REFERENCES

Adams, J. D. (2000). Dominant institutions and their responsibility. *The Futurist, 34*(2), 65.

Conhaim, W. (2001). The global digital divide. *Information Today, 18*(7).

Committee for Democracy in Information Technology (CDI). (n.d.). *Institutional profile*. Retrieved October 2001 from *http://www.cdi.org.br*

Epstein, E.M. (1998, March). Business ethics and corporate social policy. *Business and Society, 37*(1).

Igbaria, M. (1999). The driving forces in the virtual society. *Communications of the ACM, 42*(12), 64-70.

Katz, J., & Ventura-Dias, V. (2000, June). America Latina y el Caribe en la transicion hacia una sociedad del conocimiento. Una agenda de politicas publicas. *Informe, LC/L.1383/E,* 1.

Ochoa-Morales, H.J. (2001). The digital gap between the industrialized countries and the less developed ones (LDC): The transition toward a knowledgeable society in Latin America. *Journal of Issues in Information Systems, 2,* 337-342.

Ochoa-Morales, H.J. (2002). The impact of reforms in the telecommunication sector and its effects on Latin America. *Journal of Issues in Information Systems, 3,* 483-489.

Secretaria de la Comunidad Andina (S.C.A.), Cooperacion Francesa y UN-CEPAL. (1998, May 26). Multilaterismo y regionalismo. *Seminario efectuado en Santa Fe de Bogota,* 1-2.

UIT. (2000, April). Indicadores de telecomunicaciones de Las Americas 2000. *Resumen Ejecutivo,* 1-22.

UN-CEPAL. (1999a, January). America Latina: Poblacion total, urbana y rural y porcentaje urbano por paises. Cuadro 11. *Boletin Demografico, 63,* 1-6. Retrieved May 2001, from *http://www.eclac.cl/publicaciones/Poblacion/2/LCG2052/BD63.11html*

UN-CEPAL. (1999b, January). America Latina: Tasa de crecimiento de la poblacion total, urbana y rural por paises. Cuadro 12. *Boletin Demografico, 63,* 1-3. Retrieved May 2001, from *http://www.eclac.cl/publicaciones/Poblacion/2/LCG2052/BD63.12html)*

UN-CEPAL. (2000a, June21-22). *Declaracion de Florianapolis. Reunion regional en tecnologia de informacion para el desarrollo,* Brasil (pp. 1-9).

UN-CEPAL. (2000b, June). Latin America and the Caribbean in the transition to a knowledge-based society: An agenda for public policy. *LC/L.1383,* 5-25.

KEY TERMS

Convergence: The occurrence taking place with computing and telecommunications that emphasizes the preeminence of computer-based information and telecommunications networks.

Digital Gap: The disparity or breach among countries generated by the lack of communication infrastructure and computer-based power that contribute to accentuate socio-economic differences.

Knowledge-Based Society: The UN Economic and Social Council decision 1999/281 defines the "function of information technology in the context of a world economy based in knowledge". This decision determines the importance of information technology in conjunction with knowledge to create a society in which these parameters will contribute to disseminate knowledge, create participation among countries and therefore among individuals, and generate economic growth. All the latter will take place within a societal framework that helps to preserve and increase the social and economic well-being of the inhabitants of the world.

Latin America: The part of the American continent located south of the United States border, which extends to the Patagonia.

Privatization: The transfer of government, public owned, or controlled institutions and entities to private enterprise. This was a common public policy in Latin America during the 90s to generate hard currency and support deregulation.

Social Responsibility: The moral obligation of business organizations to seek goals that will provide common good for the communities that are beyond those required by business itself.

Telecommunications: The transmission of data, text, sound, and images over long distances, and the physical infrastructure that supports such endeavor.

Socio–Cognitive Model of Trust

Rino Falcone
Institute for Cognitive Sciences and Technology, National Research Council of Italy, Italy

Cristiano Castelfranchi
Institute for Cognitive Sciences and Technology, National Research Council of Italy, Italy

INTRODUCTION

Humans have learned to cooperate in many ways and in many environments, on different tasks, and for achieving different and several goals. Collaboration and cooperation in their more general sense (and, in particular, negotiation, exchange, help, delegation, adoption, and so on) are important characteristics - or better, the most foundational aspects - of human societies (Tuomela, 1995).

In the evolution of cooperative models, a fundamental role has been played by diverse constructs of various kinds (purely interactional, technical-legal, organizational, socio-cognitive, etc.), opportunely introduced (or spontaneously emerged) to support decision making in collaborative situations.

The new scenarios we are destined to meet in the third millennium transfigure the old frame of reference, in that we have to consider new channels and infrastructures (i.e., the Internet), new artificial entities for cooperating with artificial or software agents, and new modalities of interaction (suggested/imposed by both the new channels and the new entities). In fact, it is changing the identification of the potential partners, the perception of the other agents, the space-temporal context in which interaction happen, the nature of the interaction traces, the kind and role of the authorities and guarantees, etc.

For coping with these scenarios, it will be necessary to update the traditional supporting decision-making constructs. This effort will be necessary especially to develop the new cyber-societies in such a way as not to miss some of the important cooperative characteristics that are so relevant in human societies.

BACKGROUND

Trust (Ganbetta, 1990; Luhmann, 1990; Dasgupta, 1990), in the general frame described above, might be considered as a socio-cognitive construct of main importance. In particular, trust building is always more recognized as a key factor for using and developing the new interactional paradigm.

Trust should not be made indistinct with security. The latter can be useful to protect – in the electronic domain - from the intrusiveness of an unknown agent, to guarantee an agent in the identification of its partner, to identify the sender of a message (for example, by verifying the origin of a received message; by verifying that a received message has not been modified in transit; by preventing that an agent who sent a message might be able to deny later that it sent the message [He, Sycara & Su, 2001]). With sophisticated cryptographic techniques, it is possible to give some solution to these security problems.

However, more complex is the issue of trust, that must give us tools for acting in a world that is in principle insecure (that cannot be considered 100% secure), where we have to make the decision to rely on someone in risky situations. (Consider the variety of cases in which it is necessary or useful to interact with agents whose identity, history or relationships are unknown, and/or it is only possible to make uncertain predictions on their future behaviors.)

Trust should not be made indistinct with reputation, too. In fact, communicated reputation (Conte & Paolucci, 2002) is one of the possible sources on which the trustier bases its decision to trust or not.

The more actual and important example of the usefulness of trust building is electronic commerce, but we must also consider other important domains of Multi Agent Systems and Agent Theory such as Agent Modeling, Human-Computer Interaction, Computer Supported Cooperative Work, Mixed Initiative and Adjustable Autonomy, Pervasive and Ubiquitous Computing. In fact, today many computer applications are open distributed systems (with many autonomous components that are spread throughout a network and interacting with each other). Given the impossibility to rule this kind of system by a centralized control regime (Marsh, 1994), it becomes essential to introduce local tools in order to choose the right partnership and at the same time reduce the uncertainty (deriving from the nature of an open distributed system) associated with that choice.

S

TRUST IN THE NEW TECHNOLOGICAL SCENARIOS

In fact, various different kinds of trust should be modeled, designed, and implemented:

- Trust in the environment and in the infrastructure (the socio-technical system)
- Trust in personal agents and in mediating agents
- Trust in potential partners
- Trust in sources
- Trust in warrantors and authorities.

Part of these different kinds of trust have a complementary relation with each other, that is, the final trust in a given system/process can be the result of various trust attributions to the different components. An exemplary case is one's trust in an agent that must achieve a task (and more specifically in its capabilities for realizing that task) as different from one's trust in the environment (hostile versus friendly) where that agent operates, or again as different from one's trust in a possible third party (arbitrator, mediator, normative systems, conventions, etc.) able to influence/constrain the trustee and representing a guaranty for the trustier (Castelfranchi & Falcone, 1998; Falcone & Castelfranchi, 2001).

Therefore, the "sufficient" trust value of one single component cannot be established before evaluating the value of the other components. In this regard, it is very interesting to characterize the relationships between trust and (partial) control (Castelfranchi & Falcone, 2000).

It is important to underline how trust is in general oriented towards not directly observable properties. It is, in fact, based on the ability to predict these properties and to rely or not to rely on them. Thus, it is quite complex to assess the real trustworthiness of an agent/system/process, not only because - as we have seen - there are many different components that contribute to this trustworthiness, but also because the latter is not directly observable (see [Bacharach & Gambetta, 2001] about signs of trust). The important thing is the perceived trustworthiness that is, in its turn, the result of different modalities of the trustier's reasoning about direct experience; categorization; inference, and communicated reputation.

SOCIO-COGNITIVE MODEL OF TRUST

The Socio-Cognitive model of trust is based on a portrait of the mental state of trust in cognitive terms (beliefs, goals). This is not a complete account of the psychological dimensions of trust. It represents the most explicit (reason-based) and conscious form. The model does not account for the more implicit forms of trust (for example, trust by default, not based upon explicit evaluations, beliefs, derived from previous experience or other sources) or for the affective dimensions of trust, based not on explicit evaluations but on emotional responses and an intuitive, unconscious appraisal (Thagard, 1998).

The word *trust* means different things, but they are systematically related with each other. In particular, three crucial concepts have been recognized and distinguished not only in natural language but also in the scientific literature. Trust is at the same time:

- A mere *mental attitude* (prediction and evaluation) toward another agent, a simple *disposition*;
- A *decision* to rely upon the other, i.e., an *intention* to delegate and to trust, which makes the trustier "vulnerable"(Mayer, Davis, & Schoorman, 1995);
- A *behavior*, i.e., the intentional *act* of trusting, and the consequent *relation* between the trustier and the trustee.

In each of the above concepts, different sets of cognitive ingredients are involved in the trustier's mind. The model is based on the BDI (Belief-Desire-Intention) approach for modeling mind that is inspired by Bratman's philosophical model (Bratman, 1987). First of all, in the trust model only an agent endowed with both goals and beliefs can "trust" another agent. Let us consider the trust of an agent X towards another agent Y about the (Y's) behavior/action α relevant for the result (goal) g when:

- *X is the (relying) agent, who feels trust; it is a cognitive agent endowed with* internal explicit goals and beliefs (the *trustier*)
- Y is the agent or entity that is trusted (the *trustee*)
- X trusts Y about g/α and for g/α.

In the model Y is not necessarily a cognitive agent (for instance, an agent can - or cannot - trust a chair as far as to sustain his weight when he is seated on it). On the contrary, X must always be a cognitive agent: so, in the case of artificial agents we should be able to simulate these internal explicit goals and beliefs.

For all the three notions of trust defined above (*trust disposition*, *decision to trust*, and *trusting behavior*), we claim that someone trusts some other one only relatively to some goal (here the goal is intended as the general, basic teleonomic notion, any motivational representation in the agent: desires, motives, will, needs, objectives, duties, utopias, are kinds of goals). An unconcerned agent does not really "trust": he just has opinions and forecasts. Second, trust itself *consists of* beliefs.

Since *Y*'s action is useful to *X* (trust disposition), and *X* has decided to rely on it (decision to trust), this means that *X* might delegate (act of trusting) some action/goal in his own plan to *Y*. This is the strict relation between trust disposition, decision to trust, and delegation.

The model includes two main basic beliefs (we are considering the trustee as a cognitive agent, too):

- *Competence Belief*: a *sufficient evaluation* of *Y*'s abilities is necessary. *X* should believe that *Y* is useful for this goal, that *Y* can produce/provide the expected result, and that *Y* can play such a role in *X*'s plan/action.
- *Willingness Belief*: *X* should think that *Y* not only is able and can do that action/task, but *Y* actually will do what *X* needs (under given circumstances). This belief makes the trustee's behavior predictable.

Another important basic belief for trust is:

- *Dependence Belief*: *X* believes -to trust *Y* and delegate to it- that either *X* needs it, *X* depends on it (*strong dependence*), or at least that it is better to *X* to rely on it, rather than not to rely on it (*weak dependence*). In other terms, when *X* trusts someone, *X* is in *a strategic situation*: *X* believes that there is interference and that his rewards, the results of his project, depend on the actions of another agent *Y*.

Obviously, the willingness belief hides a set of other beliefs on the trustee's reasons and motives for helping. In particular, *X* believes that *Y* has some motives for helping it (for adopting its goal), and that these motives will probably prevail -in case of conflict- on other motives, negative for it. Notice that motives inducing adoption are of several different kinds: from friendship to altruism, from morality to fear of sanctions, from exchange to common goal (cooperation), and so on. This is why, for example, it is important to have common culture, shared values, and the same acknowledged authorities between trustier and trustee.

Another important characteristic of the socio-cognitive model of trust is the distinction between trust "in" someone or something that has to act and produce a given performance thanks to its *internal characteristics*, and the global trust in the global event or process and its result, which is also affected by *external factors* like opportunities and interferences.

Trust in *Y* (for example, "social trust" in strict sense) seems to consists in the first two prototypical beliefs/evaluations identified as the basis for reliance: *ability/competence* (that with cognitive agents includes knowledge and self-confidence*)*, and *disposition* (that with cognitive agents is based willingness, persistence, engagement, etc.). Evaluation about external opportunities is not really an evaluation about *Y* (at most the belief about its ability to recognize, exploit and create opportunities is part of our trust in *Y*). We should also add an evaluation about the probability and consistence of obstacles, adversities, and interferences.

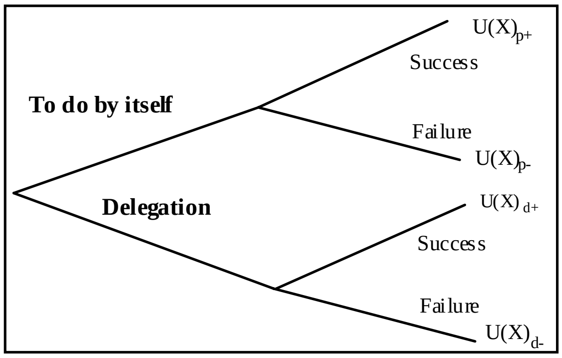
Trust can be said to consist of or better to (either implicitly or explicitly) imply the *subjective probability* of the successful performance of a given behavior a, and it is on the basis of this subjective perception/evaluation of risk and opportunity that the agent decides to rely or not to rely on *Y*. However, the probability index is based on, and derives from those beliefs and evaluations. In other terms, the global, final probability of the realization of the goal *g*, i.e., of the successful performance of a, should be decomposed into the probability of *Y* performing the action well (*internal attribution*) and the probability of having the appropriate conditions (*external attribution*) for the performance and for its success, and of not having interferences and adversities (*external attribution*). This decomposition is important because:

a) The trustier's decision might be different with the same global probability or risk, depending on its composition (for example, for personality factors);
b) Trust composition (internal vs. external) produces completely different intervention strategies: to manipulate the external variables (circumstances, infrastructures) is completely different from manipulating internal parameters.

The idea that trust is gradable is usual (in common sense, in social sciences, in Artificial Intelligence). However, since no real definition and cognitive characterization of trust is given, the quantification of trust is quite *ad hoc* and arbitrary, and the introduction of this notion or predicate is semantically empty. On the contrary, in the socio-cognitive model of trust there is a strong coherence between the cognitive definition of trust, its mental ingredients, and, on the one side, its value, on the other side, its social functions. More precisely the latter are based on the former.

A degree of trust of *X* in *Y* is grounded on the cognitive components of *X*'s mental state of trust. More precisely *the degree of trust is a function of the subjective certainty of the pertinent beliefs*. The degree of trust is used to formalize a rational basis for the decision of relying and betting on *Y*. A "quantitative" aspect of another basic ingredient is relevant: the value or importance or utility of the goal (g). In sum, *the quantitative dimensions of trust are based on the quantitative dimensions of its cognitive constituents*.

Figure 1



If we call $DoT_{XY\tau}$ the degree of trust of an agent *X* about *Y* on the task $\tau=(\alpha,g)$ we have:

- $DoT_{XY\tau} = DoC_X[Opp_Y(\alpha,g)] * DoC_X[Ability_Y(\alpha)] * DoC_X[WillDo_Y(\alpha,g)]$

Where:

- $DoC_X[Opp_Y(\alpha,g)]$, is the degree of credibility of *X's* beliefs about *Y's* opportunity of performing α to realize *g;*
- $DoC_X[Ability_Y(\alpha)]$, the degree of credibility of *X's* beliefs about *Y's* ability/competence to perform α*;*
- $DoC_X[WillDo_Y(\alpha,g)]$, the degree of credibility of *X's* beliefs about *Y's* actual performance*;*
- $DoC_X[WillDo_Y(\alpha,g)] = DoC_X[Intend_Y(\alpha,g)] * DoC_X[Persist_Y(\alpha,g)]$
(Given that *Y* is a cognitive agent)

In any circumstance, an agent *X* endowed with a given goal, has three main choices:

i) To try to achieve the goal by itself;
ii) To delegate the achievement of that goal to another agent, *Y*;
iii) To do nothing (relatively to this goal), renouncing.

Considering the simplified scenario in which only (i) and (ii) are the possible choices we have the Figure1.

Where if *U(X)* is the agent *X*'s utility function, more specifically:

- ***U(X)p+***, the utility of the *X*'s success performance;
- ***U(X)p-***, the utility of the *X*'s failure performance;
- ***U(X)d+***, the utility of a successful delegation (the utility due to the success of the delegated action);
- ***U(X)d-***, the utility of a failure delegation (the damage due to the failure of the delegated action).

In the previous scenario, in order to delegate we must have:

$$DoT_{XY\tau} * U(X)_{d^+} + (1 - DoT_{XY\tau})\, U(X)_{d^-} > DoT_{XX\tau} * U(X)_{p^+} + (1 - DoT_{XX\tau})\, U(X)_{p^-}$$

where $DoT_{XX\tau}$ is the *selftrust* of *X* about τ.

More precisely, we have: $U(X)_{p^+}$ *= Value(g) + Cost [Performance(X)],*

U(X)p- = Cost [Performance(X)] + Additional Damage for failure

$U(X)_{d^+}$ *= Value(g) + Cost [Delegation(X Y)],*

$U(X)_{d^-}$ *= Cost [Delegation(X Y)]+ Additional Damage for failure*

Where it is supposed that it is possible to attribute a quantitative value (importance) to goals and where the costs of the actions (delegation and performance) are supposed to be negative.

FUTURE TRENDS

One of the main aspects that should be analyzed in the next few years is the dynamics of trust and the possibility of introducing all the dynamic aspects in the computational setting in which humans and machines will work together. Trust is a dynamic phenomenon in its intrinsic nature (Falcone & Castelfranchi, 2001). Trust changes with experience, with the modification of the different sources it is based on, with the emotional state of the trustier, with the modification of the environment in which the trustee is supposed to perform, and so on. But, trust is also influenced by trust, itself, in the same specific interaction: for example, how trust creates a reciprocal trust; how the fact that A trusts B can actually increase B's trustworthiness; and so on. In other words, in a computational model of trust relationships we have to consider all the dynamical aspects of the trust phenomenon.

CONCLUSION

The Socio-Cognitive model of trust analyzes the basic elements on which trust is founded in terms of the cognitive ingredients of the trustier. In fact, the richness of the referred model (trust is based on many different beliefs) allows us to distinguish between internal and external attributions (to the trustee) and for each of these two attributions it allows us to distinguish among several other sub-components such as: competence, disposition, un-harmfulness, and so on.

The model introduced a degree of trust instead of a simple probability factor since it permits us to evaluate trustfulness in a rational way.

In other words, if we understand what precisely the basic ingredients of trust are, we would be able to better model and build artificial systems in which this attitude should be present.

REFERENCES

Bacharach, M., & Gambetta, D. (2001). Trust as type detection. In C. Castelfranchi & Y. Tan (Eds.), *Trust and deception in virtual societies*. Dordrecht: Kluwer Academic Publishers.

Bratman, M.E. (1987). *Intentions, plans, and practical reason*. Cambridge, MA: Harvard University Press.

Castelfranchi, C., & Falcone, R. (1998). Principles of trust for MAS: cognitive anatomy, social importance, and quantification. Proceedings of the International Conference on Multi-Agent Systems (ICMAS'98).

Castelfranchi, C. & Falcone, R. (2000). Trust and Control: A Dialectic Link. *Applied Artificial Intelligence Journal*, 14(8), 799-823.

Conte, R. & Paolucci, M. (2002). *Reputation in artificial societies. Social beliefs for social order*. Boston, MA: Kluwer Academic Publishers.

Dasgupta, P. (1990). Trust as a commodity. In D. Gambetta (Ed.), *Trust* (pp. 49-72). Oxford: Basil Blackwell.

Falcone, R., & Castelfranchi, C. (2001). The socio-cognitive dynamics of trust: does trust create trust? In R. Falcone, M. Singh, and Y. Tan (Eds.), *Trust in Cyber-societies: Integrating the human and artificial perspectives* (pp. 55-72). LNAI 2246 Springer.

Falcone, R., & Castelfranchi, C. (2001). Social Trust: A Cognitive Approach. In C. Castelfranchi & Y. Tan (Eds.), *Trust and deception in virtual societies*. Dordrecht: Kluwer Academic Publishers.

Gambetta, D. (Ed.) (1990). *Trust*. Oxford: Basil Blackwell.

He, Q., Sycara, K., & Su, Z. (2001). Security infrastructure for software agent society. In C. Castelfranchi & Y. Tan (Eds.), *Trust and deception in virtual societies*. Dordrecht: Kluwer Academic Publishers.

Luhmann, N. (1990). Familiarity, confidence, trust: Problems and alternatives. In D. Gambetta (Ed.), *Trust* (pp. 94-107). Oxford: Basil Blackwell.

Marsh, S. (1994). Formalising trust as a computational concept. Ph.D. thesis, Department of Computing Science, University of Stirling.

Mayer, R.C., Davis, J.H., & Schoorman, F.D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709-734.

Thagard, P. (1998). Emotional Coherence: Trust, Empathy, Nationalism, Weakness of Will, Beauty, Humor, and Cognitive Therapy, Technical Report, University of Waterloo.

Tuomela, R. (1995). The importance of us: a philosophical study of basic social notions. Stanford University Press.

KEY TERMS

Cyber-societies: The set of natural, artificial, and virtual agents connected and interacting with each others through natural and artificial infrastructures within virtual institutions.

Reputation: the estimated trustworthiness in an agent as derived from the communicated opinions of other parts (directly or indirectly received); the resulting and emergent "common opinion" about the agent's trustworthiness.

Task: An action and/or a goal an agent has to realize as delegated by another agent; thus – in the opposite perspective - the couple action/goal that an agent intentionally delegates to another agent; where at least the delegating agent knows one between the action and the goal.

Trust: The attitude of an agent to delegate a part of its own plan/goal to another agent and rely upon it in a risky situation (possible failure) on the basis of its own beliefs about the other agent and on the environment in which it operates.

Trustee: the trusted agent in the trust relationship.

Trustier: the trusting agent in the trust relationship.

Ubiquitous Computing: The trend of the technological development to integrate into any kind of object information processing and communication capabilities.

Softening the MIS Curriculum for a Technology-Based Profession

Glenn Lowry
United Arab Emirates University, UAE

Rodney Turner
Victoria University of Technology, Australia

INTRODUCTION

Information systems[1] (IS) professionals help to achieve business and organizational goals through the use of information technology. The information systems profession is team oriented and project based.

Students are first and foremost concerned with future employability. Employers, on the other hand, often indicate that they want new graduates who can be immediately productive in their environment.

Are the aspirations of students and employers fundamentally incompatible? How can IS educators help to find a workable and satisfying balance? How can information systems educators soften and refine the design of information systems curricula to achieve a better fit between the workplace and the university 'studyplace'?

BACKGROUND: BALANCING THE MIX OF HARD AND SOFT SKILLS IN THE IS/IT CURRICULUM

The preparation of information systems (IS) professionals must encompass a body of knowledge and a repertoire of technical and 'soft' skills identified by various professional bodies (Cheney, Hale & Kasper, 1990; Davis, Gorgone, Feinstein & Longenecker, 1997; Gorgone & Gray, 1999; Underwood, 1997; Lidtke, Stokes, Haines & Mulder, 1999; Cohen, 2000; Mulder & van Weert , 2000; ACM-AIS, 2002).

The persistent research finding that employers want graduates who possess better business skills is often interpreted by academics to mean that more traditional, formal business subjects such as accounting, economics, business finance, and marketing should be taught alongside traditional technical or 'hard' skill subjects such as systems analysis/design and programming in particular languages (Trauth, Farwell & Lee, 1993; Van Slyke, Kittner & Cheney, 1997). On the other hand the 'soft' areas such as teamwork, communication skills, ability to accept direction, and others are somehow 'picked up' along the way through an unspecified, osmotic process and not addressed as part of a curriculum.

The past few decades have been characterized by a rapidly and constantly changing business environment. Lee, Trauth, and Farwell (1995) argued that technological and sociological developments facilitated by evolving information technology and changing business needs has made it necessary for IS professionals to develop a wider range of non-technical skills than was previously the case. Similar views have been expressed by many others, including Burn, Ng, and Ma (1995), Cafasso (1996), Lowry, Morgan, and FitzGerald, (1996), and Morgan, Lowry, and FitzGerald (1998).

The work presented here is part of an ongoing research program that investigates the views of major IS curriculum stakeholders, including employers, IS practitioners, currently enrolled students, and academics. The data were gathered from a survey of IS practitioners and IS decision makers in Australia, and covered all industry sectors as well as business unit sizes. We argue that IS practitioners, employers, and students see little value in some of the more formal business subject areas that often form the core of an IS degree offered in a business or commerce faculty. These stakeholder groups see more value in the development of 'soft skills' useful in client interaction. The findings have serious implications for IS educators and IS curriculum design (Turner & Lowry, 1999a, 1999b, 2000, 2001, 2002, 2003).

IS CURRICULUM CONTENT AND DELIVERY IN THE FIRST DECADE OF THE 21ST CENTURY

In a 1999 study, the authors began to suspect that the 'other business skills' desired of new IS graduates were not synonymous with traditional business curriculum subjects. Study results indicated that of nine business subjects that are typically included in IS curricula, only three—*Accounting*, *Business Ethics*, and *Management*—were judged to be important by students and employers.

Follow-up studies were conducted in 2001 and 2002 to further explore the 'other business skills' aspect of the IS curriculum (Turner & Lowry, 1999b, 2001, 2002, 2003). Tables 1 and 2 show comparative ratings of 'hard' or traditional IS and business subjects, and of 'soft' skills by information systems and technology professionals and employers.

COMPARATIVE RATINGS OF ACADEMIC SUBJECTS BY IS PRACTITIONERS AND EMPLOYERS

The research instruments contained two sections pertaining to academic preparation of graduates. These two sections separately covered the technical areas of an IS business degree and the other academic areas that are not specific to IS. Subjects responded to each question using a seven-point Likert scale (1= irrelevant, 7 = essential). Table 1 shows the mean and standard deviation of ratings of academic subjects and soft skills by *Information Systems/Information Technology (IS/IT) Professionals* and *IS/IT Employers.*

Of the 14 subjects/skills in Table 1 that achieved a mean rating of 5.0 or more, the highest rating by both practitioners and employers was achieved by *Communications & Report Writing,* a soft skill. *Analysis and Design,* a core IS academic subject, was rated highly by both IS/IT professionals and employers. The value placed on these two subjects is consistent with earlier (1999) findings by the authors.

Eleven technical subjects and two 'other business subjects'—management and business ethics—achieved mean ratings of 5.0 or more. These higher ratings suggest

Table 1. Comparative ratings of academic subjects by IS practitioners and employers

	IS/IT Professionals		**IS/IT Employers**	
Skills	**Mean**	σ	**Mean**	σ
Communications & Report Writing	6.02	1.05	6.09	0.81
Analysis & Design	5.87	1.09	5.63	1.26
Client Server Applications	5.67	0.92	5.37	1.15
Business Applications	5.65	1.11	5.76	1.20
Use Operating Systems	5.60	1.10	5.39	1.27
Database Design	5.55	1.25	5.12	1.16
Management	5.54	1.03	5.20	1.10
Knowledge of PC Apps	5.43	1.22	5.41	1.37
Project Management	5.43	1.16	5.60	1.24
E-Commerce/E-Business Development	5.33	1.23	4.78	1.38
Apply OOPs	5.26	1.25	4.61	1.51
LAN & Data Communications	5.22	1.27	5.55	1.22
Large System Experience	5.12	1.19	4.54	1.53
Business Ethics	5.07	1.57	5.23	1.52
Web Design/Development	4.96	1.54	4.67	1.15
Organizational Behavior	4.90	1.41	4.92	1.34
Data Mining/Data warehousing	4.76	1.36	4.66	1.36
Apply 3GLs	4.70	1.41	4.15	1.58
CASE Applications	4.51	1.32	3.80	1.42
Knowledge Base/Expert Systems	4.49	1.42	4.20	1.37
ERP Implementations & Operations	4.48	1.39	4.33	1.61
Marketing	4.35	1.52	4.39	1.34
Business Finance	4.30	1.50	4.54	1.47
Operations Research	4.29	1.26	4.32	1.26
Mathematical Modeling	4.25	1.44	3.97	1.49
International Business	4.24	1.59	3.69	1.56
Business Statistics	4.18	1.40	4.33	1.38
Accounting	4.13	1.55	4.68	1.38
Business or Commercial Law	4.07	1.55	4.12	1.45
Psychology	3.70	1.76	3.85	1.46
Economics	3.63	1.50	3.68	1.47
Foreign Languages	3.15	1.78	3.04	1.46
***n*=**	**136**		**138**	

Table 2. Comparative importance of soft skills by IS/IT professionals and employers

	IS/IT Professionals		IS/IT Employers	
Skills	**Mean**	**σ**	**Mean**	**σ**
Work as a team	6.52	0.66	6.39	0.81
Problem-solving skills	6.44	0.57	6.37	0.68
Work under pressure	6.42	0.78	6.27	0.75
Quickly acquire new skills	6.37	0.64	6.15	0.73
Independently acquire new skills	6.35	0.72	6.23	0.71
Meet deadlines	6.35	0.68	6.13	0.81
Work independently	6.27	0.94	6.22	0.65
Time management	6.21	0.95	5.98	0.84
Problem definition skills	6.18	0.74	6.14	0.74
Willing to undergo ongoing professional development	6.18	0.89	5.93	0.89
Written communication skills	6.18	0.85	6.04	0.76
Client-focused service ethic	6.16	1.00	6.09	0.94
Handle concurrent tasks	6.16	0.81	6.08	0.81
Interact with people of different backgrounds	6.13	0.73	6.03	0.85
Think creatively	6.08	0.89	6.09	0.71
Work with people from different disciplines	6.04	0.74	6.10	0.82
Accept direction	6.03	0.89	5.98	0.84
Information-seeking skills	5.83	0.96	5.82	0.93
Oral presentation skills	5.79	1.07	5.56	0.88
Place organizational objectives first	5.73	0.95	5.74	0.97
Business analysis skills	5.63	1.03	5.51	1.04
Leadership potential	5.18	1.08	4.99	0.94
Good sense of humor	5.15	1.35	5.58	1.14
Able to prepare multimedia presentations	4.73	1.25	4.32	1.38
n=	**136**		**138**	

that these subjects should be included in IS curricula, ahead of other subjects that received lower ratings by practitioners and employers.

COMPARATIVE IMPORTANCE OF SOFT SKILLS BY IS/IT PROFESSIONALS AND EMPLOYERS

A third section in the survey solicited rankings of the importance of a range of so-called ‘soft skills’. The results are presented in Table 2, which shows the ratings by IS/IT professionals and employees for ‘soft’ business skills.

Table 2 shows a marked similarity between practitioners and employers in their ratings of the importance of ‘soft’ business skills. Only the ability to prepare multimedia presentations failed to achieve a mean rating of 5.0. All other soft business skills were highly rated by both IS practitioners and employers.

Unlike the academic subjects in Table 1, many of the soft skills can be identified, grouped together, and developed through formal ‘professional development’ units in IS curricula or specifically taught and assessed as part of formal coursework.

FUTURE TRENDS AND SOME CRITICAL ISSUES IN IS CURRICULUM DESIGN AND DELIVERY

Barriers to meaningful reform and evolution of the IS curriculum include confusion, tradition and inertia, scarcity of resources, and vested interests. Table 3 summarizes these issues and offers recommendations, in increasing order of difficulty.

Table 3. Issues and recommendations for IS curriculum refinement

Confusion: Research results that call for more 'business skills' have handily and traditionally been interpreted as meaning exposure of students to additional, formal business subjects. While an IS student may well gain knowledge and skill in marketing, economic analysis, or international business in that way, our findings suggest that it is the 'soft' skills, rather than formal academic skill, that is wanted by IS/IT practitioners and employers.	**To reduce confusion:** Local course advisory and professional bodies can provide invaluable insight into the mix of technical and non-technical formal courses and 'soft' skills appropriate for a given institution's service area. Focus groups and local replication of available studies should provide targeted, timely, and authoritative guidance for ongoing curriculum evolution**.**
Tradition and Inertia—Content: It is easy to offer traditional business subjects as they are already being taught anyway. In many institutions, the existing IS academic staff would have to acquire the additional academic background and skills needed to introduce a substantive 'soft skills' emphasis into the IS curriculum. In most instances, there may be insufficient time or interest to do so. **Tradition and inertia—Delivery Method:** Many, if not most, subjects are taught in a familiar lecture-practical mode. As a project- and team-oriented profession, information systems programs may well achieve a better match between the workplace *through* the 'study place' through the adoption of active-learning, student-centered delivery methods such as Problem-Based Learning (PBL) and Work-Integrated Learning (WIL) (Bentley, Sandy & Lowry, 2002).	**Overcoming Tradition and Inertia:** To some extent, prospective students have taken matters into their own hands by opting in larger numbers to bypass traditional university courses in favor of industry-sponsored/sanctioned entry gateways such as those offered by Microsoft, Oracle, SAP, and others. It is possible that students electing the non-academic alternative see more value in industry-focused training than in academic education that they see as irrelevant to their career aspirations. A large number of traditional business subjects were rated low in importance by IS practitioners and employers in Tables 1 and 2. Tertiary educators will need to reconsider the value of these traditional business subjects, and will have to develop and deliver revitalized curricula that squarely address the expressed desire for 'soft' skills that have been identified in a number of studies.
Resources: If a substantial portion of an undergraduate degree program were shifted from traditional business subjects to the acquisition and development of 'soft skills', who would develop, teach, and assess the new 'soft skills' curriculum component? We must learn to master what we teach about building client ownership to enlist influential industry partners.	**Finding Resources:** Resources are easier to obtain from a body of satisfied clients, such as the firms who employ our graduates. If we are seen to consult with, listen to, and serve the interests of those firms, they will follow their self-interest and become rich sources of guidance in curriculum planning and development, work experience for students, consultancies for academics, equipment, money, and political weight in our own institution.
Vested Interest: Some academic institutions supplement the enrollment in less relevant or popular subjects through inclusion of those subjects in a popular curriculum such as information systems. In many institutions, economic incentives exist for students to be enrolled in subjects within a single administrative unit, such as a business or IT faculty**.**	**Neutralizing Vested Interests:** Senior university managers may oppose substantive IS curriculum reform such as that discussed in this chapter for a number of reasons, including loss of revenue if students enroll in 'soft' skill courses provided by another administrative unit. It is up to IS educators to develop strategies to address the turf issues that preoccupy some administrators. Building effective partnerships with the appropriate industries in our service area can provide a powerful voice to speak on our behalf to senior management. Accrediting bodies can also help in this way.

CONCLUSION

There has long been agreement that the IS curriculum should be comprised of some combination of technical subjects and non-technical business subjects, and that graduates also need 'soft' business skills. There is far less agreement about what the mix between these should be and how best to prepare students in some areas, notably in the development of 'soft' business skills.

While we agree with the general view that 'soft' skills have become increasingly important, we argue that the traditional business subjects are *not* the business skills primarily sought in studies of the IS marketplace. Does the study of traditional business subjects such as marketing, business law, or economics directly help the students to develop a repertoire of 'soft' business skills? The findings suggest that in reality it is not more core business subjects that are needed, but an appreciation of business processes and activities that are not always covered in IS degree programs.

Some formidable barriers exist to substantive revision of IS curricula to emphasize acquisition and development of 'soft' business skills. In increasing order of difficulty, Table 3 summarizes some of the issues and barriers to meaningful reform, and evolution of the IS curriculum might be surmounted over time and with sufficient dedication. There are, of course, no easy solutions to resolve these issues.

While study after study has called for 'soft' skills acquisition and development by IS students, some IS programs have a clearer and better-developed vision than others of what those skills are, and how they may be introduced and cultivated. The growing emphasis on 'soft' skills in IS education is an indication that what began as a fundamentally technology-oriented discipline is, indeed, evolving into a technology-based profession. We can watch someone else claim that knowledge and the opportunities that it offers, or we can embrace it ourselves, while we have the initiative.

REFERENCES

ACM-AIS. (2002). *IS 2002—model curricula and guidelines for undergraduate degree programs in information systems*. Association for Computing Machinery Association for Information Systems. Retrieved November 28, 2003, from *www.acm.org/ education /curricula.html #IS2002*.

Bentley, J., Sandy, G. & Lowry, G. (2002). Problem-based learning in information systems analysis and design. In E. Cohen (Ed.), *Challenges of information technology education in the 21st century* (pp. 100-123). Hershey, PA: Idea Group Publishing.

Burn, J.M., Ng Tye, E.M.W. & Ma, L.C.K. (1995). Paradigm shift—cultural implications for development of IS professionals. *Journal of Global Information Management, 3*(2), 18-28.

Cafasso, R. (1996). Selling your soft side helps IT. *Computerworld, 18*(35), 60-61.

Cheney, P., Hale, D. & Kasper, G. (1990). Knowledge, skills and abilities of information systems professionals: Past, present and future. *Information Management, 9*(4), 237-247.

Cohen, E. (2000). *Curriculum model 2000 of the Information Management Association and the Data Administration Mangers Association*. IRMA-DAMA. Retrieved November 28, 2003, from *www.irma-international.org/ downloads/pdf/irma_dama.pdf*

Davis, G., Gorgone, J.T., Feinstein, D.L. & Longenecker, H.E. (1997). *IS'97 model curricula and guidelines for undergraduate degree programs in information systems.* Association of Information Technology Professionals.

Gorgone, J.T. & Gray, P. (1999, January 5-8). Graduate IS curriculum for the 21st century. Proceedings of the 32nd Hawaii International Conference on Systems Science, Maui, Hawaii. Retrieved November 28, 2003, from *computer.org/proceedings/hicss/0001/00011/ 0001toc.htm* and *www.cs.adelaide.edu.au/%7Echarles/ ACS/ACS0.pdf*

Lee, D.M., Trauth, E.M. & Farwell, D. (1995). Critical skills and knowledge requirements of IS professionals: A joint academic/industry investigation. *MIS Quarterly, 19*(3), 313-340.

Lidtke, D., Stokes, G., Haines, J. & Mulder, M. (1999). *ISCC '99: An information systems-centric curriculum'99: Guidelines for educating the next generation of information systems specialists*. Retrieved November 28, 2003, from *www.iscc.unomaha.edu/*

Lowry, G.R., Morgan, G.W. & FitzGerald, D.G. (1996). Organisational characteristics, cultural qualities and excellence in leading Australian-owned information technology firms. *Proceedings of the 1996 Information Systems Conference of New Zealand* (pp. 72-84), Palmerston North, New Zealand. Los Alamitos, CA: IEEE Computer Society Press.

Morgan, G.W., Lowry, G.R. & FitzGerald, D.G. (1998). development staff characteristics and service stability in leading Australian-owned information technology firms. *Proceedings of the 1998 International Conference on Software Engineering: Education and Practice* (pp. 96-103), Dunedin, New Zealand. Los Alamitos, CA: IEEE Computer Society Press.

Mulder, M. & van Weert, T. (2000). *Informatics curriculum framework 2000 for higher education* (ICF-2000). Paris: UNESCO.

Trauth, E., Farwell, D. & Lee, D. (1993). The IS expectation gap: Industry expectations versus academic preparation. *MIS Quarterly*, *17*(3), 293-307.

Turner, R. & Lowry, G. (1999a). The Compleat graduate: what students think employers want and what employers say they want in new graduates. In S. Lee (Ed.), *Preparing* *for the Global Economy of the New Millennium. Proceedings of Pan-Pacific Conference XVI* (pp. 272-274). Fiji: Pan-Pacific Business Association.

Turner, R. & Lowry, G. (1999b). Reconciling the needs of new information systems graduates and their employers in small, developed countries. *South African Computer Journal, 24*(November), 136-145.

Turner, R. & Lowry, G. (2000). Motivating and recruiting intending IS professionals: A study of what attracts IS students to prospective employment. *South African Computer Journal, 26*(4), 132-137.

Turner, R. & Lowry, G. (2001). What attracts IS students to prospective employment: A study of students from three universities. In *Managing information technology in a global economy* (pp. 448-452). Hershey, PA: Information Resources Management Association.

Turner, R. & Lowry, G. (2002). The relative importance of 'hard' & 'soft' skills for IT practitioners. In M. Khosrow-Pour (Ed.), *Issues and trends of information technology management in contemporary organizations* (pp. 1-10). Hershey, PA: Information Resources Management Association.

Turner, R. & Lowry, G. (2003). Education for a technology-based profession: Softening the information systems curriculum. In T. McGill (Ed.), *Issues in information systems education* (pp. 156-175). Hershey, PA: Idea Group Publishing.

Underwood, A. (1997). *The ACS core body of knowledge for information technology professionals*. Retrieved from *www.acs.org.au/national/pospaper/bokpt1.htm*

KEY TERMS

Hard Skills: Measurable capabilities and academic knowledge acquired through traditional tertiary study. Current MIS curriculum examples include communications and report writing, systems analysis and design, client/server applications, and business applications.

Inquiry-Based Learning (IBL): A student-centered, active learning approach focusing on questioning, critical thinking, and problem solving. IBL is expressed by the idea "involve me and I understand." The IBL approach is more focused on using and learning content as a means to develop information-processing and problem-solving skills. The system is more student centered, with the teacher as a facilitator of learning. There is more emphasis on "how we come to know" and less on "what we know." Students are involved in the construction of knowledge through active involvement. The more interested and engaged students are by a subject or project, the easier it will be for them to construct in-depth knowledge of it. Learning becomes easier when it reflects their interests and goals, and piques their natural curiosity.

Problem-Based Learning: An active learning strategy that may be suitable for better preparing information systems students for professional practice. In the problem-based approach, complex, real-world problems or cases are used to motivate students to identify and research concepts and principles they need to know in order to progress through the problems. Students work in small learning teams, bringing together collective skill at acquiring, communicating, and integrating information in a process that resembles that of inquiry.

Project-Based Learning: An active learning approach that focuses on developing a product or creation. The project may or may not be student centered, problem based, or inquiry based. Project-based learning uses open-ended assignments that provide students with a degree of choice, and extend over a considerable period of time. *Teachers* act as facilitators, designing activities and providing resources and advice to students. Instruction and facilitation are guided by a broad range of teaching goals. *Students* collect and analyze information, make discoveries, and report their results. *Projects* are often interdisciplinary.

Soft Skills: Cultivated elements of professionalism that derive from example, reflection, imitation, and refinement of attitudes, personal capabilities, work habits, and interpersonal skills, and are expressed in consistent and superior performance, characterized by a customer service and team orientation. Current MIS curriculum examples include ability to work as a member of a team, well-developed oral and written presentation skills, and the ability to work independently.

Student-Centered/Active Learning: Places the student into active, self-directed learning, learning by enquiry, and ownership of the learning goals. Active-learning strategies include Problem-Based Learning (PBL), Project-Based Learning, Inquiry-Based Learning, and Work-Integrated Learning.

Teacher-Centered Learning: Characterized by didactic teaching, passive learning, and the teacher as the 'expert'. Control of learning rests with the instructor. The learning of the students is directed by the instructor and is often based on what the instructor believes the student needs to learn. This approach can lead to students focusing on determining what knowledge they require to pass the subject, rather than the instructional objectives of the program, a phenomenon well-known to educators.

Work-Integrated Learning: A hybrid approach that achieves learning outcomes through a combination of alternating periods of traditional academic pedagogy, with extended periods of practical experience in the professional workplace. Work-Integrated Learning is a mature pedagogical strategy that is often referred to as 'sandwich' and 'end-on' courses.

ENDNOTE

1 The term "information systems" will be used to mean "management information systems" and "informatics" throughout this article.

Software Agents in E-Commerce Systems

Juergen Seitz
University of Cooperative Education Heidenheim, Germany

INTRODUCTION

The Internet introduces a new global marketplace for a large number of relatively unknown and often small companies often offering substitutive or complimentary products and services. The merchants profit from reduced costs, reduced time, and unsold stocks. Customers are attracted by increasing convenience and fast fulfillment.

Merchants offering these products and services on this new marketplace need to acquire new customers and sustain ongoing relationships. Nowadays, most merchants' sites are passive catalogs of products and prices with mechanisms for orders (Dasgupta et al., 1998). The pull strategy is also applied in auctions available over the Internet, where the seller waits passively for bids. The new push technologies for electronic commerce, like software agents, enable customers to compare a bewildering array of products efficiently, effectively, and automatically (Jennings et al., 1998). Switching costs for customers and, thereby, their loyalty to previous suppliers in the marketplace decline.

Using the Internet, the producers profit from reduced cost through direct, non-intermediated sales. The key elements to successful long-term relationships between merchants and customers will be the offering of personalized and value-added services, like one-to-one marketing services, discounts, guarantees, and savings coupons (Seitz et al., 2002, p. 209).

In this article, we will analyze possible consequences of new push and pull technologies in electronic commerce for customers' loyalty. The active technologies enable customers to purchase efficiently and the merchants to offer highly personalized, value added, and complimentary products and services. We will discuss some examples of such services and personalization techniques sustaining one-to-one relationships with customers and other actors involved.

BACKGROUND

The World Wide Web provides a great opportunity to compare products and services. Customers as well as competitors may quickly gain detailed and up-to-date data. Especially, suppliers of digital goods are in fear of declining customers' loyalty. Customers compare catalogs of products of merchants and producers, and conduct transactions independently of their geographic localization. The crucial basic factors responsible for a limited loyalty of customers are convenience, time, and cost of fulfillment. So, an electronic commerce system should support the ability to embed intelligence to automate the decision process (Dasgupta et al., 1998). The system should not only compare products and prices, but also negotiate and finally purchase products (Teuteberg & Kurbel, 2002). Nowadays, most systems still involve a substantial human element that is from the consumer's perspective neither convenient nor efficient. The human involvement should be limited to transaction specification at the beginning of the process and to the buying or refusal decision at the end of the process (Chen, 2000). This means that an appropriate technology is necessary in the intermediate stages to coordinate between customers and suppliers (d'Inverno & Luck, 2003). Mobile software agents emerge as ideal mediators in electronic commerce and thereby as an appropriate technology for an automated procurement process. Customers may specify constraints on the features of products that enable mobile agents to select products from the merchant's catalogs and finally to determine the terms of the transaction. Otherwise, software agents may be used by suppliers as market surveyors to determine the current demand and an appropriate price for the good. Software agent technology also abolishes the problem of different technological standards, like hardware platforms and operating systems of remote computers. This means that geographical or technological barriers for customers are of no significant importance anymore. The key factors are convenience, time and cost of the procurement process.

AGENT MEDIATED ELECTRONIC COMMERCE

Software agents are computer programs showing the following characteristics (Joshi & Ramesh, 1998):

- Reactivity: Agent perceives and reacts to environmental changes.

- Autonomy: Agent has its own program code, data and execution state.
- Proactivity: Agent initiates changes to the environment.

The ability of an agent to travel around in networks enhances it to a mobile agent (Brenner, Zarnekow & Wittig, 1998). Mobile software agents may be classified based on their attributes, like mobility, type of cooperation and level of interactivity (see Figure1) (Joshi & Ramesh, 1998). For further possible classifications see, for example, Nwana (1996), Sycara et al. (1996) or Kurbel and Loutchko (2001).

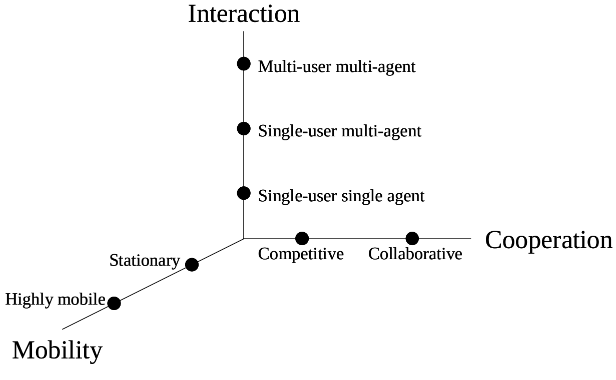
Competitive agents, mostly single-agents, maximize the interests of their owners. Collaborative agents, on the contrary, share their knowledge and try to maximize benefits of the community on the whole (Joshi & Ramesh, 1998). Mobile agents differ also in terms of the ease of the mobility between remote computers. A continuously traveling nomadic agent, like mobile sales agents (containing information of the total quantity of the product to be sold, the initial price of the product and the list of potential customers to visit), arrives at a customer's site and communicates with a stationary customer agent, which determines the quantity to be purchased at a given price (Dasgupta et al., 1998). The customer agent uses market values and demand curves of the product for its decision. The sales agent has to adjust the price dynamically during negotiations in order to maximize the gross returns. The price for the product may not be settled too low (an agent sells all of his stock at a bargain price) or too high (a given quantity of the goods may be unsold). Such a supplier-driven electronic commerce system enables merchants to maximize their gross return, but also to identify quickly the customers' needs and finally to cultivate long-term relationships with them. The architecture of the supplier-driven system was presented by Dasgupta et al. (1998).

From a customer's perspective, software agents should be highly personalized, continuously-running and autonomous mediators, that have to delegate some process management tasks (Guttman et al., 1998). A software agent should identify customers' needs at first, then retrieve information about the product from the merchants' sites, compare the offers and finally determine the terms of the transaction (Castro-Schez et al., 2004). Nowadays, customer agents are mostly used for product and merchant brokerage and for negotiation (Guttman et al., 1998).

The price of a product may also be dynamically negotiated instead of being fixed. For example, tête-à-tête agents cooperatively negotiate multiple terms of a transaction, like warranties, return policies, delivery times, and loan options (Guttman et al., 1998). The buyer agent in a tête-à-tête system negotiates towards a pareto-optimal deal with the sales agent (Fatima et al., 2004). A system like this does not maximize gross returns to suppliers or price discounts for customers (Excelente-Toledo & Jennings, 2004; Rahwan et al., 2004). However, it takes into consideration the important value-added merchant's services.

Summarizing, software agents are helping customers to compare and to purchase goods on the Internet. Most of them are agents for a simple online product price comparison or for competitive negotiation over price without considering the value-added and post-purchase services from merchants. Such agents decrease customers' loyalty to a merchant towards zero. However, additional services, like guarantees, return policies, loans, gifts, discounts and insurance are of interest to customers. Therefore, they should rather use agents for comparing or negotiation over multiple terms of a transaction (tête-à-tête). Otherwise, merchants may also send their own sales agents to potential buyers in order to acquire new customers and remind the previous customers of new sales offerings (Dutta et al., 2003).

Figure 1. Classification of software agents (Joshi & Ramesh, 1998).



FUTURE TRENDS

In general, software agents helping customers in the procurement process may minimize their loyalty to merchants. Suppliers who do not want to compete solely on the basis of price provide their customers with highly personalized and value-added services to sustain a long-term relationship.

Personalization and Privacy

Personalization is defined as the customization of a Web site to meet the particular needs of individual users (Chaffey et al., 2000; Dean, 1998). The goal of personalization technology is to encourage repeated visits and to enhance user loyalty. The identification of customers' needs occurs through the observation of their behavior and the collection of data (filling out forms or following decision-tree sets of questions).

There exist some advanced techniques supporting personalization of Web site contents, like rule-based matching and collaborative filtering. Using rule-base matching, users have to answer a set of yes/no or multiple-choice questions to settle a set of users' criteria. Collaborative filtering methods combine the personal preferences of a user with preferences of like-minded people (Dean, 1998). In regard to personalization techniques, one-to-one-marketing should be noticed. This strategy enables targeting unique offers to specific customers (Chaffey et al., 2000). Institutions offering such individualized services have to dispose of accurate user profiles before.

A critical factor of personalization is the privacy issue. Filtering and customization techniques entail the collection and the use of personal data, like name, e-mail address, postal address, age, gender, income, Internet connection, and employment status, that must be protected from abuse (Dean, 1998). Furthermore, a lot of suppliers on the Internet deriving revenues mainly from advertising need to identify their users in order to better customize the content and to attract the advertisers that might be interested. Hence, the user should be informed by suppliers how they use the personal data and how the protect them. Nowadays, there are several initiatives and standardization projects for the privacy of personal data usage. Such initiatives increase user trust and confidence in electronic commerce. However, no organization or institution has the power to enforce it to the wide usage of suppliers.

Value-Added Services

Merchants who do not want to compete solely on the basis of prices often offer their customers value-added or complimentary services. Complimentary products imply higher benefits to the customer in the case that he or she only buys the product he or she is looking for (Seitz et al., 1999). Such products or services increase the value of the primary good to the customer. Examples for value-added services in electronic commerce are sales discounts, savings coupons, additional insurance and guarantees, gifts, but also free software to test. In general, value-added services enable customers to trade at favorable terms and with confidence. They increase the attractiveness of the merchant to present customers and attract new customers.

Reduced Financial Transaction Costs

Internet merchants might achieve additional reduction of transaction costs using electronic payment systems. Electronic payment systems reduce cash handling costs for merchants and improve speed and convenience for customers. The aggregated cost of each payment consists of transformation costs, for example, the fees for conversion from assets to cash and vice versa, transport and storage costs, costs for safety measures, and search and time costs (Hakenberg, 1996).

CONCLUSION

This article discusses consequences of electronic commerce on customers' loyalty. Electronic commerce on the Internet offers the possibility to create a perfect marketplace. The intermediation in distribution will be reduced. This means lower costs for both suppliers and customers.

Software agents may be classified into different types with regard to their use on supporting electronic commerce. Most of the software agents only perform simple product price comparisons; some support the purchase of products. These software agents reduce customers' loyalty because the price is the only parameter. Quality and added values are not considered. Therefore, multi-agent systems allowing negotiation might be useful from a customer's perspective. Merchants may also send their own sales agents to potential buyers in order to remind previous customers of new sales offerings or to suppliers in order to maximize their gross return.

Personalization and customization of Web sites, value-added services and the reduction of transaction costs are instruments for increasing customers' loyalty. Personalization techniques, like rule-based matching and collaborative filtering, provide contents on Web sites that are appropriate to the customer's preferences or analyze past purchases and prior suggestions of other customers. One-to-one-marketing may be especially useful for sophisticated products demanding explanations or to enable cross-selling of other products and services. User profiles allow merchants to make customer-oriented offers or build special offers including additional services. Value-added services attract the customer to trade at favorable terms. The usage of electronic payment systems may reduce transactions costs.

REFERENCES

Brenner, W., Zarnekow, R., & Wittig, H. (1998). *Intelligente Softwareagenten. Grundlagen und Anwendungen*. Berlin: Springer.

Castro-Schez, J.J., Jennings, N.R., Luo, X., & Shadbolt, N. (2004). Acquiring domain knowledge for negotiating agents: A case study. *International Journal of Human Computer Studies* (to appear).

Chaffey, D., Mayer, R., & Johnston, K. (2000). *Internet marketing*. London: Prentice Hall.

Chen, Z. (2000). Intelligent agents. In M. Zeleny (Ed.), *The IEBM handbook of information technology in business* (pp. 561-569). London: Thomson Learning.

Dasgupta, P., Narasimhan, N., Moser, L.E., & Melliar-Smith, P.M. (1998, November). A supplier-driven electronic marketplace using mobile agents. *Proceedings of the First International Conference on Telecommunications and E-Commerce*, Nashville.

Dean, R. (1998, June 2). Personalizing your Web site. Retrieved August 29, 2003, from *http://builder.cnet.com/webbuilding/pages/Business/Personal/*

Dutta, P.S., Moreau, L., & Jennings, N.R. (2003). Finding interaction partners using cognition-based decision strategies. *Proceedings of the IJCAI workshop on Cognitive Modeling of Agents and Multi-Agent Interactions*, Acapulco, Mexico (pp. 46 – 55).

Excelente-Toledo, C.B., & Jennings, N.R. (2004). The dynamic selection of coordination mechanisms. *Journal of Autonomous Agents and Multi-Agent Systems, 9*(1-2), 55–85.

Fatima, S., Wooldridge, M., & Jennings, N.R. (2004). Bargaining with incomplete information. *Annals of Mathematics and Artificial Intelligence* (to appear).

Guttman, R.H., Moukas, A.G., & Maes, P. (1998). Agents as mediators in electronic commerce. *EM-Electronic Markets, 8*(1), 22–27.

Hakenberg, T. (1996). Elektronische Zahlungssysteme im Wettstreit mit dem Bargeld. *Sparkasse, 6*, 271–274.

d'Inverno, M., & Luck, M. (2003). *Understanding agent systems*. Berlin: Springer.

Jennings, N.R., Sycara, K., & Wooldridge, M. (1998). A road map of agent research and development. *Autonomous Agents and Multi-Agent Systems, 1*(1), 7–38.

Joshi, N., & Ramesh, V.C. (1998). *On mobile agent architectures*. Technical report. ECE Department, Illinois Institute of Technology.

Kurbel, K., & Loutchko, I. (2001). A framework for multi-agent electronic marketplaces: Analysis and classification of existing systems. *Proceedings of International ICSC Congress on Information Science Innovations (ISI 2001). Natural and Artificial Intelligence Systems Organization*. American University in Dubai, U.A.E.

Nwana, H. S. (1996). Software agents: An overview. *Knowledge Engineering Review, 11*(3), 205–244.

Rahwan, I., Ramchurn, S.D., Jennings, N.R., McBurney, P., Parsons, S., & Sonenberg, L. (2004). Argumentation-based negotiation. *The Knowledge Engineering Review* (to appear).

Seitz, J., Stickel, E., & Woda, K. (1999). Electronic payment systems: A game-theoretic analysis. In M. Khosrow-Pour (Ed.), *Managing information technology resources in organizations in the next millennium. Proceedings of the 1999 Information Resources Management Association International Conference* (pp. 564-568). Hershey, PA: Idea Group Publishing.

Seitz, J., Stickel, E., & Woda, K. (2002). Impacts of software agents in ecommerce systems on customer's loyalty and on behavior of potential customers. In B. Fazlollahi (Ed.), *Strategies for ecommerce success* (pp. 208-223). Hershey, PA: IRM Press.

Sycara, K., Decker, K., Pannu, A., Williamson, M., & Zeng, D. (1996, December). Distributed intelligent agents. *IEEE Expert*. Retrieved August 29, 2003, from *http://www-2.cs.cmu.edu/~softagents/publications.html#1996*

Teuteberg, F., & Kurbel, K. (2002, April 24–25). Anticipating agents' negotiation strategies in an e-marketplace using belief models. In W. Abramowicz (Ed.), *Proceedings of the 5th International Conference on Business Information Systems (BIS 2002)*, Poznan, Poland (pp. 91–100).

KEY TERMS

Collaborative Filtering: Collaborative filtering methods combine personal preferences of a user with preferences of like-minded people to guide the user.

Customer Loyalty: Because there is no existing ownership to service products, suppliers have to make special efforts to get long-standing customers.

Customer Profiling: Usage of the Web site to get information about the specific interests and characteristics of a customer.

Customization: The adjustment of products or services to individual needs. Basic characteristics are implemented in the product or service and may be controlled by parameters.

Disintermediation: The elimination of agents, like wholesale dealers or brokers, who built the former relationship between producer and consumer. Disintermediation allows the direct supply of the consumer.

One-to-One-Marketing: One-to-one-marketing means the direct dialog between a producer and an individual consumer or a group of consumers with similar needs.

Personalization: Web-based personalization means providing customized content to individual users using Web sites, e-mails and push technologies.

Software Agents: Computer programs that are characterized by reactivity, autonomy and proactivity. Therefore, the software agent interacts with its environment.

Software and Systems Engineering Integration

Rick Gibson
American University, USA

INTRODUCTION

With software an increasingly significant component of most products, it is vital that teams of software and systems engineers collaborate effectively to build cost effective, reliable products. This article will identify the key aspects of software engineering and systems engineering in an effort to highlight areas of consensus and conflict to support current efforts by practitioners and academics in the both disciplines in redefining and integrating their professions and bodies of knowledge.

In response to increasing concerns about software development failures, the Software Engineering Institute (SEI) pioneered a software process improvement model in 1988, with the fully developed version of their Capability Maturity Model for Software (SW- CMMâ) appearing in 1993. Since the early nineties, there have been comparable improvement models introduced in the systems engineering community as well, some of which have been published and widely accepted include: Systems Engineering Capability Maturity Model (SE-CMM), also known as the Electronic Industries Alliance Interim Standard (EIA/IS) 731, Systems Engineering Capability Model (SECM), and the Integrated Product Development Capability Maturity Model (IPD-CMM). The resulting avalanche of models and standards has been described by Sarah Sheard (Software Productivity Consortium) as a "Framework Quagmire". In December of

Table1. Software and system engineering similarities and differences

Similarities	Differences
Definition and analysis involves manipulation of symbols.	Software is not subject to physical wear or fatigue.
Highly complex aggregation of functions, requiring satisfying (though not optimizing) multiple criteria.	Copies of software are less subject to imperfections or variations.
Decisions driven by need to satisfy quality attributes such as reliability, safety, security, and maintainability.	Software is not constrained by the laws of physics.
Easy and dangerous to suboptimize solutions around individual subsystem functions or quality attributes.	Software interfaces are conceptual, rather than physical—making them more difficult to visualize.
Increasing levels of complexity and interdependency.	Relative to hardware, software testing involves a larger number of distinct logic paths and entities to check.
	Unlike hardware, software errors arrive without notice or a period of graceful degradation.
	Hardware repair restores a system to its previous condition; repair of a software fault generally does not.
	Hardware engineering involves tooling, manufacturing, and longer lead times, while software involves rapid prototyping and fewer repeatable processes.

2000, the SEI initiated the Capability Maturity Model–Integrated (CMMI^{SM}) project, which combines best practices from the systems and software engineering disciplines. (Note: CMMâ and CMMI^{SM} are copyrights and service marks of the Software Engineering Institute.)

Recent studies (Carter et al., 2003; Goldenson & Gibson, 2003) have validated the SEI's assertion the each of the disciplines benefit from incorporation of principles from the other. Moreover, there appears to be no fundamental differences between the disciplines that would prevent their integration.

BACKGROUND

There is great hope that the SEI imitative will provide the impetus to overcome some long-standing discipline boundaries. The nature of the systems and software engineering work has led to terminology differences rooted in the very descriptions of the disciplines. One important problem with software is the difficulty in understanding its inherent level of quality.

Issues and concerns regarding such an integration were articulated by Barry Boehm and Fred Brooks as early as 1975. Boehm suggested that the adoption of systems engineering reliability techniques by software engineers was counterproductive. Moreover, Brooks' Law suggests that a common systems engineering solution to schedule slippage (add more people) will only make late software projects even later.

More recently, Boehm (1994) expressed concerns that, in spite of the central function of software in modern systems, the two engineering disciplines have not been well integrated. Boehm articulated similarities and differences as shown in Table 1.

Software engineering, as defined by the Institute of Electrical and Electronics Engineers (IEEE, 2001), is: (1) the application of a systematic, disciplined, quantifiable approach to the development, operation, and maintenance of software; that is, that application of engineering to software; (2) The study of approaches as in (1)—and further identifies the body of knowledge for software engineering to be: software requirements, software design, software construction, software testing, software maintenance, software configuration management, software engineering management, software engineering process, software engineering tools and methods, and software quality.

A useful definition of systems engineering resides in an in-process body of knowledge document by the International Council on Systems Engineers (Leibrandt, 2001, p. 3), which defines systems engineering in terms of product and process: "…product oriented engineering discipline whose responsibility is to create and execute an interdisciplinary process to ensure that customer and stakeholder needs are satisfied in a high quality, trustworthy, cost effective and schedule compliant manner throughout a system's lifecycle". The process starts with customer needs, and consists of stating the problem, investigating alternatives, modeling, integrating, launching the system, and assessing performance. Moreover, the system engineer is responsible for pulling together all the disciplines to create a project team to meet customers' needs. The complete systems engineering process includes performance, testing, manufacturing, cost, schedule, training and support, and disposal. The body of knowledge recognizes that systems engineering processes often appear to overlap software and hardware development processes and project management. Thus, systems engineering is a discipline that focuses on processes; it develops structure, and efficient approaches to analysis and design to solve complex engineering problems. In response to concerns about integrated development of products, the system engineer plans and organizes technical projects and analyzes requirements, problems, alternatives, solutions and risks. Systems engineering processes are not specific to a particular discipline; they can be applied in any technical or engineering environment.

In short, software engineering is defined by IEEE Standard 610.12 as the application of a systematic, disciplined, quantifiable approach to the development, operation, and maintenance of software—that is, the application of engineering to software. Eisner (2002) adopts the International Council on Systems Engineering (INCOSE) definition of systems engineering as an interdisciplinary approach and means to enable the realization of successful systems.

When different process models are in place within developer groups, say for systems engineering and software engineering of an organization, the organizations will have communication problems, be unable to improve their processes, and if the combined performance of one advances beyond the other in capability, then the problems are even more profound (Johnson, 1998).

In 2002, the SEI released a single integrated capability model for systems engineering and software engineering, integrated product and process development and supplier sourcing. The new model, Capability Maturity Model Integrated (CMMI), is intended to improve organizations' development and maintenance of products. The CMMI will eventually replace the SEI's Software Capability Maturity Model (Phillips, 2002). In the integrated model (SEI, 2002), CMMI, the categories and processes are:

One purpose of the CMMI was to evolve the software CMM while integrating the best features of the systems engineering capability models. The combination of the

Process Management
Organizational Process Focus
Organizational Process Definition
Organizational Training
Organizational Process Performance
Organizational Innovation and Deployment

Engineering
Requirements Management
Requirements Development
Technical Solution
Product Integration
Verification
Validation
Validation

Project Management
Project Planning
Project Monitoring and Control
Supplier Agreement Management
Integrated Product Management
Risk Management
Qualitative Project Management
Integrated Teaming

Support
Configuration Management
Process and Product Quality Assurance
Measurement and Analysis
Causal Analysis and Resolution
Decision Analysis and Resolution
Organizational Environment for Integration
Organizational Environment for Integration

practices of the models into one single framework required more than just combining practices because of differences in interpretation, focus, and terminology. Compromises and intentional inefficiencies were required in order to integrate these models.

With the arrival of the CMMI, a wider continuum of the product life cycle has been targeted for possible enhancement, no longer limiting process improvement only to the development of software. This integrated approach provides a reduction in the redundancy and intricacy resulting from the use of multiple, separate process improvement models. For organizations that wish to assess their process improvement efforts against multiple disciplines, the CMMI provides some economies of scale in model training and assessment training. This one evaluation method can provide separate or combined results for the fields of software and systems engineering. Furthermore, software organizations can also focus on the amplifications for software engineering within the engineering shared process areas and take advantage of any systems engineering amplifications that are helpful. Although still subject to debate, a distinction is made between base and advanced engineering practices as model constructs. Adopting the continuous representation of CMMI not only forces software organizations to define business goals and choose process areas that should be implemented first to focus on these goals, but it also forces companies who are choosing a new subcontractor to do the same. One of the claimed benefits of a staged representation is that it facilitates comparisons among organizations (Shrum, 2000). While it may simplify comparisons, it does so at the loss of additional details. Using a continuous representation, the comparison can be done based on the process areas that are judged by the organization as important rather than simply comparing the organization's maturity score. When using a continuous representation, there is less likelihood that organizations will try to attain a specific level without reasons within their business to do so. It provides an incentive to address processes that would have the greatest impact on their business goals.

FUTURE TRENDS

Despite anticipated problems, bringing systems engineering best practices into the established software process improvement models is expected to be very beneficial. Boehm (1994) reminds us that an important reason to overcome or bridge these differences is to establish an adequate supply of people who can deal with complex systems problems. The Bureau of Labor Statistics (1997) estimates of anticipated growth in information technology jobs, shown in Table 2, provides further support for this concern.

The final job type, managers, is a significant concern addressed by Jerry Weinberg in an interview (Layman, 2001). Weinberg explains that the software development problems are growing faster than individuals' levels of competence. Moreover, he asserts that the current state of practice is one where we need to apply a few fundamentals (e.g., requirements, reviews, configuration management); that is, things known to be useful, but not adopted in the sense of consistent application. Soloman (2002) provides guidelines for using the CMMI to improve earned value management.

Table 2. Anticipated employment growth 1996-2006

Type of Job	1996 Employment	2006 Employment	% Change
Database Administrators and Computer Support	212,000	461,000	118%
Computer Engineers	216,000	451,000	109%
Systems Analysts	506,000	1,025,000	103%
Data Processing Equipment Repair	80,000	121,000	52%
Engineering, Science, and Computer Systems *Managers*	343,000	498,000	45%

It has been suggested that the systems engineering—hardware engineering interfaces have matured nicely over many years, but that the systems engineering—software engineering interface is not as mature as the various hardware engineering interfaces.

Meanwhile, the dependency on the systems engineering-software engineering interface has increased faster than it has matured. Specific concerns by discipline include:

The State of Systems Engineering

- Most successful projects rely on expertise established with similar systems.
- Lack of documented processes makes repeatability difficult.
- Development efforts for unprecedented or significantly different systems often encounter problems.

The State of Software Engineering

- The brief history of software development has been filled with problems of cost overruns, schedule slippage, and failure to achieve performance goals.
- Systems are increasingly dependent on software, yet hardware typically gets the most visibility.

CONCLUSION

Although Pierce (2000) suggests that the CMMI is more of a merged model than an integrated one, which may serve to prolong the separation of the disciplines, Rassa (2001) summarizes the benefits of the CMMI project as follows:

- Common, integrated vision of improvement for all organizational elements;
- Means of representing new discipline-specific information in a standard, proven process improvement context;
- Efficient, effective assessment and improvement across an organization's multiple process disciplines;
- Reduced training and assessment costs.

Although many software-only organizations remain adamant that they do not do systems engineering, all software must run on some computer system, and interface with others. This perceived separation of concerns exacerbates the difficulties associated with hardware/software/system tradeoff decisions, which are further complicated by terminology differences and disparate mental models. Interpretive guidance for CMMI implementation has been provided (e.g., Chrissis et al., 2003)

However, the integration potential of the CMMI℠ can allow the system and software engineering communities to get the most out of their similarities. The CMMI℠ allows organizations to tailor the model to mesh with their own mission and goal statements as well as their business objectives. Each individual project can use CMMI℠ models for individual disciplines and discipline combinations because the architecture of the CMMI℠ does not force the employment of every discipline for every organization implementing it. Before the CMMI℠, the systems engineering models shared many of the same principles as the software version of CMM, but were written to address the needs and terminology of the systems engineering community. Because the CMMI℠ includes the common and shared elements and best features of both software and system engineering together with discipline-specific elements, an organization can generate integrated capability maturity models or discipline-specific capability models. With CMMI℠, an organization can still capitalize on these similarities and improve the efficiency of and the return on investment for process improvement. The resulting integrated capability models will adapt to an organization's business purposes.

The concept of an architecture continues to serve as a theoretical link for both the software/system tradeoffs and the integration of process improvement efforts. While respecting the legitimate differences in areas such as reliability testing, it is important to sustain the hope that overlapping or underlying theories will emerge regarding areas of common concern such as: requirements, security, safety, and performance.

In order to achieve true integration of software and system engineering practices into one process improvement model, the remaining differences of terminology and model construction have to be addressed. These two communities have well-developed disparate languages and methodologies that are reflected in their different origins, models and perspectives; differences that have become entrenched in their organizational cultures. With the adoption of an integrated process improvement model, an organization can assess both software and systems engineering functions, reduce conflict and increase consensus.

REFERENCES

Boehm, B. (1994, July-September). Integrating software engineering and system engineering. *The Journal of INCOSE, I*(I).

Carter, L., Graettinger, C., Patrick, M., Wemyss, G., & Zasadni, S. (2002). *The road to CMMI: Results of the first technology transition workshop.*

Chrissis, M., Wemyss, G., Goldenson, D., Konrad, M., Smith, K., & Svolou, A. (2003). *CMMI® interpretive guidance project: Preliminary report.*

Eisner, H. (2002). *Essentials of project and systems engineering management.* New York: John Wiley & Sons.

Goldenson, D., & Gibson, D. (2003). *Demonstrating the impact and benefits of CMMI®:* An update and preliminary results.

International Council on Systems Engineering (INCOSE). Retrieved February 2002, from *http://www.incose.org/*

Johnson, K.A., & Dindo, J. (1998, October). Expanding the focus of software process improvement to include systems engineering. *Crosstalk: The Journal of Defense Software Engineering,* 13-19.

Layman, B. (2001, September). An interview with Jerry Weinberg. *Software Quality Professional, 3*(4), 6-11.

Leibrandt, R. (2001, April 22). A guide to the systems engineering body of knowledge (SEBoK). Retrieved February 2002, from *www.incose.org/orlando/sebok/attach/sebok_text.doc*

Phillips, M. (2002, February). CMMI version 1.1: What has changed. *Crosstalk: The Journal of Defense Software Engineering,* 4-6.

Pierce, B. (2000, July). Is CMMI ready for prime time. *Crosstalk: The Journal of Defense Software Engineering.*

Rassa, B. (2001, March 13). Beyond CMMI-SE/SW v1.0. Software Engineering Institute. *http://www.sei.cmu.edu/cmmi/publications/sepg01.presentations/beyond.pdf*

Shrum, S. (1999, December). Choosing a CMMI model representation. SEI Interactive. Retrieved July 17, 2001, from *http://www.stsc.hill.af.mil/crosstalk/2000/jul/shrum.asp*

Software Engineering Institute. (2002, March). *Capability maturity model integrated (CMMI), version 1.1.*

Soloman, P. (2002). *Using CMMI® to improve earned value management.*

KEY TERMS

Capability Maturity Model (CMM): Contains the essential elements of effective processes for one or more disciplines. It also describes an evolutionary improvement path from ad hoc, immature processes to disciplined, mature processes with improved quality and effectiveness.

Process Architecture: Describes the ordering, interfaces, interdependencies, and other relationships among the process elements in a standard process. Process architecture also describes the interfaces, interdependencies, and other relationships between process elements and external processes (CMMI).

Process Improvement: A program of activities designed to improve the performance and maturity of the organization's processes, and the results of such a program.

Quality: The ability of a set of inherent characteristics of a product, product component, or process to fulfill requirements of customers.

Software Engineering: The software engineering discipline covers the development of software systems.

Software engineers focus on applying systematic, disciplined, and quantifiable approaches to the development, operation, and maintenance of software.

Systems Engineering: The systems engineering discipline covers the development of total systems, which may or may not include software. Systems engineers focus on transforming customer needs, expectations, and constraints into product solutions and supporting those product solutions throughout the product life cycle.

S

Software Contracts for Component–Based Web Engineering

Martin Gaedke
University of Karlsruhe, Germany

Martin Nussbaumer
University of Karlsruhe, Germany

Emma Tonkin
University of Karlsruhe, Germany

INTRODUCTION

As an emerging technology, the Web is full of unique challenges for developers, designers – and engineers. Its use for increasingly complex applications such as e-commerce and banking, involving connection to many data sources, highlighted a number of common difficulties. During the design phase, good modeling of complex sites is difficult; during implementation, the design models are often found to be difficult to translate into an implementation model; afterwards, maintenance and long-term management of the site's evolution offer problems of their own. Qualities such as integrity, security, and usability are often afterthoughts, or left unconsidered – most particularly during the site's post-deployment evolution. Many organizations working for the Web still use ad-hoc, chaotic development and maintenance methods, even though the issues surrounding development are now widely known.

Solving these issues called for software engineering – a discipline covering design, development and use of computer software (Weik, 1989) – to be applied to the Web. Researchers in the field noticed a gap between the granularity of design models and that of implementation models for the Web (Gellersen, Wicke & Gaedke 1997) – that is to say, the models that worked best for designers did not provide the right levels of detail for the implementation stage.

BACKGROUND

Software engineering approaches have been brought to the Web through design models especially suited for Web and other hypermedia technologies, for example OOHDM (Schwabe, Rossi & Barbos, 1996), RMM (Isakowitz, Stohr & Balasubramaninan, 1995), and UML (Conallen, 1999). Nevertheless, the lack of representation of higher-level concepts in the implementation remains.

A proposed solution to this issue came from *compositional design and reuse*, a concept from the software engineering domain within which components are designed such that they may be reused as building blocks. An application may then be developed – composed – from the available components. Should additional components be required, they are developed; they may then be reused in future compositions. This concept resulted in a new sub-discipline: component-based Web engineering (CBWE) (Gaedke & Turowski, 2002).

Components seem to solve the granularity problem, as components can be described in whatever granularity suits both parties. For example, a component might represent a formatted paragraph of text, a navigational menu, an active component such as a calendar – or a single link. Components can represent a workflow as a set of pages, forms, and underlying "business logic". Once developed, such components can be stored and reused. Components can even represent an assemblage of smaller components.

The advantages of this approach, christened the *WebComposition* approach (Gaedke, 2000), were demonstrated by the development of proof-of-concept language WCML, the Web-Composition Markup Language (Gaedke, Schempf & Gellersen, 1999), which offers a convenient way to define and represent components. It uses the eXtensible Markup Language (XML), an almost ubiquitous standard, for this purpose. WCML permits XML-based definition of components, associated typed attributes (name-value pairs) referred to as *properties*, and relationships between components. Components are stored in a *repository*, a dedicated data store, and referred to by a unique ID (UUID). These component assemblies provide WebComposition services or in short services – for example, an ordering service is made up of a form interface and a product database. Additional components provide business logic to control the order process, thus enabling the concept of constructing Web applications based on the services they provide. This abstract definition allows for different implementations of the

WebComposition service concept, for example using components, Web services, or even the secretary phone in the office.

WCML was developed as part of a complete approach, which defined a disciplined procedure of composing Web applications with components based on the WebComposition component model (Gaedke, 2000). It is a synthesis of a component-oriented process model with a dedicated Web application framework, reuse management, and dedicated component technology. The details of the reuse model are discussed in Gaedke and Turowski (2002).

DROWNING IN INFORMATION – AND STARVING FOR KNOWLEDGE?

Just as in R.D. Roger's popular quotation, one component in a repository resembles a small needle in a big haystack. A designer looking for an appropriate component in the repository needs a good way to find it, so that he or she can retrieve it using its unique ID. What methods exist to sort or index components within a repository, and what techniques permit components to be most usefully described and specified?

WebComposition services, and the components of which they are built, can be represented in several ways that use searchable indexes to direct the user to the UUID of likely components: controlled indexing, uncontrolled indexing, and methods containing semantic information. Controlled indexing refers to the use of controlled vocabularies or categorizations, such as hierarchical or taxonomical classification or *faceted* classification, where objects are described by characteristic. For example, facets of this encyclopedia entry include its publication date, the expected knowledge of its readers, and its subject type. Uncontrolled indexing refers to methods without the constraint of a controlled vocabulary, such as free-text keywords added to the object's description by a human or a computer, or attribute-value pairs. Examples of both are shown in Figure 1. Controlled and uncontrolled indexing methods were compared (Frakes & Pole, 1994) and both were found to be useful – but, the authors concluded, no single method is perfect. A good result often comes from mixing several methods.

Unlike most types of data, components have an additional property – behavior. As elements in a software application, they have to perform a task, and they must work reliably and accurately – and they must interoperate correctly with other components. In order to ensure this, *software contracts* were introduced in 1988 by Meyer, as part of the Eiffel programming language (Meyer, 1988). Software contracts may include information on many levels, depending on the context of the agreement; configurable values may be negotiated as part of the contract.

Software components are obligations to which a service provider (a component) and a service consumer (a client) agree. The provider guarantees that a service it offers, under certain conditions – such as the provision of appropriate data – will be performed to a guaranteed quality. Furthermore, it also provides information about certain external characteristics, such as its interfaces. The following sections will discuss the different levels for specifying a component, as shown in Figure 2.

Figure 1. Classifying an article

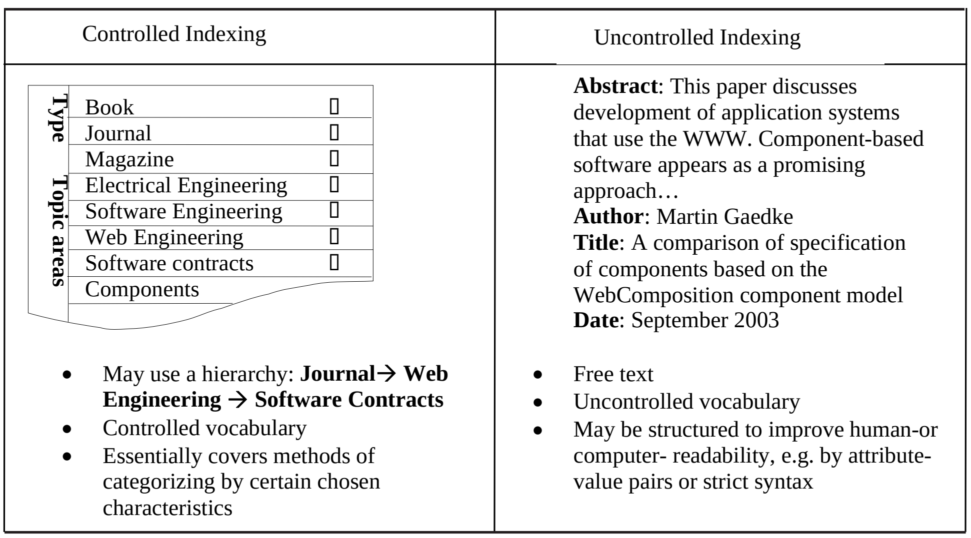
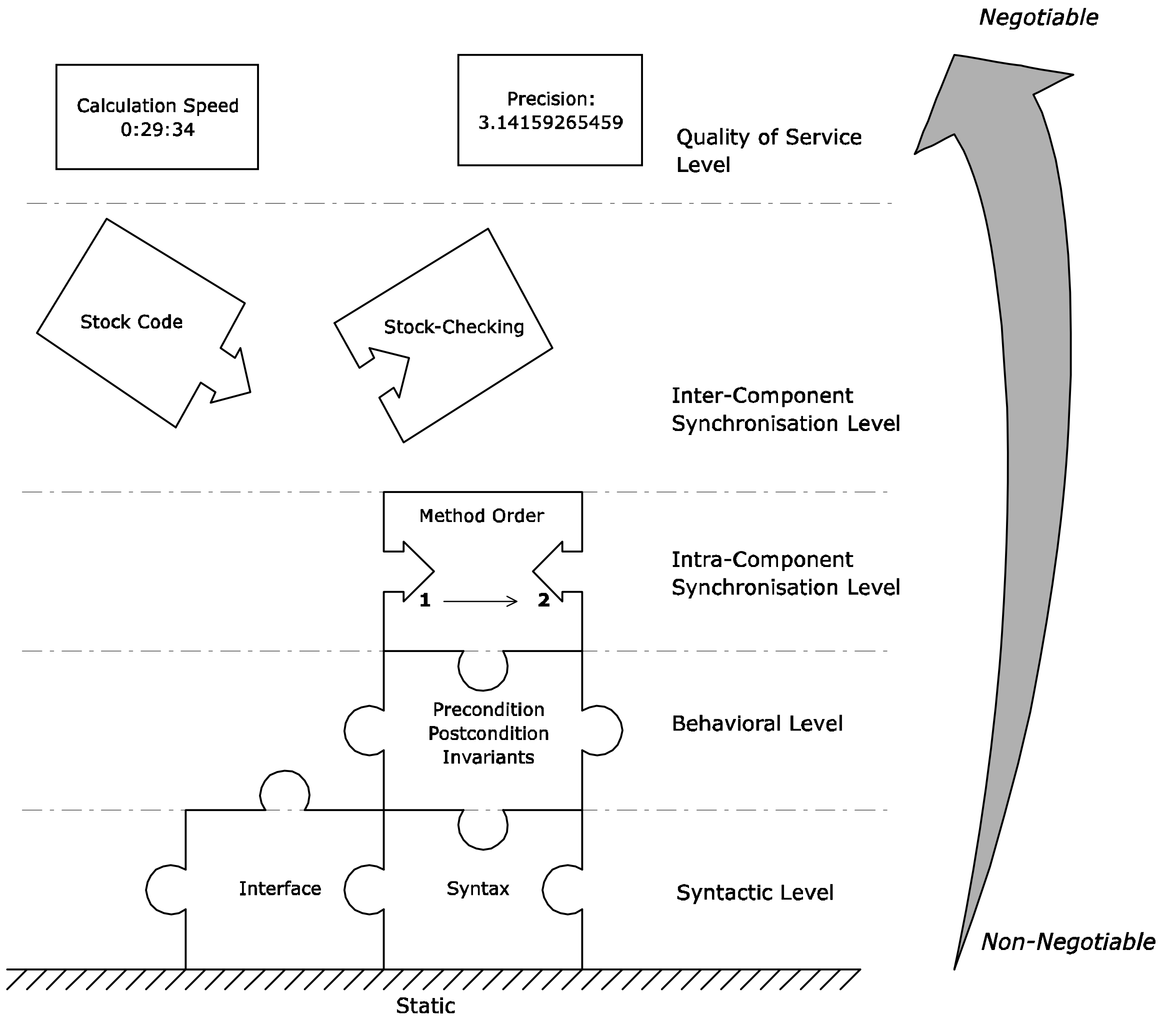


Figure 2. Software contract levels – specifications of a component



Syntactic and Behavioral Information Levels

On the most fundamental level, software contracts specify the interface and behavior of a component. This offers a way for components to provide information about what their functions are, in what ways they may be addressed, the prerequisites of use, on what data they may operate and what outputs they should provide. Components can use software contracts to describe their interfaces, to describe their behavior, and to set pre-and post-conditions on use.

Intra- and Inter-Component Synchronization Levels

The technique can also be used to describe synchronization issues, both internally and with relation to other contracts. The correct execution of a component, accessed concurrently from multiple sources, may depend on synchronization rules – or policies – ensuring that an entire transaction is correctly executed before another is begun. Externally, components may be mutually dependent; the order in which component services are executed may be controlled by software contract. A component could stipulate that another component be run before itself, such as, "Before you access the stock-checking component, you must first retrieve the stock code from the search tool".

Quality of Service Level – Negotiable Properties

Finally, software contracts could be used to negotiate the properties of a component in detail. Components may provide options, such as a trade-off between calculation speeds and result precision or, in the context of the Web, the bandwidth demanded may be traded against properties of the returned object, such as image size, resolution or detail. These are often classed as *non-functional* characteristics.

The WCML is able, by making use of several companion technologies, to include the contract levels described here. The Interface Definition Language (IDL) proposed by the Object Management Group (ISO International Standard 14750) permits the definition of component behavior on the syntactic level (OMG, 2003), to complement the interface definition capabilities provided by WCML itself on the semantic level by WCML schemas. Several IDLs permit definition of the interface between client and server processes.

The Object Constraint Language (OCL) specification (OCL, 1997) builds on the Unified Modeling Language (UML), providing a useful notation for specification of facts at the behavioral level. A temporal extension, Temporal OCL, was developed (Ziemann & Gogolla, 2002) for the purpose of describing synchronization on the inter-component and intra-component levels. Other graphical notations may be used for the purpose, such as Petri nets. Appropriate alternative notations include process algebras, developed as an algebraic approach to the study of concurrent processes (Van Glabbeek, 1987), and process calculi. Several process calculus languages exist, designed to permit descriptions or specifications of software that consists of multiple interactive elements.

Natural language may be used for many specifications at the quality of service level – for example, the details of a component's level of accessibility to the visually impaired may be specified in this way. Other services require more specialized constraints to be specifiable, in which case alternative notations should be considered (mathematics or scientific notation, for example).

FUTURE TRENDS

At this time, we envisage the next steps in software component design to begin with development of a component referencing system that combines the concepts of software contracts and traditional metadata with the Universal Description, Discovery and Integration (UDDI) protocol (UDDI, 2002). This standard, introduced in September 2000 as the result of a joint industry effort by 36 major companies, provides a standard for development of an electronically searchable directory of services, such as Web services. Web services are definable as a special case of the Web component concept, software components that communicate via the XML-based standard Simple Object Access Protocol (SOAP). UDDI, designed to enable efficient business-to-business exposure of Web services, is considered to be adaptable to the general case of software components. Once combined and adapted, the technologies are expected to provide a powerful method supporting component discovery, either manually or in an automated fashion – self-discovery.

Reflection, in the sense of the ability to discover interfaces of previously unknown components and therefore inform oneself on components' abilities and provided services, is also an interesting technique in the context of CBWE. This approach paves the way to implementation of software contracts on the Web, as well as a greater degree of automation and fact and error checking.

Thirdly and finally, a group of techniques considered to be of importance to component specification is the addition of information and processes that contribute towards successful and transparent digital rights management (DRM). Issues surround the use, deployment and control of components, which may be at least partially solved by implementing a form of rights management within the component discovery process and underlying runtime environment.

CONCLUSION

Component-based Web engineering helps to solve the problems in representing abstract design ideas in code. This raises the problem of specifying components so that they can be located in a large data repository. WCML has provided a method by which the concepts behind CBWE can be demonstrated and has been used to illustrate the principle. Furthermore, the WebComposition approach has successfully been applied to several real-world applications, and its advantages verified in a 4-year project for a large international Web-application: "E-Victor Procurement Portal" at Hewlett-Packard.

Recent efforts within Web engineering have focused on specifying components by means of software contracts. By combining this method with other classifications, and perhaps with additional information such as digital rights management markers, components may be made available and discovered by means of the various specifications. The range of information available about each component may be such that it is useful for users, developers – in the technical and in the business contexts – and even for computer-based automatic discovery of components.

REFERENCES

Conallen, J. (1999). Modeling Web application architectures with UML. *Communications of the ACM*, *42*(10), 63-70.

Frakes, W.B., & Pole, T.P. (1994). Empirical study of software reuse. *IEEE Transactions on Software Engineering, 20*(8), 617-630.

Gaedke, M. (2000). *Komponententechnik für entwicklung und evolution von anwendungen im World Wide Web*. Aachen: Shaker Verlag.

Gaedke, M., Schempf, M., & Gellersen, H.-W. (1999). WCML: An enabling technology for the reuse in object-oriented Web engineering. *Poster-Proceedings of the 8th International World Wide Web Conference (WWW8)*, Toronto, Ontario, Canada.

Gaedke, M., & Turowski, K. (2002). Specification of components based on the WebComposition component model. In S. Becker (Ed.), *Data warehousing and Web engineering* (pp. 275-284). Hershey, PA: IRM Press.

Gellersen, H.-W., Wicke, R., & Gaedke, M. (1997). WebCompostion: An object-oriented support system for the Web engineering lifecycle. *Computer Networks and ISDN Systems 29 – Special Issue on the 6th Intl. World-Wide Web Conference*, Santa Clara, CA (pp. 1429-1437).

Isakowitz, T., Stohr, E.A., & Balasubramaninan, P. (1995). RMM: A methodology for structured hypermedia design. *Communications of the ACM*, *38*(8), 34-44.

Meyer, B. (1988). *Object-oriented software construction*. Englewood Cliffs: Prentice Hall.

OCL. (1997). Rational Software, Microsoft, Hewlett-Packard, Oracle, Sterling Software, MCI Systemhouse, Unisys, ICON Computing, IntelliCorp, i-Logix, IBM, ObjecTime, Platinum Technology, Ptech, Taskon, Reich Technologies, & Softeam. *Object Constraint Language Specification: Version 1.1*. Retrieved September 1, 1997, from *http://www.rational.com/uml*

OMG. (2003). OMG IDL: Details. Retrieved September 16, 2003, from *http://www.omg.org/gettingstarted/omg_idl.htm*

Schwabe, D., Rossi, G., & Barbosa, S. (1996). Systematic hypermedia design with OOHDM. *Proceedings of the ACM International Conference on Hypertext, Hypertext'96* (pp. 116-128).

UDDI. (2002). IBM, Microsoft, Hewlett-Packard, Oracle, Intel, SAP. *UDDI v.3.0 Specification*. Retrieved September 16, 2003, from *http://www.uddi.org/specification.html*

Van Glabbeek, R.J. (1987). Bounded nondeterminism and the approximation induction principle in process algebra. In F.J. Brandenburg, G. Vidal-Naquet & M. Wirsing (Eds.), *Proceedings of 4th Annual Symposium on Theoretical Aspects of Computer Science: STACS 87* (pp. 336-347). Springer-Verlag.

Weik, M. (1989). *Communications standard dictionary* (2nd ed.). New York: Van Nostrand Reinhold Co.

Ziemann, P., & Gogolla, M. (2002). An extension of OCL with temporal logic. In J. Jurjens, M.V. Cengarle, E.B. Fernandez, B. Rumpe & R. Sandner (Eds.), *Critical systems development with UML: Proceedings of the UML '02 Workshop* (pp. 53-62). TUM, Institut für Informatik.

KEY TERMS

Component-Based Web Engineering: The application of systematic, disciplined and quantifiable component-based and reuse-oriented approaches to the understanding, construction, operation and evolution of Web-based applications and systems.

Compositional Design and Reuse: The design or composition of components from a collection of generic building-block components.

Controlled Indexing: Objects may be indexed using controlled vocabularies, hierarchies or classifications.

DRM: Digital Rights Management (DRM) refers to methods and technologies designed to control access to or use of copyrighted data.

Federated Web Services: Usually, problems like authorization are handled on a network-by-network basis – a given login may only be valid to a single network or site. Federation provides a means to transcend this limitation. In our example, groups could federate by declaring trust in each other, thus permitting authorization across multiple organizations.

Software Contracts: A means of negotiating obligations to which a service donator and client may agree. This may include information about several levels, from the technical (available communications protocol) to the non-functional (quality of service).

UDDI: The Universal Description, Discovery and Integration specification provides a platform-independent way of describing and discovering services and service providers.

Uncontrolled Indexing: The indexing of objects using uncontrolled vocabularies, such as descriptions or semi-structured metadata.

UUID: Universally Unique IDs (UUID) are typically machine-generated, and are used as world-wide unique labels. Their uniqueness is assured through such means as encoding of timestamps and hardware identifiers.

Software Requirements Risk and Maintainability

Norman F. Schneidewind
Naval Postgraduate School, USA

INTRODUCTION

In order to continue to make progress in software measurement, as it pertains to reliability and maintainability, there must be a shift in emphasis from design and code metrics to metrics that characterize the risk of making requirements changes. By doing so, the quality of delivered software can be improved because defects related to problems in requirements specifications will be identified early in the life cycle. An approach is described for identifying requirements change risk factors as predictors of reliability and maintainability problems. This approach can be generalized to other applications with numerical results that would vary according to application. An example is provided that consists of 24 space shuttle change requests, 19 risk factors, and the associated failures and software metrics.

In addition to the relationship between requirements and reliability and maintainability, there are the intermediate relationships between requirements and software metrics (e.g., size, complexity) and between metrics and reliability and maintainability. These relationships may interact to put the reliability and maintainability of the software at risk because the requirements changes may result in increases in the size and complexity of the software that may adversely affect reliability and maintainability.

BACKGROUND

Several projects have demonstrated the validity and applicability of applying metrics to identify fault prone software at the code level (Khoshgoftaar & Allen, 1998; Khoshgoftaar et al., 1996a, 1996b; Schneidewind, 2000). This approach is applied at the requirements level to allow for early detection of reliability and maintainability problems. Once high-risk areas of the software have been identified, they would be subject to detailed tracking throughout the development and maintenance process (Schneidewind, 1999).

Much of the research and literature in software metrics concerns the measurement of code characteristics (Nikora et al., 1998). This is satisfactory for evaluating product quality and process effectiveness once the code is written. However, if organizations use measurement plans that are limited to measuring code, the plans will be deficient in the following ways: incomplete, lack coverage (e.g., no requirements analysis and design), and start too late in the process. For a measurement plan to be effective, it must start with requirements and continue through to operation and maintenance. Since requirements characteristics directly affect code characteristics and hence reliability and maintainability, it is important to assess their impact on reliability and maintainability when requirements are specified. As will be shown, it is feasible to quantify the risks to reliability and maintainability of requirements changes — either new requirements or changes to existing requirements.

Once requirements attributes that portend high risk for the operational reliability and maintainability of the software are identified, it is possible to suggest changes in the development and maintenance process of the organization. To illustrate, a possible recommendation is that any requirements change to mission critical software — either new requirements or changes to existing requirements — would be subjected to a *quantitative* risk analysis. In addition to stating that a risk analysis would be performed, the policy would specify the risk factors to be analyzed (e.g., number of modifications of a requirement or *mod level*) and their threshold or critical values. The validity and applicability of identifying critical values of metrics to identify fault prone software at the code level have been demonstrated (Schneidewind, 2000). For example, on the space shuttle, rigorous inspections of requirements, design documentation, and code have contributed more to achieving high reliability and maintainability than any other process factor. The objective of these policy changes is to prevent the propagation of high-risk requirements through the various phases of software development and maintenance. The payoff to the organization would be to reduce the risk of mission critical software *not* meeting its reliability and maintainability goals during operation.

APPROACH TO ANALYZING REQUIREMENTS RISK

By retrospectively analyzing the relationship between requirements and reliability and maintainability, it is possible to identify those risk factors that are associated with

S

Table 1. Change request hierarchy

Change Requests (CRs)
1. No Discrepancy Reports (i.e., CRs with no DRs)
2. (Discrepancy Reports) or (Discrepancy Reports and Failures)
   - 2.1 No failures (i.e., CRs with DRs only)
   - 2.2 Failures (i.e., CRs with DRs and Failures)
     - 2.2.1 Pre-release failures
     - 2.2.2 Post-release failures

reliability and maintainability. In addition, risk factors are prioritized based on the degree to which the relationship is statistically significant. In order to quantify the effect of a requirements change, various risk factors were used, which are defined as the attribute of a requirement change that can induce adverse effects on reliability (e.g., failure incidence), maintainability (e.g., size and complexity of the code), and project management (e.g., personnel resources).

Table 1 shows the Change Request Hierarchy of the space shuttle, involving change requests (i.e., a request for a new requirement or modification of an existing requirement), discrepancy reports (i.e., DRs: reports that document deviations between specified and observed software behavior), and failures. In Table 1, Category 1 versus Category 2 was analyzed with respect to risk factors as discriminants of the categories.

Categorical Data Analysis

Using the null hypothesis, Ho: A risk factor is not a discriminator of reliability and maintainability versus the alternate hypothesis H_1: A risk factor is a discriminator of reliability and maintainability, categorical data analysis is used to test the hypothesis. A similar hypothesis was used to assess whether risk factors can serve as discriminators of metrics characteristics. Requirements, requirements risk factors, reliability, and metrics data from the space shuttle *"Three Engine Out"* software (abort sequence invoked when three engines are lost) were used to test the hypotheses.

Table 2. Definition of samples

Sample	Size
Total CRs	24
CRs with no DRs	14
CRs with (DRs only) or (DRs and Failures)	10
CRs with modules that caused failures	6
CRs can have multiple DRs, failures, and modules that caused failures. CR: Change Request. DR: Discrepancy Report.	

Table 2 shows the definition of the change request samples that were used in the analysis. Sample sizes are small due to the high reliability of the space shuttle software. However, sample size is one of the parameters accounted for in the statistical tests that produced statistically significant results in certain cases.

To minimize the effects of a large number of variables that interact in some cases, a statistical categorical data analysis was performed incrementally. Only one category of risk factor at a time was used to observe the effect of adding an additional risk factor on the ability to correctly classify change requests that have *No Discrepancy Reports* versus change requests that have ((*Discrepancy Reports Only*) *or* (*Discrepancy Reports and Failures*)). The Mann-Whitney test for difference in medians between categories was used because no assumption need be made about statistical distribution. In addition, some risk factors are ordinal scale quantities (e.g., modification level); thus, the median is an appropriate statistic to use.

RISK FACTORS

One of the software process problems of the NASA Space Shuttle Flight Software organization is to evaluate the risk of implementing requirements changes. These changes can affect the reliability and maintainability of the software. To assess the risk of change, the software development contractor uses a number of risk factors. The following are the definitions of the 4 out of 19 risk factors that were found to have a statistically significant relationship with reliability and maintainability. The names of the risk factors used in the analysis are given in quotation marks.

Table 3. Statistically significant results (alpha ≤ .05). CRs with no DRs vs. ((DRs only) or (DRs and Failures)). Mann-Whitney Test

Risk Factor	**Alpha**	**Median Value** CRs with no DRs	**Median Value** (DRs only) or (DRs and Failures)
issues	.0076	1	14
space	.0186	6	123
mods	.0401	0	4
sloc	.0465	10	88.5

Table 4. Selected risk factor module characteristics

Change Request	**Failed Module**	**Metric**	**Metric Critical Value**	**Metric Value**
A	1	change history line count in module listing	63	558
A	2	non-commented loc count	29	408
B	3	executable statement count	27	419
C	4	unique operand count	45	83
D	5	unique operator count	9	33
E	6	node count (in control graph)	17	66
All of the above metrics exceeded the critical values for all of the above change requests.				

- Complexity Factor
  o Number of modifications or iterations on the proposed change; “mods”.
  - How many times must the change be modified or presented to the Change Control Board (CCB) before it is approved?
- Size Factor
  o Number of source lines of code affected by the change; “sloc”.
  - How many source lines of code must be changed to implement the change request?
- Requirements Issue Factor
  o Number of possible conflicts among requirements (requirements issues); “issues”
  - Will this change conflict with other requirements changes (e.g., lead to conflicting operational scenarios)?
- Performance Factor
  o Amount of memory space required to implement the change; “space”
  - Will the change use memory to the extent that other functions will not have sufficient memory to operate effectively?

RISK ANALYSIS EXAMPLE

This section provides a risk analysis example, using the space shuttle data and performing the statistical analyses in a. and b., as shown in Tables 3 and 4, respectively. Only those risk factors where there was sufficient data and the results were statistically significant are shown.

a. Categorical data analysis on the relationship between (CRs with no DRs) versus ((DRs only) or (DRs and Failures)), using the Mann-Whitney Test.
b. Identification of modules that caused failures as a result of the CR, and their metric values.

Categorical Data Analysis

Of the original 19 risk factors, only 4 survived as being statistically significant (alpha ≤ .05); 7 were not significant; and 8 had insufficient data to make the analysis. As Table 3 shows, there are statistically significant results for (CRs with no DRs) versus ((DRs only) or (DRs and Failures)) for the risk factors “mods,” “sloc,” “issues,” and "space". The value of alpha in Table 3 is used to prioritize the use of risk factors, with low values meaning high priority. The priority order is: "issues," "space," "mods," and "sloc".

The significant risk factors would be used to predict reliability and maintainability problems for this set of data and this version of the software. The finding regarding "mods" does confirm the software developer's view that this is an important risk factor. This is the case because if there are many iterations of the change request, it implies that it is complex and difficult to understand. Therefore, the change is likely to lead to reliability and maintainability problems. It is not surprising that the size of the change "sloc" is significant because previous studies of space shuttle metrics have shown it to be an important determinant of software quality (Schneidewind, 2000). Conflicting requirements "issues" could result in reliability and maintainability problems when the change is implemented. The on-board computer memory required to implement the change "space" is critical to reliability and maintainability because unlike commercial systems, the space shuttle does not have the luxury of large physical memory, virtual memory, and disk memory to hold its programs and data. Any increased requirement on its small memory to implement a requirements change comes at the price of demands from competing functions.

Identification of Modules that Caused Failures

Table 4 shows modules that caused failures, as the result of the CRs had metric values that far exceed the critical values. The latter were computed in Schneidewind (2000). A critical value is a discriminant that distinguishes high quality from low quality software (Schneidewind, 2000). A module with metric values exceeding the critical values is predicted to cause failures and maintainability problems. Although the sample sizes are small, due to the high reliability of the space shuttle software, the results consistently show that modules with excessive size and complexity lead to failures. Not only will the reliability be low, but this software will also be difficult to maintain. The application of this information is that there is a high degree of risk when changes are made to software that has the metric characteristics shown in the table. Thus, these characteristics should be considered when making the risk analysis.

FUTURE TRENDS

Reliability predictions that are made in the test phase, when failure data are available, are useful, but it would be much more useful to predict at an earlier phase – preferably during requirements analysis—when the cost of error correction is relatively low. Thus, there is great interest in the software reliability and metrics field in using static attributes of software in reliability and maintainability modeling and prediction. Presently, the software engineering field does not have the capability to make early predictions of reliability and maintainability problems. Early predictions would allow errors to be discovered and corrected when the cost of correction is relatively low. In addition, early detection would prevent poor quality software from getting into the hands of the user. As a future trend, the focus in research and practice will be to identify the attributes of software requirements that cause the software to be unreliable and difficult to maintain.

CONCLUSION

Risk factors that are statistically significant can be used to make decisions about the risk of making changes. These changes affect the reliability and maintainability of the software. Risk factors that are not statistically significant should not be used; they do not provide useful information for decision-making and cost money and time to collect and process. Metric characteristics of modules should be considered when making the risk analysis because metric values that exceed the critical values are likely to result in unreliable and non-maintainable software. This methodology can be generalized to other risk assessment domains, but the specific risk factors, their numerical values, and statistical results may vary.

REFERENCES

Khoshgoftaar, T.M., & Allen, E.B. (1998). Predicting the order of fault-fault-prone modules in legacy software. *Proceedings of the Ninth International Symposium on Software Reliability Engineering, 7,* 344-353. Paderborn, Germany.

Khoshgoftaar, T.M., Allen, E.B., Halstead, R., & Trio, G.P. (1996a). Detection of fault-prone software modules during a spiral life cycle. *Proceedings of the International Conference on Software Maintenance*, Monterey, CA (pp. 69-76).

Khoshgoftaar, T.M., Allen, E.B., Kalaichelvan, K., & Goel, N. (1996b). Early quality prediction: A case study in telecommunications. *IEEE Software, 13*(1), 65-71.

Nikora, A.P., Schneidewind, N.F, & Munson, J.C. (1998). *IV&V issues in achieving high reliability and safety in critical control software, final report, 1; Measuring and* *evaluating the software maintenance process and metrics-based software quality control, 2; Measuring defect insertion rates and risk of exposure to residual defects in evolving software systems, 3; and Appendices.* Pasadena, California Jet Propulsion Laboratory, National Aeronautics and Space Administration.

Schneidewind, N.F. (1999). Measuring and evaluating maintenance process using reliability, risk, and test metrics. *IEEE Transactions on Software Engineering, 25*(6), 768-781.

Schneidewind, N.F. (2000). Software quality control and prediction model for maintenance. *Annals of Software Engineering, 9,* 79-101. Baltzer Science Publishers.

KEY TERMS

Failure: The inability of a system or component to perform its required functions within specified performance requirements.

Maintainability: The ease with which a software system or component can be modified to correct faults, improve performance or other attributes, or adapt to a changed environment

Metric: A quantitative measure of the degree to which a system, component, or process possesses a given attribute.

Quality: The degree to which a system, component, or process meets specified requirements.

Reliability: The ability of a system or component to perform its required functions under stated conditions for a specified period of time.

Requirement: A condition or capability needed by a user to solve a problem or achieve an objective.

Risk: The chance of injury, damage, or loss.

Software Reuse in Hypermedia Applications

Roberto Paiano
University of Lecce, Italy

INTRODUCTION

Hypermedia applications were, at the beginning, hand-coded pages with "ad-hoc" links. This production method was acceptable until a few pages had to be produced, but it became rapidly unmanageable when several hundreds of pages with complex interactive objects had to be considered. In particular, two interwoven problems rapidly became relevant: how to ensure the "usability" of modern large hypermedia-applications (Garzotto, Matera & Paolini, 1999), and how to improve the efficiency of its production/maintenance process.

In good hypermedia applications, in fact, the reader should be able to effectively exploit the information contained in the application: that is, he or she should be able to quickly locate the objects of interest, to understand the inner structure of the objects and to easily navigate from one object to another. Several factors concur to the achievement of usability: one of the most important is to have a good structuring of the information objects and a good structuring of the navigation patterns.

BACKGROUND

Several authors have recently proposed the adoption of design models (Garzotto, Mainetti & Paolini, 1995; Isakowitz, Stohr & Balasubramanian, 1995; Schwabe & Rossi, 1995)and design patterns (Rossi, Schwabe & Lyardet, 1999), in order to improve the quality of hypermedia applications, at least for those aspects concerning structure and navigation. Other authors (Conallen, 1999; Schwabe & Rossi, 2000) have proposed the use of object oriented paradigm to model this kind of application, but the navigation structures are more simple. Design models provide, in fact, the primitives that allow structuring the information objects and the corresponding navigation patterns along regular and systematic features, improving consistency, predictability (for the user), robustness of the design, and therefore improving usability. The ancestor of these models can be traced to HDM (Garzotto, Paolini & Schwabe, 1993) and its evolution: W2000 Model (Baresi, Garzotto & Paolini, 2000).

The adoption of W2000 to design the internal structure and the navigational features of hypermedia applications is desirable for three reasons:

- resulting applications are usable;
- the production process can be decomposed into sub-problems easy to manage;
- the application model can be "executed" by a suitable "interpreter" to create the application pages in a way that is independent from the specific application.

Furthermore, in several real-life projects we encountered the problem of dealing with application families. An application family is a set of applications sharing (part of) the content and also (part of the) conceptual design. The problem of application families is the typical situation where the application owner, after a successful first application, needs a second one very similar to the first one. At first it seems a simple problem of "reuse" of content: use the same pictures, use the same texts, use the same data, and so forth (Garzotto, Mainetti & Paolini, 1996). After a while it becomes apparent that not only content, but also (pieces of the) conceptual structure must be "reused". Then comes a third "similar" application, and so on. So, the truth emerges: the designer should have started from the beginning having in mind a family of applications, knowing that several specific applications could have resulted from it. In other words the designer should have optimized the activity of "carving out," from a family, a specific application, for a specific need.

Therefore it became clear that the design process, the design model and the design support system should adopt the notion of family of applications. Such kind of activity is made easily using a structured model.

BRIEF DESCRIPTION OF W2000 METHODOLOGY

The methodology was developed by the UWA Consortium (UWA), and specifically by Polytechnic of Milan.

W2000 methodology assumes that it is essential to make a clear distinction between the different aspects of the application that need to be observed during the design phase, in order to make the design itself a structured and easily controllable process, and to obtain clear modeling, suitable for different users and delivery devices.

After the Requirements Analysis phase, guided by a goal-oriented approach, the methodology suggests a sequence of steps that may be briefly summarized as follows:

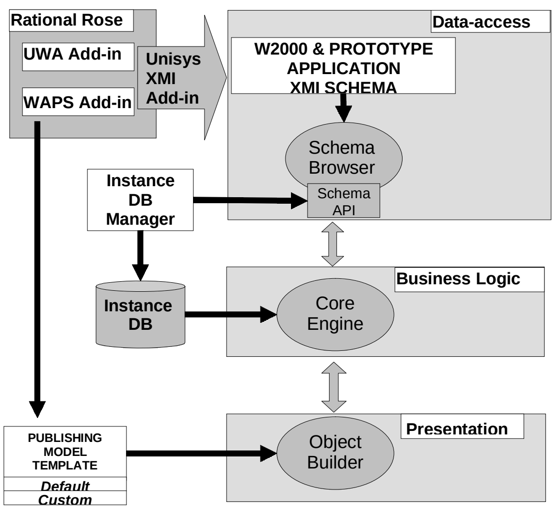
- Information Design: the goal is to describe the information that the application is going to deal with, giving it a structured organization from the user's point of view.
- Navigation Design: this reconsiders the information and its organization from the viewpoint of its fruition, defining the navigational paths the user can follow.
- Publishing Design: the results of the previous steps must be complemented with considerations on presentation and organized into "pages" and "fruition units".

According to the previous description, a database can store the application components described by the model, and then a run-time engine can extract those components from the database to display it to the reader. This kind of engine, named WAPS, is application independent, so it is really reusable and it may be defined as a W2000 methodology Interpreter. It is the last evolution of a family of navigation engines according to the evolution both of methodology (Bochicchio & Paiano, 2000; Paiano & Pandurino, 2003) and available technology (Bochicchio & Paolini, 1998; Bochicchio, Paiano & Paolini, 1999).

A REUSABLE INTERPRETER

The run-time environment, the WAPS core, has the main task of creating a mock-up application starting from the W2000 model in XMI format.

Figure 1. WAPS architecture



In accordance with the modular structure of the W2000 methodology and the various aspects of WA, as shown in Figure 1, it is possible to identify a clear n-tier architecture for the WAPS run-time environment. This choice was highly suitable for the W2000 methodology: in order to manage the complexity, each architecture layer manages a single aspect and provides services to the other levels. All the data managed in the modules are in XML format; furthermore, all interaction between modules is in the same format according to the market trend and standard.

It allows the use of transformation parsing techniques like XSL in the visualization and processing phases, also allowing the following of the evolution of methodology.

- Rational Rose Add-in: Rose helps to design the WA in graphic format using standard UML notation, in accordance with W2000 methodology, in order to obtain a "machine readable" description. Another rational rose add-in: "Unisys Rose XML Tools," produced by Unisys, exports the UML diagram into a standard XMI output.
- Schema Browser: This module allows a unique entry point to the WA schema, hiding the complexity in order to manage the XMI in raw mode. The module provides a set of schema APIs (S-API) to navigate through the WA model via W2000 primitives, bypassing the UML MOF used by XMI.
- Core Engine: This module corresponds to the business level for a three-tier application. This module has the task of understanding the requests from the Object Builder, using the S-API of the schema browser to compose the reply schema that will contain the application data taken from the Instance DB. Since this module creates the reply schema, all design customizations take effect at this stage.
- Object Builder: This module is the door to WAPS systems: the user request comes in, the prototyped page goes out. The module moves the request to the Core Engine and receives the response in XML-like format. Its main task is to apply a template to make the page visible. WAPS uses the XSLT transformation to obtain HTML or WML pages.
- Instance DB: It is an E-R database that contains the data that will be shown to the user. The E-R schema is derived from the W2000 model; thus the schema is fixed and does not change with the domain of the WA being prototyped.
- Publishing Model Template: It is an E-R database that contains the references to the visualization template to create the page.

FUTURE TRENDS

We are now assisting the revival of MVC (Model View Controller) architecture applied to the Web. There are several frameworks oriented to the prototyping of Web applications, based on the MVC paradigm, but the Model site is not fully specified.

The next step of research in this area is to join the W2000 Model definitions with the Model site of these frameworks.

To achieve this goal, the SET Lab of University of Lecce is developing this kind of interface between W2000 model and Struts framework developed by Apache software Foundation, obtaining very interesting results.

CONCLUSION

The adoption of structured approaches and conceptual models, such as W2000, is an important step in the direction to improve the quality and the reusability of hypermedia applications, reducing at the same time the costs and the time required for their development.

The construction of a reusable engine for hypermedia application, based on the W2000 model, is the logical extension to the model-based approach. The reusable engine, in fact, is able to implement the needed hypermedia software in an application-independent way. The effectiveness of this approach is based on the ability of the model to describe a wide range of complex hypermedia applications.

REFERENCES

Baresi, F., Garzotto, F., & Paolini, P. (2000). *From Web sites to Web applications: New issues for conceptual modeling.* WWW Conceptual Modeling Conference.

Bochicchio, M.A., & Paolini, P. (1998). An HDM interpreter for on-line tutorials. In N. Magnenat-Thalmann & D. Thalman (Eds.), MultiMedia modeling (pp. 184-190). Los Alamitos: IEEE Computer Society.

Bochicchio, M.A., & Paiano, R. (2000). Prototyping Web applications. *ACM Symposium on Applied Computing, Como, (IT),* 978-983.

Bochicchio, M.A., Paiano, R., & Paolini, P. (1999). JWeb: An HDM environment for fast development of Web applications. *IEEE Conference on Multimedia Computing and Systems, Firenze (IT), 2,* 809-813.

Conallen, J. (1999). Modelling Web application achitectures with UML. *Communication of the ACM, 42,* 63-70.

Garzotto, F., Mainetti, L., & Paolini, P. (1995). Hypermedia application design: A structured approach. In J.W.Schuler, N.Hannemann & N. Streitz (Eds.), *Designing user interfaces for hypermedia.* Springer Verlag.

Garzotto, F., Mainetti, L., & Paolini, P. (1996). Information reuse in hypermedia applications. *ACM International Conference on Hypermedia.* Boston: ACM Press.

Garzotto, F., Matera, M., & Paolini, P. (1999). *Abstract tasks for hypermedia usability evaluation.* Technical Report No.03-99. Dept. of Electronics and Information, Polytechnic of Milan.

Garzotto, F., Paolini, P., & Schwabe, D. (1993). HDM: A model based approach to hypermedia application design. *ACM Transactions on Information Systems, 11*(1), 1-26.

Isakowitz, T., Stohr, E.A., & Balasubramanian, P. (1995). RMM: A methodology for structured hypermedia design. *Communications of the ACM, 38*(8), 33-44.

Paiano, R., & Pandurino, A. (2003). From the design to the development: A W2000 based framework, issues and guidelines. *IRMA International Conference,* Philadelphia (pp. 500-503).

Rossi, G., Schwabe, D., & Lyardet, F. (1999). Improving Web information systems with design patterns. *International WWW Conference.* Toronto: Elsevier Science.

Schwabe, D., & Rossi, G. (1995). The object-oriented hypermedia design model. *Communications of the ACM, 38*(8), 45-46.

Schwabe, D., & Rossi, G. (2000). An object oriented approach to Web-based application design. *http://www.telemidia.puc-rio.br/oohdm/oohdm.htm*

UWA Consortium. (2001). *General definition of the UWA framework.* Technical report EC IST UWA Project. *www.uwaproject.org*

KEY TERMS

HDM: Hypermedia Design Methodology, developed by Polytechnic of Milan (Italy). It is a methodology to design hypermedia applications.

Interpreter: The traditional definition in computer science is a program that translates and executes source language statements one line at a time. The meaning in this article is: a program that accesses the model to understand the requested navigation one user action at a time.

W2000: A methodology to conceptually design Web applications, developed by Polytechnic of Milan (Italy) in the UWA project.

WAPS: A reusable W2000 interpreter to prototype Web applications.

Web Application: An application that presents the characteristics and the issues of both hypermedia applications and traditional applications. In other words, this kind of application has the navigational issues of Web sites joint to the traditional operation issues.

XMI: XML Metadata Interchange is a widely used interchange format for sharing objects using XML. It is defined by OMG.

XML: EXtensible Markup Language is a mark-up language much like HTML designed to describe data using your own tags. It is a recommendation of W3C Consortium.

Space Opera–GIS Basics

Paul H.J. Hendriks
Radboud University Nijmegen, The Netherlands

INTRODUCTION

The term "Geographical Information Systems", commonly abbreviated to GIS, is an umbrella term covering a class of computer based information systems that are typified by their focus on geographical or spatial data and information. The basic notion underlying a GIS is that every object present on the earth can be "geo-referenced", which refers to defining the spatial location of objects by referencing systems (e.g. latitude/longitude or postal zones). The key to specific GIS functionality is that GIS allow connection of geo-referenced data to attribute or non-spatial data of geographical objects. In varying ways and to varying degrees GIS facilitate the steps necessary for acquiring both classes of data and turning them into geographical information, including input (e.g. via GPS or scanning of maps), data processing and analysis (e.g. overlay analysis or shortest path calculation) and display (e.g., on maps).

The history of GIS spans roughly four decades, bringing a story that springs from many origins, and mingles many disciplines. GIS finds its roots in public administration domains and military applications, but has fanned out to many commercial, non-profit and academic areas. From its origins in the nineteen sixties, GIS has grown in size and stature, building on diverse influences concerning concepts and principles, data and issues of spatial infrastructure, software and software vendors, application areas, etc. The GIS landscape contains a multitude of proprietary and public domain GIS software packages (most notably Autodesk's AutoCAD Map, Erdas's Imagine, ESRI's Arc/Info, Intergraph's GeoMedia, Maplnfo and Smallworld). From its ongoing development history GIS emerges not as one sharply delineated concept or class of computer systems. The GIS landscape is as diverse as the multitude of roots from which it originated and the throng of influences that shaped its development.

BACKGROUND

GIS as an umbrella term covers various types of technology, ranging from relatively simple mapping facilities to advanced environments supporting spatial analysis. What combines different manifestations of GIS is that they all draw their operability from functionalities around spatial information, or geoinformation (Goodchild, 2003; Konecny, 2003). Developing an understanding of what distinguishes GIS from other information systems can be achieved by looking at GIS through four windows. Firstly, GIS can be defined through a *database window*, which involves exploring the alternative data models used in GIS and the principles and technical issues involved in defining and linking spatial and attribute data (including temporal data) based on these models. Secondly, an important distinction between different GIS refers to the types of spatial and non-spatial analysis they support, which involves looking at these systems through a *geographical analysis window*. Thirdly, applying a *visualization window* allows to assess the functionality of GIS for making the contents of their database and analysis outcomes visible, in which maps and computer cartography play an important role. Fourthly, GIS can be looked at through a *relevance window*. This fourth perspective, which involves looking at the context of system usage, is crucial for understanding GIS as information systems instead of data systems, as it should allow to understand how and when geographical analysis may make geographical data meaningful, how and when their presentation on maps will be useful, and how and when specific functionalities of GIS are necessary elements for defining GIS.

A DATABASE WINDOW ON GIS

Looking at GIS through a *spatial database* window shows GIS as a dedicated system for acquiring, storing and processing data with a spatial component (Haining, 2003; Shekhar & Chawla, 2003). Building a spatial database requires spatial data modeling and processing. Spatial modeling involves several steps. Firstly, it concerns developing a spatial understanding by distinguishing and relating the elements that define "spatiality", such as "location" or "distance". Secondly, it involves building a *conceptual spatial data model* by translating the spatial perspective into formalized data elements. Thirdly, a *logical spatial data model* must be built, which involves representing and formalizing the elements of the conceptual model in a univocal manner so that these can be entered into an automated system. Fourthly, the logical spatial data model has to be translated into the actual data storage, which refers to choosing a *physical data model*.

As the first two steps most specifically enter the spatial perspective in data modeling, we will focus on these.

Building a conceptual spatial data model consists of identifying spatial objects and characterizing these. Five aspects describe spatial objects:

1. These are objects that are in a certain place or location (aspect: *where* or *absolute location*),
2. with a certain spatial size, appearance, etc. (aspect: *spatial form*),
3. at a certain distance from other spatial objects and in a certain relationship with these (aspect: *spatial relationships* or *relative location*),
4. at a certain moment in time or during a certain period, and possibly subject to change (aspect: *when*), and
5. with other characteristics than spatial and temporal (aspect: *what*).

A basic distinction between conceptual spatial data models is whether they represent spatial objects in 2D or 3D. Commonly, four basic types of two-dimensional spatial objects are distinguished: points, lines (or specific collections of points), areas (objects identified by lines beginning and ending in the same point) and surfaces. Combinations of these basic classes and their characteristics may produce new types of spatial objects, such as spatial patterns, spatial structures, or spatial networks. For instance, a spatial network such as a road network or a network of pipelines, results from understanding points as nodes and lines as connections between nodes, and may be expanded to include areas and surfaces as well. Data models for representing these objects can be distinguished into raster models (based on fixed units of space) and vector models (based on combinations of points and lines). Three-dimensional spatial data models or geomorphometrical models, which are less common in GIS, aim to represent the height dimension of space in a more sophisticated form than by using contour lines. The most commonly used models are the triangulated network (TIN) and the digital elevation model (DEM).

Connecting the five aspects specified above with spatial objects involves three issues: firstly, defining the spatial data (aspects 1, 2, 3); secondly, defining the non-spatial data (aspects 4 and 5); and thirdly, linking spatial and non-spatial data. The non-spatial are typically called the attribute data in GIS, which include temporal data (see Christakos, Bogaert, & Serre, 2001). Describing absolute location (aspect 2) is called georeferencing. A basis distinction is between continuous (e.g., LATLONG) and discrete (e.g., postal zones) georeferencing.

Working with spatial data in GIS presumes a choice of spatial data model, facilities for getting data into the system based on that data model, and functions for processing data. Some GIS have chosen one spatial data model as the basis of their functionality for handling spatial (e.g., leading to a distinction between raster and vector GIS); others offer combinations of models and facilities for linking these. As to data input GIS users no longer need to collect all data themselves because many commercial datasets are available (e.g., based on postal zones or road networks), as well as spatial data sets built in academic or other public domain projects (see Rigaux, Scholl, & Voisard, 2002; Walford, 2002). Data input has been revolutionized by the Global Positioning System (GPS), which allows inputting both 2D and 3D spatial data from satellite data (e.g., Konecny, 2003). GIS typically contain functions for additional processing of data input from external data sets or from input devices (GPS, digitizers, etc.), such as geometric transformation (translating among methods of map projection), edge matching (ensuring that features crossing adjacent map sheets have the same edge locations, attribute descriptions and feature classes), and standard attribute editing functions (e.g., reclassifying objects based on their attribute data).

A GEOVISUALIZATION WINDOW ON GIS

The window of *visualizing spatial information* shows GIS as a system to present spatial data and information, e.g. via maps (Cartwright & Hunter, 2001). Geovisualization, and therefore also the use of maps in GIS, can refer to GIS output or input. Geovisualization as input refers to taking the visual form as the starting point for communication with GIS or further analysis steps within or outside of GIS (Maceachren & Brewer, 2004). For instance, maps can be used as interactive devices for defining queries for showing and exploring the contents of the database, e.g., to detect errors (Andrienko, Andrienko, & Gitis, 2003), or for starting a new geographical analysis. Using maps in GIS in an interactive sense presumes that the map is linked with a set of controls (e.g., it can be made clickable) allowing to change the content of the map or to take the content of the map as a step toward building a new map (e.g., Cartwright & Hunter, 2001).

Geovisualization as output refers to showing the GIS's contents, such as the data contained in the spatial databank, or the outcomes of spatial analysis. GIS can be used for producing, showing and interpreting maps and other visualizations. Producing maps involves generating data with a spatial footprint, and showing these in a new or updated map. Producers of atlases and road maps use GIS to facilitate updating the frequent changes in the location of objects (e.g., changes in the street networks, or the usually less frequent changes in the boundaries between administrative units). Geographical analysts that perform their analysis in analysis modules linked to GIS and show the analysis outcomes in GIS are another class of GIS map producers.

AN ANALYSIS WINDOW ON GIS

Through a *spatial analysis window* GIS appears as a system that offers possibilities for processing spatial and non-spatial data so that new spatial and non-spatial data and information can be inferred. GIS are best understood as toolboxes of operations that may be combined in multiple ways, guided by specific analysis needs. Additionally some GIS offer a (mostly modestly filled) spatial model base coupled to standard spatial analysis functions or offer functions for linking to external model bases (DeMers, 2002; Haining, 2003).

Typical analysis-related functions in GIS include retrieval, classification, overlay, neighborhood and connectivity operations. Such GIS functions may lead to GIS-based spatial analysis when applied and combined in different ways. Firstly, the analysis may involve generating new attribute data from existing attribute data of known spatial objects, while not affecting the location data of these objects. An example that fits within this class of spatial analysis is the overlay of two maps (e.g., private-public ownership and land use classes of lots) to produce a characterization of the areal units on the map in a third map (e.g., conveying an insight into the connections between ownership and land use of lots).

Secondly, GIS may support forms of analysis that deduce attribute properties from location data of known spatial objects. As an example consider combining network and aspect operations to assess whether a particular location is likely to run the risk of pollution in the case of a toxic spillover, because of how the downhill path runs from the spillover location (e.g., Rodriguez-Bachiller & Glasson, 2004).

Thirdly, analyzing attribute data of spatial phenomena may form the basis of the detection and subsequent characterization of new spatial objects. As an example consider a shop that is interested in what typifies its service area, to assess possibilities for adapting its range of goods. GIS allows to establish this neighborhood if customer trip data are available and to analyze prevailing demographic properties of residents within the window by connecting the spatial object with external attribute data as are available for postcode zones (e.g., Kidner & Higgs, 2003; Miller & Shaw, 2001).

Fourthly, spatial analysis may involve detecting new spatial characteristics of objects, by examining their spatial properties alone or in conjunction with associated attribute data. A GIS analyst may, for instance, perform a combined contiguity and network analysis based on address and routing information to determine how close residents are to a service such as an ambulance post or community service, to define the catchment areas of these services (e.g., Kidner & Higgs, 2003; Miller & Shaw, 2001).

A RELEVANCE WINDOW ON GIS

The *relevance window* concerns the objective for which the information stored in GIS and generated by GIS is used. A GIS is more than a collection of hardware and software. It is a specific aspect system of a larger system, namely the organization within which GIS data is managed and processed.

A view through this fourth window shows the classes of GIS applications and users (individuals, groups, departments, organizations, networks, etc.). Looking at typical GIS applications reveals a broad spectrum of uses and users (e.g., see Geertman & Stillwell, 2003; Grimshaw, 2000; Kidner & Higgs, 2003). Important business users of GIS include owners of large physical distribution networks, such as utility companies, for example gas, phone, electric, water, cable TV companies, the telecommunication industry, the transportation sector (including government transportation departments who use GIS, for instance, to store information on the state of pavement on the highway network, maintain an inventory of all highway signs, and analyze data on accidents). This sector also includes delivery companies that use GIS technology to keep track of shipments, know where they are, and plan efficient delivery routes (e.g., Miller & Shaw, 2001). GIS is also being used for siting facilities such as retail shops and warehouses, by analyzing demographic data of customer populations, access routes for customers, delivery routes for suppliers, etc. Large-scale GIS users also include farmers and foresters who increasingly use detailed maps for planning, analyzing and evaluating land use.

If we try to define GIS through a relevance window in more general terms their multicolored nature becomes even more apparent. GIS can be defined as dedicated spatial object and spatial network management tools as is customary in the AM/FM (automated mapping/facilities management) domains. GIS can be looked at as spatial decision support systems, when we see them as tools for supporting decision makers making choices that involve spatial elements, such as solving routing problems, connected location-allocation problems, or problems of service area development (e.g., Clarke & Madden, 2001; Grimshaw, 2000). GIS appear as organizational learning systems when their presence in organizations and functionalities are assessed from the perspective as to whether and when using spatial categories supported through GIS makes the organization smarter

(Hendriks, 2000, 2004). GIS can be tools in the hands of spatial scientists, including geographers, physical planners, forestry students, environmental students, and geologists, that support them in their research activities.

FUTURE TRENDS

In recent decades several developments have resulted in moving GIS from the outskirts of eclectic computing by geographical pioneers to mainstream data and information processing in private and public companies, in academia and even in private households (e.g., car navigation systems or route planners on the Internet). Many trends have supported this move. More data is available for GIS, via commercial datasets with spatial footprints and via new data collection techniques. Among these, Global Positioning System (GPS) has become a major source of new GIS data, and new GIS data comes increasingly from integrated GPS/GIS systems. The mobility of portable GIS and GPS systems has revolutionized GIS use. As GIS have been recognized for their integration facilities for databases of different kinds, GIS has quickly incorporated distributed systems and databases. Improvements in the software installation facilities user interface, improved interoperability of proprietary GIS packages, etc. have substantially altered how GIS “look and feel”. National geographical data infrastructures are being built and expanded, combining the datasets and knowledge of various agencies involved in spatial data handling (e.g., land register offices, spatial planning departments). GIS education has done its part in lowering some thresholds for recognizing the potential value of GIS, and in raising the critical awareness using GIS without profound geographical insight has limited value.

It can be expected that new developments will build on such existing trends. In the short run the further integration with GPS technology, and the development of national and transnational spatial infrastructures may be expected. In the longer run we may anticipate an increasing integration of scientific visualization and computer graphics with GIS capabilities. An increased integration of GIS with advanced modules for geographical analysis can also be foreseen. The use of animated and interactive maps, whether delivered via the Internet or not, holds much promise for enhancing the approachability of such spatial analysis modules.

CONCLUSION

Combining the four windows specified above – or adopting a more GIS-like terminology: overlaying these windows – allows drawing the contours of what GIS are. The four windows show Geographical Information Systems as sets of software tools and associated guidance principles that are based on insights as to how using spatial categories can further our understanding of phenomena and that allow the further development of such a spatial understanding through instantiations of combined spatial data, analysis and visualization models. History has moved GIS from the backwaters of pioneering academics and some individual public institutions into the floodlights. The story of GIS is an ongoing story of success, notwithstanding the critical comments GIS has received and is receiving (e.g., Sheppard, Couclelis, Graham, Harrington, & Onsrud, 1999). It presents GIS viewers with a space opera in which the acts they have seen so far have done everything to arouse the attention of what the future will hold for GIS and its users.

REFERENCES

Andrienko, G., Andrienko, N., & Gitis, V. (2003). Interactive maps for visual exploration of grid and vector geodata. *ISPRS Journal of Photogrammetry and Remote Sensing, 57*(5-6), 380-389.

Cartwright, W. E., & Hunter, G. J. (2001). Towards a methodology for the evaluation of multimedia geographical information products. *Geoinformatica, 5*(3), 291-315.

Christakos, G., Bogaert, P., & Serre, M. L. (2001). *Temporal GIS: Advanced functions for field-based applications*. Berlin: Springer.

Clarke, G., & Madden, M. (2001). *Regional science in business*. Berlin: Springer.

DeMers, M. N. (2002). *GIS modeling in raster*. New York: Wiley.

Geertman, S., & Stillwell, J. (2003). *Planning support systems in practice*. Berlin: Springer.

Goodchild, M. F. (2003). The nature and value of geographic information. In M. Duckham, M. F. Goodchild & M. F. Worboys (Eds.), *Foundations of geographic information science* (pp. 19–32). New York: Taylor and Francis.

Grimshaw, D. J. (2000). *Bringing geographical information systems into business* (2nd ed.). New York: Wiley.

Haining, R. (2003). *Spatial data analysis:Ttheory and practice*. New York: Cambridge University Press.

Hendriks, P. H. J. (2000). An organizational learning perspective on GIS. *International Journal of Geographical Information Science, 14*(4), 373-396.

Hendriks, P. H. J. (2004). Intelligence from space; Using geographical information systems for competitive intelligence. In D. J. Vriens (Ed.), *Information and communications technology for competitive intelligence* (pp. 194-226). Hershey, PA: Idea Group Publishing.

Kidner, D., & Higgs, G. (2003). *Socio-economic applications of geographic information science*. London: Taylor & Francis.

Konecny, G. (2003). *Geoinformation: remote sensing, photogrammetry and geographic information systems*. London: Taylor & Francis.

Maceachren, A. M., & Brewer, I. (2004). Developing a conceptual framework for visually-enabled geocollaboration. *International Journal of Geographical Information Science, 18*(1), 1-34.

Miller, H. J., & Shaw, S.-L. (2001). *Geographic information systems for transportation: principles and applications*. New York: Oxford University Press.

Rigaux, P., Scholl, M., & Voisard, A. (2002). *Spatial databases: With application to GIS*. San Francisco, CA: Morgan Kaufmann.

Rodriguez-Bachiller, A., & Glasson, J. (2004). *Expert systems and geographic information systems for impact assessment*. London: Taylor & Francis.

Shekhar, S., & Chawla, S. (2003). *Spatial database: A tour*. Upper Saddle River, NJ: Prentice Hall.

Sheppard, E., Couclelis, H., Graham, S., Harrington, J. W., & Onsrud, H. (1999). Geographies of the information society. *International Journal of Geographical Information Science, 13*(8), 797-823.

Walford, N. (2002). *Geographical data characteristics and sources*. Chichester, UK: Wiley.

KEY TERMS

Geographical Analysis Model: A model to derive new spatial information by combining existing spatial and attribute data sources as appears useful for a given research goal (e.g., route selection, service area definition).

Geographical Information System (GIS): Computer system for acquiring, storing, processing, analyzing and displaying combined spatial and non-spatial data for academic, societal or organizational purposes.

Georeferencing: Identifying the geographic location of features and their boundaries on the Earth surface, e.g., derived from GPS, remote sensing, mapping, and surveying technologies.

Geovisualization: Data representation using more than one dimension to convey meaning, including animation and dynamic cartography, and spatialization (using geographic metaphors to generate georeferenced displays of nonspatial data such as text, image or sound archives).

Global Positioning System (GPS): A system of determining the absolute location of features by calculating x, y, z coordinates of ground locations from the signal of satellites orbiting the Earth.

Spatial Data Model: Set of principles and concepts for describing spatial objects and associated characteristics in a formal, univocal way, at the conceptual, logical or physical level.

Spatial Decision Support System: System built on the integration of spatial database management systems, geo-analytical models, geovisualization models, and the expert knowledge of individuals and organizations facing spatial decision problems.

Spatial Analysis in a Public Health Setting

Robert Lipton
Prevention Research Center, USA

D.M. Gorman
The Texas A&M University System Health Science Center, USA

William F. Wieczorek
Buffalo State College-State University of New York, USA

Paul Gruenewald
Prevention Research Center, USA

INTRODUCTION

Spatial methods are an increasingly important tool in understanding public health issues. Spatial analysis addresses an often forgotten or misunderstood aspect of public health, namely, studying the dynamics of people in places. As advances in computer technology have continued apace, spatial methods have become an appealing way to understand the manner in which the individual relates to larger frameworks that compose the human community and the physical nature of human environments (streets with intersections, dense vs. sparse neighborhoods, high or low densities of liquor stores or restaurants, etc.). Spatial methods are extremely data-intensive, often pulling together information from disparate sources that have been collected for other purposes, such as research, business practice, governmental policy, and law enforcement. Although initially more demanding in regard to data manipulation compared to typical population level methods, the ability to compile and compare data in a spatial framework provides much additional information about human populations that lies beyond typical survey or census research. We will discuss general methods of spatial analysis and mapping which will help to elucidate when and how spatial analysis might be used in a public health setting. Further, we will discuss a practical research example focusing on the relationship between alcohol and violence.

BACKGROUND: GIS CAPABILITIES AND PREVENTION

A basic understanding of the capabilities of geographic information systems (GIS) is critical to the development of prevention activities because alcohol-related problems are not evenly distributed across space. GIS can be defined as a combination of computer hardware, software, spatial data (digital maps), and data with a geographic reference (e.g., alcohol outlets, crime locations) that facilitates spatial analysis. The key functions of GIS provide access to the broad spectrum of potential spatial analyses that can support the simple targeting of resources as well as the development of more complex models of spatial interactions. Both simple maps of problem rates or clusters and spatial interaction models may be useful for targeting traditional individual-based prevention programs or environmental interventions. Spatial interaction models, however, may be more appropriate for identifying the locations of events (e.g., assaults or crashes) that may be most amenable to environmental or regulatory prevention. In addition, GIS capabilities promote the development of a basic spatial/geographic epidemiology of alcohol use and related consequences, which is critical to the development of prevention programs (see Wieczorek, 2000, and Wieczorek & Hanson, 1997, for more details).

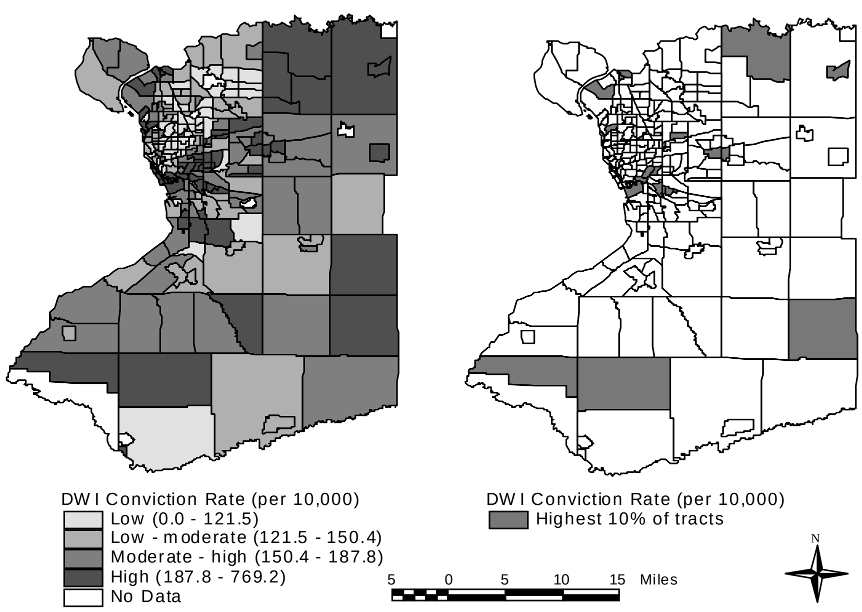
The key functions of GIS include: geocoding, data overlays, reclassification functions, and distance/adjacency measures. Geocoding is a generic term used to describe the GIS function of providing a specific location to descriptive data. Geocoding applies to point data (e.g., alcohol outlet) as well as to area data (e.g., number of assaults in a census tract). Sometimes geocoding is known as address matching because the process of matching points to addresses is very common. The advent of the Census Bureau's TIGER system has made geocoding a relatively low cost and widely available GIS function. However, professional geocoding services have developed to assist persons who are not comfortable in geocoding their own data or because of the high cost of updating digital maps based on TIGER in areas of changing population. Geocoding is the most basic of GIS functions because it transforms descriptive information into a format suitable for spatial analysis.

A GIS-based map may consist of multiple sources of data. The ability to combine multiple layers of information is known as the overlay function. An example of an overlay function is to place geographic boundaries (such as the outline of a town) on top of individual points (such as residences of DWI offenders). The points within each area can then be automatically counted to create rate-based maps such as those shown in Figure 1. To create rate-based maps from relevant point information, at least three layers of data are necessary (i.e., map of the points, a map with relevant boundaries, and census data on population). The ability to perform an intersection between separate maps, to aggregate data into meaningful geographic areas, and to link data to standard sources, such as census data, highlights some of the processing capabilities of GIS overlay functions, processes nearly impossible to accomplish by non-automated methods (see Wieczorek & Hanson, 2000, for an example using regions and mortality data).

One major contribution of GIS to prevention is its ability to provide useful visualizations of spatial data. The reclassification function of GIS allows the user to easily manipulate the number of categories or select specific information (e.g., crashes by time of day or day of the week) for display. Figure 1 shows how the reclassification function can assist in the targeting of prevention by reclassifying the same data to emphasize highest rate areas.

A second major contribution of GIS to prevention is that the technology enables the development of models of spatial patterns and interactions within and between populations and environments. These models require accurate information on the distance between individual objects (e.g., bars and traffic crashes) and their spatial relationships to one another. Distance and adjacency functions of GIS allow assessments of these relationships. Data generated from these assessments of spatial relationships can be exported from the GIS and used in spatial modeling software. Information about adjacencies of different geographic objects can be used to assess contributions of environmental features (e.g., bars) to problematic public health outcomes in surrounding areas (e.g., assaults) (Gruenewald et al., 1996; Lipton & Gruenewald, 2002). Other important GIS functions are based upon the assessment of distance relationships: neighborhood functions calculate the number of a specific characteristic (e.g., assaults) within a specific radial distance (e.g., 300 yards) of point features (e.g., bars). Buffer functions use the distance function on a complex feature such as the road network to identify points within a set distance of the feature (e.g., homes of DWI offenders within 400 yards of a bus line). These GIS functions can also be combined in complex ways to provide new insights for targeting prevention activities to areas with the greatest need (see Harding & Wittman, 1995, for additional applications in support of prevention).

Figure 1. Reclassification and targeting prevention



Spatial Clusters

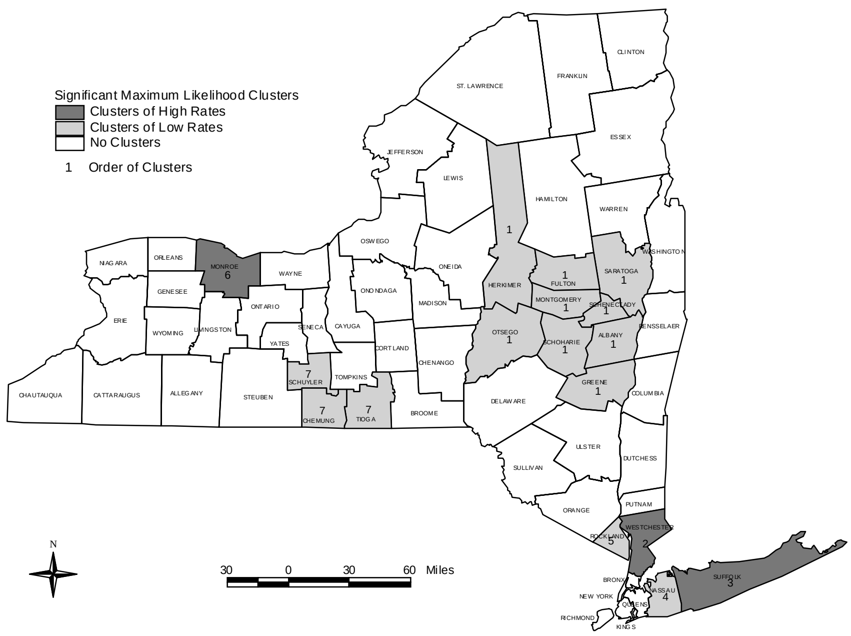
Spatial clusters are a greater than expected geographically close group of occurrences or events (e.g., deaths, crashes, alcohol outlets). Spatial clusters are a natural result of spatial dependencies in the data; by definition, spatially dependent data will have an uneven geographic distribution. The use of spatial cluster analysis was pioneered for finding cancer clusters, especially for rare cancers (Aldrich, 1990; Rickett, Savitz, Gesler, & Osborne, 1994). Specific spatial clustering techniques can be used with point or geographic area data and may also be used for space-time cluster analysis to examine temporal trends (Jacquez, 1994). Spatial cluster analysis is useful for identifying areas with significantly high or low rates of alcohol problems where services can be targeted, to identify new research questions (e.g, why are rates highest in certain areas), to empirically identify the appropriate scale of analysis in small area studies, and to examine the impact of interventions on communities over time (e.g., do the clusters change or disappear in response to interventions).

It is important to note that low rate clusters are as important to identify for prevention purposes as high rate clusters. High rate clusters clearly have problems that require prevention/intervention; however, low rate clusters may be areas at high risk of developing problems, especially if the low rate area is embedded in surrounding high rate areas. Characteristics of low rate areas may also provide important insights into factors important for prevention application in high problem locations. Figure 2 shows a map of spatial scan clusters of alcohol-related mortality in New York. Note that both high and low rate clusters are identified. The analysis of county-level data also shows the potential for regional level prevention approaches.

Other Spatial Analytic Techniques

Two additional approaches deserve mention in the context of spatial analysis for prevention. The first technique is a relatively simple method to control spatial autocorrelation in multiple regression analysis. Spatial autocorrelation, correlated measurement error between spatially adjacent units, is a substantial source of statistical bias in these analyses. The method is to use a GIS to calculate a generalized spatial potential for the dependent variable used in multiple regression of geographic area data (e.g., census tracts or zip code areas). Wieczorek and Coyle (1998) provide an example of this technique in the context of targeting the neighborhoods of DWI offenders. A generalized spatial potential (GSP) for the DWI rate was calculated for each tract by summing the ratio of DWI rates (V) and distances (D) to every other tract ($GSP_i = V_1/D_1 + V_2/D_2 + \ldots V_n/D_n$). By including the GSP as an independent variable in multiple regression, some aspects of biases due to spatial autocorrelation can be controlled (allowing more appropriate interpretations of model coef-

Figure 2. Significant spacial scan clusters of alcohol-explicit mortality



ficients and statistical tests). This approach is not as statistically complete as direct methods for assessing and controlling spatial autocorrelation (see Gruenewald et al., 1996), but it is a substantial improvement that may be implemented relatively easily.

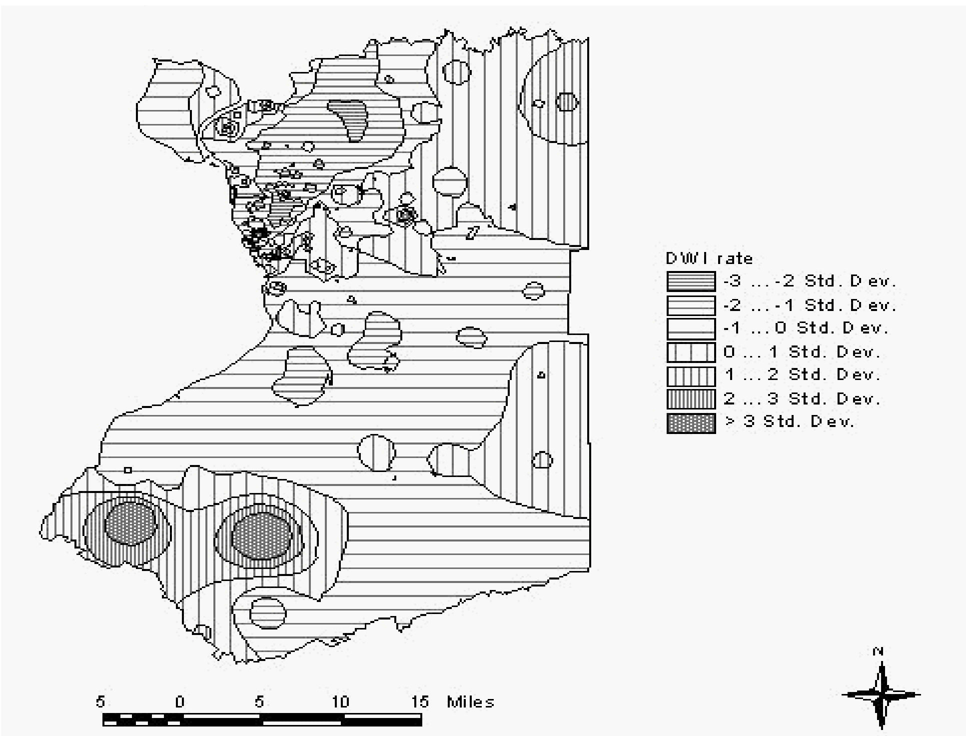
The second technique is the development of continuous surface models, or "kriging." Kriging is a modeling technique for spatial data that can be used to develop contour maps (e.g., maps that show lines of equal value such as DWI rates) from a limited number of assessment points or areas (Isaaks & Srivastava, 1989). These continuous surface models overcome a central limitation of area data; that actual rates within the geographic areas are unlikely to be as uniform as suggested in area maps. Kriging creates a continuous surface model by overlaying a grid of cells over the entire areas and calculating a weighted value for each cell based on the distance to surrounding centroids. The values calculated for the grid are then used to create a contour map. An example of kriging is provided in Wieczorek and Hanson (1997). Figure 3 shows a continuous surface model created by applying kriging to the specific tract rates used to generate Figure 1. A continuous surface model may provide a more realistic version of geographic variation that can be used to target prevention and assist in the overall planning of alcohol-related services.

SPATIAL ANALYSIS: AN EXAMPLE FROM VIOLENCE PREVENTION

Knowing where problem events take place is different from knowing why they take place where they do. And understanding the etiology of public health problems in different geographic areas requires knowing all the spatial technologies discussed to this point and their suitable application in spatial statistical analysis; the use of spatial data for the purpose of explicating the etiological dynamics of public health problems. This section will provide an example of one such application; a first approach to understanding the environmental correlates of violence.

Much of the criminology and public health literature is concerned about the determinants of violence in different community areas. A particular focus has been upon the role of alcohol outlets in violence. Similar to the work of Morenoff, Sampson, and Raudenbush (2001) and Baller, Anselin, Messner, Deane, and Hawkins (2001) our analysis includes demographic and socio-economic data so as to capture violence related to population characteristics (i.e., high unemployment, low rates of high school graduates, etc.). These population characteristics are analyzed in relation to the moderating effects of alcohol outlets on the production of violence. Moderation can simply be

Figure 3. DWI conviction rate continuous surface model



thought of as interaction of outlets with people (with a mix of characteristics). This outlet interaction could serve to increase or decrease violence, depending on the composition of population characteristics and outlet presence and type. Further, we examine spatial components of these moderating effects. Rates of violence may be affected by characteristics of populations living in adjacent areas (Gorman, Speer, Gruenewald, and Labouvie , 2001).

Our analysis assesses whether such spatial relationships exist and controls for spatial autocorrelations that may obscure the relationship between population characteristics and the production of violence. The sample comprised 766 zipcodes from four selected areas of California: Los Angeles, the Bay Area, Sacramento, and the northern section of the state. The first three areas are heavily urbanized while the last is quite rural. The three urban areas are heterogeneous with regard to ethnic, age and socio-economic composition, particularly in relation to the rural area that is more homogeneous in most population level measures. In this research, data are taken from three different sources: census data (1990), using a three-item scale representing concentrated disadvantage, immigrant concentration and residential stability, based on the work of Sampson, Raudenbush, and Earls (1997). Hospital discharge data for 1991, using patient home address contained information on assaults. California state data on alcohol outlets (1991) gave type and address of outlet.

The graphic presented in Figure 4 outlines the general conceptual framework that guides the analysis. We hypothesize that populations of people produce assaults at a given rate that is moderated by population characteristics (e.g., greater in places where there are more young people) and environmental characteristics (e.g., where there are more alcohol outlets). This adjusted rate may be further modified by the numbers and characteristics of nearby populations (not shown). It is assumed that errors in estimation are not independent, but rather are spatially autocorrelated, ρ_s.

In Figure 5, assaults per roadway mile are presented for each of the four regions. The difference in concentrations of assaults is apparent in this map, with greater densities occurring where there are greater densities of population. This is not, however, universally the case. For example, the western region of the Los Angeles basin appears to exhibit relatively greater assaults than expected from the population distribution observed. Overlays of zipcodes are represented on this particular map.

Table 1 and Figure 4 (note numbers in the figure) present the results of an analysis that includes the direct effects of population variables, alcohol outlets, and adjacent population density. Many, but not all, of the estimated coefficients from the model are significant. A direct interpretation of the coefficients of the model suggests that as population density increases, there is a reduction of approximately 2.16 assaults for every 10,000 persons in each zip code area. This rate, however, is that expected for an isolated population living in an area with no bars, restaurants or off-premise alcohol establishments (and rather unrealistically, with values of zero for the measures of population characteristics). This rate is moderated, however, by the presence of large populations in adjacent areas, the densities of restaurants, bars, and off-premise establishments, and local population characteristics. This rate is greater in areas where bar densities are greater and restaurant densities are less. This rate is further greater in areas where social disadvantage is greater, immigrant presence less, and residential stability greater. Finally, the rate at which local population density produces assaults is greater in areas surrounded by larger populations.

The results of this analysis indicate that it is possible to construct a conceptually well-framed spatial analysis of assault rates that explains to a substantial degree

Figure 4. Conceptual model and outline of model 3

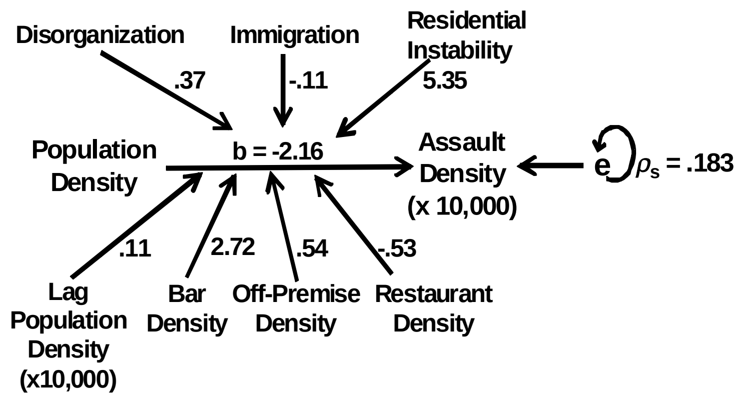
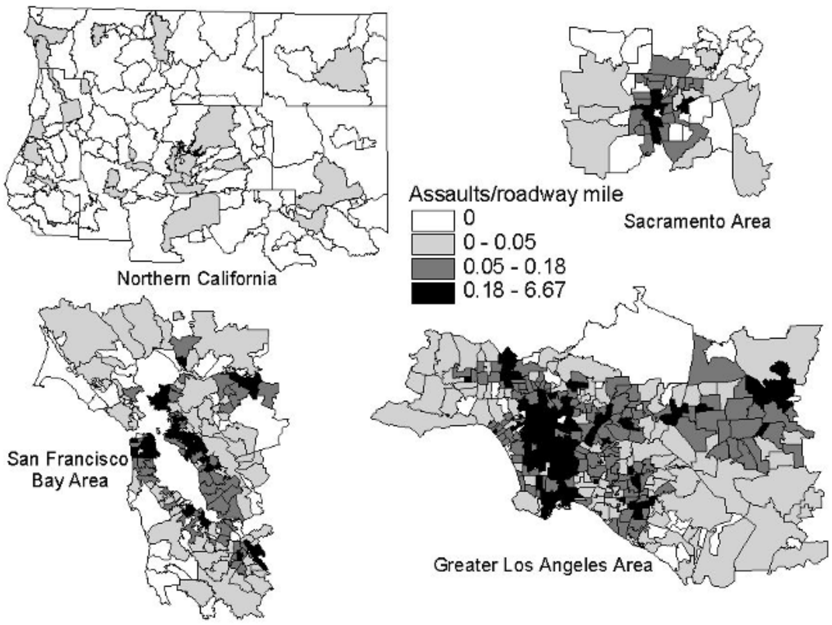


Figure 5. Assaults per mile of roadway



variation in rates of assault between places in California. Shifting from a representation that suggests that outlets on their own create violence to one that presents outlets as providing contexts for violence, the usual pattern of relationships of violence to environmental densities of alcohol outlets continue to be observed. The current analysis suggests, however, that alcohol outlets moderate rates at which violence is produced within areas and that these effects persist when controlling for spatial effects and other covariates related to the production of violence in local populations (Table 1). Notably, positive relationships continue to exist between bar and off-premise outlet densities and assaults, with no relationship to densities found for restaurants (Table 1 and Figure 4). At the geographic scale of the current study (zip codes), greater rates of violence are observed in stable non-immigrant areas with greater concentrated disadvantage. Bar densities, when controlling for other environmental or socio-demographic measures, are clearly connected to an increase in assaults. Thus, in this spatial example, beyond the obvious finding that denser populations have more assaults, we are able to observe important environmental effects that may be actionable in terms of prevention policy.

FUTURE TRENDS

Spatial analysis in public health is in its infancy. This is due to the rapid increase in computing power that has only in the last few years allowed for a more comprehensive development of mapping and spatial statistical techniques. There are two general areas of development that are interrelated: 1) GIS will become more intuitive and popular with more data becoming accessible (e.g., through geo-coding techniques that will become more consistent and "industry standard" through time); 2) Spatial statistical methods will become more available and more user friendly with statistical software companies offering spatial statistical solutions. As the comfort level grows with both mapping and spatial statistics, there will be an increasing demand for data and software that more capably handles research needs. This trend will only accelerate. The ability to add a spatial component to typical population level data opens up an entirely new approach toward thinking about public health issues. The individual will no longer be considered the sole unit of analysis, place will start to constitute its own unit of analysis or as complexly interacting with the individual. Further, as spatial methods improve we will be able to model the dynamic relationship between a target area and areas around the target. The above example is one of the first public health examples of such an approach. In the future, this dynamic modeling will become much more routine and essential to characterizing people in places. In addition, although not given in the above example, we will also have the ability to see how such dynamic spatial systems change over time. There is no methodological or theoretical obstacle to being able to, for example, analyze

Table 1. Associations of outlet densities with rates of assault hospitalizations (x 10.000)

Variable Name	b:	t:	p:*
Population Density	-2.16	02.62	.009
Outlet Densities			
Bars	2.72	4.36	<.001
Off-Premise	.54	1.64	.101
Restaurants	-.53	-3.97	<.001
Population Characteristics			
Social Disadvantage	.37	32.74	<.001
Immigrant Presence	-.11	-15.17	<.001
Resident Stability	5.35	2.41	.016
Adjacent (Lag) Population Density (x 10,000)	.11	2.21	.027
Model based estimate for spatial autocorrelation: $\rho_{s} =$ .183 ** Z = 3.51 P < .001			

how characteristics in one area, such as density of liquor stores, affects violence in an adjacent area and to see how this relationship changes through time.

SUMMARY AND CONCLUSIONS

In this chapter, we have described how spatial methods may be applied in a specific public health area, namely alcohol and alcohol-related problems. Although, an important area of research, it is by no means unique in allowing for spatial/ecological analysis. There are several important factors that should be considered when contemplating public health spatial analysis: 1) are there specific environmental features (such as the presence of bars or liquor stores) that might help explain an outcome (such as violence)? 2) Is there a dynamic relationship between individual behavior and environmental (area) setting? 3) Do environmental factors, such as alcohol outlets modify the relationship between socio-demographic factors and the outcome of interest (e.g., violence)? 4) Is there data available to support a spatial analysis? 5) Is the effect of adjacent areas likely to obscure relationships between exposures and outcomes (spatial autocorrelation)?

When studying alcohol-related problems, spatial analysis allows for the integration of disparate types of information into a meaningful story from both a public health and criminological point of view. The ability to put people in places in more than a purely descriptive framework signals a new generation in research that transcends traditional proscriptions against the use of ecological data. Further, measures of community health such as social cohesion take on a more fully realized form in a spatial analytical context. Indeed, given that most public health and criminological data is collected at a population level, spatial analysis allows researchers to more clearly observe population level effects for whatever measures chosen.

REFERENCES

Aldrich, T.E. (1990). *CLUSTER: User's manual for software to assist with investigations of rare health events.* Atlanta, GA: Agency for Toxic Substances and Disease Registries.

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis, 27*(2), 93-115.

Baller, R.D., Anselin, L., Messner, S.F., Deane, G., & Hawkins, D.F. (2001). Structural covariates of U.S. county homicide rates: incorporating spatial effects. *Criminology, 39*, 561-590.

Gorman, D.M., Speer, P.W., Gruenewald, P.J. & Labouvie, E.W. (2001). Spatial dynamics of alcohol availability,

S

neighborhood structure and violent crime. *Journal of Studies on Alcohol, 63*, 628-636.

Gruenewald, P.J., Millar, A., Treno, A.J., Ponicki, W.R., Yang, Z., & Roeper, P. (1996). The geography of availability and driving after drinking. *Addiction, 91*, 967-983.

Harding, J.R., & Wittmann, F.D. (1995). GIS enhances alcohol/drug prevention planning. *GIS World, 8* (6), 80-83.

Isaaks, E.H., & Srivastava, R.M. (1989). *An introduction to applied geostatistics*. New York: Oxford University Press.

Jacquez, G.M. (1994). *Stat! Statistical software for the clustering of health events*. Ann Arbor: BioMedware.

Lipton, R.I., & Gruenewald, P.J. (2002). The spatial dynamics of violence and alcohol outlets. *Journal of the Study of Alcohol, 63*(2), 187-195.

Morenoff, J.D., Sampson, R.J., & Raudenbush, S.W. (2001). Neighborhood inequality collective efficacy, and the spatial dynamics of urban violence. *Criminology, 39*, 517-560.

Ricketts, T.C., Savitz, L.A., Gesler, W.M., & Osborne, D.N. (Eds.) (1994). *Geographic methods for health services research*. New York: University Press of America.

Sampson, R.J., Raudenbush, S.W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science, 277,* 918-924.

Wieczorek, W.F. (2000). Using geographic information systems for small area analysis. In Wilson, R.E. & Dufour, M.C. (Eds.), *The epidemiology of alcohol problems in small geographic areas* (pp.137-162), NIH National Institute on Alcohol Abuse and Alcoholism: Bethesda (NIH Pub. No. 00-4357).

Wieczorek, W.F., & Coyle, J.J. (1998). Targeting DWI prevention. *Journal of Prevention and Intervention in the Community, 17*(1), 15-30.

Wieczorek, W.F., & Hanson, C.E. (1997). New modeling methods: Geographic information systems and spatial analysis. *Alcohol Health and Research World, 21*(4), 331-339.

Wieczorek, W.F., & Hanson, C.E. (2000). Regional patterns of alcohol-specific mortality in the United States. In R.C. Williams, M.M. Howie, C.V. Lee, & W.D. Henriques (Eds.), *Geographic Information Systems in Public Health: Proceedings of the Third National Conference* (pp. 669-676), Atlanta: Centers for Disease Control and Prevention.

KEY TERMS

Geocoding: A generic term used to describe the GIS function of providing a specific location to descriptive data. Geocoding applies to point data (e.g., alcohol outlet) as well as to area data (e.g., assaults in a census tract).

Geographical Information Systems (GIS): The geographic use of data to develop maps and statistical relationships that help describe processes like the relationship between alcohol outlets and violence or vehicle crashes and alcohol outlets.

GIS Distance and Adjacency Function: The distance between individual objects (e.g., bars and crashes) and whether areas are adjacent to one another.

Kriging: A technique that can be used to develop contour maps (e.g., maps that show lines of equal value such as DWI rates) from a limited number of points or areas (which can be given a value at the centroid).

Overlay Function: The ability to combine multiple layers of information.

Spatial Analysis: Using geographic data to mathematically model the relationship between measures such those mentioned above, that is, alcohol outlets and violence.

Spatial Autocorrelation: The measure of similarity between values (for a given variable, e.g., income) located in space. Similarity of values in spatial proximity may indicate some underlying mechanism that is spatial in nature and contributes to the spatial pattern of the predictor variable. Controlling for spatial autocorrelation reduces statistical bias in parametric modeling.

Spatial Clusters: A greater than expected geographically close group of occurrences or events (e.g., deaths, crashes, alcohol outlets).

Spatial Modeling of Risk Factors for Gender-Specific Child Mortality

Mohammad Ali
International Vaccine Institute, Korea

Christine Ashley
University of Minnesota, USA

M. Zahirul Haq
ICDDRB, Centre for Health and Population Research, Bangladesh

Peter Kim Streatfield
Center for Population and Health Research, Bangladesh

INTRODUCTION AND BACKGROUND

The global reduction of child mortality has been a priority of international and national organizations for the last few decades. Despite widespread global efforts to improve child survival, the latest UNICEF report on the *State of the World's Children 2000* (UNICEF, 2000) indicates that child mortality rates continue to remain higher in lesser developed countries (LDCs), and in some areas, girls continue to die at a greater rate than boys.

Matlab, a rural area of Bangladesh, shows child mortality has declined greatly in the last two decades (Figure 1). The rates have also declined in other rural areas across the country, though a lesser extent than in Matlab. The gender mortality differential that was notoriously high in Matlab in the 1980s virtually disappeared by the mid-1990s. Despite the large decline and elimination of gender disparity in child mortality, spatial variations in mortality continue to exist in Matlab (Ali et al., 2001). This variation most likely exists in the rest of Bangladesh, as well as in other less developed countries.

A multitude of social, demographic, economic, and environmental factors has been identified as global factors for contributing to the gender differential of child mortality in Bangladesh (Bairagi et al., 1999; Basu, 1989; Bhuiya & Streatfield, 1991; Chen et al., 1980; Fauveau et al., 1991; Islam & Ataharul, 1989; Koenig & D'Souza, 1986; Muhuri, 1995; Muhuri & Menken, 1997; Salway & Nasim, 1994). Measuring these social environments at a local geographic scale is important for identifying the environments that have a link with higher child mortality. The knowledge would eventually help in directing our effort towards the areas where it is essentially needed for child survival. Using a geographic information system (GIS), we attempted to define local level social environments, and to identify the environments that are influencing local level geographic variation of gender specific mortality in Matlab.

DATA AND METHODS

Child (one to four years) mortality for the periods 1984-86 and 1994-96 were chosen from Matlab demographic surveillance systems to examine the changes in local level spatial variation of mortality over a decade. The three-year periods were chosen to avoid temporal bias in the data.

The Matlab GIS (Figure 2) provided 7691 geographically referenced points of *baris* (a group of patrilineally households living in a geographic space). The mortality rates were smoothed at the point of *baris* by a spatially adaptive filtering that counted population (child) size close to 35. The choice of the specific population size is a trade- off between very low and very high smoothed data. Then, by using *kriging* (a method that is used to extrapolate the data at a regular spaced interval (Oliver &Webstar, 1990)) and contour mapping, gender specific surface maps of higher mortality were created for the two time periods. The temporal surface maps of each gender group were cross-classified, and an output map of each gender group was created (Figures 3 and 4) with the changes shown in: risk area remains risk area (R-R), risk

Table 1. Thresholds (deaths/1,000 children) for defining high-risk areas of child mortality

Gender	1984-86	1994-1996
Male	30.4	9.0
Female	41.4	10.5

S

Table 2. Results of the multiple logistic regression (spatial) of the male child mortality

Variables	Model: Non-risk to Risk (N-R)		Model: Non-risk to Non-risk (N-N)	
	Regression coefficient	t-test	Regression coefficient	t-test
Intercept	-3.817858	-250.14736	2.690645	166.0446
multiple groups of professionals	0.000090	39.580791	-0.000058	-26.80997
high educational status	0.000362	18.031919	-0.000044	-2.516286
high fertility	0.000106	10.347898	0.000123	13.851040
comparison area	0.170880	17.679539	-0.058213	-6.269722
outside embankment	0.469618	51.742344	-0.306266	-35.09447
cost distance to TC	0.001229	21.177443	-0.000528	-9.392699
high density of population	0.009980	1.036064	0.054325	5.873418
Hindu dominance	0.001021	17.087276	-0.000358	-6.205904
Adjusted R^2	**0.114470**		**0.034517**	

Models: Non-risk area changed to risk area (N-R) and non-risk remains risk area (N-N).

area changed to non-risk area (R-N), non-risk area changed to risk area (N-R), and non-risk area remains non-risk area (N-N). These maps were used as the dependant variable in a spatial regression model.

The data on social environment were smoothed using a fixed filtering of size 210 square meters. Here, our notion is that social environment beyond that distance has little influence on an individual's health outcome. The filtered data were used to create surface maps of the environment using the same kriging and contour mapping techniques. The social environment maps include educational status, population density, fertility rate, major occupations such as agriculture, fishing, and business, and Hindu (minority religious group) predominant areas. All of these maps were described in binary category: the dominant surface got the value "1"; else get the value "0". Finally, distance surfaces were created from the dominant surface of each of these maps; the closer a point to the dominated surface, the smaller the value of the point.

A distance map was created for embankment as an input variable of the model. The study site consists of two programmatic areas: non-intervention and intervention. The attribute of the former one was denoted by "1" and the latter one was denoted by "0". The map of accessibility to nearest health care was determined by cost (in time) distance. In computing the cost distance, rivers and canals were treated as barriers, and assigned their cost five times higher than that of the ground, which is based on waiting time and speed of movement through bodies of water.

Analytical Methods

Logistic regression was employed to determine predictive risk factors for gender-specific child mortality. The regression model takes the form:

$$\text{logit}(p)=\ln(p/1-p)=a+b_1x_1+b_2x_2+\ldots\ldots+b_nx_n$$

where p is the dependent variable expressing the probability of the outcomes.

RESULTS

The thresholds to define higher mortality areas for boys and girls are given in Table 1. The results of the multiple logistic regression show that the combined effects of the factors explain 11% of the total variations in predicting N-R for male children (Table 2). Being in an area outside an embankment is the most important factor in predicting risk for male child mortality, followed by areas of multiple groups of professionals. A high fertility rate also predicts spatial risk for male child mortality. Areas of higher population density show the lowest spatial risk among the factors. On the other hand, the same factors do not predict much (only 3% of total variations) in explaining N-N.

The results of the analysis for predicting R-R and R-N of male children are presented in Table 3. In the table, the model R-R shows that the comparison area predicts higher spatial risk for male child mortality. The effect of an

Table 3. Results of multiple logistic regression (spatial) of the male child mortality

Variables	Model: Risk to Risk (R-R)		Model: Risk to Non-risk (R-N)	
	Regression coefficient	t-test	Regression coefficient	t-test
Intercept	-4.086931	-178.4538	2.057620	71.24044
multiple groups of professionals	0.000044	12.188567	-0.000017	-5.028326
comparison area	0.808863	50.562550	-0.233676	-16.77735
outside embankment	0.382884	26.844124	-0.115347	-8.904040
cost distance to TC	0.001489	16.102747	0.000166	1.993834
Adjusted R^2	**0.162250**		**0.012257**	

Models: Risk area remains risk area (R-R) and risk area changed to non-risk area (R-N).

Table 4. Results of multiple logistic regression (spatial) for female child mortality

Variables	Model: Non-risk to Risk (N-R)		Model: Non-risk to Non-risk (N-N)	
	Regression coefficient	t-test	Regression coefficient	t-test
Intercept	-3.753035	-240.24765	2.534480	153.52201
multiple groups of professionals	0.000050	20.524385	-0.000021	-9.178230
high educational status	0.000396	20.178802	-0.000223	-12.881060
high fertility	0.000060	5.703178	0.000164	18.772718
comparison area	0.369911	38.578598	-0.190321	-20.605009
outside embankment	0.306635	32.521801	-0.114156	-12.863127
cost distance to TC	0.001503	25.077888	-0.000688	-12.130284
high density of population	0.216925	21.126863	-0.117856	-12.287079
Hindu dominance	0.001357	21.803726	-0.000698	-11.900236
Adjusted R^2	**0.101230**		**0.028542**	

Models: Non-risk area changed to risk area (N-R) and non-risk remains risk area (N-N).

Table 5. Results of multiple logistic regression (spatial) for female child mortality

Variables	Model: Risk to Risk (R-R)		Model: Risk to Non-risk (R-N)	
	Regression coefficient	t-test	Regression coefficient	t-test
Intercept	-4.393467	-194.85032	2.253702	75.933701
multiple groups of professionals	0.000100	31.636171	-0.000045	-14.123897
high educational status	0.000302	10.860458	0.000152	6.076112
comparison area	0.555219	40.760643	-0.186226	-14.012320
outside embankment	0.243221	20.090242	-0.064532	-5.284036
cost distance to TC	0.001201	13.869790	-0.000286	-3.372448
Hindu dominance	0.001133	12.658359	-0.000239	-2.746511
Adjusted R^2	**0.191666**		**0.019756**	

Models: Risk area remains risk area (R-R) and risk area changed to non-risk area (R-N).

embankment in predicting spatial risk is also significant. The distance to treatment centers, and areas dominated by multiple groups of professionals also influence risk areas to remain risk areas. Overall, the combined effect of the variables explains 16% of the total variation for male child mortality. Table 3 also shows that the factors included in predicting R-R do not predict R-N.

The results of regression analysis for N-R and N-N of female child mortality are given in Table 4. In this case, the combined effect of the factors explains 10% of the total variations in predicting N-R. The comparison area predicts the highest risk among all factors of mortality included in the model, followed by embankment and then cost distance to health care. However, modeling N-N with

Figure 1. Temporal pattern of gender-specific mortality in Matlab, Bangladesh

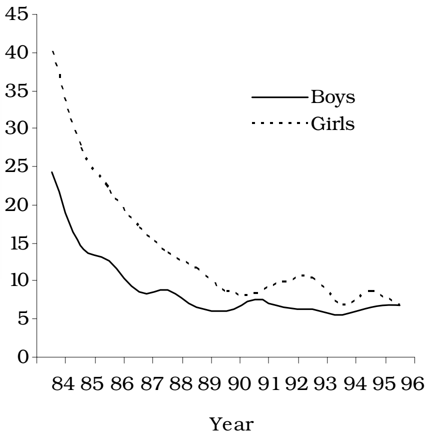
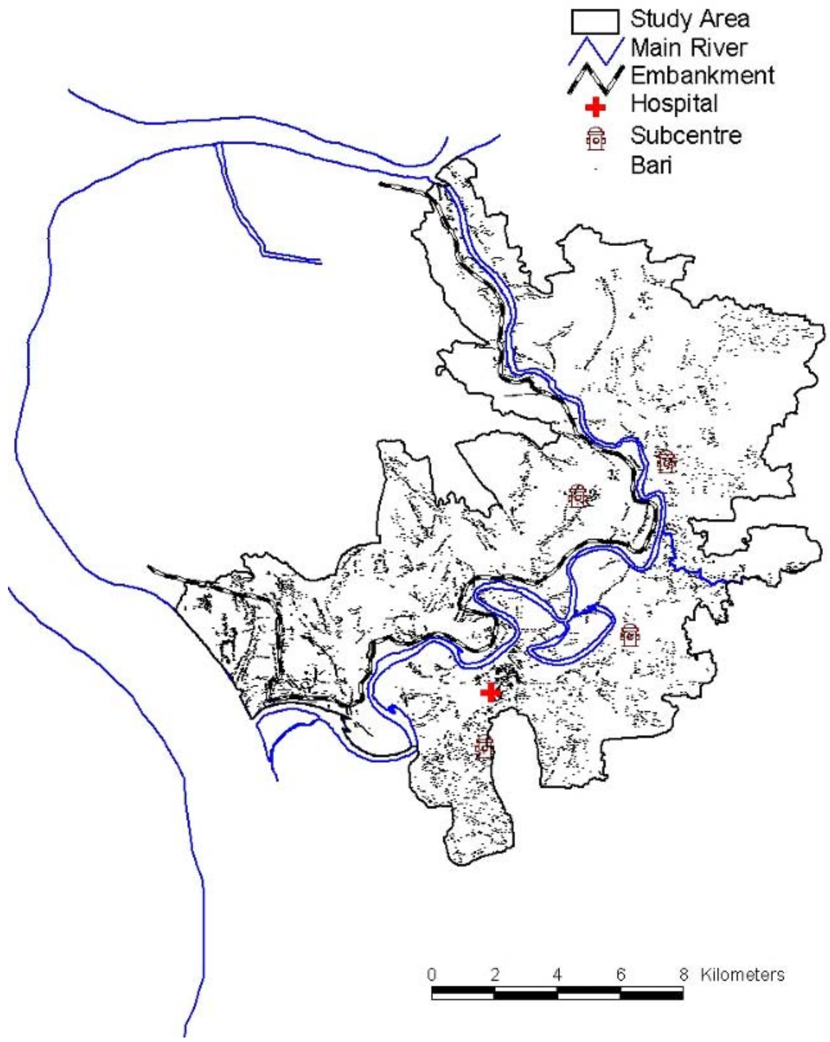


Figure 2. Matlab study area, Bangladesh



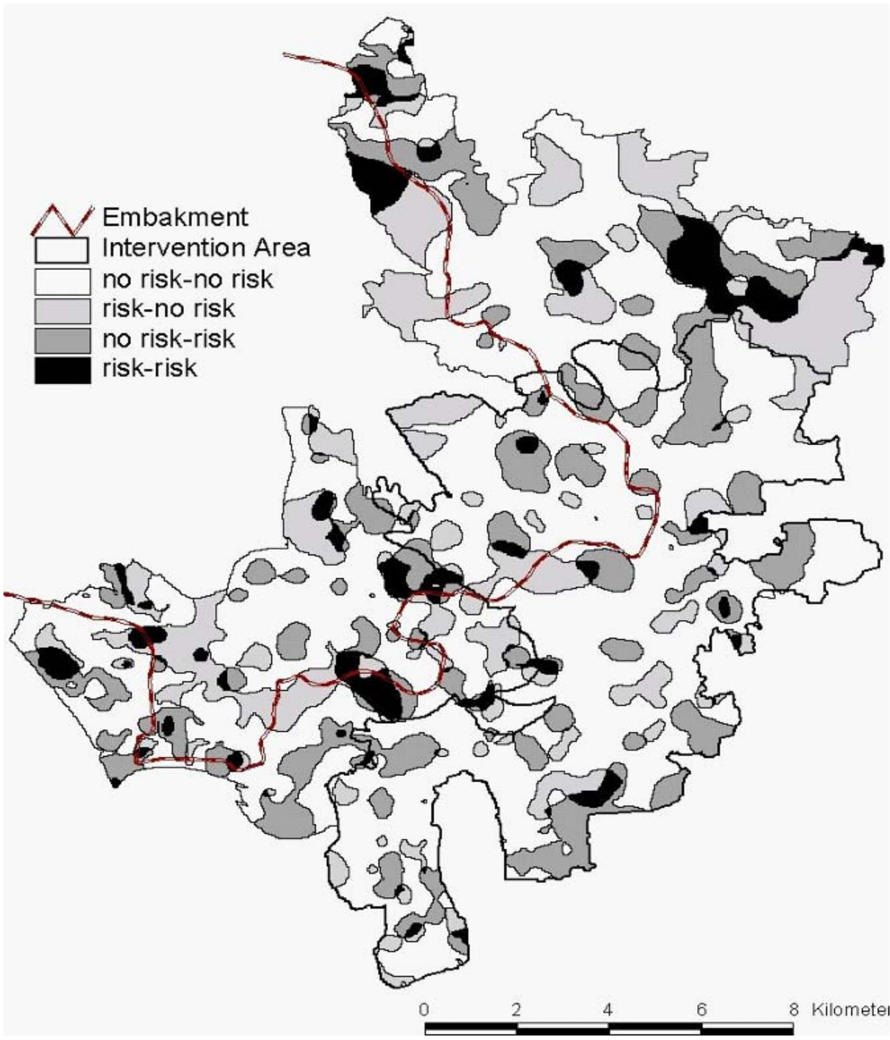
these factors has little impact on the outcome.

Table 5 presents the results of the analysis of R-R and R-N for female children. Of the factors, the comparison area influences the highest mortality, followed by multiple groups of professionals and embankment in predicting R-R. Educational status, distance to health care, and Hindu community explains 19% of the total variations in predicting R-R for girls. However, these factors did not show any influence in predicting R-N.

DISCUSSION AND CONCLUSION

The cross-classification analysis of temporal maps on higher mortality areas shows that although the mortality

Figure 3. Temporal change in risk for higher mortality of boys in Matlab, Bangladesh

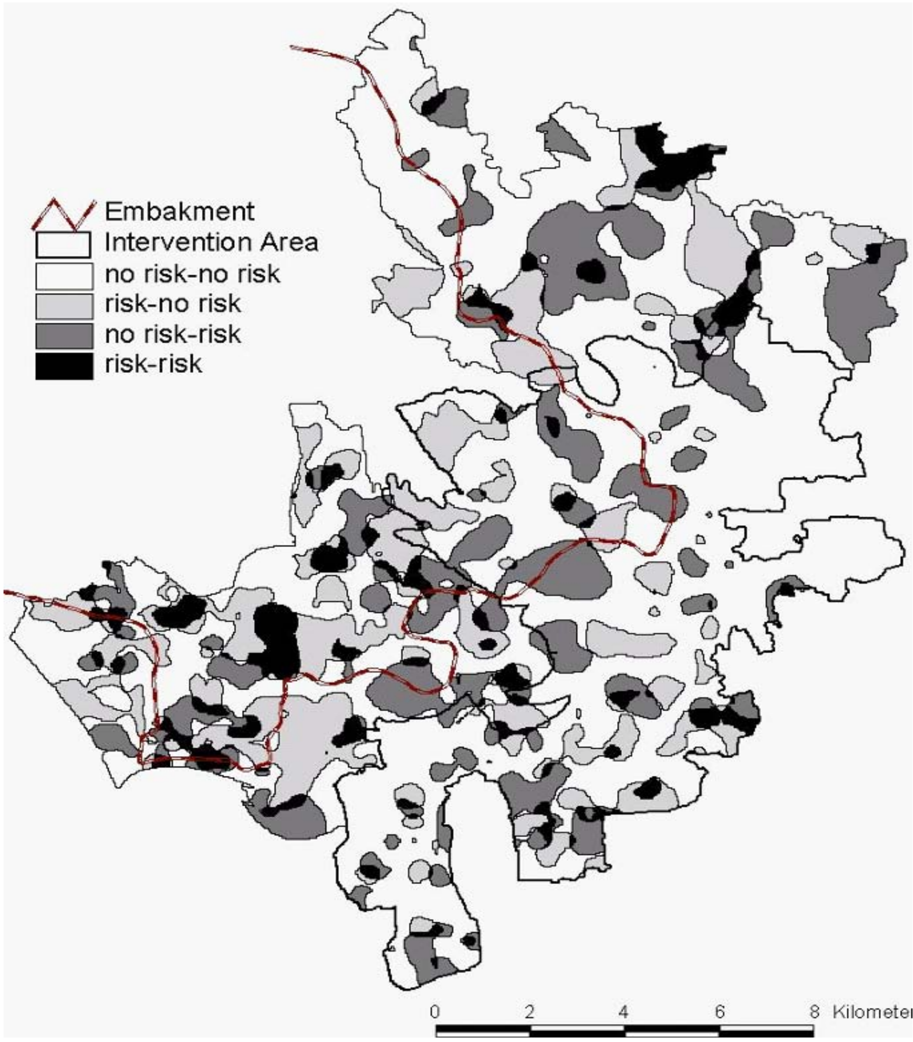


rates have declined over time, they have not declined consistently over the entire study area. About 25% of the areas that experienced lower child mortality in the 1980s shifted to higher mortality areas in the 1990s. On the other hand, a major part of the higher mortality area of the 1980s became lower mortality areas in the 1990s. The shift of one type to another type of mortality area resulted in the wide geographical variation of mortality in Matlab. The cross-classification of gender-specific mortality maps also indicates significant spatial variability in mortality between boys and girls. The variability implies that socioenvironmental processes of mortality for boys and girls are not the same.

The comparison areas where no intensive intervention was carried out, areas outside an embankment, areas distant from a treatment center, and the areas predominated by multiple groups of professionals predict a higher risk for male child mortality for both the models (N-R and R-R). The results are somewhat similar to previous studies (Myaux et al., 1997; Rahman et al., 1982). Areas of higher educational status, lower fertility level, lower population density, and non-Hindu areas predict spatial risk for male child mortality when modeling N-R.

The positive association of the areas of higher educational status with higher male child mortality indicates that neighbors' education status did not play any role in declining mortality to the extent that was anticipated. It suggests that education may not be a good predictor of mortality reduction, as it did not influence the outcome when modeling mortality with R-R. Similarly, areas of higher fertility show a negative association with higher mortality when modeling N-R, but do not show any signifi-

Figure 4. Temporal change in risk for higher mortality of girls in Matlab, Bangladesh



cant influence on mortality when modeling R-R for male children. The lower mortality in Hindu-dominant areas is perhaps related to their settlement pattern near the main river, which facilitates year-round easy access to treatment centers by their own boats. However, further study is required for a better understanding of the relationships between distance to treatment centers and mortality rates.

When predicting R-R and N-R for female children, areas of higher educational status, the comparison area of no intensive intervention, being outside an embankment, and greater distance to a TC were the common factors for higher mortality. Areas with multiple groups of professionals show a contrasting picture in predicting mortality when modeling the data with N-R and R-R for female. Predicting areas of mortality with the N-R model, the Hindu community shows a negative association for males and a positive association for female child mortality, indicating a gender bias in seeking health care services. According to Hindu laws, the daughters cannot inherit their father's property—this causes dowry systems existing during marriage. In many instances, the parents of a bride have to sell their assets to get the dowry demanded by the groom's family. Thus, a female child is not desirable in this society, and receives less attention by her parents.

Geographical analysis of health variation and its determinants have an important role in the design of equitable health programs. By modeling spatial risk, we identified several socioenvironmental predictors for gender-specific child mortality of this rural area of Bangladesh. A large amount of variation is left unexplained, modeling social environment for addressing gender-specific child mortality. This was anticipated, as we know that health is influenced by many factors (Wilkinson et al., 1998).

In conclusion, although this article describes the fact of the Matlab study area in Bangladesh, the findings of this study may provide insight into the socioenvironmental barriers of child survival elsewhere in the country and the region.

REFERENCES

Ali, M., Emch, M., Tofail, F., & Baqui, A.H. (2001). Implications of health care provision on acute lower respiratory infection mortality in Bangladeshi children. *Social Science and Medicine, 52*(2), 267-277.

Bairagi, R., Sutradhar, S., & Alam, N. (1999, June). Levels, trends and determinants of child mortality in Matlab, Bangladesh, 1966-1994. *Asia-Pacific Population Journal,* 51-68.

Basu, A.M. (1989). Is discrimination in food really necessary for explaining sex differentials in childhood mortality? *Population Studies, 43,* 193-210.

Bhuiya, A., & Streatfield, K. (1991). Mother's education and survival of female children in a rural area of Bangladesh. *Population Studies, 45,* 253-264.

Chen, L., Rahman, M., & Sardar, A.M. (1980). Epidemiology and causes of death among children in a rural area of Bangladesh. *International Journal of Epidemiology, 9*(1), 25-33.

Fauveau, V., Koenig, M.A., & Wojtyniak, B. (1991). Excess female deaths among rural Bangladeshi children: An examination of cause-specific mortality and morbidity. *International Journal of Epidemiology, 20*(3), 729-735.

Islam, M., & Ataharul, M. (1989). Assessing homogeneity and association in male and female mortality in Bangladesh. *Rural Demography, 16*(1-2), 1-7.

Koenig, M.A., & D'Souza, S. (1986). Sex differences in childhood mortality in rural Bangladesh. *Social Science and Medicine, 22*(1), 15-22.

Muhuri, P.K., & Menken, J. (1997). Adverse effects of next birth, gender, and family composition on child survival in rural Bangladesh. *Population Studies, 51*(3), 279-294.

Myaux, J., Iqbal, A., Uzma, A., Chakraborty, J., Ali, M., & Hossain, M. (1996). Environmental hazards as a leading cause of death in children from Bangladesh. *International Child Health, VII*(4), 49-52.

Oliver, M.A., & Webster, R. (1990). Kriging: A method of interpolation for geographical information systems. *International Journal of Geographical Information Systems, 4,* 313-332.

Rahman, M.M., Aziz, K.M.S, Munshi, M.H., Patwari, Y., & Rahman, M. (1982). A diarrhea clinic in rural Bangladesh: Influence of distance, age, and sex on attendance and diarrhoeal mortality. *American Journal of Public Health, 72*(10), 1124-1128.

Salway, S.M., & Nasim, S.M.A. (1994). Levels, trends and causes of mortality in children below 5 years of age in Bangladesh: Findings from a national survey. *Journal of Diarrhoeal Disease Research, 12*(3), 187-193.

UNICEF. (2000). *The state of the world's children 2000.* New York.

Wilkinson, P., Grundy, C., Landon, M., & Stevenson, S. (1998). GIS in public health. In A. C. Gatrell & M. Löytönen (Eds.), *GIS and health (GISDATA series 6).* Taylor & Francis.

KEY TERMS

Contour Map: A line connecting points of equal elevation on topographic surface. The contour map depicts continuous distribution of the phenomena on earth surface.

Cross-Classification of Maps: A method that overlays two thematic maps of binary attributes to come up with an output theme containing attributes of both the themes.

Kriging: A geostatistical gridding method that produces visually appealing maps from irregularly spaced data. The estimates of kriging are unbiased and have minimum variance. The method is based on the assumption of spatial autocorrelation of data, and the autocorrelation structure (spatial distribution of the data) is addressed through variogram modeling. The underlying assumption of the variogram is that two observations close together are more similar than those further apart.

Matlab Demographic Surveillance Systems: A demographic surveillance system of about a size 200,000 population in a rural area of Bangladesh that has been maintained since 1966. The surveillance systems collect vital demographic data of the study population in a routine basis.

Social Environment: The environment created by human beings with their knowledge, attitude and behavior in a close neighborhood.

Spatial Filter: Spatial filtering is commonly used to enhance satellite imagery for visual interpretation. It involves applying a mathematical formula such as a mean or median to a group of pixels in a raster image using a moving window. It is also being used to create smoothed maps. The method can also be used to remove random noise caused by inaccurate records or mislocated cases.

Spatial Risk: The risk that is linked to locations in geographic space, usually from physical or social environment.

Surface Map: A three-dimensional portrait of the earth surface. Here, the third dimension is the spatial processes of social environment of the study area.

Speech and Audio Signal Applications

Hector Perez-Meana
National Polytechnic Institute of Mexico, Mexico

Mariko Nakano-Miyatake
National Polytechnic Institute of Mexico, Mexico

INTRODUCTION

With the development of the VLSI technology the performance of signal processing devices have greatly improved making possible the implementation of more efficient systems to storage, transmission enhancement and reproduction of speech and audio signals. Some of these successful applications are shown in Table 1.

BACKGROUND

The signal processing has played a very important role in the development of many speech and audio systems. This section presents a review of some of these successful applications.

Echo Cancellation for Long Distance Transmission

A very successful speech signal processing application is the adaptive echo cancellation used to reduce a common but undesirable phenomenon in most telecommunications systems, called echo. Here when mismatch impedance is present in any telecommunications system, a portion of the transmitted signal is reflected to the transmitter as an echo, which represents an impairment that degrades the system quality (Messershmitt). In most telecommunications systems, such as a telephone circuit, the echo is generated when the long distant portion consisting of two one-directional channel (four wires) is connected with a bidirectional channel (two wires) by means of a hybrid transformer. If the hybrid impedance is perfectly balanced, the two one-directional channels are uncoupled and no signal returns to the transmitter side (Messershmitt). However, in general the bridge is not perfectly balanced because the required impedance to properly balance the hybrid depends on the overall impedance network. In this situation, part of the signal is reflected, producing an echo. To avoid this problem an adaptive filter is used to generate an echo replica that is then subtracted from the signal to be transmitted. Subsequently the adaptive filter coefficients are updated to minimize, usually, the mean square value of the residual echo (Madisetti & Williams, 1998). To obtain an appropriate operation, the echo canceller impulse response must be larger than the longer echo path to be estimated. Thus assuming a sampling frequency of 8kHz and an echo delay of about 60ms, an echo canceller with 256 or more taps is required (Haykin, 1991). Besides the echo path estimation, another important problem is how to handle the double talk, that is, the simultaneous presence of the echo and the near speech signal (Messershmitt).

Table 1. Main audio and speech signal processing applications

- Echo cancellation in telecommunication systems
- Acoustic echo cancellation
- Noise canceling
- Active noise cancellation
- Adaptive equalization
- Narrowband speech coding
- Broadband audio and speech coding

Acoustic Echo Cancellation

A critical problem affecting speech communication in teleconferencing system is the acoustic echo. When a bidirectional line links two rooms, the acoustic coupling between loudspeaker and microphones in each room causes an acoustic echo perceivable to the users in the other room. The best way to handle it appears to be the adaptive echo cancellation. An acoustic echo canceller generates an echo replica and subtracts it from the signal picked up by the microphones. The residual echo is then used to update the filter coefficients such that the mean square value of approximation error is kept to a minimum (Perez- Meana et al., 2002).

Adaptive Noise Cancellation

The adaptive noise canceller is a generalization of the echo canceller in which a signal corrupted with additive noise must be restored or enhanced. When a reference

Table 2. Digital speech coding standards

Rate Kb/s	Application	Type of Coder	Year
64	Public Switched Telephone Network	Pulse Code Modulation (PCM)	1972
2.4	U.S. Government Federal Standard	Linear Predictive Coding	1977
32	Public Switched Telephone Network	Adaptive Differential PCM	1984
9.6	Skyphone	Multi-Pulse Linear Predictive Coding (MPLPC)	1990
13	Pan-European Digital Mobile Radio (DMR) Cellular System (GSM)	Regular Pulse Excitation Linear Prediction Coding (RPE-LPC)	1991
4.8	U.S. Government Federal Standard	Codebook Excited Linear Prediction Coding (CELP).	1991
16	Public Switched Telephone Network	Low Delay CELP (LD-CELP)	1992
6.7	Japanese Digital Mobile Radio (DMR)	Vector Sum Excited Linear Prediction Coding (VSELP)	1977

signal correlated with the noise signal but uncorrelated with the desired one is available, the noise cancellation can be achieved by using an adaptive filter to minimize the total power of the output of the difference between the corrupted signal and the estimated noise, such that the resulting signal becomes the best estimate, in the mean square sense, of the desired signal. This system works fairly well when the reference and the desired signal are uncorrelated among them. However, appropriate reference signals are not always available. To solve this problem several noise canceling algorithms have been proposed that are resistant to crosstalk situations. A different approach, developed by Dolby Laboratories, is used in the Dolby noise reduction systems in which the dynamic range of the sound is reduced during recording and expanded during the playback (Davis, 2002). Several types of Dolby noise reduction systems have been developed, including the A, B, C, and HXpro. Most widely used is the Dolby B, which allows acceptable playback even on devices without noise reduction. The Dolby B noise reduction system uses a preemphasis that allows masking the background hiss of a tape with a stronger audio signal, especially at higher frequencies. This effect is called psychoacoustic masking (Davis, 2002).

Active Noise Cancellation

A related problem to noise cancellation is the active noise cancellation that is intended to reduce the noise produced in automotive equipment, home appliances, industrial equipment, airplane cabins, and so forth. Active noise is achieved by introducing a canceling antinoise wave through an appropriate array of secondary sources, which are interconnected through an electronic system using adaptive systems with a particular cancellation configuration. Here, the adaptive filter generates an antinoise that is acoustically subtracted from the incoming noise wave. The resulting wave is captured by an error microphone and used to update the adaptive filter coefficients, such that the total error power is minimized. Some successful applications of this technology are the earphone, electronic mufflers, noise canceling in airplane cabins, and so forth (Davis, 2002; Kuo & Morgan, 1996).

Adaptive Equalization

Digital information transmitted through physical communication channels is often distorted due to the intersymbol interference (ISI), which is mainly caused by multipath propagation of the transmitted symbols and by the no ideal characteristics of the communication channels (Proakis, 1985). To overcome this problem, several methods have been reported in the literature such as transversal equalizers (TE), decision feedback equalizers (DFE), maximum likelihood sequence estimation (MLSE), and so forth (Proakis, 1985). Most of them perform fairly well with near stationary communication channels. However, they still present difficulties with mobile communication channels that fluctuate markedly with the motion of a vehicle, ground irregularities and changing environments, because in these situations the communication channels can be assumed to be stationary only within fractions of a wavelength, that is, over 4-5m in the 900-MHz band (Madisetti & Williams, 1998).

Table 3. Some of the most used wideband speech and audio coders

Coder	Bitrate	Application
• CCITT G.722	64 kbits/s, 56 kbits/s, 48 kbits/s	Speech
• Low Delay CELP	32 kbits/s	Speech
• Compact Disc	1.41 Mbits/s	Audio
• Perceptual Audio Coder	128 kbits/s	Audio
• MP3 (MPEG-1 layer III)	96 kbits/s	Audio
• Windows Media Audio	64 kbits/s	Audio
• VQF	80 kbits/s	Audio
• Mp3PRO	64 kbits/s	Audio
• OGG Vorbis	96 kbits/s	Audio
• WAV	10 MB/min	Audio

Speech and Audio Coding

Besides interference cancellation, speech and audio signal coding are another very important signal processing applications. This is because low bit rate coding is required to minimize the transmission costs or to provide a cost efficient storage. Here we can distinguish two different groups. The narrowband speech coders are used in telephone and some video telephone systems in which the quality of telephone-bandwidth speech is acceptable, and the wideband coders are used in audio applications that require a bandwidth of at least 20kHz for high fidelity (Madisetti & Williams, 1998).

Narrowband Speech Coding

The most efficient speech coding systems for narrow band applications use analysis-synthesis based method in which the speech signal is analyzed during the coding process to estimate the main parameters of speech that allow its synthesis during the decoding process. Two sets of speech parameters are usually estimated: The linear filter system parameters, which model the vocal track, estimated using the linear prediction method, and the excitation sequence. Most speech coders estimate the linear filter in a similar way, although there have been proposed several methods to estimate the excitation sequence that determines synthesized speech quality and compression rates. Among these speech coding systems we have the multipulse and regular pulse linear predictive coding and the codebook exited linear predictive coding (CELP) and so forth that achieve bit rates among 9.6 Kb/s and 2.4 kb/s, with reasonably good speech quality (Madisetti & Williams, 1998). Table 2 shows the main characteristics of some of the most successful speech coders.

Wideband Audio and Speech Coding

Higher bandwidths than that of the telephone bandwidth results in major subjective improvements. Thus a bandwidth of 50 to 20 kHz not only improves the intelligibility and naturalness of audio and speech, but also adds a feeling of transparent communication and eases speaker recognition. However, this will result in the necessity to store and transmit a much larger amount of data, unless efficient wideband coding schemes are used. Wideband speech and audio coding intend to minimize the storage and transmission costs while providing an audio and speech signal with no audible differences between the compressed and the actual signals with 20kHz or higher bandwidth and a dynamic range equal to or above 90 dB. Four key technology aspects play a very important role in achieving this goal: the perceptual coding, frequency domain coding, window switching and dynamic bit allocation. Using these features, the speech signal is divided into a set of non-uniform subbands to encode with more precision the components that are perceptually more significant and with fewer bits the perceptually less significant frequency components. The subband approach also allows the use of the masking effect, in which the frequency components close to those with larger amplitude are masked and then they can be discharged without audible degradation. These features, together with a dynamic bit allocation, allow significant reduction of the total bits required for encoding the audio signal without perceptible degradation of the audio signal quality. Some of the most representative coders of this type are listed in Table 3 (Madisetti & Williams, 1998).

Table 4. Other successful audio and speech applications

- Signal processing for hearing aids
- Virtual musical instruments synthesis
- Alaryngeal speech enhancement
- Cross-language voice conversion
- Speech and speaker recognition

FUTURE TRENDS

The audio and speech processing have achieved an important development during the last three decades; however, there still remain several problems that must be solved, such as to develop more efficient echo canceller structures with improved the double talk control systems. In adaptive noise canceling a very important issue that remains unsolved is the crosstalk problem. To get efficient active noise cancellation systems it is necessary to cancel the antinoise wave that inside into the reference microphone that distorts the reference signal to reduce the computational complexity of ANC systems, as well as to develop more accurate secondary path estimation. Another important issue is to develop low distortion speech coders for bit rates below of 4.8 kBits/s. Another important issue is to increase the convergence speed of adaptive equalizers, to allow the tracking of fast time varying communication channels. Finally, the speech and audio processing systems will also contribute to improve the performance of medical equipments such as hearing aids and alaryngeal speech enhancement systems, as well in security through the development of efficient and accurate speaker recognition and verification systems.

CONCLUSION

Audio and speech signal processing have been fields of intensive research during the last three decades, becoming an essential component for interference cancellation and speech compression and enhancement in telephone and data communication systems, high fidelity broadband coding in audio and digital TV systems, speech enhancement for speech and speaker recognition systems, and so forth. However, despite the development that speech and audio systems have achieved, the research in those fields is increasing in order to provide new and more efficient solutions in the previously mentioned fields, and several others such as the acoustic noise reduction to improve the environmental conditions of people working in the airports, in factories, and so forth. For example, research could be used to improve the security of restricted places through speaker verification systems, and improve the speech quality of alaryngeal people through more efficient speech enhancement methods. Thus it can be predicted that the speech and audio processing will contribute to more comfortable living conditions during the following years.

REFERENCES

Bosi, M., & Goldberg, R. (2002). *Introduction to digital audio coding and standards*. Boston: Kluwer Academic Publishers.

Davis, G. (2002). *Noise reduction in speech applications*. New York: CRC Press.

Gold, B., & Morgan, N. (2000). *Speech and audio signal processing*. New York: John Wiley & Sons.

Haykin, S. (1991). *Adaptive filter theory*. Englewood Cliffs, NJ: Prentice Hall.

Kondoz, A.M. (1994). *Digital speech*. Chichester, England: Wiley & Sons.

Kuo, S., & Morgan, D. (1996). *Active noise control system: Algorithms and DSP implementations*. New York: John Wiley & Sons.

Madisetti, V., & Williams, D. (1998). *The digital signal processing handbook*. Boca Raton, FL: CRC Press.

Messershmitt, D. (1984, March). *Echo cancellation in speech and data transmission*. IEEE Journal of Selected Areas in Communications, 2(2), 283-297.

Perez-Meana, H., Nakano-Miyatake, M., & Nino-de-Rivera, L. (2002). *Speech and audio signal application. Multirate systems: Design and applications* (pp. 200-224). Hershey, PA: Idea Group Publishing.

Proakis, J. (1985). *Digital communications*. New York: McGraw Hill.

KEY TERMS

Adaptive Algorithm: Method used to modify the filter coefficients, online, in order to minimize the power of an adaptive filter output error.

Adaptive Filter: Linear system that modifies its parameters, minimizing some given criterion of the difference between its output and a given reference signal. Widely used in echo and noise canceling, equalization of communication channels, antenna arrays, and so forth.

Antinoise: Estimated replica of acoustic noise generated by an active noise canceller system, which is used to cancel an environmental noise.

Crosstalk: Interference present in a signal propagating through a communication produced by other signals present at an adjacent channel.

Double Talk: An interference produced when the speakers in both ends of a telephone line simultaneously speak. This phenomenon greatly disturbs the echo canceller performance.

Hybrid Transformer: Device used to connect two one-directional with a bi-directional channel, keeping uncoupled among them the two one-directional channels.

Narrowband Speech Signal: Speech signal with a frequency band equal to that of the telephone channel, that is, with a bandwidth of 300 to 3300 kHz.

Signal Compression: Signal coding that allows a reduction of the total number of bits required to represent a given signal without distortion or with negligible distortion.

Speaker Verification: Signal processing required for verifying the speaker identity by using his/her speech features.

Speech Enhancement: Signal processing performed in a given speech signal to improve its intelligibility and signal-to-noise-ratio.

Wideband Signal: Signal with a bandwidth wider than that of the telephone channel, usually between 50Hz to 20 kHz. This fact results in major subjective improvements.

Sponsorship in IT Project Management

David Bryde
Liverpool John Moores University, United Kingdom

David Petie
Petie Ltd., United Kingdom

INTRODUCTION

Since the 1970s academics and practitioners in the discipline of project management have sought answers to two inter-related questions: How is project success defined and measured? What are the influences on project success? To answer the first question people have studied project success criteria/key performance indicators. To answer the second, studies have focused on project critical success factors. Daniel (1961) introduced the concept of "success factors," stating that "in most industries there are usually three to six factors that determine success; these key jobs must be done exceedingly well for a company to be successful" (p.116). Approaches to the management of information have been established using Daniel's concept. For example, Rockart (1979) developed a Critical Success Factor (CSF) method for meeting the information needs of top executives. This method focused on understanding the objectives and goals of the company and the factors (CSFs) critical to their achievement, and establishing information systems to report on performance in these two areas. A key challenge has been to integrate the definitions and measures of success with CSFs, and in this respect work has been carried out to develop frameworks linking models of success criteria (the measures of success) with CSFs (see, for example, van Veen-Dirks & Wijn, 2002). The concept of CSFs has also been applied to project environments, with project CSFs being "those inputs to the management system that lead directly or indirectly to the success of the project" (Cooke-Davies, 2002, p. 185). Project management theory has also looked for a holistic answer to the questions of "How is project success defined and measured?" and "What are the influences on project success?", through the development of models linking project success criteria and project CSFs (Westerveld, 2002; Bryde, 2003).

BACKGROUND

Sponsorship as a Project Critical Success Factor

In respect of individual project CSFs, the importance of project sponsorship to achieving successful project outcomes has long been recognized. In a review of previous studies of CSFs, the sponsorship of projects by top management was highlighted as one of 8 major influences on success (Pinto & Slevin, 1987) and confirmed in a later study by the same authors as one of 10 influences (Pinto & Slevin, 1989). The importance of sponsorship is recognized through the distinction made between Macro CSFs, which involves activities in the realm of the sponsoring organisation and Micro CSFs, which are carried out in the domain of the project team (DeWitt, 1988). This crucial role of sponsorship has been identified in various manufacturing and service-related business environments, such as defence (Tishler et al., 1996), construction (Black et al., 2000), research & development (Pinto & Slevin, 1989) and management consultancy (Jang & Lee, 1998). Studies of project CSFs in IT environments have confirmed the pivotal influence of project sponsorship (see, for example, Bytheway, 1999; Fui-Hoon Nah et al., 2001; Procacinno et al., 2002).

ROLES OF PROJECT SPONSOR AND PROJECT MANAGER

A key step to delivering successful outcomes is gaining an understanding of the perspectives of all stakeholders to the project, including the sponsor and other stakeholders (Wright, 1998; Wateridge, 1995). In IT environments a lack of understanding has contributed to projects being unsuccessful. A failure by project managers to understand that users emphasized longer-term criteria relating to delivering workable systems, rather than short-term

Table 1. Role of the project sponsor

- Define the business benefit/requirements
- Understand the risks to benefit realisation
- Agree the project definition, including project objectives
- Develop the project strategy, including priorities
- Help define the project success criteria
- Specify any constraints
- Determine the relative priorities of cost, time and quality
- Monitor the project's business environment
- If necessary, re-define or cancel the project
- Monitor project performance
- Take delivery at project completion
- Monitor benefit realization
- Champion the project, including making resources availab
- Support the project manager in their role

Table 2. Role of the project manager

- Develop an effective working relationship with sponsor
- Deliver the project to time and cost, quality objectives
- Evaluate the risk profile and advise the sponsor
- Meet the defined project success criteria
- Manage the sponsor's and other stakeholders' expectations
- Define the project
- Build and lead the project team
- Monitor and control project progress
- Keep sponsor informed of progress and problems
- If necessary, recommend redefining or canceling of the project
- Hand over to the sponsor on completion

criteria linked to meeting time and cost objectives was a characteristic of IT projects perceived to be unsuccessful (Wateridge, 1998). A necessary step in achieving understanding is defining and delineating the roles and responsibilities of the project sponsor and project manager (Belassi & Tukel, 1996). *Table 1* summarizes the role of the project sponsor, drawing from the following extant literature: Snowdon (1976), Kliem & Ludin (1992, pp.163-169), Morris (1994, pp.188-189, 258-259), Briner et al. (1999, pp.65-67), Turner (1999, pp.50-53), and Hall et al. (2003).

The role of the project manager is summarized in *Table 2*. This table was constructed with reference to Gaddis (1959), Middleton (1967), Mantel et al. (2001, pp. 27-34), Anderson & Merna (2003), and Kendra & Taplin (2004).

FUTURE TRENDS

Although the importance of the sponsor to achieving project success is now well established (see earlier section "Background - Sponsorship as a Project Critical Success Factor"), and the roles of the sponsor and project manager are fairly well defined in theory, there are a number of critical issues that still need to be addressed in the future.

Awareness Among Project Sponsors of Their Role in Benefit Realization

Firstly, in some project organizations the role of the sponsor, especially in relation to their relationship with the project manager, is not clearly understood. This can lead to problems. For example, the responsibility for monitoring benefit realization (a sponsor role shown in *Table 1*) is often abdicated (without any corresponding authority) to project managers. However, it must be remembered that project managers only deliver products. A failure by the sponsor to fulfill their role in relation to benefit realization will lead to sub-optimal performance from the strategic perspective of the organisation.

Integration of Project Sponsor and Project Manager Perspectives

Secondly, there is the practical difficulty of integrating the perspectives of the sponsor and the project manager. The sponsor should be focused on the benefits of an IT system, its impacts on the organisation and its contribution to the company vision. However the project manager should be focused on delivering products, IS/IT functionality and how it will be used operationally. The need for such a strategic and tactical integration underpins work on the development of new models for the delivery of successful IT projects (for example, Byers & Blume, 1994; Ward & Elvin, 1999). Project managers, by their very nature, are backward looking and tend to focus on what has been done, what has been spent and the problems they have delivering the product. Sponsors, on the other hand, must be forward looking, trying to ensure that the benefits will be delivered and that nothing gets in the way of this.

Distinction Between "Executive" Project Sponsorship and the Sponsorship of Individual Projects

Finally there is the question of what is meant by executive sponsorship? There is a difference between the sponsorship provided by one person to an individual project and the sponsorship given by the organisation to enable a strategic vision to be linked with effective delivery of an IT system. This second type of sponsorship, which is a potentially new paradigm, involves company directors, senior executives and middle managers understanding how to lead and manage change. It focuses on ensuring the organisation is ready for a project to be undertaken. For example, an exploratory study showed that the greatest influence on achieving successful outcomes from business process re-engineering initiatives was the innovative capacity of the organisation (Teng et al., 1998). In a similar fashion, in the new paradigm, the sponsors of projects must ensure that organizational competency issues linked to project management, in such areas as programme management, benefits management, sponsoring individual projects, long-term planning and governance, training and support, are properly addressed.

To ensure success, most organizations intuitively restrict projects to a particular department or area, with little or no cross-functional interaction. This is purely a competency issue, largely relating to executive sponsorship rather than project management competencies. The line of least resistance is to manage work packages or small projects with limited impact on other areas of the business. However, successful organizations are able to manage complex and inter-related projects through effective sponsorship. It is not hard to appreciate that decisions made regarding a particular initiative will have ramifications in other parts of the business. These elements need to be managed by individuals who have the interests of the organisation as a whole at heart, rather than a specific initiative. Effective executive sponsorship achieves this. Project managers cannot be expected to manage these aspects as well as deliver the projects on time, to the right quality and at the right cost.

This executive sponsorship equates to demonstrating effective leadership and being a good corporate citizen. A real test of this is when projects get cancelled because the benefits cannot be realized or the risk profile is too great.

CONCLUSION

Over the last 30 years there has been much work carried out in establishing the importance of project sponsorship to achieving successful outcomes and in defining the theoretical roles of the project sponsor and the project manager. This has resulted in a "traditional" view of the project sponsor and project manager roles, which in the case of the sponsor tends to focus on the sponsorship of individual projects. A further element, which was discussed earlier, is that of "executive" sponsorship, which focuses on organizational competency issues linked to project management. Further study is needed to establish the extent to which the two types of sponsorship, "traditional" and "executive," are undertaken in practice and to explore the relationship between different types of sponsorship and project success. This will require the taking of a holistic view of "project success," that is, incorporating project management-related criteria (meeting cost, time and quality objectives), satisfaction-related criteria (for the customer, project team and user) and benefits-related criteria (tangible and intangible). From this holistic perspective it will be possible to investigate how variations in the nature of project sponsorship impacts on the different success criteria, with the most effective sponsorship being that which meets all three.

Another focus for future work in this area needs to be on how project sponsors and project managers work together to ensure that IT project performance is optimized and benefits are realized. To achieve such optimization it may also be necessary to redefine the sponsorship paradigm to focus on addressing organizational competency in project management. For this to be useful in terms of project management practice there will need to be accompanying education, training and awareness-raising activities of the broader sponsorship role.

REFERENCES

Anderson, D.K., & Merna, T. (2003). Project management strategy: Project management represented as a process based set of management domains and the consequences for project management strategy. *International Journal of Project Management*, *21* (6), 387-393.

Belassi, W., & Tukel, I. (1996). A new framework for determining critical success/failure factors in projects. *International Journal of Project Management*, *14* (3), 11-151.

Black, C., Akintoye, A., & Fitzgerald, E. (2000). An analysis of success factors and benefits of partnering in construction. *International Journal of Project Management*, *18*, 423-434.

Briner, W., Hastings, C., & Geddes, M. (1999). Project Leadership (2nd ed.). Aldershot: Gower.

Bryde, D.J. (2003). Modelling project management performance. *International Journal of Quality & Reliability Management*, *20* (2), 228-245.

Byers, C.R., & Blume, D. (1994). Tying critical success factors to systems development. *Information & Management*, *26* (1), 51-61.

Bytheway, A. J. (1999). Successful software projects and how to achieve them. *IEEE Software, 16* (3), 15-18.

Cooke-Davies, T. (2002). The "real" success factors on projects. *International Journal of Project Management*, *20* (3), 185-190.

Daniel, R.D. (1961). Management information crisis. *Harvard Business Review, 39* (5), 111-121.

De Witt, A. (1988). Measuring project success. *International Journal of Project Management, 6* (3), 164-170.

Fui-Hoon Nah, F., Lee-Shang Lau, J., & Kuang, J. (2001) Critical factors for successful implementation of enterprise systems. *Business Process Management Journal*, *7* (3), 285-296.

Gaddis, P.O. (1959). The project manager. *Harvard Business Review*, *37* (3), 89-98.

Jang, Y., & Lee, J. (1998). Factors influencing the success of management consulting projects. *International Journal of Project Management, 16* (2), 67-72.

Kendra, K.A., & Taplin, L.J. (2004). Change agent competencies for information technology project managers. *Consulting Psychology Journal: Practice and Research*, *56* (1), 20-34.

Kliem, R.L., & Ludin, I.S. (1992). *The people side of project management*. Aldershot: Gower.

Hall, M., Holt, R., & Purchase, D. (2003). Project sponsors under new public management. *International Journal of Project Management*, *21* (7), 495-502.

Mantel, S.J., Meredith, J.R., Shafer, S.M., & Sutton, M.M. (2001). *Project management in practice*. New York: John Wiley & Sons.

Middleton, C.J. (1967). How to set up a project organization. *Harvard Business Review*, *45* (2), 73-83.

Morris, P.W.G. (1994). *The management of projects*. London: Thomas Telford.

Pinto, J.K., & Slevin, D.P. (1987). Critical factors in successful project implementation. *IEEE Transactions on Engineering Management*, *34* (1), 22-27.

Pinto, J.K., & Slevin, D.P. (1989). Critical success factors in R&D projects. *Research Technology Management*, pp. 31-35.

Procaccino, J.D., Verner, J.M., Overmyer, S.P., & Darter, M.E. (2002). Case study: Factors for early prediction of software development success. *Information and Software Technology*, *44* (1), 53-62.

Rockart, J.F. (1979). Chief executives define their own data needs. *Harvard Business Review, 57* (2), 81-93.

Snowdon, M. (1976). Project management in the early stages. *Engineering & Process Economics*, *1* (4), 257-264.

Teng, J.T.C., Fiedler, K.D., & Grover, V. (1998). An exploratory study of the influence of the IS function and organizational context on business process reengineering project initiatives. *Omega*, *26* (6), 679-698.

Tishler, A., Dvir, D., Shenhar, A., & Lipovetsky, S. (1996). Identifying critical success factors in defense development projects: A multivariate analysis. *Technological Forecasting and Social Change*, *51* (2), 151-171.

Turner, J.R. (1999). *The handbook of project-based management* (2nd Ed.). London: McGraw-Hill.

Van Veen-Dirks, P., & Wijn, M. (2002). Strategic control: Meshing critical success factors with the balanced scorecard. *Long Range Planning, 35* (4), 407-427.

Ward, J.W., & Elvin, R. (1999). A new framework for managing IT-enabled business change. *Information Systems Journal*, *9*, 197-221.

Wateridge, J. (1995). IT projects: A basis for success. *International Journal of Project Management, 13* (3), 169-172.

Wateridge, J. (1998). How can IS/IT projects be measured for success? *International Journal of Project Management, 16* (1), 59-63.

Westerveld, E. (2003). The project excellence model: Linking success criteria and critical success factors. *International Journal of Project Management*, *21* (6), 411-418.

White, D., & Fortune, J. (2002). Current practice in project management: An empirical study. *International Journal of Project Management*, *20* (1), 1-11.

Wright, J.N. (1998). Time and budget: The twin imperatives of a project sponsor. *International Journal of Project Management, 15* (3), 181-186.

KEY TERMS

Organisational Competency in Project Management: Ensuring that the organisation is in a ready state in order for projects to be able to deliver benefits.

Project Critical Success Factors: The influences on the success, or otherwise, of a project. A distinction can be made between the underlying factors, or causes of success or failure and the symptoms of an ineffective project management process. The lack of top management support is a typical project critical success factor, which can lead to a variety of symptoms, such as adequate resources not being made available to the project.

Project Manager: An individual with the responsibility of ensuring the project objectives are delivered.

Project Performance: The degree to which the project meets its overall objectives. This compares with *project management performance*, which is the degree to which the traditional objectives of cost, time and quality are met.

Project Sponsor: An individual or group with the responsibility and authority to ensure that the project benefits are realised.

Project Stakeholder: Any person or group that has an interest in the project. The interest could be in the project outcome, outputs or the project management process.

Project Success Criteria/Key Performance Indicators: The measures of success. The terms project success criteria and project key performance indicators are used interchangeably. Traditional measures are meeting cost, time and quality objectives. Other measures are linked to the attributes used by a stakeholder to judge whether their expectations have been met.

Standards for Web-Based Integration Adapters

Bill Karakostas
City University, London, UK

INTRODUCTION

EAI (enterprise application integration) refers to the plans, methods, and tools aimed at modernizing, consolidating, and coordinating the computer applications in an enterprise. Typically, an enterprise has existing legacy applications and databases and wants to continue to use them while adding or migrating to a new set of applications that exploit the Web, e-commerce, extranet, business-to-business (B2B) commerce, and other new technologies. Enterprise application integration is difficult mainly because there is no standard infrastructure for communication between heterogeneous systems. The four types of B2B and A2A integration challenges that most organizations encounter today are user-interface integration, application integration, business-to-business integration, and data integration.

- ***User-Interface Integration:*** Employees, suppliers, and other trading partners need fast access to relevant and up-to-date information without having to search through a variety of inconsistent Web sites, navigate dissimilar user interfaces, and enter multiple passwords. To access this information, users enter their chosen realm via a *portal* that presents a consistent interface to different applications.
- ***Application Integration:*** Business processes involving multiple applications require integration of application logic and functions or tasks. To facilitate this integration, applications must communicate to exchange important business information. If a business is interested in selling goods over the Web, then its Web store must be integrated with its payment server. Although these are two separate pieces of software, they are components of the same order-taking business process.
- ***B2B Integration:*** B2B integration is imperative for businesses to participate in a marketplace or other B2B-exchange environment. These environments bring together buyers and suppliers. When properly architected, a B2B-exchange platform can integrate catalog sources such as supplier-hosted catalogs, specialized vertical exchanges, and other online information.
- ***Data Integration:*** In most B2B e-commerce environments, data integration is essential, especially when data comes from many different sources. Data integration allows applications to work with data from different computing platforms, databases, and operating systems.

BACKGROUND

Distributed Web and component-based architectures will remain strategic for most organizations for the foreseeable future. Typical enterprise systems include ERP systems, mainframe transaction-processing systems, other technology legacy applications, or big hierarchical databases. For all these strategic systems, the challenge remains the same, that is, to make them full participants in Web-based e-commerce and business-to-business transactions.

To allow such systems to connect and interoperate on the Web, a number of protocols and languages such as XML (extensible markup language) and Java and software such as application adapters have been proposed. The section below discusses their role in application integration on the Web.

XML

Using XML to represent the data being exchanged between applications presents significant advantages from an integration viewpoint because it lets the receiving application obtain the data and metadata simultaneously. The application can more easily understand the data it receives as such data are *self-describing*. Once a system is given the metadata and data, a clear and fairly automated approach can be implemented for mapping the original data into the native format of existing applications using technologies such as the extensible style sheet language (XSLT). Implementing such automated mapping requires companies to provide dictionaries of their business terms. This can help a tool analyze new input document formats and try to associate elements to an existing internal system representation. Several current integration initiatives include XML, for example, OAG, Microsoft's BizTalk, and RossettaNet, which are attempting to establish business semantics and process standards on top of XML technology.

Integration Adapters

An integration adapter is a software component used to connect an application (called *target*) to another application (called *source*) in enterprise application integration so that data from the target application can be interfaced with and transmitted to and from the source application.

Application adapters serve to transform data back and forth from the native format of the target application to that of the source application. To establish a communication line between the two applications, the adapter acts as an agent on the source machine to send and receive requests from and to the target. The purpose of the adapter is to monitor the requests that are performed from a source application and is responsible by selecting an appropriate communication mechanism to send the data to the target for further processing. Adapters can be set to send notifications when an applications data has changed, publish data to one or many other applications, and subscribe to data topics from other applications. Adaptors hide the complexity of the application's programming interface. Traditionally, this task has been completed by bespoken in-house utility software that is costly to write and maintain.

Web-Based Integration Adapters

Web-based adapters link applications in one computing environment with information or logic in another computing environment using Web-based protocols for data and communications. Adapters typically consist of run-time components that reside on each end of the communication pipe between the platforms being integrated. An application that invokes an adapter to get access to required resources is often called the source, while the target is the remote data store or application logic the application accesses. Adapters use an application programming interface (API) that shields the application (and its developer) from the complexities involved in accessing data or resources. Historically, APIs have been proprietary, such as Microsoft's component object model (COM). Other common APIs are based on industry standards such as Java database connectivity (JDBC), J2EE (Java 2 Enterprise Edition) connector, simple object access protocol (SOAP), or open database connectivity (ODBC). Transmission control protocol/Internet protocol (TCP/IP) has become the de facto transport for Web-based adapters.

Requirements for Web-Based Adapters

Web-based integration adapters must fulfill certain requirements; that is, they need to be as follows.

- **Lightweight:** small-in-size components that use a simple XML-described API and contain no application logic
- **Standards based:** adapters need to implement open ubiquitous protocols such as XML-based protocols (e.g., SOAP) over HTTP (hypertext transfer protocol)
- **Scalable and reliable:** Adapters need to use XML document handling and interpretation, connector replication, standard exception handling, and management agents with logging and performance instrumentation.
- **Intelligent:** Intelligent adapters minimize message volume and overhead by being able to be interrogated and remotely, dynamically configured to work in the most appropriate way.
- **Supporting transactional integrity:** Adapters can deal with distributed transactions in either synchronous, two-phase commit mode, or asynchronous transaction-messaging mode.
- **Supporting event-driven functionality:** Adapters have historically been unidirectional and request driven. They can facilitate the execution of a database query, detect the presence of a message on a queue for processing, or initiate some remote application logic. However, emerging adapters are bidirectional and event driven by the target system.
- **Handling multiple protocols and data conversions:** Adapters must hide the mundane conversion routines required for information exchange between two disparate platforms. This includes conversion of data types and character sets, as well as network protocols.
- **Supporting schema and metadata support:** Adapters expose schema and metadata information through the source-side API to help ensure ease of use for developers and dynamic processing, especially with updating their configuration when changes occur in the data schema of the target system, without the need for manual intervention.
- **Easy to use for developers and management:** High-quality adapters provide considerable additional value by easing implementation and providing support for system management.

FUTURE TRENDS

An emerging standard, the J2EE connector architecture (JCA), defines an architecture and adapter behavior. JCA provides the ability to create an adapter once, then use it anywhere, including in most integration and application servers. The J2EE connector architecture specifies a standard architecture for integrating Java applications with existing enterprise information systems. The JCA standard proposed by Sun and its Java Community Process partners is part of Version 1.3 of the J2EE specification. JCA defines system-level contracts for connection management, and security and transactions between an application server and a connector. The connector implements these contracts in a way that is specific for the given application. The application server implements a standard connector API, referred to as the common client interface (CCI).

CONCLUSION

The lack of common business and technical architectures makes application integration across enterprises over the Web complex. It is not just the differences in data formats or interfaces, but the lack of common definition of the business concepts that underlies the different sources to be integrated. XML-based standard vocabularies aim to overcome the problem of semantic incompatibility by standardizing the definitions of business concepts, documents, and message exchanges in business-to-business exchanges. Application integration adapters simplify the integration process to the extent that it is almost automatic or requires as little development effort to implement as the off-the-shelf application packages that they can integrate. Adapters therefore minimize the need to deal with the interface details involved in communications between a variety of different source and target systems. Adapters are becoming more important as new service-oriented paradigms and technologies such as Web services are emerging. Ultimately, a greater number of existing Web applications will become service-based, leveraging service-oriented access standards such as the Web services description language (WSDL), SOAP, and other target systems. To convert such existing legacy applications to Web services, adapters will be required to translate Web services semantics into the proprietary calls the target legacy-system supports.

Emerging standards such as JCA, discussed in this article, have the potential to make adapters reusable across applications as well as across vendors. In addition, the JCA standard is likely to evolve, accounting for needs such as metadata and scalable communications.

REFERENCES

Allen, D. W. (2001, June). Establishing an EAI Architecture. *EAI Journal*.

Balen, H. (2000, December). Deconstructing Babel: XML and application integration.

Bradbury, D. (2000, May/June). Middleware with bells on. *Application Development Advisor*.

CBDI Forum. (1999). *Application integration: A CBDI forum report.*

Christofi, S., & Karakostas, B. (2001). An XML and Java based approach to application integration in shipping. *Third International Conference on Information Integration and Web-Based Applications and Services*, 119-125.

Cresswel, M. (2002, September). What is down the road for adapters? *EAI Journal*.

Flanagan, D. (1999). *Java Enterprise in a nutshell: A desktop quick reference.* O' Reilly.

Karakostas, B., & Christofi, S. (2003). An approach to Web-based application integration using Java adapters and XML. In A. Dahanayake & W. Gerhardt (Eds.), *Web-enabled systems integration: Practices and challenges.* Hershey, PA: Idea Group Publishing.

Linthicum, D. (2002, December). The evolution of adapters. *EAI Journal*.

Sun Microsystems Inc. (2003). *J2EE connector architecture.* Retrieved from *http://java.sun.com/j2ee/connector/index.jsp*

KEY TERMS

Application Integration: Application integration is the process of bringing data or a function from one application program together with that of another application program.

Enterprise Application Integration: EAI (enterprise application integration) refers to the plans, methods, and tools aimed at modernizing, consolidating, and coordinating the computer applications in an enterprise. Typically, an enterprise has existing legacy applications and databases and wants to continue to use them while adding or migrating to a new set of applications that exploit the Internet, e-commerce, extranet, and other new technologies.

Integration Adapter: Data and application adapters (also known as data interfaces or data drivers) are native software objects that allow integration tools to retrieve data efficiently from complex, sometimes proprietary stores of information.

Integration Hub: A hub is a messaging server that exchanges important information between different applications so that the applications can interact in a meaningful way.

Messaging: In the context of an enterprise integration environment, messaging represents a communications system whereby important packets of information or messages are sent to and from queues by a software-based system known as a messaging server.

S

Staying Up-to-Date with Changes in IT

Tanya McGill
Murdoch University, Australia

Michael Dixon
Murdoch University, Australia

INTRODUCTION

Information and communications technology (ICT) has been changing rapidly over a long period, and this rate of change is likely to continue or increase (Benamati & Lederer, 2001a; Fordham, 2001). This rapid rate of change has produced many opportunities for organizations but has also brought with it many challenges (Benamati & Lederer, 2001b; Lederer & Mendelow, 1990). Among these challenges is the struggle for organizations to obtain personnel with the appropriate knowledge and skills in order to meet their ICT needs (Byrd & Turner, 2001; Doke, 1999). This is mirrored by the continual requirement for information technology (IT) professionals to keep up-to-date with the skills required by organizations (Benamati & Lederer, 2001a; Klobas & McGill, 1993).

Previous research has investigated the importance employers place on various skills and perceived deficiencies in these skills (e.g., Doke, 1999; Leitheiser, 1992; Nelson, 1991). While the call for improved communication and social skills has been consistent, the technical skills in demand have varied dramatically over time (Van Slyke, Kittner, & Cheney, 1998). Less has been written about students' perceptions of the importance of various ICT skills, though this was addressed in a study that compared Australian and American students' perceptions of ICT job skills (von Hellens, Van Slyke, & Kittner, 2000). This chapter provides an overview of a project that investigated the channels of information that ICT students use to keep up-to-date with employers' needs.

BACKGROUND

Given that the skills required by IT professionals change over time, IT professionals need effective methods to keep up-to-date. The methods used by IT professionals to keep up-to-date were studied by Klobas and McGill (1993). They identified the existence of a variety of information-gathering strategies and noted that while IT professionals tended to be diligent in their efforts to keep up-to-date, a majority found it difficult to do so. In a more recent study, Benamati and Lederer (2001a) investigated the coping mechanisms adopted by IT professionals and noted that many mechanisms were not successful.

If it is difficult for experienced IT professionals to keep up-to-date, it is likely that it is even more difficult for ICT students to do so. New graduates require marketable skills in order to gain good employment, but the skills most in demand change regularly. Little is known about how ICT students keep informed of employers' requirements or about how they ensure that they can meet those requirements. Yet, this knowledge would be of use to both educational institutions aiming to facilitate this process and to potential employers hoping to recruit students with the required skills.

Information about ICT skill requirements is available from a variety of sources in a variety of formats. Information sources include ICT suppliers, publishing companies, and universities. Formats include different types of publications, presentations, and personal contacts. The term "information channel" can be used to describe the various combinations of sources and formats of information.

HOW DO STUDENTS KEEP UP-TO-DATE?

Eighty-five information technology students at an Australian university were surveyed to investigate the channels of information that they use to keep up-to-date with employers' needs. Participants were recruited during class and completed a questionnaire on the spot.

The questionnaire listed information channels that may be used to keep up-to-date and asked participants whether they had used each channel within the last 3 months, and also asked them to rate the importance of each channel to them as a means of knowing what skills are in demand. Importance was measured on a 5-point scale ranging from (1) "Not important" to (5) "Vital." The initial list of channels of information was drawn from Klobas and McGill's (1993) report of the methods used by IT professionals to keep up-to-date with developments in

ICT. Several additional channels were included after consultation with industry contacts. Table 1 lists the information channels included in the questionnaire.

Overall, the students appeared to be diligent in their efforts to keep up-to-date with employers' skill requirements. The average number of channels used by the students during the previous 3 months was 3.8 (and the most common number used was 5). Thirteen students (15.3%) had not made any attempt to keep up-to-date during this period, and four (4.7%) had made use of all nine listed channels.

The information channels are ranked by frequency of use in Table 1. The most frequently consulted channels were newspaper employment and IT sections and Internet sources. University instructors had been consulted by about half of the participants during the previous 3 months. Other students had also been used as sources of information by a number of students (40%). This high level of use of other students to provide information about employers' skill requirements is understandable given the easy accessibility of other students (Klobas & McGill, 1993). Work colleagues were ranked seventh overall, but as only around a third of the participants had ICT work experience, this means that most of those with prior experience had consulted their colleagues (75% of those with prior ICT work experience had consulted their colleagues). The least used channels were books and vendor presentations. It is likely that students were conscious that information about employer skill requirements derived from books was not going to be sufficiently up-to-date to meet their needs.

Table 2 shows the importance rankings of the individual information channels. The most highly ranked information channel was Internet sources, such as the Cisco and Lucent sites. As well as being frequently used, newspaper ICT sections and employment pages were also considered very important (ranked two and three). University instructors were ranked fourth in importance, which was consistent with their frequency of consultation by students. Although other students were consulted by many students, they were not considered an important channel of information (ranked seventh). This suggests that students recognize that although other students are easily accessible sources of information, they are not necessarily accurate or reliable sources. Both books and vendor presentations were considered of low importance. In future research, it would be interesting to determine how well student perceptions match those of employers.

In addition to the items about methods used to keep up-to-date, participants were also asked several questions that addressed whether they believed they were, in fact, obtaining the skills employers required. A majority of participants believed that their degree would provide the skills employers require (67.1% "yes," 5.9% "no," and 27.1% "not sure"). This high level of confidence suggests that although only around 50% of students had consulted their instructors about employer skill requirements during the previous 3 months (and instructors were only given a medium ranking of importance), students implicitly accept that instructors know what skills students require. Industry certification was also seen as a very important means to ensure that students obtain the necessary skills (mean importance score was 4.18/5 for those students not yet working in the ICT industry). This is consistent with the results of a recent study on IT certification that found that students undertaking certification believe that the most important benefit of certification is that it provides "real-world" experience (McGill & Dixon, 2004).

Table 1. Information channels ranked by frequency of use

Rank	Information channel	Number	Percentage
1	Newspaper employment pages	56	65.9
2	Newspaper ICT sections	52	61.2
3	Internet sources (e.g., Cisco, Lucent)	47	55.3
4	University instructors	43	50.6
5	Other students	34	40.0
6	ICT magazines (e.g., *Packet Magazine*)	29	34.1
7	Work colleagues	24	28.2
8	Books	20	23.5
9	Vendor presentations	17	20.0

Table 2. Information channels ranked by importance

Rank	Information channel	Mean	Standard deviation
1	Internet sources (e.g., Cisco, Lucent)	3.55	1.40
2	Newspaper IT sections	3.38	1.44
3	Newspaper employment pages	3.30	1.30
4	University instructors	2.88	1.42
5	ICT magazines	2.62	1.43
6	Work colleagues	2.54	1.43
7	Other students	2.41	1.13
8	Books	2.24	1.34
9	Vendor presentations	2.13	1.32

Are There Demographic Differences in Use and Importance?

Patterns of use and perceptions of importance were further examined to determine whether gender, level of study, or previous ICT work experience had influence. Differences in use were explored using c^2 tests, and differences in importance were explored using independent sample *t*-tests. These factors had surprisingly little influence on patterns of use and perceived importance of information channels.

The first demographic factor considered was gender. No significant difference was found between the number of information channels used by male and female students. The only significant gender difference was for the levels of use and perceived importance of Internet sources. Male students used Internet sources more frequently and perceived them to be more important for keeping up-to-date with the skill requirements of employers.

The possible impact of previous ICT work experience was considered next. No significant difference was found between the number of information channels used by those with and those without previous ICT work experience. The only significant difference in usage of information channels was related to consultation with work colleagues and with other students. Those with previous work experience, not surprisingly, consulted with work colleagues more frequently, and they appeared to consider work colleagues a more important channel of information. Presumably, those with previous ICT experience would have received better quality information from their work colleagues than would those without ICT work experience who would have been receiving information from a pool of people with perhaps limited direct ICT experience.

Those without ICT work experience consulted other students more frequently, but there was no difference in perceptions of the importance of other students between those with and those without previous ICT work experience. As previously mentioned, this suggests that other students are consulted because of their accessibility rather than their credibility as sources of information. Those with previous ICT experience have other accessible sources of more credible information and, hence, do not rely so heavily upon other students.

The differences between undergraduate and postgraduate students were similar to those between students with previous ICT work experience and those without. This is consistent with postgraduate students being more likely to have previous ICT work experience than are undergraduates. (A total of 54.5% of postgraduates versus 22% of undergraduates had previous ICT work experience.) Undergraduate students consulted other students more frequently but did not value their information more highly. Postgraduate students also consulted work colleagues more frequently, but they did not value their input more highly. This finding differs from the added importance given to work colleagues by those with previous ICT experience, but the means are in the same direction, and the result may reflect the fact that 45.5% of the postgraduates did not have previous ICT work experience.

FUTURE TRENDS

The rapid rate of change in ICT is likely to continue (Benamati & Lederer, 2001b), and in fact, some authors believe that the rate of change is accelerating (Horn, 1999). This means that ICT students will continue to require access to up-to-date information about employers' ICT skill requirements. Given the increased role of electronic means of information dissemination (Bertot, 2003; Williams & Nicholas, 2001), it is likely that Internet sources of information will continue to be seen as the most important sources, and that their frequency of use will increase rapidly so that Internet sources will soon be the most frequently used. Greater broadband access will enable delivery of richer content and greater interactivity. Convergence of information technologies, such as notebooks, phones, and television, and the development of pervasive computing will provide even greater flexibility to students who wish to keep up-to-date with employer skill requirements.

CONCLUSION

New graduates require marketable skills in order to gain good employment, but as the ICT industry is subject to rapid change, the skills most in demand change regularly. The study described in this chapter investigated the approaches that a group of ICT students used to keep up-to-date with employers' skill needs. Overall, they appeared to be diligent in their efforts to keep up-to-date with skill requirements. The most commonly used channels were newspaper employment and IT sections and Internet sources. The same three channels were also rated most highly in terms of importance, with Internet sources being seen as most important.

Instructors were ranked relatively high in terms of both frequency of consultation and importance, and the results suggest an implicit confidence that the knowledge of instructors is up-to-date. While students have a wide variety of information channels available to them and do make use of them, instructors have a major role to play in providing up-to-date information about employers' needs. They need to be highly accessible and to ensure that their knowledge of employers' skill requirements remains current. Instructors should use studies of employers' requirements to assess their course offerings and to help guide their students.

REFERENCES

Benamati, J., & Lederer, A. L. (2001a). Coping with rapid changes in IT. *Communications of the ACM, 44*(8), 83–88.

Benamati, J., & Lederer, A. L. (2001b). How IT organizations handle rapid IT change: 5 coping mechanisms. *Information Technology and Management, 21*(1), 95–112.

Bertot, J. C. (2003). World libraries on the information superhighway: Internet-based library services. *Library Trends, 52*(2), 209–227.

Byrd, T. A., & Turner, D. E. (2001). An exploratory analysis of the value of the skills of IT personnel: Their relationship to IS infrastructure and competitive advantage. *Decision Sciences, 32*(1), 21–54.

Doke, E. R. (1999). Knowledge and skill requirements for information systems professionals: An exploratory study. *Journal of IS Education, 10*(1), 10–18.

Fordham, D. R. (2001). Forecasting technology trends. *Strategic Finance, 83*(3), 50–54.

Horn, P. M. (1999). Information technology will change everything. *Research Technology Management, 42*(1), 42–47.

Klobas, J. E., & McGill, T. (1993). Computing professionals and information about developments in information technology. *The Australian Computer Journal, 25*(4), 149–158.

Lederer, A. L., & Mendelow, A. L. (1990). The impact of the environment on the management of information systems. *Information Systems Research, 1*(2), 205–222.

Leitheiser, R. (1992). MIS skills for the 1990s: A survey of MIS managers perceptions. *Journal of Management Information Systems, 9*(1), 69–91.

McGill, T., & Dixon, M. (2004). Information technology certification: A student perspective. In M. Khosrow-Pour (Ed.), *Innovations Through Information Technology: 2004 IRMA International Conference,* (pp.302-306), New Orleans, Louisiana, USA. Hershey, PA: Idea Group Publishing.

Nelson, R. R. (1991). Educational needs as perceived by IS and end-user personnel: A survey of knowledge and skill requirements. *MIS Quarterly, 15*(4), 503–525.

Van Slyke, C., Kittner, M., & Cheney, P. (1998). Skill requirements for entry-level IS graduates: A report from industry. *Journal of Information Systems Education, 3*(9), 7–11.

von Hellens, L., Van Slyke, C., & Kittner, M. (2000). A comparison of Australian and American students' perceptions of IT job skills. In M. Khosrow-Pour (Ed.), *Challenges of Information Technology Management in the 21st Century. 2000 IRMA International Conference* (pp.915-916), Anchorage, Alaska, USA. Hershey, PA: Idea Group Publishing.

Williams, P., & Nicholas, D. (2001). *The Internet and the changing information environment*. London: Aslib-IMI.

KEY TERMS

Industry Certification: Certification involves passing a recognized standardized test (or set of tests) within particular subject areas. It intends to establish a standard of competency in defined areas. ICT industry certifications are designed to provide targeted skills that have immediate applicability in the workplace.

Information Channel: A term used to describe the various combinations of sources and formats of information.

Information Format: The arrangement and appearance of information. Format includes both the media used and the style of presentation.

Information-Gathering Strategies: The approaches and processes used by information seekers. Information-seeking behavior is influenced by previous experience, mental models, and preferences of information seekers.

Information Source: An organization or person from which information is obtained.

Information Technology Professionals: This is a term used to describe people for whom development and support of IT systems and related activities is their primary employment. The group includes people who design hardware, who develop and support information systems, and who train end users. It does not include people who use ICT in the course of pursuing other professions.

Information Technology Skills: All IT professionals require some computer skills; these may include particular programming languages or database skills or networking skills.

Stickiness and Web-Based Customer Loyalty

Guisseppi Forgionne
University of Maryland, USA

Supawadee Ingsriswang
BIOTEC and NSTDA, Thailand

INTRODUCTION

Between 1997-2003, Internet hosts grew from 16 million to over 233 million worldwide (www.isc.org, 2004). Of all Internet hosts, the number of commercial domain names (.com) is about 20.9%, which increased from 3.9 million in January 1997 to 48.6 million in January 2004.

Despite continued market growth, a number of Web sites have been unprofitable. From 2000 to 2003, at least 962 Internet companies ran out of money, shut down their operation and sold their businesses, with 63 5 of these enterprises operating in the business-to-consumer (B2C) sector (www.webmergers.com, 2004). Some notable failures were eToys.com, boo.com, bluefly.com, buy.com and valueamerica.com. An examination of the companies' IPO filings suggests that the collapses were caused by cut-price strategies, over-investment, incorrect expectations, and non-profitability. The surviving dot com companies in the new economy also have experienced financial difficulties. Amazon, a member of the Internet hall of fame, is still unable to generate a profit (www.webmergers.com (2004) REPORT). Surviving in the digital market has become a critical challenge for Web managers.

To face the business challenge, Web managers and marketers demand information about Web site design and investment effectiveness (Ghosh, 1998). Further, as the rate and diversity of product/service innovation declines and competition intensifies, Web managers must increase their efforts to improve their product/service and stabilize market shares. In this context, Donath (1999) and Hoffman (2000) noted the lack of, and the need for, research on Internet-related investment decisions. Web managers and marketers would benefit from reliable and consistent measuring tools for their investment decisions, and such tools are the focus of this article.

BACKGROUND

Since online consumers can switch to other Web sites or competitive URLs in seconds with minimal financial costs, most Web sites invest heavily in programs to attract and retain customers. The Web site's ability to capture consumers' attention is known widely as "stickiness".

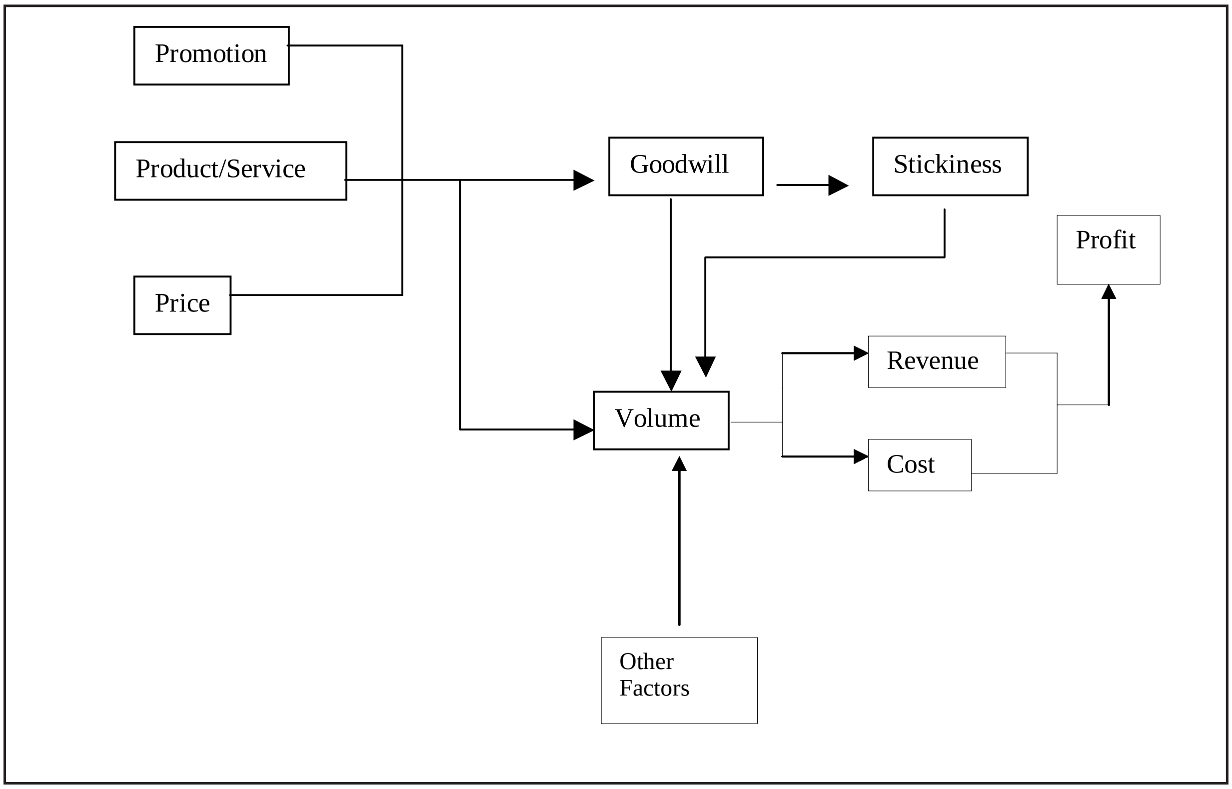
From one perspective (Rubric, 1999, p. 5), "The sticky factor refers to the ability of a web site to capture and keep a visitor's attention for a period of time." Likewise, stickiness can be described as the ability of the site in attracting longer and more frequent repeat visits or the ability of the site to retain customers (Anders, 1999; Davenport, 2000; Hanson, 2000; Murphy, 1999; O'Brien, 1999; Pappas, 1999). The various perspectives suggest that stickiness is similar to, if not the same concept as, customer loyalty.

On the other hand, Demers and Lev (2000) distinguished stickiness from customer loyalty. In their research, stickiness is represented by the average time spent at the site per visit, and customer loyalty refers to the frequency of visits. In fact, both the time duration and frequency of visits are mutually inclusive. The extension of the duration of relationships with customers is driven by retention (Reichheld, 1996). Loyal customers would visit the Web site frequently and remain with the site for a longer time period. In the process, these customers are generating not only traffic but revenue as well. Therefore, the goal of any Web site stickiness program must be to make customers loyal.

Customer loyalty has been a core element of marketing since the early eighties (Morgan & Hunt, 1994). The idea is to develop and maintain long-term relationships with customers by creating superior customer value and satisfaction. Enhanced customer satisfaction results in customer loyalty and increased profit (Anderson, Fornell & Lehmann, 1994; Reichheld & Sasser, 1990). Loyal customers, who return again and again over a period of time, also are valuable assets of the Web site. The ability to create customer loyalty has been a major driver of success in e-commerce (Reichheld & Schefter, 2000; Reichheld et al., 2000) since enhanced customer loyalty results in increased long-term profitability.

From marketing theory, then, stickiness can be viewed as the ability of a Web site to create both customer attraction and customer retention for the purpose of maximizing revenue and profit. Customer attraction is the ability to attract customers at the Web site, whereas customer retention is the ability to retain customer loyalty.

Figure 1. Stickiness conceptual model



MAIN THRUST

E-commerce customer loyalty, or stickiness, results from goodwill created by the organization's marketing efforts (Reichheld, 1996; Reichheld & Sasser, 1990), or

Stickiness = f(Goodwill) (1)
and Goodwill = f(Marketing Mix) (2)

By encouraging current and return visits, stickiness will influence the organizations' volume. Marketing theory also suggests that the mix of price (including switching costs to consumers), product/service (including site characteristics), and promotion (including banner and other Web site ads), as well as other factors (including consumer characteristics), will influence this volume (Page et al., 1996; Storbacka et al., 1994). Hence,

Volume = f(Stickiness, Promotion, Product, Price, Other Factors) (3)

Such volume will determine revenue, cost, and thereby profit for the Web site.

According to standard accounting practice and economic theory, profit is defined as the excess of revenue over cost, or

Profit = Revenue - Cost (4)

while revenue will equal volume multiplied by price, or

Revenue = Volume $\times$ Price (5)

According to standard accounting and economic theory, an organization's costs will have fixed and variable components, or

Total Cost = Fixed Cost + Variable Cost (6)

and variable cost will equal volume multiplied by unit cost, or

Variable Cost = Volume $\times$ Unit Cost (7)
with volume as defined in equation (3).

These marketing-economic-accounting-theory-based relationships are illustrated in Figure 1. This figure and equations (1) – (7) provide a conceptual model that specifies the manner in which the stickiness investment can contribute to an e-business organization's effectiveness by generating volume, revenue and profit. As such, the model identifies the information that must be collected to effectively evaluate investments in an e-commerce customer loyalty plan. These equations also provide a framework to objectively evaluate these plans and their impact on organizational operations and activities.

In practice, the model can be applied in a variety of ways. The general relationships can be used for strategy formulation at the macro-organizational level. In addition, the equations can be decomposed into detailed micro-level blocks with many variables and interrelationships. At this micro level, tactical policies to implement the macro strategies can be specified and evaluated.

EMPIRICAL TESTING

Much of the required model information is financial in nature. Such data are largely proprietary and therefore not readily available.

Summarized financial information is available from some Internet companies' quarterly (10-Q) and annual reports and other public sources. In particular, data were obtained from the 10-Q and 10-K reports for 20 Internet companies over the period 1999-2000. This pooled cross-section, time series data provided values for the revenue and marketing mix variables found in equations (1) - (7). Customer characteristics were proxied through demographic variables, and data for these variables were obtained from U.S. Census Bureau, Statistical Abstract of the United States for the period 1999-2000. The quarterly and annual report and Census data provided 120 observations to operationalize the model embodied in equations (1) – (7).

Goodwill, as defined in the economic literature (repeat business based on happy customers), is also not available from the annual and quarterly reports. However, an accounting measure of goodwill, amortization of goodwill and other intangible assets is available from the reports. Although not strictly the same concept as marketing goodwill, the accounting measure is likely to be correlated with economic goodwill and is thereby used as a proxy for economic goodwill (Chauvin & Hirschey, 1994; Jennings et al., 1996; McCarthy & Schneider, 1995).

No other data were provided by the available sources on a consistent and reliable basis. These data limitations reduced the variable list that could be used to operationalize the stickiness model specified in equations (1) – (7) to the list summarized in Table 1. To test the theory, then, it was necessary to use a truncated form of Figure 1's stickiness model.

Since, equations (4)-(7) in Figure 1's stickiness model are identities, only equations (1)-(3) must be estimated statistically from the available data. Prior research has shown that marketing investments in one period can continue to affect volume in subsequent periods (Hanson, 2000). Marketers call this phenomenon the carryover, lagged, or holdover effect. Stickiness investments and the marketing mix can be expected to create such an effect from new customers who remain with the Web site for many subsequent periods. Because of the data limitations, this important lagged effect had to be measured quarterly, and the carryover had to be restricted to one period. From the relationships in Figure 1, it seems reason-

$$\log(V_t) = \log(a_0) + a_1 \log(S_t) + a_2 \log(A_t) + a_3 \log(AI_t) + a_4 \log(GW_{t-1}) + z_1 \quad (8)$$

$$\log(S_t) = \log(b_0) + b_1 \log(GW_t) + z_2 \quad (9)$$

$$\log(GW_t) = \log(c_0) + c_1 \log(Q_t) + z_3 \quad (10)$$

Table 1. Variables with data

VARIABLE	DEFINITION
Volume **V**	total sales volume of the Internet company
Goodwill **GW**	amortization of goodwill and other intangible assets
Promotion **A**	sales and marketing expenditures
Product Quality **Q**	product development costs, which include research and development and other product development costs
Average Income **AI**	average consumer income
Stickiness **S**	number of unique visitors at the firm's Web site

Table 2. Results of the SUR estimation

Equation	MSE	R-Square	Adj. R-Square
Volume (equation 8)	0.3894	0.7408	0.7299
Stickiness (equation 9)	0.2190	0.1944	0.1862
Goodwill (equation 10)	3.5312	0.1296	0.1208

Table 3. SUR-estimated log-linear result

Variable	Volume (V_t)		Stickiness(S_t)		Goodwill (GW_t)	
	Est.	Prob>|T|	Est.	Prob>|T|	Est.	Prob>|T|
Constant	-5.09	.8282	8.74	<.0001	1.19	.0037
Stickiness	0.58	.0002				
Promotion	0.95	<.0001				
Income	0.09	.9885				
Lagged Goodwill	0.05	.1739				
Goodwill			0.17	<.0001		
Product Quality					0.77	<.0001

able to assume that the holdover will be expressed through the goodwill variable. Namely, goodwill in the previous quarter is assumed to affect revenue in the current quarter.

At time period *t*, the inputs may have joint, rather than independent, effects on volume. Moreover, elasticities (the percentage change in an output generated by a 1% change in an input) can provide useful information for policy making and analysis. To account for the nonlinearities, to facilitate the computation of elasticities, and to account for the carryover effect, equations (1) – (3) can be specified with the popular Cobb-Douglas production function, or in log-linear form in equations (8) – (10).

In equations (8) – (10), *t* denotes the considered time period (quarter), the a, b, c, and z labels denote parameters to be estimated, and the other variables are defined in Table 1. Price is excluded from the volume equation because data for the price variable are unavailable from the annual and quarterly reports.

Equations (8) – (10) below, form a simultaneous system of equations. Hence, simultaneous equation estimation techniques should be used to separate the effects of the marketing instruments and stickiness on volume. Otherwise, there will be a statistical identification problem.

There are several simultaneous estimation techniques that can be used to estimate the operational stickiness model (equations (8)-(10)). Using coefficients of determination, theoretical correctness of the estimated parameters, and forecasting errors as guidelines, the best results were generated by seemingly unrelated regressors (SUR). These results are summarized in Table 2 and Table 3.

Table 2 shows that the volume equation (8) has an R^2 = 0.74 and thereby accounts for about 74% of the variation in this variable. More importantly, Table 3 shows that the estimated coefficient of the stickiness variable in equation (9) is significant at the a = .05 level, suggesting that volume is significantly influenced by the stickiness of the Web site.

Table 2 shows that the operational stickiness model accounts for a relatively small percentage of the variation in stickiness and goodwill. Nevertheless, Table 3 indicates that goodwill has positive and significant effects on the stickiness of the Web site, while marketing instruments, such as product quality, also create positive effects on goodwill.

In short, the operational stickiness model provides reasonable statistical results. These results validate the theory that stickiness is an investment that, along with the marketing mix, will influence organizational performance. The results also indicate that volume will have inelastic responses to stickiness and promotion investments. That is, a 1% increase in such investments will lead to a less than 1% increase in volume. There are similar inelastic responses of stickiness to goodwill and goodwill to product quality.

FUTURE TRENDS

The analysis also suggests future research directions. First, it is useful to determine whether there are significant carryover effects from goodwill. If significant carryover effects are discovered, research should examine how long the investment in stickiness would contribute to the firm's short-term or long-term financial success. Second, research should evaluate what role competition plays in stickiness analysis. For example, additional research may examine how competitors' stickiness investments impact the firm.

Empirical e-commerce evaluations, then, will require the collection, capture, and retrieval of pertinent and consistent operational financial and other data for evaluation purposes. Appropriate statistical methodologies will be needed to estimate the proffered model's parameters from the collected data. Information systems will be required to focus the data, assist managers in the estimation process, and help such users to evaluate experimental customer loyalty plans. These tasks and activities offer new, and potentially very productive, areas for future research.

CONCLUSION

Figure 1's stickiness model shows how to measure the impact of online stickiness on revenue and profit, and the model also shows how stickiness and goodwill are related. The operational model can demonstrate and forecast the long-term revenue, profit, and return on investment from a Web site's products/services. In addition, the operational model: (a) provides a mechanism to evaluate Web redesign strategies, (b) helps Web managers evaluate market changes on custom loyalty plans, and (c) provides a framework to determine the "best" stickiness and other marketing policies.

In sum, in the e-commerce world, marketing still provides leadership in identifying consumer needs, the market to be served, and the strategy to be launched. Web managers must realize that desirable financial outcomes depend on their marketing effectiveness. With Figure 1's conceptual stickiness model, Web managers and marketers have a tool to evaluate their investment decisions.

REFERENCES

Anderson, E.W., Claes, F., & Lehmann, D.R. (1994, July). Customer satisfaction, marketshare and profitability: Findings from Sweden. *Journal of Marketing, 58*, 53-66.

Anders, G. (1999). The race for sticky Web sites – Behind the deal frenzy, a quest to hang onto restless clickers. *Wall Street Journal*.

Chauvin, K.W., & Hirschey, M. (1994). Goodwill, profitability, and the market value of the firm. *Journal of Accounting and Public Policy, 13*, 159-180.

Davenport, T.H. (2000, February). Sticky business. *CIO Magazine.www.cio.com/forums/ec*

Demers, E., & Lev, B. (2000). A rude awakening: Internet value-drivers in 2000. New York University. *http://www.stern.nyu.edu/~blev/newnew.html*

Donath, B. (1999). *The quest for eBusiness frameworks.* White Paper. eBusiness Center, The Pennsylvania State University.

Ghosh, S. (1998). Making business sense of the Internet. *Harvard Business Review, 76*(2), 126-135.

Hanson, W. (2000). *Principles of Internet marketing.* South-Western College Publishing.

Hoffman, D.L. (2000). The revolution will not be televised: Introduction to the special issue on marketing science and the Internet. *Marketing Science, 19*(1), 1-3.

Jennings, R., Robinson, J., Thompson, R.B., & Duvall, L. (1996, June). The relationship between accounting goodwill numbers and equity values. *Journal of Business Finance and Accounting, 23*(4), 513-533.

McCarthy, M.G., & Schneider, D.K. (1995). Market perception of goodwill: Some empirical evidence. *Accounting and Business Research, 26*(1), 69-81.

Morgan, R., & Hunt, S. (1994, July). The commitment trust theory of relationship marketing. *Journal of Marketing*, 20-38.

Murphy, K. (1999). Stickiness is the new gotta-have. *Internet World, 5*(12), 1-2.

O'Brien, J. (1999). Sticky shopping sites. *Computer Shopper, 19*(7), 121.

Page, M., Pitt, L., & Berthon, P. (1996). Analysing and reducing customer defection. *Long Range Planning, 29*(6), 821-824.

Pappas, C. (1999). Let's get sticky. *Home Office Computing, 17*(1), 90-91.

Reichheld, F.E. (1996, March-April). Learning from customer defections. *Harvard Business Review*, 56-69.

Reichheld, F.E., & Sasser, W.E. (1990, March). Zero defections: Quality comes to services. *Harvard Business Review, 14*, 495-507.

Reichheld, F.F., & Schefter, P. (2000, July/August). E-loyalty: Your secret weapon on the Web. *Harvard Business Review, 78*(4), 105-113.

Reichheld, F.F., Markey, R., & Hopton, C. (2000). The loyalty effect- the relationship between loyalty and profits. *European Business Journal, 12*(3), 134-139.

Rubric, Inc. (1999). Evaluating the sticky factor of e-commerce sites. *http://www.rubricsoft.com*

Storbacka, K., Strandvik, T., & Gronroos, C. (1994). Managing customer relationships for profit: The dynamics of relationship quality. *International Journal of Service Industry Management, 5*(5), 21-38.

www.isc.org (2004, January). Internet domain survey.

www.webmergers.com (2004, January). REPORT: Internet companies three years after the height of the bubble. *http://www.webmergers.com/data/article.php?id=67*

KEY TERMS

Customer Attraction: The ability to attract customers at the Web site.

Customer Loyalty: The ability to develop and maintain long-term relationships with customers by creating superior customer value and satisfaction.

Customer Loyalty Plan: A strategy for improving financial performance through activities that increase stickiness.

Customer Retention: The ability to retain customers and their allegiance to the Web site.

Goodwill: The amount of repeat business resulting from happy and loyal customers.

Stickiness: The ability of a Web site to create both customer attraction and customer retention for the purpose of maximizing revenue or profit.

Stickiness Model: A series of interdependent equations linking stickiness to an organization's financial performance.

Storage and Access Control Policies for XML Documents

George Pallis
Aristotle University of Thessaloniki, Greece

Konstantina Stoupa
Aristotle University of Thessaloniki, Greece

Athena Vakali
Aristotle University of Thessaloniki, Greece

INTRODUCTION

The Internet (and networks overall) are currently the core media for data and knowledge exchange. XML is currently the most popular standardization for Web document representation and is rapidly becoming a standard for data representation and exchange over the Internet. One of the main issues is XML documents and in particular, storage and accessing. Among data management issues, storage and security techniques have a particular importance, since the performance of the overall XML-based Web information system relies on them. Storage issues mainly rely on the usage of typical database management systems (DBMSs), whereas XML documents can also be stored in other storage environments (such as file systems and LDAP directories) (Amer-Yahia & Fernandez, 2002; Kanne & Moerkotte, 2000; Silberschatz, Korth & Sudarshan, 2002). Additionally, in order to guarantee the security of the XML data, which are located in a variety of the above storage topologies, the majority of implementations also provide an appropriate access control. Most storage systems cooperate with access control modules implementing various models (Joshi, Aref, Ghafoor & Spafford, 2001), whereas there are few commercial access control products available. However, there are some standardized XML-based access control languages that can be adopted by most tools.

This article focuses on presenting an overview related to both storing and securing XML documents. Such an overview will contribute to identifying the most important policies for storage and access in Web-based information systems (which extensively use the XML as the data representation format). In addition, the most well-known related implementations are presented. A more integrated survey on these issues is presented in Pallis, Stoupa, and Vakali (2004).

BACKGROUND

Storage Policies

Several solutions for storing XML data have been proposed both in the scientific literature and in commercial products. In particular, storage approaches can be classified with respect to the type of system on which they rely and on the used XML document representation model. In this framework, they can be categorized as follows:

- Relational DBMS: Uses a collection of tables to represent both data and relationships among these data. More specifically, in order to represent XML data by using tables, it is necessary to break down the XML documents into rows and columns. The tree-like structure of XML facilitates both their decomposition and storage in relational tables. However, this process is expected to cause some performance overhead mainly due to the continuous translation of trees to tables (and vice versa). Due to its popularity, several models have been proposed to store XML documents in relational DBMSs (e.g., Bohannon, Freire, Roy & Simeon, 2002; Khan & Rao, 2001; Schmidt, Kersten, Windhouwer & Waas, 2000; Tian, DeWitt, Chen & Zhang, 2000; Zhu & Lu, 2001).
- Object-Oriented (O-O) DBMS: XML documents are stored as collections of object instances, using relationships based on the O-O idea (McHugh, Abiteboul, Goldman, Quass & Widom, 1997). O-O DBMSs have been designed to work well with object programming languages such as C++, C#, and Java. Inheritance and object-identity are their basic characteristics. However, O-O DBMSs cannot easily handle data with a dynamic structure since a new

class definition for a new XML document is needed, and the use of O-O DBMSs for XML document storage is not as efficient and flexible.

- Object-Relational (O-R) DBMS: The XML documents are stored in a nested table, in which each tag name in DTD (or XML schema) corresponds to an attribute name in the nested table. Currently, researchers have showed a steadily increasing interest in O-R DBMSs, since they combine both the benefit of the relational maturity and the richness of O-O modeling (Pardede, Rahayu & Taniar, 2004). In O-R DBMSs, the procedure for storing XML data to relation mapping is modeled by an O-R model. In this context, a number of transformation steps from the XML schema to an O-R model are presented (Widjaya, Taniar & Rahayu, 2004). More specifically, each nested XML element is mapped into an object reference of the appropriate type. Then, several mapping rules are indirectly embedded in the underlying model.
- Native XML DBMS: In this case, the XML document is the fundamental unit of storage (Kanne & Moerkotte, 2000). Therefore, a native XML database defines a (logical) model for an XML document, and stores and retrieves documents according to that model. In particular, two basic steps are involved for storing the XML documents on a native XML DBMS: (1) the data are described by its structure (DTD or XML schema), and (2) a native database XML schema (or a data map) is defined. However, the native XML DBMSs have not yet become very popular, since these systems must be built from scratch.
- LDAP Directories: XML documents are stored in LDAP directories which can be considered as a specialized database (Marron & Lausen, 2001). Therefore, the internal storage model of this database system is defined in terms of LDAP classes and attributes. More details about the architecture of the LDAP model and protocol are discussed by Howes, Smith, and Good (1999). Comparing the LDAP directories with the typical DBMSs, they are more widely distributed, more easily extended, and replicated on a higher scale.
- File Systems: Since an XML document is a file, a typical storage approach is to store it simply as a flat file. In particular, this approach uses a typical file-processing system, supported by a conventional operating system (as a basis for database applications). The wide availability of XML tools for data files results in a relatively easy accessing and querying of XML data (which are stored in files). By using a flat file for XML data, we have faster storing (or retrieving) of whole documents. However, this storage format has many disadvantages, such as difficulty in accessing and updating data, since the only way to accomplish this is to overwrite the whole file (Silberschatz et al., 2002).

Access Control Policies

In order to protect XML files, we need to design authorizations defining which client (subject) can access which protected resource (object) and under which mode. Since XML files are organized according to DTDs or XML schemas, protected resources can be both XML files, DTDs, or schemas, or parts of them, such as a specific element or even attribute.

The basic access control models are:

- Discretionary Access Control (DAC): Every subject and object is enumerated, and there are authorizations connecting each subject and object. The owner is responsible for defining policies in order to protect his/her resources, and (s)he can also define who is going to be assigned some access privileges. It is the most flexible and simple access control model. However, it does not provide high levels of protection, and it cannot be used in case multiple security levels are required.
- Mandatory Access Control (MAC): This is based on the existence of one central administrator responsible for the definition of policies and authorizations. It is expressed with the use of security labels associated with both subjects and objects. Every object has a classification label defining its sensitivity, and each subject is assigned a clearance label defining its trustworthiness. This model is more secure than DAC, but it cannot be used in wide distributed Internet-based environments. Therefore, their usage is gradually decreased.
- Role-Based Access Control (RBAC): The most widely used access control model in modern environments (Sandhu, Coyne & Feinstein, 1996). Subjects are assigned roles that are categorizations of subjects according to their duties in an organization. Thus, every subject is assigned roles that in their part have some authorizations. Therefore, the number of the needed policies is highly decreased. RBAC model is a super set, since it can also express DAC and MAC policies. It is appropriate for distributed heterogeneous networks offering Internet access. Moreover, it is recommended to protect hypertext documents (HTML or XML files). Lately, a generalized RBAC model has been proposed where objects and environmental conditions are assigned roles (Moyer & Ahamad, 2001).

FUTURE TRENDS

Implementations

Some of the most popular database vendors (like IBM, Microsoft, and Oracle) have developed (mainly) database tools for the storage of XML documents, and several storage techniques have been adopted in order to maximize functionality and performance. Furthermore, for reasons of integrity, the majority of the DBMSs are supported by access control mechanisms.

In general, the most well-known software tools that employ XML document storage and access control can be categorized in the following types of databases:

- XML-Enabled DBMSs: These provide various interfaces for extracting structured data from XML documents and then to be stored in DBMSs (according to their model). Table 1 presents the basic products of such an approach.
- Native XML DBMSs: XML documents are stored in XML internally in the database. Table 2 shows some of the most popular DBMSs.

Currently, the most widely adopted technology to enforce the security guarantees of the XML-based databases is the Microsoft (MS) .NET platform. MS .NET has been recently proposed and has been adjusted to support XML Web Services. Its security mechanism is adopted by MS SQL Server 2000 (Rys, 2001), Oracle 9i, and DB2, which have been designed to cooperate with Microsoft .NET technology.

All of the above storage systems can be supported by access control modules. Most of them support the Microsoft .NET access control framework. The idea of protecting XML files is quite new, and therefore most of the implemented systems still belong to the scientific literature and they have not stepped into the market.

The core of every access control mechanism is a language used for expressing all of the various components of access control policies, such as subjects, objects, constraints, and so forth. To date several languages have been proposed. Since XML is a structured language, it can be used in the definition of access control issues, such as roles, policies, and authorizations. Therefore, the most well-known general-purpose XML-based access control languages are:

- XACML (eXtensible Access Control Markup Language): This is an OASIS standard that defines both a policy language and an access control decision request/response language (both written in XML) (www.oasis-open.org/committees/xacml/repository). The policy language is used to describe general access control requirements, and has standard extension points for defining new functions, data types, combining logic, and so forth. The request/response language permits forming a query in order to ask whether a given action should be allowed (or not), and to interpret the result that defines the response.
- XrML (eXtensible rights Markup Language): This is a language for digital rights management (Wang et al.). By using XrML, the owner of a digital resource can specify who can use the resource (principal) and under what circumstances (constraints). The owner thus produces licenses containing such information that are then given to the appropriate principal. Since XrML is mainly used for protecting digital media, it contains schemas defining the metadata of such resources, such as audio or video files. Of course, it can be easily extended for controlling access to any other type of resource.

Table 1. Storage in XML-enabled databases

Product	DBMS Model	Storage
Oracle 9i *technet.oracle.com*	Object-Relational	XML documents are stored as relational tables.
MS SQL Server 2000 *www.microsoft.com*	Object-Relational	Each XML document is stored as a relational table, and an element is created for each row.
IBM's DB2 *www-4.ibm.com*	Object-Relational	XML documents are stored as relational tables.
XIS *www.exceloncorp.com/xis/*	Object-Oriented	XML documents are stored in a B-tree-like structure.
Lore (McHugh et. al., 1997)	Object-Oriented	XML documents are stored using an Object Exchange Model (OEM).
SHORE (Hess, Schulz & Brossler, 2000)	Object-Oriented	This stores information extracted from XML documents using a variant of R-trees and B-trees structure.

S

Table 2. Storage in native XML databases

Product	Storage Model
NATIX (Kanne & Moerkotte, 2000)	Tree structure: XML documents are split into sub-trees (basic storage and query units) based on certain criteria.
SODA *www.cse.unsw.edu.au/~soda/*	Tree structure: XML documents are stored in a single tree, which preserves all XML information.
Xyleme *www.xyleme.com*	Tree structure: XML documents are stored as trees until a certain depth where byte streams are used.
Ipedo *www.ipedo.com/html/products.html*	Collection structure: XML documents are organized into collections (like directories), which can be typed or un-typed.
eXist *exist.sourceforge.net*	Collection structure: XML documents are stored using a DOM-tree (built from SAX-events) and are organized into hierarchical collections, similar to storing files in a file system.
Tamino (Schoning, 2001)	Collection structure: Each XML document is stored in exactly one collection.
Xindice *www.dbxml.org*	Collection structure: Each XML document is stored in at least one collection (a collection can be created either consisting of documents of the same type, or a single collection can be created to store all documents together).

Table 3. Examples of access control tools

Product	Type	Access Control
XENA (Tan & Zhao, 2000)	Non XML-Based	Centralized tool protecting XML repositories using relational database technology.
Author-X (Bertino, Castano & Ferrari, 2001)	XML-Based	Java-implemented tool, supporting varying security granularity levels. PULL and PUSH dissemination technology.
ACP (Access Control Processor) (Damiani, De Capitani di Vimercati, Paraboschi & Samarati, 2000)	XML-Based	Decentralized system supporting distributed XML repositories by using both local and global authorizations.
ProvAuth (Kudo & Hada, 2000)	XML-Based	Access control environment where access is allowed with some provisions. It uses XACL.

- ODRL (Open Digital Rights Management): This is a language used for Digital Rights Management (DRM) in open environments (odrl.net/1.1/). It can express rights for protecting any digital content, like electronic publications, digital images, audio and movies, learning objects, computer software, and so forth.

Except for the languages, research has to present several access control tools protecting XML files. Those can be categorized into non XML-based and XML-based. Table 3 contains some of the tools belonging to both categories.[1] The only tool using a standardized access control language is ProvAuth (Provisional Authorizations) (Kudo & Hada, 2000). As its name suggests, access is allowed to users only after (or before) the execution of some provisional actions, like logging the session, encrypting the view of a protected document, and so forth. To support such functionality, a specific-scope XML access control language (XACL) has been introduced which is part of the IBM XML Security Suite. Due to its extensibility and flexibility, the model can be integrated into a conventional Web application consisting of clients and servers.

CONCLUSION

This article has presented an overview for storage and access control of XML documents. It is important to indicate that no definite guidelines have yet been proposed for selecting an optimal solution when storing and securing XML documents. In order to improve the management of XML documents, some issues should require further research. In particular, the storage of XML documents may be improved by using some data mining techniques (e.g., specific clustering algorithms). Furthermore, the XML management techniques should further extend existing access control policies, in order to improve the security in XML documents accessing. Finally, the commercial DBMSs should be extended to support more sophisticated storage and access control techniques such as integrated methods for locating and accessing dynamic XML documents.

REFERENCES

Amer-Yahia, S. & Fernandez, M. (2002). Techniques for storing XML. *Proceedings of the 18th International Conference on Data Engineering* (ICDE 2002), San Jose, California, USA.

Bertino, E., Castano, S. & Ferrari, E. (2001). Securing XML documents with Author-X. *IEEE Internet Computing*, *5*(3), 21-31.

Bohannon, P., Freire, J., Roy, P. & Simeon, J. (2002). From XML schema to relations: A cost-based approach to XML storage. *Proceedings of the 18th International Conference on Data Engineering* (ICDE 2002), San Jose, California, USA.

Damiani, E., De Capitani di Vimercati, S., Paraboschi, S. & Samarati, P. (2000). Designing and implementation of an access control processor for XML documents. *Proceedings of 9th World Wide Web Conference (WWW9) and Computer Networks*, Amsterdam, Holland.

Hess, A., Schulz, H. & Brossler, P. (2000). *SHORE—a hypertext repository based on XML*. Technical Report, SD&M Corporation, Software Design and Management, Southfield, USA.

Howes, T.A., Smith, M.C. & Good, G..S. (1999). *Understanding and deploying LDAP directory services*. Macmillan Technical Publishing.

Joshi, J.B.D., Aref, W.G., Ghafoor, A. & Spafford, E.H. (2001). Security models for Web-based applications. *Communications of the ACM*, *44*(2), 38-44.

Kanne, C.C. & Moerkotte, G. (2000). Efficient storage of XML data. *Proceedings of the 16th International Conference on Data Engineering* (ICDE 2000), San Diego, California, USA.

Khan, L. & Rao, Y. (2001). A performance evaluation of storing XML data in relational DBMS. *Proceedings of the 3rd ACM CIKM'01 Workshop on Web Information and Data Management* (WIDM 2001) (pp. 31-38), Atlanta, Georgia, USA.

Kudo, M. & Hada, S. (2000). XML document security based on provisional authorization. *Proceedings of the 7th ACM Conference on Computer and Communications Security* (pp. 87-96), Athens, Greece.

Marron, P.J. & Lausen, G.. (2001). On processing XML in LDAP. *Proceedings of the 27th Conference on Very Large Data Bases* (VLDB 2001) (pp. 601-610), Roma, Italy.

McHugh, J., Abiteboul, S., Goldman, R., Quass, D. & Widom, J. (1997). Lore: A database management system for semi-structured data. *ACM SIGMOD Record*, *26*(3), 54-66.

OASIS eXtensible Access Control Markup Language Technical Committee. *eXtensible Access Control Markup Language (XACML) Committee Specification, Version 1.1*. Retrieved February 15, 2004, from *www.oasis-open.org/committees/xacml/repository*

Open Digital Rights Language Initiative. *Open Digital Rights Language (ODRL) Specification, Version 1.1*. Retrieved February 15, 2004, from *odrl.net/1.1/*

Pallis, G., Stoupa, K. & Vakali, A. (2004). *Storage and access control issues for XML documents*. In D. Taniar & W. Rahayu (Eds.), *Web information systems* (Chapter 4, pp. 104-140). Hershey, PA: Idea Group Publishing.

Pardede, E., Rahayu, J.W. & Taniar, D. (2004). On using collection for aggregation and association relationships in XML object-relational storage. *Proceedings of the 19th ACM Symposium on Applied Computing* (SAC 2004), Nicosia, Cyprus.

Rys, M. (2001). Bringing the Internet to your database: Using SQL Server 2000 and XML to build loosely coupled systems. *Proceedings of the 17th International Conference on Data Engineering* (ICDE 2001) (pp. 465-475), Heidelberg, Germany.

Sandhu, R.S., Coyne, E.J. & Feinstein, H.L. (1996). Role-based access control models. *IEEE Computer*, *29*(2), 38-47.

Schmidt, A., Kersten, M., Windhouwer, M. & Waas, F. (2000, May). Efficient relational storage and retrieval of XML documents. *Proceedings of the 3rd International Workshop on the Web and Databases* (WebDB 2000) (pp. 137-150), Dallas, Texas.

Schoning, H. (2001). Tamino—a DBMS designed for XML. *Proceedings of the 17th International Conference on Data Engineering* (ICDE 2001) (pp. 149-154). Heidelberg, Germany.

Silberschatz, A., Korth, H. & Sudarshan, S. (2002). *Database system concepts* (4th Edition). McGraw-Hill.

Tan, K. & Zhao, J. (2000). Effective storage and access control of XML documents with relational database system. Retrieved November 16, 1997, from *xena1. ddns.comp. nus.edu.sg:8080 /Xena/paper.pdf*

Tian, F., DeWitt, D.J., Chen, J. & Zhang, C. (2002). The design and performance evaluation of alternative XML storage policies. *ACM SIGMOD Record*, *31*(1), 5-10.

Wang, X., Lao, G., DeMartini, T., Reddy, H., Nguyen, M. & Valenzuela, E. XrML-eXtensible Rights Markup Language. *Proceedings of the ACM Workshop on XML Security* (pp. 71-79), Fairfax, Virginia.

Widjaya, N.D., Taniar, D. & Rahayu, J.W. (2004). Transformation of XML schema to object relational database. In D. Taniar & W. Rahayu (Eds.), *Web information systems* (Chapter 5, pp. 141-189). Hershey, PA: Idea Group Publishing.

Zhu, Y. & Lu, K. (2001). An effective data placement strategy for XML documents. *Proceedings of the 18th British National Conference on Databases* (BNCOD 2001) (pp. 43-56), Chilton, UK.

KEY TERMS

Database Management System (DBMS): A collection of interrelated data that is called a database, and a variety of software tools for accessing those data. The three leading commercial DBMSs are: Oracle 9i, IBM DB2, and Microsoft SQL Server.

Discretionary Access Control (DAC): A model where the owner is responsible for defining policies in order to protect his/her resources and (s)he can also define who is going to be assigned some access privileges.

DTD: A set of rules defining the element types that are allowed within an XML document, and specifying the allowed content and attributes of each element type. Also defines all the external entities referenced within the documents and the notations that can be used.

Lightweight Directory Access Protocol (LDAP): A protocol definition for accessing specialized databases called directories. An LDAP directory is organized by a simple "tree" hierarchy.

Mandatory Access Control (MAC): A model expressed with the use of security labels associated with both subjects and objects.

Native XML Database: A database that defines a model for an XML document (as opposed to the data in that document), and stores and retrieves documents according to that model. At a minimum, the model must include elements, attributes, PCDATA, and document order. Examples of such models are the XPath data model, the XML Infoset, and the models implied by the DOM and the events in SAX.

Role-Based Access Control (RBAC): An access control model where subjects are organized into roles defining their duties in an organization or environment.

XML Document: A document consisting of an (optional) XML declaration, followed by either an (optional) DTD or XML schema, and then followed by a document element.

XML Schema: An alternative to DTDs; a schema language that assesses the validity of a well-formed element and attribute information items within an XML document. There are two major schema models: W3C XML Schema and Microsoft Schema.

ENDNOTE

[1] The names of the tools, ACP and ProvAuth, are assigned by us for reasons of brevity.

Strategic Alignment of Organizational Strategies

Sam Lubbe
University of KwaZulu-Natal, South Africa

INTRODUCTION

Alignment is important to all organizations. This is confirmed by King (1995) but it is not yet clear how to achieve harmony between business strategy, information technology (IT) and e-commerce and what the impact of this alignment would be on an organization. In the past, the IT department was responsible for planning, development and management of information systems (IS). With the convergence of e-commerce, gaining an understanding of how e-commerce and IT may be jointly employed to support organizational strategies is important for managers (it is acknowledged that some organizations place e-commerce in the IT department, while others have it as a separate department). Good practice will ensure that the organization will find the best place for it. It is, however, good policy to split e-commerce and IT staff while aligning strategies. It is necessary to align the strategy for the e-commerce department as well. This alignment could be either "lawful" (wishes to create and protect) or chaotic (wishes to destroy).

Strategic alignment is defined by Papp and Fox (2002) as the use of IT in the integration and development of business strategies and corporate goals. Alignment is a dynamic process that requires close, continual assessment (because the goals keep on moving) and cooperation between achieving competitive advantage and surviving.

This article therefore discusses the changing environment of the organization in terms of strategic planning. Managers of tomorrow must understand what e-commerce is, how the approach to this concept will be, and how it will affect the leverage of the organization. The article will briefly provide a discussion of the topic and a perspective on the issues and problems as they relate to alignment. It will also briefly look at future trends and will end with a definition of some terms.

BACKGROUND

The establishment of alignment between IT and organizational objectives has consistently been reported as one of the key concerns of IS managers (Lubbe, 2001; Reich & Benbasat, 2000; Rodgers, 1997). Findings by authors (e.g., Dhillon & Orton, 2001; Hamilton & Chervany, 1981; Henderson & Sifonis, 1988; Moad, 1994; Papp, 1997; Reich & Benbasat, 2000) suggest that both practitioners and researchers should direct their efforts toward understanding of domain knowledge. On the other hand, Venkatraman (2000) argues that managers must also align their vision to dotcom to ensure that the leaders of the Industrial Age do not become the dinosaurs of the dotcom era.

Alignment should be based on different levels; that is, it can be based on the shared domain knowledge between business and IT executives, IT implementation success, communication between business and IT executives and connections between business, IT and e-commerce planning processes (Reich & Benbasat, 2000; Papp et al., 1996). Venkatraman (2000) argues that the Internet changes everything – particularly for brick and mortar organizations, branded products and services, and traditional supplier and customer relationships.

Papp (1997) argues that executives rarely agree and that they normally oppose each other while meeting to discuss alignment. These executives, however, acknowledge that there are different approaches to alignment, but are not willing to give any leeway. Each industry will also have a different type of dimension to its alignment that impacts its architecture through changes in e-commerce, IT and the business infrastructure. Therefore, with the advent of e-commerce another level had been added that impacts on effective planning of strategies.

Amadi (1998) and Cerpa and Verner (1998) state that strategic alignment and all its dimensions give value to the business on all levels of the business. Bildfell (2003) also noted that many planes or dimensions should be kept in mind while aligning strategies. These dimensions could be architectural alignment, cultural alignment, customer alignment, and so forth. Dhillon and Orton (2001) argue that strategic options such as cost leadership, differentiation, strategic alliances and globalization could further add on dimensions to the strategic alignment of the organization. They added that there will be some schizoid incoherence and that management should pay attention.

The relationship between peers inside the organization helps management visualize how alignment has been achieved within the organization. Enns et al. (2003) argue that the relationship between CIOs and their peers should successfully affect the visualization. The degree of mutual understanding of current objectives (short-term alignment) and the congruence of IT and e-commerce vision (long-term alignment) between IT, e-commerce and IT executives affects alignment as well. Alignment as a key should maximize the value and impact of the organization's IT and e-commerce investments and be flexible enough to allow the organization to act quickly to ensure that changes are taken into account.

Chan (2000) also argues that communication and understanding are factors that should be kept in mind on a business level. There should also be linked executive commitment to IS and e-commerce issues and initiatives. These links contribute to missions and dimensions for e-commerce, IT and the organization.

FUNCTIONAL INTEGRATION

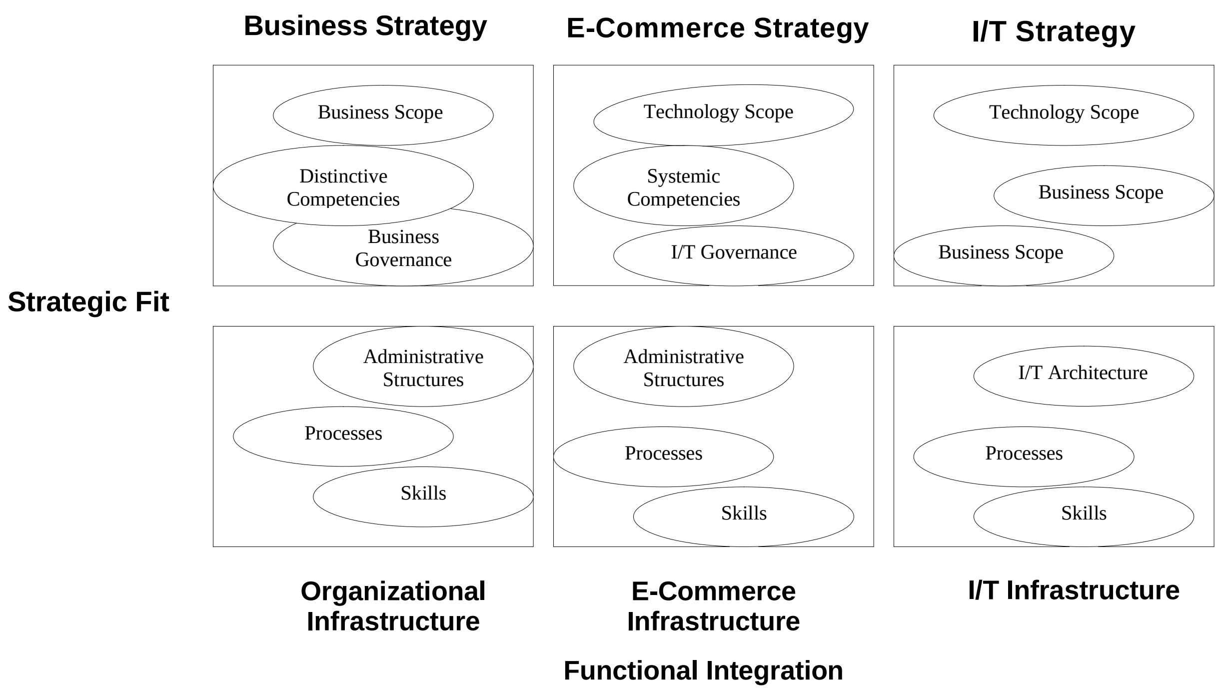
Some of the functions that should be paid attention to while attempting to align strategies are the shared domain knowledge between business, IT and e-commerce executives; IT and e-commerce implementation success; communication between business, e-commerce and IT planning process; and connections between business, e-commerce and the IT planning process. All the factors mentioned in this instance were found to influence short-term alignment, while only shared domain knowledge was found to influence long-term alignment. Reich and Benbasat (2000) established that a new factor, strategic business plans, affects both short-term and long-term alignment (supported by Dhillon & Orton, 2001).

All management functions must be aligned – especially for organizations that run e-commerce operations. If there is no alignment, then the dotcom strategy may be "hijacked" by staff members for their own purpose (Venkatraman, 2000). The functional areas will influence the ranking of enablers and inhibitors of strategic planning (Papp & Fox, 2002). The type of business (brick-and-mortar and/or e-commerce) will have a bearing on the alignment perspective of the organization. Bildfell (2003), on the other hand, argues that there are three functional areas that should be kept in mind while aligning strategies. These are relationships that accelerate the strategy; relationships that help penetrate new markets and relationships that improve competitiveness and solution capabilities.

INSTRUMENTS TO MEASURE ALIGNMENT

The Gartner Group had designed an instrument that could be used for measurement of assignment called Alignment

Figure 1. The strategic alignment model (Adapted from Papp & Fox, 2002)



Strategy Assessment (Business Wire, 1998). The instrument analyzes five key areas: IT application alignment with business goals; IT application alignment with business operations; IT infrastructure alignment with business goals; IT infrastructure alignment with business operations; and IT infrastructure alignment with IT applications. There are also other instruments (e.g., questionnaires, spreadsheets, etc.) as explained by Lubbe (2001). The instrument by the Gartner Group enables the CIOs to assess the effectiveness of IT support for key areas of business and to prioritize IT resources and applications so that they are in alignment with the organization's business goals and operations. The problem is to align e-commerce goals with IT and the organization still a concern for CIOs.

Amadi (1997) states that BPAR is a tool that can help with the visualization of alignment. Segars and Grover (1999) argue that instruments measuring alignment should take into account the different dimensions of alignment. These dimensions, according to them, include comprehensiveness, leverage, formalization, focus, and consistency. Any e-commerce or IT person needs to understand the leverage points of the industry, the history and current issues of the leverage effect, and to learn to apply "common sense" of the business units in the application of technology to business problems. Venkatraman (2000) argues that organizations should enter into alliances to explore a wide range of opportunities and leverage those that succeed. This is confirmed by Lubbe (2001). Papp and Fox (2002) support the importance of the strategic alignment by asking: Are the business strategies and plans leveraging IT?

ISSUES TO BE KEPT IN MIND

Organizations should identify strengths and weaknesses that affect alignment. Some organizations do believe that their strategies are aligned (50%) and 42% said that they were not aligned. This is a problem and could affect the flashes of commercial insight (gut feeling) that some managers exhibit. Some of the enablers and inhibitors of alignment are listed next. These were also listed and demonstrated by Moad (1994). Bildfell (2003) argues that successful alignment, keeping problems and success stories in mind, ensures that organizations will be competitive. Some enablers to alignment are important.

FUTURE TRENDS

Because e-commerce is evolving rapidly, organizations must continuously expand current business models (e.g., improve marketing) while experimenting to create new ones (Venkatraman, 2000). These new business modules will ensure that the goal posts for competitors keep on changing and ensure that the organizations stay competi-

Table 1. A summary of enablers and inhibitors of alignment of strategies

Some Enablers to Alignment
Executive support for IT and e-commerce
Strategy developed in tandem
Leadership shown by IT and e-commerce
E-commerce and firm's resources are shared
Close relationship between firm, e-commerce and IT
Some Inhibitors to Alignment
Prioritize work poorly
No close relationship with e-commerce and IT
IT and e-commerce do not know their customers
No executive support for IT and e-commerce
Strategic goals not achieved by IT and e-commerce
No communication

tive. Papp and Fox (2002) argue that there could be differences between the strategies of brick-and-mortar operations and normal business operations but that it has not been defined as yet.

CONCLUSION

IT and e-commerce groups have to earn the right to play a meaningful role in management forums. One important factor that the theory suggested is that IT and e-commerce people should devote the time necessary to develop shared domain knowledge, and thereby be able to provide some input on alignment. The organization should try to be "lawful" and not chaotic and will survive. Managers should remember that there could be alignment but still be problems preventing proper alignment.

REFERENCES

Amadi, A.A.N. (1998). Tying strategic alignment and it value to business success using business process analysis and redesign (BPAR). *AMCIS*.

Bildfell, D. (2003, May). A Websphere approach to strategic market development: Alignment of corporate goals and cultures is key. *IBM Partnerworld.*

Business Wire. (1998). *GartnerMeasurement announces alignment strategy assessment: The new service measures IT support of strategic business objectives.* Gale Group.

Cerpa, N., & Verner, J.M. (1996). Information systems strategic planning: A longitudinal case study. *AMCIS.*

Chan, Y.E. (2000, Spring). IT value: The great divide between qualitative and quantitative and individual and organizational measures. *Journal of Management Information Systems, 16*(4), 225-262.

Dhillon, G., & Orton, J.D. (2001). Schizoid incoherence, micro-strategic options, and the strategic management of new organizational forms. *M@n@gement, 4*(4), 229-240.

Enns, H.G., Huff, S.L., & Higgins, C.A. (2003, March). CIO lateral influence behaviors: Gaining peers' commitment to strategic information systems. *MIS Quarterly, 27*(1), 155-176.

Hamilton, S., & Chervany, N.L. (1981, September). Evaluating information systems effectiveness- Part 1: Comparing evaluation approaches. *MISQ, 5*(3).

Henderson, J.C., & Sifonis, J.G. (1988, June). The value of strategic IS planning: Understanding, consistency, validity, and IS markets. *MISQ,* 187-200.

Lubbe, S.I. (2001). The theory behind the role of leverage and the strategic alignment of organizations while creating new markets (Internet marketing and e-commerce). In O. Lee (Ed.), *Internet marketing research: Theory and practice* (pp. 187-208). Hershey, PA: Idea Group Publishing.

Moad, J. (1994, January). IS rises to the competitiveness challenge. *Datamation,* 16-24

Papp, R. (1997). *Strategic alignment: Firm/industry assessment.* AIS Conference.

Papp, R., & Fox, D.R. (2002). Information strategy development: The strategic alignment imperative. *Eighth Americas Conference on Information Systems* (pp. 1321 – 1325).

Papp, R., Luftman, J., & Brier, T. (1996). *Business and IT in harmony: Enablers and inhibitors to alignment.* Americas Conference on IS.

Reich, B.H., & Benbasat, I. (2000, March). Factors that influence the social dimension of alignment between business and information technology objectives. *MIS Quarterly, 24*(1), 81-113.

Segars, A.H., & Grover, V. (1999, September). A study of profiles in strategic information systems planning. *Information Systems Research, 10*(9).

Venkatraman, N. (2000, Spring). Five steps to a dot-com strategy: How to find your footing on the Web. *Sloan Management Review*.

KEY TERMS

Alignment: A dynamic process that requires close, continual assessment (because the goals keep on moving) and cooperation between achieving competitive advantage and surviving.

Competitive Advantage: Usually refers to characteristics that permit a firm to compete effectively with other firms due to low cost or superior technology, perhaps internationally.

Enablers: A factor that makes something possible (e.g., Alignment is an enabler for organizations to cut production costs by half).

Inhibitors: A factor that can stop an activity.

Leverage: Investing with borrowed money as a way to amplify potential gains (at the risk of greater losses).

Strategic Alignment: As the appropriate use of IT in the integration and development of business strategies and corporate goals.

Strategy: A detailed plan for achieving success in situations such as war, politics, business, industry or sport, or the skill of planning for such situations.

S

Strategic Experimentation and Knowledge Management

V.K. Narayanan
Drexel University, USA

INTRODUCTION

Historically, the focus of IT infrastructure has been to capture the knowledge of experts in a centralized repository (Davenport & Prusak, 1998; Grover & Davenport, 2001). These centralized databases contained knowledge that was explicit and historical (e.g., competitor pricing, market share), and the IT infrastructure served to facilitate functional decision-making or to automate routine tasks (i.e., in re-engineering). The users of technology approached the repository to obtain data in a narrowly defined domain (Broadbent et al. 1999). Consequently, IT originally played a significant yet ultimately limited role in the strategy creation process. Management information systems (MIS) arguably generated information that was less applicable to strategy creation, as noted in early writings on the linkage between MIS and strategic planning (Holmes, 1985; Lientz & Chen, 1981; Shank et al., 1985).

The active management of knowledge was similarly underdeveloped. Despite the fact that strategic decision makers had always emphasized the role of tacit knowledge, the actual importance of knowledge was not *explicitly* recognized. Formalized knowledge management (KM) (Davenport & Prusak, 1998), with its associated terminology and tools, is a recent development and, as such, did not inform the strategic planning process.

However, the shifts that have taken place in IT infrastructures over the last decade and the recent developments in knowledge management have brought them closer to the creators of strategy. Indeed, both IT and knowledge management are increasingly enablers in the contemporary strategic management practice.

1. IT infrastructure is transitioning in its focus from the functional work unit to a process orientation. Whereas computer systems were once the focal point, the new infrastructure is network-centric, with an emphasis on business knowledge (Broadbent et al., 1999). For example, traditional search engines utilized rule-based reasoning to identify elements matching specific search criteria; the "state-of-the-art" knowledge management systems employ case-based search techniques to identify all relevant knowledge components meeting the user's request (Grover & Davenport, 2001).
2. IT now takes into account contexts that include cross-functional experts that are knowledgeable in a wide variety of potentially relevant issues. Additionally, there is a greater emphasis on the integration of infrastructure with organization, structure, culture (Gold et al., 2001), and organizational roles (Davenport & Prusak, 1998). In many ways, the newer IT infrastructures have enabled the garnering of explicit knowledge throughout the organization improving the speed of strategy creation.

The objective of this article is to outline how the developments in IT and KM are facilitating the evolution of strategic management to strategic experimentation in order to create quantum improvements in strategy creation and unprecedented developmental opportunities for the field of IT.

BACKGROUND

Information Technology (IT)

For the purposes of this chapter, IT is defined as physical equipment (hardware), software, and telecommunications technology, including data and image and voice networks employed to support business processes (Whitten & Bentley, 1998). The overarching plan for IT deployment within an organization is called the *IT architecture*. Technology infrastructure refers to the architecture as including the physical facilities, the services, and the management that support all computing resources in an organization (Turban et al., 1996).

Knowledge Management (KM)

As used in this chapter, data are objective, explicit pieces or units; information is data with meaning attached; and knowledge is information with an implied element of action.

Knowledge is the fluid mix of framed experience, values, contextual information, and expert insight that provides a framework for evaluating and incorporating new experiences and information. It originates and is applied in the minds of knowers. In organizations, it often becomes embedded not only in documents or repositories but also in organizational routines, processes, practices, and norms (Davenport & Prusak, 1998, p. 5).

KM is "a set of business practices and technologies used to assist an organization to obtain maximum advantage from one of its most important assets—knowledge" (Duffy, 2000, p. 62). In other words, it is actively capturing, sharing, and making use of what is known, both tacitly, informally, and explicitly, within the organization. IT often facilitates knowledge management initiatives by integrating repositories (e.g., databases), indexing applications (e.g., search engines), and user interfaces. Davenport and Prusak (1998) note that KM also incorporates traditional management functions: building trust among individuals, allocating resources to KM, and monitoring progress.

Strategic Management

The concept of "strategy" explicated in strategic management is one of marketplace strategy (i.e., winning in the marketplace against competitors, entrenched or incipient). The underlying premise is that "to enjoy continued strategy success, a firm must commit itself to outwitting its rivals" (Fahey & Randall, 2001, p. 30). A large body of literature on strategic management has persuasively argued that effective strategy creation and execution are central to a firm's performance (Covin et al., 1994).

Strategy creation involves both goal formulation—defined in terms of external stakeholders rather than operational milestones—and crafting of the strategic means by which to accomplish these goals (Hofer & Schendel, 1978). The means typically include business scope, competitive posture, strategic intent, and the organizational mechanisms for implementation. In practice, the process of strategy creation has often taken the form of strategic planning. Comprehensive strategic planning (Gluck et al., 1978) has historically been practiced in large corporations. A celebrated example is the use of scenarios by Royal-Dutch Shell. Planning usually consisted of several sequential stages of decision-making involving diagnosis, alternative development, evaluation and choice, and implementation. In each step, the strategic planners emphasized deliberate juxtaposition of "objective data" and careful analysis with top management judgment, thus highlighting the role of tacit knowledge.

Strategic planning has evolved over the years. Writing in the 1970s, Gluck et al. (1978) identified four phases of evolution: budgeting, long-range planning, strategic planning, and strategic management. Each phase of evolution incorporated the lessons from the earlier phases, but also took into account the emerging realities faced by corporations. Gluck et al. (1978) noted that during the 1980s the "strategic management" phase would represent the cutting edge of practice in the world.

TOWARD STRATEGIC EXPERIMENTATION

The 1990s witnessed a revolution in organizational environments often characterized as "hypercompetition" (D'Aveni, 1994). These environments have created three major imperatives for organizations: time compression, globalization, and technology integration (Narayanan, 2001). The increased environmental dynamism also contributes to an increase in the degree of uncertainty confronted by strategic managers, calling into question traditional planning practices. Consequently, a new type of strategy creation process is evolving, which is termed "strategic experimentation." With this evolution, the relationship between strategy creation, knowledge management, and IT is undergoing a profound shift.

All four phases of strategic planning documented by Gluck et al. (1978) incorporated a sequential approach to strategy creation and execution, leading to the identification of one winning strategy that has the highest probability of success. Consequently, firms found it logical to commit the maximum available resources to the implementation of one wining strategy. The goal was to obtain a sustainable competitive advantage vis à vis the firm's rivals, and to reduce uncertainty ex ante using analytical forecasting techniques as well as market research. This approach to planning seems to have been effective during the 1980s when the environment was moderately dynamic.

In hypercompetitive environments, market participants frequently confront great uncertainty over technological possibilities, consumer preferences, and viable business models. This high level of ambiguity often results in a situation where (a) traditional methods of ex ante uncertainty reduction (e.g., market research) fail, and (b) the costs and risks of the traditional "big bet" strategic management approach outweigh the advantages in terms of focus, decisiveness, and concentrated resource commitment. It is in this situation that the emerging strategic experimentation approach holds significant promise.

Strategic experimentation (Brown & Eisenhardt, 1998; McGrath, 1998; McGrath & MacMillan, 2000) draws on real-options reasoning (McGrath, 1997), discussions of exploration vs. exploitation, and trial-and-error learning (Van de Ven & Polley, 1992).

1. Companies engaging in strategic experimentation continually start, select, pursue, and drop strategic initiatives before launching aggressively those initiatives whose values are finally revealed (McGrath & MacMillan, 2000).
2. Strategic initiatives serve as low-cost probes (Brown & Eisenhardt, 1998) that enable the discovery of product technology and market preferences. They also serve as a stepping stone option for future competitive activity in that particular product-market domain.
3. The role of the strategic manager is to administer a portfolio of strategic initiatives that represent an appropriate mix of high and low uncertainty projects, and to maximize the learning from these real options (McGrath & MacMillan, 2000).

Strategic experimentation represents a fundamentally different view of the practice of strategic planning and the path to competitive advantage. Movement is emphasized over position in this approach. Thus, competitive advantage is viewed as temporary at best, and hence, innovation and learning are considered crucial to success. Strategic experimentation is especially appropriate for high velocity environments such as emerging product markets with high uncertainty surrounding both technology and customer preferences (e.g., the early Personal Digital Assistant, Internet appliance, and satellite-based telephony markets). Here, low-cost probes can be very effective in gaining knowledge and reducing uncertainty while minimizing exposure to the results of faulty assumptions.

THE ROLE OF IT AND KNOWLEDGE MANAGEMENT IN THE ERA OF STRATEGIC EXPERIMENTATION

Since strategic experimentation represents the cutting edge of ideas in strategic management, we should expect significant advances in tool development and utilization in the next few years that will enable us to move the idea towards normal organizational practice.

Strategic experimentation necessitates several major functions that should be performed by an organization. KM is critical in strategic experimentation; therefore, it is not surprising that many of the tools currently moving into practice have emerged from KM. Following are the four major strategic experimentation functions and the associated KM tools.

Rapid Decision-Making

The ability to *quickly* garner tacit knowledge in all phases of decision-making is a central requirement in strategic experimentation. Current KM tools to support this include visualization and prototyping, group decision facilitation, and knowledge representation. Each method attempts to reduce the time needed for a group to progress from problem identification to solution implementation. These tools help to coordinate the use of data, systems, tools, and techniques to interpret relevant information in order to take action (Little, Mohan, & Hatoun, 1982).

Integration of Learning from Experiments

Organizational learning, another core concept in strategic experimentation, requires that appropriate learning be distilled from each experiment. This orientation combines decision-making and learning. Initiatives judged to be failures are not merely weeded out; they become occasions for discovery of root causes. Nor are successes simply alternatives to back financially; successes often generate potential best practices. Current KM tools in use for this purpose include learning histories (Roth & Kleiner, 1998), group brainstorming, and shared communication platforms.

Diffusion of Learning

Organizational learning has to be diffused throughout the organization. Since formal organizational channels may stifle transmission of tacit knowledge, diffusion may require interactions among “communities of practice” (Grover & Davenport, 2001; Davenport & Prusak, 1998). An organizational architecture incorporating relevant tools and IT infrastructure has to be designed to support theses interactions. KM tools such as knowledge maps identifying the experts in specific areas and repositories of case histories, are evolving to include dynamic updating of repositories and focused search tools to reduce information overload.

Managing a Portfolio of Strategic Experiments

Finally, unlike in previous eras, strategic experimentation requires maintenance and management of a portfolio of initiatives (Narayanan et al., 2001). This has three major implications. First, the knowledge base for decisions has to be broader and richer, simply due to the increase in the number of initiatives. Second, the knowledge base becomes much more complex, since the initiatives themselves differ in terms of the mix of tacit and explicit knowledge. Thus, newer initiatives are likely to be more dependent on tacit knowledge, whereas mature ones can be augmented by explicit knowledge. Finally, the sheer number of people involved in the process will be larger, given specialized pockets of tacit knowledge that would have grown up around specific strategic initiatives. DSS and other rich data applications, including cognitive mapping, can be used to capture the knowledge and feedback.

IT can accelerate the development of strategic experimentation by designing infrastructures that accommodate the new KM demands imposed by this new mode of planning. Consider how each of the following functions can be enhanced by IT infrastructure development.

1. Future developments can significantly reduce the time expended in solution development through real time displays, and expand opportunities for geographically dispersed collaboration. Also, advanced multimedia and communication capabilities increase the benefits of GSS and DSS tools.
2. Learning from experiments can be enriched by qualitative database construction, multimedia enhancements to communication applications, and open platforms to permit the sharing of knowledge over various communication channels, including wireless media.
3. Today, diffusion is hampered by information overload that has intensified competition for the user's attention (Hansen & Haas, 2001). To solve the problem, search tools should include separate parameters for content, rationale, and purpose of the query in order to isolate salient responses. Additionally, knowledge repositories must be maintained to ensure the contents are accurate and of high quality. Maintenance, currently provided by intermediaries (Markus, 2001), might be performed by faster automated systems.
4. Expert systems or neural networks may be developed to manage and track portfolios, promoting reuse of the knowledge captured.

The significant implication for IT infrastructure from our discussion is the need for *technology integration* (Narayanan, 2001) with both hard and soft technologies. IT infrastructure should exploit the potential for integration with other *hard* technologies such as telecommunications to enhance the organizational capacity for speed and the carrying capacity for *tacit* knowledge. Similarly, IT should seek to interface with *decision sciences* to embed AI-based processing tools, and with *cognitive theorists* to capture the tacit knowledge pervasive in organizations.

CONCLUSIONS AND IMPLICATIONS

We have argued that the technological changes of the 1990s have ushered in the need for strategic experimentation as the metaphor for planning practice. Strategic experimentation involves (a) maintaining a portfolio of strategic thrusts, (b) rapid decision-making so that successful experiments are backed and failures are weeded out quickly, (c) learning from both successes and failures, and (d) diffusion of both explicit and tacit knowledge throughout the relevant segments of an organization. This phase requires fundamental shifts in our view of knowledge management—its significance, use, and tools. Finally, we have argued that the shift to strategic experimentation requires fundamental shifts in the development of IT infrastructure. Instead of developing in relative isolation to other disciplines, IT should focus on technology integration by working in close collaboration with the telecommunication technologies, artificial intelligence community, and managerial cognition scholars.

REFERENCES

Broadbent, M., Weill, P., & St. Clair, D. (1999). The implications of information technology infrastructure for business process redesign. *MIS Quarterly*, *23*, 159-182.

Brown, S.L., & Eisenhardt, K.M. (1998). *Competing on the edge: Strategy as structured chaos*. Boston, MA: Harvard Business School Press.

Covin, J.G., Slevin, D.P., & Schultz, R.L. (1994). Implementing strategic missions: Effective strategic, structural and tactical choices. *Journal of Management Studies, 31*, 481-505.

D'Aveni, R.A. (1994). *Hypercompetition: Managing the dynamics of strategic maneuvering*. New York: The Free Press.

Davenport, T., & Prusak, L. (1998). *Working knowledge: How organizations manage what they know*. Boston, MA: Harvard Business School Press.

Duffy, J. (2000). The KM technology infrastructure. *Information Management Journal*, *34*, 62-66.

Fahey, L., & Randall, R. M. (2001). *The portable MBA in strategy (2nd ed.)*. New York: Wiley.

Gluck, F.W., Kaufman, S.P., & Walleck, S. (1978, October). The evolution of strategic management [Paper]. In McKinsey Staff Paper.

Gold, A.H., Malhotra, A., & Segars. A.H. (2001). Knowledge management: An organizational capabilities perspective. *Journal of Management Information Systems, 18*, 185-214.

Grover, V., & Davenport, T. (2001). General perspectives on knowledge management: Fostering a research agenda. *Journal of Management Information Systems, 18*, 5-21.

Hofer, C.W., & Schendel, D. (1978). *Strategy formulation: Analytical concepts*. St. Paul, MN: West Publishing.

Holmes, F.W. (1985). The information infrastructure and how to win with it. *Information Management Review, 1*(2), 9-19.

Lientz, B.P., & Chen, M. (1981). Assessing the impact of new technology in information systems. *Long Range Planning, 14*(6), 44-50.

Little, J.D., Mohan, L., & Hatoun, A. (1982). Yanking knowledge from the numbers: How marketing decision support systems can work for you. *Industrial Marketing, 67*, 46-56.

March, J.G. (1991). Exploration and exploitation in organizational learning. *Organization Science, 2*, 71-87.

Markus, M.L. (2001). Toward a theory of knowledge reuse: Types of knowledge reuse situations and factors in reuse success. *Journal of Management Information Systems, 18*, 57-93.

McGrath, R.G. (1997). A real options logic for initiating technology positioning investments. *Academy of Management Review, 22*, 974-996.

McGrath, R.G. (1998). Discovering strategy: Competitive advantage from idiosyncratic experimentation. In G. Hamel, C. K. Prahalad, H. Thomas, & D. O'Neal (Eds.), *Strategic flexibility: managing in a turbulent environment* (pp. 351-370). Chichester: Wiley.

McGrath, R.G., & MacMillan, I. (2000). *The entrepreneurial mindset*. Boston: Harvard Business School Press.

Narayanan, V.K. (2001). *Managing technology and innovation for competitive advantage*. Prentice Hall College Division.

Narayanan, V.K., Buche, M. & Kemmerer, B. (2002). From strategic management to strategic experimentation: The convergence of IT, knowledge management, and strategy. In Luis Joia (Eds), *IT-based management: Challenges and solutions*. PA: Idea Group Publishing.

Roth, G., & Kleiner, A. (1998). Developing organizational memory through learning histories. *Organizational Dynamics, 27*, 43-60.

Shank, M.E., Boynton, A.C., & Zmud, R.W. (1985). Critical success factor analysis as a methodology for MIS planning. *MIS Quarterly, 9*, 121-129.

Turban, E., McLean, E., & Wetherbe, J. (1996). *Information technology for management*. New York: Wiley.

Van de Ven, A.H., & Polley, D. (1992). Learning while innovating. *Organization Science, 3*, 92-116.

Whitten J.L., & Bentley, L.D. (1998). *Systems analysis and design methods* (4th ed.). Boston: Irwin McGraw-Hill.

KEY TERMS

Exploration and Exploitation: Exploration refers to the process of discovery of knowledge, whereas exploitation refers to utilizing the knowledge. Similar to basic and applied research.

Information Technology (IT): Refers to the physical equipment (hardware), software, and telecommunications technology, including data, image, and voice networks, employed to support business processes.

Knowledge Management (KM): A set of business practices and technologies used to assist an organization to obtain maximum advantage of its knowledge.

Options: A financial option owes the holder the right, but not the obligation, to trade in securities at prices fixed earlier. Options in the sense used here confer to a firm the rights, but not the obligations, to choose a strategic alternative.

Strategic Experimentation: A form of strategic management in which firms continually start, select, pursue, and drop strategic initiatives before launching aggressively those initiatives whose values are finally revealed.

Strategic Management: The process of strategy creation and implementation. The concept of "strategy" as used here is one of marketplace strategy (i.e., winning in the marketplace against competitors, entrenched or incipient). Strategy creation involves both goal formulation—defined in terms of external stakeholders rather than operational milestones—and crafting of the strategic means by which to accomplish these goals. Implementation refers to the means of executing the created strategy.

Strategic Knowledge Management in Public Organizations

Ari-Veikko Anttiroiko
University of Tampere, Finland

INTRODUCTION

New public management and the more recent concept of new public governance have become the dominant management doctrines in the public sector. Public organizations have become increasingly network-like units with various governance relations with actors from the public, business, and voluntary sectors. Their organization is based more on networks than on traditional hierarchies, accompanied by a transition from the command-and-control type of management to initiate-and-coordinate type of governance.

Among the most critical factors in this transformation is knowledge, for most of what has happened has increased the overall demand to create and process knowledge and to utilize it in the performance of governmental functions. The success of public organizations depends increasingly on how efficiently they utilize their knowledge assets and manage their knowledge processes in adjusting to local and contextual changes, as illustrated in Figure 1 (cf. Fletcher, 2003, pp. 82-83; Gupta et al., 2004, p. 3; Skyrme, 1999, p. 34). This requires that special attention be paid to strategic knowledge management.

Figure 1. The public organization as an institutional mediator (Adopted from Anttiroiko, 2002, p. 272)

In the early organization theories of public administration, knowledge was predominantly conceptualized within the internal administrative processes, thus to be conceived of as bureaucratic procedures, rationalization of work processes, identification of administrative functions, and selected aspects of formal decision-making. New perspectives emerged after World War II in the form of strategic planning and new management doctrines. The lesson learned from strategic thinking is that we need information on the external environment and changes therein in order to be able to adapt to and create new opportunities from these changes (see Ansoff, 1979; Bryson, 1995). As the complexity in societal life and related organizational interdependency has increased due to globalization and other trends, new challenges of managing organization-environment interaction also emerged (cf. Skyrme, 1999, p. 3).

BACKGROUND

The branch of management doctrine that became known as knowledge management (KM) reflected actual changes and new ideas in the business world. Classic works that inspired later developments included Polanyi (1966) and Drucker (1969). During the 1980s knowledge became widely recognized as a source of competitiveness, and by the end of the 1990s, knowledge management had become a buzzword. Among the best known thinkers who contributed to the rise of this field are Peter Senge (1990), Ikujiro Nonaka and Hirotaka Takeuchi (1995), Karl-Erik Sveiby (1997), and Thomas A. Stewart (1997). (On the evolution of knowledge management see Barclay & Murray, 1997; Gupta et al., 2004, pp. 8-10). It is becoming common understanding that in essence *knowledge management* is about governing the creation, dissemination, and utilization of knowledge in organizations (Gupta et al., 2004, p. 4; Lehaney et al., 2004, p. 13).

Knowledge cannot be managed in the traditional sense of management. The processing and distribution of information can surely be managed, but it is only one part of the picture. The other concerns knowledge and especially managers' ability to create conditions that stimulate active and dynamic knowledge creation, learning, and knowledge sharing within the organization (e.g., Nonaka, Toyama & Konno, 2000). To systematize this picture we may say that knowledge management includes four core areas (cf. Gupta et al., 2004; Lehaney et al., 2004):

- Information management: managing data and information, and designing information and knowledge systems
- Intellectual capital management: creating and utilizing knowledge assets, innovations, and intellectual capital
- Knowledge process management: organizing, facilitating, and utilizing sense-making and other knowledge processes
- Organizational learning: creating learning and knowledge sharing environments and practices

Traditionally the most widely applied areas of knowledge management in public organizations used to be data and transaction processing systems and management information systems serving mainly internal administrative functions. Yet, since the 1980s authorities started to facilitate the exchange of information by local area networks, followed by the Internet revolution of the 1990s. In the early 2000s the knowledge management agenda has focused increasingly on knowledge sharing and learning, and in inter-organizational network and partnership relations (see e.g., Wright & Taylor, 2003). As reported by OECD (2003, p. 4), knowledge management ranks high on the management agenda of the great majority of central government organizations across OECD member countries, followed with some time lag by regional and local authorities. Many public organizations have even developed their own KM strategies. The leading countries in this respect include France, Sweden, Finland, and Canada (OECD, 2003, pp. 28-29).

As to more operational actions, there has been a wave of intranet projects at all levels of public administration since the late 1990s. The result is that some 90% of state agencies surveyed by OECD in the early 2000s had their intranets in place. Sectors that really stand out as being well above the OECD average include organizations in charge of finance and budget, of justice, and of trade and industry (OECD, 2003, pp. 20-30). To give a concrete example from local level, New York City's Office of Technology set up extranet and intranet projects – the Human Services Extranet Project, to link the city agencies with human service contractors and the NYC Share Project, a citywide intranet that intended to improve the exchange of information among agencies – to facilitate external and internal knowledge processes. Such projects are typical in the public sector in the early 2000s. They indicate a transition from information management towards genuine knowledge management.

FOCUSING ON THE STRATEGIC ASPECT

Combining strategic thinking with knowledge management brings us to the very core of the life of organizations. *Strategic knowledge management* is a set of theories and guidelines that provides tools for managing an organization's knowledge assets and processes of strategic importance for the purpose of achieving organizational goals. The basic idea of strategic knowledge management in the public sector is to ensure that public organizations are capable of high performance by utilizing knowledge assets and knowledge processes when interacting with their environment.

What is essential in strategic knowledge management is that it needs to be strategic in the true sense of the word, as opposed to operational. Public employees have sometimes a tendency to view their knowledge requirements from the point of view of their current work practices. At an organizational level, too, there is sometimes a temptation to map out the future on the basis of current strengths and well-defined short-term challenges. The strategic approach to knowledge aims to overcome such inertia and narrow perspectives by creative knowledge processes, which help to transform views from introspective to outward-looking, from resources to outcomes, and from formal duties to actual impacts and customer satisfaction.

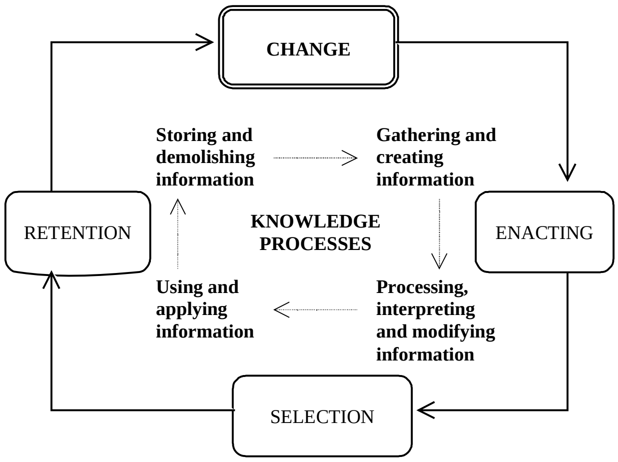
In the knowledge management literature, knowledge has primarily been approached either as an object or a process (cf. Sveiby, 2001). The main focus of public organizations is on knowledge processes framed by certain institutional arrangements. Among the most important of these are the political dimension and democratic control and legally defined functions, competencies and procedures within territorially defined jurisdictions. This theme will be discussed next.

FACILITATING STRATEGIC KNOWLEDGE PROCESSES

Public organizations possess and process a huge amount of information in their internal operations and external exchange relations. This is why the most important function of their knowledge management practice is to manage knowledge processes and to support knowledge-sharing practices.

Nonaka (1994) considers an organization's ability to accomplish the task of acquiring, creating, exploiting, and

Figure 2. Strategic sense-making and related knowledge process of the organization

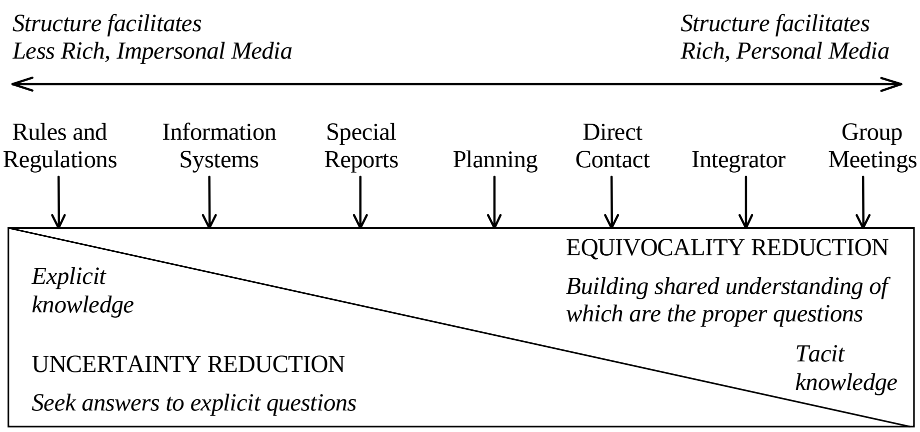


accumulating new knowledge. This formulation takes us very close to how the knowledge process can be operationalized. The *knowledge process* can be defined as an organizational process in which information is gathered, created, processed, used and dissolved in order to form an enriched orientation base for taking care of organization's basic functions (cf. Gupta et al., 2004, p. 3; Mendes et al., 2004, p. 153).

It is important to note that in the actual knowledge process it is more or less meaningless to make a clear-cut distinction between knowledge and information, for both are processed in such a process. For example, knowledge is not simply extracted from information, for knowledge is possessed by human beings and serves as a background and in-built epistemic frame to deal with complexity, novelty, and the requirements of innovativeness (cf. Wiig, 2000). Thus, genuine aspects of the category of *knowledge* are in question when we deal with statements, assumptions, and understandings and such learning and communicative processes in which these knowledge assets can be shared, assessed, and enriched. Many theorists consider that it is tacit knowledge in particular that is the most challenging and important form of knowledge in organizations (Nonaka, 1994; Polanyi, 1966). It also needs to be stressed that it is not knowledge as something abstract but a "generative dance" or interplay between (a) *knowledge* we possess and (b) *knowing* as an epistemic aspect of the interaction with the world that generates organizational innovation and strategic understanding (Cook & Brown, 1999).

Strategic knowledge processes are those aspects of knowledge processes that have the most profound and far-reaching impact on an organization's adjustment to contextual changes and on its core competencies. A paradigmatic form of a strategic knowledge process is the *sense-making or strategy process* in which an organization devotes effort to analyzing its internal attributes and external conditions and decides on that basis about the action lines in order to achieve its overall goals (cf. Weick, 1995). In such a strategic knowledge process the organization seeks information on environmental changes and utilizes this in strategy formulation, in which such tools as SWOT analysis have traditionally been used. A basic model of the organizational knowledge-based adaptation process is presented in Figure 2 (Anttiroiko, 2002). This model serves as a heuristic tool to conceptualize knowledge processes. Yet, it is important to keep in mind that this is only a starting point. When taking this idea further, clear-cut sequential stages or phases of the KM life cycle need to be "recontextualized" as a set of continuous interdependent sub-processes (cf. Mendes et al., 2004, p. 165). Thus, context-related and situational aspects of knowledge need to be integrated with all essential connections to their environments into the key functions and operations of an organization in order to assess their meaning as a part of actual strategic adaptation and sense-making processes.

Figure 3. Continuum of knowledge facilitation mechanisms (Daft & Lengel, 1986)



Applying Daft and Lengel (1986), we may ask how organization structures and systems should be designed in order to meet the need to manage knowledge processes. Well designed systems help to decrease the uncertainty and ambiguity faced by an organization by ordering the amount of relevant information and by enabling clarification of problems and challenges. Daft and Lengel propose seven structural mechanisms that can be used to deal with uncertainty and ambiguity in varying degrees, as illustrated in Figure 3. This model resembles the continuum of communication that has explicit knowledge and tacit knowledge as its extremities (Lehaney et al., 2004, p. 21).

The idea is that these mechanisms form a continuum starting from tools to be used to tackle well defined problems and thus to reduce uncertainty, and proceeding towards more communicative mechanisms designed to facilitate sense-making processes that aim at reducing equivocality or ambiguity (Anttiroiko, 2002).

As stated, a paradigmatic case for strategic knowledge management is the strategy process of an organization (on strategy and information resources, see Fletcher, 2003, pp. 82-84). What is of utmost importance is that managers ensure that people feel involved in the strategy formulation process. The staff also needs to understand the meaning of strategy in their own situations. This would help to make strategy documents living guidance that is owned by all in the organization, as concluded by Wright and Taylor (2003, p. 198).

Another important premise relates to organization culture and work practices that often impede the development of knowledge management. For example, employees may resist a new knowledge management initiative if they perceive it only as extra work. Similarly, employees may be reluctant to share their knowledge if there are no rewards or tangible benefits for themselves or their organizations. In all, the human factor is essential for improving KM practices, for most of the positive outcomes are the result of the commitment of all employees, successful structural changes in the organization, and the development of the organizational culture and climate (OECD, 2003, p. 4).

THE ROLE OF TECHNOLOGY

Information technology (IT) provides a range of new tools that can be effectively used in knowledge management. Relevant applications can support decision making, executive functions, planning, communication, and group work.

Tools and technologies available for knowledge management include generic communication tools (e.g., e-mail), computer-based information and decision support systems, document management systems, intranets and extranets, groupware, geographic information systems, help desk technology, and a range of knowledge representation tools. (Grafton & Permaloff, 2003; Gupta et al., 2004, pp. 17-24).In general, the Internet may be suggested as the KM infrastructure due to its widespread availability, open architecture, and developed interfaces (Jennex, 2003, p. 138).

In real life most of the tools applied in knowledge management are more or less conventional, such as training, seminars, meetings, and the like. Various KM-specific organizational arrangements had been adopted by about half of the organizations studied in the OECD survey on ministries, departments, and agencies of central government in the early 2000s. These measures include central coordination units for KM, quality groups, knowledge networks, and chief knowledge officers. Another important application area is the classification of information, referring to new filing mechanisms, e-archives, and new types of databases. In internal knowledge sharing Intranet projects form the mainstream, combined with wide access to the Internet and having e-mail addresses for the staff. The external knowledge sharing goes largely hand in hand with the emergence of new practices of e-governance. These practices have increased the knowledge sharing in both local and wider governance processes (Anttiroiko, 2004; OECD, 2003, pp. 17-20).

FUTURE TRENDS

The future challenge for public organizations is to increase their responsiveness to stakeholders, especially to citizens. At the same time they need to be capable of strategic institutional mediation in the increasingly turbulent environment, thus bringing an element of continuity and stability to social life and guaranteeing democratic and civic rights at different institutional levels. All this requires increasing capacity to manage knowledge of strategic importance.

CONCLUSION

Strategic knowledge management refers to the theory and practice of managing knowledge assets and processes of strategic importance. Public organizations need to create favorable organization structures and environments for knowledge sharing, organizational learning, and other aspects of knowledge management in order to create all the knowledge they require in their adjustment and trend-setting processes.

A main return of strategic knowledge management is better capability to adjust to contextual changes. This is difficult to measure, even if such tools as Balanced Scorecard (BSC), the Intangible Assets Monitor (IAM), and Intellectual Capital Index (ICI) are available. This is because they provide only a partial picture of KM performance, as claimed by Chaudhry (2003, p. 63). What seems to be needed is more process-focused assessments that are able to analyze the steps of KM processes, thus highlighting the actual changes in organizational knowledge base, capacities, and processes. As usual, there is no measurement system that fits all organizations in all situations. Rather, measurement should be tailored to the actual needs of the organization.

REFERENCES

Ansoff, H.I. (1979). *Strategic management*. London: Macmillan.

Anttiroiko, A.-V. (2002). Strategic knowledge management in local government. In A. Grönlund (Ed.), *Electronic government: Design, applications & management* (pp. 268-298). Hershey, PA: Idea Group Publishing.

Anttiroiko, A.-V. (2004). Introduction to democratic e-governance. In M. Malkia, A.-V. Anttiroiko & R. Savolainen (Eds.), *eTransformation in governance*. Hershey, PA: Idea Group Publishing.

Barclay, R.O., & Murray, P.C. (1997). *What is knowledge management?* Knowledge Management Associates. Retrieved February 11, 2004, from *http://www.media-access.com/whatis.html*

Bryson, J.M. (1995). *Strategic planning for public and nonprofit organizations. A guide to strengthening and sustaining organizational achievement* (rev. ed.). San Francisco: Jossey-Bass Publishers.

Chaudhry, A.S. (2003). What difference does it make: Measuring returns of knowledge management. In E. Coakes (Ed.), *Knowledge management: Current issues and challenges*. Hershey, PA: IRM Press.

Cook, S.D.N., & Brown, J.S. (1999). Bridging epistemologies: The generative dance between organizational knowledge and organizational knowing. *Organization Science, 10*(4), 381-400.

Daft, R.L., & Lengel, R.H. (1986). Organizational information requirements, media richness and structural design. *Management Science, 32*(5), 554-571.

Drucker, P. (1969). *The age of discontinuity. Guidelines to our changing society*. London: Heinemann.

Fletcher, P.D. (2003). The realities of the Paperwork Reduction Act of 1995: A government-wide strategy for information resources management. In G.D. Garson (Ed.), *Public information technology: Policy and management issues*. Hershey, PA: Idea Group Publishing.

Grafton, C., & Permaloff, A. (2003). Computer tools for better public sector management. In G.D. Garson (Ed.), *Public information technology: Policy and management issues*. Hershey, PA: Idea Group Publishing.

Gupta, J.N.D., Sharma, S.K., & Hsu, J. (2004). An overview of knowledge management. In J.N.D. Gupta & S.K. Sharma (Eds.), *Creating knowledge based organizations*. Hershey, PA: Idea Group Publishing.

Jennex, M.E. (2003). A survey of Internet support for knowledge management organizational memory systems. In E. Coakes (Ed.), *Knowledge management: Current issues and challenges*. Hershey, PA: IRM Press.

Lehaney, B., Clarke, S., Coakes, E., & Jack, G. (2004). *Beyond knowledge management*. Hershey, PA: Idea Group Publishing.

Mendez, M.M., Gomes, J.F.S., & Bátiz-Lazo, B. (2004). Management of knowledge in new product development in Portuguese higher education. In J.N.D. Gupta & S.K. Sharma (Eds.), *Creating knowledge based organizations*. Hershey, PA: Idea Group Publishing.

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science, 5*(2), 14-37.

Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company. How Japanese companies create the dynamics of innovation*. New York: Oxford University Press.

Nonaka, I., Toyama, R., & Konno, N. (2000). SECI, ba and leadership: A unified model of dynamic knowledge creation. *Long Range Planning, 33*, 5-34.

OECD. (2003, April 3-4). *The learning government: Introduction and draft results of the survey of knowledge management practices in ministries/departments/agencies of central government*. 27th Session of the Public Management Committee, Public Management Committee (PUMA). Paris: Organisation for Economic Co-operation and Development (OECD).

Polanyi, M. (1966). *The tacit dimension*. Garden City, NY: Doubleday & Company.

Senge, P.M. (1990). *The fifth discipline. The art and practice of the learning organization*. New York: Doubleday.

Skyrme, D.J. (1999). *Knowledge networking. Creating the collaborative enterprise*. Oxford: Butterworth-Heinemann.

Stewart, T.A. (1997). *Intellectual capital: The new wealth of organizations*. New York: Currency/Doubleday.

Sveiby, K.E. (1997). *The new organizational wealth. Managing and measuring knowledge-based assets.* San-Francisco: Berrett-Koehler Publishers.

Sveiby, K.-E. (2001). *What is knowledge management?* Retrieved February 10, 2004, from *http://www.sveiby.com/articles/KnowledgeManagement.html*

Weick, K. (1995). *Sense-making in organizations.* Thousand Oaks, CA: Sage.

Wiig, K. (2000). Knowledge management: An emerging discipline rooted in a long history. In C. Despres & D. Chauvel (Eds.), *Knowledge horizons. The present and the promise of knowledge management.* Boston, MA: Butterworth-Heinemann.

Wright, G., & Taylor, A. (2003). Strategic knowledge sharing for improved public service delivery: Managing an innovative culture for effective partnerships. In E. Coakes (Ed.), *Knowledge management: Current issues and challenges.* Hershey, PA: IRM Press.

KEY TERMS

Intellectual Capital (IC): Knowledge and know-how possessed by an individual or an organization that can be converted into value in markets. Roughly the same as the concept of intangible assets.

Intellectual Capital Management (ICM): A management of value creation through intangible assets. Close to the concept of knowledge management.

Intellectual Property (IP): Any product of the human intellect that is unique and has some value in the marketplace. It may be an idea, composition, invention, method, formula, computer software or something similar. In practice, special attention is paid to such intellectual property that can be protected by the law (e.g., patent and copyright).

Knowledge Assets (KAs): Statements, assumptions, abstract models and other forms of knowledge regarding the organization itself and its environment (markets, customers, etc.) that an organization possesses. These assets provide economic or other value to an organization when interacting within it or with its environment.

Knowledge Management (KM): Management theory and practice on managing intellectual capital and knowledge assets and also the processes that act upon them. In practical sense KM is about governing the creation, dissemination, and utilization of knowledge in organizations.

Knowledge Management System (KMS): Set of tools and processes used by knowledge workers to identify and transmit knowledge to the knowledge base contained in the organizational memory.

Organizational Learning (OL): An organizational process in which the intentional and unintentional processing of knowledge within a variety of structural arrangements is used to create an enriched knowledge and orientation base and a better organizational capacity for the purpose of improving organizational action.

Strategic Utilization of Data Mining

Chandra S. Amaravadi
Western Illinois University, USA

INTRODUCTION

In the past decade, a new and exciting technology has unfolded on the shores of the information systems area. Based on a combination of statistical and artificial intelligence techniques, data mining has emerged from relational databases and Online Analytical Processing as a powerful tool for organizational decision support (Shim et al., 2002).

A number of techniques are available to analyze warehouse data, including descriptive techniques such as data summarization; data visualization; clustering and classification; and predictive techniques such as regression, association, and dependency analyses (Jackson, 2002; Mackinnon & Glick, 1999). The technology is being extended to mine semi-structured data as well (Hui & Jha, 2000).

Applications of data mining have ranged from predicting ingredient usage in fast-food restaurants (Liu, Bhattacharyya, Sclove, Chen & Lattyak, 2001) to predicting the length of stay for hospital patients (Hogl, Muller, Stoyan & Stuhlinger, 2001). See Table 1 for other representative examples. Some of the important findings are:

1) Bankruptcies can be predicted from variables such as the “ratio of cash flow to total assets” and “return on assets” (Sung, Chang & Lee, 1999).
2) Gas station transactions in the UK average £20, with a tendency for customers to round the purchase to the nearest £5 (Hand & Blunt, 2001).
3) Sixty-nine percent (69%) of dissatisfied airline customers did not contact the airline about their problem (Marple & Zimmerman, 1999).
4) Sales in fast-food restaurants are seasonal and tend to peak during holidays and special events (Liu et al., 2001).
5) Patients in the age group > 75 are 100% likely to exceed the standard upper limit for hospital stay (Hogl et al., 2001).

BACKGROUND

A majority of data mining (DM) applications serve a managerial purpose. They are useful in finding information such as identifying loyal customers or patients who are likely to stay longer at hospitals. This usage can be extended to strategic decision making as well. According to Sabherwal and King (1991), a *strategic application* is one that has a profound influence on a firm’s success, by either influencing or shaping the organization’s strategy or by playing a direct role in the implementation or support of it. If DM could be utilized in shaping the firm’s strategy, it could have a strategic impact. Let us then consider the process of strategic decision making (SDM). An interpretive view of this process involves *scanning* the environment for important events or information, *interpreting* these events as threats or opportunities, and *formulating* a response (Daft & Weick, 1984).

The interpretation stage involves some form of consensual validation, with managers comparing notes with subordinates or with peers (Daft & Weick, 1984). It is of particular interest since it involves modifying the *belief systems*, the summary of perceptions, observations, and experiences concerning the organization’s resources, markets, and customers. For instance, an organization

Table 1. Examples of data mining applications

- Predicting supplies in fast food restaurants (Liu, Bhattacharyya, Sclove, Chen and Lattyak 2001).
- Quality of health care (Hogl, Muller, Stoyan and Stuhlinger 2001).
- Analyzing Franchisee sales (Chen, Justis and Chong 2003).
- Predicting customer loyalty (Ng and Liu 2001).
- Job shop scheduling (Koonce, Fang and Tsai 1997).
- Mining credit card data (Hand and Blunt 2001).
- Bankruptcy prediction (Sung, Chang and Lee 1999).

might have a perception that its product lines are aging. Customers switching to competitors' products could confirm this observation. There is empirical evidence that belief systems influence strategic decision making (Lorsch, 1989). In a study of 12 firms, Lorsch (1989) found that major strategic decisions were influenced by core beliefs that included financial goals, acceptable types of risks, and distinctive competence, among other things. Thus decisions regarding product lines, customers, and suppliers are influenced by perceptions concerning adequacy of product lines, type of customers, reliability of suppliers, and so forth. We will alternatively use the term *micro-theories* (MTs) to refer to these beliefs and will regard each as a strategic assumption to be tested by data mining.

The mining process, often labeled as "KDD" (Knowledge Discovery in Databases) can be "data driven" or "hypothesis driven" ("question driven"). *Data-driven* methods attempt to identify all possible patterns from the data, while *hypothesis-driven* methods attempt to verify whether or not a particular pattern exists (Hogl et al., 2001). Usually, organizations have more data than they can analyze. Question-driven approaches are computationally more tractable, especially when large data sets are involved, since the solution space is bounded. In this mode, KDD commences with a set of MTs that management is keen on verifying. The remainder of the process is the same for both approaches (Mackinnon & Glick, 1999). The next step is to select suitable data (see Figure 1). This is greatly facilitated if the analyst already has hypotheses to verify. Otherwise, data selection will involve an iterative process of selection followed by testing. The required data needs to be carefully selected from the warehouse or organizational databases. It is then cleaned and transformed by filling in missing values, changing "look up codes" (i.e., standardizing codes from numeric values to text or vice-versa: "1"–married; "2"–single), and ignoring outliers if necessary. Calculations such as totals, cost/item, and discount are also performed during this stage. The next step is testing and analysis where each MT is examined using the "selected" and "cleaned" data. The last step is the sharing of results with management, usually through formal reports or presentations or via an intranet.

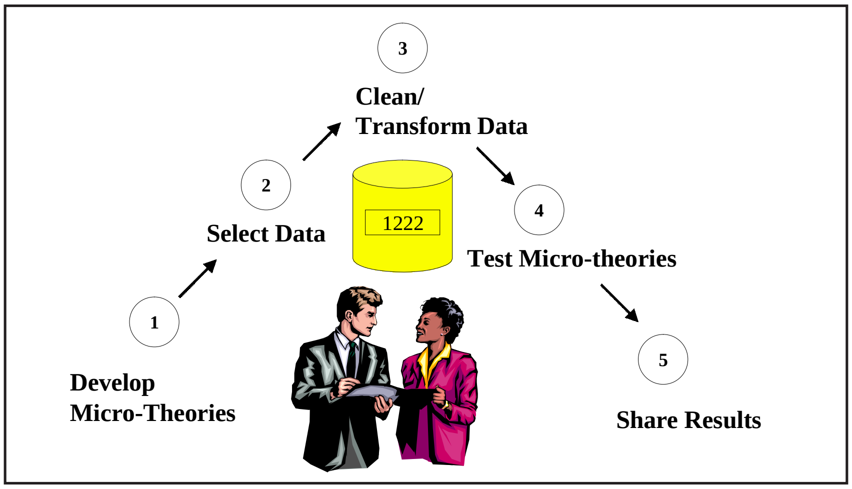
MAIN THRUST

What sort of beliefs should an analyst select for testing purposes? Porter's framework is widely used by academics and practitioners to understand organizational strategy. It summarizes key competitive forces acting on an organization as follows: a) bargaining power of customers, b) bargaining power of suppliers, c) firm rivalry, d) threat of substitute products, and e) threat of new entrants. To compete effectively, a firm has to counteract these forces.

The framework is useful in organizing micro-theories as well since it describes the entities pertaining to the organization's *task environment*, which govern its inputs and outputs and therefore affect its performance.

As shown in Figure 2, a firm's beliefs can be organized by each of the entities in the firm's task environment, including suppliers, customers, competitors, and substitute products. For instance, how do customers perceive the products of a company? How does the firm feel about a particular substitute? Ultimately, these perceptions

Figure 1. Knowledge discovery process with micro-theories (Adapted from Mackinnon & Glick, 1999)



influence strategy formulation and therefore need to be verified. The reader should note that perceptions about the firm, its management, and its employees are also included in the framework since a firm's internal resources and capabilities are very pertinent to the strategic decision-making process.

As a first step in the KDD process, the analyst needs to identify these assumptions. Decision mapping is a suitable technique here. A *decision map* is a chart depicting the decision processes in the organization (Ashworth & Goodland, 1990). For each of the task areas, the analyst should identify decisions made with a view to identifying underlying micro-theories. For instance, for the entity suppliers, *decisions faced* are: What type of a relationship should a company have with its suppliers? Should it have a long-term or a short-term relationship? Should it trust a single supplier for a critical component or should it maintain alternative sources? Should it attempt to develop offshore suppliers based on price or are there other considerations? How can it optimize a contract in terms of price/delivery time/lead time?

The beliefs that can underlie these decisions include (see Figure 2):

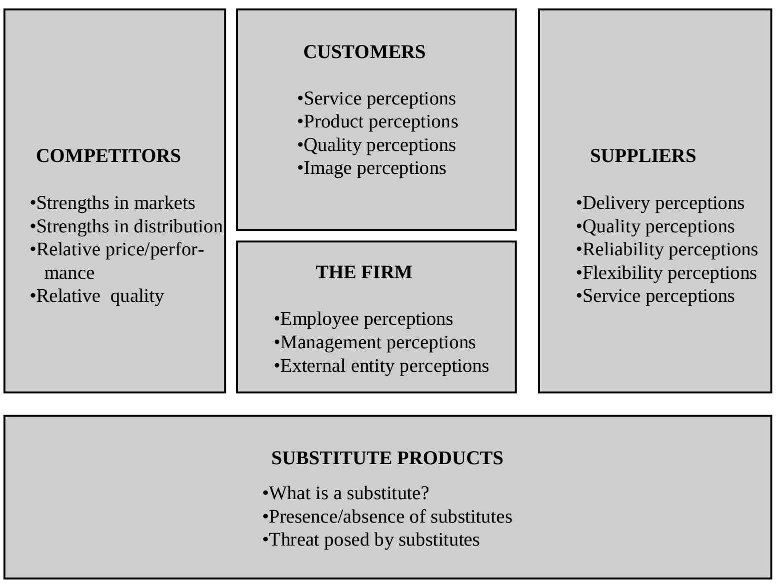
- The supplier is reliable.
- The supplier delivers on time.
- The supplier produces quality products.
- The supplier has historically offered good pricing/ delivery combination.
- The supplier is flexible in producing products to specifications.
- The supplier can operate with a small lead time.
- The supplier is quick in resolving problems.

Similar beliefs underlie other key task entities that have been identified. Typically, organizations will have dozens of such beliefs embedded in their SDM processes. To identify those that are relevant, the analyst can study the decisions, prepare a checklist of all MTs, and have senior management select the most important. This list will then drive the remainder of the KDD process.

The Testing Process

Once micro-theories are identified and data sets are selected/transformed, the next step in the KDD process is testing. Testing proceeds in two stages, first with "test data" which is usually 10-20% of the actual data to develop the model, and then with the remainder of the data to validate the model. As mentioned, the DM techniques include clustering, association, classification, and dependency analysis. The MT test list is used by the analyst as a guide in selecting a suitable technique. For instance, an assumption about the reliability of a supplier could be confirmed by an association analysis between suppliers, delivery times, and the number of times the specifications were met 100%. It should be noted that the raw data may not be available in this form, and therefore may require tabulating and aggregation, especially with respect to the variable, *specifications being met 100%*. If the association analysis confirms some vendors meeting these crite-

Figure 2. Porterian Framework for Characterizing Micro-Theories



ria, this is again tested on the remainder of the data in the second stage. A number of situations may arise with tested hypotheses: a) the hypothesis is supported in its entirety at the 90% confidence level or higher; b) the hypothesis is not supported at the 90% confidence level, but at a lower level of confidence; c) the hypothesis is not supported at any confidence level. Situations "a" and "c" are clear cut, resulting in confirmation or disconfirmation of the MT, but "b" can place the analyst in a quandary. In such cases, an alternative hypothesis may be sought by modifying the MT. For instance, an alternative hypothesis for the case above is that delivery times and specifications may be contingent on the delivery quantities. Thus testing is not always straight-forward and the strategy may need modification.

FUTURE TRENDS

The utilization of data mining in organizations will continue to increase. It is expected that some of the simpler data mining techniques will be incorporated into management support tools (Shim et al., 2002). If the assumption surfacing process could be automated (Conklin & Begeman, 1988), it will streamline the verification of micro-theories. Progress in data visualization could add to the efficacy of the process (Thearling et al., 2001). The integration of data mining with systems managing soft information is another important challenge.

CONCLUSIONS AND IMPLICATIONS

The strategic usage of data mining technology requires a hypothesis-driven approach to DM. The hypotheses to be tested are often embedded in the strategic assumptions of management. Referred to as micro-theories, or beliefs that underlie and influence critical decisions in an organization. A Porterian framework has been provided to serve as a guide to surfacing these MTs. IBIS (Conklin & Begeman, 1988) and SAST (Mason & Mitroff, 1981) are similar frameworks. IBIS was developed in the context of recording planning activities in governmental agencies, while SAST was specifically developed for strategic decision making. IBIS is concerned more with the rationale for decision making, while SAST is concerned with identifying assumptions through dialectic among stakeholders, *during* the decision-making process. Neither approach addresses the content or nature of the assumptions.

Using the framework, along with decision mapping, the analyst should surface strategic assumptions and test them with the usual techniques of mining. Typically the results will confirm the MT, but this may not always be the case. Studies have shown that managers are often too optimistic or too pessimistic, leading to divergence between MTs and conclusions from KDD. Not all MTs will be testable. For instance, the belief that a supplier is potentially valuable cannot be tested except through "soft" methods such as consensual validation. The ultimate result of such efforts is that executives can make strategic decisions with greater confidence.

REFERENCES

Ashworth, C. & Goodland, M. (1990). *SSADM—a practical approach*. Maidenhead: McGraw-Hill.

Chen, Y-S., Justis, R. & Chong, P.P. (2003). Data mining in franchise organizations. In H. Nemati & C. Barko (Eds.), *Organizational data mining*. Hershey, PA: Idea Group Publishing.

Conklin, J. & Begeman, M.L. (1988). IBIS: A hypertext tool for exploratory policy discussion. *ACM Transactions on Office Information Systems*, *6*(4), 303-331.

Daft, R.L. & Weick, K.E. (1984). Towards a model of organizations as interpretation systems. *Academy of Management Review*, *9*(2), 284-295.

Hand, D.J. & Blunt, G. (2001). Prospecting for gems in credit card data. *IMA Journal of Management Mathematics*, *12*(2), 173-200.

Hogl, O.J., Muller, M., Stoyan, H. & Stuhlinger, W. (2001). Using questions and interests to guide data mining for medical quality management. *Topics in Health Information Management*, *22*(1), 36-50.

Hui , S.C. & Jha, G. (2000). Data mining for customer service support. *Information and Management*, *38*(1), 1-13.

Jackson, J. (2002). Data mining: A conceptual overview. *Communications of the Association for Information Systems*, *8*, 267-296.

Koonce, D.A., Fang, C-H. & Tsai, S-C. (1997). A data mining tool for learning from manufacturing systems. *Computers in Industrial Engineering*, *33*(1-2), 27-30.

Lorsch, J.W. (1989). Managing culture: The invisible barrier to strategic change. In A.A. Thompson and A.J. Strickland (Eds.), *Strategy formulation and implementation*, (pp. 322-331). Homewood, IL: BPI/Irwin.

Liu, L.M, Bhattacharyya, S., Sclove, S.L., Chen R. & Lattyak, W.J. (2001). Data mining on time series: An illustration using fast-food restaurant franchise data. *Computational Statistics and Data Analysis, 37*, 455-476.

Mackinnon, M.J. & Glick, N. (1999). Data mining and knowledge discovery in databases—an overview, *Australian & New Zealand Journal of Statistics*, *41*(3), 255-275.

Marple, M. & Zimmerman, M. (1999). A customer retention strategy. *Mortgage Banking*, *59*(11), 45-49.

Mason, R.O. & Mitroff, I.I. (1981). *Challenging strategic planning assumptions.* New York: John Wiley & Sons.

Ng, K. & Liu, H. (2000). Customer retention via data mining. *Artificial Intelligence Review*, *14*(6), 569-590.

Sabherwal, R. & King, W.R. (1991). Towards a theory of strategic use of information resources. *Information and Management*, *20*(3), 191-212.

Shim, J.P., Warkentin, M., Courtney, J.F., Power, D.J., Sharda, R. & Carlsson, C. (2002). Past, present, and future of decision support technology. *Decision Support Systems*, *33*(2), 111-126.

Sung, T.K., Chang, N. & Lee, G. (1999). Dynamics of modeling in data mining: Interpretive approach to bankruptcy prediction. *Journal of Management Information Systems*, *16*(1), 63-85.

Thearling, K., Becker, B., DeCoste, D., Mawby, B., Pilote, M. & Sommerfield, D. (2001). Visualizing data mining models. In U. Fayyad, G. Grinstein & A. Wierse (Eds.), *Information visualization in data mining and knowledge discovery*. San Mateo, CA: Morgan Kaufman.

KEY TERMS

Association: A technique in data mining that attempts to identify similarities across a set of records, such as purchases that occur together across a number of transactions.

Beliefs: Summaries of perceptions that members in an organization typically share, such as “Sales are strong in the Southwest.”

Classification: A technique in data mining that attempts to group data according to pre-specified categories, such as “loyal customers” versus “customers likely to switch.”

Clustering: A technique in data mining that attempts to identify the natural groupings of data, such as income groups that customers belong to.

Data Driven: If the data drive the analysis without any prior expectations, the mining process is referred to as a data-driven approach.

Micro-Theories: Beliefs that need to be tested during the data mining process.

Multi-Dimensional Databases: A virtual database where data is organized according to dimensions or aspects of the data—such as product, location, and time for sales data—to facilitate queries, such as: “How many shoes were sold by Store #4 in January?”

OnLine Analytical Processing (OLAP): Performing high-level queries on multi-dimensional databases.

Question Driven: In question-driven or hypothesis-driven approaches, the analysis is preceded by an identification of questions of interest.

Strategic Vision for Information Technology

Mary Elizabeth Brabston
The University of Manitoba, Canada

INTRODUCTION

Today's dispersal of information technology (IT) decision making across many organizational managers and employees offers the promise of greatly increasing the extent to which IT is applied to the organization's managerial and operational work systems (Boynton & Zmud, 1987; Jarvenpaa & Ives, 1990). Along with the advantages of distributed action, however, come the potential disadvantages of inappropriate allocations of effort and resources and associated outcomes — chiefly poor investment decisions, lost opportunities, duplicated efforts, and incompatible platforms and applications.

One strategy for minimizing these disadvantages lies in the development of a common understanding among an organization's members concerning the primary roles to be served by information and by IT throughout the organization. One tactic for achieving this common understanding is through the development, articulation, and dissemination of a strategic vision for IT (Brabston, Zmud & Carlson, 2000; Collins & Porras, 1991; Nanus, 1992; Parker & Benson, 1991; Robbins & Duncan, 1988).

A 1994 study of banking showed that "while only 9% of CIOs interviewed said their firms have failed to effectively outline an information systems vision to the lines of businesses they support, a full 30% of business leaders reported a void in this area" (Hoffman, 1994). Effective development, implementation, and communication of a strategic vision for IT should significantly reduce this gap.

BACKGROUND

An organizational strategic vision is the shared understanding of what an organization should be and how it must change to get there (Schoemaker, 1997). According to Mintzberg, Ahlstrand, and Lampel (1998), "vision serves as both an inspiration and a sense of what needs to be done – a guiding idea…vision often tends to be a kind of image more than a fully articulated plan…that leaves it flexible, so that the leader can adapt it to his or her experiences" (pp. 124-125). By sharing a common strategic vision, organizational members will better understand how their individual roles contribute to their organization's strategic mission and will be more likely to act in an appropriate and consistent manner when faced with uncertain or ill-defined situations.

"Having a strategic vision redefines the rules for acting opportunistically or incrementally. The strategies, plans, and budgets should be determined by the vision" (Schoemaker, 1997). Similar benefits are believed to arise when a strategic IT vision is shared by an organization's managers and employees (Parker & Benson, 1991). If an organization's managers and employees are aware of the organization's strategic IT vision, they are more likely to make more effective IT decisions.

MAIN THRUST OF THE ARTICLE

The process by which a strategic IT vision is developed is clearly important. If the process produces an appropriate, aligned IT vision, one that its developers believe in, one that they can strongly sell to others in the organization, then the IT vision may become reality.

An organization's overall strategic vision is a shared, realistic, yet idealistic and attractive view of the organization that inspires and motivates others, through their individual and collective efforts, to move toward the vision (Collins & Porras, 1991; Nanus, 1992). Figure 1A graphically represents how an organizational strategic vision can serve as the stimulus for all organizational planning. Figure 1B depicts a similar role for the strategic IT vision. The intent is similar but with a narrower scope involving only the organizational role of information and IT. As shown by the dashed lines, the IT strategic vision must be aligned with the organizational strategic vision (Henderson & Venkatraman, 1993; Parker & Benson, 1991).

Figure 1 models the relationship between an organization's strategic vision and its subsequent planning and goal setting and those of the IT strategic vision. The typical process at the organizational level is that a group of individuals develop the organization's enterprise-wide strategic vision (e.g., Gerber's vision that "Babies are our most important product"). This vision is then communicated to others in the organization, resulting in the setting of strategic goals, from whence tactical goals are set, and finally, operational goals are developed based on the tactical goals. Each of these goal sets should be aligned with the enterprise-wide strategic vision.

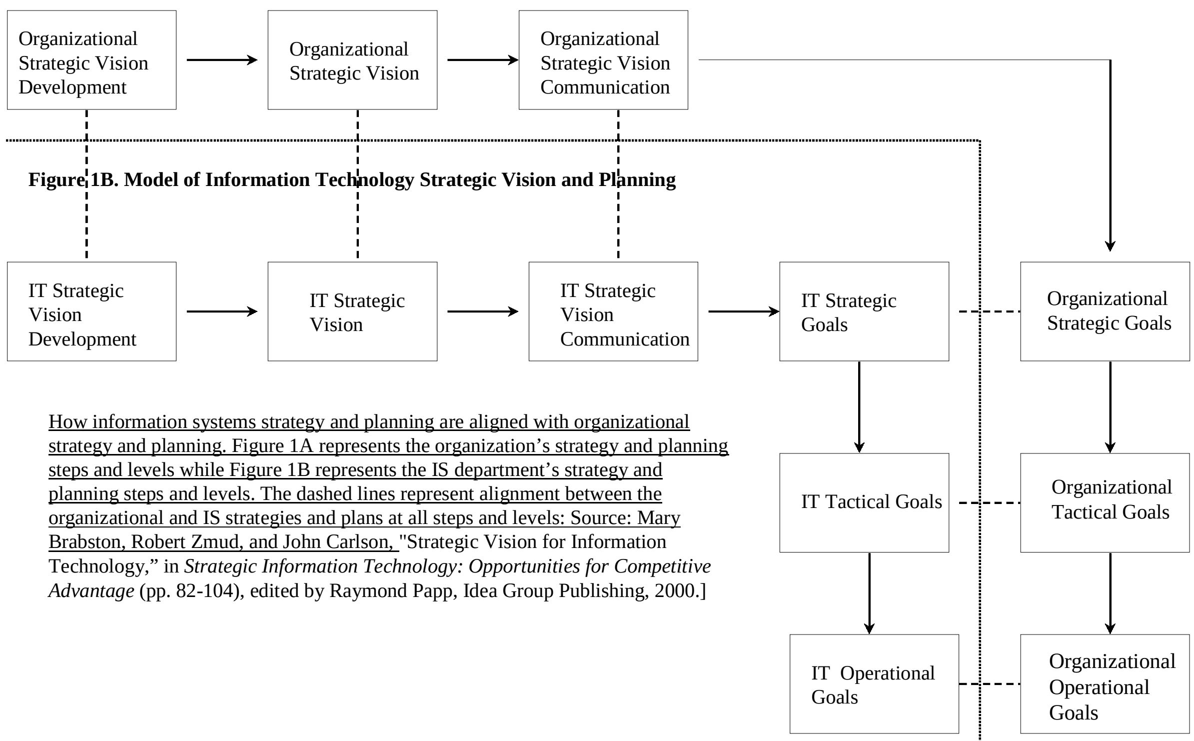
In the same way, a group of individuals would develop the IT vision, communicate it, and set goals at each level. At each stage of the IT strategic vision process, the vision and goals should be aligned with the vision and goals of the enterprise-wide strategic vision, as indicated by the dashed lines.

If appropriately developed, articulated, and disseminated, a strategic IT vision should both formally and informally influence the many IT-related decisions made throughout an organization so that these decisions, taken together, enable progress toward achieving the organization's overall strategic vision. In the absence of an enterprise-wide strategic IT vision, information and IT decision making will most likely be framed, rather than by an overarching organizational purpose, by the numerous (and potentially conflicting) mental models that are maintained in the minds of individuals, as well as those that are collectively developed by formal and informal groups throughout the organization concerning the appropriate roles of information and IT (Boland & Tenkasi, 1995).

Most often, strategic visions are phrased somewhat generically in order to increase the applicability of the vision to each organizational member (Nanus, 1992). This may be particularly important for a strategic IT vision so that this vision is not unintentionally restricted to a particular time frame, IT platform, or set of IT applications. However, although the vision is abstracted to apply across diverse organizational contexts, it should not be worded ambiguously or in such a manner as to create confusion on the part of organizational managers and employees working to apply the vision.

It is important here to distinguish between a strategic IT vision, our focus here, and a strategic vision for an organization's IT (or information services) function. The IT function's strategic vision articulates the organizational role of the IT function as well as the dominating values applied in carrying out these roles (c.f., Nanus, 1992). It may be valuable, but not necessary, that the strategic vision of the IT function be communicated across an organization; it is only necessary that it be communicated to those organizational managers and employees who significantly affect the IT function's efforts in carrying out that area's mission, for example, IT departmental employees, IT vendors, and so forth. Essentially, then, the strategic IT vision paints a picture of the what and why associated with an organization's use of

Figure 1A. Model of organizational strategic vision and planning



information and IT while the strategic vision of the IT function captures one aspect of the how.

A strategic IT vision can be explicitly or implicitly developed. With an explicit vision, a top management group conducts a formal process for developing the vision and, at the conclusion of this process, a formal vision statement is articulated and disseminated throughout the organization. With an implicit vision, the top management group evokes the vision through its ongoing decision-making processes, communication activities, and behaviors.

Ideally, strategic-level IT planning, including the development of a strategic vision for IT, should be integrated with the organizational planning process (Tapscott & Caston, 1993). This integration should ensure alignment of the strategic vision for IT with the organizational strategic vision and alignment of the IT strategic goals with organizational strategic goals (refer back to Figure 1). A number of researchers (Feigen, 1997; Harvey Jones, 1998; Tapscott & Caston, 1993) also promote the inclusion of non-strategic level employees in developing the strategic vision for IT to achieve "buy in" for the vision.

Wilson (1992) seconds the argument for a group as opposed to individual planning process. Reich and Benbasat (2000) found that "only a vague idea exists of how (IT) visions are nurtured and refined over time". This points to problems in both the development and communication of IT visions.

Although we are unaware of any objective supporting data, our experience strongly suggests that, while few organizations have developed explicit strategic IT visions, many organizations have evolved implicit strategic IT visions. Current research (Brabston, Zmud & Carlson, 2000) indicates that most organizations only develop an IT function vision, with little regard for the merits of a strategic IT vision. This is one of the reasons that information technology so frequently does not add its expected value to an organization. The lack of a strategic vision for IT results in various entities within an organization taking a shotgun or unplanned approach to their IT management and investment. Without a vision of what IT should be in the organization, development of a plan to manage IT has no foundation on which to build (c.f., James, 1985; Knorr, 1993; Nanus, 1992; Taylor, 1994; Wilson, 1992).

After a strategic vision for information technology has been developed, it must be effectively communicated to all managers and employees who need to know and understand – and hopefully, agree with – the strategic vision for information technology. Studies show that this is more difficult to accomplish than it would appear (Brabston, 1993; Brabston, Zmud & Carlson, 2000; Coulson-Thomas, 1993). Appropriate use of multiple channels for repeated messages in different media (e.g., printed, online, e-mail, meetings, etc.) can help the message to get through. It appears that those who are higher in the organization's hierarchy receive messages about the strategic vision for IT across more channels and understand the messages more readily than those lower in the hierarchy (Brabston, Zmud & Carlson, 2000). It also appears that many organizations use a "filter-down" approach, using informal communication means rather than a formal plan to communicate their strategic vision for IT (Brabston, Zmud & Carlson, 2000).

Feeny, Edwards, and Simpson (1992) found that the quality of the relationship between the CEO and the CIO is affected by the shared vision held by these two executives of "the role of IT as an agent of transformation" (p. 435). In contrast to the filter-down approach and agreeing in essence with Feeny, Edwards, and Simpson, Reich and Benbasat (2000) found that shared domain knowledge held by both business and IT executives (presumably achieved by communication) had a significant effect on long-term alignment. They proposed that substantial effort should be aimed at building this shared knowledge. In a previous study, Reich and Benbasat (1996, p. 55) had a similar finding, that "understanding of current objectives and shared vision for the utilization of information technology are the most promising potential measures for short-and long-term aspects of the 'social' dimension of linkage," which we call alignment here.

FUTURE TRENDS

There is little ongoing research about strategic vision for IT, its implementation, and its communication. Carr (2003) made a case for an approach to IT as a commodity utility, just like electricity. Carr's view rejects the importance of a strategic vision for IT, defining IT as a support tool rather than a strategic tool. Other authors came forward to offer support for the strategic view of IT, rejecting the strict IT-as-support-only view (Bartholomew, 2003; Delisi, 2003; Keefe, 2003; Schrage, 2003). This debate is ongoing among both scholars and managers.

Another issue is that some feel that IT strategy – whether general or specific – should be kept on a need-to-know basis and not communicated throughout an organization and its strategic partners. Certainly, there is good reason for keeping the competition in the dark but not at the expense of distributed decision-making capability and progress toward strategic goals within the organization. It is the author's view that strategic vision can be communicated, but perhaps, the strategic goals that are derived from that vision should be kept to a need-to-know basis.

Sabherwal and Chan (2001) found that alignment of business and IT strategies does in fact matter, but only for proactive organizations – those they categorized as prospectors and analyzers – but not for defensive organizations. This makes sense since "defender" organizations would primarily focus their IT plans on responding to their external competitive environment and not on longer-range plans of the organization as a whole.

CONCLUSION

There is little wonder, given the communication environment for communicating or implementing the strategic vision for information technology, that this has not become a strategic necessity, particularly for information-intensive organizations. In this enlightened digital age, it is surprising that so few organizations have a strategic vision for IT while so many have a strategic vision for their IT function. Yet the potential of developing and effectively communicating a strategic vision for information technology still remains high.

REFERENCES

Bartholomew, D. (2003). Yes, Nicholas, IT does matter. *Industry Week*, *252*(9), 44.

Boland, R.J., Jr., & Tenkasi, R.V. (1995). Perspective taking and perspective making in communities of knowing. *Organization Science, 6*(4), 350-373.

Boynton, A., & Zmud, R.W. (1987). Information technology planning in the 1990's: Directions for practice and research. *MIS Quarterly, 11,* 59-72.

Brabston, M.E. (1993). Effectiveness of channel use for communication of strategic vision for information technology to organization members. *Proceedings of the 24th Annual Meeting of the Decision Sciences Institute*, Washington, D.C.

Brabston, M.E., Zmud, R.W., & Carlson, J.R. (2000). Communicating strategic IT vision to organization members: A conceptual framework. In R. Papp (Ed.), *Strategic information technology: Opportunities and advantage* (pp. 82-104). Hershey, PA: Idea Group Publishing.

Carr, N.G. (2003). IT doesn't matter. *Harvard Business Review, 81*(5), 41-49.

Collins, J.C., & Porras, J.I. (1991). Organizational vision and visionary organizations. *California Management Review, 34,* 30-52.

Coulson-Thomas, C. (1993). Strategic vision or strategic con? Rhetoric or reality. *International Review of Strategic Management*, *4,* 87-104.

Delisi, P. (n.d.). Comments on: Is IT still strategic? *CIO*. Retrieved May 2, 2003, from *http://comment.cio.com/comments/11683.html*

Feeny, D.F., Edwards, B.R., & Simpson, K.M. (1992). Understanding the CEO/CIO relationship. *MIS Quarterly, 16*(4), 435-448.

Feigen, M. (1997). In discussion with Lloyd Bruce, "Real change leaders: The key challenge for management today". *Leadership and Organisation Development Journal*, *18*(1), 37-40.

Harvey Jones, J. (1998). *Making it happen*. London: Fontana.

Henderson, J.C., & Venkatraman, H. (1999). Strategic alignment: Leveraging information technology for transforming organizations. *IBM Systems Journal, 38*(2/3), 472-485.

Hoffman, T. (1994, November 14). Bank execs, IS not talking. *Computerworld*, *28,* 46.

James, P.N. (1985, Fall). A framework for strategic and long-range information resource planning. *Information Strategy: The Executive's Journal,* 4-12.

Jarvenpaa, S., & Ives, B. (1990). Information technology and corporate strategy: A view from the top. *Information Systems Research, 1,* 351-376.

Keefe, P. (2003). IT does matter. *Computerworld, 37*(19), 20.

Knorr, R.O. (1993). A strategy for communicating change. *Journal of Business Strategy*, *14*(4), 18-20.

Mintzberg, H., Ahlstrand, B., & Lampel, J. (1998). *The strategy process*. Hemel Hempstead: Prentice Hall.

Nanus, B. (1992). *Visionary leadership*. San Francisco: Jossey-Bass.

Parker, M.M., & Benson, R.J. (1991). Why business strategy should not follow financial systems. *Financial and Accounting Systems*, *6*(4), 20-29.

Reich, B.H., & Benbasat, I. (1996). Measuring the linkage between business and information technology objectives. *MIS Quarterly*, *20*(1), 55-81.

Reich, B.H., & Benbasat, I. (2000). Factors that influence the social dimension of alignment between business and information technology objectives. *MIS Quarterly, 24*(1), 81-113.

Robbins, S.R., & Duncan, R.B. (1988). The role of the CEO and top management in the creation and implementation of strategic vision. In D.C. Hambrick (Ed.), *The executive effect: Concepts and methods for studying top managers* (pp. 205-233). Greenwich, CT: JAI Press.

Sabherwal, R., & Chan, Y.E. (2001). Alignment between business and IS strategies: A study of prospectors, analyzers, and defenders. *Information Systems Research, 12*(1), 11-33.

Schoemaker, P.J.H. (1997). Disciplined imagination: From scenarios to strategic options. *International Studies of Management and Organization, 27*(2), 43-70.

Schrage, M. (2003). Why IT really does matter. *CIO, 16*(20), 1.

Tapscott, D., & Caston, A. (1993). The demise of the I.T. strategic plan. *I.T. Magazine, 25*(1), 28-35, 50.

Taylor, G. (1994). The challenge of the 90's. *Inform, 8*(5), 39, 42.

Wilson, I. (1992). Realizing the power of strategic vision. *Long Range Planning, 25*(5), 18-29.

KEY TERMS

Alignment: In this context, alignment means the alignment of the strategic vision for information technology with the organization's strategic vision, ensuring the strategic vision for information technology supports the organization's strategic vision.

Operational Goals: Very specific targets that represent the desired outcomes the organization expects to achieve over the short term; operational goals are derived from the tactical goals.

Organizational Strategic Vision: A vision statement that determines where the organization will go and how it will get there in general; contrasted to a mission statement, which is more specific and shorter in term.

Strategic Goals: Broad targets that represent the desired outcomes the organization expects to achieve over the long term.

Strategic Vision for Information Technology: A vision statement of how the organization will use information technology to achieve its organizational vision. Like the organization's strategic vision, it is not as specific as is the strategic vision for the information technology function.

Strategic Vision for the IT Function: A vision statement that articulates the organizational role of the IT function as well as the dominating values applied in carrying out these roles; affects the function of the IT department but not necessarily the rest of the organization directly.

Tactical Goals: More specific targets that represent the desired outcomes the organization expects to achieve over the middle term; tactical goals are derived from the strategic goals.

Vision: A statement of values and beliefs that sets forth the organization's overriding beliefs.

Strategically-Focused Enterprise Knowledge Management

Robert J. Mockler
St. John's University, USA

Dorothy G. Dologite
Baruch College, USA

INTRODUCTION

Many of the knowledge management systems, such as the ones described in this entry, were originally computer information systems to which were added knowledge expertise that complemented the information and data communicated. As use of the Internet expanded and intranets within companies were developed, many new knowledge expertise exchange systems were established. As seen in the following discussions, many knowledge systems are therefore mixed, that is, integrated with traditional information and decision support systems, as at the consulting firm, while many, such as Xerox's, focus specifically on expertise knowledge storage and transfer and so can be designated *focused* knowledge management systems. Expert knowledge-based systems generally are *pure* knowledge systems (Mockler, 1992; Mockler & Dologite, 1992).

BACKGROUND

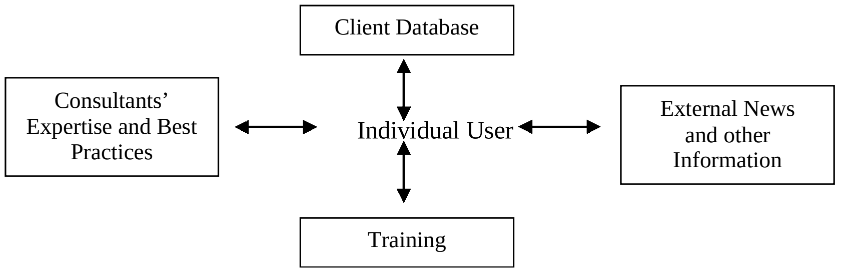
Narrowly defined, knowledge refers to practical skills or expertise gained from actual experience. In practice, however, knowledge management generally refers to the process of identifying and generating, systematically gathering, organizing and providing access to, and putting to use anything and everything that might be useful to know when performing some specified business activity. The knowledge management process is designed to increase profitability and competitive advantage in the marketplace.

As seen in the Key Terms section at the end of this article, since the knowledge management process involves keeping informed about and getting to know anything useful to doing a business task, the process can encompass data, information, and knowledge. Further, the knowledge management process can involve employing any useful and practical means of communication and storage, manual or electronic. Useful manual means might include: service manuals; professional publications; personal correspondence and conversations; special studies and reports; client correspondence and summaries; competitor role-playing; sales force feedback; current news; supplier feedback; and the like. Useful computer-based electronic technologies might include: e-mail; hierarchical, network, and relational databases and data warehouses; group decision support systems; Lotus Notes; intranets and Internet Web sites; browsers and search engines; expert and knowledge-based systems; and the like.

Because of the wide range of concepts and activities involved, the term knowledge management can more easily be understood by examples. Figure 1 outlines the knowledge management system (KMS) at a large consulting firm (Engoron, 1998).

The strategic focus is the individual consultant who needs access to data, information, and knowledge in order to do his/her job. Since consulting is its business, the

Figure 1. Knowledge management system at a major consulting firm



system is strategic. The system provides this access in large part electronically.

At the top of Figure 1 is a large computer database of information about clients, covering past assignments, consultants who worked on the assigned projects, outcomes, organized data on the company involved, and contacts who can provide further information. On the right, there is a system incorporating expert knowledge-based systems that scans news media and library resources daily and daily directs relevant intelligence material to different consultants. On the left is a database of consultants' expertise or knowledge including that acquired from experience during past assignments. This includes written summaries of what was learned from the assignments, videos in which consultants describe the highlights of their experiences or general knowledge, and contingent best practices guidelines in different areas (such as strategic alliances, all marketing and production areas, human resources management, and the like). At the bottom, there is available a bank of online training programs, which a consultant can make use of (privately) to sharpen skills needed to improve job performance.

On any given day that a consultant receives a new assignment, he/she could immediately review current relevant information in the media (intelligence) about the client and project area, gather information quickly about the client and past assignments involving the client, review the related knowledge expertise of other consultants, and brush up on needed skills. At the same time, the consultant would make use of any relevant personal knowledge sources. The system is a good example of using a knowledge system to strategically manage resources; that is, of a strategic management knowledge system.

Not all knowledge management systems are that complex or that multidimensional in scope. Some are narrowly focused on single activities. For example, Xerox in 1996 developed Eureka, an intranet communication system linked with a traditional corporate computer database that helps service representatives share repair tips; that is, knowledge. To date, more than 500 tips have been entered by Xerox technicians, and this practical knowledge is available to all via their laptops. For employees scattered around the world who travel often, the ability to share such know-how means they do not have to miss out on the kind of knowledge typically exchanged at the water cooler (Hickens, 1999).

A number of key characteristics of knowledge management can be identified from company experiences. These can apply to strategic and operational knowledge systems. First, the types of knowledge management systems vary considerably depending on the company situation requirements, a contingency perspective. Second, knowledge generation involves identifying knowledge relevant to strategic business activities, as well as its source and the way it is used or exploited. Third, *structuring* refers to designing knowledge management systems to capture and deliver the knowledge generated; such structures can range from simple ones involving individual business process areas, as at Xerox, to multidimensional complex enterprise-wide ones, as at the consulting firm. Their content can involve any company activity/ business process or combination of them. Fourth, *diffusing* or communicating any type of relevant data, information, or knowledge involves transferring and absorbing knowledge to put it to work. In the company experiences studied, the main means of diffusion was electronic and audio/video tools. Knowledge is also very often continually transferred and absorbed informally through personal interaction.

Due to the complexity of knowledge management systems, they are best defined by a description of their characteristics, as discussed previously.

ENTERPRISE-WIDE KNOWLEDGE MANAGEMENT SYSTEMS

This section begins by describing a range of systems in use by businesses followed by a discussion of the contingent situation requirement factors of knowledge management systems. This section then continues with discussing the impact of these factors and others on KMS developments as well as their implementation and use in such areas as: knowledge generation and selecting or developing strategic structure, content, and design of the system

Company Examples

The Ford Motor Company case provides an example of how at a large firm the company-wide strategic knowledge systems are closely linked to and dependent on computer information systems (Austin, 1997, 1999). As part of an integration program in the early 1990s, computer information systems at Ford were standardized across the company, which enabled installation of an external Internet network – extranet – with appropriate Web sites linking Ford with its suppliers and with its customers. Most of these were used initially for communication of information on available models, prices and availability of supplies, and other information (that is, targeted organized data). It also enabled development of an internal company intranet system, which also focused mainly on information conveyance initially.

The system also, however, served as a basis upon which to develop broader, more strategic knowledge systems. For example, in the design area, as auto design and development facilities were more closely coordinated worldwide, knowledge about solving design problems and inconsistencies could now be resolved using the intranet, a knowledge exchange process based on experiential expertise. Knowledge about lessons learned from experience in other business process areas, such as manufacturing, could also now be exchanged, since a worldwide system with Web sites was in place.

Complex strategic knowledge management systems can also focus on critical business activity areas. For example, strategic alliances are extremely important to multinational companies today (Mockler, 1999; Sparks, 1999). They involve, however, complex human and business processes whose management requires in-depth expertise gained from experience. Capturing this developing knowledge base is a knowledge management activity. As a company undertakes alliances and begins learning from successes and failures, leaders in alliance management within a company emerge. These leaders, who are essentially gurus with experience and knowledge gained from experience, are the firm's initial imbedded alliance expertise capability. This initial experiential expert knowledge base in successful firms is extended in several ways. First, formal processes and procedures and a staff capable of managing alliance processes are developed. This is the initial knowledge depository for future use. The steps taken to collect, store, and disseminate this knowledge and to train people in order to further institutionalize alliance capabilities vary at different firms (Harbison & Pekar, 1997a, b, c).

Hewlett-Packard (H-P), for example, found that general seminars for managers on alliances were not enough. Managers needed H-P specific information on the best practices guidelines developed from H-P alliance experiences. A database of case histories, tools kits, checklists, and best practices was, therefore, developed and incorporated into training seminars. This database material was supplemented with studies of the best practices of other companies (Harbison & Pekar, 1997a, b, c).

In general, such a knowledge database would include a specific company's experiences with each of its alliance partners in each of the applicable best practices guidelines areas, areas that are outlined in alliance guidebooks (Mockler, 1999). These areas include strategic planning, negotiation, alliance structures and contracts, operational planning and management, and control. Companies such as Ford, IBM, and Dun & Bradstreet, are in various stages of creating such company-specific database repositories; most often these are mixed systems – using computers and other approaches, as for example at H-P. The alliance knowledge bases are Web sites that are accessible from laptop computers by consultants or service personnel at clients' offices (Harbison & Pekar, 1997a, b, c).

As part of their strategic knowledge management systems, dissemination of this knowledge is usually supplemented through seminars and workshops. BellSouth, for example, has offered a two-day alliance workshop for 150 senior managers, a major means of developing personal information networks to encourage ongoing knowledge dissemination. H-P has conducted 50 two-day seminars on alliances for its top 1000 executives prior to 1999 (Harbison & Pekar, 1997a, b, c).

Contingent Solution

As seen from the company experiences described, KM is a contingent process. While it is stimulating and useful to study the different approaches of other so-called models, the final solution will be the one that meets specific situation requirements factors, such as:

- The company's size and nature of its business.
- The company's competitive position.
- The type of business activity(ies) supported.
- The knowledge used for strategic aspects of the business activities targeted.
- The type of knowledge.
- The people involved.
- The state of the organization's culture.
- Available company resources.
- The technologies available and needed.

Knowledge Generation

Knowledge comes from inside and outside the organization. The main criteria are whether and how it is useful in making business activities and processes more effective and efficient (Moore, 1999). Its source – whether created from scratch or borrowed – is a secondary consideration. For example, some companies have an award for the best stolen idea! Five methods of generating knowledge can be identified: acquisition; dedicated resources; fusion; adaptation; and knowledge networking.

Knowledge can be acquired in many ways – for example, through strategic alliances, acquisitions of patents or other companies, engaging consultants, outsourcing research and development, and hiring experts. While each of these sources is useful, additional steps are needed to assimilate acquired knowledge into existing organizations.

Dedicated resources were established at H-P to preserve and transmit company knowledge about strategic alliances. H-P also set up a permanent knowledge system designed to preserve and transmit knowledge to targeted users in a systematic continuing way. This is referred to as institutionalizing or imbedding knowledge within a company.

Fusion involves bringing together people of different perspectives to create new synergies. Sometimes referred to as creative chaos, such interactive collaboration is often needed to create new knowledge. There are several guidelines that help make such brainstorming cross-fertilization work: identify key knowledge workers who might be brought together; provide the time and incentives to fuse knowledge; create specific projects and project goals that inculcate attitudes encouraging such working together; reinforce a sharing culture through incentives and leadership.

Adaptation often arises from competitive market pressures, such as when new technologies are invented, competitors introduce new products, and social changes occur. These pressures can drive knowledge generation. Knowledge is also generated by informal self-organizing networks or "webs" within organizations that may over time become more formalized. There exist communities of knowers within organizations, united by common interests or a common professional vocabulary, who talk with each other in person, by phone or through e-mail or groupware – to share knowledge and solve problems, in essence by word-of-mouth. This kind of knowledge is hard to capture systematically, though at times, such as at Xerox, it can be done. Such informal webs can be nurtured through creating common meeting places, allowing time for people to interact, and generally developing a corporate sharing climate.

Developing Strategic Structure, Content, and Design for the System

These initial decisions are dictated by strategic needs. For example, first the needs of strategic operations are identified – as in the consulting operation described earlier or in the strategic alliance area at H-P. Second, the sources of this knowledge – the people and data/information reservoirs – are identified; for example, the consulting firm's partners' experiences in former assignments, the alliance managers' experiences at H-P, and the repair workers' expertise at Xerox. Third, the form it is in and its suitability to available manual and electronic tools. Five different media were used, for example, at the consulting company: video and audio tools, databases, expert systems, e-mail, and personal interviewing. These were supplemented by personal interaction webs. H-P's system included a variety of electronic and manual approaches in their system, which combined training and computer techniques in a much different way. Fourth, not all knowledge is easily systematically stored, transmitted, and absorbed; for example, knowledge involving planning in rapidly changing markets often requires tacit intuitive knowledge that is not readily articulated and codified. For this segment of a KMS, personal networking or expert systems may be more useful. In contrast, repair procedures learned from experience can be codified, as at Xerox.

Davenport (1997) has described the structure of a KMS at Monsanto. Monsanto's knowledge management architecture system aimed to allow the firm's 30,000 employees to share knowledge and information. In making global knowledge locally available, Monsanto was mimicking GE's strategic posture: combining the knowledge benefits of a large firm (quantity and diversity) with the benefits of a small one (accessibility to knowledge).

As for the system's structure, in classifying existing knowledge and information a distinction was made between quantitative structured content and relatively unstructured qualitative content. Different tools were used to store and manipulate each kind: relational databases with desktop access through appropriate query software for structured material; Web pages and Lotus Notes for unstructured material. This avoided distortions arising from forcing unstructured material into an artificially rigid structure.

The system also provided definitions of key terms, such as "customer," "product" and "material," a necessary step in organizing intellectual material into a single system. Such common definitions were also necessary common ground for communicating and sharing knowledge across boundaries, but they had to be limited to only the most necessary terms, since some local truths and nuances can be lost during the standardization process. Gatekeepers were charged with continually refining this aspect of the system, as well as with organizing and identifying the most important aspects. A large body of material randomly assembled could be counterproductive; editing was needed, as were clearly defined paths to needed knowledge. All of the major systems examined have such gatekeepers and facilitators.

FUTURE OUTLOOK AND CONCLUSION

As discussed in this article, the current and future success of KMS is and will be highly dependent on the strategic fit of structure and content with strategic requirements or critical success factors in the situation. The success of knowledge management also depends and will continue to depend on the participation of people sharing their knowledge expertise with others, which in turn can depend on the way the system is designed, implemented, and managed (the operational fit), as well as on the degree to which a firm has a "learning" organization culture (Lucier & Torsilieri, 1997). Nurturing this sharing culture in the future will require very active leadership by a knowledge management champion, since very often people are reluctant to share their expertise (Manchester, 1999b).

REFERENCES

Austin, R.D. (1999). *Ford Motor Company: Supply chain strategy*. Boston: Harvard Business School.

Austin, R.D., & Coteleer, M. (1997). *Ford Motor Company: Maximizing the business value of Web technologies*. Boston: Harvard Business School.

Engoron, F. (1998). *Organization effectiveness and development.* PricewaterhouseCoopers, New York: New York University (Center for Research on Information Systems).

Glasser, P. (1998, December 15-January 1). The knowledge factor. *CIO,* 108-118.

Harbison, J.R., & Pekar, P. (1997a). *A practical guide to alliances: Leapfrogging the learning curve*. New York: Booz-Allen & Hamilton.

Harbison, J.R., & Pekar, P. (1997b). *Cross-border alliance in the age of collaboration*. New York: Booz-Allen & Hamilton.

Harbison, J.R., & Pekar, P. (1997c). *Institutionalizing alliance skills: Secrets of repeatable success*. New York: Booz-Allen & Hamilton.

Hickens, M. (1999, September). *Management Review,* 40-45.

Lucier, C.E., & Torsilieri, J.D. (1997). Why knowledge programs fail: A CEO's guide to managing learning. *Strategy & Business*, *9*,64-79.

Manchester, P. (1999a, November 10). A marriage of culture and technology. *Financial Times,* 1.

Manchester, P. (1999b, November 10). Fundamental dilemma over ownership of knowledge. *Financial Times,* 5.

Mockler, R.J. (1992). *Developing knowledge-based systems using an expert system: A guidebook for general business managers and computer information technicians*. Upper Saddle River, NJ: Prentice-Hall/Macmillan.

Mockler, R.J. (1999). *Multinational strategic alliances*. New York: John Wiley & Sons.

Mockler, R.J., & Dologite, D.G. (1992). *Expert systems: An introduction to knowledge-based systems*. Upper Saddle River, NJ: Prentice-Hall/Macmillan.

Sparks, D. (1999, October 25). Special report: Partners. *Business Week*, 106-134.

KEY TERMS

Data: Something given or admitted as a fact on which an inference may be based. Simple observations of the world, which are often quantified, and easily structured, captured on machines, and transferred. The number of "baby boomers" born in a given year is data.

Expert Systems: A computer system that attempts to replicate what human experts normally do. Human experts make decisions and recommendations, such as what company strategy to follow or who to give a bank loan to, and do tasks, such as adjust temperature controls in a manufacturing plant. They also assist (or help) and train others to do tasks and to make decisions. So do expert systems.

Information: Knowledge derived from reading, observation or instruction, at times consisting of unorganized or unrelated facts or data. Data endowed with relevance and purpose, for example, a firm's balance sheet and income statement.

Intelligence: Information, news, and advice. Brain power or cognitive skills. IBM uses the term "business intelligence systems" to describe their mixed integrated knowledge systems.

Knowledge: Familiarity gained by actual experience, practical skill, or expertise. The act, or state, of understanding. Valuable information arising from reflection, synthesis, and other cognitive activities of the human mind. It is often, but not always, hard to structure, difficult to capture on machines, sometimes tacit, and hard to transfer. *Knowledge* includes data and information (organized data which are relevant and purposeful), and knowing how to apply and use that information and data. The term "ledge" means to put to work or apply. The word knowledge, therefore, means knowing how to put to work what we know, and so in popular usage can in certain situations encompass information and data.

Learning: In this context it means knowledge and skills acquired by instruction or study.

Technology: Applied science, systematically organized knowledge.

S

Strategies of E-Commerce Business Value Optimization

Ada Scupola
Roskilde University, Denmark

INTRODUCTION

The Internet economy is becoming an integral part of many countries' economies, creating new jobs, giving rise to new companies like the dot coms and transforming traditional jobs and traditional companies. The Internet is increasingly becoming a part of the basic business model for many companies as organizations around the world are adopting new e-business models, integrated solutions to explore new ways of dealing with customers and business partners, new organizational structures and adaptable business strategies (Singh & Waddell, 2004). There are many definitions of electronic commerce (e.g., Wigand, 1997). Here, a classic definition by Kalakota and Whinston (1996) is adopted, where e-commerce is "the buying and selling of information, products and services via computer networks today and in the future via any one of the myriad of networks that make up the 'Information Superhighway (I-way)'" (p.1). A distinction between physical and digital products can be made. A digital product is defined as a product whose complete value chain can be implemented with the use of electronic networks; for example, it can be produced and distributed electronically, and be paid for over digital networks. Examples of digital products are software, news, and journal articles. The companies selling these products are usually Internet-based "digital dot coms" such as Yahoo and America Online. On the contrary, a physical product cannot be distributed over electronic networks (e.g., a book, CDs, toys). These products can also be sold on Internet by "physical dot coms", but they are shipped to the consumers. The corporations using electronic commerce are distinguished into "bricks and mortar" companies, hybrid "clicks and mortar" companies (such as Amazon.com) and pure dot coms (Barua & Mukhopadhyay, 2000).

Many studies from the early days of deployment of information technology (IT) in organizations have struggled to measure the business value and profitability of information technology (Barua & Mukhopadhyay, 2000). Many of these studies have showed that productivity gains are small or non existent and that the effects of information technology and electronic commerce have to be often looked upon from a competitive advantage point of view (Barua, Konana, Whinston, & Yin, 2001; Porter & Miller, 1985; Scupola, 2003). Recent research has argued that increasing the business value of electronic commerce to a corporation is important to shift the focus from whether electronic commerce creates value to a company to "how to create value" and "how to optimize such value" (Barua, Konana, Whinston, & Yin, 2001). This can be achieved by exploring complementary relationships between electronic commerce, strategies and complementarity (Scupola, 2002, 2003).

BACKGROUND

Since the early days of IT use in commercial organizations, researchers and professionals have struggled with the problem of measuring the bottom line contribution of IT investments (Scupola, 2003). Six main areas of IT business value research can be distinguished: information economics-based studies; early IT impact studies; production economics studies that did not find positive impacts; microeconomics studies that found positive impacts of IT; business value studies; and studies involving complementarity between IT and non-IT factors. The information economics-based studies date back to the 1960s, and though relevant to the economic contribution of IT investments, they mainly focus on the changes in information due to IT use and their impact on the single decision-maker. Therefore, while the information economics approach is theoretically sound and rigorous, its unit of analysis, which is either the individual or team decision, makes it difficult to obtain meaningful and insightful results in broader organizational contexts (Barua & Mukhopadhyay, 2000).

In the early 1980s, a stream of research emerges focusing on assessing the contribution of IT investments to performance measures such as return on investment and market share (Barua, Konana, Whinston, & Yin, 2001; Barua & Mukhopadhyay, 2000). The majority of these studies did not find much positive correlation between IT investments and firm performance metrics up to the early 1990s. The lack of correlation between IT investments and productivity made Roach (1988, 1989) to coin the term "IT productivity paradox".

In the 1990s, research on measuring the economic and performance contributions can be divided into two main streams: one based on production economics and one based on "process-oriented" models of IT value creation. The IT production studies based on production economics hypothesize that IT investments are inputs to a firm's production function. These studies (e.g., Brynjolfsson & Hitt, 1993, 1996) finally started finding signs of productivity gains from IT. For example, Brynjolfsson and Hitt (1996) identify three sources of IT value to a corporation: productivity, consumer value, and business profitability. The study shows that information technology contributes to increases in the productivity and consumer value, but not business profitability. Simultaneously, process-oriented studies started hypothesizing relationships between IT and other input factors to performance measures at various levels of aggregation. These studies (e.g., Kauffman & Kriebel, 1988) have laid the foundation of the business value approach to the impact of IT on firm performance. This approach on the contrary of the production function-based approach might have the explanatory power to point out where and how IT impacts are created and where management should act to increase the payoff from IT investments. These explanations are more difficult to get with production function-based approaches since they operate at a very high level of aggregation, thus making it difficult to distinguish between different types of IT investments and their impacts on specific areas of business. After having dispelled the productivity paradox, new refinements to existing approaches are emerging to measure the contribution of IT to business performance. An important stream of research is pointing to complementarity theory to investigate the interactions between IT and other organizational factors (e.g., Barua, Konana, Whinston, & Yin, 2000, 2001; Barua, Lee, & Whinston, 1996; Barua & Mukhopadhyay, 2000). In fact, production economics and business value approaches have mostly ignored the synergy between IT and other related factors such as the level of fit with business strategies, employee empowerment, and team orientation of business processes. Barua and Mukhopadhyay (2000) present a generalized business value complementarity model that explores the synergies among such factors. The basic idea of their business value complementarity model (BVC) suggests that investments in IT should be first related to intermediate performance measures such as time to market, customer service, response time and extent of product mass customization to be able to see any positive results from such investments. In a second moment, the intermediate performance measures can be related to high-level performance metrics such as profitability, return on investment (ROI), market share. The focal point of a business value complementarity model is the complementarity that potentially exists at each level of the model (Barua & Mukhopadhyay, 2000; Barua, Konana, Whinston, & Yin, 2001; Scupola, 2003).

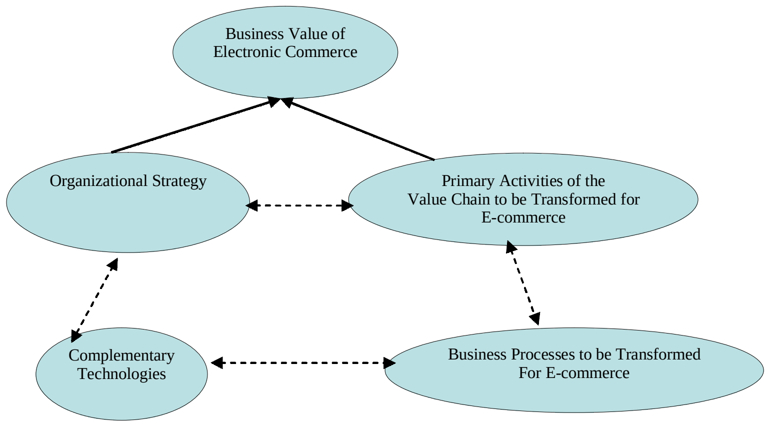
The advent of the Internet, based on open standards and a universal Web browser, raises the question of whether investing more in Internet technology lead to a better financial performance in electronic commerce. This calls for more attention to the specific business processes that have to be reengineered for online commerce and the way they should support the company strategy (Scupola, 1999, 2003).

MAIN THRUST OF THE ARTICLE

A business value complementarity (BVC) model of electronic commerce could be used as a methodology to optimize e-commerce initiatives when entering the e-commerce arena (Scupola, 2003). The BVC model presented here is based on the value chain (Porter, 1980), the theory of BVC (Barua, Lee, & Whinston, 1996; Barua & Mukhopadhyay, 2000; Barua et al., 2002; Milgrom & Roberts, 1990) and the concept of strategy (Porter, 1982). In this model, it is hypothesized that complementarity (represented in Figure 1) exists between the variables of the same level and different levels of the model. It is furthermore hypothesized that the exploration of complementarities and possible synergies between the company strategy, the primary activities of the value chain, corresponding business processes and supporting technologies should: 1) maximize the business value of electronic commerce to a corporation and 2) lead to a better fit between the overall organizational strategy, the business processes that have to be transformed for the online market place, and the information system that should be designed and implemented to support these strategies. The exploration of complementarities, it is hypothesized, can also contribute both to avoid investments into an information system that could not be used at a later point if new e-business processes should be added to the system and avoid the implementation of a business model that does not correspond to the corporation's strategy. It is argued that to succeed in electronic commerce it is important to reengineer the parts of the value chain and the corresponding business processes relevant to the product in question and the company strategy.

The main objective of the model is to make the business value of electronic commerce as close to optimal as possible in terms of one of the performance measures, such as company profitability, competitive advantage, increase in market share, shareholder value or customer satisfaction. This can be done by exploring

Figure 1. Business value complementarity model of electronic commerce



complementarities among the dependent variables of the model: the company strategy, the activities of the value chain, the corresponding business processes, and the technologies available to transform these activities and processes for the marketspace.

Furthermore, to succeed in electronic commerce, it is important to reengineer the parts of the value chain and the corresponding business processes relevant to the product in question and the company strategy. For example, the strategy or combination of strategies a company wants to pursue is relevant for the primary activities of the value chain, and the corresponding business processes that have to be implemented online. The strategy is also relevant to the classes of technologies that have to be chosen to enter the electronic marketplace. For example, a company can use electronic commerce to implement a cost leadership strategy, or to become the low cost producer in the industry. Once decided upon the strategy, it is important to explore complementarities between the strategy and the value chain activities in order to implement online all those activities that would support an optimal implementation of the strategy chosen.

The number of primary activities and corresponding business processes that should be transformed for the marketspace depends also on the company's type of product and strategy. It is important to take into consideration complementarities among the different activities of the value chain when reengineering for electronic commerce. The more activities of the value chain are simultaneously conducted online, the more likely it is that the business value of electronic commerce will be optimized. The adoption of a holistic approach in redesigning the primary activities for electronic commerce would, thus, be a more successful strategy than reengineering only one or some at a time. This is due to potential complementarities between the different activities, which lead to a better performance in one if the others are also reengineered for online commerce.

Furthermore, each business process of each activity of the value chain could be reengineered for e-commerce. This model argues that the exploration of complementarities among the different business processes and the simultaneous transformation of all the complementary processes of a particular activity for online commerce would lead to a higher business value than if only one or a casual numbers of processes were reorganized online (Scupola, 1999).

In the design phase, it is important to consider potential complementarities between the business processes that have to be redesigned for online commerce and the supporting technologies. The exploration of this complementarity should lead to an optimal system design that also offers possibilities for further expansion if other online business processes should be added in the future. For example, electronic search of the company's information will give more accurate and quicker results, the faster and more advanced the search engine is and the better built are the user interface and the repository systems.

Finally, the exploration of complementarities between the different technologies used to implement the system for electronic commerce could bring to a more robust and flexible computer system than a system built without the exploration of complementary relationships between the different component technologies. For example, end user interfaces and repositories are complementary technologies in the sense that the better designed the repository system, the simpler the user interface can be.

FUTURE TRENDS

The studies on IT productivity and business value conducted over the last decade have showed positive impacts of IT investments on firms' productivity both with respect to labor and other non-IT capital used by organizations (Barua & Mukhopadhyay, 2000). However, Internet-based technologies, with their open standards and wide applicability, raise again the issue of profitability and business value of investing in such technologies. Furthermore, the fact that Internet is giving rise to a "new economy", raises a number of questions among which: How productive are the players in this new economy? Does e-commerce increase the profitability and business value of brick and mortars and hybrid click and mortars companies? For dot coms, do more investments in Internet commerce technologies necessarily lead to a better performance of the company? And especially, if all companies have equal access to Internet-based technologies, what are the factors that differentiate their performance in e-commerce?

Recent literature investigating the business value and profitability of electronic commerce is focusing on the exploration of complementary relationships between electronic commerce technologies and other factors in order to see positive returns from investments in these technologies (Barua, Konana, Whinston, & Yin, 2000, 2001; Barua & Mukhopadhyay, 2000; Scupola, 2003). For example, Barua, Konana, Whinston, and Yin (2001) develop a framework of electronic commerce business value that identifies linkages between performance drivers such as Internet applications, processes and electronic business readiness of customers and suppliers and operational excellence and financial metrics. They argue that "firms engaged in electronic business transformation must make synergistic investments and commit resources not only in information technology, but also must align processes and customer and supplier readiness to maximize the benefits" (p.1).

Similarly, an empirical investigation of the business value of e-commerce in small, medium and large companies across Europe and USA (Barua, Konana, Whinston, & Yin, 2000) identifies a set of key e-drivers such as system integration, customer orientation of IT, supplier's orientation of IT, and internal orientation of IT. The study concludes that high performance companies have invested more effort and resources in these e-business drivers than companies who have not benefited from e-business.

To conclude, these studies show that ignoring complementarities in research on business value measurement might lead to misleading results. On the other hand, from a managerial point of view, the non exploration of complementary relationships between IT and related factors such as strategy, business processes, business models, incentives, and so forth, might lead to failure of investments in sophisticated electronic commerce systems and ventures. These considerations point to the need for more empirical as well as normative, prescriptive research on complementarity and business value of IT in general and electronic commerce technologies in particular.

CONCLUSION

Many companies are very skeptical about investing into electronic commerce technologies due to the lack of profitability, (or at least the difficulties to show positive return on IT investments) that until now has characterized the investments in IT and electronic commerce. Here, a framework that can be used as a methodology to analyze organizational strategies and technology choices in reengineering for electronic commerce has been presented. Companies should explore the potential complementarities existing between strategy, value chain activities, business processes and supporting technologies when entering the field of electronic commerce. This should lead to investments in electronic commerce systems that best support the company strategy, thus minimizing failures. This is a future challenge for corporations, industries and researchers.

REFERENCES

Barua, A., & Mukhopadhyay, T. (2000). Information technology and business performance: Past, present, and future. In R. Zmud (Ed.), *Framing the domains of IT management, projecting the future through the past*.

Barua, A., Konana, P., Whinston, A.B., & Yin, F. (2000, November-December). Making e-business pay: Eight key drivers for operational success. *IT Pro*. IEEE Publisher.

Barua, A., Konana, P., Whinston, A.B., & Yin, F. (2001, Fall). Driving e-business excellence. *Sloan Management Review*, *43*(1), 36-44.

Barua, A., Lee, S.C.H., & Whinston, A.B. (1996). The calculus of reengineering. *Information Systems Research*, *7*(4), 409-428.

Barua, A., Pinnell, J., Shutter, J., Wilson, B., & Whinston, A.B. (1999). The Internet economy indicators part II. Retrieved on June 19, 2002 from *http://www.internetindicators.com*

Brynjolfsson, E., & Hitt L.M. (1993). Information Technology and the Productivity Paradox: Review and Assessment. *Communication of the ACM*, 35, December, 66-77.

Kalakota, R., & Whinston, A.B. (1996). *Frontiers of electronic commerce.* Addison-Wesley.

Kauffman, R.J., & Kriebel, C.H. (1988). Modeling and measuring the business value of information technologies. In P.A. Strassman, P. Berger, E.B. Swanson, C.H. Kriebel, & R.J. Kauffman (Eds.), *Measuring the business value of information technologies.* Washington, DC: ICIT Press.

Milgrom, P., & Roberts, J. (1990). The economics of modern manufacturing: Technology, strategy and organization. *American Economic Review*, 511-528.

Porter, M. (1980). *Competitive advantage.* The Free Press.

Porter, M. (1982). *Competitive strategy*. The Free Press.

Porter, M., & Miller, V. (1985). How information gives you competitive advantage. *Harvard Business Review*.

Roach, S.S. (1988). Stop the Dice on Technology Spending, interview with G. Harrar, editor. *Computerworld Extra*, June 20.

Roach, S.S. (1989). *The case of the missing technology payback.* Presentation at the 10th International Conference on Information Systems, Boston, MA.

Scupola, A. (1999). The impact of electronic commerce on the publishing industry: Towards a business value complementarity framework of electronic publishing. *Journal of Information Science, 25*(2).

Scupola, A. (2002). A business value complementarity framework of electronic commerce. In M. Khosrow-Pour (Ed.), *Issues and trends of information technology management in contemporary organizations*. Hershey, PA: Idea Group Publishing.

Scupola, A. (2003). Organization, strategy and business value of electronic commerce: The importance of complementarities. In J. Mariga (Ed.), *Managing e-commerce and mobile computing technologies* (pp. 147-162). Hershey, PA: Idea Group Publishing.

Singh, M., & Waddell, D. (2004). *E-business innovation and change management*. Hershey, PA: Idea Group Publishing.

Wigand, R.T. (1997). Electronic commerce, definition, theory and context. *The Information Society, 13*, 1-16.

KEY TERMS

Business Processes: The specific processes into which each primary activity of the value chain can be decomposed.

Business Value: The overall value that an investment brings to a corporation. Examples of performance measures of the business value of electronic commerce can be: (1) profitability, that is, whether electronic commerce contributes to an increase in the profitability of the corporation; (2) competitive advantage that could be measured as an increase in market share, shareholder value or customer satisfaction.

Complementarity: Several activities are mutually complementary if doing more of any one activity increases (or at least does not decrease) the marginal profitability of each other activity in the group. Complementarities among activities imply mutual relationships and dependence among various activities whose exploration can lead to higher profitability.

E-Commerce: The buying and selling of information, products and services via computer networks and especially the Internet.

Internet Economy: A large collection of global networks, applications, electronic markets, producers, consumers and intermediaries.

Re-Engineering: The redesign of a corporation's business processes (or part of them) to take place over the Internet. The main goal is reduced costs, lower product cycle times, faster customer response, and improved service quality.

Strategy: A planning, rational process through which the company chooses a certain mode of development, among all of the possible ones, and maintains that direction through a well-defined period (design view). In the process view, strategy is a process that might change on the way, giving rise to an emergent strategy. The realized strategy might be different then the original intended strategy.

Value Chain: The activities of a corporation such as procurement, production, marketing and sales, and customer support.

Structural Text Mining

Vladimir A. Kulyukin
Utah State University, USA

John A. Nicholson
Utah State University, USA

INTRODUCTION

The advent of the World Wide Web has resulted in the creation of millions of documents containing unstructured, structured and semi-structured data. Consequently, research on structural text mining has come to the forefront of both information retrieval and natural language processing (Cardie, 1997; Freitag, 1998; Hammer, Garcia-Molina, Cho, Aranha, & Crespo, 1997; Hearst, 1992; Hsu & Chang, 1999; Jacquemin & Bush, 2000; Kushmerick, Weld, & Doorenbos, 1997). Knowledge of how information is organized and structured in texts can be of significant assistance to information systems that use documents as their knowledge bases (Appelt, 1999). In particular, such knowledge is of use to information retrieval systems (Salton & McGill, 1983) that retrieve documents in response to user queries and to systems that use texts to construct domain-specific ontologies or thesauri (Ruge, 1997).

BACKGROUND

Structural mining of texts consists of two related tasks: the task of partitioning text into components, for example, topics, sentences, terms, and so forth; and the task of finding relations among found components, for example, term and topic associations. Texts can be divided into three broad categories: free, structured, and semi-structured.

Free texts do not give the computer many road maps to the information they contain. To discover a road map in a free text requires a certain amount of data mining through parsing, statistical analysis, and/or machine learning. Novels and newspaper and journal articles are examples of free texts. Structured texts organize their content according to well understood road maps. Relational databases are structured texts where all of the relations between textual entities, that is, records, are known and can be readily obtained through well-defined queries. Semi-structured texts offer more structure than free texts but less than structured ones. HTML pages are semi-structured texts. While they offer a standard set of tags that point to the structural organization of information in them, they do not specify the types of information that the tags label or the relations among these types.

ISSUES IN TEXT MINING

The three fundamental problems in structural text mining are:

- Text Segmentation;
- Automatic Ontology (Thesaurus) Construction; and
- Information Extraction.

Text segmentation is a process of partitioning free texts into segments of content. The underlying assumption is that texts are intellectual artifacts that consist of words related to each other semantically in a number of complex ways (Bookstein, Kulyukin, Raita, & Nicholson, 2003). The intellectual process of producing texts incidentally leaves behind simple statistical regularities. Capturing those regularities through statistical analysis allows one to arrive at the structural organization of information in the texts.

The two most prominent approaches to text segmentation are statistical and qualitative. Statistical approaches to text segmentation (Hearst, 1997) first parse texts to identify primitive components, for example, sentences, and then combine those primitive components into larger segments by defining various similarity measures between pairs of components. For example, if components are represented as vectors of terms each of which is assigned a specific weight (1 or 0 in the basic case), the similarity between two components can be computed through a range of vector metrics: dot product, cosine of the angle between the vectors, a hamming distance, and so forth. Powerful as they are, statistical approaches to text segmentation have two drawbacks. First, statistical computations are based on the idea of statistical significance. Achieving statistical significance requires large quantities of data. Since many documents are small in size, the reliable discovery of their structural components using numerical methods alone is not always appropriate.

Second, numerical approaches frequently ignore the fact that text writers leave explicit markers of content structure in document texts. The presence of these markers in texts helps the reader digest the information contained in the texts. If these markers are ignored, the texts become much harder to navigate and understand. These intuitions are at the heart of qualitative approaches to text segmentation (Kulyukin & Burke, 2003). In these approaches, the structural organization of information in texts is discovered through mining free text for content markers left behind by text writers. For example, police crime reports and scientific journal papers have well defined structures that can be fruitfully mined for information. The ultimate objective of qualitative approaches is to find scalable data mining solutions for free text documents in exchange for modest knowledge engineering requirements.

Research in automatic thesaurus construction investigates ways to extract thesaurus relations from texts. A thesaurus is a set of terms plus a set of relations among them. Automatic thesaurus construction complements manual thesaurus construction, which, as the argument goes, is expensive in terms of expert time and effort and cannot respond in a timely manner to rapid changes in scientific fields. Automatic thesaurus construction is usually collection dependent, that is, it is done on a specific text collection at hand. Approaches to automatic thesaurus construction include statistical analyses of term co-occurrence data (Dagan, Lee, & Pereira, 1999), syntactic patterns used to extract semantic relations among terms (Grefenstette, 1994; Hearst, 1992), and ensemble methods that combine different information extraction techniques and rank their outputs by their utility to the task at hand, for example, manual query expansion during retrieval (Curran, 2002). Evaluation of automatic thesauri, that is, evaluation of the authenticity of found relations and their utility, remains a major challenge.

Information extraction goes hand in hand with automatic thesaurus construction. In information extraction, the problem of mining text for structure is cast in terms of extracting sets of facts, for example, a specific statistic in a crime report, and/or rules, for example, how to find a victim's name and age in crime reports, from the texts at hand. In particular, many researchers are concerned with the problem of extracting database-like structures from Web pages, in effect reverse-engineering the process of database-backed Web page generation.

Hammer et al. (1997) present a configurable tool for extracting semi-structured data from a set of HTML pages, given a declarative specification of where the data of interest are located. The machine learning approach to this problem has been labeled "wrapper induction" (Kushmerick et al., 1997). The extraction procedure, or wrapper, for a specific resource is learned from a set of representative pages from that resource.

Hsu and Chang (1999) describe a formalism to represent information extractors as Finite-State Transducers (FST). A finite-state transducer is a variation of a finite-state automaton (Hopcroft & Ullman, 1979). The input document is assumed to be tokenized before it is given to a finite-state transducer. The authors distinguish two types of transducers: single-pass and multi-pass. A single-pass transducer scans the text only once. A multi-pass transducer scans the text multiple times, each time focusing only on a specific type of object to extract. The ultimate goal of this approach is the automated construction of extractors from a set of training examples. However, the reported empirical evaluations assume that the space of possible graph structures, that is, finite-state automata, is restricted or that the structure is given to the learner in advance.

Freitag (1998) casts information extraction as a relational learning problem. Relational learning represents hypotheses as sets of if-then rules. Because sets of if-then statements can be viewed as programs in a logic programming language, such as PROLOG, relational learning is often called Inductive Logic Programming (Mitchell, 1997). Freitag describes a general purpose top-down relational learning algorithm for information extraction called "SRV". SRV takes as input a set of token-oriented features that encode most of the domain-specific information. For example, they may encode a standard set of questions that can be asked of someone's home page, such as the owner's name, affiliation, e-mail, and so forth. An answer to each question is assumed to be a text fragment from that home page. Thus, the algorithm solves the problem of finding the best unbroken fragment of text that answers a question from a given set of questions. The SRV algorithm makes no assumption about document structure. Instead, structural information is supplied as input to the system.

Jacquemin and Bush (2000) present a tool for the acquisition of named entities, for example, names of companies, from textual sources. The authors' approach combines lexical indices with formatting instructions. Lexical indices are discourse markers and formatting instructions are HTML tags. The system includes three shallow parsers for mining HTML texts for specific structures such as lists, enumerations, and anchors. The named entities are extracted from the found structures by analyzing discourse markers and HTML tags.

FUTURE TRENDS

The issues discussed in this article are likely to remain major challenges in structural text mining. The push to automation will bring an ever greater emphasis on the issue of the evaluation of automatically discovered results to achieve greater recall and precision. Human judgment, arguably the most reliable way of evaluating automatically extracted information, is not always available and, when available, is expensive in terms of time and effort. Thus, one is likely to see more research effort on automating the evaluation of obtained results through discourse analysis and machine readable dictionaries and ontologies.

CONCLUSION

The three fundamental problems in the structural mining of text are text segmentation, automatic ontology construction, and information extraction. Text segmentation investigates methods of partitioning texts into segments of content. Automatic ontology construction focuses on the problem of finding relations among concepts in texts. Information extraction studies ways to extract specific types of information from texts to create database-like records. Each of the three research areas has a number of well established approaches briefly surveyed in the article.

REFERENCES

Appelt, D. (1999). An introduction to information extraction. *Artificial Intelligence Communications*, *12*(3), 161-172.

Bookstein, A., Kulyukin, V., Raita, T., & Nicholson, J. (2003). Adapting measures of clumping strength to assess term-term similarity. *Journal of the American Society for Information Science and Technology (JASIST)*, *54*(7), 611-620.

Cardie, C. (1997). Empirical methods in information extraction. *AI Magazine*, *18*(4), 65-79.

Curran, J. (2002). Ensemble methods in automatic thesaurus extraction. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Philadelphia, PA.

Dagan, I., Lee, L., & Pereira F. C. (1999). Similarity based models of word co-occurrence probabilities. *Machine Learning*, *34*(1-3), 43-69.

Freitag, D. (1998). Information extraction from HTML: Application of a general machine learning approach. *Proceedings of the 15th Conference on Artificial Intelligence (AAAI-98)*. Menlo Park, CA: AAAI Press, pp. 517-523.

Grefenstette, G. (1994). *Explorations in automatic thesaurus discovery*. Boston: Kluwer Academic Publishers.

Hammer, H., Garcia-Molina, J., Cho, R., Aranha, A., & Crespo. V. (1997). Extracting semistructured information from the Web. *Proceedings of the Workshop on Management of Semistructured Data (PODS/SIGMOD'97)*, Tucson, Arizona.

Hearst, M. (1992). Automatic acquisition of hyponyms from large text corpora. *Proceedings of the Fourteenth International Conference on Computational Linguistics (COLING 92),* Nantes, France.

Hearst, M. (1997). TextTiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics*, *23*(1), 33-64.

Hopcroft, J.E., & Ullman, J.D. (1979). *Introduction to automata theory, languages, and computation*. Reading, MA: Addison-Wesley.

Hsu, C.N., & Chang, C.C. (1999). Finite-state transducers for semi-structured text mining. *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI) Workshop on Text Mining*. IJCAI Press, pp. 76-82.

Jacquemin, C., & Bush, C. (2000). Combining lexical and formatting cues for named entity acquisition from the Web. *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*. Hong Kong University of Science and Technology, ACM Press, pp. 189-193.

Kulyukin, V., & Burke, R. (2003). Mining free text for structure. In J. Wang (Ed.), *Data mining: Opportunities and challenges* (pp. 278-300). Hershey, PA: Idea Group Publishing.

Kushmerick, N., Weld, D., & Doorenbos, D. (1997). Wrapper induction for information extraction. *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, Providence, Rhode Island. IJCAI Press, pp. 25-34.

Mitchell, T.M. (1997). *Machine learning*. New York: McGraw-Hill.

Ruge, G. (1997). Automatic detection of thesaurus relations for information retrieval applications. *Foundations of Computer Science: Potential – Theory – Cognition, Lecture Notes in Computer Science, Volume LNCS 1337* (pp. 499-506). Berlin, Germany: Springer Verlag.

Salton, G., & McGill, M. (1983). *Introduction to modern information retrieval*. New York: McGraw-Hill.

KEY TERMS

Automatic Thesaurus Construction: The process of using a computer to automatically extract thesauri from texts.

Information Extraction: The process of using a computer to automatically extract specific facts and rules from texts.

Lexical Indices: Lexical items, for example, proper names and events, used to index texts.

Named Entities: Lexical phrases that fall into one of the three semantic categories: TIMEX, NUMEX, and ENAMEX. TIMEX includes temporal expressions, for example, *March 1st*, *noon EST*. NUMEX consists of numeric expressions, for example, *7.5%*. ENAMEX includes: proper names, for example, *Peter the Great*; locations, for example, *St. Petersburg*; and organizations, e.g., *Russian Orthodox Church*.

Ontology: A set of concepts and a set of relations among those concepts.

Statistical Methods: Methods that rely on collecting numerical data on texts and applying statistical techniques to analyze the data and/or to make statistical inferences about the texts.

Structural Mining of Texts: The process of partitioning texts into components and finding relations among the found components.

Text Segmentation: The process of partitioning a text into segments according to a given criterion.

Structure- and Content-Based Retrieval for XML Documents

Jae-Woo Chang
Chonbuk National University, South Korea

INTRODUCTION

The XML was proposed as a standard markup language to make Web documents in 1996 (Extensible Markup Language, 2000). It has as good an expressive power as SGML and is easy to use like HTML. Recently, it has been common for users to acquire through the Web a variety of multimedia documents written by XML. Meanwhile, because the number of XML documents is dramatically increasing, it is difficult to reach a specific XML document required by users. Moreover, an XML document not only has a logical and hierarchical structure in common, but also contains its multimedia data, such as image and video. Thus, it is necessary to retrieve XML documents based on both document structure and image content. For supporting the structure-based retrieval, it is necessary to design four efficient index structures, that is, keyword, structure, element, and attribute index, by indexing XML documents using a basic element unit. For supporting the content-based retrieval, it is necessary to design a high-dimensional index structure so as to store and retrieve both color and shape feature vectors efficiently.

BACKGROUND

Because an element is a basic unit that constitutes a structured (i.e., SGML or XML) document, it is essential to support not only retrieval based on element units but also retrieval based on logical inclusion relationships among elements. First, RMIT in Australia proposed a *subtree model* that indexes all the elements in a document and stores all the terms that appear in the elements (Lowe, Zobel & Sacks-Davis, 1995) so as to support five query types for structure-based retrieval in SGML documents. Secondly, SERI in South Korea proposed a *K-ary Complete Tree Structure,* which represents a SGML document as a K-ary complete tree (Han, Son, Chang & Zhoo, 1999). In this method, a relationship between two elements can be acquired by calculation because each element corresponds to a node in a K-ary tree. Thirdly, University of Wisconsin in Madison proposed a new technique to use the position and depth of a tree node for indexing each occurrence of XML elements (Zhang, Naughton, DeWitt, Luo & Lohman, 2001). For this, the inverted index was used to enable ancestor queries to be answered in constant time. Fourthly, IBM T.J. Watson research center in Hawthorne proposed ViST, a novel index structure for searching XML documents (Wang, Park, Fan & Yu, 2003). The ViST made use of tree structures as the basic unit of query to avoid expensive join operations and provided a unified index on both text content and structure of XML documents. However, these four indexing techniques were supposed to handle tree data. Finally, University of Singapore proposed D(k)-Index, a structural summary for general graph structured documents (Chen, Lim & Ong, 2003). The D(k) index possesses the adaptive ability to adjust its structure according to the current query load, thus facilitating efficient update algorithms.

There have been a lot of studies on content-based retrieval techniques for multimedia or XML documents. First, the *QBIC (Query By Image Content) project* of IBM Almaden research center studied content-based image retrieval on a large online multimedia database (Flickner et al., 1995). The study supported various query types based on the visual image features such as color, texture, and shape. Secondly, the VisualSEEk project of Colombia University in the USA developed a system for content-based retrieval and browsing (Smith & Chang, 1996). Its purpose was an implementation of CBVQ (Content-Based Visual Query) that combines spatial locations of image objects and their colors. Thirdly, the Pennsylvania State Univesity presented a comprehensive survey on the use of pattern recognition methods for content-based retrieval on image and video information (Antani, Kasturi & Jain, 2002). Fourthly, the Chonbuk National University in South Korea developed an XML document retrieval system that can support a unified retrieval based on both image content and document structure (Chang, 2002). Finally, the Chinese University of Hong Kong presented a multi-lingual digital video content management system, called iVIEW, for intelligent searching and access of English and Chinese video contents (Lyu, Yau & Sze, 2002). The iVIEW system allows full content indexing and retrieval of multi-lingual text, audio and video materials in XML documents.

STRUCTURE- AND CONTENT-BASED RETRIEVAL

A structure- and content-based XML document retrieval system consists of five main parts: a preprocessing part for parsing XML documents and doing image segmentation, an indexing part for generating index keys of XML documents, a storage manager part for storing index information into a specific database, a unified retrieval part for finding results and integrating them into an unified one, and a user interface part for answering user queries by using a Web browser.

When XML documents are given, they are parsed, and image segmentation is done through the preprocessing part. The parsed document information is transported into the structure-based indexer in order to index its document structure consisting of element units. By constructing the index, it is possible to support queries based on a document structure as well as a logical inclusion between elements. In addition, the parsed image information is transported into the content-based indexer in order to get the index information of its color and its shape. To obtain an image feature vector for shape, it is possible to use the image object produced by the preprocessing part and generate a high-dimensional feature vector based on distances between the center point and a set of edge points. To generate a color feature vector, it is necessary to generate a color histogram and normalize the color histogram by dividing it by the number of pixels.

The structure-based and content-based index information is separately stored into their index structures, respectively. The index structures for structure-based retrieval are constructed by indexing XML documents based on an element unit and consist of keyword, structure, element, and attribute index structures. The index structure for content-based retrieval is a high-dimensional index structure, like X-tree (Berchtold, Keim & Kriegel, 1996) or CBF (Han & Chang, 2000), so as to store and retrieve both color and shape feature vectors efficiently.

Using the stored index information extracted from a set of XML documents, some documents are retrieved by the retrieval part in order to obtain a unified result to answer user queries. There is little research on retrieval models for integrating structure- and content-based information retrieval. To answer a document structure query, a similarity between an element q and an element t is computed as the similarity between the term vector of node q and that of node t by using a cosine measure (Salton & McGrill, 1983). Also, a similarity between an element q and a document D is computed as Max{COSINE($NODE_q$, $NODE_{Di}$), $0 <= i < n$}. To answer an image content query, it is possible to compute a similarity between the query feature vector and the image feature vector as 1 – (Euclidean distance / maximum distance) and retrieve relevant documents with high similarity in decreasing order of the similarity. In the case of content-based retrieval based on both color and shape feature vectors, a similarity measure is required. A similarity between a query image q and a target image t in the database is calculated as (1-Distc(q,t)/Nc)*(1-Dists(q,t)/Ns) where Distc (Dists) means a color (shape) vector distance between a query image and a target image, and Nc (Ns) means the maximum color (shape) distances for normalization.

Finally, a final document set that is acquired by integrating preliminary results from both structure- and content-based retrieval is given to users through a convenient user interface, such as a Web browser. To design an efficient structure- and content-based query interface, it is necessary to classify XML queries into two types: simple and composite. The simple query can be divided into keyword, structure, attribute, and image query. The composite query is the composition of simple queries, like structure plus keyword, structure plus attribute, image plus keyword, and image plus structure. It is shown from some experiments that it takes much more time to answer the structure query, compared to the other types of simple queries (Chang, 2002).

FUTURE TRENDS

Directions for future work can be studies on new information retrieval models for integrating preliminary results acquired from both structure- and content-based retrieval, because the two types of retrieval are very different each other, in terms of their nature and property. This can be achieved ultimately by trying to handle MPEG-7 compliant XML documents (Haoran, Rajan & Liang-Tien, 2003; Westermann & Klas, 2003).

CONCLUSION

For effective Web document retrieval, it is very important to retrieve XML documents based on both document structure and image content. To support structure-based retrieval, it is necessary to index XML documents based on the basic element unit, thus generating four index structures: keyword, structure, element, and attribute. To support image content-based retrieval, it is necessary to construct a high-dimensional index structure like CBF (Han & Chang, 2000) for retrieving both color and shape feature vectors efficiently.

REFERENCES

Antani, S., Kasturi, R., & Jain, R. (2002). A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video. *Pattern Recognition, 35*(4), 945-965.

Berchtold, S., Keim, D., & Kriegel, H.-P. (1996). The X-tree: An index structure for high-dimensional data. *Proceedings of Int'l Conf. on Very Large Databases* (pp. 28-39).

Chang, J.-W. (2002). An XML document retrieval system supporting structure- and content-based queries. *Lecture Notes in Computer Science, 2465,* 320-333.

Chen, Q., Lim, A., & Ong, K.W. (2003). D(k)-Index: An adaptive structural summary for graph-structured data. *Proceedings of ACM Int'l Conf. on Management of Data* (pp. 134-144).

Extensible Markup Language (XML). (2000). *http://www.w3.org/XML/*

Flickner, M. et al. (1995). Query by image and video content: The QBIC system. *IEEE Computer, 28*(9), 23-32.

Han, S., & Chang, J.-W. (2000). A new high-dimensional index structure using a cell-based filtering technique. *Lecture Notes in Computer Science, 1884,* 79-92.

Han, S., Son, J., Chang, J.-W., & Zhoo, J. (1999). Design and implementation of a structured information retrieval system for SGML documents. *Proceedings of Int'l Conf. on Database Systems for Advanced Applications* (pp. 81-88).

Haoran, Y., Rajan, D., & Liang-Tien, C. (2003). Automatic generation of MPEG-7 compliant XML document for motion trajectory descriptor in sports video. *Proceedings of ACM Int'l Workshop on Multimedia Databases* (pp. 10-17).

Lowe, B., Zobel, J., & Sacks-Davis, R. (1995). A formal model for databases of structured text. *Proceedings of Int'l Conf. on Database Systems for Advanced Applications* (pp. 449-456).

Lyu, M.R., Yau, E., & Sze, S. (2002). A multilingual, multimodal digital video library system. *Proceedings of ACM/IEEE-CS Joint Conf. on Digital Libraries* (pp. 145-153).

Salton, G., & McGill, M. (1983). *An introduction to modern information retrieval*. McGraw-Hill.

Smith, J.R., & Chang, S.F. (1996). VisualSEEk: A fully automated content-based image query system. *Proceedings of ACM Int'l Conf. on Multimedia* (pp. 87-98).

Wang, H., Park, S., Fan, W., & Yu, P.S. (2003). ViST: A dynamic index method for querying XML data by tree structures. *Proceedings of ACM Int'l Conf. on Management of Data* (pp. 110-121).

Westermann, U., & Klas, W. (2003). An analysis of XML database solutions for the management of MPEG-7 media descriptions. *ACM Computing Surveys, 35*(4), 331-373.

Zhang, C., Naughton, J., DeWitt, D., Luo, Q., & Lohman, G. (2001). On supporting containment queries in relational database management systems. *Proceedings of ACM Int'l Conf. on Management of Data* (pp. 425-436).

KEY TERMS

Content-Based Retrieval: Retrieval based on image content. This includes retrieval based on image color, texture, shape and position of salient objects, dominant edges of image items and regions.

Cosine Measure: The vector angle between two documents that is used as a measure of similarity.

Euclidean Distance: Ordinary straight-line distance.

High-Dimensional Index Structure: The organization of indexes for storing the feature vectors of images where the number of their dimensions is large, at least over 10.

HTML: Hypertext Markup Language. A hypertext document format used on the World Wide Web. HTML is an application of SGML. Tags embedded into the text describe certain properties of the text elements and can include other elements such as links to HTML pages or to other resources such as images. HTML is a recommendation of the W3C.

SGML: Standard Generalized Markup Language. A general language for representing documents. SGML is defined in ISO 8879:1986.

Structure-Based Retrieval: Retrieval based on document structure. This includes retrieval based on element units as well as logical inclusion relationships among elements because an element is a basic unit of XML documents.

XML: Extendible Markup Language. The initial proposal for XML was based on an already existing standard called SGML. The XML recommendation was released in 1998 by the World Wide Web Consortium (W3C).

S

Students' Perceptions of Online Courses

Judith C. Simon
The University of Memphis, USA

Lloyd D. Brooks
The University of Memphis, USA

Ronald B. Wilkes
The University of Memphis, USA

INTRODUCTION

An increasing number of traditional colleges and universities, responding to marketplace pressures, are offering online courses and degree programs. According to Weil (2001), 54% of U.S. higher education institutions offer e-learning courses. Many AACSB-accredited business schools provide courses and complete degree programs online. New schools have been created that exist solely in cyberspace (Peltz, 2000). Students can complete undergraduate online degree programs in fields as diverse as nursing, business, engineering, and technology.

Possible reasons for schools offering online courses and degree programs include: increasing revenues and credit hour production; better utilization of limited resources; serving an expanded geographic area or student population; and serving the existing student population differently. Students may take online courses due to their perceptions of location convenience, time convenience, cost, and quality.

Some online courses have been implemented so quickly that insufficient time has been available to allow in-depth assessment of the desires, interests, and concerns of their potential direct customers, that is, students. The study described here was developed to identify students' expectations and current perceptions of online courses and degree programs. The results are expected to facilitate effective planning and development of these courses and programs in the future. The intent is to study the next generation of online learners. This group's perceptions of online coursework are important for course providers as they consider ways to maintain and enlarge their programs, perhaps expanding to meet the needs of younger, more traditional students.

BACKGROUND

Online courses are a form of distance education. Today's widespread availability of the Internet has made online courses and degree programs available to students around the globe. Many traditional colleges and universities have decided to add online coursework so that they can be competitive with the large number of programs now available from private enterprises set up for this purpose. Several issues have emerged related to this instructional delivery format, and studies are now being reported related to advantages and disadvantages of this type of education.

As with other forms of distance education, convenience has been widely quoted as a primary reason for students taking online courses. Another reason is the self-paced nature of many online courses, which allows for a variety of student schedules and comprehension rates. Availability of coursework regardless of the student's physical location is another reason it is used; it would be possible for students in remote locations to take these courses as long as they had the necessary technology. It is also possible that online courses provide more up-to-date materials than that found in some traditional courses and textbooks, due to the ease of making changes to Web sites rather quickly (Accetta & Wind, 2001). Another factor that may make online courses an appealing option is that a person's age or physical limitations may be less likely to prevent the person from succeeding in this format (Brown, 2000).

Faculty who teach online courses may find it rewarding if they enjoy using new technology and also enjoy the challenge of adapting to a different form of course delivery.

Critics of online courses point out that a Web-based educational program has at least one disadvantage in that it does not provide a forum for physical contact and live debate between students and faculty. Representatives of some elite universities claim that it is impossible to replicate a classroom environment with online courses. Harvard University's Professor W. Earl Sasser indicated that Harvard does not plan to offer an MBA degree online because it would distract from the residential experience (Symonds, 2001). Kumar, Kumar and Basu (2002) cite "strong evidence that students perceive interaction, student-to-student and student-to-instructor, to suffer as a result of virtual education" (p. 140). But some information technology professionals argue that there is little difference between getting a degree on-campus or over the Web. Robert Baker, a systems consultant, observes that students taking online courses are not isolated and can get to know faculty and other students by using online discussion boards (Dash, 2000). Schooley (2001) states that online university courses facilitate communication with the instructor and among classmates.

Another concern is that online coursework may worsen the so-called "digital divide" between students in higher-income versus lower-income families, as those with higher incomes will have more Internet access (Accetta & Wind, 2001). However, libraries and schools have increased Internet access availability to people at all income levels and at locations worldwide.

Some questions also exist about the cost of taking online courses. Some students believe that an online course is a less expensive method, and some course providers believe it to be a cost-effective method of course delivery. As a point of comparison, Duke University's Fuqua School of Business offers MBA degrees that provide about 65% of the work online and about 35% of the work in residency. Duke charges up to $90,000 for the program versus $60,000 for its traditional on-campus MBA degree program (Symonds, 2001).

One critical indicator of program success or failure is the extent to which graduates are accepted in the job market. Quigley (2001) cites a *BusinessWeek* magazine survey reporting that most recruiters are skeptical of skills of graduates of online business schools. Many executives do not feel that online degree programs have been offered long enough to prove themselves with on-the-job performance of MBA degree graduates.

Some educators are concerned about the rapid growth of online degree programs. Others have questioned the superiority of online courses when compared with classroom-based courses. Online courses may be developed by curriculum specialists with little or no participation from the faculty member who will conduct the class. Much e-learning is self-directed and not led directly by faculty members. A fear is that faculty members will no longer be curriculum developers and participants in intellectual debates within their disciplines. Accetta and Wind (2001) suggest that "faculty will become mere shepherds herding their passive sheep through pre-prepared fields of outdated and insubstantial information."

Faculty members are also concerned that students will not get the same campus experience that is provided to students who are enrolled in traditional on-campus degree programs. Some schools require that students check in online for a specified number of times each week. Many programs also require some amount of on-campus time, as well as meetings via conference calls. Some programs incorporate team projects that require students being together during specified points during the course. Some programs require one or more retreats in a traditional lecture/seminar format.

Many faculty members are skeptical about the quality when the time period required to obtain a degree is very short. For example, the American InterContinental University-On-line (2002) advertises that a Master's degree can be earned in as little as 10 months.

There are also concerns about technical, administrative, and pedagogical issues that arise as faculty consider moving from a traditional classroom environment into a Web-based environment. Logistical concerns relate to providing students with the same level of support (e.g., library, bookstore, advising) in both environments. Pedagogical considerations involve issues relating to management of course quality and control over the learning environment.

Wonacott (2000) states that instructional design should be the primary factor in implementing online coursework rather than the appeal of technology. He also mentions that students and instructors need appropriate training and guidance if this type of instruction is to be effective.

STUDIES OF EXISTING ONLINE COURSES AND PROGRAMS

Many reports have indicated typical students are "adult learners". Typical participants are between 25 and 50 years of age and are taking online courses either to learn something new or to update their skills (Grill, 1999).

Studies have also indicated that students were more satisfied with online courses when there was more interaction with the course instructor. Interaction with other students was also a factor leading to more student satisfaction (Fredericksen, Pickett, Shea, Pelz & Swan , 2000).

Several comparison studies have been conducted between online and traditional on-campus courses, and most have found no statistically significant difference in test performance and grades. A greater degree of active learning in both settings increases performance and student attitude (Hall, 1999).

Studies of online coursework are being conducted in several countries. For example, one study compared "Anglo-Saxon" students in five countries to "Asian" students in nine countries to see if cultural differences existed related to the use of the Web as a learning environment. Some differences existed in the comfort level of various instructional formats. More Asian students liked the Web environment because of the innovative learning possibilities, while Anglo-Saxon students were more comfortable with the use of discussion groups (Chin, Chang & Bauer, 2000).

POTENTIAL FOR ONLINE COURSEWORK FOR FUTURE STUDENTS

Schools providing online courses and programs will eventually need to expand the attractiveness of their programs to additional markets in order for their programs to remain viable. Since the current programs are predominantly used by students who are older than the traditional full-time college student, an initial step would be to determine the perceptions of the next group of college students regarding these programs. To obtain data related to perceptions of online courses by upcoming college students, a survey instrument was distributed to almost 400 students enrolled in high school business courses. Although participants in existing online courses and programs are older, educational institutions should understand the needs and perceptions of other constituencies if they want to expand their enrollments. These current high school students could represent a large potential target market in the future that will be increasingly computer literate, which would enable more students to take advantage of online coursework.

Demographics

A total of 381 high school students participated in this study. Slightly more than 71% of these students were juniors and seniors. A majority of the participants (58.78%) were female.

A high percentage of students (88.97%) indicated that they have access to a computer at home, which could have a direct effect on ability to participate in online courses. Lack of access to a computer could affect perceptions. A comparison of accessibility by gender showed that the difference between genders was slight (Table 1).

Current Status Regarding Taking Online Courses

Students were asked to describe their current status regarding taking a course online and were allowed to select more than one response. As shown in Table 2, about one-fourth of the respondents indicated that they would not take an online course. There was little difference between males and females on this issue. Not surprisingly, very few had already taken or were currently taking a course online.

Ratings of Issues' Importance

The remainder of the survey had two parts, each with a listing of issues for students to consider. Thirty-eight students failed to respond to a majority of the questions in this part of the survey and were eliminated from further calculations.

In the first section, students were asked to indicate how important each identified issue was to them in deciding whether to take a course online or on-campus. A Likert-type scale was used, with 1 representing "not at all important" and 5 representing "extremely important". A mean was calculated as a basis for determining which issues were considered rather important (defined as a mean of at least 4.0)

Ratings That a Characteristic is More Likely True for Online versus On-Campus

For the second section, students were asked to consider the same issues as in the previous section but to indicate the likelihood that each issue was a characteristic of an online versus on-campus course, with 1 representing "much more likely in an online course" and 5 representing "much more likely in an on-campus course". A mean was calculated to identify which issues were considered much more likely in an online course (defined as a mean of no greater than 2.0) and which were considered much more likely in an on-campus course (defined as a mean of at least 4.0). As shown in Table 4, three issues had a mean greater than 4. Only one issue, "Submitting assignments electronically," had a mean below 3.0.

A cluster plot or "scatter diagram" was developed to illustrate the relationship between the two parallel sets of variables used in the study. As shown in Figure 1, all but one of the items were clustered in the upper right corner, which represents the section for higher importance and higher likelihood that the characteristic would be in an on-

Table 1. Access to a computer at home by gender

Gender	Percentage with Access
Male	89.03
Female	88.50

Table 2. Current status regarding online courses

Status Regarding Online Courses	Percentage of All Respondents	Percentage of Male Respondents	Percentage of Female Respondents
I would not take a course online.	24.87	24.52	24.78
I would consider taking a course online.	58.20	56.77	58.41
I would like to take a course online.	28.04	26.45	28.76
I plan to take a course online.	6.35	6.45	6.19
I am currently taking a course online.	0.26	0.65	0.00
I have completed a course online.	0.49	1.94	0.00

campus course. As discussed earlier, the only item in the left side of the diagram was "submitting assignments electronically."

The issues identified as most important were given further review, since those issues were not given particularly strong ratings as either online or on-campus options. Issues identified as most important in Table 3 are shown again in Table 5, along with an identification of their means for the comparison of "much more likely in an online course" versus on-campus. Table 5 indicates a tendency toward likelihood in an on-campus course for all of the issues identified as most important. All had means greater than 3.0, which is the midpoint of the scale where 1 represented much greater likelihood in an online course and 5 represented much greater likelihood in an on-campus course.

FUTURE TRENDS

Large increases are occurring in the availability of online courses and college degree programs, both at the undergraduate and graduate levels. Many traditional universities are adding online coursework to their offerings in an effort to maintain student enrollments. Consequently, these institutions are concerned about the best methods for providing equivalent course content online.

Significant growth in online programs could lead to an eventual reduction in on-campus course offerings at universities.

Students are becoming increasingly competent in the use of computer technology. Future students, therefore, can be expected to be better prepared and possibly better candidates for online programs.

CONCLUSION

Students who participated in this study indicated that all the specified issues were perceived to be important (a mean greater than 3.0 using a 1 to 5 scale). All but one of the issues were perceived to be more likely to be characteristic of an on-campus course. Educational institutions need to begin considering the needs of this next generation of potential students in order to expand or maintain online course offerings in the future.

The fact that students in this study rated "knowledge gained" as their most important issue in a course or degree and also rated it more likely to be characteristic of an on-campus than an online course indicates that institutions wanting to increase the number of online students might need to look for documentation of high levels of knowledge gained through online course delivery. In addition,

Table 3. Issues considered important in making course environment decisions

Issue	Mean
Knowledge gained	4.35
Skills acquired	4.29
Access to information (resource materials)	4.20
Time required to complete coursework	4.16
Costs of tuition and fees	4.13
Schedule flexibility to accommodate work responsibilities	4.13

S

Table 4. Issues that are much more characteristic of an on-campus course

Issue	Mean
Opportunity for live interaction/discussion among students	4.07
On-campus exams	4.05
Opportunity for live interaction/discussion between faculty and students	4.04

Figure 1. Comparison of importance with likelihood of being on-campus versus online

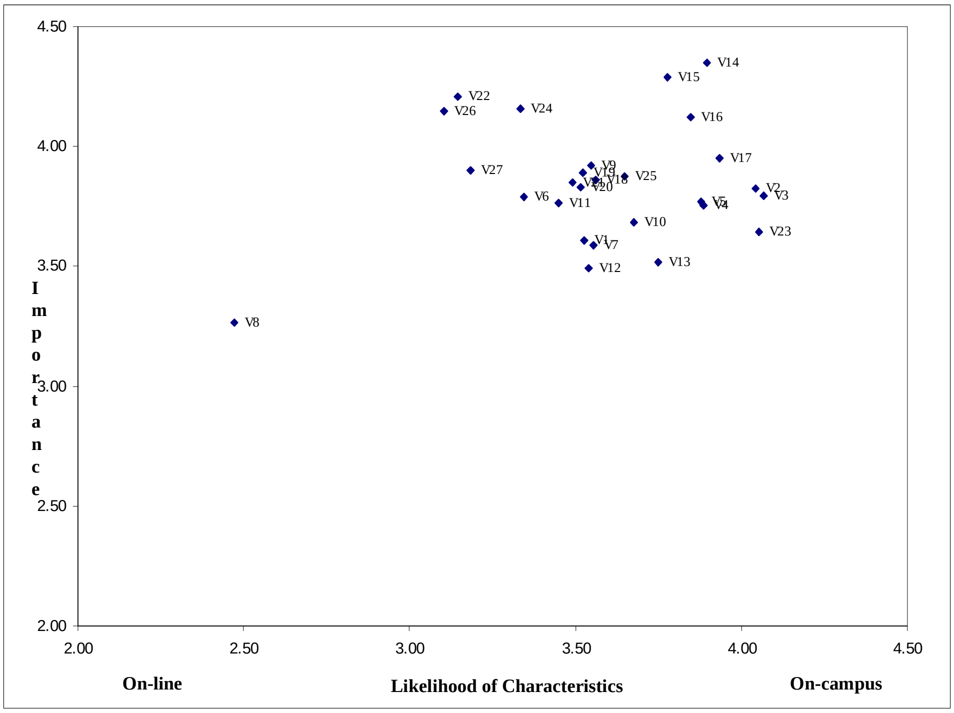


Table 5. Important issues related to online versus on-campus instruction

Issue	Mean
Knowledge gained	3.90
Skills acquired	3.78
Access to information (resource materials)	3.15
Time required to complete coursework	3.33
Costs of tuition and fees	3.85
Schedule flexibility to accommodate work responsibilities	3.10

(Much more likely in an online course = 1; much more likely in an on-campus course = 5)

studies should be done to determine the specific features and programs that would be the best fit for these students in an online learning environment.

Students' perceptions of online programs should continue to be monitored by institutions wanting to increase their participation in these programs.

REFERENCES

Accetta, R., & Wind, Y. (2001, February 26). E-learning crossfire. *INFORMATIONWEEK.com*. Retrieved June 25, 2002, from *http://www.informationweek.com/826/elearning_6side.htm*

American InterContinental University-On-line. (2002). Why AIU on-line? Retrieved July 12, 2002, from *http://aiudegreeonline.com/why.jsp*

Brown, B. (2000, October). Web-based training. *ERIC Digests*. Retrieved February 10, 2004, from *http://234www.ericfacility.net/ericdigests/ed445234.html*

Chin, K.L., Chang, V., & Bauer, C. (2000). The use of Web-based learning in culturally diverse learning environments. *AusWeb2K*. Retrieved February 9, 2004, from *http://ausweb.scu.edu.au/aw2k/papers/chin/paper.html*

Dash, J. (2000). IT pros give on-line universities high marks. *CNN.com*. Retrieved June 21, 2002, from *http://www.cnn.com/2000/TECH/computing/09/15/online.college.idg/*

Fredericksen, E., Pickett, A., Shea, P., Pelz, W., & Swan, K. (2000). Student satisfaction and perceived learning with on-line courses: Principles and examples from the SUNY Learning Network. *Journal of Asynchronous Learning Networks, 4*(2). Retrieved February 9, 2004, from *http://www.aln.org/publications/jaln/v4n2/v4n2_fredericksen.asp*

Grill, J. (1999, Summer). Access to learning: Rethinking the promise of distance education. *Adult Learning, 10*, 32-33.

Hall, R. (1999, April 19-23). *Web-based conferencing as a component of a collaborative-learning based educational psychology class*. Paper presented at the Annual Meeting of the American Educational Research Association (AERA), Montreal, Quebec.

Kumar, A., Kumar, P., & Basu, S.C. (2002). Student perceptions of virtual education; An exploratory study. In M. Khosrow-Pour (Ed.), *Web-based instructional learning* (pp. 132-141). Hershey, PA: IRM Press.

Peltz, P. (2000, November 18). Do virtual classrooms make the grade? *CNN.com*. Retrieved July 12, 2002, from *http://www.cnn.com/2000/TECH/computing/11/18/index.elearning/cover.elearning/index.html*

Quigley, A. (2001, May 21). Six degrees of separation. *eLearn Magazine*. Retrieved May 30, 2002, from *http://www.elearnmag.org/subpage/s*

Schneider, M. (2001, April 5). GM gives log-on learning a boost. *BusinessWeek Online*. Retrieved June 25, 2002, from *http://www.businessweek.com/bwdaily/dnflash/apr2001/nf2001045_517.htm*

Schooley, C. (2001). Online universities introduce alternatives for higher education. *Planning Assumption*. GIGA Information Group.

Symonds, W. C. (2001, December 3). Giving it the old online try. *BusinessWeek Online*. Retrieved May 29, 2002, from *http://www.businessweek.com:/print/magazine/content/01_49/b3760072.htm?mainwindow*

Weil, N. (2001, June 26). University net courses help pros make the grade. *CNN.com*. Retrieved July 3, 2002, from *http://www.cnn.com/2001/TECH/internet/06/26/university.net.courses.idg/index.html*

Wonacott, M.E. (2000). Web-based training and constructivism in brief: Fast facts for policy and practice No. 2. National Dissemination Center for Career and Technical Education. Retrieved February 18, 2004, from *http://www.nccte.org/publications/infosynthesis/in-brief/in-brief02/index.asp*

KEY TERMS

Distance Education: A form of instruction in which a geographical separation exists between instructor and students; it may be same time/different place or different time/different place. Various types of technology may be used as part of this form of education, with more technology required for the same-time format.

Distance Learning: The intended outcome of distance education, characterized by learning through audio-visual delivery of instruction that is transmitted to one or more other locations that may be live or recorded instruction but does not require physical presence of an instructor and students in the same location. Current technology used for distance learning includes text, images, and audio and/or video signals.

E-Learning (Electronic Learning): A form of learning that involves "electronic" or technology-based delivery of learning; examples of forms of delivery include individual computer-based delivery, Internet/Web delivery, and virtual classroom delivery. Media can be in many forms, including videotape and DVD, CD-ROM, satellite transmissions, interactive television, and various Web-based media (Internet, intranet, extranet).

Off-Campus Course: A course offered at a business location, satellite center, or other site other than the main campus of an educational institution.

On-Campus Course: A course offered wholly or in part on the main campus of an educational institution. An on-campus course is typically offered in a traditional lecture/discussion environment, but may include procedures and methodology associated with a hybrid online course.

On-Campus Degree Program: A prescribed program of studies leading to a degree conferred by an educational institution wherein the majority of courses are on-campus courses. (See definition of *on-campus course*.)

Online Course (Fully and Hybrid): An online course may be either a fully online course or a hybrid online course. A fully online course is offered in a format where its content, orientation, and assessments are delivered via the Internet and students interact with faculty and possibly with one another while using the Web, e-mail, discussion boards, and other similar aids. A hybrid online course meets one or more of the following requirements for students: (a) access to the Web to interact with the course content, (b) access to the Web to communicate with faculty and/or other students in the course, (c) on campus for orientation and exams, with other aspects of the course meeting the requirements for a fully online course.

Online Degree Program: A prescribed program of studies leading to a degree conferred by an educational institution wherein the majority of courses are online courses. (See definition for *online course*.)

Success Surrogates in Representational Decision Support Systems

Roger McHaney
Kansas State University, USA

INTRODUCTION

Rapid and frequent organizational change has been a hallmark of business environments in the past two decades. Frequently, technology and new software development are embraced as aspects of complex strategies and tactical plans. Without sufficient analysis, the unforeseen consequences of change can result in unexpected disruptions and the loss of productivity. In order to better control these contingencies, modern managers often employ a variety of decision support aids. One such aid, classified as a representational decision support system, is discrete event simulation (DES).

BACKGROUND

In its purest form, DES is considered to be a branch of applied mathematics. Its considerable popularity is due in part to the availability of computers and improvements in simulation languages and simulator software packages. DES is often the technique of choice when standard analytical or mathematical methodologies become too difficult to develop. Using a computer to imitate the operations of a real-world process requires a set of assumptions taking the form of logical relationships that are shaped into a model. These models assess the impact of randomly occurring events. Experimental designs are developed and the model manipulated to enable the analyst to understand the dynamics of the system. The model is evaluated numerically over a period of time and output data is gathered to estimate the true characteristics of the actual system. The collected data is interpreted with statistics allowing formulation of inferences as to the true characteristics of the system. Table 1 lists primary features of a DES application.

While the value of DES in organizational settings has been accepted and is evidenced by the varied and growing market of related products, not every DES application is suited for every problem domain. For this reason, information systems researchers have worked to identify salient characteristics of DES and its usage and then measure the relationship between its application and successful organizational outcomes. These studies have been conducted in different ways with focuses on independent and dependent variables.

INDEPENDENT VARIABLE RESEARCH

Much DES research has focused on identifying and evaluating software and project characteristics (independent variables) and then producing recommendations that

Table 1. DES features

Feature	Description
Statistics Collection	Tools which gather data for purposes of inferential statistics about the model
Resource Modeling	A means for the representation of a constrained resource in the model
Transaction	A means for representation of the participants in the simulatior model
Simulation Clock	Tools for analysis and step processing of the coded model
Random Number Generators	A means for producing random number streams for randomization of events within the simulation model
Model Frameworks	Generalized frameworks for the rapid development of a model

S

Table 2. List of DES success factors

Factor 1: Software Characteristics
Factor 2: Operational Cost Characteristics
Factor 3: Software Environment Characteristics
Factor 4: Simulation Software Output Characteristics
Factor 5: Organizational Support Characteristics
Factor 6: Initial Investment Cost Characteristics
Factor 7: Task Characteristics

either suggest how a project can be successfully implemented or how failure can be avoided. A variety of useful practitioner-focused articles have been published in this area (Banks, 1991; Law & Haider, 1989; Swain, 2003). These studies provide recommendations, based in part on first-hand experience, of consultants and simulation analysts. In many cases, these recommendations list the packages currently available and focus on the pros and cons of each.

Academic studies of independent variables have also been conducted. In one of the first studies using information systems as basis for DES, McHaney and Cronan (2000) extended a framework of general decision support system (DSS) success factors identified by Guimaraes, Igbaria, and Lu (1992). The developed contingency model was theoretically derived from the simulation literature and empirically tested. The results indicated a seven-factor model that was structured as shown in Table 2.

Academic research by Robinson (1999) approached the problem from the opposite perspective and identified sources of simulation inaccuracy which may result in project failure. Other studies, such as one by McHaney and White (1998), developed a DES software selection framework which matched salient DES characteristics with software package features. The importance of evaluation and selection of appropriate DES packages was determined in relation to the success of simulation implementation. As might be expected, the choice of the wrong simulation package often correlated with simulation system failure. This study provided a set of criteria to be systematically considered when evaluating DES software. The taxonomy for simulation evaluation, together with importance ratings provided by the collective expertise of a large number of DES users, was used as a guideline in deciding the relative weighting to give various software package capabilities.

Other research (McHaney, White, & Heilman, 2002) attempted to develop an understanding of DES project success by determining which characteristics of a simulation project were more likely to be present in a successful simulation effort. Potential success factors were derived from the simulation literature and used to develop a questionnaire. Based on the findings, projects perceived as failing were often characterized by high costs, model size constraints, and slow software. Successful projects were characterized by teamwork, cooperation, mentoring, effective communication of outputs, high quality vendor documentation, easily understood software syntax, higher levels of analyst experience, and structured approaches to model development. By understanding the simulation process and characteristics of successful simulations, practitioners will find it easier to avoid common mistakes that can ruin a modeling effort.

DEPENDENT VARIABLE RESEARCH

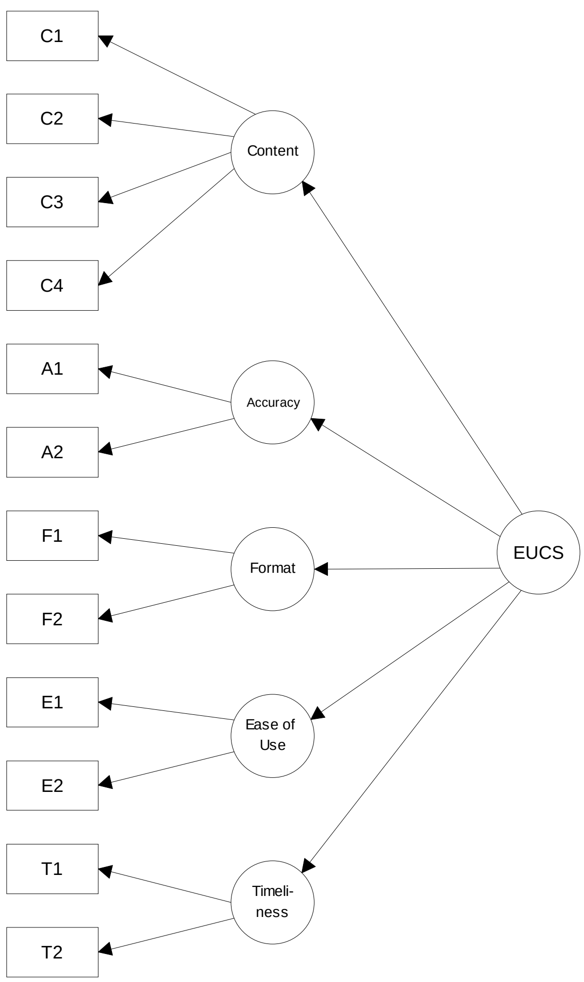
The studies mentioned in the previous section focused on understanding independent variables defining successful representational decision support system applications. A problem with research that considers only the independent side of the equation is the lack of objective measures of whether suggested recommendations correlate with desired outcomes. This dilemma has been problematic throughout information systems research in general and has been the subject of academic debate (Delone & McLean, 1992). DES researchers recognize that the identification of a meaningful, reliable, and robust dependent variable is central to being able to conduct accurate comparisons between competing tools, techniques, software implementations, project approaches, and modeling perspectives.

A wide variety of dependent variables have been investigated in the broader fields of information systems (Delone & McLean, 1992) and decision support systems. Among these, DSS researchers have focused on success surrogates and investigated the possibility of assessing DSS success or failure. Representational DSS researchers have extended the use of these surrogates.

The first information system success surrogate validated in the context of DES was the end-user computing satisfaction instrument (Doll & Torkzadeh, 1988). The EUCS instrument, shown in Figure 1, is of particular interest because most applications of discrete event computer simulation can be categorized as end-user computing. McHaney and Cronan (1998) collected data from 411 participants using a variety of discrete event computer simulation software packages. The analysis indicated the instrument retained its psychometric properties and provided a valid success surrogate for end users beyond the introductory stages of using representational DSSs. The study established the use of information systems surrogate success measures for applied instruments. The results suggest EUCS can reliably and confidently be used in the investigation of competing tools, features, and technologies in the area of DES. Later, EUCS was the subject of a test-retest reliability study that was distributed to users of DES through a mail survey. One month later, follow-up surveys were administered. The original respondents were asked to again evaluate the same system. The two data sets were compared and the results supported the instrument's internally consistency and stability (McHaney, Hightower, & White, 1999).

Another success surrogate, the technology acceptance model (TAM), was also investigated within the context of representational decision support systems. TAM was developed by Davis (1989) to provide a theoretical explanation of factors influencing technology usage. Davis' theory is derived from the theory of reasoned action (TRA) model (Fishbein & Ajzen, 1975). The TRA model explains actions by identifying connections between various psychological constructs such as attitudes, beliefs, intentions, and behaviors. TRA posits that an individual's attitude toward a given behavior is determined by the belief that this behavior will result in a particular outcome. TAM expands upon this framework to provide a theoretical explanation of factors influencing technology usage. However, instead of relying on purely attitudinal determinants, TAM hypothesizes that perceived ease of use and perceived usefulness influence a person's intention, which in turn determines actual technology usage. In essence, Davis' research identifies the external variables which influence attitude towards voluntary use of technology. TAM has been validated and tested in a variety of technology-related studies (Adams, Nelson, & Todd, 1992; Davis, Bagozzi, & Warshaw, 1989; Mathieson, 1991). Figure 2 illustrates the Davis instru-

Figure 1. EUCS instrument



C1: Does the simulation system provide the precise information you need?
C2: Does the simulation output information content meet your needs?
C3: Does the simulation provide reports that seem to be just about exactly what you need?
C4: Does the simulation system provide sufficient information?

A1: Is the simulation system accurate?
A2: Are you satisfied with the accuracy of the simulation system?

F1: Do you think simulation output is presented in a useful format?
F2: Is the simulation output information clear?

E1: Is the simulation system user-friendly?
E2: Is the simulation system easy to use?

T1: Do you get the output information you need in time?
T2: Does the simulation system provide up-to-date information?

ment and TAM. McHaney and Cronan (2001) focused on external validity aspects of TAM when applied to users of DES. The findings suggested that TAM provided a reasonable surrogate measure for success within this domain. An updated version of this study was developed into a book chapter (McHaney & Cronan, 2002).

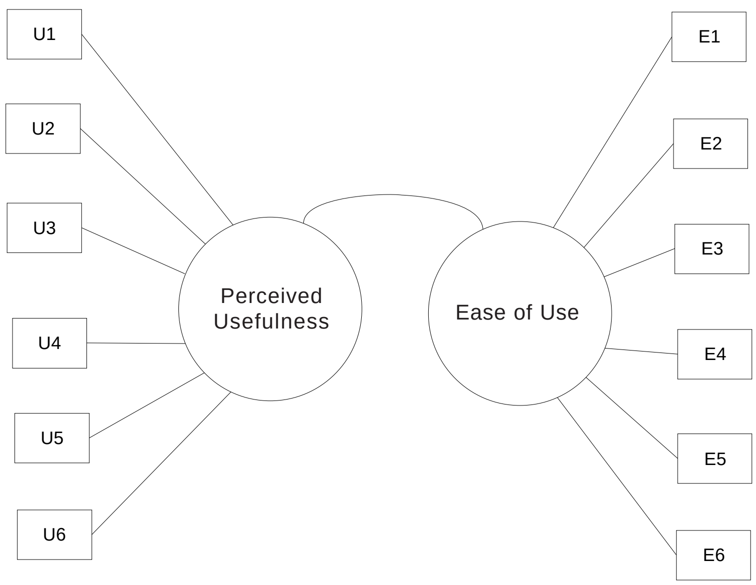
FUTURE TRENDS

Representational decision support systems will continue to become more important to managers, particularly as computing power continues to increase and costs decrease. Web-enabled DES, applications embedded in enterprise computing systems, visualization, and use of DES to fine-tune manufacturing operations and supply chain operations on-the-fly will become more commonplace. Representational decision support system research will continue to seek dependent variables and related surrogates that enable competing tools and techniques to be evaluated and compared. Mainstream DSS and information system research will continue to contribute methodologies and techniques that can be adopted for use in this field.

CONCLUSION

Representational decision support systems, particularly as manifested in DES, have become an important tool in a modern manager's decision-making arsenal. In order to determine the effectiveness of these applications, researchers have investigated characteristics of the technology and determined the suitability of various success surrogates to enable comparisons between competing packages, approaches, and methodologies. Future research in this field may determine whether other surrogates are suitable for use with DES and will provide managers with quantitative comparisons to aid in their software selection processes.

Figure 2: Davis instrument



E1: Learning to operate simulation would be easy for me.
E2: I would find it easy to get simulation to do what I want it to.
E3: My interaction with simulation would be clear and understandable.
E4: I would find simulation to be flexible to interact with.
E5: It would be easy for me to become skillful at using simulation.
E6: I would find simulation easy to use.

U1: Using simulation in my job would enable me to accomplish tasks more quickly.
U2: Using simulation would improve my job performance.
U3: Using simulation in my job would increase my productivity.
U4: Using simulation would enhance my effectiveness on the job.
U5: Using simulation would make it easier to do my job.
U6: I would find simulation useful in my job.

REFERENCES

Adams, D. A., Nelson. R. R., & Todd, P. A.(1992). Perceived usefulness, ease of use, and usage of information technology: A replication. *MIS Quarterly*, *16*(2), 227-247.

Banks, J. (1991). Selecting simulation software. In B.L. Nelson, W.D. Kelton, & G.M. Clark *Proceedings of the Winter Simulation Conference*, 21(5), 15-20.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(4), 319-340.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R.(1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, *35*, 982-1003.

Delone, W. H., & McLean, E. R. (1992). Information success: The quest for the dependent variable. *Information Systems Research*, *3*(1), 60-95.

Doll, W. J., & Torkzadeh, G. (1988). The measurement of end-user computing satisfaction. *MIS Quarterly*, *12*(2), 259-274.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

Guimaraes, T., Igbaria, M., & Lu, M. (1992) The determinants of DSS success: An integrated model. *Decision Sciences*, *23*(2), 409-430.

Law, A. M., & Haider, S. W. (1989, May). Selecting simulation software for manufacturing applications: Practical guidelines and software survey. *Industrial Engineering*, 33-46.

Mathieson, K. (1991). Predicting user intentions: Comparing the technology acceptance model with the theory of planned behavior. *Information Systems Research*, *2*, 173-191.

McHaney, R. W., & Cronan, T. P. (1998). Computer simulation success: On the use of the end-user computing satisfaction instrument. *Decision Sciences*, *29*(2), 525-536.

McHaney, R. W., & Cronan, T. P. (2000). Toward an empirical understanding of computer simulation implementation success. *Information & Management*, *37*, 135-151.

McHaney, R. W., & Cronan, T. P. (2001). A comparison of surrogate success measures in on-going representational decision support systems: An extension to simulation technology. *Journal of End User Computing*, *13*(2), 15-25.

McHaney, R. W., & Cronan, T. P. (2002). Success surrogates in representational decision support systems. In M. A. Mahmood (Ed.), *Advanced topics in end user computing* (pp. 243-262). Hershey, PA: Idea Group Publishing.

McHaney, R. W., & White, D. (1998). Discrete event simulation software selection: An empirical framework. *Simulation & Gaming*, *29*(2), 228-250.

McHaney, R. W., Hightower, R., & White, D. (1999). EUCS test-retest reliability in representational model decision support systems. *Information & Management*, *36*, 109-119.

McHaney, R. W., White, D., & Heilman, G. (2002). Simulation project success and failure: Survey findings. *Simulation & Gaming*, *33*(1), 49-66.

Robinson, S. (1999). Three sources of simulation inaccuracy (and how to overcome them). In P. A. Farrington, H. B. Nembhard, D. T. Sturrock, & G. W. Evans (Eds.), *Proceedings of the 1999 Winter Simulation Conference* (pp. 1701-1708). Piscataway, NJ: IEEE.

Swain, J. J. (2003). Simulation Reloaded: Sixth biennial sur-vey of discrete-event software tools. OR/MS Today 30(4): 46-57, INFORMS, Baltimore, Maryland.

KEY TERMS

Dependent Variable: A value representing the presumed effect or consequence of various states of related independent variables. In other words, a dependent variable is the condition for which an explanation is sought.

Discrete Event Simulation (DES): Use of a computer to mimic the behavior of a complicated system and thereby gain insight into the performance of that system under a variety of circumstances. Generally the system under investigation is viewed in terms of instantaneous changes due to certain sudden events or occurrences.

End-User Computing Satisfaction: A widely accepted information systems success surrogate that measures the degree to which a technology provides the user with a sense of satisfaction that meaningful usage has been affected.

Independent Variable: A value representing a presumed cause of a particular outcome.

Representational Decision Support System: Computer-based information system that combines models with data in a fashion that closely resembles the system that is being studied. Computer simulation is a form of representational decision support system.

Success Surrogate: A proxy for information systems success that takes the form of measurable values. A success surrogate is a dependent variable.

Technology Acceptance Model: A theoretical explanation of factors influencing technology usage that hypothesizes perceived ease of use and perceived usefulness influence a person's intention, which in turn determines actual technology usage.

Theory of Reasoned Action (TRA): This theoretical model explains actions by identifying connections between various psychological constructs such as attitudes, beliefs, intentions, and behaviors; then posits that an individual's attitude toward a given behavior is determined by the belief that this behavior will result in a particular outcome.

Successful Health Information System Implementation

Kristiina Häyrinen
University of Kuopio, Finland

Kaija Saranto
University of Kuopio, Finland

INTRODUCTION

A Standish Group (1994) study showed that only 16% of all information technology projects come in on time and within budget. The situation is not better concerning health information systems. Many health information system implementations are less than completely successful (Berg, 2001; Giuse & Kuhn, 2003; Lorenzi & Riley, 2003). In this article, the health information system means "a system, whether automated or manual, that comprises people, machines and /or methods organized to collect, process, transmit, and disseminate" data that represent user information in healthcare (Kuhn & Giuse, 2001, pp. 275). What is successful implementation and whose success is measured? Successes can be measured in many ways. Delone and McLean have been finding out the success factors of management information system which are also applicable to health information system. The success factors are: system qualities, *e.g.*, the ease of use or time savings, information quality, *e.g.*, completeness or data accuracy, usage, *e.g.*, the frequency of use or the number of entries, user satisfaction, *e.g.*, user-friendliness or overall satisfaction, individual impact, *e.g.*, changed work practices or direct benefits and organizational impact, *e.g.*, communication and collaboration or impact on patient care. Furthermore, user involvement during system development, implementation and organizational culture have been identified as possible factors measuring the success. However, the need for further research to determine which attributes are the most useful ones in measuring success has also been revealed. (van der Meijden, Tange, Troost & Hashman, 2003).

The different phases in implementation process are, in general, user needs and requirements analysis (specification), system design, initial system implementation and testing (Ahmad, Teater, Bentley, Kuehn, Kumar, Thomas & Mekhjian, 2002; Schuster, Hall, Couse, Swayngim & Kohatsu, 2003; Souther, 2001). The system requirements analysis includes workflow analysis, and the initial system implementation includes the technical installation of the information system, integration of the information system to other information systems and users' training. Project management is an important factor in every phase of the implementation project.

The purpose of this article is to highlight the health information system implementation process from end-user perspective. Which factors are crucial in the implementation process from the point of view of the end-users? How does project management contribute to the implementation process, what is the role of the end-user in system designing and how does training effect the information system implementation?

BACKGROUND

The lack of financial support was the most significant barrier to successfully implementing information technology in healthcare from both clients' and vendors' perspective. The vendors' inability to deliver products, and difficulties in achieving end-user acceptance or use were the other barriers from the point of view of the clients. (HIMSS, 2002.) Costs are often underestimated because the cost of the software is only the beginning of other expenditures, *e.g.*, person-hours for training and support have been forgotten (Ash, Stavri & Kuperman, 2003).

The social and organizational issues, not only the technical ones, are the critical issues in the implementation of information systems. The health information systems do not effectively support the health processes, and terminology for the healthcare environment is needed. (Ahmad et al., 2002; Berg & Toussaint, 2003; Berg, 2001; Giuse & Kuhn, 2003; Kuhn & Giuse, 2001; Littlejohns, Wyatt & Garvican, 2003).

Human-computer interaction is also perceived as unsatisfactory. The human-computer interaction indicates the means by which humans interact with computers, *e.g.*, users enter and retrieve data. To optimize the design of the human-computer interaction, concepts are needed (Berg, 2001; Kuhn & Giuse, 2001). Technical issues, *e.g.*, integration with other information systems and the need for open systems are also issues which must be solved (Giuse & Kuhn, 2003; Kuhn & Giuse, 2001).

The reasons for failures were that the complexity of healthcare tasks and social and professional cultures of healthcare organizations was not taken into account and, furthermore, the education of the users was insufficient and the timing of the education was wrong (Littlejohns, Wyatt & Garvican, 2003). Lorenzi and Riley (2003) report that the failures of the implementation of the health information system can be classified into four categories: technical shortcomings, project management shortcomings, organizational issues and information explosion. The technical failures contain, *e.g.*, the old system maintenance and staff training. Project management issues are, *e.g.*, project management skills. Organizational issues are concerned with constant changes. Information explosion means that knowledge has increased exponentially and new technical tools have been developed to cope with the information. Berg (2001) notes that it is important to notice that the implementation is not only a technical installation, and also that the project is not only a technical project but also an organizational development project.

The three major reasons that a project will succeed are user involvement, executive management support and the clear statement of requirements (Standish Group, 1994). Doolan, Bates and James (2003) reported that the factors associated with successful implementation are unusually strong leadership, a clearly defined long-term commitment, clear focus on improving clinical processes and gaining clinical involvement and support improving productivity. Lorenzi and Riley (2003) included technical skills, project management skills and people and organizational skills to the success factors. The skills mean knowledge, experience and abilities in each area. Ahmad et al. (2002) stated that success factors are a continuous executive support, engagement of physicians, an effective implementation team, a consistent user-friendly interface and on-going user support.

User involvement during system development, implementation process and organizational culture may explain the failure of the information system. The attributes assigned to system development were the extent of user involvement, redesigning work practices and the reconstruction of content and technical limitations. Communication, training and technical support were attributes addresses to implementation process. Organizational aspects attributes were organizational culture, *e.g.*, control and decision-making, management support, professional values as well as support and maintenance. (van der Meijden, Tange, Troost & Hashman, 2003.)

THE ROLE OF THE PROJECT MANAGEMENT

A project approach is the most common way to implement health information systems. Project management is the process of planning for organizing and controlling projects. From the end-users' point of view, the objectives of health information system projects must make explicit, *i.e.*, improve patient care or efficiency. It is recommendable to set a stage for improvements, *e.g.*, reduce the number of phone calls or move manual files to on-line files. The objectives of the project must also improve workflows and work practice, in other words the hospital managements and also clinicians involved in the project must also upgrade their work performance. The implementation of information system must add value for the end-user. Clear objectives motivate the end-users for implementation. (Ash, Stavri & Kuperman, 2003; Berg, 2001; Berg & Toussaint, 2003; Doolan, Bates & James, 2003, FitzHenry & Snyder, 1996; Giuse & Kuhn, 2003; Littlejohns, Wyatt & Garvican, 2003; Lorenzi, Riley, Blyth, Southon & Dixon, 1997; Nikula, Elberg & Svedberg, 2000; Lechleitner, Pfeiffer, Wilhelmy & Ball, 2003)

The information system implementation process must be seen as an organizational change process (Anderson & Stafford, 2002; Berg, 2001; Lorenzi, Riley, Blyth, Southon & Dixon, 1997). Change management, which means "the process of assisting individuals and organizations in passing from an old way of doing things to a new way of doing things" (Lorenzi & Riley, 2003 pp.200), should be taken into account from the start of the implementation process. Organizational resistance always occurs during the implementation of new information systems. The change management is one reason why the leader has an important role in projects. (Lorenzi & Riley, 2003; Lorenzi, Riley, Blyth, Southon & Dixon, 1997; FitzHenry & Snyder, 1996) Furthermore, the implementation process itself requires effective leadership (Ash, Stavri & Kuperman, 2003;Lorenzi & Riley, 2003; Souther, 2001). Leadership is needed at multiple levels in organizations; high-level leadership was considered the single most important factor. It was demonstrated by the long-term commitment of resources. (Ahmad et al. 2002; Doolan, Bates & James, 2003; Littlejohns, Wyatt & Garvican, 2003). At the executive level, leadership is needed to promote a shared vision the purpose of health information system, which is e.g. to improve patient care. At the project management level, the leadership is needed to make practical, effective and useful decisions (Ash, Stavri & Kuperman, 2003). The need for bottom-up participative and top-down authoritarian approaches are also useful when changing organizational behavior. (FitzHenry & Snyder, 1996; Souther, 2001). External consults can also be used as change agents. (FitzHenry & Snyder, 1996; Lechleitner, Pfeiffer, Wilhelmy & Ball, 2003; Souther, 2001)

There is a need for different people, and many types of skills, both technical and medical, in implementation process. The persons who are assigned to the projects must be influential and knowledgeable and they must represent significant organizational groups and professionals. The challenge is to find the right persons. The multidisciplinary team approach has proven advantageous in health information system implementation.(FitzHenry & Snyder, 1996; Lechleitner, Pfeiffer, Wilhelmy & Ball, 2003; Littlejohns, Wyatt & Garvican, 2003; Souther, 2001) There must be different levels of teams: an executive steering team, a project steering team and a project work team. The end-users are representatives in different levels of teams. It is important that physicians and nurses belong to the project steering team, which makes major policy decisions based on user needs and requirements, organization infrastructure and general implementation strategies. The end-users also belong to the project work team, whose focus is training, user-interface, testing, security policy and pilot implementation. (FitzHenry & Snyder, 1996; Souther, 2001) The end-users' participation to the project enhanced their involvement.

The organizational readiness for information system innovation could also be assessed in advance. The information about the organizational readiness could be used in planning implementation strategy (Nikula, 1999; Snyder-Halpern, 2001; Yetton, Sharma & Southon, 1999), *e.g.*, in what order to engage different parts of the organization or in which phase of implementation process managerial influence must be at its highest. The healthcare organizational readiness for information system innovation could be assessed by the validated eight sub-dimensions of knowledge, end-users, technology, administrative support, management structures, processes, resources, values and goals. The end-user readiness themes include, *e.g.*, skills to use computers or involvement in the change process. The information about the end-user readiness could be used in planning their training. (Snyder-Halpern, 2001)

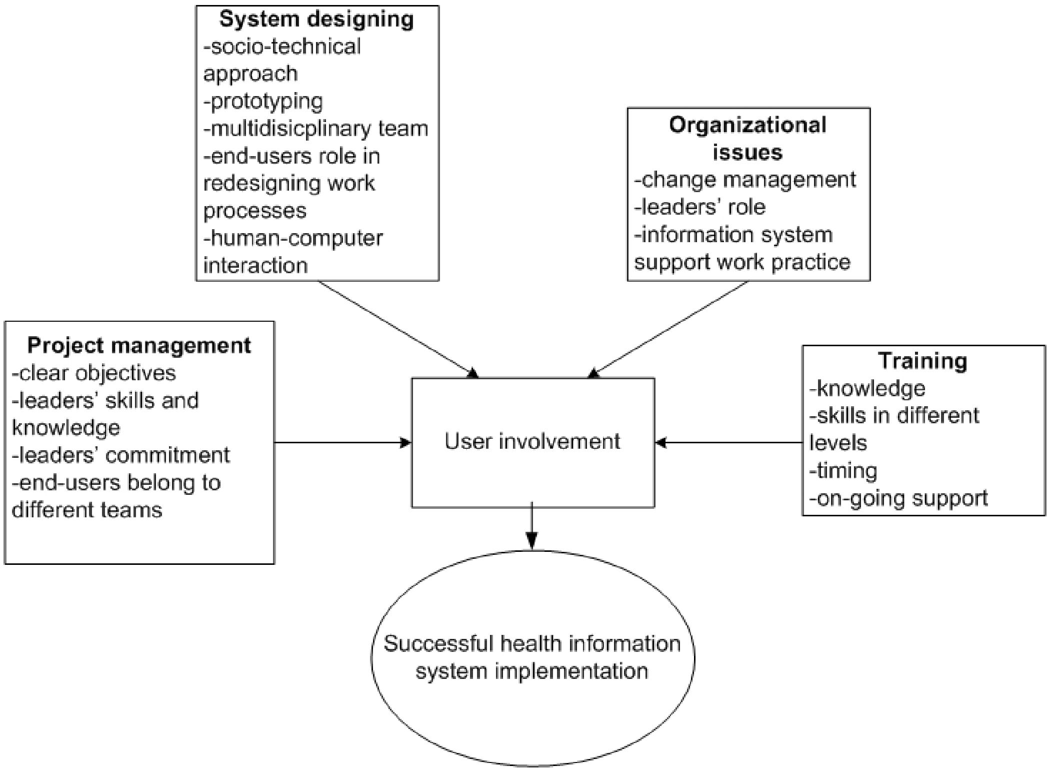
THE END-USERS' ROLE IN DESIGNING INFORMATION SYSTEM

For the design of information systems, the conventional approach has been to model a system completely before implementing it (Berg, 2001; Kuhn & Giuse, 2001; Lorenzi, Riley, Blyth, Southon & Dixon, 1997). Prototyping has proven a successful project (Ahmad et al., 2002; Doolan, Bates & James, 2003; Lechleitner, Pfeiffer, Wilhelmy & Ball, 2003; Lorenzi, Riley, Blyth, Southon & Dixon, 1997; Schuster, Hall, Couse, Swayngim & Kohatsu, 2003). Prototyping requires an iterative approach, in which information system analysis, design, implementation and evaluation merge. It is evident that the information system changes work practices and vice versa. Therefore, the first step is to model the environment of the system. A socio-technical approach is suitable for system design. It means that the work practices in which information systems will be used are the starting points for design and implementation of the information systems. This means defining the core actors, their core tasks, the flow of work and the flow of data. The present workflows are analyzed, the future workflows described and documented. (Berg & Toussaint, 2003; Berg, 1999; FitzHenry & Snyder, 1996; Schuster, Hall, Couse, Swayngim & Kohatsu, 2003; Souther, 2001) The interactive model increases end-user involvement. The need for optimally adapting information systems to the work practice in the healthcare organizations and the users' terminology are important, as well as reducing workload and simplifying the workflow (Anderson & Stafford, 2002; Berg & Toussaint, 2003; Doolan, Bates & James, 2003; Giuse & Kuhn, 2003, Kuhn & Giuse, 2001; Lechleitner, Pfeiffer, Wilhelmy & Ball, 2003; Lorenzi, Riley, Blyth, Southon & Dixon, 1997). Health professionals must develop the information system together with the vendors. The interaction between the vendor, the teams installing the information system and the users is a crucial point. Representatives must include all different health professionals and administrative personnel (Anderson & Stafford, 2002; Berg & Toussaint, 2003; Doolan, Bates & James, 2003; Kuhn & Giuse, 2001; Lechleitner, Pfeiffer, Wilhelmy & Ball, 2003; Lorenzi, Riley, Blyth, Southon & Dixon, 1997 Nikula, Elberg, Svedberg, 2000; Schuster, Hall, Couse, Swayngim & Kohatsu, 2003; Sleutel & Guinn, 1999)

TRAINING AND SUPPORT

The proper training and support are an essential part of the health information system implementation process (Ahmad et al., 2002; Anderson & Stafford, 2002; Ash, Stavri & Kuperman, 2003; Doolan, Bates & James, 2003; Littlejohns, Wyatt & Garvican, 2003; Lorenzi, Riley, Blyth, Southon & Dixon, 1997; Schuster, Hall, Couse, Swayngim & Kohatsu, 2003; Souther, 2001). An implementation of the health information system requires extensive support and training in order to achieve user acceptance. The training plan must be completed before the initial

Figure 1. The factors of successful implementation process- end-users perspective



implementation and it should also include intensive support at the time of the implementation. In the most successful implementation the end-users have more support after initial implementation than before it (Ash, Stavri & Kuperman, 2003).

There are also differences between computer literacy within healthcare professionals. (Saranto & Hovenga, 2004; Souther, 2001). Computer literacy means basic skills in computers use, *e.g.*, read and write. Information literacy consists of computer literacy, information retrieval abilities and communication skills. (Saranto & Hovenga, 2004) Therefore, the skills of the users, as well as their needs, must be taken into account when planning the training of a new information system. The quality of user training has a key role for a good user acceptance of information system (Anderson & Stafford, 2002; Quinzio, Junger, Gottwald, Benson, Hartmann, Jost, Banzhaf & Hempelmann, 2003). Disagreement on the efficiency of the system training has occurred (Lechleitner, Pfeiffer, Wilhelmy & Ball, 2003; Quinzio et al., 2003). The well-trained users were significantly more satisfied with the information system in routine use. A three-step training strategy has been recommended. First, the employees should be informed of the planned information system. The second phase includes general on-job training, and the third phase includes the opportunity to work on the individual problems of each user. (Quinzio et al., 2003) Whitman, Hamann & Vossler (1997) have also introduced a training plan. The strengths of the training plan are the contract that identified the roles and responsibilities of staff, management, and information services and staff development. Peer trainers have proven useful. Everyone must also have their own computer during training sessions. A competency test has been suggested as improvement to training plan. Ahmad et al. (2002) report a formalized training program for all users which precedes the implementation by two or four weeks. In addition, extensive on-site training occurred during the implementation period. The traditional help-desk phone has been found inadequate for user support.

FUTURE TRENDS

From the end-user perspective, the critical issues concerning the health information system implementation are project management, system designing, organizational issues and training and support. First of all, the project must have clear objectives. The teams must consist of different types of healthcare professionals, vendors, management and so on. It is important to find the right people to engage in the project. Project management task and responsibilities must be clear for each level of the teams and the interaction between the teams will also be efficient. The teams must work actively with the vendors. The management must also commit to the implementation process so the end-users notice that the project is important. The project management aids the change management, which is one of the critical issues of information system implementation process. The leaders can use the information about the organizational readiness in planning the implementation strategy.

It is crucial that both representatives from the future users and leaders are involved in the design process. They are the experts of their work. The iterative development process will be used and the information systems adapting to workflows is important. When the information systems' requirements and future workflows are modeled properly it becomes evident that the work processes change, because the healthcare professionals must consider what they do and why they do it.

The training plan must be designed carefully. The information literacy of the end-users will be taken into account. A multilevel approach is good. First, the vendor can present the main functions of the information system. After that the tutors teach the users to use the information system and the users can train the system themselves. The users are also supported, they can ask questions. After that the users themselves can practice and also contact the tutors if they have any problems using the information system.

CONCLUSION

In the first stages the point of view in health information system implementation was only technical. The implementation process paid no attention to workflow patterns prior to the implementation, or to the computer experience among personnel or aspects of the organizational culture. Strategies for the successful management of health information system implementation must be planned carefully. The personnel of an organization must learn to use the information system and also change their working processes. Education is the critical point of the implementation process. Knowledge about end-user attitudes and skills prior to computerization is useful. The training will be tailored to meet the varying needs of user groups. There is a need for research in the area of evaluating health information system implementation.

REFERENCES

Ahmad, A., Teater, P., Bentley, T.D., Kuehn, L., Kumar, R.R., Thomas, A. & Mekhjian, H.S. (2002). Key attributes of a successful physician order entry system implementation in a multi-hospital environment. *Journal of the American Medical Informatics Association: JAMIA, 9*(1), 16-24.

Anderson, L. K., & Stafford, C. J. (2002). The "Big Bang"; implementation: not for the faint of heart. *Computers in Nursing, 20*(1), 14-20.

Ash, J. S., Stavri, P. Z., & Kuperman, G. J. (2003). A consensus statement on considerations for a successful CPOE implementation. *Journal of the American Medical Informatics Association: JAMIA, 10*(3), 229-234.

Berg, M. (1999). Patient care information systems and health care work: a sociotechnical approach. *International Journal of Medical Informatics, 55*(2), 87-101.

Berg, M. (2001). Implementing information systems in healthcare organizations: myths and challenges. *International Journal of Medical Informatics, 64*(2-3), 143-156.

Berg, M., & Toussaint, P. (2003). The mantra of modeling and the forgotten powers of paper: A sociotechnical view on the development of process-oriented ICT in health care. *International Journal of Medical Informatics, 69*(2-3), 223-234.

Doolan, D. F., Bates, D. W., & James, B. C. (2003). The use of computers for clinical care: a case series of advanced U.S. sites. *Journal of the American Medical Informatics Association: JAMIA, 10*(1), 94-107.

FitzHenry, F., & Snyder, J. (1996). Improving organizational processes for gains during implementation. *Computers in Nursing, 14*(3), 171-80.

Giuse, D. A., & Kuhn, K. A. (2003). Health information systems challenges: the Heidelberg conference and the future. *International Journal of Medical Informatics, 69*(2-3), 105-114.

HIMSS (2002). 13th Annual HIMSS Leadership Survey. Retrieved June 5, 2004, from *http://www.himss.org/2002survey/*

Kuhn, K.A. & Giuse, D.A. (2001). From hospital information systems to health information systems: Problems, challenges, perspectives. *Methods of Information in Medicine*, 40, 275-287.

Lechleitner, G., Pfeiffer, K. P., Wilhelmy, I., & Ball, M. (2003). Cerner millennium: the Innsbruck experience. *Methods of Information in Medicine, 42*(1), 8-15.

Littlejohns, P., Wyatt, J. C., & Garvican, L. (2003). Evaluating computerised health information systems: Hard lessons still to be learnt. *BMJ (Clinical research ed.), 326*(7394), 860-863.

Lorenzi, N. M., & Riley, R. T. (2003). Organizational issues = change. *International Journal of Medical Informatics, 69*(2-3), 197-203.

Lorenzi, N. M., Riley, R. T., Blyth, A. J., Southon, G., & Dixon, B. J. (1997). Antecedents of the people and organizational aspects of medical informatics: review of the literature. *Journal of the American Medical Informatics Association: JAMIA*, *4*(2), 79-93.

Nikula, R. E. (1999). Organizational and technological insight as important factors for successful implementation of IT. *Proeedings of the AMIA Symposium*, , 585-588.

Nikula, R. E., Elberg, P. B., & Svedberg, H. B. (2000). Informed decisions by clinicians are fundamental for EPR implementation. *International Journal of Medical Informatics*, *58-59*, 141-146.

Quinzio, L., Junger, A., Gottwald, B., Benson, M., Hartmann, B., Jost, A., Banzhaf, A. & Hempelmann, G. (2003). User acceptance of an anaesthesia information management system. *European Journal of Anaesthesiology, 20*(12), 967-972.

Saranto, K., & Hovenga, E. J. (2004). Information literacy-what it is about?: Literature review of the concept and the context. *International Journal of Medical Informatics, 73*(6), 503-513.

Schuster, D. M., Hall, S. E., Couse, C. B., Swayngim, D. S., & Kohatsu, K. Y. (2003). Involving users in the implementation of an imaging order entry system. *Journal of the American Medical Informatics Association: JAMIA, 10*(4), 315-321.

Sleutel, M., & Guinn, M. (1999). As good as it gets? Going online with a clinical information system. *Computers in Nursing*, *17*(4), 181-185.

Snyder-Halpern, R. (2001). Indicators of organizational readiness for clinical information technology/systems innovation: a Delphi study. *International Journal of Medical Informatics*, *63*(3), 179-204.

Souther, E. (2001). Implementation of the electronic medical record: the team approach. *Computers in Nursing, 19*(2), 47-55.

Standish Group 1994. The CHAOS Report. Retrieved June 5, 2004, from *http://www.standishgroup.com/sample_research/chaos_1994_1.php*

van der Meijden, M. J., Tange, H. J., Troost, J., & Hasman, A. (2003). Determinants of success of inpatient clinical information systems: A literature review. *Journal of the American Medical Informatics Association: JAMIA, 10*(3), 235-243.

Yetton, P., Sharma, R., & Southon, G. (1999). Successful IS innovation: The contingent contributions of innovation characteristics and implementation. *Journal of Information Technology*, 14, 53-68.

Whitman, B. L., Hamann, S.K., & Vossler, B. L.(1997). A Training plan for Bedside Computers. *Journal of Nursing Staff Development*, *13*(1) 33-36.

KEY TERMS

Change Management: Change management is the process of assisting individuals and organizations in passing from an old way of doing things to a new way doing things.

Health Information System: Health Information System is the system, whether automated or manual, that comprises people, machines and /or methods organized to collect, process, transmit and disseminate data that represent user information in health care.

Implementation: The implementation of information system includes different phases: user needs and requirements analysis (specification), system design and initial system implementation and testing. The system requirements analysis includes workflow analysis; the initial system implementation includes technical installation of information system, integration the information system to other information systems and users' training.

Information Literacy: Information literacy consists of computer literacy and information retrieval abilities and communication skills. Computer literacy means basic skills in computers use, *e.g.*, being able to read and write with computers.

Project Management: Project management is the process of planning for organizing and controlling projects. The multidisciplinary team approach in project management means that there are different levels of teams: the executive steering team, the project steering team and the project work team. Different teams have their own tasks.

Prototyping: Prototyping is the development approach in which prototypes are produced. It is an iterative approach to the developing of the information system. The distinctions of the different phases of development of the information system, *e.g.*, analysis, design, implementation and evaluation are not clear.

Socio-Technical Approach: Socio-technical approach means that the work practices in which information systems will be used are the starting points for design and implementation of the information systems.

Supporting Assurance and Compliance Monitoring

Peter Goldschmidt
The University of Western Australia, Australia

INTRODUCTION

Governments and commercial organizations typically use monitoring facilities that depend on data that identify source agents and their relationships, to detect and draw attention to possible anomalies and potential non-compliance.

The assurance of compliance monitoring requires decision support and appropriate domain knowledge, relevant to the level of user, to manage the results of the surveillance. This is required in order to fulfill the necessary and sufficient evidence verifying or refuting this output.

BACKGROUND

This article discusses methods to support assurance of surveillance monitoring and output verification knowledge management (CV-KM), including a brief discussion on primary monitoring systems; the different environments in which they operate; the verification problem solving and decision making tasks; the problem structure and the coordination of the review process to facilitate truth maintenance. The surveillance operation is considered a primary monitoring function, with the analysis of the resulting output the secondary monitoring function - the assurance component.

Examples of monitoring systems range from standard data processing routines that ensure internal control, such as data input, processing and output compliance. Weber (1999) provides a comprehensive discussion on these processes, to the monitoring of events transacted in more complex environments, such as fraud detection, intrusion detection, data mining systems and the like, via sophisticated statistical, artificial intelligence and neural computing techniques, or hybrid combinations. These devices are termed primary monitoring systems (PSS).

Assuring, verifying and managing PSS information quality and integrity is fundamental to the success of modern information-dependent organizations. Concurrent with the need for surveillance is a need to maintain personal privacy, due diligence, and accountability (Cillufo, 2000).

Clarke (1988) highlights the inherent dangers of drawing conclusions resulting from the electronic monitoring of data related to individuals and groups of individuals, and points out that a major problem in "dataveillance" is the high noise to signal ratio, which may be misleading. Davis and Ord (1990) acknowledge the problem of setting threshold levels in an ever-changing environment. With any set of tolerance levels, deviant (even fraudulently motivated) behaviour may escape detection. Tightening tolerance levels limits increases the likelihood that exception conditions will trigger an alert but also increases false positive alerts since the number of instances that fall outside the tolerance increases. The cost for the analyst (the decision-maker) to review the additional non-exception condition alerts must be assessed in relation to the imputed value of identifying the additional true exceptions detected by more stringent limits (Davis & Ord, 1990).

Advances have, in general, reduced the problem of misleading results produced from "noisy data," including improvements in data processing and the increased use of sophisticated computational techniques such as statistical, knowledge-based and artificial neural computational methods. These systems are centered on the events being monitored and the events' source agents. Their results, however, may still require human judgment to determine their validity (Goldschmidt, 2001). CV-KM systems act as a secondary monitoring facility supporting, verifying and assuring data and information compliance by assisting in analyzing and categorizing exceptions, or results, generated by PSS. CV-KMs assist in assuring the fulfillment of the necessary and sufficient evidence supporting (true positive/negative) or refuting (false positive) hypotheses of non-compliance. The input to CV-KMs requires the output resulting from the organization's domain-specific PSS plus related information. Operationally, the CV-KMs are a bolt-on addition to the PSS.

WHAT ARE PRIMARY SYSTEMS?

Typically, these systems examine the integrity of transaction data as well as the entire transaction, or event, to ensure compliance with predetermined conditions. An exception report identifies any variances. This identification either fulfills the conditions of necessary and sufficient evidence and determines an instance of non-compliance, or indicates possible non-compliance. In the latter case further evidence may be sought to substantiate the hypothesis of non-compliance.

The function of PSS is twofold: identifying a variance, and producing and accumulating supporting evidence. When both these conditions are met, the evidence points to the detective, corrective or preventative actions required.

The detective function is fulfilled by recognition of the variance; correction can then be made to the data or the event, which is then reprocessed. The preventative function is fulfilled by the recognition of the variance resulting in the rejection of the event. Decision-makers must interpret ambiguous evidence to determine what action is required, or if the non-compliant indicator is a true or a false positive directive.

Examples of PSS range from standard data processing routines that ensure internal control, such as data input, processing and output compliance, to the use of sophisticated statistical (procedural) techniques, artificial intelligent (declarative) techniques and neural (associative) techniques, or hybrid combinations. In general, computational techniques are either demons or objects (O'Leary, 1991; Vasarhelyi & Halper, 1991). Demons are computerized routines that are instantiated by data or events received, as opposed to being requested by some program. "Demons add knowledge to a system without specification of where they will be used ... like competent assistants they do not need to be told when to act" (Winston, 1977, p. 380). They are data or event dependent, rather than program dependent, and provide intelligent self-activation for monitoring data triggered by compliance threshold levels. O'Leary points out that demons have been developed to monitor patterns for the purpose of auditing activities conducted on computer-based systems. Vasarhelyi and Halper describe an alternate: CPAS, Continuous Process Audit System. CPAS allows for the continuous audit of on-line systems by monitoring transactions to determine variance between monitored information and expected information.

THE PSS AND CV-KM ENVIRONMENT

PSS and CV-KM can be classified by levels of complexity, characterized by their place on the simple or complex environmental continuum in which they operate and the decisions required to determine instances of non-compliance. Constraints may take the form of an organization's predetermined policies and procedures, needed to ensure data and event integrity, contractual agreements, and statutory requirements. These constraints are not mutually exclusive and can be seen as bounds or threshold levels. The parameters used to construct these levels may change with modifications to threshold requirements such as evolutionary changes in constraints and changes in data and event requirements. A simple environment is so-called because: 1) the threshold levels either seldom change or only change over the longer term; 2) the identification of the variance fulfils the conditions of necessary and sufficient evidence to determine an instance of non-compliance; and 3) the decisions, needed to determine if events comply, lie on the structured to highly structured portion of the decision-making continuum. The degree to which the bounds of the threshold levels are set, very narrow to very broad, determines the type of decision required. Under a simple environment the bounds or threshold limits are narrow, characteristic of structured decisions such as data input integrity and customer credit checks. Decision-making in this environment is ex-ante, made of a single step, and the constraints are all predetermined.

In a complex environment, decision-making is ex-post, complex and may require multiple steps. Initial monitoring uses a priori thresholds broader than in a simple environment, that is, more granular and produces exceptions that identify suspected non-compliant events (SNCEs). Once these exceptions have been produced, the decision-maker must substantiate true positive exceptions. This task must be broken down into smaller components and subgoals must be developed (Simon, 1973) to identify, categorise and discard any false positive exceptions. False negatives do not generate an exception, and allow possible suspect events to slip through the surveillance sieve. If the threshold limits are stringent enough, marginal false negatives could be subsumed and later considered. Nevertheless, this would not necessarily reduce the occurrences of *true* false negatives, as their characteristics may not be known. True positives are those exceptions that the decision-maker has determined are indeed anomalous. Evidence for this decision uses the results of the initial monitoring as well as important information related to the event, characterized by a need for judgmental expertise. Examples of these approaches to complex environments include: Byrnes et al. (1990), Major and Riedinger (1992), Senator et al. (1995), and Kirkland et al. (1999).

CV PROBLEM SOLVING AND DECISION-MAKING TASKS

Secondary monitoring problem solving, human evaluation of the exceptions produced by the primary monitoring system, determines if a generated exception is feasible. This is similar to an analytical review (AR) conducted by auditors, characterised by Libby (1985) as a diagnostic-inference process. Koonce (1993) defines AR as the diagnostic process of identifying and determining the cause of unexpected fluctuations in account balances and other financial relationships. Similarly, secondary monitoring problem solving identifies and determines the causes of unexpected variances resulting from the primary monitoring facility. Blocher and Cooper (1988) found that analytical review (AR) typically follows four distinct diagnostic inference components: accumulation and evaluation of relevant information; initial recognition of unusual fluctuations; subsequent hypothesis generation; and information search and hypothesis evaluation.

With CV-KM, accumulation and evaluation is guided by the results of the PSS. Subsequently, a hypothesis of the potential causes of the observed variance is generated. The diagnostic approach takes the form of defeasible logic, which means that any inference made may be only tentative, as the inference may require revision if new information is presented. The decision-maker must evaluate all possible legitimate reasons for the occurrence of the variant. If none is found, the hypothesis of non-compliance is strengthened.

CV PROBLEM STRUCTURE

Following Sol (1982), structuredness of the complex problem is twofold: the variance identification is the structured component, and the accumulation of evidence supporting or refuting the non-compliant event (NCE) hypothesis is the ill-structured component. The variance is typically the product of some algorithm indicating a possible occurrence of NCEs, but in order to substantiate a true NCE the required accumulation of evidence requires judgment of agent behaviour. The agents include the source of the event, the identification of the source agents' possible motivations, the environment in which the source agent is operating and the impact this event may have on the environment.

COORDINATION: THE REVIEW PROCESS TO FACILITATE TRUTH MAINTENANCE

Coordination refers to the managing of interactions between multiple agents cooperating in some collective task. Pete et al. (1993) show that optimal organizational design depends on the task environment and, as with an audit team or group, is hierarchical. The objective is to reduce the problems discussed by Freedman (1991), to reduce any potentially redundant activities conducted by the evaluating agents, and to increase efficiency and effectiveness. The agent or agents may be human or machine based. Machine based or independent software agents function as repositories of human opinions related to the event under scrutiny.

The process of review when evaluating judgments made on accounting data and information is well established in the auditing literature (Libby & Trotman, 1993). To facilitate coordination, evaluating agents should communicate their findings via a communication protocol. Communication protocol establishes the means and modes of communication between agents. Information exchange can be either via an implicit communication mechanism such as a common memory or blackboard (Hayes-Roth et al., 1983), or via explicit communication mechanisms, such as message sending. Using the blackboard approach, the SNCE's details plus the evaluating agents' assumptions and results are posted. This facilitates the more senior agents imposing their criteria on lesser agents' results, as well as using their task-specific criteria to further refine the classifications.

Computerised decision support systems have been proposed and built to address some of the previously mentioned problems. A limited framework for a CV-KM intelligent decision support system using multi-agent technology is presented in Chang et al. (1993) and Goldschmidt (1996, 2001).

FUTURE TRENDS

With the increase in the reliance of electronic communications for business, industry, medicine, defense and government, assuring, verifying and managing the integrity of transactions and managing the results of these monitoring systems for information quality and integrity

S

Advantages of CV-KM

The company ALERT-KM Pty Ltd holds the CV-KM IP in 24 countries and is currently commercializing this technology.

Adds functionality to primary monitoring infrastructure, without modifying primary system.
Proposes a framework for compliance verification knowledge management.
Provides for the decomposition of surveillance tasks.
Provides a consistent evidence evaluation and combination structure.
Provides records of evidence from each stage.
Adds value to surveillance operations by reducing the cost of surveillance monitoring, assisting in surveillance accountability and providing transparency, when required, thereby contributing to surveillance governance and due diligence.
Employs a method that adds value to a generated exception by encapsulating and associating the event's attributes, its source agent's characteristics, the evaluating agent's analysis and the recommended remedial action plus the substantiating evidence.
Exploits an infrastructure support construct and secondary filter, allowing for collaboration, truth maintenance, audit trails and decision support, thereby facilitating decision consistency and greater processing volume.
Using the approach as a decision aid and secondary filter, analysis of results can then be used to review the analyst's decision-making processes and to refine the primary filter tolerance levels.
Supports a structured, flexible and inclusive approach to surveillance analysis.
By adding a cost function to the surveillance-monitoring infrastructure can capture cost-benefit trade-off.
Insight is gained from the knowledge acquisition component when setting up parameters and heuristics.
Assists in the development of an effective accountability structure.
Reduces distrust of surveillance monitoring systems, by reinforcing accountability and transparency.

is fundamental to the success of modern information-dependent organizations. Concurrent with the need for surveillance is a need to maintain personal privacy, due diligence, and accountability (Cillufo, 2000).

Data mining technology, distributed heterogeneous database access and information distribution has moved from a silo approach to a more pooled approach within organizations. This has led to further need for information assurance that is reliant on data monitoring. Therefore the management of the accuracy and validity of the monitoring output necessitates the assurance of decisions made based on this information.

CONCLUSION

CV-KM operates in highly complex environments, domains where the threshold granularity is high and the decision-making time factor is short may benefit from the decision support discussed. It is essential for accountability that organizations in these domains ensure transactions identified as suspected NCE are scrutinized and substantiated. This assists in minimizing false positive conclusions that may result from the speed, volume and increased complexity of transactions, and the information used to analyze them. CV-KM also addresses some of the problems highlighted by Clarke (1988), that electronic monitoring of data related to individuals and groups of individuals is subject to inherent dangers of drawing misleading conclusions from these data. Assurance and compliance monitoring team infrastructure support includes aspects of information systems, cognitive sciences, decision support and auditing judgment. Fuzzy set theory is advocated in decision environments where there may be a high degree of uncertainty and ambiguity, catering for qualitative and quantitative evidence validating and assuring the assertion of noncompliance.

Current research efforts in monitoring and assurance systems (Roohani, 2003; Schneier, 2001; SRI, 1999; UCD, 1996)still concentrate on improving the efficiency and accuracy of primary monitoring systems. Whilst this is necessary, further research opportunities exist in addressing and improving the utility and effectiveness of supporting the analysts responsible for evaluating the results of these primary systems and ensuring their accountability.

REFERENCES

Blocher, E., & Cooper, J. (1988, Spring). A study of auditors' analytical review performance. *Auditing: A Journal of Practice and Theory,* 1-28.

Byrnes, E., Thomas, C., Henry, N., & Waldman, S. (1990, Summer). INSPECTOR An expert system for monitoring world-wide trading activities in foreign exchange. *AI Review.*

Chang, A., Bailey, A., Jr., & Whinston, A. (1993). Multi-auditor cooperation: A model of distributed reasoning. *IEEE Transactions on Engineering Management*, *20*(4), 346-59.

Cillufo, F. (2000, September). Cyber attack: The national protection plan and its privacy implications. *Journal of Homeland Security*. *http://www.homelandsecurity.org/journal/*. ANSER Analytical Service, Inc, Arlington, VA.

Clarke, R. (1988). Information technology and dataveillance. *Communications of the ACM*, *31*(5), 498-512.

Davis, S., & Ord, K. (1990). Improving and measuring the performance of a security industry surveillance system. *INTERFACES*, *20*(5), 31-42.

de Kleer, J. (1986). An assumption based truth maintenance system. *Artificial Intelligence, 28*(2), 127-162.

Denning. *http://www.cs.georgetown.edu/~denning/index.html*

FBI. (1999). Digital storm. FBI Annual Report. *http://www.fbi.gov/programs/*lab/labannual99.pdf

Freedman, R.S. (1991). AI on Wall Street. *IEEE Expert, 6*(2), 2-7.

Goldschmidt, P. (1996). *Compliance monitoring for anomaly detection in a complex environment using multiple agents: Supporting capital market surveillance*. PhD dissertation. The University of Western Australia.

Goldschmidt, P. (2001). Assurance and compliance monitoring support. In G. Dhillon (Ed.), *Information security management: Global challenges in the next millennium* (ch. 10, pp. 135-154). Hershey, PA: Idea Group Publishing.

Hayes-Roth, F., Waterman, D.A., & Lenat, D.B. (1983). *Building expert systems*. Reading, MA: Addison-Wesley.

Koonce, L. (1993). A cognitive characterisation of audit analytical review. *Auditing: A Journal of Practice and Theory*, *12*(Supplementary), 57-76.

Libby, R., & Trotman, K. (1993). The review process as a control for differential recall of evidence in auditor judgements. *Accounting, Organisations and Society*, *18*(6), 559-74.

Major, J., & Riedinger, D. (1992). EFD: A hybrid knowledge / statistical-based system for the detection of fraud. *International Journal of Intelligent Systems, 7,* 687-703.

O'Leary, D. (1991). Artificial intelligence and expert systems in accounting databases: Survey and extensions. *Expert Systems with Applications, 3,* 143-52.

Roohani, S.J. (Ed.). (2003). *Trust and data assurances in capital markets: The role of technology solutions.* Bryant College, RI. PriceWaterhouseCoopers research monograph.

Schneier, B. (2001). *http://www.counterpane.com/msm.html*

Senator, T., Goldberg, H., Wooten, J., Cottini, M., Khan, A., Klinger, C. et al. (1995). Financial crimes enforcement network AI system (FAIS). *AI Magazine, 16*(4), 21- 39.

Simon, H. (1973). The structure of ill structured problems. *Artificial Intelligence, 4*(3-4), 181-201.

SRI. (1999). *http://www.csl.sri.com/%7eneumann/det99.html*

Trotman, K.T. (1985). The review process and the accuracy of auditor judgements. *Journal of Accounting Research, 23*(2), 740-52.

UCD. (1996). *http://olympus.cs.ucdavis.edu/cmad/4-1996/slides.html*

Vasarhelyi, M.A., & Halper, F.B. (1991). The continuous audit of online systems. *Auditing: A Journal of Theory and Practice, 10*(1), 110 -25.

Winston, P. (1977). *Artificial intelligence*. Reading, MA: Addison-Wesley.

KEY TERMS

Complex Environments: Complexity increases as: the granularity increase, the frequency of changes increases, the time availability decreases, and the degree of judgment required increases. Decision-making is ex-post, complex and may require multiple steps. Initial monitoring uses a priori thresholds broader than in a simple environment, that is, more granular and produces exceptions that identify suspected non-compliant events (SNCEs). Evidence for decision-making uses the results of the initial monitoring as well as important information related to the event, characterized by a need for judgmental expertise. Examples of these approaches to complex environments include: Byrnes et al. (1990), Major and Riedinger (1992), Senator et al. (1995), and Kirkland et al. (1999).

Compliance Verification: Ensuring the necessary and sufficient evidence supports the assertion of non-compliance.

Dataveillance: Surveillance of data using automated data analysis to identify variances. These typically depend on data that identify source agents and their relationships and is used to draw a compliance analyst's attention to a particular event or group of events that indicate possible anomalies.

Primary Surveillance Systems (PSS): Processes, methods or devices that monitor data identifying events, their source agents and their relationships, to draw attention to possible anomalies and potential non-compliance.

Secondary Monitoring: Secondary monitoring supports agents with verifying and assuring data and information compliance by assisting in analyzing and categorizing exceptions, or results, generated by PSS. This assists in assuring the fulfillment of the necessary and sufficient evidence supporting (true positive/negative) or refuting (false positive) hypotheses of non-compliance. The input is the results of the organization's domain-specific PSS plus related information and human judgment from either human agents or machine agents encoded with heuristics.

Simple Environments: Monitoring environments in which the threshold levels either seldom change or only change over the longer term; the identification of the variance fulfils the conditions of necessary and sufficient evidence to determine an instance of non-compliance; and the decisions, needed to determine if events comply, lie on the structured to highly structured portion of the decision-making continuum. Decision- making in this environment is ex-ante, made of a single step, and the constraints are all predetermined.

Supporting Compliance Monitoring Assurance: A process or method supporting human or machine agents in verifying and assuring the validity of generated suspected non-compliant event by assuring the necessary and sufficient evidence supporting the hypothesis of non-compliance.

Suspected Non-Compliance Event (SNCE): An event triggering an exception by the primary monitoring facility that indicates instances of possible non-compliance. The exception still requires verification.

Supporting the Evaluation of Intelligent Sources

Dirk Vriens
Radboud University of Nijmegen, The Netherlands

INTRODUCTION

To survive, organizations need to produce and process information about their environment, for instance, about customers, competitors, suppliers, governments, or all kinds of socioeconomic and technological trends. The process of obtaining this information is often called competitive intelligence (cf Fleisher & Blenkhorn, 2001; Kahaner, 1997; Vriens, 2004). An important stage in the competitive intelligence process is the collection stage. In this stage, one has to determine relevant sources, access them, and retrieve data from them (cf Bernhardt, 1994; Kahaner). For each data class, many possible sources are available, and determining the right ones is often difficult. Moreover, accessing sources and retrieving data may require a lot of effort and may be problematic (cf Cook & Cook, 2000; Fuld, 1995; Kahaner, 1997). In this chapter, we present a tool for supporting the effective and efficient use of sources: the source map. In essence, a source map links data classes to sources and contains information about these links. This information indicates the adequacy of sources in terms of ease of access, ease of retrieval, and usefulness of the retrieved data. A source map can support the selection of appropriate sources and it can support the assessment of the overall adequacy of available sources.

BACKGROUND

The process of competitive intelligence is often described as a cycle of four stages (the intelligence cycle; see Kahaner, 1997; Vriens, 2004). This cycle comprises (a) the direction stage (in which the organization determines about what aspects in the environment data should be collected), (b) the collection stage (where sources are determined and data are collected), (c) the analysis stage (in which the data are analyzed to assess whether they are useful for strategic purposes), and (d) the dissemination stage (where the data are forwarded to decision makers; Bernhardt, 1994; Gilad & Gilad, 1988; Herring, 1999; Kahaner, 1997; Sammon, 1986). The collection stage is considered to be the most time-consuming stage (e.g., Chen, Chau, & Zeng, 2002) and if it is not performed carefully, many difficulties arise (e.g., too much time spent on search, collection stage leads to irrelevant data, information overload; see, for example, Cook & Cook, 2000; Chen et al.; Teo & Choo, 2001; Vriens & Philips, 1999). For successfully carrying out collection activities, knowledge about what sources contain what kind of data and knowledge about how to approach these sources (metaknowledge regarding the collection of data) would be very helpful. This chapter presents a tool to structure and deal with this metadata: the source map.

To collect data about the environment one has to

1. identify possible sources,
2. judge the value of the source (in terms of different criteria; e.g., does it contain relevant data? What are the costs of employing this source? Is it reliable?), and
3. use value judgments to select the appropriate sources.

Many authors discuss Step 1 by pointing to a variety of available sources (cf Fuld, 1995; Kahaner, 1997; Sammon, 1986). Typical sources include the Internet, online databases, sales representatives, internal or external experts, CEOs, journals, tradeshows, conferences, embassies, and so forth.

The literature treats the valuation step more implicitly. It discusses distinctions regarding sources, such as open versus closed sources, internal versus external sources, or primary versus secondary sources (Fleisher & Blenkhorn, 2001; Kahaner, 1997). These distinctions implicitly refer to criteria used in the valuation of sources. The distinction of open versus closed sources implicitly refers to, for instance, criteria such as ease in collection or relevance. The distinction of primary versus secondary sources implicitly refers to the criterion of the reliability of the data. In our view, it is possible to value sources more precisely when the valuation criteria are stated explicitly and not implicitly in the form of these distinctions.

The selection step is even more elusive in literature (and practice). This step integrates value judgments to select appropriate sources for collecting the required data. Few methods seem to be designed for source selection.

In this article, we propose a tool to structure and support the valuation and selection of sources: the source map. This tool builds on Fuld's (1995) intelligence maps and knowledge maps (e.g., Davenport & Prusak, 1998).

The purpose of the source map is to help pin down the appropriate sources quickly and detect weaknesses in the available sources.

THE SOURCE MAP AS A TOOL FOR ASSESSING SOURCES

What is a Source Map?

A source map links data (or classes of data) to sources in such a way that the (most) appropriate sources can be selected for the collection of the requested data. If viewed as a matrix, the column entries may refer to data classes (e.g., products under development by competitor X) and the row entries to possible sources. Each column then indicates what sources may be used to gather the requested data (e.g., a patent database, economic journals, or the Internet site of competitor X). To determine what sources are (most) appropriate, the source map needs to contain information about criteria for appropriateness and their valuation. The cells in the source map (connecting the data classes to sources) should contain this information. To get this information, it should be clear (a) what the relevant criteria are, (b) how they can be given a value, and (c) how to integrate them into an overall judgment of the appropriateness of the sources. The next two sections deal with these issues.

Note that we treat the source map as a tool for supporting and structuring collection activities *given* the data classes. We assume that the data (classes) are already defined in the direction phase (the first phase of the intelligence cycle).

Criteria and Scores for Judging Sources

The criteria for assessing the appropriateness of sources link up with the three activities required to deal with sources. These activities are the following.

1. Accessing the source. Accessing means determining the exact location and approaching the source to prepare retrieval.
2. Retrieving (in interaction with the source) the data from the source.
3. Using the retrieved data in further processing (i.e., for the production of intelligence).

Referring to these activities, the appropriateness of sources depends on four dimensions: (a) ease of access, (b) ease of retrieval, (c) usefulness of the content of the retrieved data and processing ease, and (d) cost effectiveness. Below, we discuss criteria in these dimensions.

Criteria for Access and Retrieval

To assess the appropriateness of sources regarding access and retrieval, barriers in employing a source can function as criteria (cf Fuld, 1995; Davenport & Prusak, 1998). Examples of these barriers are as follows.

- A language barrier.
- A cultural barrier (i.e., a difference in culture between collector and source).
- An institutional barrier. In some (bureaucratic) organizations, it may be very hard to locate and approach certain people and documents.
- A personal barrier. Personal characteristics can make it difficult to approach and interact with someone.
- A geographical barrier. Some sources need to be dealt with on location.
- A technological barrier. Accessing some sources and retrieving data from them may sometimes be possible only by means of specific information and communications technology, requiring specific knowledge or skills.
- A fee barrier. For accessing some sources and/or retrieving data, a fee may be charged.
- A time barrier. For some sources, the response time may be very slow.
- A clarity barrier. This barrier refers to the effort one has to give to make sense of the data from the source. Factors that increase this barrier are the use of specific jargon and the lack of (requested) structure in the data.
- A stability barrier. This barrier refers to the stability of access to the source (some sources may cease to exist, some are not available at the expected moment, others may decide to stop providing their services, etc.).

In our view, these criteria can also be used to express the costs associated with using a particular source. We therefore prefer to deal with the above criteria, instead of using cost estimates that may be derived from them, because (a) it is difficult to translate the criteria into costs and (b) if only cost estimates are used, one loses information about the appropriateness of sources.

Using a barrier as a criterion to assess appropriateness, it can be scored on a five-point Likert scale where 1 means *very problematic* and 5 means *nonexistent*.

Criteria for the Use of Data

There are four criteria for assessing the appropriateness of sources regarding the use of the data for the production of intelligence. One of them is a processing criterion and three of them are content criteria.

The processing criterion refers to the ease of processing. This can be determined by the format in which the data are delivered; that is, does the source deliver the data in a format that can be used directly for the purposes of the collector or does it need reformatting? One may score this criterion on a five-point scale ranging from 1, *much reformatting needed*, to 5, *right format*.

The content criteria are completeness, reliability, and timeliness (cf O'Brien, 1998, for a summary of these criteria). When applied to the value of sources, these criteria mean the following.

- Completeness: The source can deliver all the data required to gain insight into the data class for which the source is used. This can, for instance, be measured in terms of the number of times the source was unable to deliver the requested data and/or the number of aspects of a data class for which the source could not provide data.
- Reliability: This refers to the reliability of the data from the source. It can be measured, for instance, in terms of the number of times the data from the source proved to be incorrect.
- Timeliness: The data from the source is up to date. It can be measured in terms of the number of times that the source delivered obsolete data.

In literature, one often finds relevance as an additional criterion to assess the content of data. Relevance then refers to the suitability of the data in gaining insight into the data class for which the source was used. However, relevance can be adequately expressed in terms of completeness, reliability, and timeliness. Completeness links the data provided by a source to the required data defined by the data class. Given the completeness, the data should further be correct and up to date to be relevant. Relevance, therefore, can be treated as an overarching concept, referring to the other three content criteria.

The content criteria can, again, be scored on a five-point scale, where 1 means *very incomplete*, *very unreliable*, and *very obsolete*, respectively, and 5 means *very reliable*, *very complete*, and *very timely*.

Content of Source Map Cells

The criteria for the appropriateness of sources and their scores should be put in the source map. To this end, each cell in a source map contains the following information (see also Figure 1).

1. General information about the source, consisting of the name of the source, the data-carrier (human, data or electronic) and (if known) the exact or default location.
2. Scores on the criteria for access, retrieval, content and processing of the (data from the) source.
3. Information about what data could not be delivered if the source was incomplete. This is useful for analyzing the appropriateness of the sources (see next section).
4. Remarks concerning one of the above aspects.

Figure 1. Content of cells in a source map (the shaded areas are not applicable)

Name:
Carrier:
Location:

	Language barrier	Cultural barrier	Institut. barrier	Personal barrier	Geogr. barrier	Techno. barrier	Time barrier	Fee barrier	Clarity barrier	Stability barrier
Access	1…5									
Retrieval										

Content:
Completeness: 1…5 If incomplete: What data could not be delivered?
Timeliness: 1…5
Reliability: 1…5

Process/format: 1…5

Remarks: …

Using the Source Map

A source map allows for two different uses. First, it is used to find appropriate sources for a particular data class. Second, it is used to assess the overall adequacy of the sources. For both types of use, it is necessary to compare the sources. In this section, we discuss how to compare sources and how to use this method for comparison for the two different uses.

Comparing Sources

Sources can be compared using a single criterion (e.g., which source scores highest on completeness?). It is also possible to integrate the values of (several) individual criteria and compare these integrated scores. To integrate these values into an overall score, we propose the following procedure.

1. Define two classes of criteria: efficiency criteria and effectiveness criteria. The class of efficiency criteria consists of the access criteria, the retrieval criteria, and the ease-of-processing criterion. The class of effectiveness criteria consists of the criteria completeness, reliability, and timeliness.
2. Estimate weights for the criteria in the two classes. A possible way of determining the weights of the individual criteria is to have CI professionals produce a rank order of the criteria (in each of the two classes) expressing their ideas about the relative relevance of the criteria. Next, one could discuss the results and produce one rank order for each class. (This procedure could be supported by groupware, such as Group Systems; cf Nunamaker, Dennis, Valacich, Vogel, & George, 1991).
3. Compute, for each source, the overall scores for the two classes. For both classes, we suggest taking the weighted average score for the given criteria. To compute these scores, the scores on the individual criteria should be available. These scores might be obtained initially by asking CI professionals. From that moment on, they should be evaluated every time a source is used and updated when necessary.

Finding Appropriate Sources

The most straightforward use of the map is to find out what sources are available for a particular data class. A step beyond merely enumerating available sources is to give a judgment about the appropriateness of these sources in terms of the criteria presented in the previous section. To this end, we use the efficiency and effectiveness scores of the sources. For a particular data class, all the available sources can be plotted regarding these two scores (see Figure 2).

The figure states that Source 5 scores best on effectiveness, Source 4 best on efficiency, and Source 1 scores lowest on both classes of criteria.

Figures like the above can help in analyzing the appropriateness of a source for a particular data class. As a general heuristic for ranking the sources, we suggest that sources in the upper right quadrant should be preferred to those in the lower right quadrant, and these should be preferred to the ones in the upper left quadrant. Sources in the lower left quadrant should probably be discarded.

Sources that come up as appropriate should be checked for completeness. If they are complete, they can be added to the list with preferred sources. If they are incomplete, it is necessary to find out if there are sources that can compensate for this lack. To this end, information about what data the source is unable to deliver can be used. This information directs the search for an appropriate compensating source.

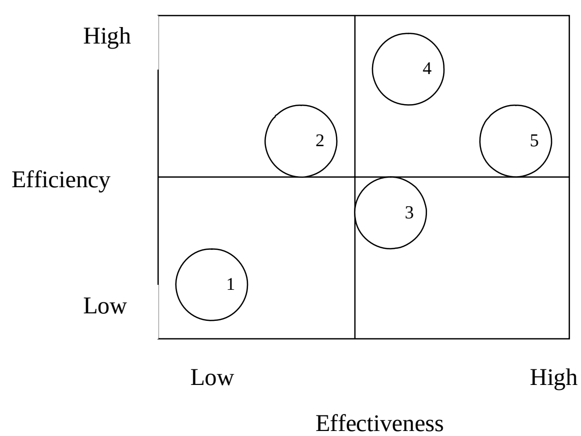
For sources that score high on effectiveness but low on efficiency (lower right quadrant), it should be examined (a) whether the relevance of the data class makes the effort (and costs) worthwhile and/or (b) whether measures can be taken to make the use of the source more efficient, for example, the efficiency of scale in gathering data (a subscription to an often-used online database).

For sources that score in the two left quadrants, it can be established what exactly causes the score. Dependent on the outcome of this investigation, it may be decided to stop using the source.

Assessing the Adequacy of Sources

To judge the overall adequacy of the sources, the map may help in answering the following questions.

Figure 2. Scores of five sources regarding their appropriateness for a particular data class (see text)



1. Do the sources cover all data classes?
2. Do we have adequate sources for the required data classes? If some data classes only have sources that have scores in the lower left quadrant of Figure 2, problems may arise. If a rank order of the data classes regarding their relevance exists, one can also establish whether the most relevant data classes are covered by appropriate sources.
3. Do we have enough different sources for the (most important) data classes? This question refers to the flexibility in collecting data. If a source is suddenly unavailable, one needs to have adequate alternatives. It is also useful to have different sources for the purpose of validating the data.

Answers to these questions help intelligence officers to identify weaknesses in the available sources and direct their efforts to repair them.

Implementing a Source Map

To build, maintain, and use a source map does not require exceptional resources. IT applications for implementing the map range from sophisticated applications to simple solutions. An example of a simple solution is an implementation of the map by means of Microsoft Excel sheets. However, it is also possible to use more sophisticated application, for instance, Web-based applications of the map. Making the map available via an intranet, for instance, can enhance its use and maintenance. In addition to these technological issues, it is important to define and allocate tasks and responsibilities regarding maintenance and use of the map. Finally, data collectors should be motivated to use the map to define their search strategies. In our experience, data collectors see the benefits of a good map and will be inclined to use and maintain it.

FUTURE TRENDS

To aid intelligence officers in their task to evaluate sources, the source map was introduced. Given the increasing need for organizations to collect data about their environment, it can be expected that the need for tools to evaluate sources (like the source map) will also increase. In order to deal with this, information technology tools may be tailored to support the implementation of source maps and the process of keeping them up to date (see, for instance, Philips, 2004).

CONCLUSION

To produce actionable intelligence, the efficient and effective use of sources is imperative. However, up until now, little attention has been paid to supporting the selection of sources. In this paper, we deal with this omission by presenting the source map as a support tool. Properly implemented source maps can be valuable instruments in the support of collection activities. In our view, they can aid in both the everyday use of sources and in the assessment of the overall adequacy of available sources.

REFERENCES

Bernhardt, D. C. (1994). I want it fast, factual, actionable: Tailoring competitive intelligence to executive needs. *Long Range Planning, 27*(1), 12-24.

Chen, H., Chau, M., & Zeng, D. (2002). CI-spider: A tool for competitive intelligence on the Web. *Decision Support Systems, 34*, 1-17.

Cook, M., & Cook, C. (2000). *Competitive intelligence*. London: Kogan Page.

Davenport, T. H., & Prusak, L. (1998). *Working knowledge*. Boston: Harvard Business School Press.

Fleisher, C. G., & Blenkhorn, D. L. (2001). *Managing frontiers in competitive intelligence*. Westport: Quorum, CT.

Fuld, L. M. (1995). *The new competitor intelligence*. New York: Wiley.

Gilad, B., & Gilad, T. (1988). *The business intelligence system*. New York: Amacon.

Herring, J. P. (1999). Key intelligence topics: A process to identify and define intelligence needs. *Competitive Intelligence Review, 10*(2), 4-14.

Kahaner, L. (1997). *Competitive intelligence*. New York: Touchstone.

O'Brien, J. (1998). *Introduction to information systems* (2nd ed.). New York: McGraw-Hill.

Nunamaker, J. F., Dennis, A. R., Valacich, J. S., Vogel, D. R., & George, J. F. (1991). Electronic meetings to support group work. *Communications of the ACM, 34*(7), 40-61.

Philips, E. A. (2004). Building a competitive intelligence system: An infrastructural approach. In D. Vriens (Ed.), *Information and communication technology for competitive intelligence* (pp. 227-247). Hershey, PA: IRM Press.

Sammon, W. L. (1986). Assessing the competition: Business intelligence for strategic management. In J. R. Gardner, R. Rachlin, & H. W. Sweeny (Eds.), *Handbook of strategic planning* (pp. 4.12-4.19). New York: Wiley.

Teo, T. S. H., & Choo, W. Y. (2001). Assessing the impact of using the Internet for competitive intelligence. *Information & Management, 39*, 67-83.

Vriens, D. (2004). *Information and communication technology for competitive intelligence.* Hershey, PA: IRM Press.

Vriens, D., & Philips, E. A. (1999). Business intelligence als informatievoorziening voor de strategievorming. In E. A. Philips & D. Vriens (Eds.), *Business intelligence.* Deventer, Netherlands: Kluwer, 11-44.

KEY TERMS

Collection Stage: Stage of the intelligence cycle. In this stage, sources regarding the required environmental data are located and accessed, and the data are retrieved from them.

Competitive Intelligence: In the literature, two definitions are used: a product definition and a process definition. In the product definition, competitive intelligence is defined as information about the environment, relevant for strategic purposes. The process definition highlights producing and processing this environmental information. Process definitions often refer to the intelligence cycle.

Intelligence Cycle: Cycle of four stages (collections of intelligence activities). The stages are direction (also referred to as planning, in which the strategic information requirements are determined), collection (determining sources and retrieving data), analysis (assessing the strategic relevance of data), and dissemination (of the intelligence to strategic decision makers).

Source: Something or someone containing data and from which the data can be retrieved. Many distinctions regarding sources are given in the competitive intelligence literature, for instance, open versus closed sources, primary versus secondary sources, internal versus external sources, and a distinction referring to the carrier of the data (human, electronic, or paper).

Source Evaluation: The process of assessing the efficiency and effectiveness of a source or several sources, given certain criteria. The result of this process can be (a) a judgment about the usefulness of a particular source for collecting data and/or (b) an insight into the relative usefulness of all available sources. See also "Source map."

Source Identification: Identifying suitable sources (i.e., efficient and containing the relevant data) given a certain data need. See also "Source map."

Source Map: A source map is a matrix linking data classes to sources. In the cells of the matrix, the sources are valued according to different criteria (e.g., accessibility, costs, timeliness of the data, etc.).

Survey of 3D Human Body Representations

Angel D. Sappa
Computer Vision Center, Spain

Niki Aifanti
Informatics & Telematics Institute, Greece

Sotiris Malassiotis
Informatics & Telematics Institute, Greece

Nikos Grammalidis
Informatics & Telematics Institute, Greece

INTRODUCTION

The problem of human body modeling was initially tackled to solve applications related to the film industry or computer games within the computer graphics (CG) community. Since then, several different tools were developed for editing and animating 3D digital body models. Although at the beginning most of those tools were devised within the computer graphics community, nowadays a lot of work proceeds from the computer vision (CV) community. In spite of this overlapped interest, there is a considerable difference between CG and CV human body model (HBM) applications. The first one pursues realistic models of both human body geometry and its associated motion. On the contrary, CV seeks more of an efficient than an accurate model for applications such as intelligent video surveillance, motion analysis, telepresence, 3D video sequence processing, and coding.

Current work is focused on vision-based human body modeling systems. This overview will present some of the techniques proposed in the bibliography, together with their advantages or disadvantages. The outline of this work is as follows. First, geometrical primitives and mathematical formalism, used for 3D model representation, are addressed. Next, a brief description of standards used for coding HBMs is given. Finally, a section with future trends and conclusion is introduced.

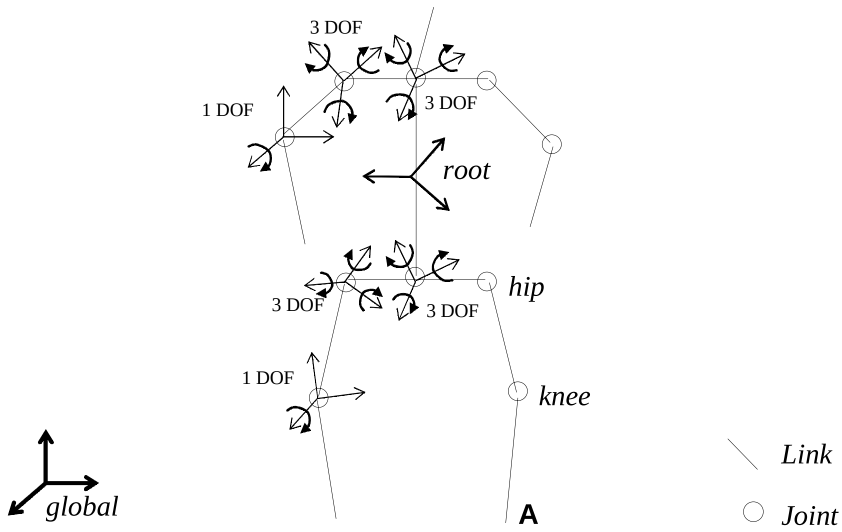
3D HUMAN BODY REPRESENTATIONS

Modeling a human body implies firstly the definition of an articulated 3D structure, in order to represent the human body biomechanical features. Secondly, it involves the choice of an appropriate mathematical model to govern the movements of that articulated structure.

Several 3D articulated representations and mathematical formalisms have been proposed in the literature to model both the structure and movements of a human body (Green & Guan, 2004). Generally, a HBM is represented as a chain of rigid bodies, called *links*, interconnected to one another by *joints*. Links can be represented by means of sticks (Yoo, Nixon & Harris, 2002; Taylor, 2000), polyhedron (Saito & Hoshino, 2001), generalized cylinders (Sidenbladh, Black & Sigal, 2002), or superquadrics (Marzani, Calais & Legrand, 2001). A joint interconnects two links by means of rotational motions about the axes. The number of independent rotation parameters will define the *degrees of freedom* (DOF) associated with a given joint. Figure 1 presents an illustration of an articulated model defined by 12 links (sticks) and 10 joints. Other HBM representations, which do not follow the aforementioned links-and-joints philosophy, have been also proposed in the literature to tackle specific applications. For example, Douros, Dekker and Buxton (1999) present a technique to represent HBMs as single entities by means of smooth surfaces or polygonal meshes. This kind of representation is only useful as a rigid description of the human body. On the contrary, Plänkers and Fua (2003) and Aubel, Boulic, and Thalmann (2000) present a framework that retains an articulated structure represented by sticks, but replaces the simple geometric primitives by soft objects. The result of this soft surface representation is a realistic model where body parts such as chest, abdomen, or biceps muscles are well modeled.

The simplest 3D articulated structure is a stick representation with no associated volume or surface (Liebowitz & Carlsson, 2001). Planar 2D representations, such as cardboard models, have also been widely used (Huang & Huang, 2002). However volumetric representations are

Figure 1. Stick representation of an articulated model defined by 22 DOF



preferred when more realistic models need to be generated. In other words, there is a trade-off between accuracy of representation and complexity. The utilized models should be quite realistic, but they should have a low number of parameters in order to be processed in real-time. Table 1 presents a summary of some of the approaches followed in the literature.

Each of the aforementioned geometrical structures is complemented by means of a motion model that governs its movements (Rohr, 1997); the objective is that the full body performs realistic movements. There is a wide variety of ways to mathematically model articulated systems from a kinematics and dynamics point of view. A mathematical model will include the parameters that describe the links as well as information about the constraints associated with each joint. A model that only includes this information is called a *kinematics model* and describes the possible static states of a system. The state vector of a kinematics model consists of the model state and the model parameters. A system in motion is modeled when the dynamics of the system are modeled as well. A *dynamics model* describes the state evolution of the system over time. In a dynamics model the state vector includes linear and angular velocities as well as position.

After selecting an appropriate model for a particular application, it is necessary to develop a concise mathematical formulation for a general solution to the kinematics and dynamics problems, which are non-linear problems. Different formalisms have been proposed in order to assign local reference frames to the links. The simplest approach is to introduce joint hierarchies formed by independent articulation of one DOF, described in terms of Euler angles. Hence, the body posture is synthesized by concatenating the transformation matrices associated with the joints, starting from the root. In order to illustrate this notation, let us express the coordinates of point **A** in the global reference frame associated with the root of the model (see Figure 1):

$$\mathbf{A}_{global} = Trans_{root\text{-}global} \ x \ Trans_{hip\text{-}root} \ x \ Trans_{knee\text{-}hip} \ x \ \mathbf{A}_{knee}$$

where $\mathbf{A}_{knee}$ represents the coordinates of points **A** relative to the local reference frame placed in the *knee-joint*; $Trans_{i\text{-}j}$ is the corresponding transformation matrix to express reference frame *i* in reference frame *j;* this matrix is defined as:

$$Trans_{i-j} = \left[\begin{array}{c|c} R & T \\ \hline 0 & 1 \end{array}\right] = \begin{bmatrix} C\phi C\theta C\psi - S\phi S\psi & -C\phi C\theta S\psi - S\phi C\psi & C\phi S\theta & t_x \\ S\phi C\theta C\psi + C\phi S\psi & -S\phi C\theta S\psi + C\phi C\psi & S\phi S\theta & t_y \\ -S\theta C\psi & S\theta S\psi & C\theta & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

C and *S* represent the cosine and sine respectively, and (ϕ, θ, ω) are the Euler angles. This kind of matrix concatenation can be used to express every body part in the body global reference frame.

3D HUMAN BODY CODING STANDARDS

In order to animate or interchange HBMs, a standard representation is required. Related standards, such as

Web3D H-anim standards, the MPEG-4 face and body animation, as well as MPEG-4 AFX extensions for humanoid animation, allow compatibility between different HBM processing tools (e.g., HBMs created using an editing tool could be animated using another completely different tool).

The Web3D H-anim working group was formed so that developers could agree on a standard naming convention for human body parts and joints. This group has produced the Humanoid Animation Specification standards, describing a standard way of representing humanoids in VRML. These standards allow humanoids created using authoring tools from one vendor to be animated using tools from another. H-anim humanoids can be animated using keyframing, inverse kinematics, performance animation systems, and other techniques. The three main design goals of H-anim standards are:

- *Compatibility:* Humanoids should be able to display/animate in any VRML compliant browser.
- *Flexibility:* No assumptions are made about the types of applications that will use humanoids.
- *Simplicity:* The specification should contain only what is absolutely necessary.

For this reason, a H-anim file defines a hierarchy of *Joint* nodes, each defining the rotation center of a Joint, which are arranged to form a hierarchy. The most common implementation for a Joint is a VRML Transform node, which is used to define the relationship of each body segment to its immediate parent. Each Joint node can contain other Joint nodes, and may also contain a *Segment* node, which contains information about the 3D geometry, color, and texture of the body part associated with that joint. Each Segment can also have a number of *Site* nodes, which define specific locations relative to the segment. Joint nodes may also contain additional hints for inverse-kinematics systems that wish to control the H-Anim figure.

The hierarchy of H-anim Joint and Segment hierarchy is shown in Figure 2.

Furthermore, the MPEG-4 SNHC (Synthetic and Natural Hybrid Coding) group has standardized two types of streams in order to animate avatars:

- The Face/Body Definition Parameters (FDP/BDP) are avatar specific and based on the H-anim specifications.
- The Face/Body Animation Parameters (FAP/BAP) are used to animate face/body models. More specifically, 168 Body Animation Parameters (BAPs) are defined by MPEG-4 SNHC to describe almost any possible body posture. Thus, a single set of FAPs/BAPs can be used to describe the face/body posture of different avatars. MPEG-4 has also standardized the compressed form of the resulting animation stream using two techniques: DCT based or prediction based. Typical bitrates for these compressed bitstreams are 2 kbps for the case of facial animation or 10 to 30 kbps for the case of body animation.

In addition complex 3D deformations that can result from the movement of specific body parts (e.g., muscle contraction, clothing folds, etc.) can be modeled by using Face/Body Animation Tables (FAT/BATs), which specify sets of vertices that undergo non-rigid motion and a function to describe this motion with respect to the values of specific BAPs/FAPs. However, a significant problem with using such tables is that they are body model-dependent and require a complex modeling stage. In order to solve such problems, MPEG-4 addresses new animation functionalities in the framework of AFX group by including also a generic seamless virtual model definition and bone-based animation. Particularly, the AFX specification describes state-of-the-art components for rendering geometry, textures, volumes and animation. A hierarchy of geometry, modeling, physics and biomechanical

Table 1. Human body structure representations

Authors	***DOF***	***Geometrical Model Representation***
Delamarre and Faugeras (2001)	22	Truncated cones (arms and legs), spheres (neck, joints, and head), and right parallelepipeds (hands, feet, and torso)
Gavrila (1999)	22	Superquadrics
Barron and Kakadiaris (2000)	60	Sticks
Cohen, Medioni and Gu (2001)	32	Generalized cylinders
Ning, Tan, Wang and Hu (2004)	12	Truncated cones (torso, arms, and legs) and a sphere (head)

models are described, along with advanced tools for animating these models (Figure 3).

Specifically, the new Humanoid Animation Framework, defined by MPEG-4 SNHC (Preda, 2002; Preda & Prêteux, 2001) is defined as a biomechanical model in AFX and is based on a rigid skeleton made of bones. The skeleton consists of bones, which are rigid objects that can be transformed (rotated around specific joints), but not deformed. Attached to the skeleton, a skin model is defined, which smoothly follows any skeleton movement.

FUTURE TRENDS AND CONCLUSIONS

Vision-based applications have been growing considerably fast during the last two decades. As a result of that growing, the current technology can tackle—at the moment only under well-defined constraints—complex tasks such as human body modeling. In addition, the knowledge collected during this time from different research areas (e.g., video processing, rigid/articulated object

Figure 2. The H-anim 1.1 Joint and Segment hierarchy (from H-anim Website). Three sets of joints are identified, classified according to their significance, so that H-anim models of different complexity can be produced. Segments are shown with dark grey color and Sites with light grey color. Each object beginning with l_ (left) has a corresponding object beginning with r_ (right). The chart was produced by J. Eric Mason and Veronica Polo, VR Telecom Inc.

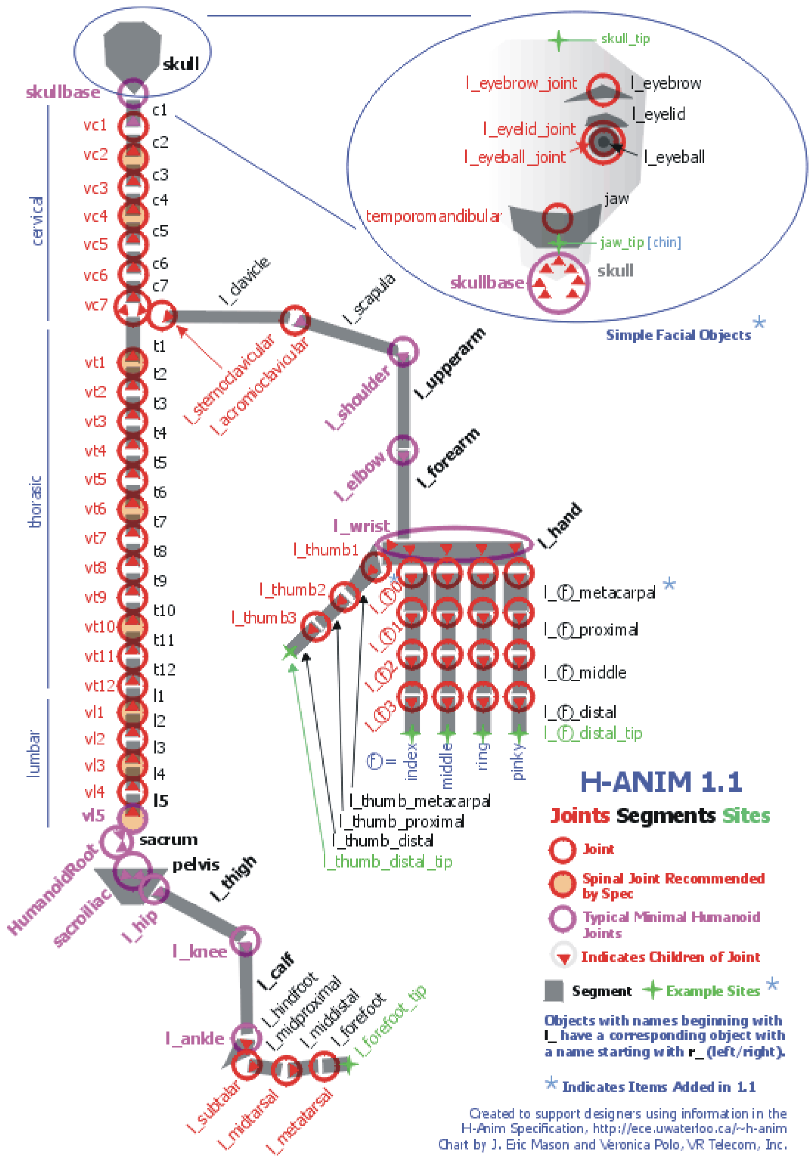
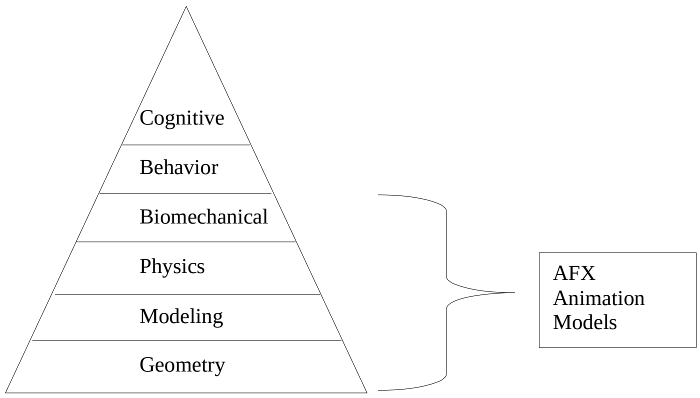


Figure 3. Hierarchy of AFX animation models



modeling, human body/motion models, etc.) also helps to face up to vision-based human body modeling. However, in spite of all this large amount of work, many issues are still open. Problems such as development of models including prior knowledge, modeling of multiple person environments, and real-time performance still need to be efficiently solved.

In addition to the aforementioned issues, the reduction of the processing time is one of the milestones in the non-rigid object modeling field. It is highly dependent on two factors: on one hand the computational complexity, and on the other hand the current technology. Taking into account the past few years' evolution, we can say that computational complexity will not be significantly reduced during the next few years. On the contrary, improvements in the current technology have become commonplace (e.g., reduction in acquisition and processing times, increase in the memory size). Therefore, algorithms that nowadays are computationally prohibitive are expected to have good performance with next technologies. The latter gives rise to a promising future for HBM applications and, as an extension, to non-rigid object modeling in general.

REFERENCES

Aubel, A., Boulic, R. & Thalmann D. (2000). Real-time display of virtual humans: Levels of details and impostors. *IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on 3D Video Technology*, *10*(2), 207-217.

Barron, C. & Kakadiaris, I. (2000). Estimating anthropometry and pose from a single camera. *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition,* Hilton Head Island, SC, USA.

Cohen, I., Medioni, G. & Gu, H. (2001). Inference of 3D human body posture from multiple cameras for vision-based user interface. *Proceedings of the World Multiconference on Systemics, Cybernetics and Informatics,* USA.

Delamarre, Q. & Faugeras, O. (2001). 3D articulated models and multi-view tracking with physical forces. *Special Issue on Modeling People, Computer Vision and Image Understanding*, *81*, 328-357.

Douros, I., Dekker, L. & Buxton, B. (1999). An improved algorithm for reconstruction of the surface of the human body from 3D scanner data using local B-spline patches. *Proceedings of the IEEE International Workshop on Modeling People*, Corfu, Greece.

Gavrila, D.M. (1999). The visual analysis of human movement: A survey. *Computer Vision and Image Understanding, 73*(1), 82-98.

Green, R. & Guan, L. (2004). Quantifying and recognizing human movement patterns from monocular video images—part I: A new framework for modeling human motion. *IEEE Transactions on Circuits and Systems for Video Technology*, *14*(2), 179-190.

Huang, Y. & Huang, T. (2002). Model-based human body tracking. *Proceedings of the 16th International Conference on Pattern Recognition,* Quebec City, Canada.

Liebowitz, D. & Carlsson, S. (2001). Uncalibrated motion capture exploiting articulated structure constraints. *Proceedings of the IEEE International Conference on Computer Vision,* Vancouver, Canada.

Marzani, F., Calais, E. & Legrand, L. (2001). A 3-D marker-free system for the analysis of movement disabilities—an application to the legs. *IEEE Transactions on Information Technology in Biomedicine*, *5*(1), 18-26.

Ning, H., Tan, T., Wang, L. & Hu, W. (2004). Kinematics-based tracking of human walking in monocular video sequences. *Image and Vision Computing, 22,* 429-441.

Plänkers, R. & Fua, P. (2003). Articulated soft objects for multiview shape and motion capture. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, *25*(9), 1182-1188.

Preda, M. (Ed.). (2002). *MPEG-4 Animation Framework eXtension (AFX) VM 9.0.* ISO/IEC JTC1/SC29/WG11 N5245.

Preda, M. & Prêteux, F. (2001). Advanced virtual humanoid animation framework based on the MPEG-4 SNHC Standard. *Proceedings of the Euroimage ICAV 3D 2001 Conference*, Mykonos, Greece.

Rohr, K. (1997). Human movement analysis based on explicit motion models. In M. Shah & R. Jain (Eds.), *Motion-based recognition* (Chapter 8, pp. 171-198). Dordrecht, Boston: Kluwer Academic Publishers.

Saito, H. & Hoshino, J. (2001). A match moving technique for merging CG and human video sequences. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing,* Salt Lake City, USA.

Sidenbladh, H., Black, M.J. & Sigal, L. (2002). Implicit probabilistic models of human motion for synthesis and tracking. *Proceedings of the European Conference on Computer Vision,* Copenhagen, Denmark.

Taylor, C. (2000). Reconstruction of articulated objects from point correspondences in a single uncalibrated image. *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition,* Hilton Head Island, South Carolina, USA.

Yoo, J., Nixon, M. & Harris, C. (2002). Extracting human gait signatures by body segment properties. *Proceedings of the 5th IEEE Southwest Symposium on Image Analysis and Interpretation,* Santa Fe, New Mexico, USA.

KEY TERMS

Articulated Object: Structure composed of two or more rigid bodies interconnected by means of joints. The degrees of freedom associated with each joint define the different structure configurations.

H-Anim: VRML Consortium Charter for Humanoid Animation Working Group. This group has recently produced the International Standard, "Information technology—computer graphics and image processing—humanoid animation (H-anim)," an abstract representation for modeling three-dimensional human figures.

Human Body Modeling: Digital model generally describing the shape and motion of a human body.

MPEG: Moving Picture Experts Group; a group developing standards for coding digital audio and video, as used for example in video CD, DVD, and digital television. This term is often used to refer to media that is stored in the MPEG-1 format.

MPEG-2: A standard formulated by the ISO Motion Pictures Expert Group (MPEG), a subset of ISO Recommendation 13818, meant for transmission of studio-quality audio and video. It covers four levels of video resolution.

MPEG-4: A standard formulated by the ISO Motion Pictures Expert Group (MPEG), originally concerned with similar applications as H.263 (very low bit rate channels, up to 64kbps). Subsequently extended to encompass a large set of multimedia applications, including over the Internet.

MPEG-4 AFX: MPEG-4 extension with the aim to define high-level components and a framework to describe realistic animations and 2D/3D objects.

MPEG-7: A standard formulated by the ISO Motion Pictures Expert Group (MPEG). Unlike MPEG-2 and MPEG-4, which deal with compressing multimedia contents within specific applications, it specifies the structure and features of the compressed multimedia content produced by the different standards, for instance to be used in search engines.

Rotation Matrix: A linear operator rotating a vector in a given space. A rotation matrix has only three degrees of freedom in 3D and one in 2D. It can be parameterized in various ways, usually through Euler angles, yaw-pitch-roll angles, rotation angles around the coordinate axes, and so forth.

Virtual Reality: 3D digital world, simulating the real one, allowing a user to interact with objects as if inside it.

VRML: Virtual Reality Modeling Language, a platform-independent language for virtual reality scene description.

Surveying Mobile Commerce Environments

Jari Veijalainen
University of Jyväskyä, Finland

Mathias Weske
University of Potsdam, Germany

THE EMERGENCE OF MOBILE COMMERCE

During the last five years, the term mobile commerce (m-commerce) has appeared in the vocabulary of business people and researchers. Historically and conceptually, m-commerce can be regarded a new phase in electronic commerce (e-commerce). Although the term was introduced without a clear meaning and it is still lacking a single widely accepted definition, most people would say that the term m-commerce refers to e-commerce activities performed by people while on the move. Thus, m-commerce involves e-commerce transactions where a mobile terminal and a wireless network are used to conduct them. Therefore, m-commerce takes advantage of the e-commerce infrastructure developed for Internet e-commerce. Although in some cases an m-commerce transaction might be an alternative to a regular e-commerce transaction (such as buying a book) performed using a workstation and wired network, in many cases this is not the situation. The limitations of the mobile device - for instance, user interface limitations - are such that it is not attractive to perform typical Internet e-commerce transactions on them. Wireless technologies, combined with so-called 'Internet-enabled' terminals, constitute an ideal platform to realize new types of e-commerce transactions that are not possible or reasonable for wired terminals. The small and light, yet powerful, mobile terminals are almost always carried by their owners, just like wallets or watches. They can indeed also store electronic cash, credit card information, tickets, certificates of the Public Key Infrastructure (PKI), and so forth. Thus, they can assume the role of an e-wallet, as well as function as authentication and authorization devices in various contexts.

The wireless mobile terminals can be *positioned* either by using satellite technologies, terrestrial network facilities, or indoor mechanisms. Recent developments in these technology areas seamlessly extend the positioning of wireless devices into several environments. Thus, services based on the current position of the terminals, and their past or future movement patterns on earth, have become widely possible. These so-called location-based services (LBSs) are a new service class for mobile networks and make new kinds of location-related transactions viable. Thus, LBSs represent a new business area in e-commerce that is typical of m-commerce. For the above reasons, the telecom industry has begun to call portable wireless terminals with the above functionalities Personal Trusted Devices (PTDs).

It is worth noticing that a user can do many things with a PTD that are not m-commerce in the above, rather restricted, sense. Some of these activities can still be closely related with business, such as browsing catalogs, or receiving offers and discount coupons (see Funk, 2004). Those activities that precede or follow the actual m-commerce transactions belong rather to m-business activities (Kalakota & Robinson, 2002). We confine ourselves in the sequel mostly on the m-commerce aspects, although the network infrastructure and many other aspects we are modeling below support m-business and activities that fall outside of both. This is natural, because the wireless infrastructure is indeed common to m-commerce and many other activities.

All in all, there are hundreds of papers written on various aspects of m-commerce so far, and published in journals and conference proceedings. Regarding other approaches to model the entire m-commerce environment, it seems that there are only a few so far. One was presented in Varshney and Vetter (2001) and refined in Varshney and Vetter (2002), but the emphasis was on technical aspects. Another characterization is presented in the preface of Mennecke and Strader (2003) in order to differentiate between general e-commerce and m-commerce. The book (Sadeh, 2002) contains a balanced snapshot of m-commerce applications and business models, but does not present a clear overall model of the m-commerce environment. The E-Factors network in Europe (E-Factors, 2003) has addressed a wide variety of issues in e-commerce and m-commerce, but has not presented a succinct model for the environment. Mobile Internet Technical Architecture (Nokia, 2002) models the entire wireless Internet environment primarily from the network architecture perspective, and m-commerce is an activity supported by certain applications. An application-centric view is also adopted in Varshney (2003). Camponovo and Pigneur (2003) present a rather good overview of the pertinent issues, published literature, and also a description of their view of the m-business environment.

MODELING GLOBAL MOBILE COMMERCE ENVIRONMENT

A model in general is an artificially produced object that reflects the structure and other properties of an object of interest (e.g., a part of reality) in a simplified manner and thus makes it easier to gain knowledge of it or produce the actual object of interest. Creating a model (i.e., *modeling*) is always performed from a certain perspective and with certain goals in mind. These determine what kind of a model is created (physical, conceptual, computer simulation, etc.) and what kind of questions it is expected to answer. The original questions in our modeling attempt were: Are there any entities, structures, or phenomena that are *persistent* in the m-commerce environment? If there are, what could they be and what are the persistent *relationships* between those entities and phenomena? What we have in mind are concepts and relationships with a lifetime of, say, tens of years, instead of a couple of years. Should we be able to identify those, we could write down a framework that could be used in further research and practice for several years to come. This framework should make it possible to distinguish between entities, structures, and phenomena that are persistent and at the same time also inherent in m-commerce, and those that are not. We show below that there are persistent and inherent properties in the environment that are desirable, but others that are not desirable.

Wireless Terminals: Personal Trusted Devices

The first evident thing that is rather persistent and is embedded into the very definition of m-commerce is the terminal. It is indeed immediately clear and trivial as such that m-commerce cannot be performed without a terminal, where terminal is meant in an abstract sense. What forms the concrete terminals adopt in their development over time have of course much influence on the concrete m-commerce transactions that can be performed. But always a terminal is needed in m-commerce, be it a powerful one mounted into a car, be it a personal area network with many wirelessly connected nodes, or a telecom terminal or a PDA as we know them now. As indicated above, the current top-end terminals are called Personal Trusted Devices or PTDs.

Wireless Access Networks, Core Networks, and the Internet

A wireless access network and a larger e-commerce infrastructure including m-commerce servers and authentication servers is needed. Again, the concrete forms these take vary over time, but they are needed in some form.

The main components of the infrastructure are mobile telecom networks - access networks and core networks (Kaaranen et al., 2001) - Internet, and various servers providing the contents and services for e-commerce. The important additional persistent property of the infrastructure is that it consists of pieces that are controlled by different autonomous organizations. There are hundreds of separately controlled 2G telecom networks in the world and roughly the same number of mobile telecom operators that control the networks. Further, there are millions of servers providing for e-commerce services.

Persistent Organizational Structures

Autonomous organizations make autonomous decisions. This leads to heterogeneity of the infrastructure, as concerns the services, protocols, and so forth. A further consequence is that the development speed and stage of the network infrastructure in different parts of the world vary considerably. This situation will prevail because there is no one external force in the world that could stop this development. Only if the organizations themselves decide on a voluntary basis to provide interoperability of their systems can the problems arising from the different stages and development speeds be solved or at least mitigated.

Concerning the regional development, there are currently three large regions in the world, namely the EU, Japan, and the USA, that have reached the m-commerce phase. The rules are more or less homogeneous inside a region, but not necessarily between the regions. Other large regions, like China, might develop their own rules, or they might adopt those of the leading regions. We expect the three most advanced regions to be the dominant ones also in the foreseeable future, although new regions with new rules might emerge. We call such a region *regulatory area*.

Japan and the USA (as well as China, Korea, and Australia) are homogeneous in natural language, whereas Europe is divided into tens of smaller and larger language areas. Thus, although the rules applied would be the same, as is roughly the case in the European Union, the concrete m-commerce services could still be obtainable in different language. Thus there is no guarantee of m-commerce in Europe in a strong sense.

So far in Japan, the mobile telecom operators have taken the initiative in the mobile arena, and government has observed the development. The Japanese approach has been most successful in m-commerce, whereas in the EU, the Commission has established rules for e-commerce including m-commerce, but the development of the latter lags behind the Japanese market. Operators in the U.S. have relied mostly on market actors; the situation is somewhat similar to Europe. The other side of the coin is, however, that the Japanese market is fragmented, because the operators have created "walled gardens," and the competition happens between these. To come to a different market structure, that is, relationships between m-commerce companies and customers, requires different actions from the players than what were taken in Japan. The newest development in the Japanese market can be found in Funk (2004).

Mobility as a Persistent Property

Mobility is inherent in the definition of m-commerce, and it is one of the persistent properties of m-commerce. As is understood now, different mobility patterns have different consequences for the service provisioning. The 2G and 3G access networks support continuous service provisioning for moving customers through handovers within a network controlled by one operator. *Roaming* is also supported, that is, services are offered for customers of other network operators. From an m-commerce point of view, the above regulatory areas and the homogeneous language areas are important while roaming. Thus, if a person travels within a language area, he or she would not encounter language problems while using the m-commerce services. In theory, roaming within a regulatory area should be possible without service access problems, but in practice the geographic areas where certain m-commerce services are currently offered are within individual countries and even in a certain mobile telecom network as indicated above. Roaming from one regulatory area to another means that currently m-commerce services are not at all available, or if they are, they are designed in a local language and thus difficult to use for foreigners. This and problems with lower level protocols in the protocol stack needed in m-commerce can be perceived as an interoperability problem that is unfortunately rather persistent. This is one of the few persistent properties that should be made non-persistent in m-commerce environments, should the market become truly global and accessible for large numbers of roaming customers. Whether this happens depends largely on the economic feasibility of providing services for "overseas" roaming customers, because it is unlikely that authorities would enforce interoperability by rules.

Business Model as a Persistent Entity

M-commerce is a form of business, and is also a persistent and inherent property. In our modeling we go a step further and say that what is persistent, but at the same time varying, is a *business model* (Timmers, 1998). Our idea is that in a concrete m-commerce environment, there is always a business model, but depending on the regulatory area, technology in use, culture, and other factors, it can and will be different. Business models are designed by business actors. We perceive them to be primarily companies.

Continuous Change as a Persistent Phenomenon

Just like in other areas of modern information technology, in m-commerce the only stable property is change. The concrete elements change over time, including base technology, business models, and rules and regulations. New technology that has been deployed, especially in the mobile telecom networks, has made m-commerce possible, and we envision it to be the main driver, at least in the first phase, although business models and emerging needs have quickly begun to form the technology to fulfill needs (e.g., camera and video phones). We tackle this movement from a new technology to an established and deployed one by introducing the concept of *enabling technology*. It is technology that is being developed in laboratories and will perhaps be deployed in the future. Partially it is technology that is necessary in the current situation (e.g., security and privacy protection), partially explorative (wireless transmission technology with very high transmission capacity, measured in Gigabits-per-second, Gbps). It is worth noticing that a concrete enabling technology can exist long before it is actually used in m-commerce (for instance, message encryption), or it can be developed in a short time as a response to the emerging needs of m-commerce, for instance mobile digital rights management.

THE FRAMEWORK IN MORE DETAIL

The basic structure of the m-commerce framework is presented in Figure 1; components of this framework are explained in more detail below. A more elaborate treatment is provided in Veijalainen and Weske (2003).

- Regulatory Frameworks*:* The organizational and technical aspects of laws, standards, and recommendations, as well as the bodies involved in their definition. The main actors here are international

Figure 1. Spheres of concern

<table>
<tr><td rowspan="3">Regulatory Frameworks</td><td>Business Models</td></tr>
<tr><td>Global Infrastructure</td></tr>
<tr><td>Enabling Technologies</td></tr>
</table>

organizations, such as OECD and EU, governments, standardization bodies (ISO, ANSI, ETSI), and interest groups formed by diverse industry sectors to develop standards, such as 3GPP (2003) and Open Mobile Alliance (2003). The persistent entities and their relationships are presented in a more detailed way in Figure 2.

- Business Models*:* Business aspects, including business players, provided services, business protocols, revenue sharing, and code of conduct are important artifacts in this sphere (Timmers, 1998). This sphere captures the pertinent business aspects, persistent concepts and structures. At any moment and in different parts of the world, the business models have a different shape, due to differences in regulatory environment, local (business) culture, economic strength of the individuals, and companies in a particular region. Concrete business models change over time, once new technologies and new business opportunities can be concretely deployed. The persistent entities and their relationships are presented in detail in Figure 3.
- Global Infrastructure: The global infrastructure sphere deals with the global network and the concrete terminals that facilitate m-commerce, as well as the real services. The real global infrastructure is a patchwork of many wireless access networks and backbones, including the Internet. It is and will be heterogeneous at different architectural levels at any point of time. This is because in different parts of the world the development will progress with different pace. The sphere is represented in more detail in Figure 4.
- Enabling Technologies: This sphere includes emerging technologies for user terminals and network technologies, such as new cryptography, privacy-protecting technologies, positioning with high accuracy, and new batteries and other energy sources for the terminals. It also includes standardization and business interest groups and other organizations developing these emerging technologies. Typical examples are 4G (and beyond) technologies, which are currently being researched and which will move later into standardization. The sphere is represented in more detail in Figure 5.

The division of the m-commerce environment into four spheres of concerns, as we call them, is motivated by the observation that these spheres seem to be very persistent and their life span is much longer than that of any of the concrete ingredients. The reasons are basically simple. Regulatory frameworks address in our model the organizational aspects that are necessary to regulate the business environment, the deployed networks, and the development of new technologies. Enabling technologies are in a way a basis for everything, because their change is a necessary condition for the deployment of new technologies in the global infrastructure and business models. Should the development of new technologies stop, this would mean stagnation in the upper spheres in Figure 1.

Global infrastructure is the totality of the terminals, access networks, backbones, servers, and other components that are deployed at a certain moment. It also contains the technical support for concrete m-commerce transactions, that is, implementation support for business models in place. Business models also address many and mainly other aspects than the technical level, including the business partners, money and contents flow, and so forth. They live their own life within organizations and as interaction patterns between organizations.

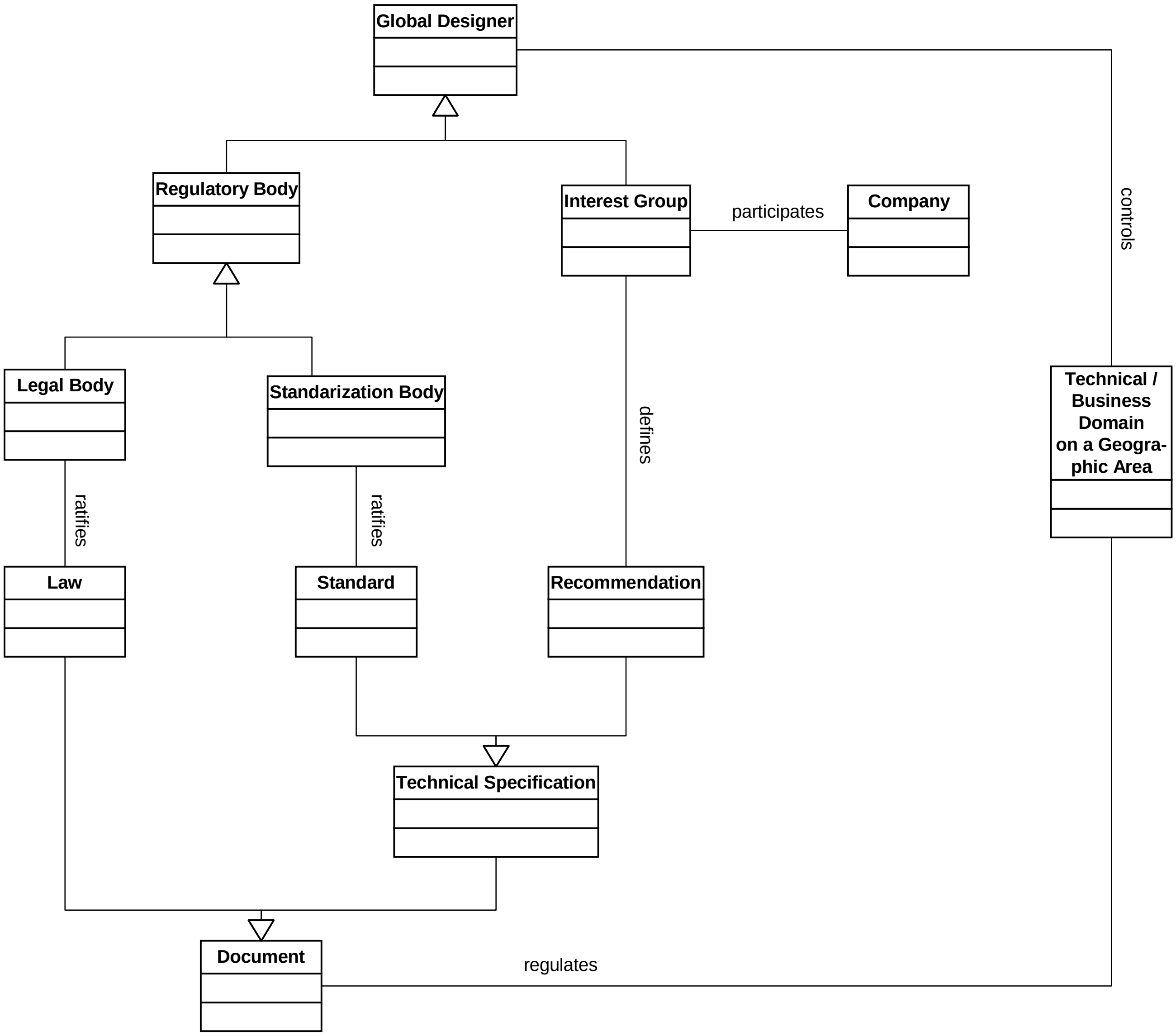
The identified entities and the persistent relationships between these are presented in the diagrams below. They were obtained from Veijalainen and Weske (2003) and explained more thoroughly there.

CONCLUSION

We have attempted to describe the most common persistent properties of the m-commerce environment. Some of them are non-desirable and require concerted action if one wants to avoid them. Others are inherent properties of the very concept of m-commerce, as we have defined it in this context. The main contribution of the framework is perhaps the separation of the four rather persistent spheres of concern that have their own, relatively independent rules of development, but which are still tightly coupled with the other spheres. We also identify some entities within the spheres that are more persistent than the others. This does not mean, however, that these would be the only important entities for understanding current m-commerce.

The interesting questions are related to the relationships between the spheres and their development. For instance, which regulations inhibit the business model, infrastructure, and enabling technology development?

Figure 2. Regulatory frameworks (Figure 3 from Veijalainen & Weske, 2003)



S

Figure 3. Enabling technologies (Figure 4 in Veijalainen & Weske, 2003)

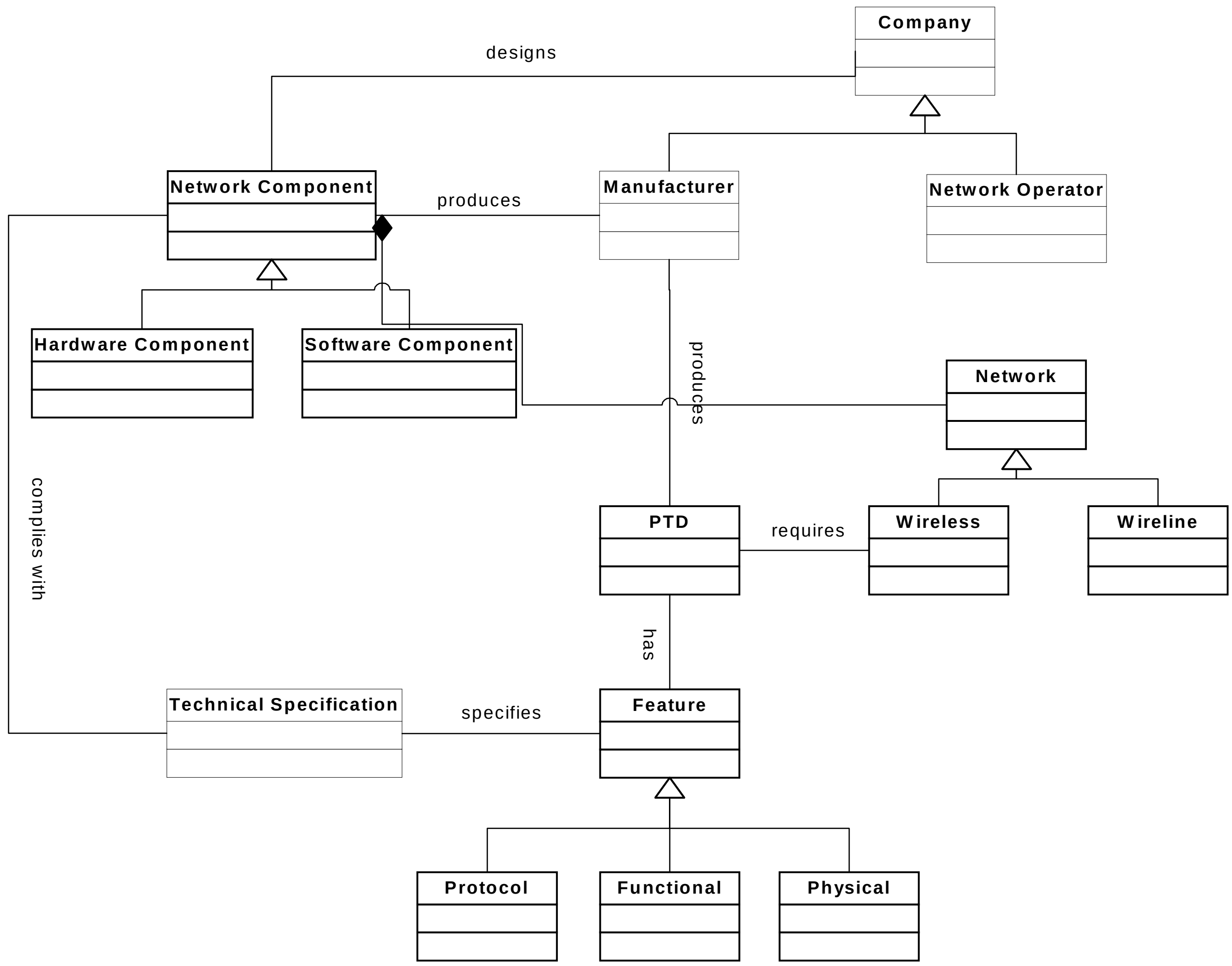


Figure 4. Global infrastructure (Figure 5 in Veijalainen & Weske, 2003)

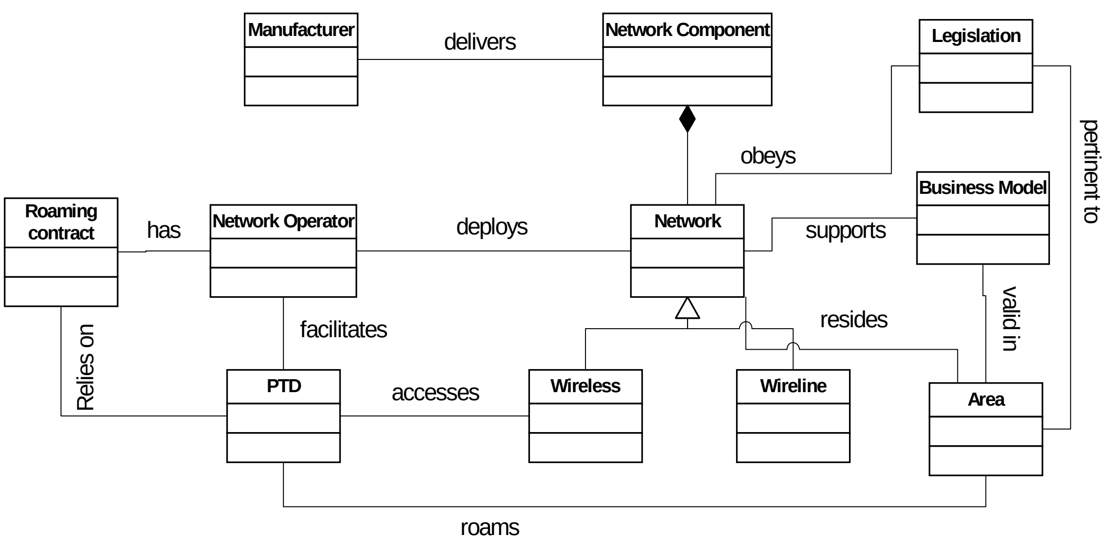


Figure 5. Business models (Figure 6 in Veijalainen & Weske, 2003)

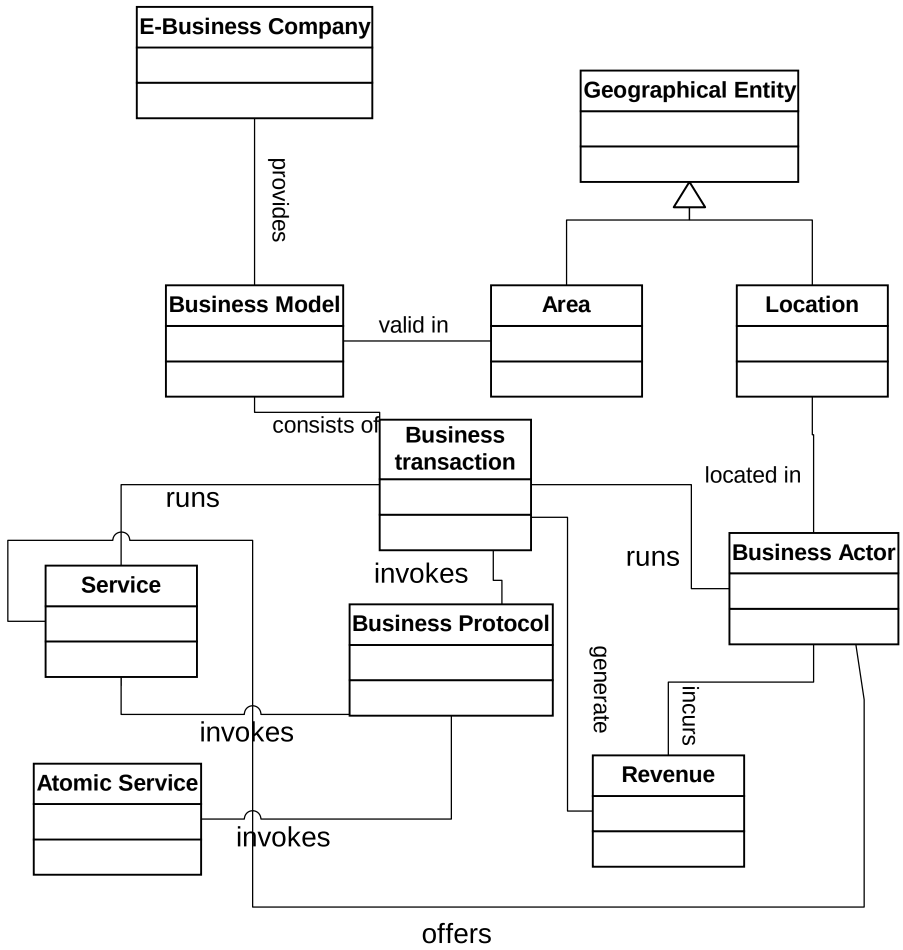
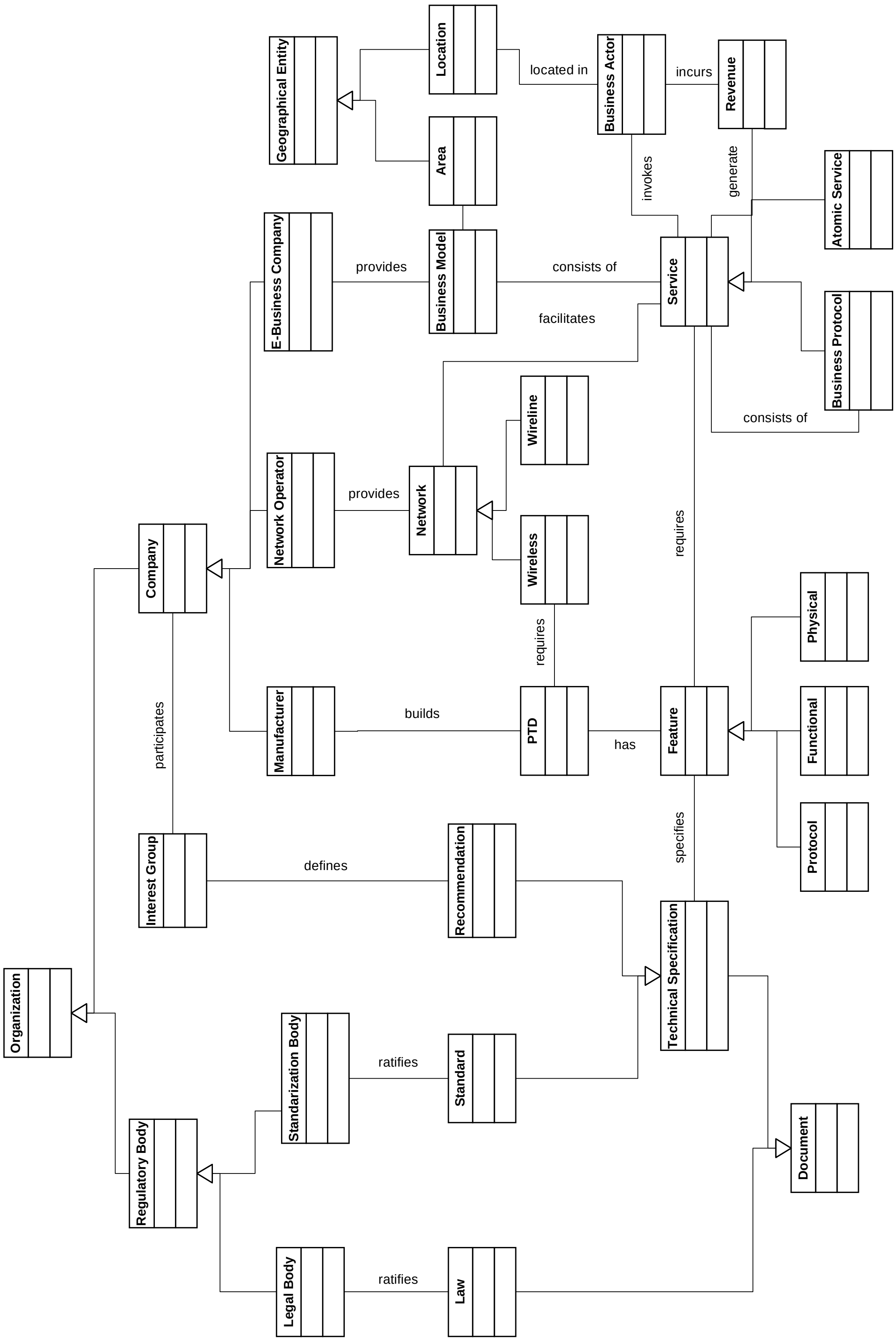


Figure 6. Overall framework of m-commerce environment



Which would foster it? We have discussed privacy in Veijalainen and Weske (2003) as a concept that penetrates the spheres, and treated privacy protection in Veijalainen et al. (2003) as a requirement that addresses different issues in different spheres. Further, what are the concrete interdependencies between the companies and other actors within the regulatory frameworks sphere? What are such concrete business models that foster or inhibit the development of new enabling technologies? What are the economic parameters of the different spheres: how much does the infrastructure cost, how much the development of the new enabling technologies, and how much should the m-commerce yield in economic terms? An important question is: which are such entities and issues within various spheres that are shaped under competition? Which issues should be agreed upon globally or regionally among business actors and regulators?

REFERENCES

3rd Generation Partnership Project (3GPP). Retrieved November 1, 2003, from www.3gpp.org.

Campnovo, G. & Pigneur, Y. (2003). Analyzing the m-business landscape. *Annals of Telecommunications, 58*(1-2).

E-Factors. *European network centering on e-commerce issues.* Retrieved November 1, 2003, from www.e-factors.net.

Funk, J. (2004). *Mobile disruption; The technologies and applications driving the mobile Internet.* New York: John Wiley & Sons.

Kaaranen et al. (2001). *UMTS networks; Architecture, mobility and services.* London: John Wiley & Sons.

Kalakota, R. & Robinson, M. (2002). *M-business, the race to mobility.* McGraw-Hill.

Mennecke, B.E. & Strader, T.J. (Eds.). (2003). *Mobile commerce: Technology, theory and applications.* Hershey, PA: Idea Group Publishing.

MeT. (2002). MeT terminology specification; version 2.2, 14-11-2002. *Mobile Electronic Transaction Forum Ltd.* Retrieved November 7, 2003, from *www.mobiletransaction.org/documents.html*

Nokia. (2002). *Mobile Internet technical architecture.* IT Press.

OECD, measuring the information economy. Annex 4 The OECD definitions of Internet and e-commerce transactions. Retrieved October 20, 2003, from *www.oecd.org*

Open Mobile Alliance. Retrieved November 1, 2003, from www.openmobilealliance.org.

Sadeh, N. (2002). *M-commerce; technologies, services, and business models.* Norman Sadeh/John Wiley & Sons.

Timmers, P. (1998). Business models for electronic markets. *Electronic Markets, 8*(2), 2-8.

Varshney, U. & Vetter, R. (2001). *A framework for emerging mobile commerce applications.* In R. Sprague (Ed.), *Proceedings of the 34th Hawaii International Conference on System Sciences* (IEEE 2001). Retrieved from *www.hicss.org*

Varshney, U. & Vetter, R. (2002). Mobile commerce: Framework, applications and networking support. *ACM/Kluwer Mobile Networks and Applications (MONET), 7*(3), 185-198.

Varshney, U. (2003). Location management for mobile commerce applications in wireless Internet environment. *ACM Transactions on Internet Technology, 3*(3), 236-255.

Veijalainen, J. & Weske, M. (2003). Modeling static aspects of mobile electronic commerce environments. In L.E. Peng & K. Siau (Eds.), *Advances in mobile commerce technologies* (Chapter 7, pp. 137-170). Hershey, PA: Idea Group Publishing.

Veijalainen, J., Yamakawa, P., Markkula, J., Tsalgatidou, A., Hadjiefthymiades, S. & Matsumoto, M. (2003, June 23-24). On requirements for mobile commerce. In G.M. Giaglis, H. Werthner, V. Tschammer & K.A. Froeschl (Eds.), *Proceedings of the 2nd International Conference on Mobile Business* (M-Business 2003) (pp. 453-464), Vienna, Austria. Vienna: Austrian Computer Society.

KEY TERMS

Business Model: A sphere of concern; addresses business aspects, as perceived by companies, including business players, provided services, business protocols, revenue sharing, and code of conduct are important artifacts in this sphere of concern (Timmers, 1998). A concrete business model is facilitated by (a part of) the global infrastructure (that is, the concretely deployed technology), and simultaneously restricted and enabled by the legislation in force.

Electronic Commerce Transaction (broad definition): An (e-commerce) electronic transaction is the sale or purchase of goods or services, whether between businesses, households, individuals, governments, and other public or private organizations, conducted over computer-mediated networks. The goods and services are ordered over those networks, but the payment and the ultimate delivery of the good or service may be conducted online or off. Includes orders received or placed on any online application used in automated transactions such as Internet applications, EDI, Minitel, or interactive telephone systems (OECD, 2002).

Enabling Technologies: A sphere of concern; consists of those technologies that are under development at a certain moment and that have potential to become part of the global infrastructure in the future. Currently, new 3G terminals and 4G network technologies are typical examples.

Global Infrastructure: A sphere of concern; consists of the wireless access and backbone networks, as well as servers and terminals that are necessary for e-commerce in general and m-commerce in particular at a certain moment. Global infrastructure is shaped by the organizational entities within the regulatory framework sphere, and controlled and operated by companies and governments. Global infrastructure consists of heterogeneous components (WLANs, 2G and 3G telecom network, wireline telecom networks, Internet, proprietary IP networks, and a great variety of wireless terminals) and evolves over time as new technologies are adopted from the enabling technologies sphere, where they are developed.

Mobile Electronic Commerce (M-Commerce): Activity that consists of launching and controlling mobile electronic commerce transactions. Note: m-commerce also includes pure machine-to-machine e-commerce transactions, although human involvement is assumed in most cases at the mobile terminal.

Mobile Electronic Commerce Transaction: An e-commerce transaction that is conducted using a mobile terminal (e.g., a PTD) and a wireless network or link.

Personal Trusted Device (PTD): A mobile communications device that can be personalized (through the registration of service certificates), and trusted by the user and service providers requiring application-level public key security (MeT, 2002).

Regulatory Frameworks: A sphere of concern; organizational level that is required to control the interoperable technical domain or legislation pertinent in m-commerce. Most important current concrete entities within this sphere of concern are governments and intergovernmental organizations (EU, OECD, UN, etc.), industry associations (OMA, W3C, etc.), and proper standardization bodies (ISO, ETSI, ANSI, etc.).

Sphere of Concern: A set of issues and entities relevant for a particular modeling process, in this case modeling of m-commerce (c.f., Universe of Discourse (UoD) in conceptual modeling).

System Development for E-Business

Theresa A. Steinbach
DePaul University, USA

Linda V. Knight
DePaul University, USA

INTRODUCTION

This article first introduces both the technical and business environments in which Web information systems are developed and the unique project variables of these systems. Characteristics are identified where e-business projects differ from traditional information systems projects. The next section presents various system development methodologies, provides a categorization framework and analyzes these categories for situations where each is best utilized. The synthesis of applying the environmental and project variables to the system development methodology categories provides the main contribution of the article. Conclusions are drawn that a project is unlikely to align perfectly with any one methodology. The project manager should select the methodology that is the "best fit" from the model presented here. Finally, the project manager should customize the process based on the unique characteristics of the organization, project, and team.

BACKGROUND

The speed demanded of e-business developers and the ever-fluctuating nature of the technical and business environments in which they must function combine to demand new approaches to the system development process. Traditional methodologies were designed for vastly different environments and cannot be easily transplanted into this new Web setting. In the practitioner press, Yourdon has called for a new "light" version of traditional methodologies (2000) for use in e-business development. Avison and Fitzgerald (2003) have found many Web-based applications being developed in an ad-hoc, trial-and-error manner.

The E-Business Environment

E-business projects are developed in a unique environment that is characterized by "perpetual ambiguity and rapid change" (Nadherny and Stuart, 2000). A Web Information System (WIS) is usually tightly integrated with other non-WISs, such as databases and transaction processing systems (Isakowitz and Bieber, 1998) to support the work of the organization. "Internet time," the perception that product development and consumer acceptance now occur in a fraction of the time that they traditionally took (Odlyzko, 2001), has increased time-to-market pressure in WIS projects. These distinctive characteristics set e-business projects apart from other types of information systems (IS) projects.

(1) Web Technical Environment

The technical environment for a Web Information System is vastly expanded compared to traditional IS projects. In most traditional projects, an organization builds their information systems in a closed setting, limiting external organizations' access based upon the controlling organization's technology choices. Web Information Systems are different in that both external and internal users often access WISs using a variety of different hardware, software, and networking technologies. To further complicate matters, compatibility issues arise when new WISs must interface with existing legacy systems.

(2) Competitive Marketplace

The competitive marketplace plays a far greater role in Web Information Systems than it does in most traditional projects. WIS marketplace impact can be direct and immediate. This situation offers the promise that an organization could quickly gain competitive advantage through its Web Information Systems. However, according to Porter (2001), the five underlying forces of competition: the intensity of rivalry among existing competitors, the threat of substitute products or services, the barriers to entry for new competitors, the bargaining power of suppliers, and the bargaining power of buyers, have been negatively affected by the shift away from quality, features, and service toward price. This has resulted in a dampening of overall profitability, and has reduced the ability of any company to establish an operational advantage that can be sustained. Thus the e-business competitive environment is characterized by the ability of any competitor to rapidly change marketplace dynamics, coupled with the likelihood that any such change will be quickly be supplanted by another competitor's actions.

(3) Organization's Culture

Each organization brings to every project its own culture. An organization's culture is comprised of values, behaviors, and attitudes (Hatch and Schulz, 2001). It provides continuity, structure, common meaning, and order, giving rise to stable patterns of interaction within the organization. However, that stability can be upset by the creative, more liberal culture characteristics of many Web development efforts. Developers who challenge the advisability of traditional project phases and approvals also challenge the established culture and order.

(4) Organizational Strategy

An organization's Web strategy should be viewed as a continuous cycle that builds on the current strategy of the organization while creating new business models (Venkatraman, 2000). Yet, for some organizations, e-business brings with it a change in strategic emphasis from internal operations to customers. In the world of Web Information Systems, branding has become increasingly important. A brand is the relationship between a company and its customers that is based on vision, culture, and image (Hatch and Schultz, 2001). Personalization of the Web interface and creation of online communities can extend the traditional brand relationship and increase customer loyalty and profitability. The customer, viewing their experience with the organization, should see only a seamless integration of the organization's overall strategy and their Web initiatives.

(5) Organization's Technical Environment

Web Information Systems complicate the organization's technical environment. In the 1970s mainframe era, most organizations had single-site implementations, where all hardware and software came from one manufacturer, and there were designed to work together. The introduction of client/server technologies in the early 1990s replaced this single source with the challenge of integrating hardware and software from multiple vendors. Web Information Systems operate in a distributed world with a multitude of different manufacturers' components, many of which were not engineered to work together. Organizations have the option of retaining this technology in-house or outsourcing, or a combination of both.

(6) Project Characteristics

Web Information Systems concentrate more heavily on Information Architecture than the average traditional IS

Table 1. E-business development methodology variables (Knight et al., 2003)

Organization			
Culture:	Conservative	→	Innovative
Strategy:	Committed	→	Evolving
Technology:	Stable	→	Experimental
	Non-integrated	→	Integrated
Project			
Objectives:	Clear	→	Unclear
Requirements:	Stable	→	Changing
Users:	Known	→	Unknown
Implementation:	Long	→	Rapid
Team			
Skills:	Technical	→	Creative
Composition:	Stable	→	Changing
Member Experience:	Less Experience	→	Highly Experienced
Leadership:	Less Experience	→	Highly Experienced

project. In a Web Information System context, Information Architecture refers to the design of organization, labeling, navigation, and searching systems to help people find and manage information more successfully (Rosenfeld, 2000). WIS Information Architecture concentrates heavily on the human-computer interface.

A project budget for traditional information systems is often based upon data gathered through historical experience. Web Information Systems generally do not have this kind of past experience to draw upon for estimating. Further, the question of which department should sustain the cost of a project becomes more complex for many WISs, where project ownership issues are often less clear.

Whereas traditional IS projects are largely limited to the IT department, with some involvement of primary users, the development of Web Information Systems requires teams that include a broad range of technical, user, and creative talent. Many WIS teams include marketing staff, business analysts, graphical artists, human-computer interaction specialists, and branding specialists.

System Development Methodology

System development methodology refers here to the framework that is used to structure, plan, and control the process of developing an information system. Most models can be placed into one of four major categories: linear, iterative, parallel, or agile.

(1) Linear Models

While Boehm (1986) tracks system development methodology to Bennington's 1956 Stagewise model and Royce's 1970 Waterfall model, other authors suggest that the System Development Life Cycle (SDLC) evolved primarily during the 1970s (Harrison, 1985), in response to organizations' needs to better organize, plan, schedule, and control projects (Hall, 1980). The SDLC is based upon two principles: dividing projects into phases, and using written documentation and approvals to maintain control. While the exact phases in the cycle vary from one author or organization to the next, they generally follow along these lines: initial investigation, conception or feasibility study; requirements definition; system design; coding, unit testing, integration and system testing; implementation; and ongoing system maintenance. The Waterfall model, which allows some overlap and splashback between the phases, is a variation of the SDLC that loosens this type of control slightly. Another variation on the SDLC is the V Model (Plogert, 1996), which emphasizes system quality assurance. In this model, the stages are partnered in a sort of V shape, with, for example, system design on the left leg of the V paired with system testing on the right leg, and requirements definition on the left leg paired with acceptance testing on the right leg.

Table 2. Situations where linear development models are most beneficial

- Clear project objectives
- Stable project requirements
- Knowledgeable user
- No immediate need to install
- Inexperienced team members
- Fluctuating team composition
- Less experienced project leader
- Need to conserve resources
- Strict requirement for approvals

(2) Iterative Models

Prototyping is an iterative process that lets users work with a small-scale mock-up of their system, experience how it might function in production, and request changes until it meets their requirements. As Janson and Smith (1985) noted, "Prototyping addresses the inability of many users to specify their information needs, and the difficulty of systems analysts to understand the user's environment, by providing the user with a tentative system for experimental purposes at the earliest possible time." Maturity of the prototype methodology led to a further type of iterative development model, Rapid Application Development (RAD). RAD methodologies aim at producing high quality systems quickly, primarily through the use of iterative prototyping, active user involvement, and computerized development tools.

The Spiral model developed originally by Boehm (1986) is an iterative approach that is focused on minimizing project risk. Each trip around the Spiral traverses four basic quadrants: (1) determine objectives, alternatives, and constraints of this iteration; (2) evaluate alternatives; identify and resolve risks; (3) develop and verify deliverables from this iteration; and (4) plan the next iteration (Boehm, 1986; 1988). Boehm's more recent work reflects an expanded view of the Spiral, requiring identification of stakeholders and their win conditions at the very start of each cycle, and ending the cycle with review and commitment. This broader view of the Spiral merges elements of Boehm and Ross's Theory W into the Spiral Model (Boehm, 2000). Under Theory W, the primary job of the software project manager is to insure that all of the project's stakeholders become winners (Boehm and Ross, 1989).

Table 3. Situations where prototyping/RAD is most beneficial

- Project objectives are unclear
- Functional requirements are changing
- User is not fully knowledgeable
- Immediate need to install something
- Experienced team members (particularly if the prototype is not throw-away)
- Stable team composition
- Experienced project leader
- No need to absolutely minimize resource consumption
- No strict policy or cultural bias favoring approvals
- Analysts/users appreciate business problems involved, before they begin project
- Innovative, flexible designs that will accommodate future changes are not critical

(3) Parallel Models

Parallel models resemble those used by some advertising agencies when they are under competitive pressure to produce the best possible presentation for their client in the least possible time. The advertising agency's "three creative comps" becomes three creative treatments of a Web site. With parallel models, different approaches are tried at the same time by different individuals or teams, and then as the project progresses, less productive paths are pruned. Alternative paths may be followed simultaneously just at the project start, or new alternative paths may be introduced when the project reaches critical points. An alternative path model creates multiple paths and prunes as the project progresses.

(4) Agile Development Models

During the late 1990s and early 2000s, a variety of new methodologies were put forth, most aimed at creating a lighter, faster, more flexible and responsive approach to development. Beck's Extreme Programming (1999) user requirements are noted on index cards and programmers work in pairs, with two developers sharing one computer workstation. Bad code is weeded out regularly through a process called refractoring. Initial studies have indicated that pair coding results in less initial productivity, but that pairs do write code with fewer defects (Radding, 2001; Williams et al., 2000). Highsmith's Adaptive Software Development (1999) emphasizes developing a flexible team that can function in a complex, rapidly changing environment. Adaptive development grew out of Highsmith's attempt to apply Rapid Application Development techniques to larger, more complex projects. His adaptive development cycle consists of three major phases or stages: (1) speculate (project mission statement, initial requirements and cycle plan); (2) collaborate (short cycles where system components are developed concurrently and delivered); and (3) learn (review work and team's performance.)

APPLYING THE MODEL TO SPECIFIC PROJECTS

Table 7 is an enhancement of *Table 1*, showing the relationships between various development methodologies and organizational, project, and team variables. These were culled from the background section.

When an organization applies the model in *Table 7* to a particular project, considering all their organization, project, and team variables, the model is unlikely to align

Table 4. Situations where the spiral model is most beneficial

- Risk avoidance is a high priority
- No need to absolutely minimize resource consumption
- Project manager is highly skilled and experienced
- Policies or cultural bias favor approvals
- Project might benefit from a mix of other development methodologies
- Organization and team culture appreciate precision and controls
- Delivery date takes precedence over functionality, which can be added in later versions

Table 5. Situations where parallel models are most beneficial

- Rapid installation is a primary goal
- Solid, experienced team
- Strong project management
- Excellent project-related communications
- Stable team composition
- Experienced project leader
- Little pressure to conserve resources
- Uniquely flexible development team

perfectly with any one methodology. At this point, the project leader may select a methodology model that is a close fit, cognizant of that model's limitations when applied to his or her project. For example, an innovative Web Information System might line up well with a agile method except for the lack of experience of the project team. By using the model, the project leader is able to identify this discrepancy and make plans in advance to compensate. Such plans may include team-building exercises, mentoring, weekly status meetings, and skill-building training, for example. Alternatively, when there is not a single methodology that is clearly the best fit, the project leader may elect to create his or her own "best practice" by combining various aspects of those models that most effectively address the organization, project, and team variables.

FUTURE TRENDS

Various development methodologies have risen to the forefront at different times in history. Royce proposed his linear model in the 1970s. Iterative development was widely discussed in the 1980s. Parallel models were popular in the early 1990s. Agile development emerged in the late 1990s. Currently, various agile methods are in vogue. As each new methodology has emerged, its proponents have espoused its virtues. Over time, each methodolody's limitations have become known as well. The model presented here recognizes both the advantages and the limitations of each methodology. As new methodologies continue to emerge, the model provides a way for project managers to incorporate these emerging methodologies in a way that is appropriate for their organization, team, and project.

CONCLUSION

The rapidly changing business and technical environment that characterizes Web-enabled e-business, coupled with the unique nature of Web development teams themselves, combine to demand a new approach to system development methodology. No existing methodology is ideally suited for all e-business development endeavors. However, existing methodologies do provide a storehouse of options from which project managers can and should select and customize methodologies best suited to their organizations, their projects, their teams, and the unique nature of Web-enabled e-business. When a project most closely aligns with a single methodology, the project leader can compensate in the management of those few variables for which the chosen methodology is not a good fit. If project variables do not clearly indicate a preferred methodology, the project leader must employ those tools and techniques from each methodology that address the "best practice" of each variable. The model proposed here for Web-based systems could be applied to all types of information systems.

REFERENCES

Avison, D.E., & Fitzgerald, G. (2003, January). Where now for development methodology? *Communications of the ACM*, 46(1), 79-82.

Beck, K. (1999, October). *Extreme Programming Explained: Embrace Change (The XP Series)*. Upper Saddle River, NJ: Addison-Wesley Publication.

Table 6. Situations where agile development models are most beneficial

- Rapid installation of the bulk of the system is not a critical goal
- Users have ability to make rapid, binding decisions
- Users are flexible and willing to work through many small implementations
- Collaborative team atmosphere
- Team with substantial system design experience
- Experienced project leader
- Minimal pressure to conserve resources with some models

Table 7. Model for evaluating e-business development methodology (Knight et al., 2003)

<u>*Organization*</u>			
Culture:	Conservative (Linear, Spiral)	→	Innovative (Parallel, Agile)
Strategy:	Committed (Linear)	→	Evolving (Prototyping, RAD, Agile)
Technology:	Stable (Linear, Spiral)	→	Experimental (Parallel, Agile)
	Non-integrated (Linear, Spiral)	→	Integrated (Parallel, Agile)
<u>*Project*</u>			
Objectives:	Clear (Linear, Spiral)	→	Unclear (Prototyping)
Requirements:	Stable (Linear, Spiral)	→	Changing system fundamentals (Parallel, Agile)
Users:	Known (Linear)	→	Unknown (Spiral, Prototyping, RAD)
Implementation:	Long (Linear, Spiral)	→	Rapid (Prototyping, RAD, Parallel, Agile)
Budget/schedule risk: *(risk of not meeting)*	Low (Disruptive)	→	High (Linear, Spiral)
Requirements risk: *(risk of not identifying)*	Low (Linear, Spiral)	→	High (Prototyping, RAD, Parallel, Agile)
<u>*Team*</u>			
Skills:	Technical (Linear, Spiral, Prototyping)	→	Creative (Parallel, Agile)
Composition:	Stable (Spiral, Prototyping, RAD, Parallel, Agile)	→	Changing (Linear)
Member Experience:	Less Experience (Linear)	→	Highly Experienced (Spiral, Agile)
Leadership:	Less Experience (Linear)	→	Highly Experienced (Spiral, Prototyping, RAD, Parallel, Agile)

Boehm. B. (1986, August). A Spiral Model of Software Development and Enhancement. *ACM SigSoft Software Engineering Notes*, 11(4), 21-42.

Boehm. B. (2000, July). Spiral Development: Experience, Principles, and Refinements. *Spiral Development Workshop February 9, 2000, Special Report CMU/SEI-2000-SR-008.*

Boehm, B. (1988, May). A Spiral Model of Software Development and Enhancement. *Computer*, 61-72.

Boehm. B., & Ross, R. (1989, July). Theory-W Software project management: principles and examples. *IEEE Transactions on Software Engineering*, 15(7), 902-915.

Hall, T.P. (1980, April). Systems Life Cycle Model. *Journal of Systems Management*, 31(4), 29-31.

Harrison, R. (1985, August). Prototyping and the Systems Development Life Cycle. *Journal of Systems Management*, 36(8), 22-25.

Hatch, M.J., and Schultz, M. (2001, February). Are the strategic stars aligned for your corporate brand? *Harvard Business Review*, 79(2), 128-134.

Highsmith, J.A. III (1999). *Adaptive Software Development; A Collaborative Approach to Managing Complex Systems*. New York: Dorset House.

Isakowitz, T., & Bieber, M.V. (1998, July). Web Information Systems. *Communications of the ACM*, 41(7), 78-80.

Janson, M.A., & Smith, L.D. (1985, December). Prototyping for Systems development: A Critical Appraisal. *MIS Quarterly*, 9(4), 305-315.

Knight, L.V., Steinbach, T.A., & Kellen, V. (2003). System development methodologies for Web-Enabled E-Business: A Customization Framework. In N. Shi & V.K. Murthy (Eds.), *Architectural Issues of Web-Enabled Electronic Business* (pp. 213-226). Hershey, PA: Idea Group Publishing.

Nadherny, C., & Stuart, S. (2000, February). The new e-leaders. *Chief Executive*, 59(152), 59.

Odlyzko, A. (2001, April). The myth of "Internet time". *Technology Review*, 104(3), 92-93.

Plogert, K. (1996, December 31). The tailoring process in the German V-Model, *Journal of Systems Architecture*, 42(8), 601-609

Porter, M. (2001, March). Strategy and the Internet. *Harvard Business Review*, 79(3), 62-78.

Radding, A. (2001, April 2). Simplicity, but with control. *Informationweek*, 831, 71-74.

Rosenfeld, L. (2000, April). Making the Case for Information Architecture. *Presentation, ASIS Summit 2000 Information Architecture*. Retrieved from the World Wide Web at: *http://www.asis.org/Conferences/Summit2000/rosenfeld/index.htm*

Venkatraman, N. (2000, Spring). Five steps to a dot.com strategy: How to find your footing on the Web. *Sloan Management Review*, 41(3), 5-28.

Williams, L., Kessler, R.R., Cunningham, W., & Jeffries, R. (2000, July/August). Strengthening the Case for Pair-Programming. Retrieved from the World Wide Web at: *http://collaboration.csc.ncsu.edu/laurie/Papers/ieeeSoftware.PDF*

Yourdon, E. (2000, September 18). The 'light' touch. *Computerworld*. Retrieved September 15, 2001 from the World Wide Web at: *http://www.computerworld.com/cwi/story/0,1199,NAV47_STO50363,00.html*

KEY TERMS

E-Business Environment: The business environment that is characterized by rapid time-to-market pressures, heterogeneous technical access, and quick IT strategic response requirements.

Extreme Programming: An agile system development methodology that decomposes very large projects into many small projects with minimal requirements and taking just a few weeks to complete, characterized by pair programming and refractoring.

Internet Time: The perception that product development and consumer acceptance now occur in a fraction of the time that they traditionally took (Odlyzko, 2001).

Prototyping: A tool used in system development methodology that lets users iteratively work with a small-scale mock-up of an information system, experience how it might function in production, and request changes until it meets their requirements.

Rapid Application Development: A system development methodology that aims to produce high quality systems quickly, primarily by using iterative prototyping, active user involvement, and computerized development tools.

Spiral Model: A system development methodology that utilizes an iterative process that is focused on minimizing project risk.

System Development Life Cycle: A system development methodology that divides projects into phases and uses written documentation to maintain control.

System Development Methodology: A model that is used to structure, plan, and control the process of developing an information system.

V Model: A modified system development life cycle which emphasizes quality assurance.

Waterfall Model: A modified system development life cycle which allows some overlap and splashback between phases.

Web Information System: An information system that is accessed through the Internet and usually tightly integrated with databases and transaction-processing systems.

Systems Thinking and the Internet

Kambiz E. Maani
University of Auckland, New Zealand

INTRODUCTION: EVOLUTION OF MODERN THINKING

Despite our most impressive advances in science and technology, our prevailing worldview and the way we work and relate are deeply rooted in the thinking that emerged during the Renaissance of the 17th century. This thinking was influenced by the sciences of that era and, in particular, by Newtonian physics. Newton viewed the world as a machine that was created to serve its master—God (Ackoff, 1993). The machine metaphor and the associated mechanistic (positivist) worldview, which was later extended to the economy, the society, and the organization, has persisted until today and is evident in our thinking and vocabulary. The mechanistic view of the enterprise became less tenable in the 20th century, partly due to the emergence of the corporation and the increasing prominence of human relation issues in the workplace. As the futurist Alvin Toffler (1991) declared, "the Age of the Machine is screeching to a halt" (Toffler, 1991).

In the early part of the 20th century, a new breed of scientists, in particular, quantum physicists such as Werner Heisenberg (Uncertainty Principle) and Norbert Weiner (Cybernetics), began to challenge the Newtonian precepts (Zohar & Marshal, 1994). The 1960s saw the publication of Austrian biologist Von Bertalanffy's seminal text, *General Systems Theory* (1968), a major milestone in this field. Later, Jay Forrester of MIT introduced and demonstrated the applications of feedback control theory in simulation models of organizations (Forrester, 1958). Forrester's seminal work marks the birth of the professional field known as System Dynamics. System Dynamics is concerned with applications of systems theory and computer modeling in business, economics, and environment. System Dynamics is the forerunner and the scientific foundation of Systems Thinking.

BACKGROUND: CRITICAL ISSUES

Machine-age thinking, still prevailing today, is based on the following assumptions.

- Complete understanding of the universe is possible.
- All relationships can be described through simple cause-and-effect relationships.
- The world could be understood through analysis (breaking the wholes into pieces).

For well over a century, the western world has subscribed to a way of thinking known as analysis (Ackoff, 1995). In analysis, in order to understand something—a concept, a product, a law, an organization, the human body—we break it into pieces and study the pieces separately. This approach tends to overlook the interdependencies and connections among the constituent parts that are responsible for change and dynamics in systems.

On the one hand, this divide-and-conquer approach has served us well. It has enabled efficient mass production of goods and services, which has brought a new social and economic order and has produced unprecedented wealth and standards of living in the industrialized world. On the other hand, this thinking has resulted in the fragmentation of functions and has created complexity and cross-purposes within organizations.

The major intellectual and philosophical precepts that form the bedrock of our modern society, such as division of labour, free-market economics, mass production, and scientific management, are characterised by the following (Zohar & Marshal, 1994).

- The hierarchy
- Need for certainty, stability, and the absolute
- Treating organizations and the society as consisting of isolated, separate, and interchangeable parts
- Relationships based on conflict and confrontation
- Desire for control and bureaucratic methods
- Persistence of single points of view leading to friction and polarisation
- Overemphasis on specialist expertise, leading to fragmentation and loss of relevance

SYSTEMS THINKING

Systems Thinking (ST) is a discipline for understanding the dynamics of change and complexity underlying business, economic, scientific, and social systems. Systems Thinking has three distinct but related dimensions: para-

Table 1. Why we need Systems Thinking (Maani & Cavana, 2000)

• Increasing complexity in our lives
• Growing interdependence of the world
• Revolutions in management theories and practice
• Increasing global consciousness and yet local decision-making
• Increasing recognition of learning as a key organizational capability

digm, language, and methodology. These dimensions are outlined below (Maani & Cavana, 2000),

- Paradigm: Systems Thinking is a way of thinking about the world and relationships. This paradigm relates to the dynamic relationships that influence the behaviour of complex systems. A number of expressions that we use in daily language reflect the Systems Thinking paradigm—vicious/virtuous cycle, ripple effect, snowballing, spiral effect, domino effect, and chronic behaviour.
- Language: As a language, Systems Thinking provides a tool for understanding complexity and dynamic decision-making. The Systems Thinking language is known as Causal Loop (or Influence) Diagrams.
- Methodology: Systems Thinking provides a sophisticated computer modeling technology and associated learning environments for group interactions and learning.

FUTURE TRENDS: SYSTEMS THINKING AND THE INTERNET

In the past few decades, two movements have had a profound influence on the way we think and communicate—the Internet and Systems Thinking. Both are grounded in science and technology and complement each other in principle and practice. While one has become a household name, the other still remains relatively obscure. The Internet was born in military and academic quarters in the late 1960s. In the 1990s, the Internet moved into the public domain and rapidly became a mass movement. Today, the Internet and its associated e-commerce is the engine driving the globalization and convergence of various markets, services, and industries (Query et al, 2003).

Systems Thinking likewise originated in scientific circles and is slowly growing in appeal and applications. It offers a way of thinking based on the primacy of the whole and relationships. Systems Thinking deals with hidden complexity, ambiguity, and mental models. It provides tools and techniques to leverage change and to create lasting interventions (Maani, 2001).

For centuries, information and knowledge were the preserve of the clergy and the aristocrats who used them to dominate and manipulate the masses. In the past century, knowledge privilege extended to the boss, the manager, and the teacher who assumed this as part of their role and superiority. This knowledge divide, for its part, has strengthened the hierarchy and widened the gap between the haves and have-nots.

Although they may be regarded as purely technical advances, both Systems Thinking and the Internet challenge the age-old paradigms and the ways information and knowledge are disseminated. At a more fundamental level, they challenge the hierarchy and authority, power, and leadership. In essence, the Information Age has ushered in a new culture, new social movements, and new politics around the globe (Webster, 2001).

The Internet, through its unimpeded access and speed, has brought down in effect the boundaries that define business, trade, and even nationhood. Likewise, through compelling and coherent scientific principles, Systems Thinking breaks down the superficial dichotomies of the whole vs. the part; the individual vs. the community; integration and autonomy; and business, nature, and society. Together, the Internet and Systems Thinking can provide powerful synergies blending new concepts, tools, and technologies.

Over the past 20 years, new concepts and models have emerged that have dramatically challenged the prevailing assumptions and practices in business and management. Among these are the just-in-time philosophy and techniques, total quality management, and, more recently, supply chain management and enterprise systems. These paradigms have progressively removed the conventional boundaries between the organization, the customer, the supplier, and, to some extent, the competition.

An example of the Internet-Systems Thinking synergy is the supply chain management (SCM) practiced in businesses worldwide. The conceptual underpinning of SCM is systemic in nature in that the business or organizational boundaries span to cover the entire chain of supply. In this model, the stakeholders regard themselves as partners and collaborators in an enterprise system who seek the good of the whole. This notion stands in sharp contrast to the business models preceding it, which were characterized by control, competition, and the optimization of the parts in isolation.

In turn, the reach and speed of the Internet have enabled such cross-boundary and integrative business models as supply chain management and e-business/e-commerce, such as business-to–customer (B2C) and business-to–business (B2B), to rapidly grow and multiply. With the daily spread of globalization and e-commence, there is a dire need for reliable and efficient ways to manage, monitor, and control global large-scale projects that contain thousands of workflows and hundreds of organizations located at different sites. "Companies are faced with an increasingly competitive and hostile environment. The terms 'information exchange,' 'compatibility,' and 'interoperability' have become ubiquitous in this environment" (Badir et al, 2003). To survive, companies will have to constantly develop innovative ways for managing and controlling information and for doing business. These can be accomplished efficiently and effectively with Web-based models that automate the business processes and extend them beyond the organization.

The astronomical growth of the Internet is epitomized in China where, between July 2000 and January 2002, the number of Internet users doubled to 34 million. This, for example, has enabled over 50% of domestic Chinese insurance companies to offer policies online (Query et al, 2003).

In summary, the Internet has steadily pushed our thinking and our practices closer to a systemic model of the enterprise within the organizations and industries. At a higher level, the accelerated interdependence and interconnectedness of the global economy, trade, and governance calls for a systemic (holistic) view of the global community, the environment, and nationhood. To this end, the Internet has substantially shifted the distribution of economic activity around the world through its impact on transportation, uniform pricing, and marketing strategies that target individuals rather than regions, creating in essence a new economic geography (Anderson et al, 2003) and a single invisible continent (Ohmae, 2004). Thus, Systems Thinking and the Internet can play a synergistic role in the transformation of social, cultural, and political paradigms toward a global socioeconomic order. Given the evident trends, this appears not only plausible, but also inevitable.

CONCLUSION

Contemporary society is characterised by constant change, increasing complexity and speed, greater interdependence, and a rapid breakdown of social and political norms. In order to succeed (or even to survive) collectively, we need new paradigms and tools. According to Ackoff (1995), there is a progression from information to knowledge, to understanding, and to wisdom. So far, the Internet has provided for part of this progression—information abundantly and freely available to all, as well as a mechanism of instant and global communication transcending the conventional barriers of the past.

Systems Thinking, on the other hand, addresses a compelling need to gain a deeper understanding of the ever increasing complexity and interdependence of the world and the underlying forces shaping this interdependence. In this regard, Systems Thinking makes it possible to do the following.

- Examine and foresee the consequences of policy and strategic decisions
- Implement fundamental solutions to chronic problems
- Avoid mistakenly interpreting symptoms as causes
- Test assumptions, hypotheses, and scenarios
- Find long-term solutions and avoid "fire-fighting" behaviour
- Boost employee morale and improve productivity
- Implement change management without adverse side effects

The most urgent and fundamental dilemmas that we face today in the organization and in society require a collective will that comes from a deeper understanding of our place in history and our collective destiny towards an organic unity. Systems theory provides a scientific way of thinking about the world and relationships. It offers the art and the science of seeing the tree *and* the forest; it reconciles the superficial dichotomy between the individual and the collective, and focuses attention on the common good. The principles of systems theory collectively provide a paradigm and a language for deeper understanding of the chronic issues besetting our times. Used appropriately and with care, these principles can transform our age-old views and assumptions and move us towards a shared understanding. In this respect, the challenge for the Internet is to transcend its status as an information repository and a business tool, and become a channel to enhance understanding and goodwill. The evidence to date is promising inasmuch as the Internet continues to break down national, political, and social boundaries, thus bringing us closer to a holistic and systemic understanding of our global society.

REFERENCES

Ackoff, R. (1993). *Systems thinking and thinking systems*. Interact.

Ackoff, R. (1995). *Proceedings of Systems thinking in action conference*, Boston.

Anderson, V., & Johnson, L. (1997). *Systems thinking basics*. Pegasus Communications, Inc.

Anderson, W.P, Chatterjee, L., & Lakshmanan, T.R. (2003). E-commerce, transportation, and economic geography. *Growth and Change*, *34*(4), 415.

Argris, C. (1990). *Overcoming organizational defences – Facilitating organizational learning*. Boston: Allyn and Bacon.

Badir Y.F, Founou, R., Stricker, C., & Bourquin, V. (2003). Management of global large-scale projects through a federation of multiple Web-based workflow management systems. *Project Management Journal*, *34*(3), 40.

Checkland, P. (1981). *Systems thinking, systems practice*. John Wiley.

De Geus, A. (1995). *Proceedings of Systems thinking in action conference*, Boston.

Forrester, J. (1958). Industrial dynamics – A major breakthrough for decision makers. *Harvard Business Review, 36*(4).

Kauffman, Jr., D.L. (1980). *Systems one, an introduction to systems thinking*. Future Systems, Inc.

Maani, K. (2001). Systems thinking and the Internet. In *Internet management issues*. Hershey, PA: Idea Group Publishing.

Maani, K. & Cavana, R. (2000). *Systems thinking and modelling – Understanding change and complexity*. Prentice Hall.

Ohmae, K. (2004). *The invisible continent – Four strategic imperatives of new economy*. HarperCollins/Nicholas Brealey Publishing.

Query, J.T., &Jin, Z. (2003, August). Walking tiger. *Best's Review*, *104*(4), 119.

Senge, P. (1990). *The fifth discipline –The art and practice of the learning organization*. Doubleday/Currency.

Senge, P. (1992). Building learning organizations. *Journal for Quality and Participation.*

Shiba, S., Walden, D., & Graham, A. (1994). *A new American TQM*. Productivity Press.

Toffler, A., (1991). *The third wave*. Bantam Books.

Von Bertalanffy, L. (1968), *General system theory: Foundations, development, applications.* New York: George Braziller, Inc.

Webster, F. (2001). *Culture and politics in the Information Age: A new politics?* New York: Routledge.

Zohar, D., & Marshal, I. (1994). *The quantum society*. Morrow Press.

KEY TERMS

Balancing (Counteracting) Feedback: A systemic pattern that is responsible for stability, balance, and control. It represents adjusting, correcting, and counteracting processes that resist, slow down, or impede change and growth.

Causal Loop Diagram (CLD): A tool that captures the causal interrelationships among a set of variables. CLDs reveal systemic patterns underlying complex relationships and highlight hidden causes and unintended consequences.

Flow or Rate: The amount of change in a variable over time. Flow represents the change in the status or the quantity of a variable over a specified time unit.

Leverage: Knowing which actions may yield long lasting outcomes. Knowing where and when to intervene or influence a system to gain long lasting desired change using minimal effort and energy.

Reinforcing (Positive) Feedback: A systemic pattern that represents growth and self-feeding processes. It explains the dynamics underlying vicious and virtuous cycles with downward and upward spiral effects.

Stock: The accumulation of a variable such as asset, debt, energy, morale, and reputation. Stock represents the status or the quantity of a variable at a given point in time (i.e., a snapshot of reality).

System: A purposeful entity whose parts interact with each other to function as a whole. Thus, a system is not the sum of its parts; it is the product of their interactions (Ackoff, 1993). A system can be part of a larger system.

System Dynamics: A scientific tool that embodies principles from biology, ecology, psychology, mathematics, and computer science to model complex and dynamic systems.

Systems Archetype: A generic pattern of relationships that occurs in a wide range of systems and circumstances—natural, biological, political, social, and economic. A powerful tool for seeing high-level dynamics.

Systems Delay: The time lapse between action and response. Delays often destabilise the system and slow down a system's progress towards reaching its goal. Systems delays often mask anticipated outcomes as well as unintended consequences of actions since the intervening time lapse is often longer than expected.

Systems Thinking: Thinking holistically and conscientiously about the world by focusing on the interaction of the parts and their influence within and over the system.

Tacit Knowledge and Discourse Analysis

Michele Zappavigna-Lee
University of Sydney, Australia

Jon Patrick
University of Sydney, Australia

INTRODUCTION

Much of human experience is *below-view*, unattended to as we operate in the world, but integral to our performance as social creatures. The tacit knowledge involved in our practice allows us the experiential agility to be at once efficient and creative, to assimilate the novel and the familiar: in essence, to develop expertise. The possessors of skilful practice, the artisan, the witchdoctor or the physician, have occupied a position of both importance and mystery in most cultures since ancient times. Our interest over the ages in such hidden knowledge has caused us to mythologise expertise, placing it beyond the common by constructing it as unspeakable. Thus, in contemporary times it is not surprising that the dominant research perspective on tacit knowledge maintains that it is ineffable, that is, tacit knowledge cannot be understood by looking at what and how people communicate verbally. Indeed the word tacit has its origins in the Latin, *tacitus*, meaning silent.

As information technologies have begun to alter the way in which we think about our own processes while looking for ways to automate and retain our practices, we have been compelled to consider how the experience of the artisan mentioned above can engage with the constraints of the computational world. Capturing and sharing tacit knowledge has thus been a consistent problem in information systems and knowledge management research (Boisot, 1995; Nonaka & Takeuchi, 1995; Tsoukas, 2002; Wenger, 1998). Polanyi's Theory of Tacit Knowing (TTK) is the dominant theoretical perspective in this research. In this theory, Polanyi (1958) suggests that tacit knowledge is inherently personal, underlying our ability to perform tasks we find difficult to explain, such as facial recognition. Concepts in TTK have been made available to the information systems (IS) community largely through the work of Nonaka & Takeuchi (1995) who reinterpreted the theory, precipitating research directions in information systems that are misaligned with Polanyi's theses. A notable example is the movement in IS research to differentiate tacit and explicit knowledge (Johnson & Lundvall, 2001). In contradistinction, Polanyi asserts that explicit knowledge cannot be adequately separated from its tacit coefficient:

Now we see tacit knowledge opposed to explicit knowledge; but these two are not sharply divided. While tacit knowledge can be possessed by itself, explicit knowledge must rely on being tacitly understood and applied. Hence all knowledge is either tacit or rooted in tacit knowledge. A wholly explicit knowledge is unthinkable. (Polanyi, 1969, p. 144)

The emphasis in information systems research is typically on converting tacit knowledge into explicit knowledge (Hershel, Nemati, & Steiger, 2001). Attention is also given to setting up a dichotomy of tacit and explicit knowledge in terms of articulation (can it be carried in language?), codification (can it be turned into an artifact?) or judgment (is it objective or subjective?).

This article is structured to critique the dominant position in information systems research that tacit knowledge is ineffable. The background section provides an introduction to the extensive interdisciplinary literature on tacit knowledge, providing context for the subsequent section that deconstructs the assumptions that this literature makes about what it means to, in Polanyi's (1966, p. 4) terms, "know more than we can tell." To conclude, the role of linguistic and semiotic analysis in realising the growing trend toward theorising "community knowing," rather than knowledge as an artifact, is suggested in the final sections.

BACKGROUND: UNDERSTANDING TACIT KNOWLEDGE, AN INTERDISCIPLINARY PURSUIT

Both prior and adjacent to the popularisation of Polanyi's TTK, there has been extensive interdisciplinary research in tacit knowledge and a very large body of research looking at the implicit, situated nature of practice. These disciplines include philosophy, psychology, linguistics, semiotics, sociology, history, philosophy of science, and, most recently, knowledge management. A theme running through all these domains is uncovering the below-view. For example, Freud's (1970) theory of psychoanalysis and his corresponding concept of the unconscious locates a significant aspect of human experience below the view of awareness. Similarly, a movement in psychology investigates implicit learning (Reber, 1993), that is, learning that is below-view in the sense that it occurs without the attention of the subject on the process of learning. Recently, there have been attempts by psychologists to create psychological metrics for tacit knowledge. Wagner & Sternberg (2000) develop a method for measuring tacit knowledge in psychology and business management. Their "Tacit Knowledge in Management" (TKIM) inventory seeks to "identify individuals whose 'street smarts' indicate the potential for excellent performance in managerial and executive careers" (Wagner & Sternberg, 1991, p. 1). Conceiving of tacit knowledge as a facet of *practical intelligence*, they construct a triadic view of tacit knowledge as "managing self," "managing tasks" and "managing others," and test subjects in each area through multiple-choice responses to scenarios.

In the social sciences, the focus has been on identifying the implicit subject positions which have been naturalised by culture (Bernstein, 1971; Bourdieu, 1990). This naturalisation means that these positions remain below scrutiny. This is part of a research program that suggests that our experience is constructed through implicit social processes and parallels a body of research in psychology that claims that social behaviours are encoded automatically and without intention (Bargh, 1999). Bourdieu (1990) suggests that the individual internalises the cultural habitat in which they reside, their *habitus*. This means that they form dispositions to behave and construe their experience in certain ways. The acquisition of these structural constraints is a process of acculturation into specific socially-established groups or classes (Bourdieu, 1990, p. 130) and is akin to Bernstein's (1971) notion of *code orientation*. This is a different conceptualisation to the structuralist position, which argues that we follow unconscious rules in enacting our practice. Instead, it is a view of socially constructed dispositions that are inscribed in the habitus but which may shift with changes in context: "Agents to some extent fall into the practice that is theirs rather than freely choosing it or being impelled into it by mechanical constraints" (Bourdieu, 1990, p. 90).

Research in the history and philosophy of science examines science as a discipline that operates to efface the tacit component of its practice in order to maintain an ideology of objectivity (Collins, 1974, 2001a; Kuhn, 1962; Ravetz, 1971; Turner, 2001). In their investigation of science as a social practice, theorists in this domain have sought to reveal the tacit nature of the methods scientists employ. In fact, Polanyi's model of tacit knowledge itself was directed at critiquing the scientific method and to providing "a stable alternative to…[science's] ideal of objectivity" (Polanyi, 1966, p. 25). It provided a conceptual basis for understanding scientific activity as the practice of a craft (Ravetz, 1971). This is part of a program which deconstructs what Schuster (1984) refers to as the mythic construction of method. Method discourses in science typically try to generate the idea that they are systematic, explicit and objective (Collins, 2001a, 2001b; Ravetz, 1971). As such, they present themselves as incommensurable with other practices such as astrology, which they claim to be pseudoscience. Here we have an example of an institutional discourse, science, operating to efface the role of tacit knowing in the production of knowledge.

IS TACIT KNOWLEDGE INEFFABLE?

The attribute most consistently ascribed to tacit knowledge across the disciplines is ineffability (Baumard, 1999; Collins, 2001a; Nonaka & Takeuchi, 1995; Reber, 1993). The strong position is that tacit knowledge cannot be articulated in any linguistic form, while the weak position holds that it is difficult to articulate. Polanyi's (1966:4) widely cited suggestion that "we know more than we can tell" asserts the epistemological significance of tacit knowing in terms of its ineffability. In assessing this proposal, it is important to consider what it means "to tell." If telling means making explicit, codified artifacts that are directly transferred to the mind of the listener, then this kind of telling is not a possible means of exposing tacit knowledge. However, if we allow that telling involves processes of which the speaker is not necessarily aware and which are, in turn, subject to both unconscious and conscious interpretation by the listener, linguistic structure is reinstated as relevant to understanding tacit knowledge. These below-view processes are akin to Peirce's notion of the interpretant in semiosis, introduced in the previous section.

Thus, it appears that Polanyi's statement needs to be refined. We know more than we can tell only if we think about telling as making explicit knowledge. Such an assumption utilises an impoverished model of communication. This model, referred to as the mathematical model of communication (Shannon & Weaver, 1949), presupposes that meaning in communication is absolute and, as such, may be seamlessly transferred from the mind of the speaker to that of the listener. It applies what Reddy (1979) terms the *conduit metaphor*, that is, the notion that words are boxes with meanings inside that are unpacked by the person to which they are directed. Reddy (1979:287) argues that the metalingual resources of English privilege this kind of view, as the following examples suggest:

Whenever you have a good idea, practice capturing it in words
You have to put each concept into words very carefully

Just as in uttering the sentences above we are unlikely to focus on the presuppositions about communication that they presume, when we speak, that which we utter cannot be viewed as an overt object. We may well articulate what we know implicitly through patterns and features of language to which we do not directly attend. This is an argument that articulation does not produce a form that by definition is explicit, or in alternative terms, that articulation is not the equivalent of codification. However, many studies in information systems research equate these two modes of meaning-making.

There is a substantial tradition within psychotherapy that has approached language as a way of understanding a person's unconscious experience (Ferrara, 1988, 1994; Freud, 1960; Labov & Fanshel, 1977; Lentine, 1988; Parker, 1995; Pittenger, Hockett, & Danehy, 1960). This notion is further specified in linguistics by Halliday & Webster (2002, p. 303) who assert the significance of the relationship between grammar and the unconscious. Meaning-making with grammar is, according to this view, implicit meaning-making:

Conscious language achieves its creative force mainly by lexical means; and lexical items are semantically close to experience. Unconscious language depends much more for its creative force on grammar – and grammatical categories are far removed from experience. (Halliday & Webster, 2002, p. 303)

This appears in accord with the argument that description of the grammatical features in a subject's discourse will give us insight into the nature of their unconscious experience. It follows from this, that if the features of a subject's grammar involving meanings that are effaced are explicated, then the knowledge to which they point may be elicited. In looking at implicit meaning, Hasan et al. (1996) introduce the notion of implicit style in discourse. They give the example of the clause, "they will," which they argue is an example of maximal implicitness as it does not contain a string that is not implicit. The clause raises the questions: whom "will" and what "will" they? Hasan et al. (1996, p. 194) argue that we may distinguish between implicit and explicit ways of saying and that, when an implicit style is adopted by a speaker, "precise meanings become available only if certain additional conditions are met; the average working knowledge of the language is necessary but not sufficient." It is at this point that we require the services of a linguist.

Zappavigna-Lee & Patrick (2004) present a method for eliciting tacit knowledge from the language of interviewees through a process of directed interviews based on a linguistic model of tacit knowledge. This process centers upon the interviewer identifying semantic and grammatical features in the interviewee's language that suggest knowledge that the participant possesses, which remains "below-view." The knowledge is below-view in the sense that the linguistic-choice that the interviewee has made indicates *under-representation*. Under-representation occurs when components of knowledge are effaced in discourse, as they have been automatised by the individual. For example, the agent in a clause may be omitted. In addition to simply being left out of discourse, the knowledge may be effaced through generalisation that construes it as unavailable for deconstruction. For example, a verb may be nominalised meaning that something that was a process with component steps is rendered as a static object This means that there is less potential for these steps to be analysed. Tacit knowledge is subsidiary in the sense that we do not attend to such obfuscation. The directed interview method entails:

1. Identifying the semantic feature that suggests knowledge that is under-represented in the interviewee's discourse and important for their current knowledge management task.
2. Asking a question that elicits a more delicate response from the interviewee and which prompts them to elaborate on this feature.

In applying this interview method in an on-going case study in an Australian broadcasting organisation, Zappavigna-Lee & Patrick (2004) demonstrate that identifying such features will contribute in eliciting a more delicate description of the interviewees' meaning than a strategy based solely on eliciting content. The description is more delicate not only in the sense of being more specific lexically, but also in the sense of being increasingly precise lexicogrammatically.

In a directed interview with a senior manager in the digital media division of the organisation the interviewer noted that the manager possessed tacit knowledge that was embedded in nominalizations. For example, the manager described his division as a "service area" that "provide[s] services including IT services." "Service" is the nominalisation of a range of processes involving understanding, communicating and delivering feedback to clients. An underlying component of these processes is negotiating shared cultural experience. Through elaborating this nominalisation in the directed phase of the interview the interviewer uncovered that the manager believed that the greater the shared cultural experience and active cultural processes that he was able to foster, the greater the shared knowledge and cohesion of his employees. This information was not elicited merely by asking the manager to be more specific, as this would simply have elicited a more detailed rendering of his explicit style. This may have merely produced a taxonomy of the IT services in the organisation. Instead, embedded phenomena about the manager's beliefs and practices were uncovered by analysing his implicit style. This phenomenon is more "specific" in a particular way: it is the elaboration of parts of the interviewee's language of which they were not aware. This is an exercise below the surface of the text, and below the content plane on which most interviews are conducted. As such it involves a richer elicitation of the interviewee's experience.

FUTURE TRENDS: TACIT KNOWING OR TACIT KNOWLEDGE?

There is a growing movement in the IS literature that argues that it is not possible to capture tacit knowledge but that we should instead manage processes which facilitate its social transfer (Boland & Tenkasi, 1995; Stenmark, 2001; Wenger, 1998). This is part of a movement which acknowledges that that knowledge is human and social rather than an artefact that can be abstracted in a database (Weick, 1995; Wenger, 1998). It suggests that tacit knowledge is dynamic and carried in communities of practice (Huysman, 2004; Wenger, 1998). Rather than taxonomising tacit knowledge as if it were an object, this movement adopts a community-oriented model of knowledge management in which IT services are aimed at connecting people with relevant experts rather than attempting to externalize and codify this expertise (Swan, Newell, Scarborough, & Hislop, 1999).

The community-oriented model is aligned with the post-critical epistemological orientation of Polanyi's thesis, which deals with "knowing" rather than "knowledge;" with a process rather than an object. This conceptual position is in accord with the movement in semiotics and other disciplines concerned with theorising knowledge, such as philosophy and linguistics, away from reification: that is, away from a constituency-based view of knowledge as an object, toward a view of knowledge as dynamically produced by human subjects. Future research adopting this kind of framework should seek to employ existing tools such as linguistic analysis to understand how humans "do knowing" as opposed to "construct knowledge."

CONCLUSION

Polanyi's maxim that "we know more than we can tell" remains the dominant perspective on tacit knowledge in information systems research. However, few studies consider what it means to "tell" and the implications this has on asserting the ineffability of tacit knowledge. As this research begins to shift to considering tacit knowledge a process rather than an object and draws upon the large body of existing scholarship in other disciplines, attention should be given to what kind of meaning-making tacit knowledge involves. Tacit knowledge is potentially not as taciturn as we have assumed.

REFERENCES

Bargh, J.A. (1999). The unbearable automaticity of being. *American Psychologist, 54* (7), 462-479.

Baumard, P. (1999). *Tacit knowledge in organizations*. London: Sage.

Bernstein, B.B. (1971). *Class, codes and control*. London: Routledge and K. Paul.

Boisot, M. (1995). *Information space. A framework of learning in organizations, institutions and culture*. London: Routledge.

Boland, R. J., & Tenkasi, R.V. (1995). Perspective making and perspective taking in communities of knowing. *Organization Science, 4* (4), 350-372.

Bourdieu, P. (1990). *In other words: Essays towards a reflexive sociology*. Stanford, CA: Stanford University Press.

Collins, H.M. (1974). The TEA set: Tacit knowledge and scientific networks. *Science Studies, 4*, 165-186.

Collins, H.M. (2001a). Tacit knowledge, trust and the Q of Sapphire. *Social Studies of Science, 31* (1), 71-85.

Collins, H.M. (2001b). What is tacit knowledge? In T.R. Schatzki, K. Knorr-Cetina & E.V. Savigny (eds.), *The practice turn in contemporary theory* (pp. 107-119). London, New York: Routledge.

Ferrara, K. (1988). Variation in narration: Retellings in therapeutic discourse. In J. Baugh (ed.), *Linguistic Change and Contact*. Austin, TX: University of Texas.

Ferrara, K. (1994). *Therapeutic ways with words*. New York: Oxford University Press.

Freud, S. (1960). *Jokes and their relation to the unconscious* (J. Strachey, Trans.). New York: W.W. Norton.

Freud, S., & Strachey, J. (1970). *An outline of psychoanalysis*. New York: W.W. Norton.

Halliday, M.A.K., & Webster, J. (2002). *On grammar*. London: Continuum.

Hasan, R., Williams, G., Butt, D., & Cloran, C. (1996). *Ways of saying, ways of meaning: Selected papers of Ruqaiya Hasan*. London, New York: Cassell.

Hershel, R.T., Nemati, H., & Steiger, D. (2001). Tacit to explicit knowledge conversion: Knowledge exchange protocols. *Journal of Knowledge Management, 5* (1).

Huysman, M. (2004). Communities of practice: Facilitating social learning while frustrating organizational learning. In H. Tsoukas & N. Mylonopoulos (eds.), *Organizations as Knowledge Systems: Knowledge, Learning, and Dynamic Capabilities* (pp. 67-85). New York: Palgrave Macmillan.

Johnson, B., & Lundvall, B. (2001). Why all this fuss about codified and tacit knowledge? *Industrial and Corporate Change, 11* (2), 245-262.

Kuhn, T.S. (1962). *The structure of scientific revolutions*. Chicago, London: University of Chicago Press.

Labov, W., & Fanshel, D. (1977). *Therapeutic discourse: Psychotherapy as conversation*. New York: Academic Press.

Lentine, G. (1988). Metaphor as cooperation in therapeutic discourse. In *16th Annual Conference on New Ways of Analyzing Variation: Linguistic Change and Contact* (Vol. 30). Austin, TX: University of Texas, Dept of Linguistics.

Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. New York: Oxford University Press.

Parker, I. (1995). Everyday behavior(ism) and therapeutic discourse: Deconstructing the ego as verbal nucleus in Skinner and Lacan. In J. Siegfried (ed.), *Therapeutic and Everyday Discourse as Behavior Change: Towards a Micro-Analysis in Psychotherapy Process Research* (pp. 447-467). Norwood, NJ: Ablex Publishing Corporation.

Pittenger, R.E., Hockett, C.F., & Danehy, J.J. (1960). *The first five minutes: A sample of microscopic interview analysis*. Ithaca, NY: P. Martineau.

Polanyi, M. (1958). *Personal knowledge: Towards a post-critical philosophy*. Chicago: U.P.

Polanyi, M. (1966). *The tacit dimension*. London: Routledge & K. Paul.

Polanyi, M. (1969). *Knowing and being: Essays*. Chicago: University of Chicago Press.

Ravetz, J.R. (1971). *Scientific knowledge and its social problems*. Oxford: Clarendon Press.

Reber, A.S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. New York, Oxford: Oxford University Press & Clarendon Press.

Reddy, M. (1979). The conduit metaphor. In A. Ortony (ed.), *Metaphor and Thought*. Cambridge: Cambridge University Press.

Schuster, J. (1984). Methodologies as mythic structures: A preface to future historiography of method. *Metascience: Annual review of the Australasian Association for the History, Philosophy and Social Studies of Science, 1* (2), 17.

Shannon, C.E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.

Stenmark, D. (2001). Leveraging tacit organizational knowledge. *Journal of Management Information Systems, 17* (3), 9-24.

Swan, J., Newell, S., Scarborough, H., & Hislop, D. (1999). Knowledge management and innovation: Networks and networking. *Journal of Knowledge Management, 3* (4), 262-275.

Tsoukas, H. (2002). Do we really understand tacit knowledge? In M. A. Lyles (ed.), *Handbook of Organizational Learning and Knowledge* . Blackwell.

Turner, S. (2001). Throwing out the tacit rule book: Learning and practices. In T. R. Schatzki, K. Knorr-Cetina & E. V. Savigny (eds.), *The practice turn in contemporary theory* (pp. ix, 239). London, New York: Routledge.

Wagner, R.K., & Sternberg, J. (1991). *Tacit knowledge inventory for managers: User manual*. The Psychological Corporation.

Wagner, R.K., & Sternberg, J. (2000). Tacit knowledge and management in the everyday world. In R. J. Sternberg (ed.), *Practical Intelligence in Everyday Life* (pp. xiv, 288). Cambridge, New York: Cambridge University Press.

Weick, K.E. (1995). *Sensemaking in organizations*. Thousand Oaks: Sage Publications.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge, New York: Cambridge University Press.

Zappavigna-Lee, M., & Patrick, J. (2004). *Literacy, tacit knowledge and organisational learning*. Paper presented at the The 16th Euro-International Systemic Functional Linguistics Workshop, Madrid.

KEY TERMS

Below-View: Elements of experience which are not available to direct inspection without applying some semiotic tool such as linguistic analysis.

Conduit Metaphor: A metaphor about communication, which suggests that an addresser's ideas are objects contained in packages, known as words, that are directly sent to the addressee.

Focal Awareness: Polanyi's term for conscious perception that the individual can directly access. Contrast with *subsidiary awareness*.

Habitus: Bourdieu's term for the cultural context in which an individual resides and which influences their practice.

Subsidiary Awareness: Polanyi's term for perception to which an individual does not have direct access, as it is not part of their focal awareness. Contrast with *focal awareness*.

Tacit: From the Latin *tacitus*, meaning silent.

Tacit Knowing, Tacit Integration: Polanyi's concept of the process of implicit integration of subsidiary and focal elements by an individual. The individual attends "from" the element in their subsidiary awareness "to" the element in their focal awareness.

Tacit Knowledge: Implicit understanding of which the individual is not directly aware and which is involved in their skilful practice.

Under-Representation: Zappavigna-Lee & Patrick's term for a set of specific linguistic features that indicates the presence of tacit knowledge in an individual's talk.

Tailorable Information Systems

D. S. Stamoulis
National and Kapodistrian University of Athens, Greece

D. Theotokis
University of Peloponnese, Greece

D. I. Martakos
National and Kapodistrian University of Athens, Greece

INTRODUCTION

The quest for the business fit of information systems is as old as the history of computers. In the past, computers used to lag behind the daily needs of the users because they lacked the capabilities of tackling the issues raised by the users. In the late 80's, the fourth generation of computers was a decisive step toward the acquisition of such system capabilities that allowed powerful software to be developed (O'Brien, 2002), rendering them able to tackle almost any business problem.

However, the previous decade has been marked by failures of big information systems (IS) projects, despite the sound accumulated knowledge on various information systems development methodologies and the advent of software engineering tools. The examples of IS that failed to deliver their premise are abundant, to name but a few: the U.S. IRS tax system modernization (Hughes & Cotterell, 2002); the London Ambulance Service's information system and the UK automated stock brokerage system TAURUS (Beynon-Davis, 1995); the automated train seats' reservation system SOCRATE in France (Eglizeau, Frey, & Newman, 1996). Of course, a variety of reasons could lead an IS project to a disaster, and the sources of the failure could stem from project management malpractice to poor software engineering techniques, to wrong leadership choices, and so forth. Apart from them, research has shown (e.g., Fitzgerald, 1990; Oei, Proper, & Falkenberg, 1994; Robertson, 1997) that any IS, no matter how "perfect" it can be built, is immediately being undermined for the forces of change. These forces of change inherently exist in any organization and have to do not only with the constant expansion or modification of the user requirements day-to-day, but also with the user's computer and information literacy. The latter means that the familiarization of the user with an IS creates new needs that IS maintenance comes to respond to. Maintenance is a corrective action, aiming at endowing the IS with improved functionality, in contrary to the irrational inference that a "finished" IS product is supposed to operate in the same business/organizational context meeting always to the same set of initial requirements (Bjerknes, Bratteteig, & Espeseth, 1991). No matter how precise and complete the elicitation of user requirements may be, the behaviour of an IS is constantly diverging from the evolving needs of the business/operational context (Stamoulis, Martakos, & Introna, 1998).

BACKGROUND

All methods and techniques used so far to produce flexible enough IS which can achieve business fit have all fallen victims of the fixed-point pseudo-theorem (Paul, 1993): "There exists some point in time when everyone involved in the system knows what they want and agrees with everyone else." This statement can never be verified in practice, and subsequently, describes a dead-end situation. User requirements are the result of the evolving business/organizational context of the IS, whereas the IS itself has been built to an exact specification which had been based on a snapshot of its business context's operation. Ideally, IS should behave like living organisms (Paul, 1993) that have the ability to adapt to changes of their operating environment. But, the software engineering technologies along with the way of thinking of the IS development discipline cannot escape the destiny of their determinism. This is where ateleology and tailorable IS (TIS) come into play. The former negates the mentality of constructing an IS as a product developed to meet a specific telos (purpose, definition) and the latter provides the theoretical framework for a new breed of truly flexible IS to appear. The questions of whether TIS are feasible and if software can even be tailorable are answered by means of the Atoma architecture, which has already been implemented.

MAIN THRUST OF THE ARTICLE

Tailorability is not another word for customization. IS flexibility has been based for years on how many parameters of the IS functionality can be configured by the end-user without taking any maintenance action. The tailorable IS thinking immediately focused on adaptivity. A program is adaptive when it changes its behaviour depending on its current contextual situation (Gouda & Herman, 1989). If everything in an IS may change, then tailorability affects the user interface in terms of the user dialogue, the screen layout, the help function, the icons and the functions themselves, as well as the functionality in terms of input, processing and output (Patel, Gardner, & Paul, 1995). Of course, this flexibility should be realized in real-time, without disrupting the IS operation, upon user's demand. If all of these features of an IS could change, then IS would really behave like a living system, and then it could be named as a tailorable IS (Patel et al., 1995) whose purpose is to "provide control over systems to a user". Looking at the problem at the macro level, an abstract systems architecture model has been proposed (Stamoulis, Patel, & Martakos, 1996), which describes the need for tailorability at three software layers: IT infrastructure (hardware, operating system, data sources); application logic (workflow, process logic, and facilities layers); and presentation logic layer. For each one of these layers, emerging software technologies that demonstrate tailorability traits or characteristics have been identified (Stamoulis, Kanellis, & Martakos, 2000).

In the old vein of thought, an observer was looking at an IS as an autonomous organ within a living organism's context, that is, its business environment, expecting from the IS to follow the needs and the commands of its context. From a different perspective, the user-IS system co-evolutes as a closed system (Stamoulis et al., 1998) towards a destination, a purpose, a telos that cannot be known in advance. Having used the self-organizing theory coming from the physical sciences to explain the theoretical foundations of TIS (ibid), the embracement of the ateleology ideas comes as a natural subsequence. At about the same time, research work (Baskerville, Travis, & Truex, 1992, 1996; Truex, Baskerville, & Klein, 1999) in the same area – but not starting from the need for true flexibility – started converging to the need for ateleological IS development, by means of a methodical systems development, where a methodical means is not obeying to well-structured techniques through which compliance with a specific goal can be achieved. In theory, the problem was well-defined. In practice, there was a missing link that would connect the tailorability concept with possible implementation routes. The definitive characteristic of a TIS – the key to achieve tailorability – is the deferred design decisions (Patel, 1999, 2003) mechanism.

Design decisions translate the user requirements into rigid specifications, which render an IS short-living by definition. Since an IS must constantly adapt to changing user requirements, design decisions must be deferred until run-time, when the ultimate judge, the user, determines the preferable IS behaviour and, consequently, fixes the appropriate design decisions for the IS to run. If ever these need to be changed, this IS behaviour change must occur at the user's control, at run-time, while the IS remains operating. To prove realistic, this new way of thinking had also to define other issues such a method of work for implementing TIS, a security and stability framework for TIS that constantly changes and so forth. As far as the former is concerned, the concept of an IS blueprint has been devised (Stamoulis, Theotokis, Martakos, & Gyftodimos, 2003). In line with the ateleology and a methodical IS development, TIS are initially constructed as blueprints with minimal features that correspond to a vague problem definition within a problem domain, as the user starts using the TIS, co-evolutes towards an end, a goal, a telos, and takes design decisions that implement a specific, well-defined IS behaviour, pertinent to each and every problem situation. To arrive at an IS blueprint, the Soft Systems Methodology (Checkland & Scholes, 1990) can be perfectly used, as explained in Theotokis (2003). As far as the latter is concerned, a model with separations of concerns in terms of security and change authorization domains has been proposed to guarantee for the stability of TIS (Farmakis, Kanellis, Martakos, & Stamoulis, 2000).

In the following, an innovative implementation of the deferred design decisions mechanism is been presented called the Atoma architecture and new developments towards tailorability are being discussed. This approach is based on the concept of separation of concerns as proposed by Morch (2003) and Randall and Lycett (2003). Concern separation is key in realising deferred design decisions as it facilitates the notion of "injectable" behavioural adjustments in existing operational IS. As such, Atoma was selected as representative of systems supporting concern separation and consequently behavioural variations.

The Atoma model was developed in order to better model cross-cutting behavioural changes related to the evolution of the behavioural landscape of an information system, by alleviating the problem associated with traceability and monitoring, scattering and tangling and, at the same time, enable the incorporation of modelled behavioural changes into an existing system without affecting in any form the system's stability. An implementation of the model exists and is based on the Java programming language. The key idea behind Atoma is to improve the effectiveness in realizing modifications, of the behavioural landscape of object, in a dynamic, transparent and natural way, so that systems are considered as "living", evolving and tailorable entities, both from a conceptual viewpoint, as well as that of implementation.

The Atoma model allows object-oriented design and code to be decomposed into units, describing basic behaviour, as this is captured during the initial design phase from the contractual requirements, and units that specify either variations or changes to these requirements, or in fact new requirements, as these emerge in time. Both at design and implementation level, the former are represented as standard object-oriented classes, while the latter are roles that, when composed with classes, realize the ever-emerging requirements. Each role can, therefore, be refined separately to a code artefact, and the details of the code composition can be derived from those of the design composition. There is excellent traceability and monitoring at this level, because code and design units correspond directly, and so do requirements and design. Furthermore, within a single composition, standard object-oriented design or code is used. This traceability and monitoring facilitates both evolution and "round-tripping": Projecting design changes into the code and requirements or, for that matter, reflecting the changes made in the code back into the design and requirements.

When a new requirement is added or an existing one is changed, a new design aspect, a role, is created to address it. The new design aspect can then be composed with the existing design. Then, the design aspect can be refined to a code aspect (role implementation), which is similarly composed with the existing code. The changes are localized without risking the rest of the system's stability, so there is no tangling, and both traceability and monitoring are preserved. In addition to this, dealing with the alignment problem, the decomposition into roles alleviates the monolithic nature of the design, and allows for concurrent development, while the composition underlying the Atoma model provides a powerful mechanism for integration, evolution, customization, adaptation, tailorability and improved reuse.

It is, in fact, this composition mechanism that provides a remedy to the problem of constructing rigid systems. The necessary, inflexible mapping of requirements to design units is no longer required, and design decisions, which fix requirements into design, and subsequently code, are no longer necessary. By separating the various cross-cutting aspects of a system, modelling them independently and composing them back to deliver the required function of the information system, we advocate that deferred design decisions can be realized, and thus enable the construction of truly tailorable information systems.

FUTURE TRENDS

Although a number of approaches towards behavioural tailorability have and are being proposed including, but not limited to role modelling, aspect-oriented programming, multidimensional separation of concerns, not much work has been done addressing the issue of data tailorability The area of tailorability plays a key role in achieving a fully tailorable system. Currently, techniques such as data migration are being employed for tailoring data according to new requirements. However, thorough consideration must be given to techniques and approaches that will allow data modelling to take into account the need for change, whether such change is anticipated or not. Modelling techniques need to be limited to data, but they should address roadmaps and methodological procedures that will enable analysts, designers, and developers to design tailorable systems. Another important issue that must be considered in great depth concerns the degree to which platform (software and hardware) architectures are tailorable or support tailorability. Model driven architectures take into account the need for change, but not as their first priority. However, if tailorable systems are to be successful, all aspects of software construction, operation and maintenance should focus on the issues governing and characterizing tailorable information system.

CONCLUSION

Tailorable information systems are becoming a necessity as they support the evolving nature and world an IS models. Based on the concept of deferred design decisions, they provide the means to accommodate change as it occurs, when it occurs. In as such, they differ from traditional IS in that they do not embed any rigid design decisions. Instead they support the notion of behavioural evolution in terms of "injectable" behavioural modifications. Role-based models and role-based programming are seen as the vehicle for this purpose, since they are underpinned by the concept of separation of concerns, a concept that supports deferred design decisions.

REFERENCES

Baskerville, R., Travis, J., & Truex, D. (1992). Systems without method: The impact of new technologies on information systems development projects. In K.E. Kendall et al. (Eds.), *Impact of computer supported technologies on information systems development*. Elsevier Science Publishers B.V., IFIP.

T

Baskerville, R., Travis, J., & Truex, D. (1996, January). *A methodical systems development: The deferred meaning of systems development methods*. Working Paper, Department of Informatics and Management Accounting, Copenhagen Business School.

Beynon-Davis, R. (1995). IS failure and risk assessment: The case of the London ambulance computer-aided dispatch system. In G. Doukidis et al. (Eds.), *Proceedings of the 3rd European Conference on Information Systems*, Athens.

Bjerknes, G., Bratteteig, T., & Espeseth, T. (1991). Evolution of finished computer systems: The dilemma of enhancement. *Scandinavian Journal of Information Systems, 3*.

Checkland, P.B., & Scholes, J. (1990). *Soft systems methodology in action*. Addison-Wesley.

Eglizeau, C., Frey, O., & Newman, M. (1996). SOCRATE: An implementation debacle. In J. Dias Coelho et al. (Eds.), *Proceedings of the 4th European Conference on Information Systems*, Lisbon.

Farmakis, C., Kanellis, P., Martakos, D., & Stamoulis, D. (2000). Infrastructural considerations for the deployment of tailorable information systems. *Proceeding of the ICIMIT 2000*, November, Singapore.

Fitzgerald, G. (1990). Achieving flexible information systems: The case for improved analysis. *Journal of IT, 5*, 5-11.

Gouda, M.G., & Herman, T. (1989). *Adaptive programming*. Working paper CS_TR_89_29. University of Texas, Austin.

Hughes, B., & Cotterell, M., (2002). *Software project management*. McGraw-Hill.

Morsh, A.I. (2003). Aspect-oriented software components. In N. Patel (Ed.), *Adaptive and evolutionary information systems* (pp. 105-124). Hershey, PA: Idea Group Publishing.

O'Brien, J. (2002). *Managing information systems: Managing information technology in the e-business enterprise* (5th ed.). McGraw-Hill.

Oei, J.L.H., Proper, H.A., & Falkenberg, E.D. (1994). Evolving information systems: Meeting the ever-changing environment. *Information Systems Journal, 4*, 213-233.

Patel, N.V. (1999). The spiral of change model for coping with changing and ongoing requirements. *Requirements Engineering, 4*, 77-84.

Patel, N.V. (2003). Deferred system's design: Countering the primacy of reflective IS development with action-based information systems. In N. Patel (Ed.), *Adaptive and evolutionary information systems* (pp. 1-28). Hershey, PA: Idea Group Publishing.

Patel, N.V., & Paul, R.J. (1998). Towards systems tailorability: Comparative analysis of business change and systems flexibility. In *Proceedings of the Association for Information Systems 1998 Americas Conference*, August 14-16, Baltimore, Maryland (pp. 122-124).

Patel, N.V., Gardner, L.A., & Paul, R.J (1995). Moving beyond the fixed point theorem with tailorable information systems. In G. Doukidis, R. Galliers, T. Jelassi, H. Krcmar, & F. Land (Eds.), *Proceedings of the Third European Conference on Information Systems*, ECIS '95, Athens, June.

Paul, R.J. (1993). Dead paradigms for living systems. *Proceedings of the First European Conference on Information Systems Henley-on-Thames*, March 29-30. The Operational Research Society, Birmingham (pp. 250-255).

Paul, R.J. (1994). Why users cannot get what they want. *International Journal of Manufacturing System Design, 1*(4), 389-394.

Randall, P., & Lycett, M. (2003). Adaptive and evolutionary systems: Lessons from object, component and agent approaches. In N. Patel (Ed.), *Adaptive and evolutionary information systems* (pp. 59-80). Hershey, PA: Idea Group Publishing.

Robertson, P. (1997). Integrating legacy systems with modern corporate applications. *Communications of the ACM, 40*(5), 39-46.

Stamoulis, D., Kanellis, P., & Martakos, D. (2000). Tailorable information systems: Resolving the deadlock of changing user requirements. Accepted for publication to the *Journal of Applied System Studies*.

Stamoulis, D., Martakos, D., & Introna, L. (1998). Systems for users, not for observers: The tailorability concept. In W.R.J. Baets (Ed.), *Proceedings of the 6th European Conference on Information Systems, ECIS '98*, Marseille, vol. 3 (pp. 1011-1024).

Stamoulis, D., Patel, N., & Martakos, D. (1996). A systems architecture model and implementation platforms for tailorable infromation systems. *Proceedings of the 4th European Conference of Information Systems*, Lisbon, Portugal 1996 (pp. 313-322).

Stamoulis, D., Theotokis, D., Martakos, D., & Gyftodimos, G. (2003). Ateleological development of design decisions independent information systems. In N. Patel (Ed.), *Adaptive and evolutionary information systems* (pp. 81-104). Hershey, PA: Idea Publishing Group.

Theotokis, D. (2003). Approaching tailorability in object-oriented information systems through behavioural evolution and behavioural landscape adaptability. To appear in M. Lycett & D. Martakos (Eds.), *Journal of Applied Systems Studies Special Issue on Living, Evolutionary and Tailorable Information Systems: Development Issues and Advanced Applications* (vol. 5, no. 1, 2004).

Theotokis, D., Gyftodimos, G., & Geogiadis, P. (1996). Atoms: A methodology for component object oriented software development applied in the educational context. In D. Patel, Y. Sun, & S. Patel (Eds.), *Proceedings of the Third International Conference on Object Oriented Information Systems*, December 16-18. South Bank University, London, Springer (pp. 226-242).

Theotokis, D., Gyftodimos, G., Geogiadis, P., & Philokyprou, G. (1997). Atoma: A component object oriented framework for computer based learning. In G.M. Chapman (Ed.), *Proceedings of the Third International Conference on Computer Based Learning In Science (CBLIS'97)*, July 4-8 1997, De Montford University, Leicester, UK, ISBN 80-7040-217-2 (pp. B1(15)).

Theotokis, D., Kapos, G.-D., Vassilakis, C., Sotiropoulou, A., & Gyftodimos, G. (1999). Distributed information systems tailorability: A component approach. Én the *Proceedings of the 7th IEEE Workshop on Future Trends of Distributed Computing Systems (FTDCS'99)*, Cape Town, South Africa, December 20-22 (pp. 95-101).

Theotokis, D., Sotiropoulou, A., Gyftodimos, G., & Georgiadis, P. (2001). Are behavioural design patterns enough for behavioural evolution in object-oriented systems? *Proceedings of the 8th Panhellenic Conference on Informatics*, Nicosia, Cyprus, November 8-10, 2001.

Truex, D.P., Baskerville, R., & Klein, H. (1999). Growing systems in emergent organisations. *Communications of the ACM, 42*(8), 117-123.

KEY TERMS

Ateleology: In the context of IS development, refers to the development of an IS which does not aim at the attainment of a specific set of goals only, expressed in the form of user requirements as they have been articulated at a specific point of time, but rather refers to the developing an IS that addresses a range of issues within a specific problem domain and, at the same time, is flexible enough to accommodate changes and extensions of the IS behaviour while it is being operated.

Behavioural Landscape: The set of behavioural characteristics an object possesses at a given time during its lifetime.

Behavioural Variations: Modifications on the functionality exposed by an object. Behavioural variations occur as a result of changes in the state, context, perspective and use of an object, and as such, they can be categorized as internal or external ones.

Deferred Design Decisions: The cornerstone of implementing tailorable information systems, the DDD mechanism through which decisions traditionally are taken during the design phase of an information system development project, as a result of a set of user requirements elicited during analysis, are now being deferred at run-time, until after the user decides about the required information system behaviour. The information system can then follow the new design specifications without any maintenance or redevelopment action, and execute them to provide the user-defined behaviour.

Information System Blueprint: An initial, functioning skeleton of a tailorable information system, built to reflect a problem solving framework within a specific problem domain which, by means of an inherent deferred design decision mechanism, can be tailored by the end-user to acquire or amend its existing behaviour, in order to meet the specific needs of a particular business/organizational context. A blueprint of a tailorable information system can truly achieve flexibility.

Role: A concept that its individuals stand in relation to other individuals, and that they can enter and leave the extent of the concept without losing their identity.

Tailorable Information System: An information system whose behaviour is dynamically user-controlled at run-time, without causing instability and disruption to the rest of the information system, not affected by the change(s) initiated by the user, aiming at endowing the information system with the capability to meet new or altered user requirements, and, thus, rendering it adaptable to any contextual change within a problem domain.

Teams and Electronic Technologies

Sharmila Pixy Ferris
William Paterson University, USA

Maureen C. Minielli
Indiana University-Purdue University at Indianapolis, USA

INTRODUCTION AND BACKGROUND

Teams[1] are integral elements in today's corporate and industrial worlds, considered by some to be the fundamental units of organizations, and technology has become essential to teamwork. Just as it is nearly impossible to engage in any type of organizational task without team support today, it is equally impossible to engage in effective teamwork without technological support. Technology, particularly computer and digital technology, has become essential to the functioning of both traditional teams (face-to-face) and virtual (distributed or geographically separated) teams. Technology and virtuality arguably change work groups in three important ways: they introduce new dimensions of communication among members by breaking down traditional barriers of space and time; they modify traditional group processes; and they enormously enhance the group's capacity to access, share, manipulate, retrieve, and store information.

There are inherent advantages of technology for teams. For example, technology frees team members from geographical constraints. People no longer must work in the same physical location to work together; participants can contribute from any part of the world, at any time of the day or night. By breaking down the barriers of space and time, technology enables teams to fully utilize the expertise of the members of an organization (or of several organizations) without pulling them from other projects or incurring relocation expenses. Hence, at least notionally, such teams have immense potential for improving organizational effectiveness.

Although there is a substantial amount of research on teams and technology, the focus on the technologies available to team members has hitherto remained the domain of Information Technology. This article takes an interpersonal focus to consider the software and communication technologies that allow individuals and teams to communicate, exchange information, interact collaboratively, and manage data within and outside of a group, synchronously and asynchronously. The technological support systems discussed here cover a wide range of technological applications, primarily, but not limited to, computer, audio-visual, and phone applications. To organize our discussion, we will use the classification scheme of messaging and conferencing systems (e.g., e-mail, discussion lists, electronic bulletin boards, Web logs and short message services), information-exchange and data-management systems (e.g., Internet, File Transfer Protocol, Gopher, Telnet, the World Wide Web, and Internet alternatives), and commercial groupware packages. Note that as our focus is on the team member as end-user rather than on the hardware, installation, programming, and setup as considerations of hardware and programming will not be discussed[2].

MESSAGING SYSTEMS

Messaging systems enable and facilitate communication among team members, which an be one-to-one, one-to-many, and synchronous or asynchronous. The earliest methods using computers and the Internet were largely asynchronous, with limited synchronous communication. The advent of the World Wide Web and wireless technology have increasingly allowed people to use messaging systems to communicate in real time. In this section, we will consider asynchronous messaging (e.g., e-mail, discussion lists, bulletin and message boards, Web logs, and short message service [SMS]), synchronous messaging (e.g., chat and instantaneous interactive messaging), and conferencing.

ASYNCHRONOUS COMPUTER MESSAGING SYSTEMS: FROM E-MAIL TO SMS

E-mail is the simplest form of messaging, with information-management features that allow for easy storage and retrieval, as well as easy editing, replying, and attaching. Perhaps because e-mail is transparent, it is the primary means of communication in organizations today (Barnes, 2003). Wireless access to the Internet, through personal digital assistants and cell phones, frees users from the constraints of a networked computer.

Evolutions in technology increasingly facilitate vocal and visual features for e-mail. Verbal elements require software and a microphone, and visual add-ons utilize Web cameras. The advent of cheaper technology and freeware may see catalyzed change in traditional text-based e-mail. For example, video software like CyberLink's VideoLive Mail (www.cyberlink-usa.com/), Cornell University's CU-See Me (www.cuworld.com/) is inexpensive, and Talk99 from MediaRing.com (http://www.mediaring.com) is a freeware telephony program melding voicemail and e-mail.

Electronic bulletin boards or BBSs are "worldwide, posted public messages on a wide variety of subjects" (Lamb, 1999, p.23), posted by individuals and read by groups. BBSs provide connected or disconnected individuals with a means of sharing knowledge and information, and help close the distance and isolation gap between individuals. Team members can utilize BBSs through links to their personal Web pages or through free task-focused BBSs like InsidetheWeb (http://www.insidetheweb.com) or NgBook (www.bigfoot.com/~huangng).

Discussion Lists provide an asynchronous communication medium with some similarities to BBS and e-mail. While listservs and newsgroups are rarely a primary means of communication for teams, moderated listservs can be useful for threaded discussion. Discussion lists can be set up through any networked computer system.

A final asynchronous messaging system is Web logs or blogs, Web-based individually maintained journaling sites. Like other technological innovations, Web logs have integrated themselves into the corporate world. Business software like Manilla (http://manila.userland.com) and Traction (http://www.tractionsoftware.com) allows corporate intranet teams to converse with each other and serves as a "community-building and coordinating tool" (Herman, 2003). Business blogs also allow for instantaneous communication between team members, tracking of team communication and decision-making, and communication with clients and other business partners; they serve as informational references for team members and clients to consult (Herman, 2003).

SMS is a growing phenomenon in the United States and Canada. SMS allows individuals to exchange short (less than 160 characters) wireless text messages with one to several recipients (Devi, 2003). At five to 20 cents per message, SMS messages are typically cheaper than voicemail messages (*Computer Weekly*, 2003), with limited character length allowing for straight-to-the-point messages and discouraging superfluous text (Yap, 2003). SMS is growing in popularity in the business sector for its many benefits to teams, such as global applications like general staff and other internal communication, specific applications like reminders of, or changes, in meeting times, or quick answers to questions (Mills, 2003).

SYNCHRONOUS MESSAGING SYSTEMS: FROM IM TO CHAT

Synchronous interactive communication programs (SCs) are available in many different computer platforms—UNIX platforms "Talk" and "nTalk," BM VM/CMS systems "Tell," VAX systems "Send," and the popular America Online software "Instant Messenger" or "IM." All these systems allow for real-time communication among communicators, and, to facilitate use, most systems allow participants to check whether others are logged on at the same time, employing user IDs or preestablished lists. Since its adoption by such subscription services as America Online, Yahoo, and MSN, synchronous communication using freeware has achieved great popularity and expanded into use in business and industry. In fact, 42% of business Internet users report use of IM in the workplace, even though 70% of IT departments don't support it (Schwartz, 2002); it is predicted to rise from 5.5 million users in 2000 to 180 million in 2004 (Schwartz. 2002). A number of SC software programs exist for business use: Lotus' Sametime (used by the U.S. Navy), QuickConference (http://www.quickconference.com), and Ikimbo's Omniprise (http://www.ikimbo.com).

A sophisticated software protocol that enables participants to interact virtually is Chat. Internet Relay Chat (IRC) is Internet-based network that enables multiple communicators to synchronously interact in an online environment. IRC is easily undertaken by connecting to a server on the IRC network. While IRC can be a convenient and powerful virtual meeting tool, it is a public forum and can lead to lack of privacy. This issue can be circumvented by using cyber-meeting places on commercial networks, proprietary groupware and course management software, or by using chat rooms designed specifically for business and industry such as Webtrain (http://www.webtrain.com), Magma Communication's Chat Server (http://www1.magma.ca), ParaChat (http://www.parachat.com), Volano Chat (http://www.volano.com), and Divine's eShare Expressions (http://www.divine.com).

SYNCHRONOUS CONFERENCING SYSTEMS

Conferencing systems are more specifically designed to facilitate synchronous virtual meetings than Chat or IM. While many corporate and higher education facilities maintain on-site teleconference rooms with computer controlled audiovisual transmission between locations, teams lacking this equipment can use computer-enabled conferencing to work together in real time using free proprietary software. For example, Microsoft Corporation's Netmeeting[3] (http://www.microsoft.com/windows/netmeeting) allows users to engage in audio and video conferencing, as well as white boards, online chats, file transfer, and program sharing. Another example is Sun Microsystems (freeware.thesphere.com) Java-oriented (mostly educational) software programs and iVisit (http://www.ivisit.org), as well as the Java Shared Data Toolkit, Java Web Server, and Java WorkShop.

Videoconferencing adds a visual dimension to computer conferencing, making it "richer" as it "returns" some of the dimensions of traditional F2F communication lost in computer conferencing. High-end business videoconferencing vendors include First Virtual Communications (http://www.fvc.com) and Polycom iPower (formerly PictureTel) (http://www.polycom.com), which operate over private ISDN lines or private IP networks; lower-end vendors include GrassRoots Communications' Grass Roots Live (http://www.grassrootscommunication.com) and SeeSaw.com (http://www.realityfusion.com) (Metz, 2001).

Videoconferencing systems can also be built around Chat systems, incorporating audio and visual elements. WebCast ProServer (http://bant.ms.ornl.gov) is an example of a computer program that allows users to broadcast live audio and video and participate interactively. A related service is WebCast Personal ICQ (http://www.icq.com), an Internet telephony software package that weds message boards, chat rooms, data conferencing, and file transfer. Other software programs that allow for real-time business collaboration include Netopia's Timbuktu Pro (http://www.netopia.com), Genesys' Meeting Center (http://www.genesys.com), Altiris' Carbon Copy (http://www.altiris.com), and Tandberg's Management Suite (http://www.tandberg.net).

INFORMATION-EXCHANGE/ DATA-MANAGEMENT SYSTEMS

Finding, exchanging, and managing information, as well as sharing data and information resources are essential to the functioning of teams. Team functioning has been revolutionized by computer technology.

THE INTERNET

The Internet as it exists today is a vast system of networked computers, reaching across the globe. The Internet originated in the 1970s with the first wide-area computer network, the Advanced Research Projects Agency Network (ARPANET), a government sponsored experiment that linked research universities and the military. ARPANET grew to encompass other networks, including the National Science Foundation Network (NSFNET) and tens of thousand of local, regional, national, and international networks (Barnes, 2003; Sudweeks et al., 1995). The Internet utilized packet switching and communication protocols to enable the transmission and sharing of information across the Internet (Barnes, 2003). Today, the global network that is the Internet houses and accesses enormous amounts of information and is the single largest information resource in existence.

The immense amount of information on the Internet requires shared communication protocols and necessitates software programs to search, retrieve, and manage information. Before the World Wide Web, such programs as File Transfer Protocol (FTP) and Gophers were used.

FTPs are programs for transferring information and public domain software from one computer to another. The host computer stores (archives) and maintains information, which can be obtained by any user using anonymous FTP. The program ARCHIE is a useful related service that can be used to search the contents of most FTP archive sites. Post-Web iterations of FTP include Netmanage (www.ftp.com). Gopher was developed by the University of Minnesota to find information on many different computers on the Internet, including public access databases, online books, news, and phone directories. The program VERONICA works with Gopher, providing the ability to do keyword searches (Sudweeks et al., 1995). Telnet is a software application that can be used with FTP and Gopher. It connects remote computers, generally through a telnet port. Users "telnet in" and execute commands remotely through a telnet interface (Telnet, 2003). Telnet's usefulness as an information management tool comes from its ability to obtain information from public access databases.

INTERNET ALTERNATIVES

To bypass the delays, slowdowns, and interruptions that plague the Internet, a number of alternative networks are in development. Access to these alternatives can expedite the working of virtual teams. Some of these alternatives are Internet II, National Science Foundation's very high performance Barebone Network Service (vBNS), the Department of Energy's Energy Sciences network (Esnet), and NASA's Research and Education Network (NREN). The alternative network that has received the most publicity is Internet II. This faster, more powerful network was set into motion in 1997 by 34 universities. Internet II was projected to run up to 1,000 times faster than the existing Internet, with correspondingly increased capacity and potentials for communication, information management, and audio and video capabilities.

WORLD WIDE WEB

One aspect of the Internet demands special consideration. The World Wide Web is the largest part of the Internet and globe-spanning (still growing), and the single largest source of information in existence. The Web is an Internet medium that uses hypertextual links, an "interconnected assortment of Internet computer servers that conform to the same network interface protocols" (Barnes, 2003, p. 11). Put more simply, it is "a collection of commercial, educational, and personal 'Web sites' that contain electronic pages of text and graphics"(Cozic, 1997, p. 6). The Web brings together to a hitherto unprecedented extent research, history, literature, art, science, news, and entertainment, as well as commercial information and personal Web pages, all in text and multimedia form. Through the use of search engines, virtual teams can access data, information, and records on almost any issue of interest. As of 2001, there were over 26,000,000 host sites on the Web (Barnes, 2003).

The Internet and the Web enable access to thousands of online databases, which have made collections of information available online. Library catalog systems (OPACs), specialized information (scientific, legal, medical, etc.), news services, business and academic databases, to name but a few, are all online. Library systems like the New York Public Library (http://www.nypl.org), newspapers (e.g., the *New York Times* at www.nytimes.com), and television news sources (e.g., CNN at www.cnn.com) all provide access to archived information online. Similarly, many universities make academic databases available to both account holders and nonaccount holders. The sheer number of databases available today makes a detailed consideration impossible here. One suggested resource for learning more about online databases is Radford, Barnes, and Barr's (2002) *Web Research: Selecting, Evaluating and Citing*.

The remarkable amount of information on the Web necessitates specialized software tools for information locating and retrieval. Browsers enable the finding, interpretation, and display of documents in hyptertext markup language (HTML) through a point-and-click function. The first text-based browser evolved into today's popular Netscape Navigator, with rival browsers such as Internet Explorer (Barnes, 2003). Miniature browsers in PDAs and cell phones make possible wireless access to the Web. Search engines are related software tools that help users locate and retrieve information on the Web. Search engines (i.e., Yahoo! and Alta Vista) and meta-search engines (i.e., Google, Dogpile, Profusion, Webcrawler, Metacrawler, and Ask Jeeves) are free to all users, with some competition among providers.

COMMERCIAL/PROPRIETARY GROUPWARE

The pivotal role commercial groupware plays in teams calls for a discussion, but so abundant are the proprietary groupware systems now available that a detailed consideration is beyond the scope of this article. Therefore, this section will provide a discussion somewhat limited in scope by considering representative commercial groupware from three perspectives: comprehensive groupware packages as provided by commercial giants; specialized and smaller providers; and course management software. Additional information on the host of other commercial software systems can be found on the Groupware yellow pages (http://www.csua.berkeley.edu/~mogul/groupware/) that provides links to academic and business resources, or through discussion on the Group Support Systems List.

COMPREHENSIVE PROPRIETARY GROUPWARE PACKAGES

Several software companies, both large and small, have groupware packages, some reaching back across three decades. In the category of software packages developed specifically as groupware are Lotus Notes and Domino (http://www.lotus.com), Novell Groupwise (http://www.novell.com/products/groupwise), Netscape Collabra (http://wp.netscape.com/collabra/v3.5/index.html), Fujitsu Teamware (http://www.teamware.com), and NCSA Habanero (http://www.isrl.uiuc.edu/isaac/Habanero/). Other software systems (i.e., Microsoft Exchange (http://www.microsoft.com/exchange/), Oracle Collaboration Suite (http://www.oracle.com/ip/deploy/cs/, and various Course Management Software) have larger functions, but can also be applied as collaborative tools.

Two representative software giants are Lotus and Microsoft. Lotus Notes and Domino are the most widely used groupware systems in terms of market penetration (Lotus, 2003). Microsoft Exchange is the major competitor. Both systems conveniently provide synchronous and asynchronous messaging, collaborative application development, information management through shared databases, data exchange, Web functions, and other tools essential to team functioning.

Specialized conferencing tools are a rapidly growing sector. For instance, one growing trend is the use of Web services or application service providers (ASPs) instead of comprehensive proprietary groupware packages. Web services and ASPs use the World Wide Web to allow for project collaboration among multiple partners and organizations (New Straits Times-Management Times, 2002). Alwang (2001) points out that "web-based collaboration provides a cost-effective solution without the headaches inherent with supporting complex in-house systems." Services like Thurport's HotOffice (http://www.hotoffice.com), OnProject's OnProject (http://www.onproject.com), NetDocuments Enterprise (http://www.netdocuments.com), Punch Web-Groups (http://www.punchnetworks.com), and TeamDrive (http://www.teamstream.com) all narrowly focus on document management, as do numerous other specialized services; whereas services like Entellium (http://www.entellium.com), eRoom (http://www.documentum.com), and BlueStep (http://www.bluestep.net) are broader in nature and offer several services and subservices, including communication services.

Other specialized conferencing tools include "a new generation of universally accessible collaborative environments built by developing and extending existing robust frameworks" (Narayan, 1997, p. 3.), such as the National Center of Supercomputing Applications' Habanero, or Vermics, which bridges course management software and commercial Groupware. A different example is eStudioLive, which specializes in a desktop Webcasting system and multimedia online presentations with interactive features.

COURSE MANAGEMENT SOFTWARE

In higher education, most of the commonly used instructional technology software packages contains features that are useful to teams. Although such software was created for educational use, primarily as course support, or as a vehicle for online courses, its uses go beyond the classroom. Team members with access to the software can find useful tools, both for groupware and other features.

The three most widely used course management systems are WebCT, Blackboard, and eCollege. All three packages offer team options comparable to comprehensive proprietary groupware, including messaging tools, conferencing tools, information management, and data resources. Messaging tools include e-mail to individuals and groups, with file attachment options, and some searchability. Discussion lists options are numerous, with posts possible in plain text, formatted text, or html, with attachments and URLs. Threaded discussions can be viewed by date, thread, and topic. Conferencing tools are provided with Chat and internal groups. Private chat rooms can be created, and all chats are archived. Internal small groups can be created with their own whiteboards, discussion forum, and synchronous tools. Information management and data resource options include file exchange, with options to upload files to a private or shared group folder. Whiteboards also exist, with image and PowerPoint uploading and group Web browsing. Most information, including whiteboard sessions, is archivable (Edutools, 2003).

Although lacking in some of the features of the three packages considered above, free instructional technology software programs are available (e.g., ThinkWave Educator, ClassBuilder, NgBook, and Class Information Manager).

CONCLUSION

"As businesses become more interconnected and more global, they must learn to make faster and smarter strategic decisions, and to take advantage of technological advancements" (Solomon, 2001). Teams today need to take advantage of all available technological resources. We hope that this survey of the existing support systems and technologies will assist teams, both traditional and virtual, to maximize their effectiveness, development, and growth, and to minimize problems and errors that can arise when working with technology.

REFERENCES

Alwang, G. (2001, October 30). Web collaboration [Electronic version]. *PC Magazine*. Retrieved May 23, 2003 from *http://www.pcmag.com/article2/0,4149,71267,00.asp*

Barnes, S.B. (2003). *Computer-mediated communication*. New York: Allyn & Bacon.

Business Wire. (2002, July 29). Vermics breaks new ground with distance learning solution enabling onsite delivery of live, interactive, instructor-led courses. Retrieved July 19, 2003 from Lexis-Nexis database.

Business Wire. (2002, October 21). Avistar launches vBrief video email messaging product. Retrieved July 19, 2003 from Lexis-Nexis database.

Comparison of course management software [Electronic version](2003). *Edutools*. Retrieved July 19, 2003 from *http://www.edutools.info/course/productinfo/index.jsp*

Computer Weekly. (2003, March 20). Cashing in on the SMS phenomenon [Electronic version]. *Computer Weekly*. Retrieved July 19, 2003, from EBSCOHost.

Cozic, C.P. (1997). Introduction. In *The future of the Internet*. San Diego: Greenhaven Press.

Devi, C. (2003, March 17). Trends from text messaging culture [Electronic version]. *New StraitsTimes – Management Times*. Retrieved July 19, 2003, from EBSCOHost.

Ferris, S.P. (1995). *An investigation of a role of computer-mediated communication as a media choice in the facilitation of task performance in small groups*. Unpublished dissertation.

Habanero news release (2003). Retrieved July 19, 2003 from *http://www.ncsa.uiuc.edu/News/Access/Briefs/00Briefs/000328.Habanero.html*

Herman, J. (2003, April). Blogs for business [Electronic version]. *Business Communication Review*. Retrieved May 5, 2003, from EBSCOHost.

Lamb, S.E. (1999, October-December). E-mail and online communications [Electronic version]. *Business and Economic Review*. Retrieved May 3, 2003, from EBSCOHost.

Metz, C. (2001, November 27). Virtual meetings [Electronic version]. *PC Magazine*. Retrieved May 3, 2003 from *http://www.pcmag.com/article2/0,4149,13339,00.asp*

Mills, K. (2003, February 25). 3G unlikely to deliver for business, Deloitte predicts [Electronic version]. *The Australian*. Retrieved July 19, 2003, from EBSCOHost.

Radford, M.L., Barnes, S.B., & Barr, L.R. (2002). *Web research: Selecting, evaluating and citing*. New York: Allyn and Bacon.

Schwarz, M. (2002, January 7). The instant messaging debate [Electronic version]. *Computerworld*. Retrieved May 3, 2003, from EBSCOHost.

Solomon, C.M. (2001, June). Managing virtual teams [Electronic version]. *Workforce*. Retrieved May 3, 2003, from EBSCOHost.

Sudweeks, F., Collins, M., & December, J. (1995). Internetwork resources. In Z. L. Berge & M. P. Collins (Eds.), *CMC and the online classroom: Volume 3- Distance Learning* (pp. 193-212). Cresskill, NJ: Hampton Press.

Telnet. (2003). Telnet tech support. Retrieved April 7, 2003, from http://telnet.org

TOPCALL helps businesses discover the value of text; Research reveals the top ten applications that are paving the way for SMS use in critical enterprise communications [Electronic version] (2003, June 10). *Business Wire*. Lexis-Nexis Database.

Web services gain popularity [Electronic version] (2002, March 18). *New Straits Times-Management Times*. Retrieved May 3, 2003, from EBSCOHost.

Yap, C. (2003, March 11). The SMS tease and please [Electronic version]. *New Straits Times – Management Times*. Retrieved July 19, 2003, from EBSCOHost.

Young, J.R. (1997, October 17). Demonstration gives lawmakers and researchers a taste of Internet 2. *Chronicle of Higher Education, A28*.

KEY TERMS

Commercial Groupware Packages: Proprietary software that provides bundled software, including synchronous and asynchronous messaging, collaborative application development, information management through shared databases, data exchange, Web functions, and other tools.

Conferencing Systems: Conferencing systems are specifically designed to facilitate synchronous virtual meetings by phone or computer. Teleconferencing utilizes computer-controlled audiovisual transmission. Computer conferencing uses computer-enabled conferencing to work together in real-time using free proprietary software.

Course Management Software: Instructional technology software packages created for educational use, primarily as course support or as a vehicle for online courses. Groupware features include messaging tools, conferencing tools, and information management and data resources.

Group Support Systems (GSS): Computer software and hardware used specifically to support group functions and processes. GSS can be defined as tools for group communication, process, and task achievement. It should be noted that GSS often comprise technologies that are not necessarily limited to computers and are often used in everyday life, such as telephones and videoconferencing. Non-computer and computer technologies may be determined to be GSS if they possess one of the following characteristics: it fosters collaboration between people; it fosters the sharing of information; or it enables the communication between groups of people (Burns, 1995, in Brusic, 2003, p .13).

Information-Exchange/Data-Management Systems: Systems designed for finding, exchanging, and managing information, and sharing data and information resources. They include the Internet (a globe-spanning system of networked computers, utilized packet switching and communication protocols), and the World Wide Web, an Internet medium that uses hypertextual links (Barnes, 2003).

Messaging Systems: Messaging systems enable and facilitate communication among team members. Communication among team members can be one-to-one, one-to-many, and synchronous (IM and Chat) or asynchronous (e-mail, bulletin boards, discussion lists).

Team: A collection of four to 12 individuals collaboratively working on a common and interdependent task or goal. The goal is often one requiring a decision or a solution to some problem. The elements of common tasks, goals, and interdependence are integral to our definition of a team, at least with respect to an imposed need to arrive at a collective position on a matter under consideration.

ENDNOTES

1 We use the term *team* in this book in the commonly accepted sense; that is, a team refers to a collection of four to twelve individuals collaboratively working on a common and interdependent task or goal. The goal is often one requiring a decision or a solution to some problem. The elements of common tasks/goals and interdependence are integral to our definition of a team, at least in respect to an imposed need to arrive at a collective position on a matter under consideration.

2 Note also that the technologies discussed here are current as of spring 2004. Technology rapidly changes and grows, and we have no doubt that within the next year new developments will emerge.

3 Burns, N. (1995). Groupware: Myths and realities. In I. Brusic. Groupware classification schemes. *Human-Computer Interaction CS6751*.

Technological Collaboration and Trust in Virtual Teams

Steven A. Morris
Middle Tennessee State University, USA

Thomas E. Marshall
Auburn University, USA

R. Kelly Rainer
Auburn University, USA

INTRODUCTION

A hypercompetitive environment of global competition, reduced product cycle times, and rapid change has forced organizations to seek new structures and practices that allow them to react with speed and agility. One structure that has gained increasing support among organizations seeking to improve organizational agility has been the virtual team (Mowshowitz, 1997). Virtual teams are the application of virtual organization concepts to team environments. Virtual teams use information technology (IT) to remove barriers of time, location, and organizational boundaries to allow the integration of skills through collaborative relationships based on trust and professionalism to meet specific objectives. Despite the growing popularity of virtual teams among businesses, more needs to be known about the nature of these teams (Bell & Kozlowski, 2002). For example, some research has found virtual teams to outperform traditional face-to-face teams (e.g., Majchrzak, Malhotra, Stamps & Lipnack, 2004; Schmidt, Montoya-Weiss & Massey, 2001), while other research indicates otherwise (Cramton, 2002). It is important that organizations understand the implications of these virtual teams so that they may be used in the most effective manner, and under the most appropriate circumstances.

BACKGROUND

Virtual teams appear to exhibit many of the same characteristics of virtual organizations. These common characteristics include opportunism, excellence in technology, excellence in skills, borderlessness, and reliance on trust (Goldman, Nagel & Preiss, 1995). Opportunism indicates that the virtual team is created to meet a specific need or leverage a specific opportunity. Once the objective has been met, the virtual team is typically disbanded (Jarvenpaa, Knoll & Leidner, 1998). While it is the nature of virtual teams to be temporary, they can exist as long as the opportunity that they were created to exploit exists, which can result in practical applications of virtual teams that last for decades. Excellence in technology is achieved by the contribution of the best technologies ("best" in this case meaning the most appropriate for the needs of the team) from each of the virtual team members. This contribution of technologies is not only critical to the functioning of the virtual team, but is also of interest to business researchers as virtual teams place a greater reliance on technology for the completion of team tasks and activities than previously experienced. Excellence in skills refers to the manner in which participants in the virtual team are selected. Typically, each virtual team member is selected to participate on the team because that member has, in a high degree, a skill that will be needed by the team. Borderlessness refers to the manner in which virtual teams tend to cross traditional, functional, and even organizational boundaries. Often, individuals performing in these teams may be internationally dispersed and include constituents from other firms. These geographically and organizationally dispersed teams are made viable through the use of information technology (Dube & Pare, 2001; Townsend, DeMarie & Hendrickson, 1998). The interorganizational and international aspects of the virtual team interfere with, and often invalidate, traditional mechanisms for control of team members. Dube and Pare (2001) report that global virtual teams rarely, if ever, meet in a face-to-face environment. As a result, trust is believed to become a critical component in effective virtual team operation (Larsen & McInerney, 2002; Jarvenpaa & Leidner, 1998).

The reliance on trust is dictated due to the lack of authoritative controls found in virtual structures. In place of traditional command-and-control authority structures, virtual teams rely on trust and professionalism to create interdependent, collaborative relationships. Given the importance of establishing and maintaining control with the organization, it is understandable that the departure from traditional control structures to a trust-based environment has generated interest among researchers and practitioners alike. Unlike the other characteristics of virtual structures, such as duration, and geographical and organizational dispersion that can be directly observed and controlled, the reliance on trust as a characteristic of virtual teams is a logical conclusion that is being explored in the research literature (Majchrzak, Rice, Malhotra, King & Ba, 2000). Researchers have already recognized that the unique characteristics of virtual teams will require the re-examination of existing theories and have called for additional research into areas such as trust, organizational context, and team norms (Bell & Kozlowski, 2002; Majchrzak et al., 2000; Townsend, DeMarie & Hendrickson, 1998; Warkentin, Sayeed & Hightower, 1997).

TRUST IN VIRTUAL TEAMS

Trust has frequently been assumed to be one of the key ingredients necessary for a virtual organization or team to be successful (Suomi, 1988; Konsynski, 1993; Duffy, 1994; Handy, 1995; Cohen, 1997). Currall and Judge (1995) defined trust as "an individual's behavioral reliance on another person under a condition of risk" (p. 153). Empirical research into the role of trust in the virtual team setting, however, is just emerging.

The role of trust, or the lack of trust, in human behavior as it pertains to the development of organizations and teams can be clarified using transaction cost economics (TCE) (Williamson, 1975). TCE explains the development and growth of organizations and teams as a solution to the costs of conducting transactions in a market. The principle components of TCE are opportunism, bounded rationality, small-numbers bargaining, and uncertainty. Opportunism is the behavioral tendency of an entity, either an individual, a team, or an organization, to act in its own self-interest, even at the expense of another entity. Bounded rationality describes the limitations placed on an entity in determining behaviors that are in its self-interest due to limited or imperfect knowledge. Small-numbers bargaining refers to a shift in negotiating power that occurs when only a small number of entities in a market can meet the needs of another entity. Uncertainty is defined as the inability to predict relevant future events (Williamson, 1975).

Transaction cost economics posits that there are costs associated with conducting transactions in a market, such as searching costs, transportation costs, and communication costs (Gurbaxani & Whang, 1991). These costs are complicated by the presence of uncertainty so that entities have difficulty in predicting future demand and supply. To reduce transaction costs, entities may create standing relationships with each other in the form of contracts. Even though transaction costs are associated with the creation and enforcement of the contract, use of the contract may still be more economical than transactions in a market. Because each entity in the relationship has limited knowledge of the activities of the other entity, the potential for opportunistic behavior still exists. That is, the entity is restricted by its bounded rationality, so its transaction costs increase while the other entity enjoys greater profit. To prevent this opportunistic behavior, the contract must account for contingencies, causing an increase in the costs of writing and enforcing the contract (Perrow, 1986; Gurbaxani & Whang, 1991). Opportunistic behavior may also occur if either entity in the relationship must deal with uncertainty in a fluctuating industry. To prevent opportunistic behavior, contractual costs are increased as contingencies are taken into account.

During the course of the relationship, entities make investments in the relationship, such as time spent learning procedures and the development of communication channels. If these investments are significant, they can be a catalyst for small-numbers bargaining at the time of renegotiations. Small-numbers bargaining allows one entity to act opportunistically against the other to achieve greater profit, while still engaging in a relationship that is more economical to both entities than leaving the relationship.

In order to reduce the costs of opportunistic behavior, and the costs of protecting against opportunistic behavior, one party of the relationship may assume authoritative control over the other through acquisition. While opportunism, bounded rationality, uncertainty, and small-numbers bargaining still exist as influences within the organization created by that acquisition, they can be curbed through the exercise of authority.

In terms of transaction cost economics, virtual organizations and teams represent the transition from large hierarchical structures to flatter, collaborative structures (Drucker, 1988). By losing the authoritative control of the hierarchical structure, the members of the virtual organization and team must deal with greater risks of opportunism from the individuals that they must collaborate with and rely upon. Researchers have proposed that the environment created by a virtual structure will force its members to rely more heavily on trust instead of relying on control structures to ensure the performance of others (Clemons & Row, 1992; Konsynski, 1993; Bleecker, 1994; Handy, 1995; Barner, 1996; Cohen, 1997).

In addition to the application of TCE to the issue of trust, research efforts have begun investigating trust in virtual teams. Larsen and McInerney (2002) looked at inter-university virtual teams developing information products and found that team performance was closely related to issues of trust. Lurey and Raisinghani (2001) found that relations among virtual team members had a very strong relationship to virtual team performance and virtual team member satisfaction. Interestingly, Lurey and Raisinghani (2001) found little relationship between internal group dynamics, and team performance and team member satisfaction. This is in direct contrast to Montoya-Weiss, Massey, and Song (2001), who found that management of internal conflict was critical to virtual team performance. Jarvenpaa, Knoll, and Leidner (1998) found that the development of trust is different from the development of trust in traditional face-to-face teams. They suggested a model for the development of "swift" trust, whereby the virtual team members assume an attitude of trust instead of engaging in traditional trust-building. Markus (2004) found interpersonal issues such as trust to be the greatest challenge facing virtual teams.

TECHNOLOGY IN VIRTUAL TEAMS

Researchers have long theorized about potential impacts, such as the elimination of middle management levels, that information technology could have on organization structures (Leavitt & Whisler, 1958). Advances in IT have supported the development of IT-enabled organization design variables that were not previously available. Table 1 summarizes findings from an exploration of IT-enabled design variables that organizations can use to supplant traditional design variables in the design of new organization forms, such as virtual organizations and virtual teams (Lucas & Baroudi, 1994).

The IT-enabled nature of virtual teams causes employees to rely heavily on IT to complete and coordinate tasks (see e.g., Legare, 2001; Barnatt, 1997). Therefore, the degree to which the system meets the employee's information needs is of critical importance. Markus (2004) investigated the IT tools used by virtual team members. She reported virtual team members had to employ a number of unintegrated tools since there was no single IT that met all of the team members' needs. However, the virtual team members overwhelmingly indicated the vital nature of IT for virtual team performance. Suchan and Hayzak (2001) reported in their case study of the use of virtual teams in a Fortune 500 organization that team members' attitude toward the technology was vital to the success of the virtual team structure. Larsen and McInerney (2002) found that virtual team members' perceptions of the virtual team experience were influenced by their perceptions of the technology used to implement the virtual structure. Malhotra, Majchrzak, Carman, and Lott (2001) conducted a case study of virtual teams at Boeing-Rocketdyne. They reported that the technology used by the virtual team was so critical to the virtual team members that 23 versions of the technology, based on recommendations of the team members, were developed throughout the life of the virtual team, and a dedicated technology facilitator was deployed to each teleconferencing session.

FUTURE TRENDS

Despite the many unanswered questions about virtual structures, it is clear that market demands for agile competition will continue to push organizations to seek out and exploit these types of structures. Initial research has

Table 1. Conventional and IT-enabled design variables (Adapted from Lucas & Baroudi, 1994)

Class of Variable	*Conventional Design Variable*	*IT-Enabled Design Variable*
Structural	Definition of organizational subunits	Virtual components
	Linking mechanisms	Electronic linking
	Staffing	Technological leveling
Work Process	Tasks	Production automation
	Workflows	Electronic workflows
	Buffers	Virtual components
Communications	Formal channels	Electronic communications
	Informal channels/collaboration	Technological matrixing
Interorganizational Relations	Make versus buy decision	Electronic customer/ supplier relationships
	Exchange of materials	Electronic customer/ supplier relationships
	Communications mechanisms	Electronic linking

Table 2. Trust and technology issues in virtual teams

- What factors influence the development of trust among virtual team members?
- How can trust be strengthened among virtual team members?
- Are there any viable substitutes for trust?
- What roles do cultural, national, and organizational diversity play in establishing and maintaining trust?
- What types of authoritative control structures are still viable and appropriate in a virtual team setting?
- What technologies best support virtual teams?
- What is the impact of the intensified reliance on technology within a virtual team setting?
- Can information technology be used to create systems to provide a substitute for trust?
- What impact does the mandatory use of IT in a virtual team have on team member motivation and job satisfaction?
- How does one assess the success of an information system used to support a virtual team?

shed some light on the complex nature of virtual teams. However, a great deal of future research will be necessary before the full impacts of these structures on the organization, and on the individuals involved, will be understood. Empirical research has supported some of the previously untested theories about virtual teams, and the roles that information technology and trust play within them. Other theories have not been supported, and these findings continue to give birth to additional theories and hypotheses. Future research in this area will have to find the missing pieces of the puzzle that explain the equivocal results of existing research on the performance expectations for virtual teams (Cramton, 2002; Majchrzak et al., 2004). The nature of trust in virtual teams has been shown to differ from traditional teams both in its formation and its role in shaping virtual team member expectations; however, the exact nature of the differences and why the differences exist is still under investigation. While the impact of reliance on information technology is not a new subject in the research literature (e.g., Lucas & Baroudi, 1994), the increased dependency on information technology in a virtual team setting raises several issues regarding the impact of IT on individuals, best practices for IT usage in a virtual team setting, and task interdependency. Future research will have to address these issues and many more. Table 2 lists a small sample of the questions that organizations need answered that researchers are investigating.

CONCLUSION

Virtual teams promise to help organizations improve their agility in responding to an ever-changing marketplace that demands rapid response. This agility, however, comes with a price: organizations are committing themselves to a structure whose dynamic is still largely unknown. Two areas of interest in the efforts to understand virtual teams are the issues of reliance on trust and reliance on information technology. Research and practical experience have both shown that the nature of trust is different in virtual teams, and that virtual team structures require a new level of intensity of reliance on IT. The consequences of these changes are not yet understood. Jarvenpaa et al. (1998) noted a difference in trust development that researchers are exploring. Current research is concerned with better understanding these differences in trust development, and the consequences that it may have for the maintenance of trust. This research makes it important to reexamine many of the more established trust-related concepts, such as cultural differences in trust. Other researchers are investigating the role that IT currently plays in virtual teams, as well as exploring new roles that IT could play to improve virtual team performance.

REFERENCES

Barnatt, C. (1997). Virtual organization in the small business sector: The case of Cavendish Management Resources. *International Small Business Journal, 15*(4), 36-47.

Barner, R. (1996). New millennium workplace: Seven changes that will challenge managers¾and workers. *The Futurist, 30*(2), 14-18.

Bell, B.S. & Kozlowski, S.W. (2002). A typology of virtual teams: Implications for effective leadership. *Group & Organization Management, 27*(1), 14-49.

Bleecker, S.E. (1994). Virtual organization. *The Futurist, 28*(2), 9-14.

Clemons, E.K. & Row, M.C. (1992). Information technology and industrial cooperation: The changing economics of coordination and ownership. *Journal of Management Information Systems*, *9*(2), 9-28.

Cohen, S. (1997). On becoming virtual. *Training & Development, 51*(5), 30-37.

Currall, S.C. & Judge, T.A. (1995). Measuring trust between organizational boundary role persons. *Organizational Behavior and Human Decision Processes*, *64*(2), 151-170.

Driscoll, J.W. (1978). Trust and participation in organizational decision making as predictors of satisfaction. *Academy of Management Journal, 21*(March), 44-56.

Drucker, P.F. (1988). Coming of the new organization. *Harvard Business Review,* (January-February), 45-53.

Dube, L. & Pare, G. (2001). Global virtual teams. *Communications of the ACM, 44*(12), 71-73.

Duffy, M. (1994). Ten prescriptions for surviving and thriving in the virtual organization. *Public Relations Quarterly, 39*(2), 28-31.

Goldman, S.L., Nagel, R.N. & Preiss, K. (1995*). Agile competitors and virtual organizations' strategies for enriching the customer*. New York: Van Nostrand Reinhold.

Gurbaxani, V. & Whang, S. (1991). Impact of information systems on organizations and markets. *Communications of the ACM, 34*(1), 59-73.

Handy, C. (1995). Trust and the virtual organization. *Harvard Business Review,* (May-June), 40-50.

Jarvenpaa, S.L., Knoll, K. & Leidner, D.E. (1998). Is anybody out there? Antecedents of trust in global virtual teams. *Journal of Management Information Systems, 14*(4), 29-64.

Jarvenpaa, S.L. & Leidner, D.E. (1998). Communication and trust in global virtual teams. *Journal of Computer-Mediated Communication, 3*(4). Retrieved from *www.ascusc.org/jcmc*.

Konsynski, B.R. (1993). Strategic control in the extended enterprise. *IBM Systems Journal, 32*(1), 111-142.

Larsen, K.R.T. & McInerney, C.R. (2002). Preparing to work in the virtual organization. *Information & Management, 39*(6), 445-456.

Leavitt, H. & Whisler, T. (1958). Management in the 1980s. *Harvard Business Review,* (November-December), 41-48.

Legare, T.L. (2001). How Hewlett-Packard used virtual cross-functional teams to deliver healthcare industry solutions. *Journal of Organizational Excellence, 20*(4), 29-38.

Lucas, H.C. & Baroudi, J. (1994). The role of information technology in organization design. *Journal of Management Information Systems, 10*(4), 9-23.

Lurey, J.S. & Raisinghani, M.S. (2001). An empirical study of best practices in virtual teams. *Information & Management, 38*(8), 523-544.

Majchrzak, A., Malhotra, A., Stamps, J. & Lipnack, J. (2004). Can absence make a team grow stronger? *Harvard Business Review, 82*(5), 131.

Majchrzak, A., Rice, R.E., Malhotra, A., King, N., & Ba, S. (2000). Technology adaptation: The case of a computer-supported inter-organizational virtual team. *MIS Quarterly, 24*(4), 569-600.

Malhotra, A., Majchrzak, A., Carman, R. & Lott, V. (2001). Radical innovation without collocation: A case study at Boeing-Rocketdyne. *MIS Quarterly, 25*(2), 229-249.

Markus, M.L. (2004). Reality of virtual teams. *Optimize*, (April), 69.

Montoya-Weiss, M.M., Massey, A.P. & Song, M. (2001). Getting it together: Temporal coordination and conflict management in global virtual teams. *Academy of Management Journal, 44*(6), 1251-1262.

Mowshowitz, A. (1997). Virtual organization. *Communications of the ACM, 40*(9), 30-37.

Perrow, C. (1986). *Complex organizations: A critical essay* (3rd edition). New York: Random House.

Rich, G.A. (1997). The sales manager as a role model: Effects on trust, job satisfaction, and performance of salespeople. *Journal of the Academy of Marketing Science, 25*(4), 319-328.

Schmidt, J.B., Montoya-Weiss, M.M. & Massey, A.P. (2001). New product development decision-making effectiveness: Comparing individuals, face-to-face teams, and virtual teams. *Decision Sciences*, 32(4), 575-600.

Suchan, J. & Hayzak, G. (2001). The communication characteristics of virtual teams: A case study. *IEEE Transactions on Professional Communication, 44*(3), 174-186.

Suomi, R. (1988). Inter-organizational information systems as company resources. *Information & Management, 15*(2), 105-112.

Townsend, A.M., DeMarie, S.M. & Hendrickson, A.R. (1998). Virtual teams: Technology and the workplace of the future. *The Academy of Management Executive, 12*(3), 17-29.

Warkentin, M., Sayeed, L. & Hightower, R. (1997). Virtual teams versus face-to-face teams: An exploratory study of a Web-based conference system. *Decision Sciences, 28*(4), 975-996.

Williamson, O. (1975). *Markets and hierarchy: Analysis and antitrust implications*. New York: The Free Press.

KEY TERMS

Design Variables: Characteristics of an organization, its processes, control, and coordination structures that can be varied to produce a specific organization design.

IT-Enabled: Indicating that a design variable or team can only function through the use of information technologies.

Opportunism: In terms of virtual teams, the characteristic of existing to capitalize on a specific opportunity or meet a specific, temporary need. In terms of TCE, the characteristic of an entity that indicates the entity will act in its own best interest even at the expense of other entities in the market.

Organizational Agility: An organization's ability to respond quickly and appropriately to changes in the organization's environment to solve problems and capitalize on opportunities.

Transaction Cost Economics (TCE): An economic paradigm of organization analysis that explains the development of organizations and organizational hierarchies as a function of the costs of conducting transactions in a market.

Trust: An entity's willingness to rely on another entity under a condition of risk.

Virtual Team: A team constructed of collaborative relationships supported by information technology to overcome boundaries of time, location, and organizational boundaries.

Technology and Knowledge Management

Helen Mitchell
Unitec, Auckland, New Zealand

INTRODUCTION

Rapid and extensive advances in technology, particularly in the area of communications, have had a considerable impact on the way organizations operate and opened pathways to access vast amounts of information. Information, however, is static unless knowledge is applied to translate it into something meaningful and with the potential to be actionable. From the time organizations commence business, they accumulate information about the markets in which they operate, yet often, knowledge is not applied in a way that it can be exploited to bring benefit. The ability to share knowledge, to develop ideas, and to become more innovative is increasingly important for businesses, and the range of technologies now available provides a conduit for knowledge to flow through the organization to enable sharing to occur. Technology is frequently referred to as "just an enabler," but it can also be identified as a value-adder.

INFORMATION AND KNOWLEDGE

In their paper, Evans and Wurster (1997, p. 71) referred to changes that had taken place over the previous 10 years as organizations adapted their "operating processes" to "information technologies," recognizing that accessing information was going to have an important bearing on where industries would be going in the future. During this period, technology was moving forward at a rapid rate, and organizations were investing huge sums of money in information technology.

Information is defined as facts and data organized to describe a particular situation or problem. The definition used for knowledge is that by Davenport and Prusak (1998):

Knowledge is a fluid mix of framed experience, values, contextual information, and expert insight that provides a framework for evaluating and incorporating new experiences and information. It originates and is applied in the minds of knowers. In organizations, it often becomes embedded not only in documents or repositories but also in organizational routines, processes, practices, and norms. (p. 5)

Connectivity, it is suggested by Evans and Wurster (1997), provided the most important change in the information revolution: "What is truly revolutionary about the explosion in connectivity is the possibility it offers to unbundle information from its physical carrier" (p. 73). Amidon (1997) referred to the major change from information processing to knowledge processing that has taken place, "which includes the concepts of learning tools, intelligent electronic coaching, decision-making systems, and more" (p. 87). The availability of such tools, and their ongoing development, add considerable value to an organization by providing a means through which more efficient and effective management techniques can be introduced. This evolving of expectations indicates that organizations are not only anticipating more from technology but are also becoming more reliant upon it.

As interest grows beyond information per se, organizations are looking to technology to progress toward the development of knowledge management systems. Bhatt (2001) referred to business managers who believe computers and communication technologies provide an avenue for the harvesting of knowledge from data repositories of organizations, while other managers say knowledge "resides in human minds, and therefore, employee training and motivation are the key factors to knowledge management" (p. 68). The development of technology has taken the drudgery out of searching, analyzing, and converting data into information to which knowledge can be applied.

Technology and knowledge, however, do not stand in isolation. There are many interacting factors, not least of which is the environment in which the organisation operates. According to Bhatt (2001), the "…pattern of interaction between technologies, techniques and people is unique to an organization" (p. 69). Such uniqueness is important, because it is not easy to replicate. The organisation that promotes the value of the knowledge, skills, and competencies of its people, and recognizes the importance of technology, is providing well for its future (Carneiro, 2000; Bhatt, 2001).

From a productivity perspective, Grant (2000) indicated the value of digital technology. He referred to knowledge no longer being held exclusively by people. Codification and use of technology provides the opportunity for knowledge replication. While costly to create, replication and distribution can be reduced to almost "zero marginal cost" (Grant, 2000, p. 32). In the long term, and with the arrival of new technologies, the rate of productivity growth is likely to accelerate (Grant, 2000). People have skills and knowledge, but it is technology that enables them to access in a timely manner huge amounts of data available both internally and externally to the organization. While technology enables people to access data, the added value comes in its ability to allow the data to be assembled in a meaningful way for the purpose of decision making.

SHARING KNOWLEDGE

An environment in which knowledge sharing is encouraged leads to the creation of new knowledge, but as Marshall, Prusak, and Shpilberg (1996) indicated, it is not easy to encourage voluntary sharing of knowledge by employees. An organisation that develops a knowledge-sharing environment increases opportunities for the creation of new ideas that have the potential to add value to it.

Barriers to sharing knowledge exist, and for managers advocating knowledge sharing, they should examine the existing practices within the organization. It is possible that somewhere in the organization, the "knowledge is power" syndrome exists, and if so, it does not auger well for knowledge sharing. There may also be concern over job security, "if I share my knowledge, I am no longer of value to the organization, and therefore, my job will be in jeopardy." A barrier to knowledge sharing may simply rest on the premise that a person may not realize that knowledge acquired in another situation could be of value to the organization. Managers may need to look to changing the organizational culture to one that recognizes, and acknowledges, the value of shared knowledge.

Bhatt (2001), referring to the need for the distribution and sharing of knowledge throughout the organisation, suggested that interactions among technologies, techniques, and people have a direct bearing on the effectiveness of the distribution of knowledge.

RANGE OF TECHNOLOGIES

According to Frappaola and Capshaw (1999): "Knowledge management refers to the practices and technologies that facilitate the efficient creation and exchange of knowledge on an organizational level" (p. 44).

From the literature, it appears that there are a number of applications specifically designed for sharing knowledge (Davenport, 1997; LaPlante, 1997; Fahey & Prusak, 1998). The benefits of applications, such as groupware, data warehousing, knowledge maps, client server systems, and chat rooms, are recognized as providing the means through which knowledge can be shared. A useful aspect of such applications is that they allow for interaction and sharing of information that is highly unstructured in nature (Shani, Sena, & Stebbins, 2000). Research carried out by Mitchell in 1999 indicated that people are finding groupware applications useful mechanisms for sharing knowledge.

A number of writers, Allee (1997), Amidon (1997), Marshall (1997), Watt (1997), and Davenport and Prusak (1998), refer to the intranet as providing channels through which organizational knowledge can flow. E-mail is now well accepted as a useful and valuable medium of communication, and for knowledge to flow through an organization (Bontis, Fearon, & Hishon, 2003). The advent of electronic conferencing provides an avenue through which people geographically dispersed can interact without the time and cost involved in traveling. Networking, bulletin boards, and the establishment of virtual teams also provide opportunities for people to interact over distances and to share knowledge.

Reference is made by Beckett (2000) to the "Data warehousing knowledge set" containing reference data, defined as market trends, operational data, and customer performance needs. From this knowledge set, actions can be taken that ultimately benefit stakeholders and customers. While the knowledge set may be considerable, ideally, its value only becomes realistic when there is in place the means to allow for the "free" flow of knowledge throughout the organisation and when there are people who can recognize that value can be gained from it. It is most unlikely that organizations allow for a free flow of knowledge, because issues such as privacy and confidentiality, "need to know," and other constraints restrict access to knowledge. However, providing access to knowledge that can be made available should be the aim of organizations wanting to encourage a knowledge-sharing environment.

Technology has opened avenues to the customer to search for opportunities and products that may better serve their needs. The opportunity, however, is also open to the organization to exploit the potential of technology to create greater value for its customers.

KNOWLEDGE MANAGEMENT STRATEGY

Hansen, Nohria, and Tierney (1999) identified management consultants as being among the first to recognize the potential of information technology for the capture and dissemination of knowledge. It was found that organizations tend to employ two very different strategies for the management of knowledge. One is the "codification strategy," where maximizing the use of technology shows a strong focus on codifying knowledge and storing it in databases for access by anyone in the company. The second approach viewing knowledge being shared through person-to-person contacts is identified as the "personalization strategy" (Hansen, Nohria, & Tierney, 1999, p. 107).

From their research, Hansen et al. (1999) found that organizations pursuing "an assemble-to-order product or service strategy emphasized the codification and reuse of knowledge. Those that pursued highly customized service offerings, or a product innovation strategy, invested mainly in person-to person knowledge sharing" (p.112). Whichever strategy is used, there is inevitably the inclusion, as a support mechanism, of an element of the other strategy. Hansen et al. suggest an 80:20 split, that is, "80% of their knowledge sharing following one strategy, 20% the other." They go on to say that "Executives who try to excel at both strategies risk failing at both" (p. 112).

In his article, Beckett (2000) referred to the fastest knowledge-transfer vehicles, relating them to the work of Nonaka and Takeuchi (1995). The tacit/tacit transfer is fast, "because people use multiple sensors in working together in teams" (p. 317). Using electronic media for the transfer of explicit/explicit transfer can also be fast. However, there is evidence that "tacit/explicit transfers are slow, as specialist skills are needed to draw out and carefully enunciate tacit knowledge" (p. 317). Becket indicated that explicit/tacit transfers are similarly positioned, and provides the example of formal education.

Giving consideration to the perspectives of Hansen et al. (1999) and Beckett (2000), and others, such as Ericsson and Avdic (2003), Mockler and Dologite (2002), Jentzsch and Prekop (2002), and Frappaolo and Capshaw (1999), determining a knowledge management strategy needs careful thought if it is to meet the requirements of the organization.

TECHNOLOGY: ENABLER OR VALUE-ADDER?

Lloyd (1996) referred to the use by organizations of worldwide networks and said that new technology is "the catalyst which is forcing all organizations to re-evaluate what they know, what they do with that knowledge, and how they continually add value (or not) to that knowledge in meeting changing customer needs" (p. 576). Technology has advanced considerably since Lloyd made this comment, and it will continue to evolve in the years ahead, providing for greater enrichment of organizational operations. While the cost of "keeping up" with technological developments has always been a problem for organizations that have made a strong commitment to technology, others recognize that they need to work smarter with what they have.

If technology is just an enabler, what is it that adds value to the organisation? Binney (2001, p. 33) addressed the question of knowledge-management (KM) investments. What has emerged is the KM spectrum, in which he identifies KM applications and places them into "six common categories to establish the elements of the KM spectrum" (p. 34). Binney identified from the literature six categories to provide the elements of the KM spectrum: transactional, analytical, asset management, process, developmental, and innovation and creation. He then mapped KM applications to the elements:

1. Transactional—Order entry applications; help desk applications
2. Analytical—Data warehousing, customer relationship management, DSS, MIS
3. Asset management—Document management, intellectual property
4. Process—TQM, process automation, benchmarking
5. Developmental—Skills development, training
6. Innovation and creation—Communities, virtual teams, networking

The next stage of the process added to the various elements, enabling technologies:

1. Transactional—Expert systems, probability networks
2. Analytical—Intelligent agents, data analysis and reporting tools
3. Asset management—Document management tools, knowledge maps
4. Process—Workflow management, process-modeling tools
5. Developmental—Computer-based training, online training
6. Innovation and creation—Search engines, voice mail, groupware

The KM spectrum provides organizations with the means to identify their present positions and to make use of the framework to map their future investments in knowledge management. The examples given above illustrate the wide range of applications available to organizations, and their potential and value emerges through maximizing their use for the purpose of effective decision-making in an increasingly competitive environment.

Technology has provided the impetus for the growth of the information age, but it should not be regarded as a dominant partner. As Pemberton (1998) commented, "The IT exponents of KM tend to downplay the central role of human factors in KM…IT doesn't itself create knowledge any more than does a library, an office, or a classroom" (p.60). But, as Carneiro (2000) indicated, the combining of knowledge and information technology is a major success factor in strategic planning formulation. The future will bring ever more sophisticated advancements in technology providing new avenues of exploration for seeking, creating, and sharing knowledge.

CONCLUSION

Is technology just an enabler, or is it also a value-adder? From the literature, technology, while recognized as being important, tends to be regarded as an enabler rather than a value-adder. Yet the continual movement in technological progress, as shown in Binney's KM spectrum, clearly identifies the developments that have taken place in technology to enhance the operation of business. People provide value through the application of their knowledge, but their ability to do so is considerably enhanced by the availability of technology. Technology enriches opportunities for disseminating and sharing knowledge, and through providing the means to access information for decision making. While the role of technology may be seen as that of enabler to access information, value is added through the ability to assemble data, put it into meaningful information, and manipulate it into various scenarios before determining the best decision. Technology is adding value by allowing organizations to do this. Without the range of technology available to organizations, they would be considerably restricted in their ability to effectively manage their business. Therefore, technology should be identified as an adder of value.

REFERENCES

Allee, V. (1997). *The knowledge evolution: Expanding organizational intelligence*. Oxford: Butterworth-Heinemann.

Amidon, D. M. (1997). *Innovation strategy for the knowledge economy: The Ken Awakening*. Oxford: Butterworth-Heinemann.

Beckett, R. C. (2000). A characterization of corporate memory as a knowledge system. *Journal of Knowledge Management*, *4*(4), 311–319.

Bhatt, G. D. (2001). Knowledge management in organizations: Examining the interaction between technologies, techniques, and people. *Journal of Knowledge Management*, *5*(1), 68–75. Available at *http://angelina.emeraldinsight.com/v1=.../rpsv/cw/mcb/13673270/v5n1/s6/p68.html*

Binney, D. (2001). The knowledge management spectrum—Understanding the KM landscape. *Journal of Knowledge Management*, *5*(1), 32–42.

Bontis, N., Fearon, M., & Hishon, M. (2003). The e-flow audit: An evaluation of knowledge flow within and outside a high-tech firm. *Journal of Knowledge Management*, *7*(1), 6–19.

Carneiro, A. (2000). How does knowledge management influence innovation and competitiveness? *Journal of Knowledge Management*, *4*(2), 87–98.

Davenport, T. H. (1997, June 15). Known evils: Common pitfalls of knowledge management. *CIO*, 34–36.

Davenport, T. H., & Prusak, L. (1998). *Working knowledge*. Cambridge, MA: Harvard Business School Press.

Ericsson, F., & Avdic, A. (2003). Knowledge management systems acceptance. In E. Coakes (Ed.), *Knowledge management: Current issues and challenges*. Hershey, PA: IRM Press.

Evans, P. B., & Wurster, T. S. (1997, September–October). Strategy and the new economics of information. *Harvard Business Review*, 71–82.

Fahey, L., & Prusak, L. (1998, Spring). The eleven deadliest sins of knowledge management. *California Management Review*, *40*(3), 265–275.

Frappaola, C., & Capshaw, S. (1999, July). Knowledge management software: Capturing the essence of know-how and innovation. *The Information Management Journal*, 44-48.

Grant, R. M. (2000). Shifts in the world economy: The drivers of knowledge management. In C. Despres & D. Chauvel (Eds.), *Knowledge horizons*. Oxford: Butterworth-Heinemann.

Hansen, M. T., Nohria, N., & Tierney, T. (1999, March–April). What's your strategy for managing knowledge? *Harvard Business Review*, 106–116.

Hurley, M. A., & Harris, R. (1997). Facilitating corporate knowledge: Building the data warehouse. *Information Management and Computer Security*, *5*(5), 170–174.

Jentzsch, R., & Prekop, P. (2002). A conceptual model of collaborative information seeking. In D. White (Ed.), *Knowledge mapping and management*. Hershey, PA: IRM Press.

LaPlante, A. (1997, June 2). Sharing the wisdom. *Computerworld*, 73–74.

Lloyd, B. (1996). Knowledge management: The key to long-term organizational success. *Long Range Planning, 29*(4), 576–580.

Marshall, C., Prusak, L., & Shpilberg, D. (1996, Spring). Financial risk and the need for superior knowledge management. *California Management Review, 38*(3), 77–101.

Marshall, L. (1997, September/October). New opportunities for information professionals. *Online*, 93–98.

Martiny, M. (1998, Autumn). Knowledge management at HP consulting. *American Management Association, Organizational Dynamics, 27*(2), 71–77.

Mockler, R. J., & Dologite, D. G. (2002). Strategically-focused enterprise knowledge management. In D. White (Ed.), *Knowledge mapping and management*. Hershey, PA: IRM Press.

Pemberton, J. M. (1998, July). Knowledge management (KM) and the epistemic tradition. *Records Management Quarterly*, 58–62.

Shani, A. B., Sena, J. A., & Stebbins, M. W. (2000). Knowledge work teams and groupware technology: Learning from Seagate's experience. *Journal of Knowledge Management, 4*(2), 111–124.

Watt, P. (1997, August). Knowing it all. *Intranet*, 17–18.

KEY TERMS

Adding Value: This is giving something that was not anticipated, e.g., selling a customer a product, and then giving additional information about gaining maximum benefit from it, or giving advice about safety issues relating to the product.

Codification: Putting knowledge into a form that makes it accessible to others, e.g., writing down a formula.

Groupware: Computer software allowing groups, teams, and people in different locations to work together and share information.

Knowledge Management (management perspective): This is the process whereby knowledge is managed to provide a competitive advantage for the organization.

Knowledge Management (technology perspective): Practices and technologies that facilitate the efficient creation and exchange of knowledge on an organizational level.

Knowledge Management (KM) Spectrum: This is a management tool for organizations that provides insight into knowledge management options and assists with the assessment of the many applications and technologies that are available.

Knowledge Management System: Using technology to manage knowledge systematically to enable it to be gathered, organised, and shared for the benefit of the organisation.

Productivity: A measure of the performance of the production process in transformation of the input of resources to the output of goods and services.

T

Technology and Work in the Virtual Organization

Paul M. Leonardi
Stanford University, USA

INTRODUCTION

With the proliferation of industrial manufacturing plants in the United States during the first half of the 20th century, work was traditionally understood not only as something you did to earn a living, but someplace you went. The industrial revolution fundamentally changed the nature of work as it tied the practices of manual laborers to specific technologies such as the assembly line and the printing press, which were located in specific factories, plants, and offices. Today when we talk about work as a place we go rather than a thing we do, our rhetoric is a remnant of an image of manual work that dominated the 20th century.

BACKGROUND

Coinciding with the influx of post-war university enrollments in the '40s and '50s, in the latter half of the 20th century there was an explosion of jobs requiring "knowledge work." This work included computational and analytical functions not tied to a specific workplace technology and that could be performed from anywhere. By the 1970s the combination of widespread knowledge work, high gasoline prices, and new capabilities for data communication spurred a wave of telecommuting. Working from a remote location as one would do in the office allowed people to work at home while still maintaining contact with colleagues and clients (Nilles, 1975). Today, numerous terms are used to describe such work arrangements: "Telework," "distance work," "mobile work," and "virtual work" are among the most common.

The word "virtual" has gained increasing currency as a way of talking about organizational interactions that occur outside traditional barriers of time and space. In fact, today many organizations describe themselves as virtual organizations. As Dutton (1999) describes, a virtual organization is "composed of private firms or public agencies that have employed information and communication technologies (ICTs) to transform business processes within the organization or among themselves and other organizations" (p. 474). Many organizations that have decided to enable virtual work have been enthusiastic about the possibilities of virtual meetings, work teams, offices, factories, firms, and alliances (Burn & Barnett, 1999; DeSanctis & Monge, 1999). Table 1 summarizes some of the reasons organizations have made it possible for employees to work virtually.

ROLE OF TECHNOLOGY

Digital telecommunications technologies have been the primary enablers of virtual work.

While there is growing evidence that technology does not *drive* virtual work—that is to say, influence people to work virtually (Jackson, Leonardi, & Marsh, 2004; Scott & Timmerman, 1999)—technological infrastructures and applications do make virtual work possible. The earliest modern infrastructural technology was the telephone. Through the use of the telephone, workers could conduct their work individually from remote locations and report back to the office when necessary. Soon, dial-up modems became an efficient way for a virtual worker to log in to the organization's server and access information. Today, residential broadband services provide "always on" connections that enable online access with similar speeds and bandwidth to what workers are used to in traditional workplace offices. Certain applications have also been

Table 1: Some reasons organizations enable virtual work

Employee-Driven	Organizationally Driven
• Increased control of schedule	• Decreased maintenance/facility costs
• Eliminate/reduce commute	• Leverage diverse expertise
• Freedom from interruptions	• Decentralization
• Less formal work environment	• Employee health/welfare

key enablers of virtual work. Early database technologies allowed workers who worked remotely to access information collected and stored in computers at the organization. Aside from word processing applications, e-mail is perhaps the most important tool most used by modern virtual workers (Scott & Timmerman, 1999). E-mail allows the virtual worker to send and receive important documents and messages to colleagues and clients. Today we are beginning to see the use of instant messaging (IM) and teleconferencing technologies, with which virtual workers can communicate more efficiently with one another.

Digital telecommunications technologies must perform two important tasks if they are to enable successful virtual work. First, they must allow the worker to establish a telepresence. Steuer describes telepresence as "the experience of presence in an environment by means of a communication medium" (Steuer, 1992, p. 76). Although virtual workers are not physically copresent with their colleagues and clients, they must be able to work with them as if they were. For this reason, when communicating important information those who work with others at a distance often chose to communicate through the electronic medium that most mimics a face-to-face context (Trevino, Daft, & Lengel, 1990). Second, for technologies to effectively enable virtual work they must be transparent to the user. Virtual workers are typically not technological innovators and thus prefer to use technologies that do not interfere with the ways in which they would normally work at the office (Jackson et al., 2004). To the extent that the virtual worker has to think about when and how to use a technology to complete a work task, he or she will prove to be more ineffective. The most valuable technologies are those that the virtual worker does not have to think about when using and that do not interfere with normal work practices.

FUTURE TRENDS: PRACTICE OF VIRTUAL WORK

Work is the fundamental constitutive feature of the organizing process. To organize virtually, organizational members work in similar ways as they would in the office, but also in ways specific to their virtual arrangement. Specifically, virtual work makes explicit the task-based and relational aspects of ordinary office work and adds the dimension of presence practices. Table 2 summarizes these virtual work practices.

- **Task-Based Practices:** Work practices related to a specific job description that must be carried out to produce a deliverable product, process, or idea.
- **Relational Practices:** Work practices related to communication with others, including formal and informal interaction.
- **Presence Practices:** Work practices that establish the presence of a worker in the organization as experienced by coworkers and managers.

Task-Based Practices

Although stories persist in the popular and trade presses that the invention of new telecommunications technologies will fundamentally change the way individuals work, research on work in virtual organizations has shown that the task-based work practices of those who work virtually differs little from practices conducted in the office. In fact, knowledge workers who adopt virtual work arrangements must often convince their managers that they will be able to conduct the same work, of the same quality, and at the same speed from a virtual location as they would if they remained working at their workplace office (Leonardi, Jackson, & Marsh, 2004). As a result, virtual workers rarely talk about the practices of their task-based work when discussing their work arrangements, and when they do, they discuss how little has changed from the way they worked in the office. Instead, when reflecting upon their virtual work practices, most virtual workers cite the schedule flexibility, freedom from interruptions, and time saved in commuting as major changes wrought by their virtual work arrangements (Bailey & Kurland, 2002). Task-based work practices change little as a result of a move toward virtual work, but the nature of the task itself does affect the type of technology used to conduct those work practices (Haythornthwaite & Wellman, 1998).

Table 2: Types of virtual work practices

Task-Based Practices:	Work practices related to a specific job description that must be carried out to produce a deliverable product, process, or idea.
Relational Practices:	Work practices related to communication with others, including formal and informal interaction.
Presence Practices:	Work practices that establish the presence of a worker in the organization as experienced by coworkers and managers.

Relational Practices

All work involves the maintenance of relationships. In the virtual organization relational work practices are made explicit. Because there is no "water cooler" around which members can gather, virtual workers must strategically plan to communicate with coworkers. Typically, virtual workers communicate with other virtual workers more often than they communicate with office employees (Belanger, 1999). To maintain these relationships many virtual workers plan communication time into their interactions with others. For example, virtual workers often create special occasions for coworkers to come together for conversation and interaction in a face-to-face context. Inviting colleagues to social gatherings and holding annual meetings are ways that relationships are maintained. Also, virtual workers will exchange digital images and call a colleague before a conference call or meeting in order to "chat" before beginning to work (Leonardi et al., 2004). The distance inherent in virtual work requires that virtual workers pay specific attention to the relational elements of their work. Instead of taking social interaction for granted, virtual workers must communicate strategically with others in order to maintain healthy and productive interpersonal relationships. Organizational structures and practices that encourage and facilitate such relational work improve the effectiveness of the virtual worker.

Presence Practices

Perhaps the most salient fear for virtual workers is that not being physically copresent with colleagues and managers will prohibit them from "getting ahead" in the office (Leonardi et al., 2004; O'Mahony & Barley, 1999; Orlikowski & Barley, 2001). The major driver of this fear is the distance inherent in virtual work arrangements. To combat this problem, virtual workers conduct a variety of practices that help them to establish a virtual presence. For example, virtual workers communicate via e-mail much more frequently than do traditional office employees, and their e-mails are typically shorter and more informal (Jarvenpaa & Leidner, 1999; Scott & Timmerman, 1999). Communicating often via e-mail allows the virtual worker to establish his or her presence with coworkers by letting them "see," figuratively, the virtual worker hard at work and letting them "see," physically, his or her name in their e-mail in box. In this way the virtual worker is present in the work of others. Additionally, many virtual workers establish their presence in the organization by recreating the sense of an office virtually through the use of electronic lists and online newsletters (Leonardi et al., 2004). Moreover, using technologies traditionally deemed as informal and used out of the workplace, such as IM and chat groups, in order to communicate with colleagues allows the virtual worker to be present with others while enjoying the advantages of solitary work.

CONCLUSION

With the proliferation of virtual organizations, traditional images of work are shifting. "Work" in a virtual setting is enabled by digital telecommunications technologies, but not driven by them. As a consequence, virtual work makes explicit the task and relational elements of work practices. Virtual workers must proactively manage these two as distinct from one another. Additionally, virtual work must include practices that serve to establish the worker's presence in the organization. The enactment of these work practices creates and recreates virtual organizations through the content and structure of virtual work.

REFERENCES

Bailey, D. E., & Kurland, N. B. (2002). A review of telework research: Findings, new directions, and lessons for the study of modern work. *Journal of Organizational Behavior, 23*(4), 383-400.

Belanger, F. (1999). Communication patterns in distributed work groups: A network analysis. *IEEE Transactions on Professional Communication, 42*(4), 261-275.

Burn, J., & Barnett, M. (1999). Communicating for advantage in the virtual organization. *IEEE Transactions on Professional Communication, 42*(4), 215-222.

DeSanctis, G., & Monge, P. (1999). Communication processes for virtual organizations. *Organization Science, 10*(6), 693-703.

Dutton, W. H. (1999). The virtual organization: Teleaccess in business and industry. In J. Fulk (Ed.), *Shaping organization form: Communication, connection, and community* (pp. 473-495). Thousand Oaks, CA: Sage.

Haythornthwaite, C., & Wellman, B. (1998). Work, friendship, and media use for information exchange in a networked organization. *Journal of the American Society for Information Science, 49*(12), 1101-1114.

Jackson, M. H., Leonardi, P. M., & Marsh, N. N. (2004). *Communications technologies and changing telework practices: Considering the influence of broadband* (Working Paper). University of Colorado, Department of Communication.

Jarvenpaa, S. L., & Leidner, D. E. (1999). Communication and trust in global virtual teams. *Organization Science, 10*(6), 791-815.

Leonardi, P. M., Jackson, M. H., & Marsh, N. N. (2004). The strategic use of "distance" among virtual team members: A multi-dimensional communication model. In S. P. Ferris (Ed.), *Virtual and collaborative teams: Process, technologies, and practice* (pp. 156-172). Hershey, PA: Idea Group Publishing.

Nilles, J. M. (1975). Telecommunications and organizational decentralization. *IEEE Transactions on Communications, 23*, 1142-1147.

O'Mahony, S., & Barley, S. R. (1999). Do telecommunications technologies affect work and organizations? The state of our knowledge. In R. Sutton (Ed.), *Research in organizational behavior* (Vol. 21, pp. 125-161). Greenwich, CT: JAI Press.

Orlikowski, W. J., & Barley, S. R. (2001). Technology and institutions: What information systems research and organization studies can learn from each other. *MIS Quarterly, 25*, 145-165.

Scott, C. R., & Timmerman, C. E. (1999). Communication technology use and multiple workplace identifications among organizational teleworkers with varied degrees of virtuality. *IEEE Transactions on Professional Communication, 42*(4), 240-260.

Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of Communication, 42*, 73-93.

Trevino, L. K., Daft, R. L., & Lengel, R. H. (1990). Understanding managers' media choices: A symbolic interactionsist perspective. In C. Steinfield (Ed.), *Organizations and communication technology* (pp. 71-94). Newbury Park, CA: Sage.

KEY TERMS

Knowledge Work: Conceptual and analytical practices that require domain-specific knowledge and can be conducted from a variety of locations and, thus, are not tied to processes or technologies of specific workplace locations.

Presence Practices: Work practices that establish the presence of a worker in the organization as seen by coworkers and managers.

Relational Practices: Work practices related to communication with others, including formal and informal interaction.

Task-Based Practices: Work practices related to a specific job description that must be carried out to produce a deliverable product, process, or idea.

Telepresence: The experience of presence in an environment by means of a communication medium.

Transparent Technology: Technology that workers use on a daily basis without thinking about it. Most often a technology is considered transparent when a worker forgets that he or she uses it. When a worker must think about how and when to use a specific technology it is no longer transparent.

Virtual Work: The combination of task-based, relational, and presence practices conducted by workers who work from a remote location.

Technology in the Foreign Language Classroom

William J. Switala
Duquesne University, USA

INTRODUCTION

America is a country made up of people from all corners of the globe. Although this is the case, few Americans can communicate in a language other than English. The major reason for this is that Americans do not study foreign languages to any great extent in school, and those who do, have not developed a facility to speak the language they have studied. The American Council on the Teaching of Foreign Languages found, in its survey entitled "Foreign Language Enrollments in U.S. Public High Schools, 1890-2000," that there was a steady decline in the numbers of students studying foreign languages from 1976 to 1994. From 1995 to 2000 this trend was reversed and the number of students learning new languages in the year 2000 almost matched that of the enrollment for 1974. However, this still only accounted for 42.5% of the total number of students attending American high schools (ACTFL, 2004). A possible explanation for this low number may rest in the methodology used to teach foreign languages in our schools. (Brecht, 2002).

BACKGROUND

The Evolution of Technology in Foreign Language Teaching

The approach used for teaching foreign languages in the past was to provide the student with enough vocabulary and grammar so that they could read a language. Speaking the language was not as important as reading it. The current philosophy in language instruction emphasizes speaking over reading. Today to meet this end every effort is made to engage the students in more productive activities, resulting in greater use of communication skills. A way to do this is to incorporate more technology in the foreign language classroom. Technology, for the purposes of this discussion, will consist of software programs, the Internet and distance learning satellite transmissions.

The use of technology in foreign language teaching began over 60 years ago. Starting with the use of language labs by the military in World War II, the utilization of technology has gotten much greater since then. From 1950 to 1980 the teaching of foreign languages entered what has been called the "Age of Methods". Language labs with their interconnected tape recorders moved from the military setting to high schools and colleges. During the 1960s and 1970s, the audio-lingual method, in which the student imitated correct pronunciation by listening to native speakers on tape recordings, became the rage in foreign language instruction (Rodgers, 2001). The use of videotape based language programs became popular in the 1980s. The most famous of these was the Capritz series for learning the French language entitled "French in Action". It was also during the 1980s that the computer as a device for enhancing language study appeared. Initially, consisting of simple vocabulary development programs, this technology expanded to include complete language learning programs, authoring programs for teachers, and computerized language lab systems (Earp, 2001). In the 1990s the advent of distance learning opportunities through language programs broadcast and received over satellite transmissions added to the technological possibilities in foreign language instruction.

The applications of technology in the teaching of foreign languages today fall into three major categories: basic language instruction; reinforcement of language skills; and enhancement of the target language. Each of these categories is open to multiple applications of technology. Each of them have available support from computer programs, Web sites and satellite transmissions. What follows is a discussion of each of these aspects of teaching a foreign language and how technology can augment that teaching.

BASIC LANGUAGE INSTRUCTION

Basic language instruction deals with the acquisition of a new language. Three fundamental modes of delivery exist in using technology to accomplish this. They are: software programs; the Internet; and distance learning, using a program of instruction transmitted via satellite. Each of these has special requirements, advantages, disadvantages and potentials. There are numerous software applications on the market today that aid the individual student in learning a new language. There are two ways to use them. First, the teacher can place students in a computer lab setting and share the instructional disk through a file-server. All of the students may then receive the basic instruction at the same time. In this approach the teacher will need to receive special training in the operation of the program. It is also advisable to have a trained computer aid on hand to help with the operation of the program. A second possibility is to have students work individually in an independent setting to learn the language. In this situation the teacher acts more as a mentor than as a primary instructor. The advantage to the CD-ROM program approach is that the student can proceed at his/her own pace and not be restricted by the progress of other students. A major disadvantage to this technique is that if the students are studying the language as individuals, there is no one with whom they can interact, especially in practicing the oral skills of the language. Two excellent sources for information on software applications of this nature are Applause Learning Resources (Web site: *applauselearning.com*) and Teacher Discovery Company (Web site: *teacherdiscovery.com*).

The Internet is also a source of opportunities for learning a foreign language. The same practical considerations listed for using software programs also apply to the use of the World Wide Web. Having computers available that can access the Internet is a must if one is to use this resource. The configuration of these computers is the same as for the software packages: they may be used in a lab setting for group instruction of the language, or as individual stations for one-on-one learning. In either case the need for a teacher or a mentor is also the same. There are two major Web sites that offer foreign language course instruction. The first of these, about.com, is the most comprehensive one, offering the greatest number of languages. The full address for the site is *http:// (the language you want).about.com/homework/(the language you want)*. For example, if you were looking for the program that teaches French, you would enter: *http:// french.about.com/homework/french*. The languages offered on this Web site include: Chinese, French, German, Italian, Japanese, Latin, Russian and Spanish. Parlo.com is another site that offers instruction in French, Italian and Spanish. The address for this site is *www.parlo.com*. As with the software applications, a teacher should be present to monitor the students.

The final possibility for teaching a foreign language is to use one of the distance learning offerings sent by satellite transmission. The major provider of this service is the Satellite Educational Resources Consortium (SERC). SERC was the pioneer in offering foreign languages via satellite transmission. It offers instruction in high school level German, French, Japanese and Latin. It also provides a Spanish program designed for elementary and middle school students.

There are several factors that one must attend to when using this technology. The most important of these is the obtaining of the equipment necessary for the operation of the program. A satellite dish must be purchased, located in an appropriate place at the school, and protected from vandalism. Reception and distribution equipment are needed to acquire the signal from the satellite dish and to channel the program to a specific classroom. A television monitor, speakerphone and a dedicated phone line are also needed in order to operate the program successfully. The dedicated phone line and speakerphone facilitate the two-way communication between the students and the teacher presenting the course. Finally, it is essential that someone be trained in the proper operation of the entire system.

REINFORCEMENT

Reinforcement is the process by which learned material is ingrained to a greater depth. It increases retention of the subject matter. This is especially important in the study of foreign languages, where learning is cumulative. Technology can achieve this goal and focus student attention to a degree often superior to regular classroom instructional techniques. Once again, there are three major formats through which this can occur: software programs, the Internet and satellite transmissions.

Most foreign language textbook series come with computer disks containing exercises that are drill and practice in nature. These are specially designed to reinforce concepts, usually grammatical and vocabulary, covered in the chapters and units of the textbook. Basically they consist of students manipulating objects or statements presented with well-developed graphics, usually with sound accompaniment, to complete the exercises. Self-tests are typically part of the program and students can obtain instant feedback on their performance. The teacher can use these programs, either as a general reinforcement for the entire class, or as a tutorial for individual students. In addition to programs that come with the textbook, other sources are also available. Offerings appear in several formats: drill and practice for grammar, vocabulary builders, reading comprehension programs, dialogues, puzzles and games. A wide variety of packages of this nature for classes in German, French, Latin and Spanish are available in the catalogues of the Discovery and Applause Companies.

Another option for reinforcement is to access programs on the Internet. Many textbook companies have a Web site related to their foreign language series. These sites typically offer self-progress assessments, remedial activities and review exercises. Glenco (*www.foreign language.glenco.com*), Holt Rinehart Winston (*www.hrw.com*), and Prentice Hall (*www.phschool.com/foreign_languages/index.html*) have especially effective sites.

The third option for providing reinforcement is to make use of selected satellite transmissions. These are not found on the satellite venues used for basic instruction. Instead, they involve commercial television programs. Offerings such as documentaries, soap operas, travelogues and commercials are broadcast over satellites in native languages, especially Spanish and French. Native speakers provide the dialogue and students can hear the correct pronunciation, idiomatic expressions and proper grammar of their target languages. Teachers can have students watch these transmissions in real time, or the teacher can videotape them and show them at a more convenient or effective time.

ENHANCEMENT

Enhancement is whatever the teacher does to make the learning more vivid and interesting. It enriches the students' understanding of the culture they are studying. Opportunities for enhancement are exceptionally varied today. They include offerings in art, music, customs, foods, dress styles, geography, history and the political structure of the native lands related to the languages students are studying. They are also available in a software format, through Web sites, and via satellite transmissions.

The most readily available source of enhancement programs is the software package. They accompany many textbook series on the market today and they are also obtainable from independent sources such as Applause Learning Resources. The Internet contains a number of Web sites that also offer effective and attractive learning opportunities in this area such as interaction with speakers from other countries through discussion groups, chat rooms and e-mail pen pals (Pufahl, 2002). Other excellent sources for enhancement programs are the University of Wisconsin Web site (*http://polyglot.lss.wisc.edu/lss/lang/langlink.htm*), the University of Iowa Web site (*www.uiowa.edu/~flesled/resources/resources.html*) and the site of the American Classical League (*www.aclclassics.org*). Distance learning transmissions also provide a source for enhancement possibilities. Many programs of a cultural nature are broadcast in foreign languages via satellite. These include travel shows, tours of museums and visits to other cultural areas. Students may view these programs in real time, or later, through tapes of the offerings made by the teacher.

CONCLUSION

Whatever option the teacher may choose to use, technology can provide a wide range of opportunities and vistas that make the teaching of foreign languages much more effective in the 21st century than at any other time in history.

REFERENCES

American Council for the Teaching of Foreign Languages (ACTFL). (2004). *Foreign language enrollments in U.S. public high schools, 1890-2000. www.actfl.org*

Brecht, R.D. (2002, May). Tapping a national resource: Heritage languages in the United States. *ERIC Digest.* (*www.ericfacitilty.net/ericdigests/ed464515.html*)

Earp, S. (2001). More than just the Internet: Other technology for language teaching. *ERIC Review, K-12 Foreign Language Education, 6*(1). *www.eric.ed.gov/resources/ericreview/vol6no1/tech.html*

Pufahl, I., Rhodes, N.C., & Christian, D. (2001, September). What we can learn from foreign language teaching in other countries. *ERIC Digest. www.ericfacility.net/ericdigest/ed456671.html*

Rodgers, T.S. (2001, September). Language teaching methodology. *ERIC Digest. www.cal.org/ericcll/digest/rodgers.html*

KEY TERMS

Audio-Lingual Method: This is the method of studying a foreign language that stresses hearing the language spoken by native speakers and then imitating the patterns of pronunciation and grammar heard in the example.

CD-ROM Programs: These are standard computer disk operated software that run on a stand- alone computer, or on multiple computers through a file server.

Computerized Language Labs: Foreign language instructional programs operated via computer disks and distributed to multiple computers by a file server and accessed by an entire class of students.

Distance Learning Satellite Transmissions: Learning programs transmitted via television signals from remote sources and accessed locally by students through a receiver dish antenna.

File Server: A piece of equipment that allows one to distribute a software program from a single source to a number of computers.

Internet: The electronic system that links various computer networks around the world. It includes educational venues, databases, electronic mail, news agencies and chat rooms.

Stand-Alone Programs: Foreign language instructional programs operated by an individual computer in a classroom.

Technology of Formal Education

Donald A. Hantula
Temple University, USA

Darleen M. DeRosa
Right Management Consultants, USA

INTRODUCTION

Internet distance education is a natural consequence of *fin de siecle* industrial transformations from a manufacturing economy, in which standard educational practices are based, to an information economy, in which greater autonomy, collaboration, flexibility and a project orientation to work are the norm. The Internet did not cause changes in education, but rather enabled educators to meet new demands for instructional practices and outcomes and adapt to a rapidly changing economic and social environment that was beginning to outpace the academy. Today, just as 100 years ago, educational institutions and practices are modeled on prevailing industrial examples of work and organization. This is especially the case in the United States where an overriding intended effect of formal education is to prepare students to fill roles within the prevailing economic system. Against this backdrop, it is only those components of education that reflect and reinforce the prevailing industrial system that are incorporated into the technology known as formal education. Components of education such as teaching machines and distance learning existed throughout the 20th century but never became standard educational practice until fairly recently because they were not acceptable in terms of preparing students to enter the prevailing industrial system.

BACKGROUND

Educational institutions customize many of their services according to what is dictated by industry, "manufacturing" employees who are suitable for the workplace (Jacques, 1996), thereby, completing a system of supply and demand. The classroom was designed as an industrial entity as it mirrored organizational practices and education emulated the factory. Straight lines of desks (often bolted to the floor), uniform curricula, standardized forms and procedures for evaluating students and faculty, strict scheduling, student achievement indexed according to hours worked and units completed all bear more than an accidental resemblance to the manufacturing process. As formal education grew in the United States in the early 20th century, the scientific management movement informed and inspired educators to view schools in the same terms as manufacturing businesses (Spring, 2001), or as "…essentially time- and labor-saving devices, created by us to serve democracy's needs" (Cubberly, 1919, p. 355). Education satisfied these industrial "needs" with a standard "product"—a graduate who not only was trained in the basics of reading, writing, and arithmetic (skills of practical usefulness), but who was also socialized to industry (Robbins, 1997). Educators were trained to consider themselves as administrators or managers, seeking the most efficient ways to teach attendance, punctuality, attentiveness, conformity, rote learning and an acceptance of standardized work, piece-meal production and adherence to a hierarchical order (Spring, 2001). These were the lessons to be learned so that the "industrial capabilities and character" could be shaped (Cubberly, 1909, p. 41). Principals were akin to factory managers, setting general policies and procedures under which teachers — shop managers of their own classrooms — made the process work. Thus, it is not surprising that the physical design of school buildings and their interiors reflected the design of factories; the practices occurring within them attempted to replicate, as closely as possible, the prevailing industrial order.

With the concurrent rise of both formal education and the factory system, it might be reasonable to assume that various technologies would have been quickly applied to produce more efficient education. However, this was not the case. Despite the prevailing machine age, schools for the most part did not adopt mechanized methods of education such as teaching machines. Instead, a more teacher-driven, craft model of education was the norm. Within the constraints of the classroom, teachers as skilled craftspeople assembled education from centrally approved and provided pieces in a custom shop. The craft of teaching was realized through regulating the flow and progress of students through mass-produced mandated material by explaining, illustrating, and answering questions. Teaching filled in the gaps between a standard curriculum and the individual needs of the students. Technologies such as the overhead projector, which could be easily incorporated into the classroom under the teacher's control, were accepted because they did not threaten the status quo (Kipnis, 1994). Table 1 summarizes some of the major educational technologies that had bright promise but were never widely adopted.

The classroom/factory in which the compliant worker-consumer is the end product is no longer acceptable because factories are no longer the dominant models for most business organizations. The transformation from an industrial economy to an information economy has altered the way that organizations are run and the way education is configured (Sumner, 2000). Flat organizational structures, a project versus job orientation to work, less-centralized control and flexible scheduling are current configurations that enable rapid response, new innovations, and the development of new global alliances (Alavi, Wheeler, & Valacich, 1995). In this new economic model, outcomes depend not on goods but on information, and technology is the normative tool. We have seen a precipitous decline in the importance of spatio-temporal constancy; people commonly are not in the same place at the same time when "work" occurs. Because of globalization and the rapid pace of technological change, there is now an imperative to redraw the physical boundaries of the classroom, allowing learning to be continuous and education to occur in any place or at any time. With the rise of knowledge work and increased autonomy, the work model emerging is one of collaborative, rather than individual effort. Because knowledge work requires more flexibility and adaptability, individual employees have freer reign to determine how tasks will be performed. Part of this self-direction is the ongoing option to seek assistance and to reciprocate when the opportunity arises. Because computer technology is now ubiquitous in industry, computers are no longer the tools of the few. Combining the technological imperative with the nearly appliance-like nature of computers, the social and structural determinants are in place for computer-mediated distance education to become the norm. Educational institutions have changed the way education is accomplished in order to "manufacture" the needed graduates who have the requisite skills that the new workplace demands (Jacques, 1996). Trends such as the greater need for life-long learning, the demand for more part-time educational resources, and demographic changes such as an increase in older workers, techno-literacy, networked and team-based learning will continue to pressure educational institutions to adapt their offerings to fit this new industrial order (Sumner, 2000).

FUTURE TRENDS

Networked and team-based learning will become more important in the future as there has been a proliferation of the use of collaboration and teamwork in most organizations, and organizations rely on numerous types of teams to accomplish various tasks and goals (DeRosa, Hantula, Kock, & D'Arcy, 2004). The rise in virtual teams is the result of the growth of teamwork in organizations and increased geographic dispersion of workers (Lipnack & Stamps, 2000). As a result, an emphasis on teamwork and collaboration in educational settings will better prepare employees for the business world (Zaccaro & Bader, 2003). In addition, organizations are also relying more heavily on self-managed work teams (Yeatts & Hyten, 1998). Due to the increase of these teams in organizational settings, it is plausible that SMWTs will become more prevalent in educational institutions of the future.

CONCLUSION

Previous distance and technologically-based educational innovations have not necessarily failed, but did not match prevailing economic and social conditions. The current rise in Internet and technologically-based education mirrors a much larger change in the industrial order. Privateer (1999) asserts that technology should be viewed as a tool to redesign educational curricula, rather than simply as a replacement for traditional instructional methods, and

Table 1. Some promising educational technologies that were not widely adopted

- 1800s postal mail correspondence courses (described in Nasseh, 2002)
- 1920s Pressey's (1926) teaching machine
- 1930s instructional radio (described in Wright, 1991)
- 1950s Midwest program on Airborne Television Instruction (described in Reiser, 1987)
- 1960s Skinner's (1968) teaching machines and programmed instruction; Keller's (1968) personalized system of instruction
- 1980s computer-based educational applications were introduced (Crowell, Quintanar, & Grant, 1981)

Bereiter (2002) further argues that such a profound shift should usher in an entirely new theory of knowledge and learning. Indeed, if a century of educational research and technological experiments have taught us anything, it is that it is the not the medium, but it is the method that makes a difference (Clark & Zuckerman, 1999). Availability of technological media does not guarantee that these media will be properly integrated into instructional techniques. Instead, the greater challenge appears to be the human component of this process. Humans should endeavor to shape technology to fit both their physical and psychological needs (Kock, 2001) and not simply produce advanced technologies that are not comfortable. Ultimately, technology will continue to dramatically redefine learning and teaching and will have a profound impact on the educational institutions that enable these processes, and future research must begin to move beyond simplified models to more theoretically rich ones in order to appropriately deal with the complex set of variables in computer-supported cooperative work (Coovert & Thompson, 2001). Internet-based education is not necessarily a threat to formal education but instead is another form of the larger technology of formal education, a natural part of the evolving social order.

REFERENCES

Alavi, M., Wheeler, B.C., & Valacich, J. S. (1995). Using IT to reengineer business education: An exploratory investigation of collaborative telelearning. *MIS Quarterly, 19(3)*, 293-313.

Bereiter, C. (2002). *Education and mind in the knowledge age*. Mahwah, NJ: Erlbaum.

Clark, R. C., & Zuckerman, P. (1999). Multimedia learning systems: Design principles. In H.D. Stolovich & E.J. Keeps (Eds.), *Handbook of human performance technology* (2nd ed.) (pp. 546-588). San Francisco, CA: Jossey-Bass Pfeiffer.

Coovert, M.D., & Thompson, L.F. (2001). *Computer supported cooperative work: Issues and implications for workers, organizations, and human resource management*. Thousand Oaks, CA: Sage Publications, Inc.

Crowell, C.R., Quintanar, L.R., & Grant, K.L. (1981). PROCTOR: An on-line student evaluation and monitoring system for use with PSI format courses. *Behavior Research Methods & Instrumentation, 13*, 121-127.

Cubberly, E.P. (1909). *Changing conceptions of education*. Boston, MA: Houghton-Mifflin.

Cubberly, E.P. (1919). *Public education in the United States*. Boston, MA: Houghton-Mifflin.

DeRosa, D., Hantula, D., Kock, N., & D'Arcy, J. (2004). Trust and leadership in virtual teamwork: A media naturalness perspective. *Human Resource Management, 42*, 219-232.

Jacques, R. (1996). *Manufacturing the employee: Management knowledge from the 19th to the 21st centuries*. Thousand Oaks, CA: Sage Publications.

Keller, F.S. (1968). Good-bye teacher…. *Journal of Applied Behavior Analysis, 1*, 79-89.

Kipnis, D. (1994). Ghosts, taxonomies, and social psychology. *American Psychologist,* 52, 205-211.

Kock, N. (2001). The ape that used email: Understanding e-communication behavior through evolution theory, *Communications of the AIS, 5*(3), pp. 1-29.

Lipnack, J., & Stamps, J. (2000). *Virtual teams: People working across boundaries with technology*. (2nd ed.). New York: John Wiley & Sons, Inc.

Nasseh, B. (2002). A brief history of distance education. *Adult Education in the News*. http://www.seniornet.org/edu/art/history.html

Pressey, S.L. (1926). A simple apparatus which gives tests and scores - and teaches. *School and Society, 23*(586), 373-376.

Privateer, P.M. (1999). Academic technology and the future of higher education: Strategic paths taken and not taken. *The Journal of Higher Education, 70*(1), 60-79.

Reiser, R.A. (1987). Instructional technology: A history. In R. Gagne (Ed.), *Instructional technology: Foundations*. Hillsdale, NJ: Lawrence Erlbaum.

Robbins, S.A. (1997). Implications of distance education as an agent of sociocultural change. *Educational Telecommunications*, 1791-1798.

Skinner, B.F. (1968). *The Technology of Teaching*. New York: Prentice-Hall.

Spring, J. (2001). *The American school 1642-2000*. New York: McGraw-Hill.

Sumner, J. (2000). Serving the system: A critical history of distance education. *Open Learning, 15*, 267-285.

Wright, S.J. (1991). Opportunity lost, opportunity regained: University independent study in the modern era. In B.L. Watkins, & S.J. Wright (Eds.), *The foundations of American distance education: A century of collegiate correspondence study*. Dubuque, IA: Kendall/Hunt.

T

Yeatts, D.E., & Hyten, C. (1998). *High-performing self-managed work teams: A comparison of theory to practice*. Thousand Oaks, CA: Sage Publications, Inc.

Zaccaro, S.J., & Bader, P. (2003). E-Leadership and the challenges of leading E-teams: Minimizing the bad and maximizing the good. *Organizational Dynamics, 31*, 377-387.

KEY TERMS

Educational Technology: Applications of one or more electronic or computer-based devices used in delivering education.

Information Economy: An economic system characterized by knowledge based work and processing and exchange of information and data as its primary activities.

Knowledge Work: Work accomplished through the manipulation of data and symbols.

Self-Managed Work Teams (SMWTs): Autonomous teams that are responsible for managing and performing various tasks.

Technoliteracy: The ability to understand, use, critically analyze, and improvise solutions within an information technology infrastructure.

Virtual Teams: Teams of people who work interdependently across space, time, and organizational boundaries through the use of technology to facilitate communication and collaboration.

Technology Planning in Schools

Julie Mackey
Christchurch College of Education, New Zealand

INTRODUCTION

Information and communication technology (ICT) planning as it relates to schools can be defined as the process of identifying the information and communication technologies used to support the educational and administrative goals of schools and of deciding how these technologies will be developed and managed (Lederer & Sethi, 1988; Smits & van der Poel, 1996). This article presents a model reflecting the evolution of ICT planning maturity in schools and identifies the factors that influence and characterize integration between ICT planning and educational strategy. The model suggests a common evolutionary pathway for ICT planning in schools and provides a foundation on which to propose a "stages of growth" model for characterizing and evaluating ICT planning in these settings.

BACKGROUND

Information systems (IS) planning[1] and the degree to which that planning is clearly linked into an organization's strategic goals have been widely recognized as key factors in the successful use of information technology (IT)[2] within the organization (see, for example, Watson, Kelly, Galliers, & Brancheau, 1997; Segars & Grover, 1998). Over a decade ago, Telem (1993) suggested that the use of IT as a management tool in educational contexts was a neglected area of research, particularly in terms of lack of an underlying knowledge base. Although in the ensuing years, an increased research focus on ICT in educational management has been seen, this focus has concentrated more on evaluative analysis of the efficacy of management information systems, computerized school information systems, and specific applications of information technology (see, for example, Barta, Telem, & Gev, 1995; Fung, Visscher, Barta, & Teather, 1997; Visscher, Wild, & Fung, 2001) than on understanding how schools develop their ICT systems and integrate them into their management systems and practices. Passey (2002) noted the continuing lack of research relating to ICT and school management within the school sector.

The notion that organizations evolve is encapsulated in various stages of growth models that are widely used in both organizational and IS research. In IS literature, these models are based on the premise that organizations move through various stages of maturity in their use and management of IS (Nolan, 1973; Huff, Munro, & Martin, 1988; King & Teo, 1997; Teo & King, 1997). For example, King and Teo (1997) proposed a four-stage model conceptualizing the integration of IS planning and business planning over time, to better enable the effective support of business strategies. These four stages are separate planning with administrative integration, one-way linked planning with sequential integration, two-way linked planning with reciprocal integration, and integrated planning with full integration between business planning and IS planning.

Although researchers debate the accuracy and completeness of these various models, the models provide organizations with useful benchmarks to determine their current state of maturity and planning for future growth. More specifically, Robson (1997), for example, suggested that such knowledge provides organizations with a base from which to develop appropriate IS-related strategies, management styles, control approaches, and investment levels. Schools similarly can use such models to determine their ICT planning maturity and to develop appropriate strategies for future growth.

ICT PLANNING IN SCHOOLS

The complexity of the educational environment provides an interesting context in which to consider ICT planning. First, unlike businesses, schools tend to focus on the use of ICT to support learning objectives rather than business objectives, yet, like businesses, they work with limited resources and financing. Second, technology in schools is not always well established (Latham, 1998; Knezek & Christensen, 1999), and an Educational Review Office (ERO) report stated that "the overall implementation of ICT was still in its early stages" (ERO, 2001, p. 2). As such, schools provide a contemporary context in which to demonstrate how ICT usage evolves within an organization. They also provide a venue to examine how this process is affected when many of the people within an organization have only a limited understanding of ICT development and planning, as tends to be the case in schools. Teachers and school administrators often acquire their IT skills "on the job" yet are expected to initiate the planning and implementation of ICT within the school (Ministry of Education, 1999). Third, much of the small amount of research that has been conducted on the integration of ICT plans and educational strategic plans focuses on large tertiary institutions (see, for example, Barta et al., 1995; Fung et al., 1997; Rice & Miller, 2001), which have very different organizational structures, more complex IS needs, and greater numbers of specialist IT staff than do elementary and secondary schools.

Identifying how ICT planning evolves within schools and the characteristics of each stage of that evolution enables school administrators to better understand the factors contributing to ICT educational strategy alignment and successful ICT integration and growth.

STAGE THEORY OF ICT PLANNING

The proposed model is based on an interpretive case study that examined the status of ICT planning and its integration with educational strategic planning in eight New Zealand state elementary schools. The model is based on data gathered from interviews with the principals, and analysis of the schools' strategic and ICT plans. The data identified factors that seemed to influence ICT planning in the eight schools, and highlighted a pattern across the schools in which ICT planning appeared to evolve from an unplanned state through to some form of alignment with educational objectives set in the schools' strategic plans. This pattern provided the basis of the proposed four-stage model of ICT planning maturity. The model, detailed in Figure 1, presents not only the characteristics that appear to define stages of planning maturity within schools, but also the evolutionary progression through these stages over time. In addition to outlining the four stages of the model, Figure 1 provides a set of benchmarks that corresponds to each of the stages. These benchmarks are organized across the four categories of *curriculum*, *professional development*, *infrastructure development*, and *school administration*.

The four stages of the model are not intended to be discrete but are signal points along a continuum of planning maturity. The benchmark characteristics associated with each stage are, therefore, similarly, indicators of progress along an evolutionary pathway rather than examples of discrete categories.

Progress in planning appeared to be influenced by internal, rather than external, factors. Even though all of the schools in the study were state elementary schools operating within similar parameters, they were demonstrably different in terms of their ICT-related competence and commitment. The factors that appeared to influence ICT planning and development were related to school size (which impacted human and physical resources); ICT competence and attitudes of top management; ICT competence and attitudes of teachers; the existing ICT infrastructure (relating to the effectiveness of past planning and legacy systems); and the alignment between the school's ICT-related strategies and educational strategies.

Figure 2 identifies some of the drivers that appear to enable schools to progress in their ability to plan effectively for the alignment of ICT and strategic educational goals.

FUTURE TRENDS

The development of a stage model of ICT planning maturity for schools provides a tool for analyzing current planning processes and assessing maturity vis-à-vis each of the planning categories proposed by the model. In this regard, it also offers further evidence of the usefulness of stage theory in understanding organizational development and IS strategy and business/educational strategy alignment. The model descriptors (Figure 1) provide benchmarks for highlighting areas of deficiency, assessing future directions, and providing assurance to school leaders as they work toward planning maturity. The stage drivers (Figure 2) go beyond the basic stage model to provide direction on aligning ICT planning with school strategic planning.

In similar vein to the study conducted by King and Teo (1997), the model depicts a progression from unformulated ICT plans with piecemeal implementation of hardware and software, through to the development of coherent and focused plans that emphasize the strategic goals of the school in the areas of teaching, learning, and administration. It seems that as schools develop skills in planning for ICT, they tend to identify opportunities to enhance the core business of teaching and learning through the innovative use of ICT in administrative, collaborative, and support roles.

The model provides a basis for investigating and analyzing ICT planning maturity in small organizations and, in particular, education contexts. There are opportunities to refine and enhance the model as schools mature in their planning and integration of ICT to achieve educational goals.

CONCLUSION

ICT planning and strategic planning in schools is not a clearly defined area, and numerous factors, both internal and external, impact the planning process and outcomes.

Figure 1. Four-stage model of ICT planning in schools

	Stage 1: Little or No Planning	**Stage 2: Early Formal Planning**	**Stage 3: Focused Formal Planning**	**Stage 4: Mature Integrated Planning**
Descriptor	*Planning has a technical and operational focus; Plans may be informal and not clearly documented*	*Planning becomes more formal and the focus shifts from technical aspects to the educational value of ICT*	*Planning becomes increasingly focused and linked to strategic objectives and teaching and learning*	*ICT planning is aligned with, and is integral to, the school's strategic planning process*
Categories				
Curriculum	Links between ICT use and curriculum objectives are weak or nonexistent	Some focus on how ICT can support teaching and learning	A maturing understanding of how ICT can support and enhance teaching and learning across curriculum areas is evident	ICT is an integral tool in the teaching and learning process
Professional development (PD)	PD develops basic computing skills and confidence for novice ICT users	PD targeted to varying staff needs and links skill development to possible curriculum applications	PD focuses on contextual skill development and curriculum applications as teacher competency increases	Teachers confidently use a range of technologies, and PD is customized and specific
Infrastructure development	Piecemeal technology acquisition and stand-alone configurations	Structured development with emphasis on compatibility, upgradability, and connectivity	ICT infrastructure planning complements school vision and general planning	Strategic planning explores innovations to enhance capability and infrastructure
Administration	Basic office functions are computerized	Communication and additional administrative functions are computerized	ICT enhances efficiency and effectiveness of administrative systems, teacher planning, and wider school objectives	Innovative solutions enhance communication, resource sharing, and student recording and reporting

Figure 2. Summary of stage drivers

Stage 1: Little or No Planning	Stage 2: Early Formal Planning	Stage 3: Focused Formal Planning	Stage 4: Mature Integrated Planning

- Developing a shared vision for ICT with all staff.
- Formal planning and review.
- Principal and staff willing to learn.
- Appropriate professional development.
- Some commitment to resourcing the ICT infrastructure.

- Increasing acceptance and understanding of the role of ICT to support learning.
- Top management commitment (e.g., from Principal and Board of Trustees).
- Planned and effective professional development for all staff.
- Commitment to ICT resourcing for medium-term development.
- Growing knowledge, experience, and understanding of technical aspects.

- A school culture that believes in the value of ICT to enhance and support both curriculum and administrative areas.
- Strong vision and knowledgeable leadership by the principal.
- Commitment to specific and appropriate professional development.
- Innovative approach to ICT use—willing to be leaders.
- Ability to maximize support from external agencies, e.g., business partnerships, professional development contracts.

In spite of this complexity, it is evident that a school's ICT planning matures as the focus shifts from technology and skills acquisition to the effective use and integration of technologies in line with the school's strategic plans.

REFERENCES

Barta, B., Telem, M., & Gev, Y. (1995). *Information technology in educational management*. London: Chapman & Hall.

Education Review Office. (2001). *The implementation of information and communication technologies (ICT) in New Zealand schools 2001*. Wellington, New Zealand: Education Review Office Publications.

Fung, A. C. W., Visscher, A. J., Barta, B., & Teather, D. C. B. (Eds.). (1997). *Information technology in educational management of the future*. Paper presented at the International Conference on Information Technology in Educational Management (ITEM), 22–26 July 1996, Hong Kong.

Huff, S. L., Munro, M. C., & Martin, B. H. (1988). Growth stages of end-user computing. *Communications of the ACM*, *31*(5), 542–550.

King, W. R., & Teo, T. S. H. (1997). Integration between business planning and information systems planning: Validating a stage hypothesis. *Decision Sciences*, *28*(2), Spring, 279–308.

Knezek, G., & Christensen, R. (1999). Stages of adoption for technology in education. *Computers in New Zealand Schools*, November, 25–29.

Latham, A. (1998). Strategic information systems planning: A necessary evil for schools? *Journal of Applied Management Studies*, *7*(2), 267–273.

Lederer, A. L., & Sethi, V. (1988). The implementation of strategic information systems planning. *MIS Quarterly*, *12*(3), 445–461.

Ministry of Education. (1999). *Learning technologies planning guide for schools: Using ICT to improve teaching and learning*. Wellington: Learning Media.

Ministry of Education. (2003). *Digital horizons: Learning through information and communication technology (Revised Ed.)*. Wellington: Learning Media.

Nolan, R. L. (1973). Managing the computing resource: A stage hypothesis. *Communications of the ACM*, *16*(7), 399–405.

Passey, D. (2002). *ICT and school management: A review of selected literature*. Lancaster University. Retrieved May 7, 2004, from http://www.becta.org.uk/research/reports/docs/ict_sm.pdf

Rice, M., & Miller, M. T. (2001). Faculty involvement in planning for the use and integration of instructional and administrative technologies. *Journal of Research on Computing in Education*, *33*(3), 328–337.

Robson, W. (1997). *Strategic management and information systems*. London: Pitman Publishing.

Segars, A. H., & Grover, V. (1998). Strategic information systems planning success: An investigation of the construct and its measurement. *MIS Quarterly*, *22*(2) 139–163.

Smits, M. T., & van der Poel, K. G. (1996). The practice of information strategy in six information intensive organizations in The Netherlands. *Journal of Strategic Information Systems*, *5*, 93–110.

Telem, M. (1993). Information technology: A missing link in educational research. *Journal of Research on Computing in Education*, *26*(1), 123–143.

Teo, S. H., & King, W. R. (1997). Integration between business planning and information systems planning: An evolutionary-contingency perspective. *Journal of Management Information Systems*, *14*(1), 185–214.

Visscher, A. J., Wild, P., & Fung, A. C. W. (Eds.). (2001). *Information technology in educational management: Synthesis of experience, research and future perspectives on computer-assisted school information systems.* The Netherlands: Kluwer Academic Publishers.

Watson, R. T., Kelly, G. G., Galliers, R. D., & Brancheau, J. C. (1997). Key issues in information systems management: An international perspective. *Journal of Management Information Systems*, *13*(4), 91–115.

KEY TERMS

Computerized School Information Systems: Specific information systems for educational management, for example, enrollment systems and assessment and reporting systems.

ICT Integration: The degree to which ICT is used to achieve organizational goals. In the educational context, integration means that ICT is used transparently to achieve curriculum objectives and is not taught for its own sake.

ICT Planning in Schools: Specific school planning relating to the provision, acquisition, implementation, curriculum integration, maintenance, and replacement of ICT, and the professional development of staff relating to ICT.

Information and Communication Technology (ICT): *Information technology* is the term used to describe the items of equipment (hardware) and computer programs (software) that allow us to access, retrieve, store, organize, manipulate, and present information by electronic means. *Communication technology* is the term used to describe telecommunications equipment through which information can be sought, sent, and accessed, for example, phones, faxes, modems, and computers (Ministry of Education, 2003, p. 7).

Information Systems Planning: The term used to describe the process of determining objectives for organizational computing and identifying potential applications that should be implemented in the organization (Teo & King, 1997).

Integrated ICT Planning: Planning that blends ICT planning or information systems planning with strategic planning.

Legacy Systems: Information systems from an earlier period of technology and development.

Stages of Growth Models: Models that attempt to explain evolutionary stages in the maturity of systems or organizations.

Strategic Goals: Specific goals that provide direction for the organization to meet its mission in the medium term.

ENDNOTES

1 Information systems planning is the term used to describe the process of determining objectives for organizational computing and identifying potential applications that should be implemented in the organization (Teo & King, 1997).

2 Information technology is defined as the hardware and software that allow the access, retrieval, storage, organization, manipulation, and presentation of information by electronic means (Ministry of Education, 1998).

Technology-Mediated Progressive Inquiry in Higher Education

Hanni Muukkonen
University of Helsinki, Finland

Minna Lakkala
University of Helsinki, Finland

Kai Hakkarainen
University of Helsinki, Finland

INTRODUCTION

In higher education, students are often asked to demonstrate critical thinking, academic literacy (Geisler, 1994), expert-like use of knowledge, and creation of knowledge artifacts without ever having been guided or scaffolded in learning the relevant skills. Too frequently, universities teach the content, and it is assumed that the metaskills of taking part in expert-like activities are somehow acquired along the way. Several researchers have proposed that in order to facilitate higher level processes of inquiry in education, cultures of education and schooling should more closely correspond to cultures of scientific inquiry (Carey & Smith, 1995; Perkins, Crismond, Simmons & Under, 1995). Points of correspondence include contributing to collaborative processes of asking questions, producing theories and explanations, and using information sources critically to deepen one's own conceptual understanding. In this way, students can adopt scientific ways of thinking and practices of producing new knowledge, not just exploit and assimilate given knowledge.

BACKGROUND

The best practices in the computer-supported collaborative learning (CSCL) paradigm have several features in common: consideration, in an interrelated manner, of the development of technological applications, use of timely pedagogical models, and attention to the social and cognitive aspects of learning. Emphasis is placed on creating a collaborative community that shares goals, tools, and practices for taking part in an inquiry process.

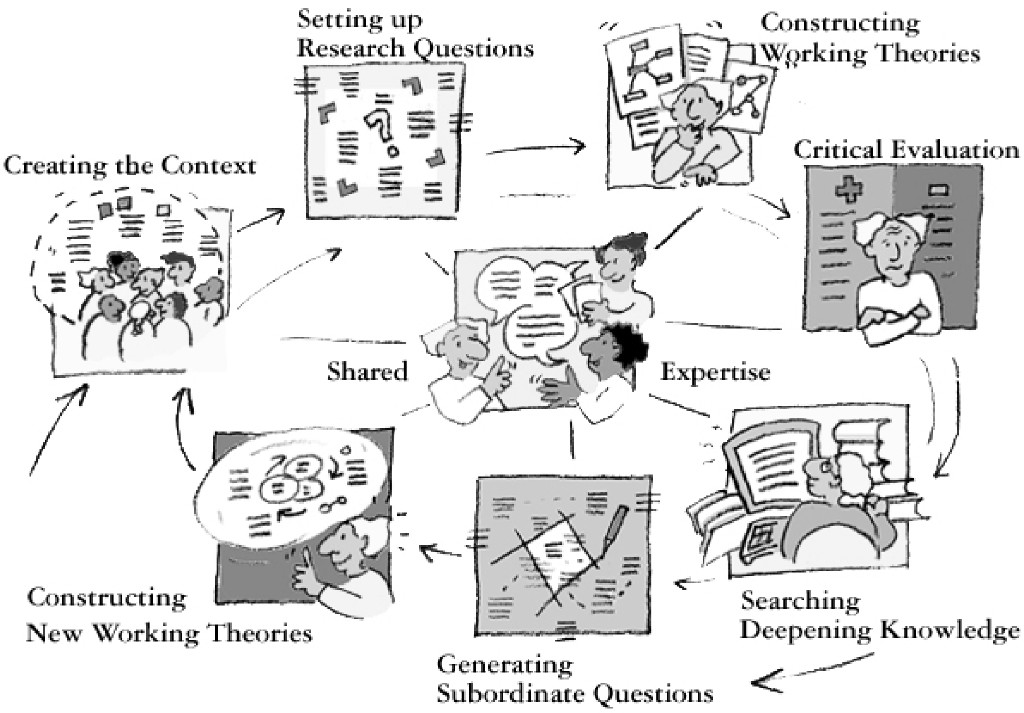
Synthesizing these demands, Kai Hakkarainen and his colleagues at the University of Helsinki have developed a model of *progressive inquiry* as a pedagogical and epistemological framework. It is designed to facilitate expert-like working with knowledge in the context of computer-supported collaborative learning. It is primarily based on Carl Bereiter and Marlene Scardamalia's (Scardamalia & Bereiter, 1994) theory of knowledge building, on the interrogative model of scientific inquiry (Hakkarainen & Sintonen, 2002; Hintikka, 1999), and on the idea of distributed expertise in a community of learners (Brown & Campione, 1994). The model has also been implemented and studied in various educational settings from elementary to higher education (see e.g., Hakkarainen, Järvelä, Lipponen & Lehtinen, 1998; Lakkala, Ilomäki, Lallimo & Hakkarainen, 2002; Lipponen, 2000; Veermans & Järvelä, in press).

THE PROGRESSIVE INQUIRY MODEL

In progressive inquiry, students' own, genuine questions and their previous knowledge of the phenomena in question are a starting point for the process, and attention is drawn to the main concepts and deep principles of the domain. From a cognitive point of view, inquiry can be characterized as a question-driven process of understanding; without research questions, there cannot be a genuine process of inquiry, although in education, information is frequently conveyed or compiled without any guiding questions. The aim is to explain the phenomena in a deepening question-explanation process, in which students and teachers share their expertise and build new knowledge collaboratively with the support of information sources and technology.

The progressive inquiry model specifies certain epistemologically essential processes that a learning community needs to go through, although the relative importance of these elements, their order, and actual contents may involve a great deal of variation from one setting to another. As depicted in Figure 1, the following elements have been placed in a cyclic, but not step-wise succession

Figure 1. Elements of progressive inquiry (Reprinted by permission from Muukkonen et al., 2004)



to describe the progressive inquiry process (Hakkarainen, 2003; Muukkonen, Hakkarainen, & Lakkala, 1999; 2004).

a) *Distributed expertise* is a central concept in the model. Progressive inquiry intends to engage the community in a shared process of knowledge advancement, and to convey, simultaneously, the cognitive goals for collaboration. Diversity in expertise among participants, and interaction with expert cultures, promotes knowledge advancement (Brown et al., 1993; Dunbar, 1995). Acting as a member of the community includes sharing cognitive responsibility for the success of its inquiry. This responsibility essentially involves not only completing tasks or delivering productions on time, but also learners' taking responsibility for discovering what needs to be known, goal-setting, planning, and monitoring the inquiry process (Scardamalia, 2002). There should be development of students' (and experts') social metacognition (Salomon & Perkins, 1998)—students learning to understand the cognitive value of social collaboration and gaining the capacity to utilize socially distributed cognitive resources.

b) The process begins by *creating the context* to anchor the inquiry to central conceptual principles of the domain or complex real-world problems. The learning community is established by joint planning and setting up common goals. It is important to create a social culture that supports collaborative sharing of knowledge and ideas that are in the process of being formulated and improved.

c) An essential element of progressive inquiry is *setting up research questions* generated by students themselves to direct the inquiry. Explanation-seeking questions (Why? How? What?) are especially valuable. The learning community should be encouraged to focus on questions that are knowledge-driven and based on results of students' own cognitive efforts and the need to understand (Bereiter, 2002; Scardamalia & Bereiter, 1994). It is crucial that students come to treat studying as a problem-solving process that includes addressing problems in understanding the theoretical constructs, methods, and practices of scientific culture.

d) It is also important that students explain phenomena under study with their own existing background knowledge by *constructing working theories* before using information sources. This serves a number of goals: first, to make visible the prior (intuitive) conceptions of the issues at hand; second, in trying to explain to others, students effectively test the coherence of their own understanding, and make the gaps and contradictions in their own knowledge more apparent (e.g., Hatano & Inakagi, 1992; Perkins et al., 1995); third, it serves to create a culture in which knowledge is treated as essentially evolving objects and artifacts (Bereiter, 2002). Thoughts and ideas presented are not final and unchangeable, but rather utterances in an ongoing discourse (Wells, 1999).

e) *Critical evaluation* addresses the need to assess strengths and weaknesses of theories and explanations that are produced, in order to direct and regulate the community's joint cognitive efforts. In part, it focuses on the inquiry process itself, placing the process as the center of evaluation, not only the end result. Rather than focusing on individual students' productions, it is more fruitful to evaluate the community's productions and efforts, and give the student participants a main role in this evaluation process. Critical evaluation is a way of helping the community to rise above its earlier achievements, creating a higher level synthesis of the results of inquiry processes.

f) Students are also guided to engage in *searching deepening knowledge* in order to find answers to their questions. Looking for and working with explanatory scientific knowledge is necessary for deepening one's understanding (Chi, Bassok, Lewis, Reiman & Glaser, 1989). A comparison between intuitive working theories produced and well-established scientific theories tends to show up the weaknesses and limitations of the community's conceptions (Scardamalia & Bereiter, 1994). The teacher of a course must decide how much of the materials should be offered to the students and how much they should actually have to search out for themselves. Questions stemming from true wonderment on the part of the students can easily extend the scope of materials beyond what a teacher can foresee or suggest. Furthermore, searching for relevant materials provides an excellent opportunity for self-directed inquiry and hands-on practice in struggling to grasp the differences between various concepts and theories.

g) *Generating subordinate questions* is part of the process of advancing inquiry; learners transform the initial big and unspecified questions into subordinate and more specific questions, based on their evaluation of produced new knowledge. This transformation helps to refocus the inquiry (Hakkarainen & Sintonen, 2002; Hintikka 1999). Directing students to return to previously stated problems, to make more subordinate questions, and answer them are ways to scaffold the inquiry.

h) *Developing new working theories* arises out of the fresh questions and scientific knowledge that the participants attain. The process includes publication of the summaries and conclusions of the community's inquiry. If all productions to the shared database in a collaborative environment have been meaningfully organized, participants should have an easy access to prior productions and theories, making the development of conceptions and artifacts a visible process.

CASES OF PROGRESSIVE INQUIRY IN HIGHER EDUCATION

Progressive Inquiry in a Cognitive Psychology Course

In a study reported by Muukkonen, Lakkala, and Hakkarainen (2001), the progressive inquiry model was implemented in a cognitive psychology course with the use of the Future Learning Environment (FLE). The FLE-environment (http://fle3.uiah.fi) is an open-source collaborative tool that has the progressive inquiry model embedded in its design and functionality (Muukkonen, Hakkarainen, & Lakkala, 1999). All the students in the course were guided, during the first two lectures, to formulate research problems. In the beginning, they individually produced these formulations. They continued by discussing their research problems with a peer and, finally, within a small group, selected the most interesting questions to pursue. These questions were then presented to all the participants in the lecture. After this initial problem setting, the technology-mediated groups (three groups of four to seven volunteers) were instructed to continue their inquiry processes between the weekly lectures in the FLE-environment. The tutor-facilitators took part in the FLE-environment, whereas the teacher conducted the weekly lectures without participating in the database discourse. The rest of the students also formed groups based on their questions, but continued their inquiry process by writing learning logs and commenting on the logs produced by other members of their group without collaborative technology.

A comparative analysis of the knowledge produced by the students in the two conditions provided evidence that the technology-mediated groups were more engaged in problem-setting and redefining practices. Further, they reflected on the process they had undertaken, with respect to the collaboration and their individual efforts. In the productions of the groups who had not used collaboration tools, the social and communal aspects of inquiry and knowledge building were not evident at all in their learning logs, although they were engaged in collaboration during the lectures. The type of the comments they provided to two of the learning logs written by other members of their group were very general, and they concentrated mainly on evaluating the level of writing, not on advancement of ideas. However, many of the learning logs were conceptually well-developed and integrated. Discourse interaction within the FLE environment was different in that the participants sometimes engaged in extensive dialogues about ideas presented by the fellow students.

Progressive Inquiry in a Design Course

Two studies carried out by Seitamaa-Hakkarainen and her colleagues (Seitamaa-Hakkarainen, Lahti, Muukkonen & Hakkarainen, 2000; Seitamaa-Hakkarainen, Raunio, Raami, Muukkonen & Hakkarainen, 2001), analyzed a collaborative design process as it occurred in the complex and authentic design task of designing clothing for premature babies. The framework of the studies was based on evidence from cognitive research on expertise that indicated that novices in design tend to generate problem solutions without engaging in extensive problem structuring; experts, by contrast, focus on structuring and restructuring the problem space before proposing solutions (Glaser & Chi, 1988). The studies described in this case were designed to examine whether an expert-like engagement in design process would be supported in the FLE-environment. Features of the environment were used to encourage the users to engage in expert-like designing and to enable graphic presentation of the knowledge artifacts in the form of importing students' drafts and prototypes into the collaborative environment and developing multiple versions of the designs.

During the collaborative design course, the students were first guided to find out information about the constraints of their design task, such as the size of the babies, special needs for the usability of the clothing, and about the materials. Then they were asked to produce their own sketches and work in small groups to share design ideas and develop their designs. Following this development, each group produced a prototype, which was tested by actual end-users in hospital. Feedback and suggestions were then used to develop advanced design ideas.

In these studies of designing with the support of a networked collaborative environment, Seitamaa-Hakkarainen and her colleagues (2001) found that a key aspect of these environments is its provision of tools for progressive discourse between the designers and users of the future products. Further, the environments offer shared spaces and tools to elaborate conceptual knowledge related to the design problem. The collaborative technology made design thinking more explicit and accessible to the fellow designers and enabled participants to share their ideas and construct a joint understanding of design problems and solutions.

TUTOR'S ROLE AND ACTIVITY

A special question in implementing progressive inquiry and knowledge building practices in higher education is the teacher's or tutor's role in supporting and guiding students' collaborative inquiry. In progressive inquiry, the traditional role of a teacher as an expert who delivers the essential information by lecturing is radically changed. The important roles of the teacher and the facilitators of collaboration are to create the context for collaboration, and provide anchors between the theoretical representations, world knowledge, and the real-life experiences that students report (Muukkonen et al., 2004). It is also necessary for the teacher to structure and scaffold the process, keep it active and in focus during the progression of the course, and to help students to gradually take upon themselves the responsibility for the higher level cognitive processes (Scardamalia, 2002).

FUTURE TRENDS

Productive changes in educational systems towards establishing inquiry-based approaches in studying and teaching call for an alignment of epistemic, pedagogical, and institutional goals and actions (Muukkonen et al., 2004). Learning technologies also need to be critically viewed for their role in fostering expert-like skills in advancing knowledge. For instance, availability of scaffolding, support for multiple forms of collaboration, and shared development of knowledge objects are challenges for designing learning technologies.

CONCLUSION

The progressive inquiry model may be utilized in a variety of educational settings to provide a heuristic framework for the key activities and epistemic goals of a knowledge-building community. The community may provide multiple levels of expertise and, equally important, social support for engaging in a strenuous quest for learning and advancing knowledge. In higher education, a progressive inquiry approach may support the development of academic literacy, scientific thinking, and epistemic agency, particularly when integrated with the use of appropriate collaborative technology and supportive arrangements in curriculum design.

T

REFERENCES

Bereiter, C. (2002). *Education and mind in the knowledge age*. Hillsdale, NJ: Erlbaum.

Brown, A.L. & Campione, J.C. (1994). Guided discovery in a community of learners. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 229-287). Cambridge, MA: MIT Press.

Brown, A.L., Ash, D., Rutherford, M., Nakagawa, K., Gordon, A., & Campione, J. (1993). Distributed expertise in the classroom. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations* (pp. 188-228). Cambridge,UK: Cambridge University Press.

Carey, S. & Smith, C. (1995). On understanding scientific knowledge. In D.N. Perkins, J.L. Schwartz, M.M. West, & M.S. Wiske (Eds.), *Software goes to school* (pp. 39-55). Oxford, UK: Oxford University Press.

Chi, M.T.H., Bassok, M., Lewis, M.W., Reiman, P., & Glaser, R (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145-182.

Dunbar, K. (1995). How scientist really reason: Scientific reasoning in real-world laboratories. In R.J. Sternberg, & J. Davidson (Eds.), *Mechanisms of insight* (pp. 365-395). Cambridge, MA: MIT Press.

Geisler, C. (1994). *Academic literacy and the nature of expertise*. Hillsdale, NJ: Erlbaum.

Glaser, R., & Chi, H. T. M. (1988). Overview. In H.T.M. Chi, R. Glaser, & M. Farr (Eds.), *The nature of expertise* (pp. xv-xxviii). Hillsdale, NJ: Erlbaum.

Hakkarainen, K. (2003). Emergence of progressive inquiry culture in computer-supported collaborative learning. *Learning Environments Research, 6*, 199-220.

Hakkarainen, K., & Sintonen, M. (2002). Interrogative model of inquiry and computer-supported collaborative learning. *Science & Education, 11*, 25-43.

Hakkarainen, K., Järvelä, S., Lipponen, L. & Lehtinen, E. (1998). Culture of collaboration in computer-supported learning: Finnish perspectives. *Journal of Interactive Learning Research, 9*, 271-287.

Hatano, G., & Inagaki, K. (1992). Desituating cognition through the construction of conceptual knowledge. In P. Light & G. Butterworth (Eds.), *Context and cognition: Ways of knowing and learning* (pp. 115-133). New York: Harvester.

Hintikka, J. (1999). Inquiry as inquiry: A logic of scientific discovery. *Selected papers of Jaakko Hintikka*, Volume 5. Dordrecht, The Netherlands: Kluwer.

Lakkala, M., Ilomäki, L., Lallimo, J. & Hakkarainen, K. (2002). Virtual communication in middle school students' and teachers' inquiry. In G. Stahl (Ed.), Computer support for collaborative learning: Foundations for a CSCL community. *Proceedings of CSCL 2002* (pp. 443-452). Hillsdale, NJ: Erlbaum. Available online: *http://newmedia.colorado.edu/cscl/97.html*

Lipponen, L. (2000). Towards knowledge building discourse: From facts to explanations in primary students' computer mediated discourse. *Learning Environments Research, 3*, 179-199.

Muukkonen, H., Hakkarainen, K., & Lakkala, M. (1999). Collaborative technology for facilitating progressive inquiry: future learning environment tools. In C. Hoadley & J. Roschelle (Eds.), *Proceedings of the Computer Support for Collaborative Learning (CSCL) 1999 Conference* (pp. 406-415). Mahwah, NJ: Erlbaum.

Muukkonen, H., Hakkarainen, K., & Lakkala, M. (2004). Computer-mediated progressive inquiry in higher education. In T.S. Roberts (Ed.), *Online collaborative learning: Theory and practice* (pp 28-53). Hershey, PA: Information Science Publishing.

Muukkonen, H., Lakkala, M., & Hakkarainen, K. (2001). Characteristics of university students' inquiry in individual and computer-supported collaborative study process. In P. Dillenbourg, A. Eurelings & K. Hakkarainen (Eds.), *European perspectives on computer-supported collaborative learning. Proceedings of the first European conference on CSCL* (pp. 462-469). Maastricht, The Netherlands: Maastricht McLuhan Institute.

Perkins, D.A., Crismond, D., Simmons, R., & Under, C. (1995). Inside understanding. In D.N. Perkins, J.L. Schwartz, M.M. West, & M.S. Wiske (Eds.), *Software goes to school* (pp. 70-87). Oxford, UK: Oxford University Press.

Salomon, G. & Perkins, D. N. (1998). Individual and social aspects of learning. *Review of Research of Education 23*, 1-24.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67-98). Chicago, IL: Open Court.

Scardamalia, M. & Bereiter, C. (1994). Computer support for knowledge-building communities. *The Journal of the Learning Sciences, 3*, 265-283.

Seitamaa-Hakkarainen, P., Lahti, H., Muukkonen, H., & Hakkarainen, K. (2000). Collaborative designing in a networked learning environment. In S.A.R. Scrivener, L.J. Ball, & A. Woodcock (Eds.), *Collaborative design: The proceedings of CoDesigning 2000* (pp. 411-420). London, UK: Springer.

Seitamaa-Hakkarainen, P., Raunio, A.M., Raami, A., Muukkonen, H., & Hakkarainen, K. (2001). Computer support for collaborative designing. *International Journal of Technology and Design Education, 11*, 181-202.

Veermans, M. & Järvelä, S. (in press). Generalized achievement goals and situational coping in inquiry learning. *Instructional Science*.

Wells, G. (1999). *Dialogic inquiry: Towards a sociocultural practice and theory of education*. Cambridge, UK: Cambridge University Press.

KEY TERMS

Distributed Expertise: Cognition and knowing are distributed over individuals, their tools, environments, and networks.

Epistemic Agency: Taking responsibility for one's own learning efforts and advancement of understanding.

Knowledge Building: A framework for collective knowledge advancement and development of knowledge artifacts.

Learning Community/Community of Learners: All participants in a learning process (students, teachers, tutors, and experts) have valuable expertise and skills, which benefit collective efforts.

Metaskills: Skills involved in academic literacy as well as metacognitive skills related to planning, monitoring, and regulating comprehension-related activities.

Progressive Inquiry: A pedagogical model for structuring and supporting a group of learners in a deepening question-explanation process.

Scaffolding: Providing support that enables a learner to carry out a task that would not be possible without that support, and enabling the learner gradually to master that task without support.

Technology's Role in Distance Education

T

Murray Turoff
New Jersey Institute of Technology, USA

Caroline Howard
Touro University International, USA

Richard Discenza
University of Colorado at Colorado Springs, USA

INTRODUCTION

Learning is enhanced by the physical and social technologies typically used in distance education. Students in distance programs typically have access to tools that allow them to repeat lectures and interact with their fellow students and faculty. Students in all classes, including face-to-face and blended courses, benefit from having similar tools and technologies available. This article will review common tools and technologies used in distance education, and demonstrate why they can facilitate learning and expand the educational opportunities for both distant and traditional students.

BACKGROUND

For many years technologies have been used to facilitate learning. In the early 1980s a group of researchers at the New Jersey Institute of Technology (NJIT) realized the enormous potential of the technology to enhance learning when they used a computer-mediated system to facilitate a regular face-to-face class. The system was introduced to students in a number of Computer Science and Information System courses. Due to the amount of material covered in lectures, there was not much time for dialogue, and only a few students participated when there was a class discussion. The instructors introduced asynchronous group communication technologies to communicate discussion questions and assigned grade point credits for student participation. One hundred percent of the students participated in these discussions outside of regular classroom hours. The extent and depth of the discussions changed the nature of the classes. Most importantly, student contributions were comprehensive, with more well-thought-out comments, because students had the time to reflect on the ongoing discussion before participating. Also very significant was that students, for whom English was a second language, became equal participants. They could reread the online discussion as many times as needed before replying. The computer-based activity monitoring and transcripts, electronic recordings of the discussions, showed that foreign students spent two to three times more time in a reading mode and reread many discussions, far more than the American students.

In addition, professors now have the ability to monitor activities and review the electronic transcripts of student involvement which gives the instructor insights into how students are learning. By reviewing the transcripts of the online discussions, it becomes obvious what and how students are learning. For courses with a high pragmatic content, such as upper-level and graduate courses in topics like the design and management of computer applications, students are required to utilize problem-solving approaches to evaluate the tradeoffs between conflicting objectives. In a traditional classroom environment, especially in large classes, it is very difficult to detect whether students are accurately incorporating the problem-solving mental models that the instructor is attempting to convey. When instructors review the transcripts of class discussions, they give insights into the approaches students are taking to master the material. Unfortunately, in the early 1980s few wanted to hear about a revolution in normal classroom teaching or were willing to expend the effort to dramatically improve classroom education. It was only the rise of distance education that generated interest in learning about the educational potential of the technology.

Hiltz (1994) performed quasi-experimental studies that compared a population of NJIT students (only familiar with face-to-face classroom education) to a population of students taking the same courses in pure face-to-face sections, with pure distance sections using only CMC technology. The students in the matched sections had the same material, the same assignments, the same exams, and the same instructor. They found no significant difference in the amount of learning or the rate of student satisfaction. This finding is much more significant than a determination based on a study that included a population of distance learners already familiar with traditional correspondence classes.

Two critical underlying variables driving the success of this approach were identified by Hiltz (1994). First, the role the instructor needed to take was different from the traditional classroom role. The instructor acted more as an active and dedicated facilitator, as well as a consulting expert on the content of the course, rather than traditional teacher. Second, collaborative learning and student teamwork were the educational methodology (Hiltz, 1994) shown in later studies to be a key factor in making distance courses as good as or better than face-to-face courses (Hiltz & Wellman, 1997). These results show that distance courses can be as effective as face-to-face courses when using any of the traditional measures, such as exams and grades.

Creative, interactive software programs accompanied by background tutoring can effectively teach students to master the skills currently taught in many undergraduate courses. When these courses are automated, the costs incurred are far below typical college tuition. In the future, colleges and universities will not be able to continue to charge current tuition costs for introductory courses that are largely skill oriented. For example, there are many stand-alone and Web-based software programs that offer introductory programming courses, as well as skills in many other areas. These courses are comparable to college courses and some are even based upon a textbook used on some college campuses. They are available for a few hundred dollars. The major difference is that they do not carry college credits.

The technology allows senior professors or department chairs to effectively evaluate and mentor all instructors of particular courses, whether they are teaching traditional classroom courses or distance courses. The ability to review whole class discussions after the class is over gives senior faculty the ability to evaluate distance instructors hired to teach previously developed courses, as well as to review on-site instructors and junior faculty. Thus, they can improve and extend their mentorship and apprenticing relationships.

Today's technology for distance education allows faculty members to live anywhere they want to. Unique benefits will be available to outstanding teaching faculty. For example, one of the best full-time instructors for NJIT, which is located in beautiful downtown Newark, is a mother with two small children who never has to be on campus. She is teaching other instructors how to teach remotely. Similarly, a University of Colorado accounting professor, on sabbatical in Thailand, is able to teach a course in the Distance MBA program. There have been a few master's programs where some or all of the instructors are located anywhere in the world. It is technically feasible for those wanting to escape winter cold to teach in places such as Hawaii that we could only dream about. The technology makes it feasible, but various administrative policies, unions, insurance companies, benefit programs, and so forth have not yet caught up to the technology. There is increasing emphasis by accrediting agencies on treating remote instructors the same as faculty are treated. This is likely to bring about a greater degree of equality between instructors and tenured track faculty. The outcome is uncertain, but it may mean that the costs for remote and traditional classes will equalize so that the profit margin in online classes will not be quite so high.

SPECIFIC FUNCTIONS OF TECHNOLOGY THAT FACILITATE LEARNING

Asynchronous Discussions

In the online environment, students can take as much time as they need to reflect on a discussion and polish their comments. This improves the quality of the discussion, and changes the psychology and the sociology of communications. Students can address topics in the sequence they choose rather than in a predefined order. This leads to the development of different problem-solving strategies among the individual members of the class. Sometimes courses include synchronous conferences, videoconferencing, and/or video presentations to supplement asynchronous discussions.

Instructor Control of Online Conference and Roles

With online course conferences (many per course), instructors control the membership of each, assign roles, and enable other instructors to monitor conferences for joint teaching exercises involving more than one course. Groups within courses are able to set up private online conferences for team and collaborative work group assignments. Joint editing of items facilitates teamwork.

Question-and-Answer Communication Protocol

Instructors are able to ask questions during discussions. They can control who views the answer and prevent other students from seeing the answer of the others or engaging in the resulting discussion until they have entered their answer. In studies of Group Decision Support Systems, it has been shown that asynchronous groups in an online Delphi mode generate many more ideas than unstructured discussions or face-to-face groups of similar size (Cho, Turoff, & Hiltz, 2003). This area has proven to be a valuable tool in forcing equal participation. Use of question-and answer communication protocol can be used to force each student to think independently through their answer without being influenced by the other students.

Anonymity and Pen Name Signatures

When students with work experience are part of a discussion, they can use their real-life experiences to illustrate the concepts the professor is presenting. Such comments from fellow students, rather than the professor, often make the instructor's message more meaningful to the students. A student confirming the theory presented by a faculty member through real-life examples is more effective in making a point than "dry" data from an instructional article. Furthermore, students can talk about disasters in their companies with respect to decisions in any area, and they can provide detail—including costs—when they are not identified and the anonymity of the company they work for is preserved. Also, the use of pen names allows individuals to develop alternative personas without divulging their real identity and is extremely useful in courses that wish to employ role playing as a collaborative learning method.

Membership Status Lists

The monitoring of activities, such as students' reading and responding to communications, allows the professor to know what each individual has read and how up-to-date they are in the discussion. This allows the instructor to detect when a student is falling behind. Student collaborative teams can make sure that everyone in the team is up to date. Furthermore, students can easily compare their frequency of contributions relative to other students in the course.

Voting

Instant access to group and individual opinions on resolutions and issues are enabled by voting capabilities. This is useful for promoting discussion, and the voting process is continuous so that changes of views can be tracked by everyone. Voting is not used to make decisions. Rather, its function is to explore and discover what are the current agreements and disagreements or uncertainties (polarized vs. flat voting distributions) so that the class can focus the continuing discussion on the latter. Students may change their votes at any time during the discussion.

Special Purpose Scaling Methods

These useful methods show true group agreements and minimize ambiguity. Currently we have a system that allows each student at the end of the course to contribute a statement of what they think is the most important thing they have learned in the course; then, everyone votes by rank ordering all the items on the list. The results are reported using Thurstone's scaling, which translates the rank order by all the individuals to a single group interval scale. In this interval scale, if 50% prefer A to B and 50% prefer B to A, the two items will be at the same point on the scale. It has been surprising what some of the results have been in some courses. For example, in a Management of Information Systems course, the concept of "runaway" software projects was felt to be twice as important as any other topic. The professor was quite surprised by this result, until he began to realize that the students were using this concept as a mental model in which to integrate many of the other things they had learned.

Information Overload

This occurs when enthusiastic discussions by students that are meant to augment the quality of the learning process augment only the quantity of the number of comments, instead leading to the problem of "information overload." Currently this phenomenon limits the size of the group in a single CMC class. Online discussions allow individuals to enter comments whenever it is convenient for them, without waiting for someone else to finish the point they were trying to make. This makes it physically possible and also very likely that a great deal more discussion will take place and much more information will be exchanged among the group than if only one person can speak at a time, as in the face-to-face classroom environment. Anything that reduces the temptation of some students to "contribute" comments or messages that have nothing to do with the meaningful discussions underway will increase the productivity of the discussion without information overload setting in. Among such functional tools the computer can provide are:

Class Gradebooks

This eliminates a tremendous amount of electronic mail traffic that would become very difficult for an individual instructor to manage with a large class.

Selection Lists

The instructor can set up lists of unique choices so that each student may choose only one item and others can see who has chosen what. This is very efficient for conveying individualized assignments and reduces a large portion of communications.

Factor Lists

Members of a class or group can add ideas, dimensions, goals, tasks, factors, criteria, and other items to a single, shared list which may then be discussed and modified based upon that discussion and later voted upon.

Notifications

Short alerts notify individuals when things occur that they need to know about. For instance, students can be notified that a new set of grades or vote distribution has been posted, eliminating the need for individuals to check for these postings. People can attach notifications to conference comments from a select list that provides alternatives like: *I agree, I disagree, I applaud, Boo!* Such appendages significantly reduce the need to provide paralinguistic cues of reinforcement as additional separate comments.

Calendars, Agendas, or Schedules

Students have access to a space to track the individual and collaborative assignments and their due dates. These are listed in an organized manner that links detailed explanations for each assignment, as well as questions and answers related to the assignments.

The authors have seen these technologies facilitate learning beyond what can be assessed using traditional measures. Some of the more subtle intangible benefits of technology that we have observed are:

- Due to social pressures, students tend to be more concerned with how other students view their work quality than how the professor views it. They are significantly more motivated to participate in a meaningful way when their fellow students can view their contributions.
- When equality of communications is encouraged, students cannot get away with being passive or lazy. The transcript or electronic recording of the discussions shows who is and is not participating. It is visible to both the instructor and other students that someone is being lazy. (In fact, students seem to be more concerned with what the other students will think of their performance than what the professor will think.)
- The scope of what the outstanding students learn becomes even more noticeable.
- The performance of students at the lower end of the distribution is improved. The communications systems permit them to catch up, because they are able to obtain a better understanding of the material with which they are most uncomfortable or have the least background knowledge.
- The instructor can become more aware of his/her successes or failures with individual students because of the reflective nature of the student contributions to the discussion.

While these dimensions and concepts need confirmation through long-term longitudinal studies of student performances, the marketplace is also providing confirmation of the beliefs held by many experienced in teaching these classes. We are seeing that collaboratively oriented programs offer a solution to the problems, which are inherent in traditional correspondence courses. Students benefit from the ability to electronically store lectures alone or in chunks integrated into other material on the Web. Electronic storage of lectures gives all students the power to choose freely whether they want to attend a face-to-face class or take the same course remotely. Traditional face-to-face students can later hear a lecture missed due to illness or travel. Students with English as a second language can listen to a lecture multiple times. Face-to-face students who have to travel or fall sick can use the same tapes to catch up and/or review material prior to exams.

In our view a student in a face-to-face class that is not augmented by a collaborative learning approach and by asynchronous group communications technology is not getting as good of an education as the distance student who has those benefits. It is the face-to-face student who may be suffering from the segregation of the college system into separate face-to-face and distance courses. These observations about the past and the present lead to some speculations about the future.

THE STATE OF THE TECHNOLOGY

The technology available today includes at least 250 versions of group communication software. However, many of them may not survive the decade. There are a growing number of software packages for course management. The online learning product landscape is changing at a rapid pace as companies are acquiring their competitors to expand functionality. Gray (2002) gives an excellent summary of the popular platforms and the evolving nature of e-learning.

There are only a few of these that have wide usage, and they are beginning to raise their prices to capitalize on their popularity. Most of these packages charge a fee per user, which is not the desirable fee structure for the customer. Many of the older conference systems charge on a per-server basis and it does not matter how many students one has. It is far cheaper to spend more on the hardware and a get a more powerful server. Also, the course management systems do not provide many of the useful software features one would like to have for group communications. Given the way prices are going, it might be better to pay some of the undergraduate students to educate some of the faculty on how to create their own Web sites and have their own pages for their courses that they update and maintain directly. This also has desirable long-term consequences in raising the ability of the faculty in this area. Once you have committed all your content to one vendor's system, you are a captured customer and will have to pay whatever they want to charge. Right now, software development is undergoing rapid evolution, and no customer should put themselves in the box of only being able to use one vendor. If it is clear that you are using a number of vendors, you may even be able to get some breaks on pricing and will certainly get the top level of service when each of them knows there is an alternative service readily available to the customer. In the coming decade, one can expect major upgrades for these software systems every few years, and the best one today may not be the best one tomorrow.

COURSE DEVELOPMENT AND DELIVERY

Unfortunately, many faculty do not know how to use the technology to design a successful distance course. As the historical record shows, it is a mistake when transferring an application to computers to just copy the way it used to be done onto the computer. Utilizing the methodology of collaborative learning is the key to designing courses using group communications technology. Simple systems, which attempt to impose a discussion thread on top of what is electronic mail technology, allow the student or the teacher only to view one comment at a time. This approach does not allow an individual to grasp the totality of any complex discussion. Only by placing the complete discussion thread in a single scrolling page can a person review and understand a long discussion. One can browse the discussion and cognitively comprehend it without having to perform extra operations and lose their cognitive focus. Users of such simple systems cannot generate a large complex discussion and have no way of realizing that complex discussion is even possible. When online discussions are successful, they can easily go from enthusiastic wonderful discussions to information overload. Current technology must evolve to fully support collaborative learning.

FUTURE TRENDS

To facilitate collaborative learning, critical development directions for the future should include:

- Tailorability of communication structures by instructor
- Tailorability of communication protocols by instructor
- Anonymity and pen name provisions
- Delphi method tools and the availability of scaling and social judgment
  (voting methods)
- Tools for collaborative model building
- Powerful information retrieval capabilities
- Tailorability by instructor of application-oriented icons and graphical components
- Tools for the analysis of alternative diagrams

Instructors also need to allow students to extend the discourse structure and to vote on the significance of incidents of relationships among factors in the problem domain by using Group Decision Support processes. These systems should allow students not only to develop their own conceptual maps for understanding a problem, but also to detect disagreements about elements of the conceptual map and the meanings of terms. This is valuable preparation for problem solving in their professional life, a process that requires removing inherent ambiguities and individual meanings in the language used to communicate about a problem with others from diverse backgrounds. Routines should be included that are based upon both scaling and social judgment theories, which improve the ability of larger groups to quickly reach mutual understanding. Currently, few tools exist in current systems that support the use of collaborative model building, gaming, and Delphi exercises. The current generation of software does not often include the functions of anonymity and pen names. Course instructors need to have complete control over course communication structures and processes, and should be able to use their recently acquired knowledge for future offerings of the course. Currently, systems lack the needed integration of functions to easily evolve the changes in both the relationships and the content in a given field. A long-term advantage of teaching in the collaborative electronic environment is that the students create useful material for future offerings and can aid the instructor in monitoring the new professional literature. Future technology will allow faculty to organize their material across a whole set of courses into a collaborative knowledge base available to the faculty teaching those courses. This would allow students and faculty to create trails for different objectives and weave the material in that knowledge base to suit a group of students or a set of learning objectives.

Individual learning teams would be able to progress through a degree program's knowledge base at the rate best for them, rather than setting the same timeframe for all learning teams or faculty teams. Faculty, individuals, or teams would take responsibility for a specific domain within the Web of knowledge representing a degree program.

CONCLUSION

Collaborative technologies are changing the concept of what constitutes a course. Program material could be an integrated knowledge Web, based largely on semantic hypertext structures. Over time, the domain experts—the faculty—would continue to develop and evolve their parts of the Web, and wait for learning groups, composed of any mix of distance and regular students sharing the same learning objectives and needs.

Current vendor systems focus on the mass market and concentrate on tools to standardize and present course content. Group communication tools are usually just disguised message servers that offer only a discussion thread capability and little more, certainly not the complex capabilities discussed above. Vendors have not yet recognized the primary importance of group communications and how faculty members can guide and facilitate the process and be available for consultation as needed. Based upon the conceptual knowledge maps they design, faculty should be encouraged to develop content structures that are characteristic of their subject matter. In the end, faculty should have the ability to insert group communication activities anywhere in their professional knowledge base (e.g., question/answers, discussion threads, lists, voting, etc.).

REFERENCES

Cho, H.K., Turoff, M. & Hiltz, S.R. (2003, January). The impacts of Delphi communication structure on small and medium sized asynchronous groups. *HICSS Proceedings (Hawaii)*,Piscataway, NJ: IEEE Press.

Discenza, R., Howard, C. & Schenk, K. (Eds.). (2002). *The design and management of effective distance learning programs*. Hershey, PA: Idea Group Publishing.

Gray, S. (2003). Moving—e-learning vendors take aim in the changing environment. *Syllabus, 16*(1), 28-31.

Harasim, L., Hiltz, R., Teles, L. & Turoff, M. (1995). *Learning networks: A field guide to teaching and learning online.* Boston: MIT Press.

Hiltz, S.R. (1993). Correlates of learning in a virtual classroom. *International Journal of Man-Machine Studies*, *39*, 71-98.

Hiltz, S.R. (1994). *The virtual classroom: Learning without limits via computer networks.* Human Computer Interaction Series. London: Intellect Press.

Hiltz, S.R. & Turoff, M. (1993). *The network nation: Human communication via computer*. Boston: MIT Press (original edition 1978).

Hiltz, S.R. & Turoff, M. (2002). What makes learning networks effective? *Communications of the ACM*, (April), 56-59.

Hiltz, S.R. & Wellman, B. (1997). Asynchronous learning networks as a virtual classroom. *Communications of the ACM*, *40*(9), 44-49.

Howard, C. & Discenza, R. (1996, October). *A typology for distance learning: Moving from a batch to an on-line educational delivery system. Proceedings of the Information Systems Educational Conference* (ISECON), St. Louis, Missouri.

Howard, C. & Discenza, R. (2001). The emergence of distance learning in higher education: A revised group decision support system typology with empirical results. In L. Lau (Ed.), *Distance education: Emerging trends and issues.* Hershey, PA: Idea Group Publishing.

McIntyre, S. & Howard, C. (1994). Beyond lecture-test: Expanding focus of control in the classroom. *Journal of Education for Management Information Systems,* (November).

Nelson, T.H. (1965). A file structure for the complex, the changing and the indeterminate. *ACM 20th National Conference Proceedings* (pp. 84-99).

Turoff, M. (1995). A marketplace approach to the information highway. *Boardwatch Magazine,* (April).

Turoff, M. (1996). Costs for the development of a virtual university. *Journal of Asynchronous Learning Networks*, *1*(1).

Turoff, M. (1997). Virtuality. *Communications of the ACM*, *40*(9), 38-43.

Turoff, M. (1998). Alternative futures for distance learning: The force and the dark side. *Online Journal of Distance Learning Administration*, *1*(1).

Turoff, M. (1999). Education, commerce, & communications: The era of competition. *WebNet Journal: Internet Technologies, Applications & Issues*, *1*(1), 22-31.

Turoff, M. & Hiltz, R.S. (1986). Remote learning: Technologies and opportunities. *Proceedings of the World Conference on Continuing Engineering Education.*

Turoff, M. & Hiltz, R.S. (1995). Software design and the future of the virtual classroom. *Journal of Information Technology for Teacher Education*, *4*(2), 197-215.

Turoff, M & Discenza, R. (forthcoming). Distance learning: Really a better education? in C. Howard, K. Schenk & R. Discenza (Eds.), Distance learning and university effectiveness: Changing educational paradigms for online learning. Hershey, PA: Information Science Publishing.

Turoff, M., Hiltz, R., Bieber, M., Rana, A. & Fjermestad, J. (1999). Collaborative discourse structures in computer mediated group communications. Reprinted in *Journal of Computer Mediated Communications on Persistent Conversation*, *4*(4).

KEY TERMS

Asynchronous Group Communication Technologies: Allow participants to send and respond to messages without being online simultaneously.

Distance Education: Consists of learning situations in which the students and instructor are located in different localities at least for a portion of the class.

Distributed Learning: Consists of learning situations in which the students and instructor are located in different localities. A bit broader than distance education as it can be used to refer to both education and training.

E-Learning: Using technology to assist in the educational process. It is often used to refer to learning situations (both education and training) in which the students and instructor are located in different localities . However, the instructor and teacher can be in close proximity.

E-Learning Technologies: The technologies used for e-learning.

Pen Name Signatures: Names participants choose for online participation which may or may not allow other participants to identify them.

Synchronous Group Communication Technologies: Allow real-time, interactive communications and require participants to be online simultaneously.

ENDNOTES

1 A great deal of recent evaluation studies are beginning to confirm ourearlier findings based upon extensive and large-scale studies at such placesas SUNY, Drexel, Penn State, and others. Some of these may be found inthe *Journal of ALN (www.aln.org)* and on the ALN EvaluationCommunity Web site *(www.alnresearch.org)*

T

Telemedicine in Healthcare Organisations

Nabeel A.Y. Al-Qirim
United Arab Emirates University, UAE

INTRODUCTION

Diminishing funds from the government and cost control led many health care providers across the globe to search for alternative and more cost-effective means of providing care (*Edelstein, 1999;* Neame, 1995). In many cases, this has become necessary for survival *(Edelstein, 1999) in order to sustain the increased competition as well amongst health care providers*. The business of health care has become so competitive that many small rural hospitals are trying to align themselves with larger tertiary care centres in a community health information network, a telemedicine network, or some other type of partnership in order to survive and to retain their local patients (Huston & Huston, 2000).

Telemedicine means medicine from a distance where distant and dispersed patients are brought closer to their medical providers through the means of telecommunication technologies (Charles, 2000; Noring, 2000; OTA, 1995; Perednia & Allen, 1995; Wayman, 1994). Noring (2000) provided an interesting comparison between the former definition of telemedicine and tele-health, where the term tele-health is defined as expanding the capacity of telemedicine to provide the full continuum of care, from health promotion and disease prevention through curative treatment and terminal care. Tele-health also implies including non-physician based health care providers.

Some researchers even envision telemedicine to be an important building block in the strategic plan of many health care organizations (Charles, 2000). Within these challenges, telemedicine emerges as one possible solution to health providers in reaching out to rural patients (Charles, 2000; Harris, Donaldson & Campbell, 2001), to areas where patient volumes for certain services are limited *(Edelstein, 1999)*, to conduct administrative and clinical meetings, and to conduct different training courses to: patients (smoke treatment centres), doctors, nurses, and other medical staff (Perednia & Allen, 1995; Wayman, 1994).

This research is interested in introducing the concepts underlying the telemedicine technology. The research will highlight the different advantages/disadvantages of this technology, and hence project different motivators and inhibitors to the adoption and use of this innovative technology in health care delivery.

BACKGROUND AND IMPLICATIONS

The first telemedicine initiative emerged in 1959 by employing video conferencing (VC) sessions for medical purposes by using microwave link for telepsychiatry consultations between the Nebraska Psychiatric Institute in Omaha and the State Mental Hospital 112 miles away (Perednia & Allen, 1995). Since then, telemedicine started to grow in different parts in the world. For example, in the late 1980s telemedicine was being used routinely to deliver general health services to remote regions of Norway (Noring, 2000). In the 1990s, telemedicine started to emerge in New Zealand and to be used successfully in Australia. In the United States, interest in telemedicine was initially focused on its use in the military, in space programs, on offshore oilrigs, in prisons, and in rural areas (Noring, 2000).

Telemedicine covers a wide spectrum of benefits through its use in areas such as consultations, diagnostics, therapeutics, transfer of patient related records, case management, training, and meetings. In a rural setting, telemedicine could help health care providers in supplying quality, fast, and economical medical services to rural patients, and hence saves doctors and patients valuable time wasted in commuting large distances. Specialists could utilise this extra time in seeing more patients at the main hospital.

The applications of telemedicine vary from full-motion and interactive video consultations to “store and forward” technologies where static images or audio-video clips are electronically captured, stored, and transmitted to a remote server such as electronic mail (e-mail) using public or private communication channels. The advantage of store-and-forward technology is that it prevents the need for simultaneous availability of the consulting parties. Radiology, dermatology, and pathology are especially suited to a store-and-forward format. Let’s not forget that the earliest form of telemedicine technology was the telephone system, where physicians used to call their patients (e.g., psychiatry) to follow-up a treatment or to check whether they are taking their medicine on time.

However, the telemedicine technology was not successful in its initial stages. Perednia and Allen (1995) reported limited telemedicine growth and pointed to the fact that only few telemedicine projects were instituted in the 1970s and 1980s at several sites in North America and Australia. They confirmed that none of the programs begun before 1986 has survived. Although data are limited, the early reviews and evaluations of those programs suggest that the equipment was reasonably effective at transmitting the information needed for most clinical uses and that users were for the most part satisfied. However, when external sources of funding (donations) were withdrawn, the programs disappeared, indicating that the single most important cause of their failure was the inability to justify these programs on a cost-benefit basis and the failure of these hospitals in addressing the importance of the telemedicine technology as one important tool amongst the other important medical tools in the hospital. Other issues, such as limited physician acceptance, played a less significant role in their downfall (Perednia & Allen, 1995).

However, there are other reasons that affect telemedicine success in health care organisations. Technological limitations were one of the major impediments. For example, until recently, transmission of a high-resolution, real-time video signal was possible only through the use of expensive equipment such as satellite systems and microwave towers. Recent developments in digitisation and data-compression technologies allow transmission of enormous amounts of information needed for video with much less bandwidth at a much lower price. While the equipment cost for video conferencing systems is relatively high, decreasing costs of hardware, software, and transmission suggest that high-resolution, full-motion VC may soon be available to medical clinics and offices at very economical prices. The wide diffusion of wired and wireless telecommunications infrastructure in rural areas has assisted many rural health care centres to interlink with speciality hospitals very easily.

Due to such technological facilitations and enhancements and increased awareness and knowledge about telemedicine technologies and applications, telemedicine witnessed another growth-phase in the 1990s. The innovative return of telemedicine since the early 1990s inspired many innovative ideas. For example, BMI British Midland has become the first airline to install telemedicine technology on planes making long-haul flights to the U.S. Virgin Atlantic has also purchased the system and will start installing it in long-haul aircraft this year. The device monitors blood pressure via a wrist cuff, pulse rate, temperature via an ear probe, electrocardiogram, blood oxygen and carbon dioxide levels. Using a modem, these are sent to physicians who can advise the crew on what action to take (Anonymous, 2002).

However, this growth was not in a straightforward manner and the telemedicine technology started to face different forms of complications. As the technology penetration in health care increases, different organisational, technological, environmental, political, individual and economical factors started to surface and to influence its success. Initially, the views about VC effectiveness in the medical area vary from one adopter to another. Depending on one's viewpoint, VC can be seen as a valuable tool for providing immediate specialty care services to rural areas, a more efficient use of existing medical resources, a way to attract patients living outside a hospital's normal service area, and a way of bringing international health care dollars. On the other hand, others could see it as a serious misallocation of increasingly scarce health care dollars (Perednia & Allen, 1995).

In review of the literature it was observed that despite the rapid growth and high visibility of VC projects in health care (Grigsby & Allen, 1997), few patients are seen through the VC for medical purposes. In almost every VC project, tele-consultation accounts for less than 25% of the use of the system (Perednia & Allen, 1995). The majority of the online time is used for medical education and administration (Hassol, 1996; Perednia & Allen, 1995; Wayman, 1994). Such result eliminates one of the most important functions of the telemedicine technology in health care delivery. The low level of usage can be explained in part by the federal government's position on reimbursement for telemedicine consultations (Hassol, 1996). However, Hassol (1996) pointed to other issues that need to be resolved first before the significance of VC could be realised. The important unresolved issues revolve around how successful VC can be in providing quality health care at an affordable cost and whether it is possible to develop a sustainable business model that would maintain profitability over time. This depends on (Perednia & Allen, 1995):

i. Clinical expectations: health care organisations need to be clear about the intended use of the telemedicine technology in their organisations; for example, will telemedicine be used for medical or administrative purposes; will it be used for general medical purposes or in certain medical specialities.
ii. Matching technology to medical needs: upon satisfying the preceding condition, the telemedicine technology needs to be integrated in the health care area in the sense that telemedicine is being seen as an integral medical tool, not an optional one. A complete protocol needs to be devised that could oversee the effective integration and use of the telemedicine technology in the health care area.
iii. Economic factors like reimbursement are major issues in countries such as the U.S. with vast geographical areas and different interstate laws and regulations. Accepting telemedicine encounters by health insurance companies could assist in increasing telemedicine success and diffusion.

iv. Legal concerns could be one of the largest impediments to the large-scale success of telemedicine in the long term. These concerns range from restrictions of medical practices across state lines (licensure), which are possible to resolve, to issues of liabilities due to malpractice or misdiagnosis. Recent research (Oakley et al., 1997) concluded that VC could be used with a reasonable degree of accuracy, and hence could not be relied on for initial diagnosis or treatment. This finding may cancel the use of telemedicine as a primary medical tool, but that does not cancel its importance as a viable follow-up medical tool and as an effective medical tool in a rural setting.
v. Social implications such as changing physician behaviours and traditional practices and workflow could impede telemedicine adoption (Anderson, 1997).
vi. Organisational factors: as indicated earlier, those are organisational specific, and hence vary from one health care provider to another. Issues such as the decision-making process concerning telemedicine adoption, whether telemedicine is considered part of an organisational strategy or direction, competing departments on resources and adopting/non-adopting telemedicine, motivation and empowerment are some of the factors that could influence telemedicine success in the long term in health care organisations.

CONCLUSION

This research provided different insights pertaining to the different factors that influence telemedicine use and success in health care organisations. Telemedicine offers health care providers many opportunities to provide timely and quality care to rural patients and this process could save them valuable time that could be used to see more patients in the main speciality hospital. There are other clinical and administrative advantages of telemedicine and health care organisations need to identify these opportunities and then devise means by which they could benefit from the different resources offered by telemedicine.

REFERENCES

Anderson, J. (1997). Clearing the way for physicians: Use of clinical information systems. *Communication of the ACM, 40*(8), 83-90

Anonymous. (2002). Telemedicine flying high. *Professional Engineering, 15*(9), 47.

Charles, B. (2000). Telemedicine can lower costs and improve access. *Healthcare Financial Management Association, 54*(4), 66-69.

Edelstein, S. (1999). Careful telemedicine planning limits costly liability exposure. Healthcare Financial Management, *53*(12), 63-69.

Harris, K., Donaldson, J., & Campbell, J. (2001). *Introducing computer-based telemedicine in three rural Missouri countries*. Journal of End User Computing, 13(4), 26-35.

Hassol, A. (1996). Surprise from the rural telemedicine survey. *Telemedicine Today,* 5-41.

Health Waikato. Retrieved October 11, 2001, from http://www.waikatodhb.govt.nz/HW/

Huston, T., & Huston, J. (2000). *Is telemedicine a practical reality?* Association for Computing Machinery. *Communications of the ACM, 43*(6), 91-95.

Neame, R. (1995). *Issues in developing and implementing a health information system*. Ministry of Health: Wellington.

Noring, S. (2000). Telemedicine and telehealth: Principles, policies, performance, and pitfalls. American Journal of Public Health, 90(8), 1322.

Oakley, A., Astwood, D., Loane, M., Duffill, M., Rademaker, M., & Wootton, R. (1997, February). Diagnosis accuracy of teledermatology: Results of a preliminary study in New Zealand. *New Zealand Medical Journal, 5*1-53.

Office of Technology Assessment U.S. Congress. (1995). *Bringing health care on line: The role of information technologies*. Washington, D.C.: US Government Printing Office.

Perednia, D., & Allen, A. (1995, February 8). TMVC technology and clinical applications. *The Journal of the American Medical Association (JAMA), 273*(6), 483-488.

Wayman, G. (1994). The maturing of TMVC technology Part I. *Health Systems Review*, *27*(5), 57-62.

KEY TERMS

Licensure: Legal restrictions of medical practice across state lines and the difficulty of resolving issues of liabilities due to malpractice or misdiagnosis. This is a major issue in countries such as the U.S. with vast geographical areas and different interstate laws and regulations.

Reimbursement: Accepting telemedicine encounters by health insurance companies as one of the different medical outlets and accordingly reimbursing doctors and physicians for their encounters with, for example, rural patients, could assist in increasing telemedicine success and diffusion.

Telecommunication Technologies: All technologies (telephone, telex, fax, video/audio conferencing, e-mail, chat, etc.) and related peripherals (digital camera, scanner, printer, etc.) that communicate with each other via telecommunications channels or networks (e.g., either wired or wireless: local public telephone lines or more sophisticated channels and networks such as ISDN, Frame Relay, DSL, microwave, satellite and LAN & WAN technologies, etc., which could provide up to several gigabits of data transfer per second (Gbps).

Telemedicine: Medicine from a distance where distant and dispersed patients are brought closer to their medical providers through the means of telecommunication technologies.

Tertiary Education and the Internet

Paul Darbyshire
Victoria University, Australia

Stephen Burgess
Victoria University, Australia

INTRODUCTION

For many years, information technology (IT) has been used to find ways to add value for customers to entice them to purchase the products and services of a business. Many educators use the Internet to supplement existing modes of delivery. Importantly, the Internet is providing a number of added value supplemental benefits for subjects and courses delivered using this new, hybrid teaching mode. There are two aspects to subject delivery to where added value benefits may be applied, and that is in the *administrative tasks* associated with a subject and the *educational tasks*. In both instances, IT solutions can be employed to either fully or partially process some of these tasks. Given the complex and often fluid nature of the education process, it is rare that a fully integrated solution can be found to adequately service both aspects of subject delivery. Most solutions are partial in that key components are targeted by IT solutions to assist the subject coordinator in the process. If we examine closely the underlying benefits gained in the application of IT to these tasks, there is a strong parallel to the benefits to be gained by business organizations with similar applications of IT. While the actual benefits actually sought by academics depend on the motivation for the IT solution, the perceived benefits can be classified using standard categories used to gauge similar commercial applications. This article examines the possibility of translating the benefits of added value to the use of the Internet by tertiary educators for subject and course delivery. A brief discussion will occur on aspects of course and subject delivery in tertiary education and the use of information technology for added value. These concepts are drawn together to indicate how the Internet may be used for added value in tertiary education. Finally, these concepts were tested with a survey of members of the IS World list serve.

BACKGROUND

Aspects of Course and Subject Delivery in Tertiary Education

For the purposes of this article, when the authors refer to tertiary education they mean university level education. There are two overall aspects to course and subject delivery, the educational and administrative components (Darbyshire & Wenn, 2000). Delivery of the educational component of a subject to students is the primary responsibility of the subject coordinator, and this task is the most visible from a student's perspective. However, the administration tasks associated with a subject form a major component of subject coordination, but these responsibilities are not immediately obvious or visible to the students.

It is essential that all aspects of subject delivery be carried out as efficiently as possible. To this end, IT, and in particular, Web-based solutions can be applied to both aspects of subject delivery. That Web-based solutions are a suitable vehicle to use has been almost universally accepted by students, teachers and academic administrators (Scott Tillett, 2000). Other advantages are the ease with which information can be disseminated, its interactivity, its use as a real-time communication medium and the ability to use text, graphics, audio and video (Kaynama & Keesling, 2000).

There are a number of administrative tasks associated with subject coordination for which IT solutions can be applied in the application. These include (Byrnes & Lo, 1996; Darbyshire & Wenn, 2000):

- *Student enrollment*. While most universities have a student enrolment system administered at the institute level, there are often local tasks associated with enrolment such as user account creation and compilation of mail lists, and so forth. Some of these tasks can be automated (Darbyshire & Wenn, 2000).

- *Assignment distribution, collection and grading*. The written assignment remains the basic unit of assessment for the vast majority of educators, and there have been many initiatives to computerize aspects of this task. Some of these include *Submit* (Hassan, 1991), *NetFace* (Thompson, 1988), *ClassNet* (Boysen & Van Gorp, 1997) and *TRIX* (Byrnes & Lo, 1996).
- *Grades distribution and reporting*. Techniques for this range from email, to password protected Web-based database lookup.
- *Informing all students of important notices*. Notice boards and sophisticated managed discussion facilities can be found in many systems. Examples include products such as *TopClass*, *Learning Space*, *Virtual-U*, *WebCT*, and *First Class* (Landon, 1998).

Many of the tasks viewed as educational can also employ IT solutions in order to gain perceived benefits. Some of these include: *online class discussions*; *learning*; *course outline distribution*; *seminar notes distribution*; *answering student queries*. Just how many of these are actually implemented will relate to a number of factors, such as the amount of face-to-face contact between lecturers and students. However, using the Internet for many of these can address the traditional problems of students misplacing handouts, and staff running out of available copies.

Discussion management systems are being integrated into many Web-based solutions. These are usually implemented as threaded discussions, which are easily implemented as a series of Web pages. Other tools can include chat rooms or listserv facilities. Answering student queries can take place in two forums, either as part of a class discussion or privately. Private discussions online are usually best handled via an email facility, or in some instances, store and forward messaging systems may replace email.

Implementing IT solutions to aid in the actual learning process is difficult. These can range from intelligent tutoring systems (Cheikes, 1995; Ritter & Koedinger, 1995), to facilitated online learning (Bedore, Bedore, & Bedore, 1998). However the major use of IT solutions in the learning process is usually a simple and straight forward use of the Web to present hypertext-based structured material as a supplement to traditional learning.

Information Technology: Efficiency and Added Value

There are a number of reasons for using IT in organisations today (O'Brien, 1999):

- *For the support of business operations*. This is usually to make the business operation more efficient (by making it faster, cheaper and more accurate).
- *For the support of managerial decision making*, by allowing more sophisticated cost benefit analyses, providing decision support tools and so forth.
- *For the support of strategic advantage*. This refers to the use of Porter and Millar's (1985) three generic strategies as a means of using information technology to improve competitiveness by adding value to products and services.

It has been recognised for a number of decades that the use of computers can provide cost savings and improvements in efficiencies in many organisations. Porter & Millar (1985) have generally been credited with recognising that the capabilities of information technology can extend further to providing organisations with the opportunity to add value to their goods. Value is measured by the amount that buyers are willing to pay for a product or service. Porter & Millar (1985) identify three ways that organisations can add value to their commodities or services (known as *generic strategies for improving competitiveness*):

- Be the lowest cost producer.
- Produce a unique or differentiated good (providing value in a product or service that a competitor cannot provide or match, at least for a period of time). If an organisation is the first to introduce a particular feature, it may gain a competitive advantage over its rivals for a period. Some ways in which information technology can be used to differentiate between products and/or services are (Sandy & Burgess, 1999):
  - Quality
  - Product Support
  - Time
- Provide a good that meets the requirements of a specialised market.

The next sections examine the possibility of translating the benefits of added value to a particular application of IT, the use of the Internet by tertiary educators, to assist with subject and course delivery.

Using Internet Technologies to Improve Efficiency and Add Value

With the recent explosion in Internet usage, educators have been turning to the Internet in attempts to gain benefits by the introduction of IT into the educational process. In this article, subject delivery at the university level is only considered. The benefits sought from such activity depend on the driving motivation of the IT solution being implemented. While many may not perceive a university as a business (and it is not advocated here), it is nonetheless possible to match the current uses of the Internet in tertiary education with traditional theory related to the reasons why firms use IT.

Internet technologies in education, which are used for the learning process itself, target the student as the main stakeholder. While the motivation may be the enhancement of the learning process to achieve a higher quality outcome, we can loosely map this to the "*support of managerial decision making*" concept identified earlier. Such technologies allow educators to obtain a far more sophisticated analysis of individual student's learning progress, and thus provide them with decision support tools on courses of action to take to influence this process.

Technology solutions, which target the academic as the stakeholder (Darbyshire & Wenn, 2000, Central Point), implement improvements or efficiencies that can be mapped to the "*support of the business operation*" previously identified. Improvements or efficiencies gained from such implementations are usually in the form of automated record keeping and faster processing time, ultimately resulting in lower costs in terms of academic time, and added value to the students.

By default, the university also becomes a stakeholder in the implementation of either of the above types of technology enhancements. Benefits gained by students and staff by such uses of technology translates ultimately to lower costs for the institution or the provision of more and/or better quality information. The benefits of such systems can be mapped onto the "*support of strategic advantage*" concept (as Porter's low cost and differentiation strategies), previously identified as a reason for using technology in business. If these institutions are to regard themselves as a business, then the successful use of IT in subject delivery could give the university a strategic advantage over other universities, which it would regard as its business competitors. Most of the reported advantages gained from online supplementation of teaching relate to cost savings in terms of efficiency, flexibility and/or convenience. These represent the traditional added value benefits of lower cost and faster access to goods in the commercial world. Thus, we can use the measures of *money savings*, *time savings*, *improved quality* and better *product information* as categories to measure the benefits gained from the introduction of IT to supplement teaching.

A SURVEY OF IS WORLD MEMBERS

The authors were interested to investigate the extent of appreciation of the added value benefits that the Internet can offer to tertiary educators, institutions and their students. In the first instance, a simple survey was conducted through the IS World discussion list to gain an initial idea of the level of appreciation that existed.

IS World is a Web-based resource which has been set up for the benefit of information systems academics and researchers around the world. A general email was posted to the IS World discussion list on January 29, 2001. A request was made for tertiary educators to respond, outlining their uses of the Internet in tertiary education and how the uses added value for the institution and for students. There were 43 responses to the survey between January 29, and February 4, 2001. Most of these (33) were within one day of the initial email.

All respondents to the survey identified as least one type of Internet usage to assist them. Approximately seven out of 10 adopted administrative uses and roughly the same proportion adopted educational uses. This supports the notion identified in the literature that the technology would be accepted in the tertiary education field. The following findings support the notion that educators identify the added value uses of the Internet in tertiary education.

The most common benefit for administrative uses was to save time for the institution and for students. Most administrative benefits were similar for both groups, except for "save money" (where more than twice the respondents felt that the institution saved money than students). The "information provision" administrative usages were the most commonly used (important notices, schedules/ timetables, assignment and grade distribution). Less common were the more "interactive" options, assignment collection and student enrolment. Educational uses of the Internet were seen as providing slightly more benefits for students than institutions. Their use were seen as providing more information and improving quality more on average than the administrative uses. As with administrative usages, the easiest educational features to set up were the most commonly used (distribute course/ subject notes, provide external links). Less common were the more "interactive" options, discussion lists and online chat groups. About three quarters of respondents used the Internet to answer student queries (probably by email). As with administrative uses, most of the benefits are similar for students and the institution, with (again) some differences for instances where the benefits save money more for the institution than students. More respondents saw the differences in the benefits of educational uses flowing to students than to institutions than with administrative uses. In three of the uses, saving time was not the most common benefit identified. These were the provision of external links to additional resources, discussion lists and online chats, where improved quality of information and more information were more commonly identified.

FUTURE TRENDS

We believe that research into the use of the Internet for educational purposes will continue to expand, as the "added value" opportunities to provide a better service to students begin to mature. At the same time, we believe that the administrative uses will continue to expand, without perhaps the recognition or body of research that is devoted to learning outcomes. Although this is understandable, the administrative improvements that the Internet can provide should not be ignored completely.

CONCLUSION

The majority of tertiary educators use the Internet to supplement existing modes of delivery. Importantly, the Internet is providing a number of added value supplemental benefits for subjects and courses delivered. There are two aspects to subject delivery to where added value benefits may be applied, and that is in the *administrative tasks* associated with a subject and the *educational tasks*. Most of the reported advantages gained from online supplementation of teaching relate to cost savings in terms of efficiency, flexibility and/or convenience. These represent the traditional added value benefits of lower cost and faster access to goods in the commercial world. The measures of *money savings*, *time savings*, *improved quality* and better *product information* can be used as categories to measure the benefits gained from the introduction of IT to supplement teaching.

A survey of 43 tertiary educators, conducted through the IS World discussion list, revealed similar usage levels of administrative and educational features to aid tertiary education on the Internet. The administrative uses showed slightly more benefits for the institution than for students and vice-versa for educational uses. In both types of uses, their adoption seemed to be based upon how difficult the feature was to set up as well as the added value benefits it provided.

REFERENCES

Bedore, G.L., Bedore, M.R., & Bedore, Jr., G.L. (1998). *Online education: The future is now*. Socrates Distance Learning Technologies Group, Academic Research and Technologies.

Boysen, P., & Van Gorp, M. J. (1997). ClassNet : Automated support of Web classes. Paper presented at the *25th ACM SIGUCCS Conference for University and College Computing Services*, Monterey, California USA.

Byrnes, R. & Lo, B. (1996). A computer-aided assignment management system: Improving the teaching-learning feedback cycle. Retrieved on December 2, 1999 from *http://www.opennet.net.au/cmluga/byrnesw2.htm*

Cheikes, B.A. (1995). GIA: An agent-based architecture for intelligent tutoring systems. Paper presented at the *Proceedings of the CIKM'95 Workshop on Intelligent Information Agents*.

Darbyshire, P. (1999). Distributed Web based assignment submission and access. *Proceedings- International Resource Management Association*, IRMA '99, Hershey, USA.

Darbyshire, P. & Wenn, A. (2000). A matter of necessity: Implementing Web-based subject administration. Chapter in *Managing Web enabled technologies in organizations*. Hershey, PA: Idea Group Publishing.

Hassan, H. (1991). The paperless classroom. Paper presented at *ASCILITE '91*. University of Tasmania, Launceston, Australia.

Kaynama, S.A. & Keesling, G. (2000, August). Development of a Web-based Internet marketing course. *Journal of Marketing Education, 22*(2), 84-89.

Landon, B. (1998, 10/4/98). On-line educational delivery applications: A Web tool for comparative analysis [Web Page]. Centre for Curriculum, Transfer and Technology, Canada. Retrieved on October 10, 1998 from *http://www.ctt.bc.ca/landonline/*

O'Brien, J.A.. (1999). *Management information systems, managing information technology in the Internet worked enterprise* (4th ed.). Irwin MaGraw Hill.

Porter, M.E. & Millar, V E., (1985, July-August). How information gives you competitive advantage, *Harvard Business Review, 63*(4), 149-160.

Ritter, S., & Koedinger, K.R. (1995). Towards lightweight tutoring agents. Paper presented at the *AI-ED 95—World Conference on Artificial Intelligence in Education*, Washington, D.C.

Sandy, G. & Burgess, S. (1999). Adding value to consumer goods via marketing channels through the use of the Internet. *CollECTeR'99: 3rd Annual CollECTeR Conference on Electronic Commerce*, Wellington, New Zealand, November.

Scott Tillett, L. (2000). Educators begin to reach out – The net cuts costs, simplifies management and could make distance learning a winner. *InternetWeek*, Manhasset, Iss.835, October 30, pp.49-56.

Thompson, D. (1988, 14/3/98). WebFace overview and history [Web page]. Monash University. Retrieved on February 1, 1999 from *http://mugca.cc.monash.edu.au/~webface/history.html*

KEY TERMS

Administrative Tasks: The tasks that support educational tasks (such as enrolment, recording results, and so forth).

Educational Tasks: Those tasks directly associated with the delivery of the educational component to students (e.g., lecturers, tutorials, assessment, and so forth).

Efficiency: From an IT viewpoint, this usually relates to improvements within the business, so for a business it may mean IT systems that reduce costs or perform tasks more reliably or faster.

Internet Technologies: That group of technologies that allow users to access information and communication over the World Wide Web (Web browsers, ftp, email, associated hardware, internet service providers and so forth).

Value: The amount a "buyer" is willing to "pay" for a product or service. A business can add value by being a low cost providing, providing a unique or differentiated product or service or filling a niche market.

Text Mining in the Context of Business Intelligence

Hércules Antonio do Prado
Brazilian Enterprise for Agricultural Research and Catholic University of Brasília, Brazil

José Palazzo Moreira de Oliveira
Federal University of Rio Grande do Sul, Brazil

Edilson Ferneda
Catholic University of Brasília, Brazil

Leandro Krug Wives
Centro Universitário Feevale and Federal University of Rio Grande do Sul, Brazil

Edilberto Magalhães Silva
Brazilian Public News Agency, Brazil

Stanley Loh
Catholic University of Pelotas and Lutheran University of Brazil, Brazil

INTRODUCTION

Information about the external environment and organizational processes are among the most worthwhile input for business intelligence (BI). Nowadays, companies have plenty of information in structured or textual forms, either from external monitoring or from the corporative systems. In the last years, the structured part of this information stock has been massively explored by means of data-mining (DM) techniques (Wang, 2003), generating models that enable the analysts to gain insights on the solutions for organizational problems. On the text-mining (TM) side, the rhythm of new applications development did not go so fast. In an informal poll carried out in 2002 (Kdnuggets), just 4% of the knowledge-discovery-from-databases (KDD) practitioners were applying TM techniques. This fact is as intriguing as surprising if one considers that 80% of all information available in an organization comes in textual form (Tan, 1999).

In their popular model to explain the phases of technology adoption (Figure 1), Moore and McKenna (1999) discuss the existence of a chasm between the "early adopters, visionaries," and the "early majority pragmatists" phases that a technology has to cross in order to become extensively adopted. From our point of view, TM is crossing this chasm yet. Although there is the existence of mature tools in the market, and an increasing number of successful case studies have been presented (Ferneda, Prado, & Silva, 2003; Fliedl & Weber, 2002; Dini & Mazzini, 2002; Prado, Oliveira, Ferneda, Wives, Silva, & Loh, 2004), it seems that the community is still leaving the second phase. However, the results presented in the case studies point out that the broad adoption of TM will happen in the near future.

BACKGROUND

When studying the relations between TM and BI, it is necessary to take into account an important intermediate layer between them: the knowledge-management (KM) process. KM refers to the set of activities responsible for carrying the information along the organization and making knowledge available where it is necessary.

To clarify the relations between TM and BI, under the point of view of a KM model, we adopted the generic KM model (Figure 2) proposed by Stollenwerk (2001). The model is made up of seven processes: (a) identification and development of the critical abilities, (b) capture of knowledge, skills, and experiences to create and maintain skills, (c) selection and validation that filter, evaluate, and summarize the acquired knowledge for future use, (d) organization and storage to assure the quick and correct recovery of the stored knowledge, (e) sharing that makes easy the access to information and knowledge, (f) application in which the knowledge is applied in real situations, and (g) creation that comprises the activities of sharing tacit knowledge, creating concepts, building archetypes, and cross-leveling knowledge. Involving the mentioned processes, there exist the aspects of leadership, organiza-

Figure 1. Moore's and McKenna's (1999) life cycle of technology adoption

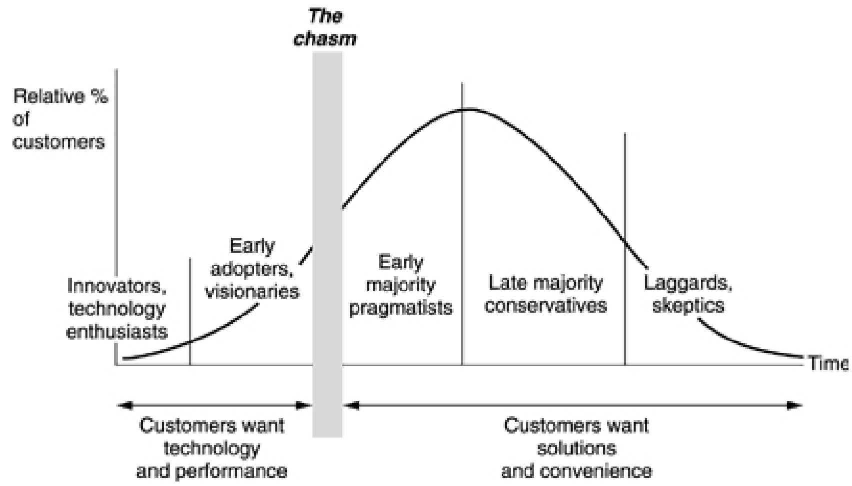


Figure 2. Generic KM model of Stollenwerk (2001)

tional culture, measuring and compensation, and technology. The main relation between TM and KM is located in the creation process. By applying the TM techniques discussed in the next section, it is possible to find patterns that, adequately interpreted, can leverage the concept-creation activity.

METHODS AND TECHNIQUES FOR TEXT MINING

Text mining can be defined as the application of computational methods and techniques over textual data in order to find relevant and intrinsic information and previously unknown knowledge.

Text-mining techniques can be organized into four categories: classification, association analysis, information extraction, and clustering. Classification techniques consist of the allocation of objects into predefined classes or categories. They are used to identify the class or category of texts in tasks such as *topic spotting* (the identification of a known topic or subject in a document) and *document routing* or *filtering* (the selection of relevant documents to a process or to someone).

Association analysis is used to identify correlation or dependencies among elements or attributes (words or concepts present in documents). It helps the identification of words or concepts that co-occur together and, consequently, to understand the contents of a document or set of documents and their relationships.

Information-extraction techniques are able to find relevant data or expressions inside documents. Typical uses of these techniques involve the creation of databases from texts or the identification of specific information (like names, dates, and e-mails) in a large set of documents.

Clustering is the process of finding relationships among texts or words and putting them together in groups of related documents. Clustering techniques are used to understand how the information or knowledge of an entire collection of documents is organized.

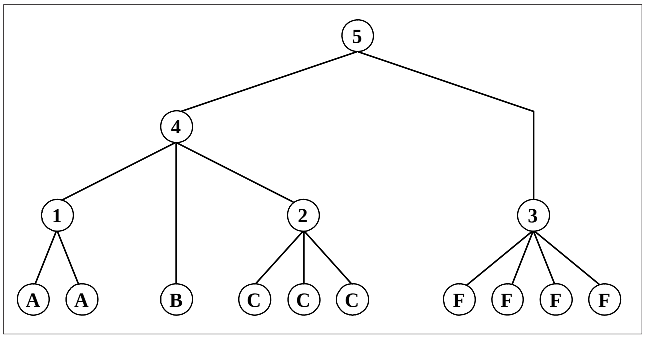
In this chapter we focus on clustering techniques because it is more appropriate to the (concept) creation phase since it helps the user analyze and understand previously unknown data, exploiting the relationship among documents and words present in a large collection of documents. The next section describes the clustering process with more detail.

CLUSTERING

Clustering is a knowledge-discovery process that identifies relationships among objects and builds clusters of objects based on these relationships (Jain, Murty, & Flynn, 1999; Willet, 1988). It is based on the cluster hypothesis (Rijsbergen, 1979), which states that similar objects tend to remain together in the same cluster as a consequence of a specific concept distance metric.

Clustering is a widely employed tool to analyze data in many fields (Everitt, Landau, & Leese, 2001; Jain et al., 1999). The idea behind cluster analysis is to find knowledge about previously unfamiliar data. Clustering methods are able to give suggestions about how specific sets of data are organized or correlated. It is possible to identify the similarity and the dissimilarity among many objects or data patterns and, based on that, construct classes or categories. Categories or classes are very important as they are the basic elements to build new concepts, and concepts are the basis of human knowledge (Aldenderfer & Blashfield, 1984).

Figure 3. Hierarchic scheme of clusters



However, since clustering indicates relationships among objects, it can be used for many other objectives. Aldenderfer and Blashfield (1984), for example, classify the goals of cluster analysis into four categories: (a) to develop a typology or classification, (b) to investigate useful conceptual schemes for grouping entities, (c) to aid in the generation of hypotheses, and (d) to test hypotheses, verifying if types defined through other procedures are really present in the data set.

Clustering Types

There are many clustering methods and this fact generates many types or schemes of clusters. According to Aldenderfer and Blashfield (1984), these methods can be classified into seven families: hierarchical agglomerative, hierarchical divisive, iterative partitioning, density search, factor analytic, clumping, and graph theoretic. Each of them creates a type or scheme of clusters that is very peculiar.

For the sake of generality, we choose to classify and detail the clustering methods according to the categories proposed by Everitt et al. (2001) and Schütze and Silverstein (1997). These categories are (a) *hierarchical* and (b) *nonhierarchical* (or *partitioning* clustering).

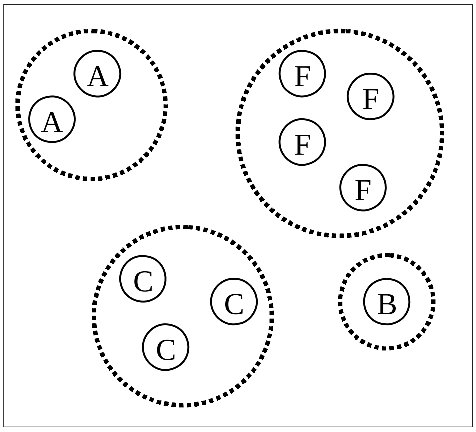
Hierarchical Clustering

In hierarchical clustering, the resulting scheme of clusters is very similar to a tree (see Figure 3). Each node represents a cluster. The intermediate clusters are clusters of clusters and the leaves are objects. The relationship among clusters is of paramount importance as it shows the specificities and abstractions among groups of objects. If the user goes up in the tree of clusters, it is possible to identify more abstract or generic groups. On the other hand, if the user goes down, more specific groups will be identified until the objects themselves are reached.

Nonhierarchical or Partitioning Clustering

When working with nonhierarchical clustering, the objects are allocated in isolated clusters and no relationship among clusters can be found. This type of clustering is also known as partitional clustering, and it is said that it generates flat (without structure) partitions of clusters (see Figure 4).

Figure 4. Flat partition of isolated clusters



Clustering Algorithms

As already stated, there are many clustering methods. More detail on these and other clustering methods can be obtained in Aldenderfer and Blashfield (1984), Jain et al. (1999), Kowalski (1997), and Willet (1988). In this section we will describe only the methods implemented in the tool used in our experiments: the Eurekha tool (Wives, 1999). The algorithms implemented come from the graph theoretic family of algorithms and are described next.

Stars

The stars algorithm analyzes the objects and tries to find groups of similar elements where the resulting aspect or format is like a star of correlated objects. In this case, the center of the star is the element that has a relation with all the other objects in the cluster, linking them together. It means that the other elements should be near or similar to this central element, but not necessarily to the others. To minimize the dissimilarity among an element that is on one side of the star with another element in another side of the star, it is defined a similarity threshold. A larger threshold among all elements in relation to the center makes the group more coherent. The more they are similar to the center (or near to the center), the more they will be similar to each other.

The algorithm starts selecting any element in the set of elements. This selection can be performed randomly or by any other method. However, the selection order influences the resulting clustering scheme.

The selected element is then elected as the center of the star (the center of the cluster). Then, this element is compared to all other elements not yet clustered (i.e., allocated to a cluster). If a relation is found, meaning that it is greater than a previously user-defined similarity threshold, the element being compared to the center is allocated to the cluster. Once all elements are compared to the star center, another unclustered element is selected and the process continues until all elements are analyzed. The elements in which the similarity to another element is not greater than the established threshold are said to be unclustered and are ignored or allocated to an isolated cluster—one to each element.

Best Star

The main problem of the star algorithm is that the order in which the elements are selected as centers influences the clustering result. Another problem is that the user has to select a threshold of minimum similarity between objects and the center, and there is not an optimal threshold to be used as a usual value. Each data set may have a different threshold. These are the greatest problems of cluster analysis that uses this kind (or family) of algorithms[1].

The best-star algorithm intends to solve these problems, allocating an element, even if it is already clustered, to the star where it is more similar (the nearest star). Somehow, a side effect is that the user does not need to establish a threshold. In this case, the elements will be reassigned to the cluster where they are more similar (i.e., nearer to the star's center).

Cliques

This algorithm is similar to the star algorithm. However, the elements are added only if they satisfy the threshold of similarity among all elements already in the cluster and not only with the central element. In this case, the elements are more tightly coupled and the quality of the resulting clusters is better.

Full Stars

Sometimes the user must know all the clusters where an element would be allocated. All the other algorithms discussed in this chapter allocate the element in the best cluster for it, according to its algorithmic restrictions. This algorithm solves this necessity, allocating an element in all clusters it has a relationship with greater than the threshold established by the user.

FUTURE TRENDS

With regard to the trends in the use of TM in BI, we can see the idea of *concept* replacing the usual approach of TM based on the words of a text. Words can lead to semantic mistakes, known as the vocabulary problem, for example, when people use synonyms or word variations.

In the conceptual approach for clustering, concepts represent the content of a textual document in a higher level, minimizing the vocabulary problem. Concepts talk about real-world events and objects, and are used by people to express ideas, ideologies, thoughts, opinions, and intentions through the language (in talks, texts, documents, books, messages, etc.).

In previous works (Loh, Oliveira, & Gastal, 2001; Loh, Wives, & Oliveira, 2000), concepts were used with success in mining processes of textual documents. Thus, using concepts as document attributes in the clustering process contributes to generate better results than using words since the resulting clusters have elements with more cohesion, besides being more understandable.

CONCLUSION

The advent of the knowledge society has imposed an important change in the context of organizations. Business competitiveness is significantly affected by the availability of knowledge about the organizational processes and the external environment. The importance of the information existing in organizations as raw material to create knowledge has been recognized since the late ’80s. As a matter of fact, the use of such knowledge for leveraging the business has led to an increasing number of KDD applications. However, the majority of these applications has been addressed to process structured data rather than the unstructured that is, by far, the biggest part of the organizational information.

The existence of mature tools to develop TM applications and the amount of textual information available in the organizations seem to be a strategic opportunity that cannot be ignored. In this chapter it was discussed the role of TM in BI, clarifying the interface between them.

REFERENCES

Aldenderfer, M. S., & Blashfield, R. K. (1984). *Cluster analysis*. Beverly Hills, CA: Sage.

Dini, L., & Mazzini, G. (2002). *Opinion classification through information extraction*. Proceedings of the Third International Conference on Data Mining (Data Mining III).

Everitt, B. S., Landau, S., & Leese, M. (2001). *Cluster analysis* (4th ed.). New York: Oxford University Press Inc.

Ferneda, E., Prado, H. A., & Silva, E. M. (2003). *Text mining for organizational intelligence*. Proceedings of the Fifth International Conference on Enterprise Information Systems, Angers, France.

Fliedl, L., & Weber, G. (2002). *NIBA-TAG: A tool for analyzing and preparing German texts*. Proceedings of the Third International Conference on Data Mining (Data Mining III).

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys*, *31*(3), 264-323.

Kdnuggets. (2002). *Poll: Which data mining techniques do you use regularly?* Retrieved June 25, 2002, from *http://www.kdnuggets.com/polls/2002/data_mining_techniques.htm*

Kowalski, G. (1997). *Information retrieval systems: Theory and implementation*. Boston: Kluwer Academic Publishers.

Loh, S., Oliveira, J. P. M. d., & Gastal, F. L. (2001). Knowledge discovery in textual documentation: Qualitative and quantitative analysis. *Journal of Documentation, 57*(5), 577-590.

Loh, S., Wives, L. K., & Oliveira, J. P. M. (2000). Concept-based knowledge discovery in texts extracted from the Web. *ACM SIGKDD Explorations, 2*(1), 29-39.

Moore, G., & McKenna, R. (1999). *Crossing the chasm: Marketing and selling high-tech products to mainstream customers*. HarperBusiness.

Prado, H. A. do, Oliveira, J. P. M., Ferneda, E., Wives, L. K., Silva, E. M., & Loh, S. (2004). Transforming textual patterns into knowledge. In M. Raisinghani (Ed.), *Business intelligence in the digital economy: Opportunities, limitations and risks*. Hershey, PA: Idea Group Publishing.

Rijsbergen, C. V. (1979). *Information retrieval* (2nd ed.). London: Butterworths.

Schütze, H., & Silverstein, C. (1997). *Projections for efficient document clustering*. Proceedings of the Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, Philadelphia, PA.

Stollenwerk, M. F. L. (2001). Knowledge management: Concepts and models. In K. Tarapanoff (Ed.), *Organizational and competitive intelligence.* Brazil.

Tan, A.-H. (1999). *Text mining: The state of the art and the challenges*. Proceedings of the Workshop on Knowledge Discovery from Advanced Databases (PAKDD’99), Beijing, China.

Wang, J. (Ed.). (2003). *Data mining: Exposed and examined*. Hershey, PA: Idea Group Publishing.

Willet, P. (1988). Recent trends in hierarchic document clustering: A critical review. *Information Processing & Management, 24*, 577-597.

Wives, L. K. (1999). *A study about arrangement of textual documents applied to unstructured information processing using clustering techniques*. Master's thesis, PPGC/Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil.

Wives, L. K. (2000). *Technologies for knowledge discovery in texts applied to competitive intelligence*. Qualification exam, PPGC/Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil.

KEY TERMS

Association Analysis: Use of statistics criteria to measure the proximity of two distinct objects or texts using some of their properties or attributes. A method for identifying corelation or dependencies among elements or attributes, using statistical techniques.

Classification: A systematic arrangement of objects (texts) on groups according to (pre) established criteria, or the process of allocating elements in predefined classes. The classification needs a predefined taxonomy in contrast with the clustering technique that works without previous knowledge. Sometimes it is also associated with the process of identifying classes, that is, discovering attributes that characterize one class and that distinguish this from others.

Cluster analysis: The process that includes the clustering method and the analysis of its results in order to discover and understand the contents of a set of elements, texts, or objects, and the relations among them.

Cluster: A group of elements that have some characteristics or attributes in common

Concept: It is an abstract or generic idea, opinion, or thought generalized from particular instances by the selection of meaningful terms. The concept may be identified by the use of text-mining techniques. They are used to explore and examine the contents of talks, texts, documents, books, messages, and so forth. Concepts belong to the extralinguistic knowledge about the world, representing real things in formal ways.

Knowledge Discovery: A computer-supported process that uses computational algorithms and tools as visualization methods to help a user to discover knowledge from stored data.

Knowledge Discovery from Texts: A computer-supported process that uses computational algorithms and tools as visualization methods to help a user to discover knowledge from stored textual data. It can be understood that this process gives to the user information that would not be recovered by traditional queries since the information is not explicitly stated or declared in the textual data.

Text Mining: A computer-supported process that uses computational algorithms and tools over textual data with the objective of discovering statistical patterns. Most common methods include *clustering*, *classification*, and *association analysis*. Most of the time, the expression is interchangeable with knowledge discovery from texts, however, the last is a larger process where the first one is involved.

Text Mining by Concepts: The application of text-mining methods over textual documents that are represented or modeled by concepts instead of words

ENDNOTE

[1] Another problem, already stated, is related to the algorithms that do not work if the user does not indicate the number of clusters that must be found (*k-means* and *buckshot*). This indication or selection is not necessary in the algorithm described and in all the algorithms of the graph theoretic family.

T

The Changing Library Education Curriculum

Vicki L. Gregory
University of South Florida, USA

INTRODUCTION

Libraries of the 21st century are very different places from those that existed at the beginning of the 20th century, and very different as well from the libraries of only 25 years ago. Library education has striven to keep pace with all the myriads of changes. Within the last 100 years, fortunately and necessarily in order to retain its relevance, professional library education and practice has evolved from the centrality of teaching and writing the "library hand" to providing modern curricula such as services for distance learners and Web-based instruction using course management systems such as Blackboard, WebCT, and so forth. Along the way, the library profession has often been first not only to accept but also to adopt and apply the technological innovations now common to modern civilization. Throughout, library educators have paved the way to the acceptance of innovation in libraries by instructing students to use and apply new technologies.

BACKGROUND

The revolutionary changes over the past 25 years in the educational curriculum for schools of library and information science, which are necessitated by the exponential expansion of computer-based technologies, requires an almost constant and continuous reexamination of the skills and expertise that need to be acquired by the next generation of librarians. Although much has changed in libraries, the core of who we are and what we are truly remains the same. Librarianship is and will continue to be a profession devoted to bringing users and information together as effectively and efficiently as possible. To meet that ideal, librarians have used technology to enhance and create services. In addition, it is important to meet emerging educational needs of our increasingly multicultural and diverse society. Librarians have recognized that changing expectations and lean budgets require organizations to call upon the talents of everyone (Butcher, 1999). And librarians have become more engaged in teaching and research in order to serve the needs of users better (Bahr & Zemon, 2000).

Importance of People and People Skills

Computer technologies and communication systems have had an undeniable impact on society as a whole and on our profession, but it is also critical to remember the importance of the individual and of the need for interpersonal skills in our profession, which at its heart remains basically a "people profession." We harness technology for a reason—to promote learning and the dissemination of information—and we do not simply revere technology for its own sake. With the aid of computer specialists, we could design the best information system imaginable, but unless it operates in a manner that is accessible to people, nobody will use it. The ability of librarians (whether through collecting, organizing, or retrieving information) to act as intermediaries between users and the world's information resources will, in my opinion, never become outdated (Gilbert, 1998). In sum, the rapid changes in all types of libraries and the burgeoning of new technologies for librarians to learn, while increasing the amount of information that students need to have under their "academic belts" if they are to enter successfully into a library career, nevertheless remain rooted in the need to carry out the traditional librarian roles—though hopefully faster, cheaper, smarter, and more effectively.

Preparing Students in Traditional Areas of Library Responsibilities

The traditional heart and sole of a library is and remains, of course, its collections—from the time of the great Alexandrine library of the Classical era, libraries have been, in essence, civilization's repositories of learning, and hence the materials through which learning is transmitted down the generations. Current students preparing for the future (and indeed the present) electronic library cannot be permitted to overlook the continued, lasting importance of print publications in the library's carrying out of its role, but by necessity, they must be equipped to deal with the rapidly expanding world of digital medium. Thus, collection development courses must reflect an appropriately balanced approach, emphasizing the latest technology not as an end in itself, but rather as simply another tool to use in addressing the problems arising in acquiring adequate resources for a library collection in whatever format is most appropriate for the particular library and the "task at hand" (Thornton, 2000).

As librarians and information professionals go about the process of acquiring electronic information resources in carrying out their collection development role, they must also continue to recognize and care about the important questions that have always concerned libraries, in respect to questions of future accessibility and preservation of library resources. Electronic materials—with their typical provision to libraries only through a licensing regime rather than through outright purchase—present altogether different problems for the library than do print materials. Collection development and preservation must remain an important part of the library school curriculum, no matter how dominated the library may become with electronic materials (Kenney, 2002).

In most conceptions of the libraries of the future, reference librarians may expect to continue to play many of the same reference roles that they have traditionally performed in interacting with their libraries' users. Reference librarians will continue to serve in an intermediary role to assist users in finding needed information and providing important "value-added" services through the production of instructional materials and guides to information resources. However, many of these functions, out of necessity, will be performed in media other than those that have been traditionally utilized. Collaboration and instruction may be expected to take place in a Web-based "chat" environment or by e-mail, rather than through a face-to-face meeting over the reference desk (Abels, 1996; Domas, 2001).

Reference librarians of the future must therefore acquire teaching skills as well as informational skills. They will need to be able to teach information literacy skills as students discover that just finding some online information on a topic and pushing the "print" or "download" button is not enough. In the electronic information world, librarians must be prepared to evaluate resources in a somewhat more in-depth way than was necessary when they could often depend upon refereed print journals for the majority of their information (Grassian & Kaplowitz, 2001).

In addition to all the vagaries involved with the classification and cataloging of traditional print materials, technical services librarians today, and doubtless more so in the future, will have to be prepared to cope with all the exponential varieties and forms that electronic resources may take. Technical services professionals are increasingly dealing with so many different formats and kinds of materials that may defy classification and are often not traditionally cataloged; other approaches, such as indexing and abstracting techniques and the development of in-house library-constructed databases, as well as Webliographies, may be undertaken as methods of organizing the access and retrieval process. Future graduates planning a career in the technical services areas should place a much greater focus than is presently typically allowed for in most library school curriculums on the technological aspects of information provision. Concurrently, library and information science schools need to take steps to provide for the programs and/or the courses that will include building student skills in document creation for the digital library environment. Unfortunately, all this cannot be allowed to serve as a replacement for the traditional knowledge and skills involved in cataloging and classification. As a minimum, students will need to gain a hands-on knowledge of the architecture of the infrastructure and databases behind a digital library. This means that LIS schools must develop additional specific courses, rather than trying to make room in the already overstuffed basic "organization of knowledge" classes that most schools currently offer (Vellucci, 1997).

FUTURE TRENDS

In the foreseeable future, it is probable that more and more instruction will be provided in a distance mode utilizing Web delivery, videoconferencing, and other technological means of providing instruction. A burden on many LIS faculty members at present is how to adapt a course, originally designed for a face-to-face classroom encounter to a Web-based encounter. Although the goals, objectives, and major assignments for a class might remain the same, the overall means of delivery puts more pressure on faculty members to devise new ways of delivering material (Gregory, 2003). Both virtual and print reserve materials may become problematic as distance from the home site increases. Compounding the traditional instructional component, there is the additional element of computer support on a 24-hour, 7-days-a-week basis (Young, 2002). Increasingly, when something goes wrong with the computer on a student's end, the faculty member is expected to be able to do computer troubleshooting over the telephone or by e-mail. It is common for programs and universities to provide technical support, but even so the faculty member usually gets caught up in the technical support problems, obviously much more so than when the class is taught in the traditional manner (Carey & Gregory, 2002; Newton, 2003, p. 418). Of course, when the academic computing staff person or the faculty member is unavailable, the next major organization on the campus that fields these questions is—you guessed it—the library. Librarians must be able to deal with technical, computing, or network issues and attempt to aid the beleaguered student (or faculty member). So although these issues primarily affect the teaching of library and information studies classes, they also have a major impact on the services demanded of the library (Barron, 2003).

CONCLUSION

The rapidly changing requirements in the educational curriculum of schools of library and information science resulting from the exponential expansion of computer-based technologies naturally result in a reexamination of the knowledge and skills that need to be acquired by the next wave of library and information professionals. Skills in the use of new technologies are not only important in professional work, but in the education process itself, as more and more LIS courses are being offered via the Web, with faculty and students utilizing course management software.

REFERENCES

Abels, E.G. (1996). The e-mail reference interview. *RQ*, 35(Spring), 345-358.

Bahr, A.H. & Zemon, M. (2000). Collaborative authorship in the journal literature: Perspectives for academic librarians who wish to publish. *College and Research Libraries, 61*(5), 410-19.

Barron, D.D. (Ed.). (2003). *Benchmarks in distance education: The LIS experience*. Westport, CT: Libraries Unlimited.

Butcher, K. (1999). Reflections on academic librarianship. *Journal of Academic Librarianship, 25*(5), 350-353.

Carey, J.O. & Gregory, V.L. (2002). Students' perceptions of academic motivation, interactive participation, and selected pedagogical and structural factors in Web-based distance education. *Journal of Education for Library and Information Science, 43*(1), 6-15.

Domas White, M. (2001). Digital reference services: Framework for analysis and evaluation. *Library & Information Science Research, 23*(3), 211-231.

Gilbert, B. (1998). The more we change, the more we stay the same: Some common errors concerning libraries, computers, and the Information Age. In M.T. Wolf, P. Ensor & M.A. Thomas (Eds.), *Information imagineering: Meeting at the interface* (pp. 219-227). Chicago, ALA.

Grassian, E.S. & Kaplowitz, J.R. (2001). *Information literacy instruction: Theory and practice*. New York: Neal-Schuman.

Gregory, V.L. (2003). Student perceptions of the effectiveness of Web-based distance education. *New Library World, 104*(10), 426-433.

Kenney, A.R. et al. (2002, January). Preservation risk management for Web resources. *D-Lib Magazine,* 8(1). Retrieved December 12, 2003, from *www.dlib.org/dlib/january02/kenney/01kenney.html*

Newton, R. (2003). Staff attitudes to the development and delivery of e-learning. *New Library World, 104*(1193), 312-425.

Thornton, G.A. (2000). Impact of electronic resources on collection development, the roles of librarians, and library consortia. *Library Trends, 48*(4), 842-856.

Vellucci, S.L. (1997). Cataloging across the curriculum: A syndetic structure for teaching cataloguing. *Cataloging and Classification Quarterly, 24*(1/2), 35-39.

Young, J.R. (2002). The 24-hour professor: Online teaching redefines faculty members' schedules, duties, and relationships with students. *The Chronicle of Higher Education, 38*(May 31). Retrieved from *chronicle.com/weekly/v48/i38/38a03101.htm*

KEY TERMS

Collection Development: The portion of collection management activities that has primarily to do with selection decisions.

Collection Management: All the activities involved in information gathering, communication, coordination, policy formulation, evaluation, and planning that result in decisions about the acquisition, retention, and provision of access to information sources in support of the needs of a specific library community.

Course Management System (CMS): Computer software system that provides a course shell with a number of integrated tools, which may include chat software, a threaded discussion board, online grade books, online testing, and other classroom functions.

Digital Libraries: Organized collections of digital information.

Distance Education: A planned teaching and learning experience that may use a wide spectrum of technologies to reach learners at a site other than that of the campus or institution delivering the course.

Information Literacy: An integrated set of skills and the knowledge of information tools and resources that allow a person to recognize an information need and locate, evaluate, and use information effectively.

Videoconferencing: Conducting a conference between two or more participants at different geographical sites by using computer networks to transmit audio and video data.

T

The CRM–KDD Nexus

Nikhilesh Dholakia
University of Rhode Island, USA

Jounghae Bang
University of Rhode Island, USA

Lutz Hamel
University of Rhode Island, USA

Seung-Kyoon Shin
University of Rhode Island, USA

INTRODUCTION OF KNOWLEDGE DISCOVERY IN DATABASES AND CUSTOMER RELATIONSHIPS

Customer relationships are increasingly central to business success (Kotler, 1997; Reichheld & Sasser, 1990). Acquiring new customers is five to seven times costlier than retaining existing customers (Kotler, 1997). Simply by reducing customer defections by 5%, a company can improve profits by 25% to 85% (Reichheld & Sasser, 1990). Relationship marketing—getting to know customers intimately by understanding their preferences—has emerged as a key business strategy for customer retention (Dyche, 2002).

Internet and related technologies offer amazing possibilities for creating and sustaining ideal customer relationships (Goodhue, 2002; Ives, 1990; Moorman, 1992). Internet is not only an important and convenient new channel for promotion, transactions, and business process coordination; it is also a source of customer data (Shaw et al., 2001). Huge customer data warehouses are being created using advanced database technologies (Fayyad et al., 1996).

Customer data warehouses alone offer no competitive advantages; insightful customer knowledge must be extracted from such data (Kim, 2002). Valuable marketing insights about customer characteristics and their purchase patterns, however, are often hidden and untapped (Shaw, 2001). Data mining and knowledge discovery in databases (KDD) facilitate extraction of valuable knowledge from rapidly growing volumes of data (Mackinnon, 1999; Fayyad et al., 1996).

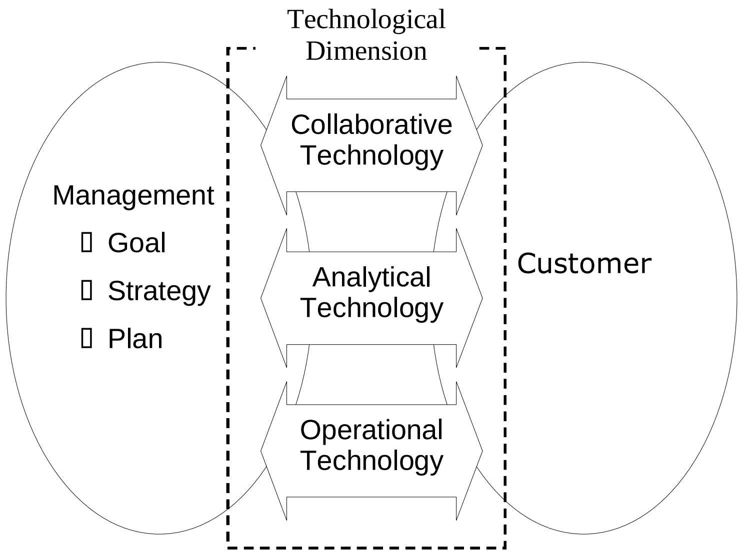
This chapter provides a brief review of customer relationship issues. The chapter focuses on (1) customer relationship management (CRM) technologies, (2) KDD techniques, and (3) key CRM-KDD linkages in terms of relationship marketing. The chapter concludes with the observations about state-of-the-art and future directions.

BACKGROUND: CRM TECHNOLOGIES

CRM is interpreted in a variety of ways (Goodhue et al., 2002; Winer, 2001; Wright, 2002). In some cases, CRM simply entails direct e-mails or database marketing. In other cases, CRM refers to customer interaction centers (CICs) and online analytical processing (OLAP), which are types of online query-driven analyses for examining stored data. Overall, CRM can be seen as a core business strategy to interact with, create, and deliver value to targeted customers to improve customer satisfaction and customer retention at a profit. It is grounded in high quality customer data and enabled by information technology (Ang & Buttle, 2002).

Three core dimensions characterize buyer-focused CRM systems: customers, management, and technologies. *Customer* service and related issues must be included in the design, implementation, and operation of any CRM system. Organizations benefit from CRM particularly when such systems benefit their customers; using CRM merely as a sales or customer service solution is a recipe for failure (Davids, 1999). *Management*'s articulation and tracking of customer relationship goals, plans, and metrics is an essential CRM component (Ang & Buttle, 2002; Greenberg, 2002). Successful CRM implementations rely on management goals, strategies, and plans that reflect customer commitment and promote a customer-responsive corporate culture at all levels of the organization (Ang & Buttle 2002; Smith 2001). *Technologies* for facilitating collaborative, operational, and analytical CRM activities are the manifest aspects of CRM (Goodhue, 2002).

Figure 1. Alignment of three dimensions of CRM



A collaborative CRM system is any CRM function that provides a point of interaction between the customer and the marketing channel (Greenberg, 2002). E-commerce and, in some cases, mobile commerce systems, offer multiple "touch points" for reaching the customers. In employing the Web and mobile technologies, it is important to ensure that such technologies enhance older, preexisting channels (Johnson, 2002). *Operational CRM* systems are technologies that span the ordering-delivery cycle (Goodhue et al., 2002). Operational CRM is concerned with automating the customer-facing parts of the enterprise (Ang & Buttle, 2002). Since the sales process depends on the cooperation of multiple departments performing different functions, integration of all such functions is critical for operational CRM systems (Earl, 2003; Greenberg, 2002). *Analytical CRM* systems analyze customer data warehouses so that the firm can detect valuable patterns of customers' purchasing behaviors. Offline data mining of customer data warehouses as well as online analytical processing (OLAP) can aid in applications such as campaign management, churn analysis, propensity scoring, and customer profitability analysis (Goodhue et al., 2002). It is this component of CRM that has a clear linkage to KDD methods.

BACKGROUND: KDD TECHNIQUES

Since multiple data formats and distributed nature of knowledge on the Web make it a challenge to collect, discover, organize, and manage CRM-related customer data (Shaw et al., 2001), KDD methods are receiving attention in relationship marketing contexts (Fayyad et al., 1996; Mackinnon, 1999). Massive databases are commonplace, and they are ever growing, dynamic, and heterogeneous (Mackinnon & Glick, 1999). Systematic combining of data mining and knowledge management techniques can be the basis for advantageous customer relationships (Shaw et al., 2001).

KDD is defined as the process of data selection, sampling, pre-processing, cleaning, transformation, dimension reduction, analysis, visualization, and evaluation (Mackinnon, 1999). As a component of KDD (Fayyad et al., 1996), data mining is defined as the process of searching and analyzing data in order to find latent but potentially valuable information (Shaw et al., 2001).

KDD constitutes the overall process of extracting useful knowledge from databases. It is a multidisciplinary activity with the following stages (Brachman et al., 1996; Bruha et al., 2000; Fayyad et al., 1996)

- Select the problem area and choose a tool for representing the goal to be achieved
- Collect the data and choose tools for representing objects (observations) of the dataset
- Preprocess the integrating and cleaning data
- Data mine—extract pieces of knowledge
- Postprocess the knowledge derived: test and verify, interpret and apply the knowledge to the problem area at hand

In Web-based relationship marketing, three distinct categories of data mining have emerged: Web content mining, Web structure mining, and Web usage mining (Jackson, 2002). Web usage mining is also referred to as clickstream analysis (Edelstein, 2001). Valuable information hidden in the clickstream data of many e-commerce sites can provide sharp diagnostics and accurate forecasts, allowing e-commerce sites to profitably target and reach key customers (Moe, 2001). Such Web-based CRM systems require large, integrated data repositories and advanced analytical capabilities. Even though there are many success stories, Web-based CRM projects continue to be expensive and risky undertakings. OLAP refers to the various types of query-driven analyses for analyzing stored data (Berry & Linoff, 1997). Data mining and OLAP can be seen as complementary tools (Jackson, 2002). Both Web-based CRM systems and OLAP, in general, involve vast volumes of both structured and unstructured data. One common challenge with managing this data is to incorporate unstructured data into a data warehouse. Traditional database systems are not designed for unstructured data.

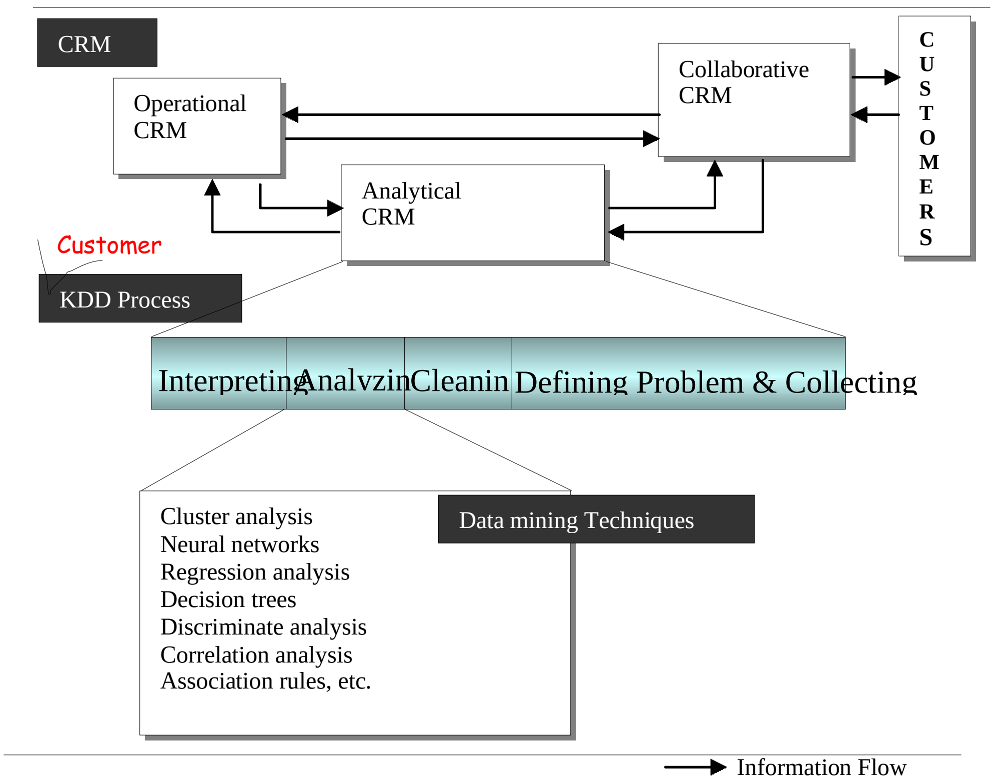
In general, research in KDD is intended to develop methods and techniques to process a large volume of unstructured data in order to retrieve valuable knowledge that is hidden in these databases, and that would be compact and abstract; yet understandable and useful for managerial applications (Bruha et al., 2000).

EXPLORING THE CRM-KDD CONNECTION

Figure 2 explains the CRM-KDD linkage from a process point of view. The importance of gaining knowledge has been well recognized. In line with this notion, CRM starts with understanding customers and gaining in-depth knowledge about customers. Therefore, the intersection between KDD and CRM can be seen as the analytical CRM part of CRM systems and customer knowledge discovery in database processes of overall KDD process as shown in Figure 2. Collaborative CRM systems help collect accurate information from customers, while operational CRM can capitalize on the result of the analyses. The problem definition stage of KDD process can be completed also in the management dimension of CRM. Following the definition of KDD and data mining, techniques for data mining are included under the analysis stage of KDD.

In relation to marketing issues, with help of CRM and KDD technologies, database marketing and one-to-one marketing methods have come to the fore. The strategic goal of database marketing is to use collected information

Figure 2. CRM & KDD process connection



to identify customers and prospects, and to build continuing personalized relationships with them, leading to greater benefits for the individuals and greater profits for the corporation (Kahan, 1998). Database marketing anticipates customer behavior over time and reacts to changes in the customer's behavior. Database marketing identifies unique segments in the database reacting to specific stimuli such as promotions (McKim, 2002).

One-to-one marketing represents the ultimate expression of target marketing—market segments with just one member each—or at least one at a time (Pitta, 1998). It relies on a two-way communication between a company and its customers in order to enhance a true relationship, and it allows customers to truly express the desires that the company can help to fulfill (Dyche, 2002). A promising solution to implementing one-to-one marketing is the application of data mining techniques aided by information technology. Data mining allows organizations to find patterns within their internal customer data. Uncovered patterns can lead to actionable and meaningful target segments.. Armed with such information, organizations can refine their targets and develop their technology to achieve true one-to-one marketing (Pitta, 1998).

As an extension of one-to-one marketing, the concept of permission marketing is focused on seeking customers' consents about desired marketing methods. Customers not only need to communication, they need to be able to stipulate how and when they wish to be approached (Newell, 2003). One-to-one and permission marketing rely heavily on information technology to track individual customers, understand their differences, and acknowledge their interaction preferences (Dyche, 2002).

Data mining methods allow marketers to sift through growing volumes of data and to understand their customers better. Shaw et al. (2001) introduce three major areas of application of data mining for knowledge-based marketing: (1) customer profiling, (2) deviation analysis, and (3) trend analysis. Also, Jackson (2002) notes that data mining may be used as a vehicle to increase profits by reducing costs or raising revenue. Following are some of the common ways to use data mining in customer relationship contexts.

- Eliminate expensive mailings to customers who are unlikely to respond to an offer during a marketing campaign.
- Facilitate one-to-one marketing and mass customization opportunities in customer relationship management.

Many organizations use data mining to help manage all phases of the customer lifecycle, and CRM systems can benefit from well-managed data analyses based on data mining. Table 1 summarizes the relationship marketing issues and includes the possible customer analyses and relevant data mining techniques.

Table 1. Customer relationship related data analysis and data mining tools

	CUSTOMER RELATIONSHIP MARKETING ISSUES		
	Database Marketing	One-to-One Marketing	Permission Marketing
Issue	Understanding customers with the database on customer behavior over time including reactions to changes	Communicating with customers as individuals Developing custom products and tailored messages based on customers' unspoken needs.	Seeking customers' agreement about desired marketing methods.
Challenge	Identifies unique segments in the database	Find patterns within the internal customer data. Track individual customers Understand their differences	Track individual customers Understand their differences Acknowledge their interaction preferences Stimulate the customer's response
Possible analysis	Segmentation	Classification Prediction	Classification Dependency Analysis
Datamining Technique most likely used	Descriptive and visualization Cluster Analysis Neural networks	Regression analysis Neural networks Decision Trees Discriminant Analysis	Descriptive and visualization Neural networks Regression analysis Correlation Analysis Decision Trees Discrimination Analysis Case-Based Reasoning Association Rules

FUTURE TRENDS

Due to the advance of information technology, there are more opportunities to collect data about customers. The Internet provides and promotes interactive communications between a business and its customers and leads to an increasing volume of rich data about customers.

However, in interactive marketing contexts, customers are also able to block out the intrusive marketing actions, and, therefore, appropriate depth and width of permissions should be obtained (Godin, 1999; Krishnamurthy, 2001). Understanding customers will become more critical as new information technology is developed.

Furthermore, companies and customers will have opportunities to co-create products, pricing, and distributions. Information technology provides these opportunities by allowing companies to assess each customer individually, and then to determine whether to serve that customer directly or through a third party and whether to create an offering that customizes the product or standardizes the offering (Sheth et al., 2000). All of the decisions should be based on thorough analyses of customer data and accurate knowledge generation about customers.

In addition to knowledge generation, knowledge sharing and dissemination through the organization should be considered. Shaw et al. (2001) argued that ownership and access to the marketing knowledge, standards of knowledge interchange, and sharing of applications become critical. In various organizational environments, both managing KDD processes to generate customer knowledge and managing customer relationship based on the knowledge generated and shared through organizations are challenges for the future.

CONCLUSION

This chapter offered a brief review of customer relationship issues. CRM systems consist of management, technology, and customer dimensions. CRM technologies are divided into three categories: analytical, operational, and collaborative CRM. As the importance of knowledge increases, KDD techniques are receiving attention as systematic processes to generate knowledge. Although CRM and KDD began separately, the two concepts have points of convergence. This chapter highlighted some of the intersections between the two.

Different relationship marketing issues have emerged; these issues rely increasingly on CRM and KDD technologies, especially for in-depth analysis. Various data mining techniques and KDD processes exist and provide the right tools to solve relationship marketing problems. While companies are eager to learn about their customers by using data mining technologies, it is very difficult to choose the most effective algorithms for the diverse range of problems and issues that marketers face (Kim et al., 2002). Even though Table 1 illustrates the main relationship marketing issues, the challenges, and the potential analytic data mining tools, it is the analyst who decides creatively which tool is appropriate for what situation and how to interpret the results.

From a process point of view, gaining customer knowledge becomes critical for managing customer relationships, and systematic knowledge generating processes are of great benefit. For effective customer-centric marketing strategies, the discovered knowledge has to be managed in a systematic manner.

REFERENCES

Berry, M.J.A., & Linoff, G. (1997). *Data mining techniques: For marketing, sales, and customer support*. New York: John Wiley & Sons, Inc.

Brachman, R.J., Khabaza, T., Kloesgen, W., Piatetsky-Shapiro, G., & Simoudis, E. (1996). Mining business databases. *Communications of the ACM, 39*(11), 42-48.

Bruha, I., Kralik, P., & Berka, P. (2000). Genetic learner: Discretization and fuzzification of numerical attributes. *Intelligent Data Analysis, 4*, 445-460.

Davids, M. (1999). How to avoid the 10 biggest mistakes in CRM. *The Journal of Business Strategy*, *20*(6), 22.

Dyche, J. (2002). *The CRM handbook: A business guide to customer relationship management*. Boston: Addison-Wesley.

Edelstein, H.A. (2001, March 12). Pan for gold in the clickstream. *Informationweek.com*

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Association for Computing Machinery. Communications of the ACM, 39*(11), 27.

Godin, S (1999). *Permission marketing*. New York: Simon and Schuster.

Goodhue, D.L., Wixom, Barbara H., & Watson, Hugh J. (2002). Realizing business benefits through CRM: Hitting the right target in the right way. *MIS Quarterly Executive, 1*(2), 79-94.

Greenberg, P. (2002). *CRM at the speed of light: Capturing and keeping customers in Internet real time* (2nd ed.). Berkeley and London: McGraw-Hill.

Ives, B., & Mason, Richard O. (1990). Can information technology revitalize your customer service? *Academy of Management Executive, 4*(4), 52-69.

Jackson, J. (2002). Data mining: A conceptual overview. *Communications of the Association for Information Systems, 8*(2002), 267-296.

Johnson, L.K. (2002). New views on digital CRM. *Sloan Management Review, 44*(1), 10.

Kim, E., Kim, Wooju, & Lee, Yillbyung (2002). Combination of multiple classifiers for the customer's purchase behavior prediction. *Decision Support Systems, 34*(2002), 167-175.

Krishnamurthy, S. (2001). A comprehensive analysis of permission marketing. *Journal of Computer-Mediated Communication, 6*(2).

Mackinnon, M.J., & Glick, Ned. (1999). Data mining and knowledge discovery in databases—an overview. *Australia & New Zealand Journal of Statistics, 41*(3), 255-275.

Moe, W.W., & Fader, Peter S. (2001). Uncovering patterns in cybershopping. *California Management Review, 43*(4), 106-117.

Moorman, C., Zaltman, Gerald, & Deshpande, Rohit (1992, August). Relationships between providers and users of market research: The dynamics of trust within and between organizations. *Journal of Marketing Research, 24*, 314-328.

Newell, F. (2003). *Why CRM doesn't work: How to win by letting customers manage the relationship*. NJ: Bloomberg Press.

Shaw, M.J., Subramaniam, Chandrasekar, Tan, Gek Woo, & Welge, Michael E. (2001). Knowledge management and data mining for marketing. *Decision Support Systems, 31*, 127-137.

Sheth, J.N., Sisodia, Rajendra S., & Sharma, Arun (2000). The antecedents and consequences of customer-centric marketing. *Journal of Academy of Marketing Science, 28*(1), 55.

Wright, L.T., Stone, Merlin, & Abbott, Julie (2002). The CRM imperative—Practice vs. theory in the telecommunications industry. *Journal of Database Marketing, 9*(4), 339-349.

KEY TERMS

Clickstream Data: Web usage data. A virtual trail that a user leaves behind while surfing the Internet (e.g., every Web site and every page of every Web site that the user visits; how long the user was on a page or a site).

CRM – Customer Relationship Management: A core business strategy that promotes interactions, creates and delivers value to targeted customers to improve customer satisfaction and customer retention at a profit. It is grounded in high quality customer data and enabled by information technology.

DM – Data mining: The process of searching and analyzing data in order to find latent but potentially valuable information and to identify patterns and establish relationships from a huge database.

E-Commerce: Any business done electronically. The electronic business where information technology is applied to all aspects of a company's operations.

KDD – Knowledge Discovery in Databases: The process of data selection, sampling, pre-processing, cleaning, transformation, dimension reduction, analysis, visualization, and evaluation for the purpose of finding hidden knowledge from massive databases.

OLAP – Online Analytical Processing: Various types of online query-driven analyses for examining stored data. OLAP enables a user to easily and selectively extract and view data from different points-of-view.

The Impact of IT on Business Partnerships and Organizational Structures

T

Fang Zhao
RMIT University, Australia

INTRODUCTION

Over the past decade, with the advent of the Internet, organizations have changed the way they communicate internally and externally, the way they are configured, and the way they build partnerships. Today's complex and volatile business world calls for changes and alternatives to the old and conventional paradigm of organizational design and new ways of doing business with others. E-business becomes one of the most important forces shaping today's business. Virtual corporations and e-partnerships become increasingly popular in the perception of managers and in business operations.

This article explores the changing nature, process, and practice of business relationships and network form of organizations in cyberspace. It also identifies and discusses a series of management issues raised in the processes of e-partnerships and virtual organizations.

BACKGROUND

The virtual organization, which is actually a network form of organization, is a revolution in organizational design and has changed the definitions, boundaries, and forms of inter-organizational collaboration and partnerships. The network form of organizations is the defining business transformation of this generation (Malone & Davidow, 1994; Jin, 1999; Hagel & Singer, 2000). Cisco, for example, is a network of suppliers, contract manufacturers, assemblers, and other partners which is connected through an intricate Web of information technology. Seventy percent of Cisco's product is outsourced to its e-partners through Cisco's network (McShane & von Glinow, 2000).

As shown above, virtual organizations rely on IT network and e-partnership. Theoretically, e-partnership refers to a partnership relying on electronic (information) technologies to communicate and interact amongst partners. In practice, the term e-partnership is mostly associated with e-commerce or e-business partnerships. It may take different forms and involve various partners from or between virtual enterprises and brick-and-mortar companies, depending on the nature of e-business activities. It flourishes in particular in supply chains through adding electronic components to the business partnership across firms (O'Toole, 2003). For example, in the manufacturing industry, the e-partners may include raw materials providers, component manufacturers, final assembly manufacturers, wholesalers, distributors, retailers, and customers (Cheng, Li, Love & Irani, 2001). This supply chain may involve a number of—even hundreds or thousands of—suppliers and distributors. The use of the Internet and other electronic media and the introduction of inter-organizational information systems are constitutive to e-partnerships and lead to the growth of virtual organizations.

E-partnerships share some common characteristics with traditional inter-organizational partnerships. But they are different in many ways and thus require different strategies and structures to manage them. Bell (2001) and de Man, Stienstra, and Volberda (2002) studied the differences and found that e-partnerships are generally entrepreneurial and less planned in nature, must move at Web speed, require flexible network structure, and have a short lifespan.

E-partnerships as "a new breed of online alliances, are fast emerging as the result of an incredible amount of Internet business in recent years" (Trask, 2000, p. 46). Entering an e-partnership is no longer a soft option, but a vital need for gaining a competitive advantage and customer satisfaction in the trend of economic globalization (meaning an increasing trend of economic integration worldwide). On the other hand, globalization is pushing companies to build informal network organizations, such as virtual organizations, that are able to work in a faster, cheaper, and more flexible way. E-partnerships and virtual organizations are products of the globalization and IT advancement over the past decade and they have fundamental synergy between them. They interrelate and interact with each other in this digital era.

ADVANTAGES

The greatest advantage of e-partnership and virtual organization lies in the fact that they eliminate the physical boundaries of organizations, and that cross-functional teams and organizations are able to operate and collaborate across space and time by communicating with each other via electronic channels. The Internet becomes the most important interface between participating organizations, teams, and individuals. E-partnerships and virtual organizations enable businesses to sell and deliver products and services across the world in a more efficient way in terms of speed and cost. Amazon.com, Priceline.com, and E*Trade are some of the successful e-businesses that have depended on, and maximized profits from, e-partnerships.

Other perceived benefits of e-partnerships and virtual organizations may include greater business opportunities, better integration of suppliers and vendors, better management information, lower operational costs, better market understanding, and expanded geographical coverage (Damanpour, 2001). E-partnerships and virtual organizations may also offer the opportunity of consolidating resources of all partners and organizational flexibility as other forms of inter-organizational partnerships and alliances do.

In this rapidly changing competitive landscape, few organizations can rely on their internal strengths only to gain a competitive advantage in national and/or international markets. Inter-organizational collaborations, alliances, joint ventures, partnering, and the like are gaining unprecedented momentum, regardless of their organizational and management structures and styles and communication channels. An organization's resources are limited in one way or another. Forming a business e-partnership and taking a network form of organization is increasingly one of the most popular strategies available to an organization to take advantage of an Internet highway on the one hand and share risks, capabilities, and revenue with partners on the other. The driving forces behind building an e-partnership and a virtual organization share a lot in common with those driving any other form of inter-organizational collaboration. They include:

- to gain a competitive advantage or increase market share in national and/or global markets;
- to share revenue and expand sales between merchant and partners;
- to prevent competition loss;
- to meet changing demands of customers and markets; and
- to gain core competencies from competitors (Sierra, 1994; Dussauge & Garrette, 1999; Trask, 2000).

KEY ISSUES

However, like e-business and e-commerce, e-partnership is also facing a range of issues related to the use of the Internet, as well as the reliance on inter-organizational interfaces. The key issues identified are:

- challenges and risks of e-partnerships and virtual organizations;
- productivity and revenue sharing in e-partnerships and virtual organizations;
- transferring and sharing core competencies between participating organizations;
- power disparity; and
- quality and effectiveness of communication.

Addressing each of these issues has posed a formidable task in front of e-managers of various kinds of inter-organizational collaboration through electronic technologies and e-networks. The following discussion explores each of these issues in detail.

Challenges and Risks

On the technological side, companies that are involved in e-partnerships must participate in external business relationships by using computer interactions. This forces e-managers to re-engineer their IT strategies and resources, and re-think their ways of communicating and doing business with e-partners in a virtual environment. Main issues to be considered are IT infrastructure and managers' and operatives' knowledge and skills associated with e-business.

On the human resources' side, e-managers are surely confronting management complexities of making cooperation work. One of the biggest challenges is conflict and differences in organizational and country cultures and systems. Each organization has its own culture developed from its own particular experience, its own role, and the way its owners or managers get things done (Hellard, 1995). In addition to the cultural differences at organizational level, multi-national e-partnerships inevitably encounter barriers caused by cultural differences between nations such as clashes between western and eastern cultures. Differences exist in systems including taxation systems, online intellectual property, and online trade and law. For example, EU member states must enact legislation to ensure that transfers of data outside their boundaries are allowed only to jurisdictions that can offer adequate protection of the data. The U.S. believes that minimal domestic regulation would foster cross-border Internet trade (Damanpour, 2001). Managing the culture and system differences across organizations and across nations is one of the high agendas that challenge managers of e-partnerships and virtual organizations.

While the Internet and network organizations facilitate improved communication of data, information, and knowledge, they give rise to issues and problems of privacy, data security, and intellectual property protection in the Internet. The information database created through Internet transactions may lead to legal disputes among e-partnerships over ownership of the IP and possible loss of the potential profit generated from the IP (Greif, 2000). Moreover, electronic research projects usually involve new technologies and innovative development, which creates a high level of technological, commercial, and legal risks for every organization involved.

Productivity and Revenue Sharing

The primary aim of building e-partnerships and virtual organizations is to generate more profit and achieve the best business results through taking advantage of online resources and extensive e-networks. Trask (2000, p. 46) considered that “a well-designed revenue-sharing program may be the best and fastest way to generate online business.” Revenue sharing becomes the most important issue in e-partnerships and virtual organizations when productivity increases and revenue goes up. The nature, timing, and amount of compensation (in the form of referral fees, royalty, and commission), together with the financial stability and honesty of commission reporting, are core considerations of e-partners and crucial factors of success in sustaining e-partnerships.

Transferring and Sharing Core Competencies

According to Lei, core competencies comprise a company’s specific and special knowledge, skills, and capabilities to stand out amongst competitors. They are intangible and an integrated part of a company’s intellectual capital and un-tradable asset rather than legally protected intellectual property (Lei, 1997, p. 211). Inter-organizational collaboration provides an opportunity for participating organizations to acquire and absorb the core competencies from each other (Couchman & Fulop, 2000). This opportunity is particularly valuable for innovative business such as e-business. However, the greatest barrier is competitive concerns over information leakage. This is an unavoidable dilemma facing e-partnerships, which makes it difficult for e-partners to achieve the potential that IT technology can offer.

Power Disparity

It is normal that a decision-making body of inter-organizational collaboration and partnership is proportionately represented by participating organizations in terms of equity holdings in an online joint venture. It should be noted that due to difference in equity holdings, power disparity occurs and is likely to affect performance of inter-organizational collaboration, although division of power and responsibility has been clearly defined in legally binding agreements between e-partners.

Quality and Effectiveness of Communication

Networking and communication play a key role particularly in coordinating and liaising inter-organizational collaboration. Expanding e-networks and achieving effective communication amongst e-partners are a top priority. Like culture and commitment, communication is a soft outcome of a total quality partnership approach and the foundation for inter-organizational collaboration (Rounthwaite & Shell, 1995; Hellard, 1995; Aggarwal & Zairi, 1998). Effective networking and communication help to eliminate barriers to collaboration. Therefore, continuous improvement of quality and effectiveness of communication amongst partners is another key issue in the agenda of e-partnerships and virtual organizations.

OPTIONS

While sufficient support of IT infrastructure and resources are definitely important to successful e-partnerships and virtual organizations, reducing potential financial, commercial, and legal risks, and effectively dealing with human and cultural factors exceed the complexities of technical setup and support in building e-partnerships and virtual organizations. Organizational culture and human resources are increasingly becoming important sources of competitive advantage, as they are difficult for competitors to imitate. How to foster a robust culture and capitalize on the core competence derived from e-partnerships is a crucial and tough issue for managers. Other critical success factors for e-partnerships and virtual organizations include:

- level of accessibility, security, and compatibility of inter-organizational information systems;
- level of traffic in collaborative e-commerce activities;
- level of customer service and e-partner support service;
- level of transferring and sharing information and knowledge between e-partners;
- building and sustaining an effective virtual network structure amongst e-partners;
- level of individual and organizational commitment to e-partnerships;

- level of mutual trust, understanding, respect, and openness;
- level of corporate and business ethics and integrity;
- level of credibility of e-partners in relation to financial situation and business experience;
- level of mutual benefit through revenue sharing;
- effectiveness and efficiency of a real-time commission reporting system;
- level of performance and productivity of e-partners;
- actively pursuing and sharing core competencies;
- willingness to share power and empower amongst e-partners; and
- quality and effective networking and continuous improvement of communication.

Achieving the best collaboration among e-partners requires more than tangible resources like IT infrastructure and support. Successful e-partnership needs a high level of intangible commitment and efforts to understand the needs and values of e-partners and customers. By resorting to a total quality partnership approach, it means that business ethics, integrity, honesty, trust, and sharing are required of e-managers of inter-organizational entities at the top and of the individuals and teams throughout the entire virtual organization (Rounthwaite & Shell, 1995; Hellard, 1995; Aggarwal & Zairi, 1998). Disputes and conflicts caused by culture and system differences, like those illustrated in this article, could be reduced if e-partners could accept the differences and maintain a flexible and realistic attitude towards the differences.

FUTURE TRENDS

"The world has finally become a Global Village, not just in rhetoric, but in reality" (Hennessey, 2000, p. 34). In today's corporate world, it will be more difficult for businesses to survive without joining an e-partnership and taking advantage of the Internet. The fast expansion of Amazon.com, travel.com, and the like through e-partnerships and e-networks and their business success reinforce the importance of online strategic alliance. Corporate e-partnerships and network-based organizations will be a crucial factor and play a key role in the future development of online business activities. The future trends will be characterized by:

- a more mature (rather than experimental) nature of e-commerce and e-business practices in terms of the scope, quality, and credibility of online customer services and products;
- more needs for devising and popularizing e-supply chains due to the needs for integrating the flow of information with the flow of goods (van Hoek, 2001; Kotzab, Skjoldager & Vinum, 2003);
- greater monopoly of the flow of e-commerce and e-business by bigger online syndicates like Amazon.com through building extensive online alliances with online retailers and suppliers (Werbach, 2000);
- more brick-and-mortar companies moving online to expand their business scope and capitalize on the abundance of the Internet;
- greater reliance on joint efforts across nations in online legislation to protect IP, security, and privacy of e-commerce and e-business; and
- greater challenge for dot.com industries to achieve sustainability, due to a more uncertain economic environment and the increasing complexities of new technologies and the more globalized economy.

CONCLUSION

In today's e-business context, technology, customers, competitors, and partners can change rapidly. Technology can become obsolete in the blink of an eye, and customers can appear and disappear with a keystroke. There are practically no barriers to new entrants (competitors) in an e-business world. E-business partnerships and virtual organizations can become ephemeral and opportunistic in nature, depending on the interdependencies between partners and the extent to which their objectives have been achieved (Robbins, Bergman, Stagg & Coulter, 2003). In such circumstances, it is important that e-business managers have an insightful knowledge and are well prepared to deal with the complexities of e-partnerships and virtual organizations. This article provides a better understanding of the crucial issues in cross-firm business processes and inter-organizational partnerships in cyberspace.

Running inter-organizational partnerships implies multiplication of decision-making bodies from each participating organization, and a potential clash of interest and values amongst participants. As illustrated in this article, total quality partnership embodies the fundamental principles for managing collaborative partnerships including e-partnerships, and can be developed and extended to help inter-organizational collaboration to achieve desired outcomes. However, managing e-partnership and virtual organization is more complex than managing intra-organizational collaboration and collaboration between brick-and-mortar companies due to the IT issues, human and cultural issues, and inter-organizational partnership issues as discussed in the article. Failure to consider the complexities of any of these issues will lead to a divorce of e-partnerships and a collapse of virtual organizations.

REFERENCES

Aggarwal, A.K. & Zairi, M. (1998). Total partnership for primary health care provision: A proposed model–part II. *International Journal of Health Care Quality Assurance*, *11*(1), 7-13.

Bell, J. (2001). E-alliances: What's new about them? In A.P. de Man, G.M. Duysters & V. Vasudevan (Eds.), *The allianced enterprise* (pp. 25-30). Singapore: Imperial College.

Cheng, W.L.E., Li, H., Love, E.D.P., & Irani, Z. (2001). An e-business model to support supply chain activities in construction. *Logistic Information Management*, *14*(1/2), 68-78.

Couchman, P., & Fulop, L. (2000). Transdisciplinary research bodies: The changing nature of organizational networks and R&D in Australia. *Journal of World Business*, *8*, 213-226.

Damanpour, F. (2001). E-business e-commerce evolution: Perspectives and strategy. *Managerial Finance*, *27*(7), 16-32.

de Man, A.P., Stienstra, M., & Volberda, H.W. (2002). E-partnering: Moving bricks and mortar online. *European Management Journal*, *20*(4), 329-339.

Dussauge, P., & Garrette, B. (1999). *Cooperative strategy: Competing successfully through strategic alliances*. New York: John Wiley & Sons.

Greif, J. (2000). Risky e-business. *Association Management*, *52*(11), 55.

Hagel, J., & Singer, M. (2000). Unbundling the corporation. In N.G. Carr (Ed.), *The digital enterprise: How to reshape your business for a connected world* (pp. 3-20). Boston: Harvard Business School.

Hellard, R.B. (1995). *Project partnering: Principle and practice*. London: Thomas Telford Publications.

Hennessey, A. (2000). Online bookselling. *Publishing Research Quarterly*, *16*(2), 34.

Jin, Z. (1999). Organizational innovation and virtual institutes. *Journal of Knowledge Management*, *3*(1), 75-83.

Kotzab, H., Skjoldager, N., & Vinum, T. (2003). The development and empirical validation of an e-based supply chain strategy optimization model. *Industrial Management & Data Systems*, *103*(5), 347-360.

Lei, D.T. (1997). Competence building, technology fusion and competitive advantage: The key roles of organizational learning and strategic alliances. *International Journal of Technology Management*, *14*, 208-237.

Malone, M., & Davidow, B. (1994). Welcome to the age of virtual corporations. *Computer Currents*, *12*(1), 12-24.

McShane, L.S., & von Glinow, A.M. (2000). *Organizational behavior*. Sydney: Irwin-McGraw-Hill.

O'Toole, T. (2003). E-relationships–emergence and the small firm. *Marketing Intelligence & Planning*, *21*(2), 115-122.

Robbins, S.P., Bergman, R., Stagg, I. & Coulter, M. (2003). *Management* (3rd edition). Sydney: Pearson Education Australia.

Rounthwaite, T., & Shell, I. (1995). Techniques: Designing quality partnerships. *The TQM Magazine*, *7*(1), 54-58.

Sierra, M.C.D.L. (1994). *Managing global alliances: Key steps for successful collaboration*. Wokingham: Addison-Wesley.

Trask, R. (2000). Developing e-partnerships. *Association Management*, *52*(11), 46.

van Hoek, R. (2001). E-supply chains–virtually non-existing. *Supply Chain Management: An International Journal*, *6*(1), 21-28.

Werbach, K. (2000). Syndication: The emerging model for business in the Internet era. In N.G. Carr (Ed.), *The digital enterprise: How to reshape your business for a connected world* (pp. 21-34). Boston: Harvard Business School.

KEY TERMS

Brick-and-Mortar Organization: An organization located or serving customers in a physical facility as opposed to a virtual organization.

E-Business (electronic business): A comprehensive term used to describe the way an organization interacts with its key constituencies, including employees, managers, customers, suppliers, and partners through electronic technologies. It has a broader connotation than **e-commerce** because e-commerce is limited to business exchanges or transaction over the Internet only.

E-Partnership: A partnership relying on electronic (information) technologies to communicate and interact amongst partners. It is mostly associated with e-commerce or e-business partnerships.

E-Supply Chain: The physical dimension of e-business with the role of achieving a base level of operational performance in the physical sphere (for more detail see van Hoek, 2001).

Intellectual Property (IP): A product of the intellect (intangible property) that has commercial value such as patents, trademarks, copyrights, and so forth.

Online Syndicate: Association of firms with a common interest formed to engage in e-business. Syndication has become an important e-business strategy of many companies.

Real Time: In the context of e-commerce, a real-time commission reporting system refers to a system in which a commission request is processed within milliseconds so that a commission report is available virtually immediately to an online salesperson.

Virtual Organization: A network form of organization that operates across space, time, and organizational boundaries through an intricate Web of information technology.

T

The Impact of Sound Relationships on Achieving Alignment

Awie Leonard
University of Pretoria, South Africa

INTRODUCTION AND BACKGROUND

International Data Corporation (CIO, 1997) surveyed 283 top executives across three vertical industries: finance, manufacturing, and retail/wholesale. They found "a strong correlation between the effectiveness of the IT department (IS organization) and the relationship between the CIO and the CEO". "We suspect that this relationship, if it is close, permits the CIO to develop the IT department (IS organization) into a service that delivers competitive advantage for the company, thus enhancing the careers of every IT professional in the organization." In other words, "a certain amount of mutual esteem will help IT (IS) function as a business partner".

In terms of alignment, sound relationships between IT and the business become even more important. Boar (1994) states that aligning with anything other than the customer leads to but momentary success. For the IT function to achieve a state of alignment with the business, it must align with the business scope, and through that business scope enable all business functions and processes to serve the customers in a superior manner.

In their research, Reich and Benbasit (1999) point out that there are two dimensions to strategy creation: the intellectual dimension and the social dimension. Research into the intellectual dimension is more likely to concentrate on the contents of plans and on planning methodologies. Research into the social dimension is more likely to focus on the people involved in the creation of alignment. The social dimension of alignment is defined as "the state in which business and IT executives within an organizational unit understand and are committed to the business and IT mission, objectives, and plans".

Another theoretical perspective supporting the concept of the social dimension of alignment is the social construction of reality. This view would suggest that, in addition to studying artifacts (such as plans and structures) to predict the presence or absence of alignment, one should investigate the contents of the players' minds: their beliefs, attitudes and understanding of these artifacts.

This article focuses on the social dimension in terms of the construction and nature of sound IT/end-user relationships and the role such relationships play in aligning IT with the business. Research in this field has shown that relationships between IT professionals and their end users are intriguing and complex, and should be seen and managed as a multidimensional environment.

IT–END-USER RELATIONSHIPS: HISTORICAL FOUNDATIONS

For many years the *culture gap* between IT departments and their end users has been characterized by unfortunate differences like distrust, skepticism and cynicism. This situation impacts negatively on the relationship of IT departments with their end users, and as such on their ability to produce service and support of high quality.

Historically, the gap was caused mainly by the difference in management culture, as well as human behaviour problems on both sides. Umbaugh (1991) states in his argumentation of organizational imbalances that too often IT exists as an adjunct to the organization and not as an integral part of the whole. This situation unfortunately still exists today and contributes to the so-called *culture gap* between IT departments and their end users. Du Plooy (1995) explains this gap as follows:

"...the 'culture gap' should be understood as a gap of misunderstanding in the sense of two different organizational 'cultures' that, according to Grindley, coexist in most organizations. The two cultures under discussion here are the 'culture' of the IT profession and the 'culture' of the rest of the organization."

The culture on both the IT department and the business side is also an important obstacle in building mutual trust, and eventually in building sound relationships between IT and its end-user environment, and as such in creating alignment between IT and the business. According to Moad (1994), the IT professional has been fighting for recognition and relevance at the CEO level for the last 25 years. He gives many examples illustrating the kind of culture that exists, which could be described as the main reason for misunderstandings and misconceptions about IT amongst today's end users.

When a user initially gets involved with the IT department, he/she is introduced to one or more IT professionals who will specifically deal with his/her problem(s). Normally a sense of mutual understanding and trust grows out of this relationship, which will definitely get disturbed the moment elements of such a relationship change without the knowledge or approval of the role players. In practice, end users very seldom get involved in the management of change which will influence a relationship in which they are involved, or even get properly informed of changes that take place on the IT side. Practice has indicated that this is a typical reason for distrust and criticism against IT departments from the end-user environment.

A review of literature on the history of relations between end users and their IT departments in the data processing industry and how they were treated tells a very sad tale. The attitude or behaviour of IT departments or the so-called DP professionals was one of "we know the best," or "we know what the end user needs and therefore we don't need to try and get the end user involved". Furthermore, even today we get end users on the business side of an organization that are illiterate in terms of computer technology. These end users are normally very uncertain when getting involved in systems development projects and in many cases are the "prey" of an IT department. As a result of this attitude, few attempts were made to keep communication with the end user on a sound basis while developing a system.

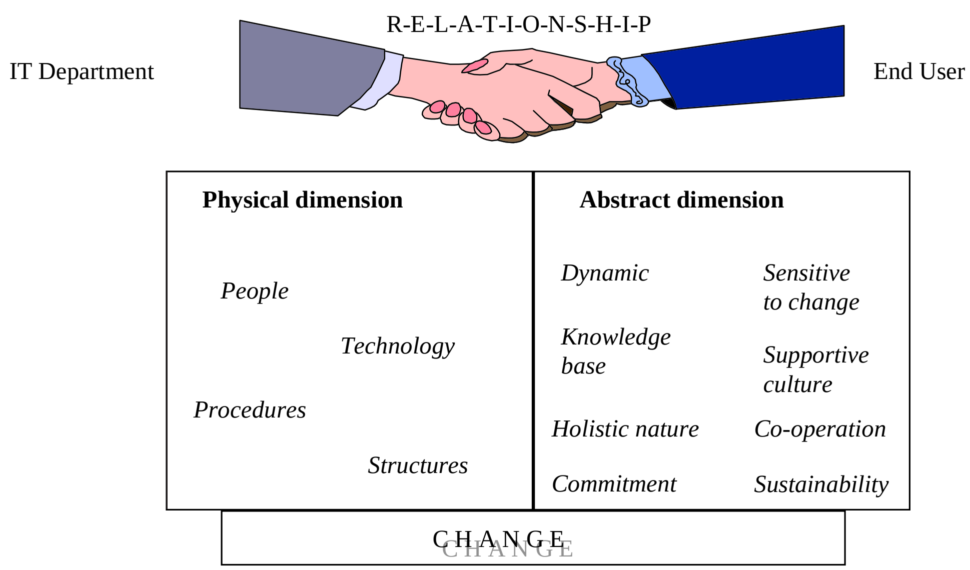
Jackson (1986) describes the traumatic period in systems development in the 1960s and early 1970s. In those days programming was considered an "art," and cost and time overruns were the rule rather than the exception. As systems development methodologies developed, control was tightened and the development process assumed many of the characteristics of an engineering discipline. In order to develop a quality product, especially in a highly complex application system, considerable skill and management expertise were required. One result of the new procedures was, if anything, a lengthening of the time it took to deliver a product. This, coupled with increased user awareness of computer potential, resulted in an increased rather than diminished backlog of applications awaiting development. According to Jackson, a study done by the Sloan School of management found that IT managers considered their application backlog to average 68% of their existing applications software portfolio. However, a survey of users found an additional 112% unbeknown to the IT managers. This illustrates, amongst other things, the typical result of poor relations between IT and its end users.

Although many efforts were made in the past to address these issues, the emphasis mainly fell on putting structures and procedures together in order to get out of the end user what his or her basic needs are. Thereafter, the IT department normally followed a lonely journey through the last phases of the systems development life cycle.

THE NATURE OF IT/END-USER RELATIONSHIPS

The preceeding paragraphs briefly describe the history of how poor relationships emerged over the years between IT departments and their end users, as well as some basic characteristics of such poor relationships. The question one can ask is, what are the characteristics of sound

Figure 1. The basic components of an IT/end-user relationship



relationships between IT departments and their end users, and how are they established? To answer the question, this section gives a definition of IT/end-user relationships and briefly discusses the nature of the different elements.

A relationship between an IT professional and an end user consists of two dimensions, namely a physical dimension and an abstract dimension. The physical dimension describes those elements that are necessary in order to enable contact between IT professional and its end users, whereas the abstract dimension describes the soft issues of such a relationship. These two dimensions enable one to fully describe the holistic nature of such a relationship and encapsulate the important elements of a support-oriented organization, namely mutuality, belonging, and connection, as mentioned by Pheysey (1993) in her book *Organizational Cultures.* The basic components of such a relationship are illustrated in Figure 1.

Without going into all the detail of the different elements of the physical and abstract dimensions as described by Leonard (2002), the article focuses on describing the most important characteristics of these elements. This will give the reader enough understanding of the social nature of IT/end-user relationships.

Physical Elements

As far as the physical dimension is concerned, the following elements could be seen as the most important:

- People: A relationship consists of all the responsible people who are involved in the systems development life cycle at a given time. “Responsibilities are negotiated and shared between systems developers and users” (Dahlbom & Mathiassen, 1993).
- Technology: Technology may be seen as one of the most important elements in such a relationship, enabling the people who participate in the relationship to communicate with one another. The importance of proper communication structures, both vertically and horizontally, are emphasized by Bommer et al. (1991) and could be seen as one of the most important organizational characteristics associated with unethical activity. Apart from the normal communication technology, facilities like help desks and Internet are of the most important factors in this regard.
- Procedures: Two types of procedures are of importance, namely organizational procedures (such as standards and policies) which already exist and which can be seen as a given, and new procedures that are being created by people because of their interaction with the given procedures and technology (DeSanctis & Poole, 1994).
- Structures: Depending upon the “type” of end user, and therefore the service and support that will be offered, relationships will differ in content as far as formal and informal social communication structures are concerned. The most common of these structures are project meetings, JAD sessions and end-user group meetings.

Abstract Elements

As far as the abstract dimension is concerned, the following elements are the most important:

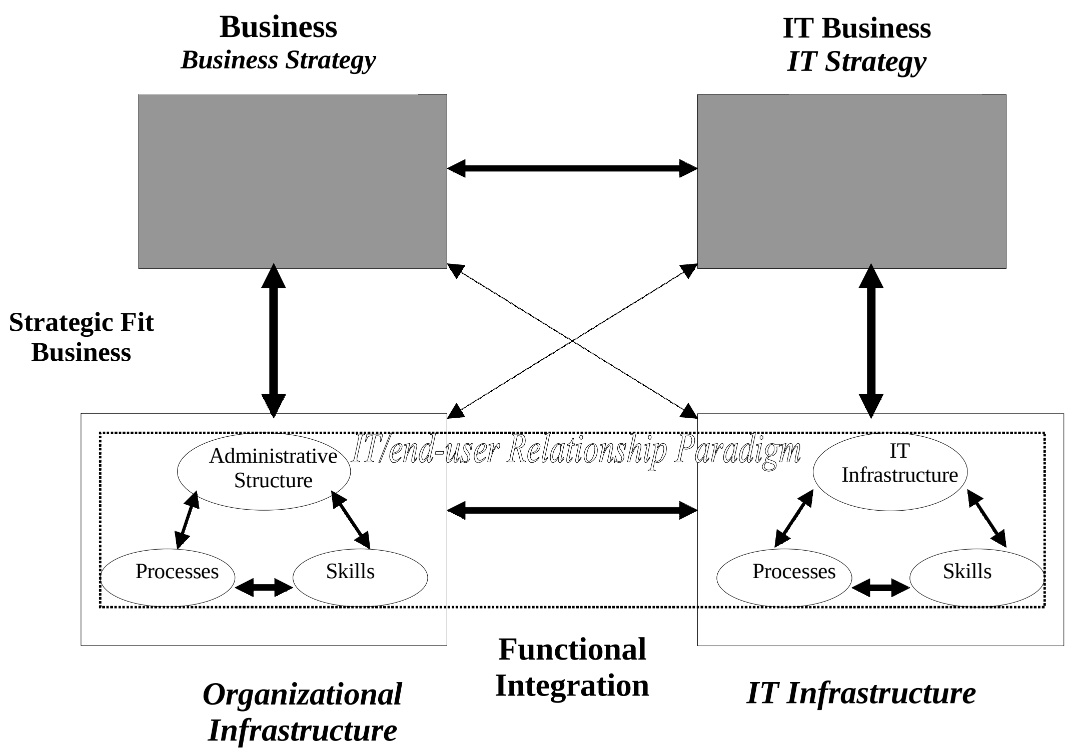
- They are dynamic: The nature of the relationships between the IT department and its end users will, amongst other things, depend upon the type of end user, as well as upon regarding the end user as a human being. According to Stokes (1991), when talking to end users, the IT professional should always bear in mind their concerns, problems, environment, and responsibilities in terms of opportunities for IT services and support.

Furthermore, he says, continuous contact with end users gives IT the opportunity to gain more insight into their problems.

- They are sensitive to change: Because of the social nature of relationships, any form of change initiated on either the IT or the end-user side may disturb a relationship. It is argued that any kind of change having an effect on any of the elements of both the physical and abstract dimensions of a relationship will in fact disturb the relationship because of its holistic nature, which will be described later.
- They have a knowledge base: The complex world of perceptions, attitudes and approaches towards developing software products by IT professionals for the end user forces us to a point where it can be said that in order to overcome the most serious problems during this communication process in a relationship, a knowledge base of some kind is required before entering a relationship.
- They have a supportive culture: In order for a relationship to be sound, continuous support and mutual understanding, amongst other things, need to be elements of such a relationship. According to Pheysey, a support-oriented organization has the elements of mutuality, belonging, and connection. Furthermore, an appreciative form of control should be applied, which means: “management is seen to be a process focused on maintaining balance in a field of relationships” (Pheysey, 1993).

- A co-operative behaviour pattern is followed by the participants: Cooperation is not a fixed pattern of behaviour, but a changing, adaptive process directed to future results. The representation and understanding of intent by every party is therefore essential to co-operation, and as such emphasizes the importance of communication during co-operation (Clarke & Smyth, 1993). Co-operation can also create new motives, attitudes, values and capabilities in the co-operating parties, which will help create and maintain a supportive culture.
- They have a holistic nature: The important elements making up a relationship between the IT department and its end users at a given time should be organized together as a whole. If any of these elements are disturbed in a negative sense, the whole relationship between the IT department and its end users is undermined. In other words, the relationship as a whole is more than the sum of its elements and therefore one can say that it has a holistic nature.
- Sustainability: A most obvious characteristic of the abstract dimension is its sustainability over a period of time. In this regard time refers to the life span of an IT/end-user relationship. One can therefore argue that out of an information systems viewpoint, a relationship of this kind will only last until the product or service reaches the end of its life cycle. In this regard, Introna (1994) states: "Structures as relationships are contingent, it appears and disappears. It could be brief (a few seconds) or long lasting (several years)."
- Commitment: Kinlaw (1989) states that one of the primary tasks of a manager is to create commitment and focus in employees. He furthermore states that managers who help employees increase their knowledge, skill and experience also are building employee commitment. In this regard it is important that managers should take note of the four sturdy supports of commitment, namely: (a) clarity of goals and values; (b) employee competencies that ensure success; (c) the degree of influence that employees have; and (d) the expressed appreciation given to employees for their contributions. Commitment should be seen as a solid block that rests on these four sturdy supports or legs (Kinlaw, 1989). Commitment has been defined by Newman and Sabherwal (1996) as a state of mind that holds people and organizations in the line of behaviour. It encompasses psychological forces that bind an individual to an action. Commitment has been argued to greatly affect the persistence of behaviour (Salancik, 1977, cited in Newman & Sabherwal, 1996).

Figure 2. Role and impact of IT/end-user relationships in the alignment of business and IT departments (based on the work of Henderson & Venkatraman, 1992)



All the elements described previously form important sub-dimensions of the physical and abstract dimensions. Each of these elements plays a specific social role in an IT/end-user relationship environment, which impacts on the soundness of such a relationship as well as the success of alignment between IT and the business. The way in which the application of this paradigm could enhance alignment is addressed in the following paragraphs.

ALIGNMENT MODEL FOR APPLYING THE IT/END-USER RELATIONSHIP PARADIGM

The theoretical construct of strategic alignment (Henderson & Venkatraman, 1992) indicates that in terms of alignment there are two distinct linkages, namely a strategic fit and functional integration. According to the model, strategic fit is the vertical linkage concerned with the integration of the external environment in which the firm competes (e.g., partnerships and alliances) and the internal environment, which focuses on administrative structure (e.g., human resources and product development). Functional integration, according to the model, is the corresponding horizontal link between business and IT. These two linkages are used to determine the relationships between IT and business (Papp, 2001).

It is clear that the paradigm of IT/end-user relationships, which is based on two dimensions, namely the physical and abstract dimensions (as described earlier), addresses the two lower domains indicated by the dotted rectangle in Figure 2. In other words, the paradigm enhances alignment in terms of organizational infrastructure and processes, and IT infrastructure and processes. The physical dimension addresses structures, skills, and processes whilst the abstract dimension addresses all the soft issues required to ensure that relationships prevail. Therefore, it is argued that if the paradigm of IT/end-user relationships is applied when service and support activities1 are performed by IT professionals, it will enhance the functional integration between IT and the business. This is the case because all the elements of the physical and abstract dimensions are of a sound nature, which directly impacts on structures, processes and skills in the infrastructure domains of the alignment model in Figure 2.

FUTURE TRENDS

Theories in terms of how relationships between an IT department and its end users could be managed are scarce. Those who do address issues in this regard (Beard & Peterson, 1988; CSC research foundation, 1994; Wike et al., 1984) do not look into soft issues, nor do they give substance to the contents of such relationships. Furthermore, none of these theories deal with the important issue of how to understand and manage the soft issues involved during the establishment and maintenance of sound IT/end-user relationships.

Managing relationships and alignment is especially critical in the outsourcing of information technology services or extending of business applications to systems in other enterprises in order to form extended enterprises. It is clear that in the future much more research needs to go into the management of relationships and service level agreements to ensure that sound relationships and alignment will be maintained.

CONCLUSION

In this article the paradigm of IT/end-user relationships was defined in terms of its physical and abstract dimensions. It was argued that these two dimensions enable one to fully describe the holistic nature of such relationships. Furthermore, in terms of business IT alignment, it was argued that the paradigm applies in the infrastructure domains (both IT and the business) and will therefore enhance alignment in terms of functional integration.

REFERENCES

Boar, B.H. (1994). *Practical steps for aligning information technology with business strategies: How to achieve a competitive advantage*. New York: John Wiley & Sons, Inc.

Bommer, M., Gratto, C., Gravander, J., & Tuttle, M. (1991). A behaviour model of ethical and unethical decision making. In R. Dejoie, G. Fowler & D. Paradice (Eds.), *Ethical issues in information systems*. Boyd & Fraser.

CIO. (1997). *Relationship matter*.

Clarke, A.A., & Smyth, M.G.G. (1993). A co-operative computer based on the principles of human co-operation. *International Journal of Man-machine studies*, *38*, 3-22.

Dahlbom, B., & Mathiassen, L. (1993). Computers in context; the philosophy and practice of systems design. Cambridge, UK: Blackwell Publishers.

DeSanctis, G., & Poole, M.S. (1994). Capturing the complexity in advanced technology use: Adaptive structuration theory. *Organization Science*, *5*(2).

Du Plooy, N.F. (1995). *Overcoming the culture gap between management and IT staff*. Paper presented at Conference on HR Management of IT staff, IEC, Jan Smuts.

Jackson, I.F. (1986). *Corporate information management.* London: Prentice-Hall.

Henderson, J.C., & Venkatraman, N. (1992). Strategic alignment: A model for organizational transformation through information technology. In Transforming Organizations. Kochan TA, Useem M. Oxford University Press. New York.

Introna, L.D. (1994). *Giddens, emergence and social intervention.* Paper presented at the International Conference on Systems Thinking and Progressive Social Change, University of Cape Town, South Africa.

Kinlaw, D.C. (1989). *Coaching for commitment: Managerial strategies for obtaining superior performance.* University Associates, Inc.,USA

Leonard, A.C. (2002). A conceptual framework for managing relationships between all participants during IT service and support activities. *South African Journal of Industrial Engineering*, 81-96.

Moad, J. (1994). Does your CEO get it? *Datamation.*

Newman, M., & Sabherwal, R. (1996, March). Determinants of commitment to information systems development: A longitudinal investigation. *MIS Quarterly*, 23-54.

Papp, R. (2001). Introduction to strategic alignment. In R. Papp (Ed.), *Strategic information technology: Opportunities for competitive advantage.* Hershey, PA: Idea Group Publishing.

Pheysey, D.C. (1993). *Organizational cultures.* New York: Routledge.

Reich, B.H., & Benbasit, I. (1999). *Factors that influence the social dimension of alignment between business and information technology objectives.* Society of Information Management (SIM) and the Management Information Systems Research Center (MISRC).

Stokes, S.L., Jr. (1991). A marketing orientation for end-user computing support. In Umbaugh R E, 1991: Handbook of IS Management (Third Edition), Auerbach Publishers, Boston and New York, 125-134.

Umbaugh, R.E. (1991). Handbook of IS management (3rd ed.). Boston: Auerbach Publishers.

KEY TERMS

Abstract Dimension: The abstract dimension describes the soft issues of such a relationship.

Commitment: A state of mind that holds people and organizations in the line of behaviour. It encompasses psychological forces that bind an individual to an action.

Culture Gap: It should be seen as a gap of misunderstanding in the sense of two different organizational cultures that coexist in most organizations. The two cultures under discussion here are the culture of the IT profession and the culture of the rest of the organization.

Holistic Nature of an IT/End-User Relationship: The important elements making up a relationship between an IT professional and its end user(s) at a given time should be organized together as a whole. If any of these elements are disturbed in a negative sense, the whole relationship between the IT professional the end user(s) is undermined. In other words, the relationship as a whole is more than the sum of its elements.

IT/End-User Relationship: A relationship between IT and the end user consists of two dimensions, namely a physical dimension and an abstract dimension. The physical dimension describes those elements that are necessary in order to enable contact between IT and its end users, whereas the abstract dimension describes the soft issues of a relationship.

Physical Dimension: The physical dimension describes those elements that are necessary in order to enable contact between IT professional and its end users.

Social Dimension of an IT/End-User Relationship: Refers to all the elements in the abstract dimensions. Each of these elements plays a specific social role in an IT/end-user relationship environment, which impacts on the soundness of such a relationship as well as the success of alignment between IT and the business.

ENDNOTE

1 Normal system development activities or any other types of support IT can give its end users.

The Organizational Context in the Use of a Workflow System

Anabela Sarmento
ISCAP/IPP, Portugal and University of Minho, Portugal

INTRODUCTION

Organizations try to implement technological solutions that will improve the handling of relevant information in order to be able to respond more effectively to the opportunities and challenges presented by the environment, as they are aware of the importance of learning and knowledge management. Among the solutions offered to achieve this improvement are workflow systems (WS). The adoption of a new technology always means change but the implementation of the same information system (IS), by different organizations, does not usually result in the same changes. This implies that there are some factors that mediate the relationship between the adoption of an IS and the organization where it is implemented.

This article begins with a brief description of WS, their major characteristics and potentialities. Secondly, it describes the main organizational factors that might mediate the relationship between the technology and the organization where it will be implemented. This is followed by the description of a case study conducted with the intention of understanding, in depth, the influence of the organizational factors identified. Finally, we draw some conclusions and point up some future research trends.

BACKGROUND

Workflow Systems

WS are defined as proactive systems that manage the flow of work, during which documents, information or tasks are passed from one participant (persons, other systems, departments or even enterprises) to another, according to predefined procedures, which constitute the tasks (Plesums, 2002; WfMC, 2003). The main focus of these systems is the way the work usually evolves, that is, the *process* and not the specific information that may be contained in it (or in the documents used as a support that circulate in the process).

There are several types of WS. The categories most commonly used are: i) Production WS (these support rigid predefined rules and are used in critical processes where nothing should be allowed to fail); ii) Administrative WS (mostly based in electronic mail systems, characterized by the routing of reports, adequate for routine and low value processes), and, finally, iii) Collaborative WS (these focus mostly on the sharing of information and therefore allow joint working, used in areas with more creative needs).

Companies are especially interested in these systems because they enable changes in several organizational domains, namely: I) Economic, ii) Process Management and iii) Knowledge and Organizational Learning (see Table 1) (Hales, 1997; List, Schiefer & Bruckner, 2001; Orlikowski, 1996; Plesums, 2002;Sarmento, 2002).

These changes, although presented in isolation, should be seen as interconnected. For instance, the automatic routing of forms and documents (coordination) helps reduce task completion time (productivity). Also, the standardization of procedures and the capture and the performance of tasks according to predefined rules (coordination) contribute to the reduction of the number of mistakes (productivity), since all possible situations are analyzed and allowed for in the system.

Organizational Factors

The adoption of a new technology is usually mediated and conditioned by several organizational factors, namely technological, structural, social and individual, political and cultural (Bertrand & Guillemet, 1988; Bolman & Deal, 1997). These factors are summarized in Table 2.

The adoption of a new technology is a continuous process, as not all organizational factors have the same influence at the same time. Furthermore, all the factors mediating the adoption of a technology interact with each other, influencing the effects of the adoption and use of new IT. Thus, the complex relationships between these factors must be analyzed in a dynamic way. A new technology changes the organization (its structure and/or the individual roles and tasks) through the new technological characteristics it conveys and helps to establish new power relations. However, these changes themselves are also conditioned by the characteristics of the existing structure and the individuals who will regulate the choice of the technology.

Table 1. Synthesis of the most important changes due to WS, in the different organizational domains

Domains		**Changes**
Economic	Productivity	• Reduction of task accomplishment time, number of mistakes, costs, paper volume • Better product and information quality
Process Management	Coordination / control	• Automatic routing • Division of work • Standardization of procedures • Performance of tasks according to predefined rules • Decentralization • Control improvement • Monitoring
	Collaboration	• Possibility of collaboration, amongst distant participants, in time and space • Possibility of working from home
	Communication	• Elimination of time and geographic constraints • Elimination of human barriers • Computer mediated communication
Knowledge and Organizational Learning	Knowledge	• Capture rules and procedures of processes and tasks • Gather, process and distribute information • Make quality information available • Contribute to the constitution of a partial organizational memory • Help update organizational memory easily

Table 2. Summary of the organizational factors able to influence the use of an IS

Technological factors	Characteristics of the technology to be adopted and the technology already existing in the organization
Structural factors	Organizational design, the complexity, the number of hierarchical levels, the number of departments, the centralization or decentralization of power and decision making, the coordination of tasks, the formalization of procedures, the design of tasks and jobs and the degree of specialization
Social and individual factors	Multidisciplinary work teams, their distribution in time and space, their education, training, work satisfaction, skills and individual characteristics
Political factors	Who decides on the kind of technology to adopt, its design and implementation, who is going to use it, its purposes and objectives
Cultural factors	Culture, norms, rules and the reaction to change, knowledge and organizational learning capacity

LESSONS FROM A CASE STUDY

We have conducted a case study with the aim to identify changes that occurred due to the adoption of a workflow system and the organizational factors that enabled and constrained such changes. Results are summarized in Table 3.

As already stated in the introduction to this article, organizational changes that occurred due to the adoption of a workflow system are influenced either by the characteristics of the organization itself and the nature of the technology adopted, or by the relationship that is established between all the factors involved that can act as constrainers or enablers. The following paragraphs describe and discuss the role of the organizational factors that acted either as constrainers or enablers of changes.

Constrainers

The use of the WS in this organization was not implemented entirely without problems. For instance, there was an employee that had difficulties in using the system, as her computer crashed constantly due to system overload. Besides, this person also had some difficulties in understanding the functioning of the system at the beginning. Especially problematic was the system requirement of formally finishing any purchase process by expressing satisfaction. This step, although relatively simple, was, in fact, creating some difficulties, and was seen as unnecessary.

The educational background of the employees, as well as the state of their previous knowledge of how to operate computers may also have constrained the use of the

Table 3. Organizational impact due to the adoption of a workflow system

Domains		**Changes**
Economic	Productivity	• Some users consider that the system helped to shorten the process-accomplishment time; others consider that the time of the process increased (tasks are not the same for everybody; we can have different perspectives in the process). There was also a reduction in the paper volume circulating.
Process Management	Process Management (general)	• No changes in the process design. The new system was applied over the existing flow of work. However, some tasks performed before were eliminated (were seen as redundant). Tasks are performed from just one place. The support document used became electronic (instead of paper)
	Coordination/ Control	• Employees can check the status of the orders placed electronically. Forms became standardized. The number of exceptions was reduced.
	Communication and Collaboration	• Computer now mediates communication. Employees do not have to move physically from one department to another to deliver papers. The number and the frequency of informal contacts were reduced.
Knowledge and Organizational Learning	Knowledge	• The system helped to build an organizational memory easy to update and always updated. Information gathered in the system is available to all the staff members.

system. The members of staff who experienced more difficulties in adapting to the new system were also those with fewer schooling years. We cannot say for sure that their difficulties were due to lack of a college education background, as the data available do not allow for such extrapolation. However, the fact that those with a college degree had fewer problems in using the WS does corroborate this hypothesis, which in turn seems to point to the possibility that a college education may help develop some competencies and skills that are easily transferred whenever changes and obstacles have to be dealt with.

Finally, some characteristics of tasks may constrain the use of the system. In this case, the process where the WS was implemented was rather simple, not a crucial one and not performed regularly. Tasks were done sporadically, and this fact hinders the acquisition of a work routine. Thus, the obstacle is not the actual filling in of the document – a very straightforward procedure – but remembering all the steps needed to perform the task.

Enablers

In spite of the difficulties described previously, there were some factors that seemed to facilitate the use of the system. One of these factors is the educational background of the staff members. Most of the employees have a college degree. We would like to suggest that their educational background seems to allow them to develop some skills and competencies that helped to adapt and respond positively to the new system.

Individual personalities may also play a role in the successful adaptation to the use of the system. Although there were employees who experienced difficulties in using the system, at least one of them tried to overcome those difficulties and expressed interest in knowing and in exploring the system, without being told to do so.

Another aspect that seemed to enable the use of this system was the fact that the WS adopted runs on a Web-based platform. To use it, employees only need to know how to navigate on the Internet.

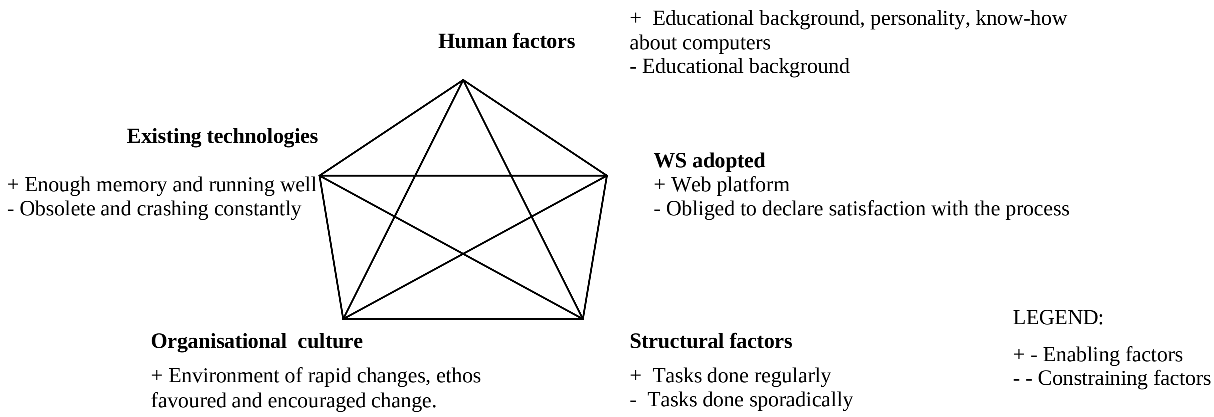
It was mentioned previously that the sporadic nature of the tasks does not help the acceptance of the system. Conversely, regular tasks enable the use of the system. So, those employees who used the system regularly became experts in the use of the WS. Interestingly, these members were also those who experienced more difficulties at the beginning of the process.

Finally, one should refer to the culture. This enterprise works in a market where changes are rapid, which means constant updating in order to remain competitive. So, changes are not unusual, but rather inevitable. And this situation might have helped the organization (and the employees involved) to better accept the new system and the changes that occurred.

Figure 1 summarizes the findings described previously.

The factors described earlier are not isolated, but interdependent. We also cannot say that one is more important than the other, as they interact constantly and simultaneously. For example, although some of the existing technologies were crashing, constantly preventing the employee from performing her job adequately, she was curious and committed enough to overcome her difficulties, even though her educational background is not as good as that of some of her colleagues. On the other hand, even those with a degree experienced some difficulties in

Figure 1. Summary of the organizational factors that enabled or constrained the use of WS



performing their tasks and learning how to use the application, probably because these tasks were infrequent. Moreover, the ethos of the company favored and encouraged change and a willingness to adopt the new system. Any expressed dissatisfaction was counteracted by the unanimous recognition that the new WS did reduce paper volume, a welcome change.

FUTURE TRENDS

The adoption of a WS in an organization should take into consideration not only the virtues and potentialities of the new system, but also consider the organizational reality of where it is going to be implemented. These characteristics of the organization are precisely what will mediate and condition those changes that might occur. Only a complete and profound knowledge of potential effects and organizational factors involved will allow a correct vision of changes caused by these systems. Furthermore, such knowledge will no doubt contribute to a better management of the process of change.

In this article we have only described a case study of an organization with particular characteristics. Also, the WS adopted was an administrative one, belonging to a specific producer. We recommend that other case studies, with different organizations and categories of WS, should be done in order to identify all the constrainer and enabler factors.

Management and application of knowledge are essential elements in the innovation process. The organizational knowledge domain is particularly important when considering the adoption and implementation of a WS. In this context, it is necessary to examine the flow of work through groups and organizations and explore how knowledge is or could be captured, managed and shared. Moreover, with the advent and expansion of e-business, another important area of research is the exploitation and use of workflow systems in the automation of business processes via the Internet. New implementation strategies are already emerging that offer e-commerce trading partners the opportunity to communicate and share information directly.

CONCLUSION

This article briefly explored workflow systems and described the application and use of such systems.

As these systems have an impact on the organization in the domains discussed earlier, we believe that future research will include the following: (1) the development of applications to be used in the different types of processes (production, administrative and collaboration/ad hoc) and (2) the exploitation of the different ways in which these systems might contribute to a better management of organizational knowledge. Some of the organizations that are concerned with these subjects and are exploring, developing and discussing issues related to workflow systems are WARIA - Workflow and Reengineering International Association (http://www.waria.com/), WfMC - the Workflow Management Coalition, (*http://www.wfmc.org/*) and the AIIM – Enterprise Content Management Association (*http://www.aiim.org/index2.asp*).

REFERENCES

Bertrand, Y., & Guillemet, P. (1988). *Organizações: Uma abordagem sistémica*. Lisboa: Instituto Piaget.

Bolman, L., & Deal, T. (1997). *Reframing organizations: Artistry, choice and leadership*. San Francisco: Jossey-Bass Publishers.

Hales, K. (1997). Workflow in context. In P. Lawrence (Ed.), *Workflow handbook 1997* (pp. 27-32). John Wiley & Sons.

List, B., Schiefer, J., & Bruckner, R. (2001). Measuring knowledge with workflow management systems. Second International Workshop on Theory and Application of Knowledge Management (TAKMA 2001). In *12th International Workshop on Database and Expert Systems Applications (DEXA'01)* (pp. 467-471). Munich, Germany: IEEE Computer Society Press. Retrieved September 25, 2001, from *http://www.ifs.tuwien.ac.at/~bruckner/pubs/takma2001_knowledge.pdf*

Orlikowski, W. (1992). *Learning from notes: Organizational issues in groupware implementation*. Working Paper 3428-92. Technical Report 134. Massachusetts, EUA: MIT Sloan School.

Orlikowski, W. (1996). Improvising organizational transformation over time: A situated change perspective. *Information Systems Research, 7*(1), 63-92.

Plesums, C. (2002). Introduction to workflow. In L. Fischer (Ed.), *Workflow handbook 2002* (pp. 19-38). Retrieved September 25, 2002, from *http://www.wfmc.org/information/introduction_to_workflow02.pdf*

Sarmento, A. (2002). *The organizational impact of collaborative systems – Case studies of the adoption and use of workflow systems (Impacto dos sistemas colaborativos nas organizações - Estudos de caso de adopção e utilização de sistemas workflow)*. PhD Thesis. Braga: University of Minho (Portuguese language).

WfMC. (2003). Introduction to the workflow management coalition. Retrieved September 25, 2003, from *http://www.wfmc.org/about.htm*

KEY TERMS

BPR – Business Process Reengineering: Analysis and redesign of workflow within and between enterprises.

Case Study: A detailed analysis of a person or group from a social, psychological or medical point of view. A careful study of some social unit (a corporation or a division within a corporation) that attempts to determine what factors led to its success or failure.

Content: Corporate knowledge stored in any form (paper, data, reports, correspondence, e-mail, multimedia, etc.).

ECMS – Enterprise Content Management System: It can be defined as an integrated approach to managing documents, Web content and digital assets. It combines the capabilities of enterprise document management system (EDMS) and content management systems with the ability to manage the full content life cycle across a growing assortment of content types.

EDMS – Enterprise Document Management System: Programs, procedures and/or software that manage, control and provide access to electronic documents.

Knowledge: Knowledge can be perceived as a discrete, objective, largely cognitive entity, susceptible of being classified as tacit and explicit, or it can be considered as a process or practice of knowing, being socially constructed and embedded in practice.

Knowledge Management: An organization that considers knowledge as a discrete entity will develop knowledge stores (repositories) and will try to capture the organization's knowledge by software. An organization that puts in perspective knowledge as something social will consider that knowledge resides in the minds of employees and cannot be captured. And thus, managing knowledge means managing people and the interactions among them.

Workflow: Term used to describe the tasks, procedural steps, organizations or people involved, required input and output information, and tools needed for each step in a business process. The workflow management focuses on processes rather than documents.

The Past, Present, and Future of End-User Performance

I. M. Jawahar
Illinois State University, USA

INTRODUCTION

Over the last decade, end-user computing has become an integral part of the organizational landscape. The emergence of end-user computing can be attributed to the necessity to manage and to effectively use information to function in a knowledge-based economy. Because of the increased organizational computing needs, computer literacy requirements have skyrocketed for clerical and support staff and for many middle and senior management positions (Bowman, Grupe, & Simkin, 1995). The proliferation of microcomputers and the availability of sophisticated user application tools (Shayo, Guthrie, & Igbaria, 1999) have facilitated the widespread implementation of end-user computing technology.

End-user computing has the potential to enhance productivity. However, for this potential to be realized, end users must learn EUC skills and perform at high levels. Given the significance of end-user performance to organizations, literally hundreds of studies have examined factors with potential to influence end-user performance. The purpose of this review is threefold: to review what we know about end-user performance, discuss some of the limitations of research on end-user performance, and to offer suggestions for future research on end-user performance.

BACKGROUND: FACTORS KNOWN TO INFLUENCE END-USER PERFORMANCE

Research in the areas of psychology and organizational behavior has clearly established that characteristics of the individual and the environment influence behavior and performance (e.g., Terborg, 1981). This approach, labeled the interactional approach, is an effective framework to review the factors that influence end-user performance (Jawahar, 2002).

Characteristics of End Users

To be sure, research has examined a variety of individual difference factors with potential to influence end-user performance. For instance, previous research has investigated the influence of attitudes (e.g., Jawahar & Elango, 1998; Kernan & Howard, 1990; Szajna, 1994), aptitudes (e.g., Evans & Simkin, 1989), learning styles (e.g., Bohlen & Ferratt, 1997), cognitive styles (e.g., Davis & Davis, 1990), self-efficacy (e.g., Jawahar & Elango, 2001), goal setting (e.g., Jawahar & Elango, 2001), experience (e.g., Dambrot, Silling, & Zook, 1988), education (e.g., Davis & Davis), age (e.g., Czara, Hammond, Blascovich, & Swede, 1989), and sex (e.g., Harrison & Rainer, 1992) on end-user performance. Organizations and managers can influence some of these individual difference factors, such as attitudes, aspiration or goals, and self-efficacy, more than other factors (e.g., aptitudes, learning/cognitive styles, and demographics). Therefore, this review focuses on the former set of factors.

Attitudes. The preponderance of research on end-user performance has focused on attitudes toward computers to predict end-user performance. However, these studies have generally reported inconsistent results. About one half of the studies that examined the relationship between attitudes and end-user performance have reported a relationship. While some of these studies reported a positive relationship (e.g., Nickell & Pinto, 1986), others have reported a negative relationship (e.g., Hayek & Stephens, 1989). Alternatively, roughly one half of the studies failed to find a relationship between attitudes and end-user performance (Kernan & Howard, 1990; Szajna, 1994). After reviewing these studies, Jawahar and Elango (1998) attributed the inconsistent results to the fact that many of these studies had *incorrectly* used the constructs of computer anxiety and negative attitudes toward computers interchangeably (see Kernan & Howard) and had relied on global attitudes to predict end-user performance.

Drawing on Ajzen and Fishbein's (1980) behavioral intentions model, Jawahar and Elango (1998) theorized that behaviors or outcomes could be best predicted by attitudes that specifically relate to those behaviors than by more global and general attitudes. They proposed that attitude toward working with computers is much more specific and relevant to performance of tasks which require the use of computer skills than the more general attitudes toward computers. Individuals who hold favorable attitudes toward working with computers are more likely to practice and learn end-user computing skills, and evidence higher levels of performance on tasks that require the use of those skills than those who hold less favorable attitudes. As expected, Jawahar and Elango found that attitudes toward working with computers but not attitudes toward computers explained unique variance in end-user performance. These results were replicated in a follow-up study (Jawahar & Elango, 2001). These two studies together with previous research indicate that attitudes that are specific to the task of working with a computer or a particular software package or packages are more likely to be predictive of end-user performance.

Goals. The positive effect of goal setting on task performance is one of the most robust and replicable findings in the psychological literature (Locke & Latham, 1990; Locke, Shaw, Saari, & Latham, 1981). Literally, hundreds of studies have been conducted on goal setting in a variety of settings and with a wide range of subjects including managers, engineers, and scientists (Locke & Latham). Research on goal setting has documented that specific and difficult or challenging goals lead to higher levels of performance than the absence of goals, easy goals, or "do your best" goals (Locke et al.). Locke and Latham have shown that goal setting, when combined with feedback or knowledge of results, leads to high levels of performance. Thus, goal setting is most likely to improve task performance when the goals are specific and sufficiently challenging, and feedback is provided to show progress in relation to the goal. In a series of two studies, Jawahar and Elango (Jawahar, 2002; Jawahar & Elango, 2001) found that end users' goals to learn and master a software package is in fact strongly related to their performance with the software package.

Self-Efficacy. Self-efficacy is the belief in one's ability to effectively complete a task or exhibit a specific behavior (Bandura, 1982). Theory and research on self-efficacy suggests that, in contrast to individuals with low levels of self-efficacy, the highly efficacious are less apprehensive of change, set more challenging goals, exert more effort, persist in the face of difficulty, and achieve higher levels of performance (Jawahar, Stone, & Cooper, 1992). Prior research has also documented that self-efficacy influences diverse behaviors and performance on various tasks including tasks involving end-user computing (e.g., Gist, Schwoerer, & Rosen, 1989; Jawahar, 2002; Jawahar & Elango, 2001; Stajkovic & Luthans, 1998). For instance, in one study, Gist et al. studied managers and administrators undergoing two types of training in the use of computer software. Trainees with higher self-efficacy prior to training performed better than their low self-efficacy peers on a timed computer task at the end of training. In another study, Jawahar and Elango reported that self-efficacy explained more unique variance in end-user performance than attitudes and goals or aspirations of end users. Enhancing self-efficacy of end users might very well be the single most effective approach to enhancing end-user performance.

Characteristics of the Environment

In contrast to the voluminous research on end-user characteristics, very little research has focused on characteristics of the environment. Even so, several different characteristics have been studied. In deciding which characteristics to include in the review, I relied on two criteria: the quality of research support for the characteristics and/or the relative ease with which organizations and managers can change the characteristic.

End-User Training. Several scholars have acknowledged end-user training as an essential contributor to the productive use of computer systems in organizations (e.g., Compeau & Higgins, 1995). The practitioner literature also supports the view that training is essential for effective use of computer technology (Finley, 1996). Because training can affect the success or failure of end-user computing in organizations (Bostrom, Olfman, & Sein, 1990; Rivard & Huff, 1988), training employees to use information technology productively has become a high priority in many organizations (Aggarwal, 1998). It is now well established that training end users enhances their performance (e.g., Gist et al., 1989). The next step, then, is to identify characteristics of training that facilitate end-user learning and performance.

Opportunity to Practice. Opportunity to practice in the training environment and in the posttraining environment is likely to affect how well trainees learn and use their newly acquired skills to perform their jobs. While the importance of practice for learning new skills is widely acknowledged (Tannenbaum & Yukl, 1992), training and work environments differ in the extent to which they provide trainees opportunities to practice newly acquired skills. For instance, Ford, Quinones, Sego, and Speer (1991) studied Air Force technical trainees after they completed training and found significant differences in opportunity to apply the training; they also noted wide variations in the lengths of time before trainees first performed the tasks for which they had been trained. Opportunities to practice newly acquired skills are likely to strengthen learning and influence how well trainees use those skills. In a study of IRS managers, Pentland (1989) found that attempts to practice trained computer skills immediately upon returning to the job had a major impact on long-term retention. These studies suggest that the opportunity to practice trained skills will significantly influence how well trainees learn and perform with those skills.

Time Constraint. In one study, Peters, O'Connor, and Rudolf (1980) used the "critical incidents" method to identify environmental constraints relevant to performance outcomes. In that study, 62 participants employed in a wide variety of jobs identified situational variables that adversely affected performance. Peters et al. identified time availability as a potent constraint. Trainees are likely to become frustrated and not learn as well when they are expected to learn and become proficient in new skills in a short amount of time. The time allotted to learn and become proficient in new skills is likely to affect how well trainees learn and use their skills. Consequently, trainees who perceive time constraints are likely to evidence lower levels of performance relative to those who do not perceive such time constraints. In a recent study, Jawahar (2002) reported that the perception of time constraint had a negative effect on end-user performance.

SHORTCOMINGS OF RESEARCH ON END-USER PERFORMANCE

The use of user satisfaction as a surrogate of end-user computing success and the lack of systematic research on characteristics in the end users' environment are two of the most obvious limitations. Because these two limitations are likely to limit the usefulness of research findings for researchers and practitioners, they are discussed next.

Surrogate Measures of End-User Performance. To date, user satisfaction has served as the most popular measure in the literature for measuring end-user computing success (Mahmood, Burn, Gemoets, & Jacquez, 2000; Munro, Huff, Marcolin, & Compeau, 1997). Use of user satisfaction as the primary measure of end-user computing success is based on the implicit assumption that satisfied users perform better than dissatisfied users (Amoroso, 1992). Unfortunately, evidence to support this assumption is lacking. Equating user satisfaction with end-user computing success is problematic because it does not tell us anything about productivity. Munro et al. aptly noted that a better measure of end user computing success than user satisfaction is necessary to justify the substantial investments in end-user technologies and in end-user training. Given that the primary reason organizations computerize their operations is to improve productivity, competitiveness, and profits, unless end users learn the skills and utilize those skills to improve their job performance, the expected benefits of end-user computing are unlikely to accrue. While perceptions of end-user satisfaction is an important measure, it is not equivalent to end-user performance, and therefore, future research should not use end-user satisfaction as a surrogate measure for end-user performance.

Characteristics of End-Users' Environment. Characteristics in the end users' environment could directly, or indirectly through their effects on characteristics of the end user himself or herself, influence end-user performance. After all, the prediction that inhibiting situational conditions could adversely affect performance is straightforward and has been expressed in several models of performance (e.g., Peters & O'Connor, 1980; Terborg, 1981). Situational factors that restrict or constrain performance are referred to as situational constraints. These include factors beyond the control of individual employees (e.g., faulty equipment, lack of resources, time constraints) that restrict the range of individual performance. Research has reported the existence of situational constraints (e.g., Peters, O'Connor, & Rudolph, 1980), and the adverse effects of situational constraints on performance is well documented (e.g., Hatcher, Prus, Englehard, & Farmer, 1991).

However, prior research has relied almost exclusively on characteristics of end users to predict end-user performance even though several models suggest that both person and environmental factors influence behavior and performance. For instance, in one study, Jawahar (2002) found that the variance in end-user performance explained by characteristics of end users reduced from 32% to 23% when they were faced with situational constraints. These results highlight the importance of examining the influence of characteristics in end users' environment with potential to influence their performance.

FUTURE RESEARCH ON END-USER PERFORMANCE

First, as noted above, studies should investigate the influence of both characteristics of the end user and the environment on end-user performance. Characteristics of end users' environment that merit research attention include attributes of the training (e.g., user friendliness of the training itself), software packages used (e.g., user friendliness, tutorials provided), and availability of posttraining support. Second, most research on end-user training has focused on training provided by the organization. Most learning, though, occurs on the job and with help from peers. The extent to which peers play a role in an end user's performance needs to be investigated. Third, not all end users are the same. Twenty years ago, Rockart and Flannery (1983) provided a framework for classifying end users. Unfortunately, many previous studies do not adequately describe the participants, making it difficult to ascertain the population to which results might generalize. It is critical that studies adequately describe the participants with reference to

Rockart and Flannery's framework. Finally, research that is likely to be of most theoretical and practical significance is research focused on the construct of end-user performance itself. For instance, it is one thing to show that X, Y, and Z factors influence performance on end-user computing tasks in a training environment or in the (end user's) job environment. What is more important is to examine if an end user's performance on computing tasks in the job environment enhances his or her overall job performance. Demonstrating the latter is likely to serve as not only the best justification for investments in end-user computing technology, but is likely to ensure organizational competitiveness, which after all, is the most important and primary reason for the emergence and need for end-user computing.

CONCLUSION

The primary purpose of introducing end-user computing technology is to improve performance and productivity, and to obtain a sustainable competitive advantage. This article reviewed research on factors that are known to influence end-user performance, discussed two significant limitations of extant research, and pointed out fruitful areas for future research. It is hoped that future research will overcome the limitations identified and address the important concerns offered as suggestions for future research on end-user performance.

REFERENCES

Aggarwal, A. K. (1998). End user training: Revisited. *Journal of End User Computing, 10*(3), 32-33.

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall.

Amoroso, D. L. (1992). Using end user characteristics to facilitate effective management of end user computing. *Journal of End User Computing, 4*(4), 5-15.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*, 122-147.

Bohlen, G. R., & Ferratt, T. W. (1997). End user training: An experimental comparison of lecture versus computer-based training. *Journal of End User Computing, 9*(3), 14-27.

Bostrom, R. P., Olfman, L., & Sein, M. K. (1990). The importance of learning style in end user computing. *MIS Quarterly, 14*(1), 100-119.

Bowman, B. J., Grupe, F. H., & Simkin, M. G. (1995). Teaching end user applications with computer-based training: Theory and an empirical investigation. *Journal of End User Computing, 7*(2), 12-18.

Compeau, D. R., & Higgins, C. A. (1995). Application of social cognitive theory to training for computer skills. *Information Systems Research, 6*(2), 118-143.

Czara, S. J., Hammond, K., Blascovich, J. J., & Swede, H. (1989). Age related differences in learning to use a text-editing system. *Behavior and Information Technology, 8*, 309-319.

Dambrot, F. H., Silling, S. M., & Zook, A. (1988). Psychology of computer use: Sex differences in prediction of course grades in a computer language course. *Perceptual and Motor Skills, 66*, 627-636.

Davis, D. L., & Davis, D. F. (1990). The effect of training technique and personal characteristics on training end users of information systems. *Journal of Management Information Systems, 7*, 93-110.

Evans, G. E., & Simkin, M. G. (1989). What best predicts computer proficiency? *Communications of the ACM, 32*(1), 1322-1327.

Finley, M. (1996, January). What's your techno type—and why you should care? *Personnel Journal*, 107-109.

Ford, J. K., Quinones, M., Sego, & Speer, J. (1991). Factors affecting the opportunity to use trained skills on the job. Paper presented at the Annual Conference of the Society of Industrial and Organizational Psychology, St. Louis, MO.

Gist, M. E., Schwoerer, C. E., & Rosen, B. (1989). Effects of alternative training methods on self-efficacy and performance in computer software training. *Journal of Applied Psychology, 74*, 884-891.

Harrison, A. W., & Rainer, R. K. (1992). The influence of individual differences on skill in end user computing. *Journal of Management Information Systems, 9*(1), 93-111.

Hatcher, L., Prus, J. S., Englehard, B., & Farmer, T. M. (1991). A measure of academic situational constraints: Out-of-class circumstances that inhibit college student development. *Educational and Psychological Measurement, 51*, 953-962.

Hayek, L. M., & Stephens, L. (1989). Factors affecting computer anxiety in high school computer science students. *Journal of Computers in Mathematics and Science Teaching, 8*, 73-76.

Jawahar, I. M. (2002). The influence of dispositional factors and situational constraints on end user performance: A replication and extension. *Journal of End User Computing, 14*, 17-36.

Jawahar, I. M., & Elango, B. (1998). Predictors of performance in software training: Attitudes toward computers versus attitudes toward working with computers. *Psychological Reports, 83*, 227-233.

Jawahar, I. M., & Elango, B. (2001). The effects of attitudes, goal setting and self-efficacy on end user performance. *Journal of End User Computing, 13*(2), 40-45.

Jawahar, I. M., Stone, T. H., & Cooper, W. H. (1992). *Activating resources in organizations. In R. W. Woodman & W. A. Pasmore (Eds.),* Research in organizational change and development *(Vol. 6, pp. 153-196). JAI Press.*

Kernan, M. C., & Howard, G. S. (1990). Computer anxiety and computer attitudes: An investigation of construct and predictive validity issues. *Educational and Psychological Measurement, 50*, 681-690.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal-setting and task performance*. Englewoods, NJ: Prentice-Hall.

Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal-setting and task performance: 1969-1980. *Psychological Bulletin, 90*, 125-152.

Mahmood, A. A., Burn, J. M., Gemoets, L. A., & Jacquez, C. (2000). Variables affecting information technology end-user satisfaction: A meta-analysis of the empirical literature. *International Journal of Human-Computer Studies, 52*, 751-771.

Munro, M. C., Huff, S. L., Marcolin, B. L., & Compeau, D. R. (1997). Understanding and measuring user competence. *Information and Management, 33*, 45-57.

Nickell, G., & Pinto, J. (1986). The computer attitude scale. *Computers in Human Behavior, 12*, 301-306.

Pentland, B. T. (1989). The learning curve and the forgetting curve: The importance of time and timing in the implementation of technological innovations. *Paper presented at the Annual Academy of Management Meetings, Washington, DC.*

Peters, L. H., & O'Connor, E. J. (1980). Situational constraints and work outcomes: The influence of a frequently overlooked construct. *Academy of Management Review, 5*, 391-397.

Peters, L. H., O'Connor, E. J., & Rudolf, C. J. (1980). The behavioral and affective consequences of performance-relevant situational variables. *Organizational Behavior and Human Performance, 25*, 79-96.

Rivard, S., & Huff, S. L. (1988). Factors of success for end user computing. *Communications of the ACM, 31*(5), 552-561.

Rockart, J. F., & Flannery, L. S. (1983). The management of end user computing. *Communications of the ACM, 26*(10), 776-784.

Shayo, C., Guthrie, R., & Igbaria, M. (1999). Exploring the measurement of end user computing success. *Journal of End User Computing, 11*(1), 5-14.

Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin, 124*(2), 240-261.

Szajna, B. (1994). An investigation of the predictive validity of computer anxiety and computer attitude. *Educational and Psychological Measurement, 54*(4), 926-934.

Tannenbaum, S. I., & Yukl, G. (1992). Training and development in work organizations. *Annual Review of Psychology, 43*, 399-441.

Terborg, J. R. (1981). Interactional psychology and research on human behavior in organizations. *Academy of Management Review, 6*, 569-576.

KEY TERMS

End-User Performance: Performance of end users

End Users: Individuals who use software packages or computer systems that do not require programming

End-User Training: Teaching end users to learn and use organizational computing technology

Environmental Constraints: Factors in the environment with potential to inhibit or constrain one's behavior and/or performance

Interactional Approach: An approach that acknowledges that both characteristics of the personand the environment influence behaviors and outcomes.

Self-Efficacy: The belief in one's ability to successfully complete a task or execute a behavior

Sustainable Competitive Advantage: Achieving and maintaining an advantage over other competing firms

T

The Social Contract Revised

Robert Joseph Skovira
Robert Morris University, USA

INTRODUCTION

Corporate and personal behaviors express responsibility and obligation in doing business. Usually, accountable and obligated actions are explicitly given in legal contracts. The legalized versions that enforce responsibility and obligation are dependent upon moral expectations framed by a social contract.

Obligations and responsibilities influence and frame our everyday experiences. While some obligations and responsibilities are shaped into legal contracts, many of the obligations and responsibilities that govern our daily situations never see the light of day except in felt senses of duty and concerns under girding actions. Obligations and responsibilities are localized social and moral constructs arising from a communal sensibility of people in the ebb and flow of everydayness. Obligations are norms that structure human interaction and locate human relationships within a moral universe of discourse. Obligation is doing one's duty in the face of situational demands. Responsibility is owning up to the effects and consequences of actions in a social situation. Responsible behavior is local and personal.

Responsibility, also, is related to a constellation of supporting values, of which accountability and privacy are significant in their social reach. Accountability works best when private and confidential arenas of action are recognized. Being responsible carries an expectation of privacy. Other constellated expectations are access to information, freedom of speech, and information ownership, both physical and intellectual (Dhillon, 2002).

Obligations and responsibilities are social expectations framing relationships. These social relationships are based on a silent and intuitive communal agreement, the social contract. A social contract is implicit to any society and culture. Although the foundation for many written social and legal documents, the social contract is never written. It is a set of common understandings and assumptions framing situational behavior. A social group is an eventful result of a social contract: A social contract entails association. There are three great theorists on the social contract: Hobbes, Locke, and Rousseau.

BACKGROUND

Previous Ideas on the Social Contract

Hobbes

The social contract for Hobbes is a tacit affair that allows a communal association to form. Community is a natural restraint on humanity's natural perversity and propensity to selfishness and conflict. The social contract is the basis of societal authority, or the power of a society to sanction behavior. The social contract is a restraint on humanity's Darwinian-like tendency to reduce all interactions to power relationships and to coercion (Hobbes, 1988). The social contract confers the rights of citizenship and the duties—being responsive (responsible) to the now common or mutual interests (Halverson, 1972; Hobbes; Solomon, 1992, 1993).

Locke

Persons begin, for Locke, with the freedoms to act, to do what they will with their property, and to live as they see fit. All people are equal to one another. This is a natural state; it rests on natural law. The mutual agreement to join together as a community and political body rests on a social contract (Kendall, 1959; Locke, 1960). The social contract for Locke is based on natural law which argues that persons have certain rights. The notion of the social contract, according to Locke, is expressive of the values of freedom and fairness (Barker, 1970).

Rousseau

For Rousseau, each person naturally tends to self-preservation and maintenance of personal concerns. Socially agreed-upon norms are the basis of any authority within any group. The social contract is conventions of association; it is a solution to situations created when people join together to defend self and property (Rousseau, 1998). Individuals establish a society by agreeing to give up any natural rights for rights based on the collective. Individuals are dependent upon the community; the common will overrides individual will (Crocker, 1968).

Information Technology as a Frame for E-Social Contract

Information technology is a metaphor of hardware and software applications to informational situations and flows, that is, information systems, corporate or personal. Technology is an ideological stance that argues that all solvable problems are defined within technological space (Barbour, 1993; Postrel, 1998). Responsible living in today's world requires understanding technology (Monsma, Christians, Dykema, Leegwater, Schuurman, & van Poolen, 1986). Information technology enhances accountability or responsibility in business affairs (Galvin, 2000; Stoll, 1999).

Views of IT

There are several different views of information technology. These views frame our sense of the social contract. Information technology insinuates itself into our social existence and affects the nature of the social contract (Friedman, 2000). One approach sees information technology as a mere tool, a subsystem within its sociocultural environment. Societal members use information technology framed by the cultural sensibilities of the users. This view considers information technology to be value neutral. A second frame is that information technology is the culture which orients the person and the society. Information technology is the world in which we live, by which we live, and through which we live. Our mode of being in the world is technological; everything is technique. Information technology provides a new context for living. This frame changes the notion of the social contract. This paradigm is the foundation of the information society. A third sense, a hybrid, is that information technology contributes in major ways to how we understand and live in the world, but it is only one aspect of our experiences. In different ways, and to varying degrees, these perspectives frame how we are in the information society.

The Internet and E-Everydayness

The Internet and the World Wide Web are manifestations of information technology as cultural paradigm. Where there is a culture, there is a morality. Where there is a moral sense, there are obligations and responsibilities. The World Wide Web is a digital society that is intentionally created by its users. Users of the World Wide Web must be self-regulators according to accepted standards and rules (Berners-Lee, 1999; Chapman & Dhillon, 2002; Stoll, 1995). The Internet is viewed as a society of multiple local communities. These local communities encapsulate social and ethical values. The Internet is based on a social contract of sorts. The Internet expresses a normative architecture. It is a structure for responsible self-governance. The architecture is made up of free choice, free speech, honesty, and openness or disclosure (Dyson, 1997; Negroponte, 1996; Rheingold, 1993). The information highway, the Internet and World Wide Web, reflect a new social contract (Gates, 1995). The new social contract is digital and its major foci are about decentralization and globalization (Negroponte, 1995; Poster, 1995).

Information Society and E-Lebenswelt

The information society is a significant context for understanding social contract. The information society shows up in the Internet and the World Wide Web (Webster, 1995). The information society is the global connectivity of commercial interests, governmental venues, as well as individual and personal ways of being linked together globally (Friedman, 2000). The information society shapes and frames not only sociocultural environments, but also human thinking (Webster, 1995).

The information society reduces the physical social aspects of human interaction to nothing more than individuals and information flows (Borgmann, 1999). The information society is nothing but information and people. Ultimately, even people exist as information entities. Information is the focus of everything; it is everything. Because informational flows are the basis, society is not centralized; there is a diffusion of information (Brown & Dugwid, 2000). Worldwide connectivity is the information society. Contemporary life for citizens in the information society is digital behavior. Cultural behavior in the information society is the semantic behavior based on digital existence (Borsook, 2000). Information society is a cultural force, a shaper of obligations manipulating the social fabric; its foundation is the e-social contract.

MAIN THRUST OF CHAPTER

The E-Social Contract

The e-social contract is a digital affair. Cyberspace revises the social contract. The digital social contract resides in the expectations, assumptions, and constraints of online actions, such as conversations carried on in e-mails, chat rooms, and instant messaging services. The e-social contract resides in guest books and other forms of relationship building on the Web. The e-social contract informs behavior on the Internet; these are the tacit conventions of right or appropriate behavior at the right times in the right places to the right people. The digital social contract assumes a virtual trust of others encountered in cyberspace. This revision of the social contract in cyberspace as the e-social contract has at least three different faces to show the world, virtual though it may be. They are virtual communitarianism, virtual radical individualism, and virtual social capitalism.

Virtual Communitarianism

Virtual communities are communities of practice which are constituted by a set of expectations. These expectations are the values of reciprocity and trust (Blanchard & Horn, 2000; Stoll, 1995). Davenport (2002) argues that anonymity does not lead to responsible behavior. Responsibility maintains claims of rights and freedoms. The consequences of anonymity are inappropriate behaviors that take advantage of the situation. Responsible behavior is trustworthy because it is public or visible behavior. Public behavior is framed by the purpose of public space (or the virtual community or cyberspace) and what the public space has been created for (Erickson, Halverson, Kellog, Laff, & Wolf, 2002). Each virtual community generates its own set of sanctioned online behaviors and underlying policies which are expressions of the e-social contract. A virtual community has jurisdiction only over individuals who freely elect to join the group. Such normative prescriptions freely entered into have more moral force than civil governments. Communal norms or rules are tacit and implicit in the relationships that create the group and remain so until individuals abuse them or raise issues, which make the normative structures visible (Dyson, 1997). The Internet constructs a social contract for the empowering of individuals (Kinney, 1995).

Virtual Individualism

The new virtual individualism, or e-individualism, is a radical individualism and libertarian in its approach. It is a form of neo-Darwinism. This neo-Darwinism is digital in nature. Digital Darwinism is an approach that argues that the economic, biological, and Internet environments influence one another. This is a philosophy that is the ideology of technolibertarianism. The worldview of the individualist or the technolibertarian holds human nature to be a model of rational economic behavior that is always guided by enlightened self-interest. The individual is responsible or accountable for holding his or her way within the virtual marketplaces and locales in cyberspace. Digital existence is a matter of being able to compete, that is, to innovate and adapt to the situation. This worldview rejects an implicit social contract, electronic or otherwise, unless you count enlightened self-interest. Technological adaptation to its environment means pushing individual adaptation in societal and economic spheres. These spheres of influence are evolutionary. Evolutionary change is a consequence of technological change (Borsook, 2000; Schwartz, 1999; Kinney, 1995; Stewart & Williams, 2000).

Virtual Social Capital

Virtual social capital refers to an implicit willingness of individuals to work together in online settings. An important element that shows up is trust. Virtual social capital maintains digital communal behavior. A commonality of purpose is the stuff of virtual social capital. Commonalities of purpose show up in the interrelationships among virtual individuals on the Internet. Such interrelationships are social networks; they are cooperative relationships (Preece, 2002). The e-social contract is the structure of virtual social capital. Virtual social capital is the stuff of communal relationships in cyberspace. A sense of community implies that trust is an important element of social association. The unspoken and unwritten obligations of social living online show up in common trust and loyalty. Societal endeavors are grounded in shared beliefs, mores, and collaborative activities. The virtual community of practice (the basis for virtual social capital) demands adherence to its worldview. This is apparent even in cyberspace. Virtual social contract shows up in a common semantic universe and worldview. This worldview and its common senses are grounded in the e-social contract, which is a tacit set of reciprocal expectations and commitments among individuals in cyberspace. The e-social contract is the tacit set of presuppositions grounding appropriate virtual behaviors, commonly held beliefs, and expectations of benefits in the interactive settings of cyberspace (Cohen & Prusak, 2001; Dyson, 1997; Preece, 2002).

FUTURE TRENDS

As with any social setting, the tenuous and intuitive frame of the e-social contract will continue to bound and channel what is considered to be proper behavior in cyberspace. Where rifts and breaks happen in the delicate social fabric woven out of the e-social contract, the cybercitizens will strive to patch the tears in the online social cloth. This patching logically will be in the guise of generating formal codes of ethics and laws. Much of this has happened already, even from the beginnings of human interaction in cyberspace. Because of the dialectic between virtual individual actions and virtual communal sensibilities, need for articulated formal boundaries (and interpretations of the e-social contract) will continue.

CONCLUSION

This discussion has looked at the foundational contexts of behavioral expectations of travelers of the Internet and the World Wide Web. The journey in this essay began with a discussion of responsibility and obligation. These ideas are framed by the ideas of the social contract, information technology, and the information society. Information technology as a cultural frame of reference is discussed as a ground for the e-social contract. The essay sees the Internet, cyberspace, as an electronic *lebenswelt*.

REFERENCES

Baker, E. (1970). *Social contract: Essays by Locke, Hume, and Rousseau* (With an introduction by Sir E. Barker). London: Oxford University Press.

Barbour, J. (1993). *Ethics in an age of technology* (Vol. 2). New York: HarperCollins.

Berners-Lee, T. (1999). *Weaving the Web*. New York: HarperSanFransico/HarperCollins.

Blanchard, A., & Horn, T. (2000). Virtual communities and social capital. In G. D. Garson (Ed.), *Social dimensions of information technology* (pp. 6-21). Hershey, PA: Idea Group Publishing.

Borgmann, A. (1999). *Holding on to reality: The nature of information at the turn of the millennium*. Chicago: University of Chicago Press.

Borsook, P. (2000). *Cyberselfish: A critical romp through the terribly libertarian culture of high tech*. New York: PublicAffairs.

Brown, J. S., & Dugwid, P. (2000). *The social life of information*. Boston: Harvard Business School Press.

Chapman, S., & Dhillon, G. S. (2002). Privacy and the Internet: The case of DoubleClick, Inc. In G. S. Dhillon (Ed.), *Social responsibility in the information age: Issues and controversies* (pp. 75-88). Hershey, PA: Idea Group Publishing.

Cohen, D., & Prusak, L. (2001). *In good company: How social capital makes organizations work*. Boston: Harvard Business School Press.

Crocker, L. G. (1968). *Rousseau's social contract: An interpretative essay*. Cleveland, OH: The Press of Case Western Reserve University.

Davenport, D. (2002, April). Anonymity on the Internet: Why the price may be too high. *Communications of the ACM*, *45*(4), 33-35.

Dhillon, G. S. (2002). Understanding social responsibility issues in the information age. In G. S. Dhillon (Ed.), *Social responsibility in the information age: Issues and controversies* (pp. 1-11). Hershey, PA: Idea Group Publishing.

Dyson, E. (1997). *Release 2.0: A design for living in the digital society*. New York: Broadway Books.

Erickson, T., Halverson, C., Kellog, W. A., Laff, M., & Wolf, T. (2002, April). Social translucence: Designing social infrastructures that make collective activity visible. *Communications of the ACM*, *45*(4), 40-44.

Friedman, T. L. (2000). *The Lexus and the olive tree*. New York: Anchor Books/Random House.

Galvin, J. (2000, June). The new business ethics. *Smart business for the new economy*, *13*(6), 86-99.

Gates, B. (1995). *The road ahead*. New York: Penguin.

Halverson, W. H. (1972). *A concise introduction to philosophy*. New York: Random House.

Hobbes, T. (1988). *The leviathan*. Amherst, NY: Prometheus Books.

Kendall, W. (1959). *John Locke and the doctrine of majority-rule*. Urbana, IL: University of Illinois Press.

Kinney, J. (1995, September). Anarcho-emergentist-republicans. *Wired*, *3*(9), 90-95.

Locke, J. (1960). *Two treatises of government*. In W. Ebenstein (Ed.), *Great political thinkers: Plato to the present* (pp. 393-413). New York: Holt, Rinehart and Winston.

Monsma, S. V., Christians, C., Dykema, E. R. Leegwater, A., Schuurman, E., & van Poolen, L. (1986). *Responsible technology: A Christian perspective*. Grand Rapids, MI: William B. Eerdmans Publishing.

Negroponte, N. (1995). *Being digital*. New York: Alfred A. Knopf.

Negroponte, N. (1996, November). Being local. *Wired*, *4*(11), 286.

Poster, M. (1995, November). The Net as a public sphere? *Wired*, *3*(11), 135-138.

Postrel, V. (1998). *The future and its enemies: The growing conflict over creativity, enterprise, and progress*. New York: The Free Press.

Preece, J. (2002, April). Supporting community and building social capital. *Communications of the ACM*, *45*(4), 37-39.

Rheingold, H. (1993). *The virtual community*. New York: HarperCollins.

Rousseau, J.-J. (1998). *The social contract or principles of political right* (H. J. Tozer, Trans.). Hertfordshire, England: Wordsworth Editions Ltd.

Schwartz, E. I. (1999). *Digital darwinism*. New York: Broad Way Books/Random House.

Solomon, R. C. (1992). *Morality and the good life*. New York: McGraw-Hill.

Solomon, R. C. (1993). *Ethics and excellence*. Oxford, England/New York: Oxford University Press.

Stewart, J., & Williams, R. (2000). The co-evolution of society and multimedia technology: Issues in predicting the future innovation and use of ubiquitous technology. In G. D. Garson (Ed.), *Social dimensions of information technology* (pp. 46-62). Hershey, PA: Idea Group Publishing.

Stoll, C. (1995). *Silicon snake oil: Second thoughts on the information highway*. New York: Doubleday.

Stoll, C. (1999). *High-tech heretic*. New York: Doubleday.

Webster, F. (1995). *Theories of the information society*. London/New York: Routledge.

KEY TERMS

Digital Darwinism: An ideology framing economic and social situations that argues that only the economically and socially fit will survive because of their ability to adapt to the digital world.

E-Social Contract: Tacit set of conventions underlying the information society and any kind of digital interaction across the Internet and the World Wide Web.

Information Society: Communal interaction based on the global informational systems of the Internet and the World Wide Web. Social relationships are informational flows and people are informational entities or links.

Information Technology: Hardware and software that constitute informational systems, as well as techniques for handling informational content and flows.

Obligation: The moral construct or concept of duty framing intentional behavior and its effects in communal situations.

Responsibility: The moral construct or concept of taking ownership of acts and the consequences in social situations.

Social Contract: Tacit and assumed set of conventions about human social involvement that frames all communal associations.

Technolibertarianism: A companion political ideology to digital Darwinism that argues in an age of global connectivity for radical individualism and no governmental regulation of digital life.

The Software Industry in Egypt

Sherif Kamel
The American University in Cairo, Egypt

INTRODUCTION

During the 1960s computing was introduced to Egypt. Its use and applications were limited to the government and the public sector. During the 1980s the introduction and diffusion of computing was widespread following the global personal computer evolution. Personal computers effectively impacted organizational development and growth due to the continuous developments in the information technology industry and caused by hardware penetration, software innovations, and the build-up of the telecommunications infrastructure. This article assesses the software industry in Egypt as a major building block of the information technology industry and a possible active contributor to business and socioeconomic development at large.

BACKGROUND

Although computing started in Egypt in the 1960s, it was only in 1985 that the active role played by the government caused a change in the way information technology was perceived as a vehicle for socioeconomic development and a tool to improve the decision-making process (Kamel, 1999). This change was accelerated by the continuous development of new tools and techniques that had direct and concrete effects on socioeconomic development. Therefore, it is perceived that the way developing countries will manage the computer driven process of change will influence whether their development goals will be promptly achieved. This will be bound to the continuous ability to invest in emerging technologies, the provision of skilled human resources and state-of-the-art information and communication technology infrastructure. Many researchers have identified information technology as the combination of information, computing and communication technologies.

Today, with the evolution and diffusion of the Internet, the integration of these technology elements is invaluable to societies around the world and is strongly contributing to globalization. The importance of information technology has been greatly emphasized in most developing countries in a deliberate effort to ensure that they do not lag behind. In most developing nations, the government has played the most important role in the diffusion of information technology, being the largest user of computers (Moussa & Schware, 1992), and through its policies, laws and regulations it still exerts the largest influence on the diffusion of information technology throughout different organizations (Nidumolu & Goodman, 1993).

In Egypt, since 1985, the driving force for the diffusion of information technology has been the government-private sector partnership. A large number of informatics projects were formulated targeting the use of information technology to leverage managerial and administrative performance in the government. These projects targeted key sectors such as education, health, employment, trade and local administration with a focus to introduce computer-based technologies with its different building blocks, including human and information resources, software, hardware, and networking resources to contribute to socioeconomic development (American Chamber of Commerce in Egypt, 2001).

A COUNTRY PROFILE

Egypt is the cradle of an ancient civilization dating back to 3000 B.C. With a population of about 70 million, it is the most populous country in the region (*www.mcit.gov.eg*). About 28% of its population is enrolled in education programs (schools and universities education) and 19 million represent its workforce (*www.idsc.gov.eg*). Egypt is trying to expand its industrial base and modernize itself technologically, with agriculture accounting for 17% of the gross domestic product, industry for 32% and a large service sector (51%) mainly built around tourism and transportation. A comprehensive economic reform program was implemented that enabled its current economic growth rate to stand at 5.7% annually with an inflation of 6% (*www.economic.idsc.gov.eg*). Estimates show that unemployment is standing at 8% and the labor force is growing at around 2.7% annually (ITU, 2001). The government of Egypt is more determined than ever to build up the national infrastructure and keep pace with the IT evolution worldwide. In October 1999, the government established a ministry for communication and information technology to embark on a master plan to build Egypt from an information and communication technology perspective that is based on the fact that as an emerging market, Egypt has already made considerable achievements in terms of economic development and is ready to move aggressively into the global market, and the only vehicle to realize that objective is through a state-of-the-art information and communication technology infrastructure (Osman, 2000).

THE SOFTWARE INDUSTRY

The software industry in Egypt is still in its infancy stage. However, it is diverse and heterogeneous in nature, with the presence of local vendors and multinationals like most mature markets. Most of the software development companies provide training services to support their products and clients. The software industry is divided into four categories, including (a) software tools, (b) packaged applications, (c) tailored applications and multimedia applications and (d) Arabization of applications. The total number of staff employed in the industry in October 2000 was estimated to be around 6,000, including managers, programmers and project managers mostly involved in the development and delivery of information systems to local and international markets (Osman, 2000). However, the figure excluded IT support staff working in the industry, the government and the private sector, which was estimated to be around 1,000 and an additional 1,000 experts and consultants in the IT training field. Software companies range in size between one to five staff members in start-ups, through to relatively mature firms with around 50 to 150 employees. The majority of firms are located in and around Cairo or Alexandria. However, recently, some of the new start-ups were located in the new industrial areas to benefit from the tax holidays they offer.

Software is the fastest growing segment of the IT industry. It represented 14% of the total market in 2000. The industry was valued at 105 million US dollars with 27% growth rate from 82.75 million US dollars in 1999 (American Chamber of Commerce in Egypt, 2002). Moreover, exports from the industry were valued at 50 million US dollars and are expected to reach 500 million US dollars by 2005 (American Chamber of Commerce in Egypt, 2001). The number of companies having software development as part of their activities is estimated to be 300 (*www.expolink.org.eg*). While these numbers are not impressive if compared with more developed software industries, they do provide a foundation from which to start a serious development of the industry. Moreover, the ministry of communication and information technology has embarked since May 2000 on a plan to train 5,000 fresh graduates annually on the recent IT and communication applications. The plan aims, over the next five years, to train 25,000 fresh graduates that could represent the core of the development of a high-tech industry (Osman, 2000). The expectation for growth in the domestic marketplace for IT products and services is expected to be in the range of 35% for services and products. Imported software accounts for 55% of software market revenues while the remaining sales comprise 19% for locally developed software, 16% for tailored software and 10% for Arabization (Osman, 2000). The market for Arabized software is large in Egypt but there is also a great potential elsewhere, with 300 million Arabic speakers in the Arab countries and worldwide to be served with language-specific software produced in Egypt.

The industry distribution channels in Egypt are still relatively underdeveloped, with around 63% of software sales without intermediaries; 50% of tailored applications are sold bundled with niche products and services. Moreover, software sales through system integrators are low because of limited subcontracting, technical cooperation and interchange of skill and specializations between local companies. Finally, function-oriented software is sold primarily through dealers. With regard to software demand, the government purchases generate 25% of total software revenues, making it the largest demand segment, with two major purchase determinants, which are quality and after-sales service for fear of system failure with cheaper systems (American Chamber of Commerce in Egypt, 1998). However, only 6% of revenues are from sales to small office and homes, which is in part due to the widespread piracy rate of 86% that plagues this segment. This figure is gradually decreasing due to the newly introduced laws against violators of software piracy laws. Also, the number of software applications sold to households is increasing due to the boom in PC sales for household usage and the spread of Internet among younger generations. The competitive advantages of Egypt's domestic software production environment have attracted numerous international producers to subcontract programming of tailored applications. Currently, 6% of subcontracting carried in Egypt is destined for export, and 15 to 20 companies are currently considered active exporters (IDSC, 2000). With the presence of over 50 Internet service providers; there is an expected significant growth in services and software applications that are Internet-based. Additionally, there is an expected increase in the development of applications of a number of key sectors in the economy, including the financial, petroleum, tourism and health sectors (Loch, Straub & Kamel, 2000).

SOFTWARE SWOT ANALYSIS

The software industry in Egypt is better analyzed through a SWOT analysis to be able to understand where it stands and where it is heading with an overview on its strengths and weaknesses and an identification of the opportunities available and the threats faced. Table 1 demonstrates the software industry in Egypt SWOT analysis.

FUTURE TRENDS

Based on the assessment of the software industry in Egypt, an action plan needs to be formulated to attempt to capitalize on the opportunities available and overcome the challenges in the market. The plan can assist Egypt in improving its software industry performance and could include (a) identifying and penetrating international target markets to realize growth, (b) investing in people to leverage the capacities of the key building block in the industry, (c) introducing incubator programs to link the industry to educational institutions and the government, (d) improving the infrastructure to ensure that firms have the vehicle to operate their business; and, (e) increasing government support role to demonstrate the value of software to Egypt's future (Kamel, 2003).

Table 1. Software industry in Egypt SWOT analysis

Strengths	**Weaknesses**
▪ Growing generation of well educated fresh graduates interested in the fields of information technology. ▪ Accumulated, though limited, experience, technical skills and capacities in the industry. ▪ Young, educated and competitive workforce with required knowledge. ▪ Low and competitive labor costs. ▪ Good command of English for dealing with overseas customers. Other languages are also learnt, such as French and Spanish. ▪ Same time zone advantage with Europe and provides a second shift for the United States. ▪ Geographically well located for most African and European cities and some Asian countries. ▪ No shortage of entrepreneurs willing to take risks in the development of their businesses. ▪ Encouragement of the government by facilitating procedures and logistics related to the software industry.	▪ Technical skills are too broad and thin. ▪ Lack of sufficient expertise in any one technology. More focus needs to be achieved ▪ Software companies spend around 6 months to turn graduates into productive contributors. ▪ Lack of skills related to project management, marketing and managing start-ups. Skills are purely related and oriented to software development. ▪ Domestic demand for software is small. ▪ Lack of management recognition to the value of using IT as a business vehicle. ▪ Minimal government role over the last decades. Some changes are starting to be introduced; however, the impact would need some time to become effective. ▪ Infrastructure level and cost is high compared to the capacities of manufacturers and beneficiaries. ▪ Corporate tax rates are high as well as taxes on print and TV advertising for promotional and diffusion purposes.
Opportunities	**Threats**
▪ Creation of software business incubators such as the smart villages' model. ▪ More proactive role played by educational institutions and training centers. ▪ Internships and scholarships from software vendors both local and multinationals. ▪ Promotional role played by software associations to activate the role of software development companies. ▪ Government support role needs to be more at the macro and micro levels. ▪ Changes in tax treatment, reduction on telephone tariffs and the introduction of new IP laws. ▪ Penetration of PCs in homes and businesses.	▪ Copyright violations. ▪ Piracy rates are relatively high. ▪ Government regulations need to be firm and enforced. ▪ Perception that software has little intrinsic value among commercial and government customers. ▪ Lack of understanding for software products and their implications on business development. ▪ Lack of market research in domestic and overseas markets that Egyptian companies could target. ▪ Lack of financial support to the industry. ▪ Limited distribution skills to serve the international software market.

CONCLUSION

Egypt has an excellent opportunity to develop the ingredients of a small but effective software industry. The level of effort and support given by government and supported by industry, finance and education will ultimately determine the level of success of the industry, making the industry profitable and contributing to business and socioeconomic development. There is a wealth of opportunities for Egypt to improve all aspects of development for the software industry. Egypt could dramatically increase the level of revenue and growth in the software business with a relatively small investment and focus to be able to attain the levels of achievement realized in nations such as Ireland and India.

REFERENCES

American Chamber of Commerce in Egypt. (1998, June). *Information technology in Egypt.* Business Studies and Analysis Center.

American Chamber of Commerce in Egypt. (2001, May). *Annual general meeting agenda.* Ministry of Communications and Information Technology.

American Chamber of Commerce in Egypt. (2002, April). *Information technology in Egypt.* Business Studies and Analysis Center.

Economic News Bulletin. (2003). Retrieved September 20, 2003, from *www.economic.idsc.gov.eg*

Egyptian Exporters Association. (2003). Retrieved September 20, 2003, from *www.explolink.org.eg*

Information and Decision Support Center. (2000). *Report on Egypt.*

Information and Decision Support Center. (2003). Retrieved September 20, 2003, from *www.idsc.gov.eg*

International Telecommunication Union. (2001). *Internet on the Nile: Egypt case study.*

Kamel, S. (1999, July 25-29). Information technology transfer to Egypt. *Proceedings of the Portland International Conference on Management of Engineering and Technology (PICMET), Technology & Innovation Management: Setting the Pace for the Third Millennium,* Portland, OR.

Kamel, S. (2003, April 6-7). The implications of the digital economy on a growing digital divide in developing nations. *Proceedings of the 8th American University in Cairo Research Conference on Globalization Revisited: Challenges and Opportunities,* Cairo, Egypt (pp. 60-70).

Loch, K.D., Straub, D.W., & Kamel, S. (2000, June 11-13). Use of the Internet: A study of individuals and organizations in the Arab world. *Proceedings of the 1st Annual Global Information Technology Management World Conference,* Memphis, TN (p. 191).

Ministry of Communications and Information Technology. (2004). Retrieved January 20, 2004, from *www.mcit.gov.eg*

Moussa, A., & Schware, R. (1992). Informatics in Africa: Lessons from World Bank Experience. *World Development, 20*(12).

Nidumolu, S.R., & Goodman, S. (1993, June). Computing in India: An Asian elephant learning to dance. *Communications of the ACM, 236*(4).

Osman, H. (2000, February). Editorial. *Business Today.*

KEY TERMS

Arabization: The transformation of software applications into the Arabic language in terms of usage as well as interface to be able to cater for a community that stands in 2003 at around 300 million people.

Building Blocks: Reflects all the critical success factors of the information technology industry that include: hardware, software, human resources "humanware," networking and information.

Diffusion of Information Technology: Reflects the spreading of information technology concepts among the society of implementation, whether within an organization or within the community at large.

Government-Private Sector Partnership: The teaming of different entities in the government and the private sector to realize a change and a transformation in the development of information technology at large and in the software industry in specific.

Incubator Programs: A form of collaboration usually between the industry, corporations, and the business community and the educational sector aiming at identifying industry and market needs, catering for these needs and creating employment opportunities for the society, especially young graduates.

Informatics Projects: The projects that involve in any way possible the use, design, delivery, implementation and management of information technology irrespective of the element involved, including software, hardware, and so forth.

Information Technology Industry: The accumulation of all elements of information technology design, delivery and management.

Software Industry: Focuses on the needs of the software development industry in terms of infrastructure, know-how, capacities, and development.

Tailored-Applications: Applications based on industry or organizational needs to complement the off-shelf software applications available in the marketplace.

Theoretical Framework for CRM Outsourcing

Babita Gupta
California State University Monterey Bay, USA

Lakshmi S. Iyer
The University of North Carolina at Greensboro, USA

INTRODUCTION

Customer service is emerging as a key differentiator among competitors as the explosive growth of e-commerce is changing the nature of competition among companies. This has changed the customer-related business requirements for all types of companies. With firms increasing their online operations, customers now have the ability to contact organizations through a variety of interactive and noninteractive means (such as e-mail, fax, call centers, FAQs, online chats, newsletters, snail mail, retail stores, and Web-based forums). This has led companies to consider customer relationship management (CRM) as an important part of their competitive strategy. As the focus is shifting to retention rather than acquisition of customers, companies are looking for ways to identify and engage their most profitable customers.

CRM is generally stated as a strategy that companies use to identify, manage, and improve relationships with their customers. Thompson (2001) defined CRM as a company's activities related to developing and retaining customers through increased satisfaction and loyalty. *eCRM* is a term generally used when a company's customer service operations are online. According to Cahners In-Stat Group (2001), worldwide revenues from CRM software application will increase from $9.4 billion in 2001 to approximately $30.6 billion in 2005.

CRM initiatives have high costs of implementation, typically $1 million, as well as high failure rates, estimated to be in the 55–75% range (Ericson, 2001). For large-scale CRM implementations at very large businesses, reported spending on CRM software, hardware, services, internal labor, and training have been as high as $30–90 million (Young, 2001).

For effective CRM, companies must generally adopt a customer-centric philosophy to achieve goals, such as increases in sales, customer loyalty, customer service and support, and better and effective distribution of products. CRM can produce additional profits for firms through cross selling, up selling, reduced product marginal costs, and lower customer acquisition costs (Winer, 2001). CRM applications must be designed to integrate all the customer communication points across various functional areas of a company. CRM products are generally classified into three categories (Karimi, Somers, & Gupta, 2001). These are as follows:

1. *Operational:* For improving sales, marketing, and customer service efficiency through marketing campaign management, service request management, and automation that integrate with existing processes and infrastructure.
2. *Analytical:* For collecting better customer data that contain customer buying histories and demographics data, mining this data to generate customer profiles and anticipate their needs, and thus formulating more effective customer-centric strategies.
3. *Collaborative:* For integrating communications across various channels to improve information sharing across the organization to build an integrated view of the customer, ensuring consistency of message to customer; for eCRM, one-to-one personalized Web marketing, customized product and services offerings for individual customers, etc.

Despite the rise in popularity of CRM outsourcing, there is little literature in information system (IS) outsourcing specific to CRM functions. This article presents an integrated framework of resource dependence, transaction cost economics, and social exchange theory. This framework is used to study CRM outsourcing partnerships and is an extension of the work of Grover, Cheon, and Teng (1996) and Lee and Kim (1999). This framework could be used to evaluate how these companies select a CRM vendor, how they adopt and integrate CRM technologies into their existing infrastructures, and what factors affect CRM outsourcing success.

OUTSOURCING OVERVIEW

Outsourcing is more a reflection of the strategic partnerships in the digital economy. While there are many definitions of IS outsourcing in the literature, there are three common components to these definitions, as stated in Yang and Huang (2000): "first an external provider takes over part or all of an organization IS functions; second, external provider should take the responsibility; and third, customers transfer IS functions to external provider as well as employee and part of computer facilities" (p. 227).

The nature and extent of IS outsourcing have evolved over the past few decades. IS functions of increasingly high asset specificity involving responsibilities for not just the technology but also for the business processes are being outsourced (Grover, Cheon, & Teng, 1994; Grover et al., 1996; Gurbaxani, 1996; Nam, Rajagopalan, Rao, & Chaudhury, 1996; Lee, Huynh, Chi-wai, & Pi, 2000). The reasons cited for widespread use of IT-enabled outsourcing by large U.S. companies are: slowing domestic economic growth, need to conserve costs, ability to focus on core capabilities and providing bottom-line benefits (Banking on Outsourcing, 2003).

Increasingly, the nature of the client–vendor relationship has shifted from just a contractual relationship to a tightly integrated partnership relationship for mutual benefit between the vendor and the outsourcing firm, as firms consider outsourcing a key strategic choice (Grover et al., 1996; Lee & Kim, 1999; Lee et al., 2000).

CRM implementations require far more coordination among functionally disparate organizational units, including IT and senior management involvement in IT planning, organization, control, and integration (Karimi et al., 2001). Because CRM initiatives are not confined to a particular function but rather cut across various functions of an organization, the process of CRM implementation through outsourcing becomes complicated. Because of the complex technologies involved in CRM, companies are choosing to outsource to vendors that specialize in CRM. IDC research shows that the worldwide CRM services market will increase from $19.4 billion in 2001 to $45.5 billion in 2006 (Morphy, 2002).

Outsourcing offers various strategic, technological, and economic advantages to a firm (Grover et al., 1996; Smith, Mitra, & Narsimhan, 1998; Lankford & Parsa, 1999; Ngwenyama & Bryson, 1999; King & Malhotra, 2000; Lee et al., 2000; Yang & Huang, 2000). These are as follows:

- *Strategic benefits* by allowing a firm to focus on its core competencies by outsourcing routine IT functions, being able to acquire state-of-the-art knowledge that the firm otherwise would not have the resources to acquire and improve IS service quality.
- *Technological benefits* through acquisition of complex technologies through an external vendor that otherwise would have high internal acquisition and coordination costs, high obsolescence risks, and longer time-to-market.
- *Economic benefits* through reduced costs and improved efficiencies in the long term by utilizing external vendors' expertise and economies of scale, and through favorable allocation of fixed costs.

Cost reductions through outsourcing to external vendors result through two primary approaches. One approach is adversarial, involving bidding among competing vendors to drive down prices. The other approach is through collaboration between the vendor and firm to lower transactional costs (Canon & Homburg, 2001). Outsourcing strategies involve contracting with either a single vendor or multiple vendors, though developing a relationship with a single vendor is more cost effective and leads to better outsourcing performance over the long term (Ngwenyama & Bryson, 1999).

Notwithstanding the advantages, outsourcing IS functions may fail due to complexities involved in managing long-term relationships with its vendor, resulting in loss of control of organizational assets, loss of firms' internal IS expertise and capacity to learn new skills and technologies, threat of opportunism from vendor, uncertainties and lack of decision models in choosing outsourcing vendor, and loss of morale and performance among firm's employees (Rao, Nam, & Chaudhury, 1996; Ngwenyama & Bryson, 1999; King & Malhotra, 2000). High-asset-specific IS outsourcing may also result in competitive threat from the vendor, and in addition, a firm may lose out on future business opportunities if the skills and competencies being outsourced appreciate in value (King & Malhotra, 2000).

CRM OUTSOURCING FRAMEWORK

While there are a number of studies for evaluating outsourcing of IS functions (Grover et al., 1996; Lee & Kim, 1999; Maltz & Ellram, 1999; King & Malhotra, 2000; Kini, 2000; Lee et al., 2000; Yang & Huang, 2000), there is very little literature specific to the outsourcing of CRM functions. This framework contributes to the IS field by studying CRM outsourcing partnerships through an integrated framework of resource dependence, transaction cost economics, and social exchange theories.

Resource-dependency theory refers to outsourcing to external vendors to fill resource gaps within the firm for the purpose of providing the firm with strategic competitive advantage (Lee et al., 2000).

Transaction cost theory provides an economic viewpoint to outsourcing via a set of principles for analyzing buyer–supplier (outsourcer–vendor) transactions and determining the most efficient mode of structuring and managing them (Nam et al., 1996; Ngwenyama & Bryson, 1999; Lee et al., 2000). Yang and Huang (2000) considered five factors—management, strategy, technology, economics, and quality—when determining the outsourcing benefits to the firm.

Social exchange theory is used to explain the close partnership style of an outsourcing relationship between the firm and vendor by taking into account prior relationships, trust, and culture, and their effects on an ongoing outsourcing relationship (Lee & Kim, 1999; Lee et al., 2000; Yang & Huang, 2000). Trust is perceived as the firm's belief that the vendor has in mind the best interests of the firm and will perform accordingly to achieve desired goals for the firm (Lee & Kim, 1999).

MEASURING CRM OUTSOURCING SUCCESS

The model (Figure 1) proposes to study the relationship between the level of CRM outsourcing and its success, how the CRM outsourcing success is influenced by the quality of service provided by the CRM vendor, and the partnership quality between the outsourcing company and the external vendor.

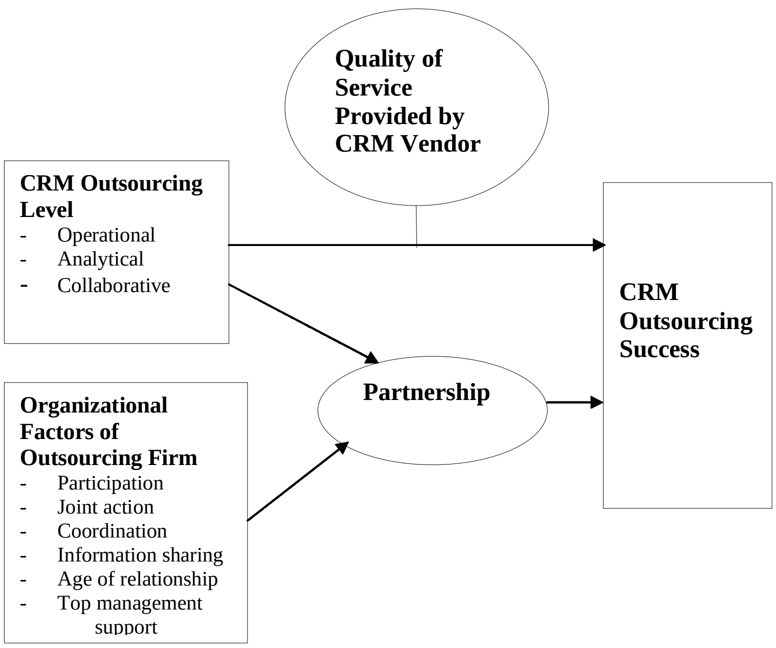
Generally, partnership relationship between vendor and firm is categorized as being of either transactional or partnership style. A transactional-style relationship is well specified in its contractual obligations. In contrast, the partnership-style relationships include sharing risk and benefit with a view to establish long-term commitment (Henderson, 1990; Lee & Kim, 1999; Lee et al., 2000).

The model also considers the effect of internal management issues and other organizational factors of the outsourcing company on the partnership quality and the success of CRM outsourcing. We consider prior outsourcing experience of an outsourcing firm, as this may affect outsourcing decisions (Nam et al., 1996; Lee & Kim, 1999; Lee et al., 2000).

MODEL CONSTRUCTS AND PROPOSITIONS

1. *CRM outsourcing level:* We adopt the classification of CRM products into three levels—operational, analytical, and collaborative (Karimi et al., 2001) and explore the relationship between level of CRM outsourcing and outsourcing success.
2. *Quality of service provided by CRM vendor:* The relationship between CRM outsourcing level and CRM outsourcing success will be stronger if service quality is higher. We adopt and adapt the two dimensions of the SERVQUAL instrument construct used in the Grover et al. (1996) study, tangibles and reliability, and add a new one, CRM technology, that pertains to product features, does one vendor provide better fit with the outsourcing company's existing technology than others, etc.
3. *Partnership:* We treat partnership as an intermediate variable. To evaluate the company–CRM vendor partnership relation, we use four dimensions: communication, trust, cooperation, and satisfaction (Grover et al., 1996; Lee & Kim, 1999).

Figure 1. Research model to measure the success of CRM outsourcing



4. *Organizational factors of outsourcing firm:* These include internal management issues that influence partnership quality, which in turn, affects outsourcing success (Lee & Kim, 1999).
5. *CRM outsourcing success:* High partnership levels between the CRM outsourcing company and the CRM vendor will help achieve the organizational objectives and provide a competitive advantage. The dependent variable, success of CRM outsourcing, is measured by using the widely used concept of satisfaction due to benefits gained by the CRM outsourcing company from a business perspective (Grover et al., 1996; Chin & Lee, 2000; Lee et al., 2000).

SUMMARY

We present a theoretical model that examines the relationship between CRM outsourcing levels, organizational factors of the outsourcing firm, partnership quality, and the success of CRM outsourcing from a business perspective. The model considers how quality of service provided by a vendor influences CRM outsourcing success and seeks to answer questions like how companies select CRM vendors, how they adopt and integrate CRM technologies into their existing infrastructures, and what factors affect CRM outsourcing success.

REFERENCES

Banking on Outsourcing. Retrieved October 20, 2003, from *http://www.tatatelecom.com/outsourcing/concept/boutsource.asp*

Cahners In-Stat Group. (2001, July). Retrieved September 29, 2001, from *http://www.instat.com/pr/2001/ec0105st_pr.htm*

Canon, J. P., & Homburg, C. (2001). Buyers–suppliers relationships and customer firm costs. *Journal of Marketing, 65*(1), 29–43.

Chin, W. W., & Lee, M. K. O. (2000, December 10–13). A proposed model and measurement instrument for the formation of IS satisfaction: The case of end-user computing satisfaction. In *Proceedings of the 21st International Conference on Information Systems*, Brisbane, Australia.

Ericson, J. (2001, August). The “failure” of CRM. Line 56. *e-Business Executive Daily*. Retrieved July 12, 2002, from *http://line56.com/print/default.asp?ArticleID=2808*

Grover, V., Cheon, M. J., & Teng, J. T. C. (1994). An evaluation of the impact of corporate strategy and the role of information technology on IS functional outsourcing. *European Journal of Information Systems, 3*(3), 179–190.

Grover, V., Cheon, M. J., & Teng, J. T. C. (1996). The effect of service quality and partnership on the outsourcing of information systems function. *Journal of Management Information Systems, 12*(4), 89–116.

Gurbaxani, V. (1996). The new world of information technology outsourcing. *Association for Computing Machinery. Communications of the ACM, 39*(7), 45–46.

Henderson, J. C. (1990). Plugging into strategic partnerships: The critical IS connection. *Sloan Management Review, 30*, 37–18.

Karimi, J., Somers, T. M., & Gupta, Y. P. (2001). Impact of information technology management practices on customer service. *Journal of Management Information Systems, 17*(4), 125–158.

King, W. R., & Malhotra, Y. (2000). Developing a framework for analyzing IS sourcing. *Information and Management, 37*(6), 323–334.

Kini, R. B. (2000). Information systems outsourcing evaluation strategy: A precursor for outsourcing. *International Journal of Management, 17*(1).

Lankford, W. M., & Parsa, F. (1999). Outsourcing: A primer. *Management Decision, 37*(4).

Lee, J., Huynh, M. Q., Chi-wai, K. R., & Pi, S. (2000, January). The evolution of outsourcing research: What is the next issue? In *Proceedings of the 33rd Hawaii International Conference on System Sciences.*

Lee, J., & Kim, Y. (1999). Effect of partnership quality on IS outsourcing: Conceptual framework and empirical validation. *Journal of Management Information Systems, 15*(4), 29–61.

Maltz, A., & Ellram, L. (1999). Outsourcing supply management. *Journal of Supply Chain Management, 35*(2), 4–17.

Morphy, E. (2002, April). CRM Web services trend building. *CRMDaily.com*. Retrieved July 12, 2002, from *http://www.ecommercetimes.com/perl/story/17437.html#story-start*

Nam, K., Rajagopalan, S., Rao, H. R., & Chaudhury, A. (1996). A two-level investigation of information systems outsourcing. *Association for Computing Machinery. Communications of the ACM, 39*(7), 36–44.

Ngwenyama, O. K., & Bryson, N. (1999). Making the information systems outsourcing decision: A transaction cost approach to analyzing outsourcing decision. *Euro-*

pean Journal of Operational Research, 115(2), 351–367.

Rao, H. R., Nam, K., & Chaudhury, A. (1996). Information systems outsourcing. *Association for Computing Machinery. Communications of the ACM, 39*(7), 27–28.

Smith, M. A., Mitra, S., & Narasimhan, S. (1998). Information systems outsourcing: A study of pre-event firm characteristics. *Journal of Management Information Systems, 15*(2), 61–93.

Thompson, B. (2001). What is CRM? Retrieved January 29, 2001, from *http://www.CRMguru.com*

Winer, R. S. (2001). A framework for customer relationship management. *California Management Review, 43*(4), 89–105.

Yang, C., & Huang, J. (2000). A decision model for IS outsourcing. *International Journal of Information Management, 20*(3), 225–239.

Young, D. (2001). CRM: Miscalculated ROI projections can result in multimillion-dollar IOUs. *Wireless Review, 18*, 10–14.

KEY TERMS

Customer Relationship Management (CRM): This is a strategy that companies use to identify, manage, and improve relationships with their most profitable customers to create long-term value for the firm.

Data Mining: Process of using statistics, artificial intelligence, and sophisticated algorithmic techniques to discover useful hidden relationships or patterns in large amounts of data.

eCRM: This is a strategy that companies use to identify, manage, and improve relationships with their most profitable online customers to create long-term value for the firm when a company's customer service operations are on the Internet, using e-mail, fax, Internet call centers, FAQs, online chats, and Web-based forums.

Information System Outsourcing: It is defined as transfer of an organization's IS function(s) to an external provider who takes the responsibility for managing these functions on behalf of the organization.

Resource-Dependency Theory: This refers to outsourcing of a firm's IT operations to external vendors to fill resource gaps within the firm for the purpose of providing the firm with strategic competitive advantage.

Social Exchange Theory: It is used to explain the close partnership style of outsourcing relationship between the firm and vendor by taking into account prior relationships, trust, and culture, and their affect on ongoing outsourcing relationships.

Transaction Cost Theory: It provides an economic viewpoint to outsourcing via a set of principles for analyzing buyer–supplier (outsourcer–vendor) transactions and determining the most efficient mode of structuring and managing them.

Transferring Technology to the Developing World

Stacie Petter
Georgia State University, USA

Galen Sevcik
Georgia State University, USA

Detmar Straub
Georgia State University, USA

INTRODUCTION

The potential for technology growth in lesser developed countries (LDCs), particularly within the Arab World, is enormous (Loch, Straub, & Kamel, 2003). Within the Middle East, the projected demand for information technology hardware and services will be $8.9 billion by 2005 (Pyramid Research, 2001). Yet, in spite of this significant promise, information technology transfer (ITT) has been slow in this region. Clearly, there is more than money at issue in diffusing computing and the Internet throughout the Arab world, and the less developed world in general.

What are the reasons for this? As with most LDCs, resources for purchasing systems may be scarce, both for businesses and individuals. The public infrastructure for electricity and telecommunications may be lacking. Political instability, disease, and warfare may limit the ability of a people to absorb and effectively utilize technology, especially in many African countries.

Yet these reasons are not sufficient to explain the slow growth of IT in developing and least developed countries. There must be something more at stake. The scientific literature, although not as extensive as work in the developed countries, suggests several explanations for this phenomenon.

First, it is clear that much of the research related to IT has taken place in developed countries, primarily in North America. But as the business environment becomes more global, it becomes increasingly important to examine the influence of IT in other contexts and cultures (Checchi, Hsieh, & Straub, 2003). Naturally, countries that are less developed economically do not invest heavily in information infrastructure, limiting a country's ability to make efficient use of resources and effective choices for investments opportunities (Checchi, Sevcik, Loch, & Straub, 2002). Evidence is mounting, however, that other reasons may affect ITT in developing countries (Straub, Loch, & Hill, 2001). Loch, Straub, and Sevcik (2000) offer two reasons why the transfer of IT is difficult in the developing countries and in the Arab world:

- Cultural differences affecting system development and implementation
- Government policies influencing ITT

BACKGROUND

Before examining the transfer of IT to the Arab world, it is important to understand the impact of culture on IT. Identifying the culture of a nation is difficult; therefore, culture is often defined by national boundaries. However, the distinction between cultures is more sophisticated than geography. Researchers have defined culture based on shared values, problem solving and outcomes, and general, all-encompassing definitions, but culture is a complex, multi-layered phenomenon (Straub, Loch, Evaristo, Karahanna, & Srite, 2002). To classify culture, Ein-Dor, Segev, and Orgad (1992) identified factors, or "constants", that influence the culture and are consistent over time. Examples of constants identified by Ein-Dor et al. (1992) include geography, language, social norms, and traditions. Factors that change over time, "changeables", include technology, GDP, and level of education. Understanding the impact of the constants on IT solutions is critical for system architects as technology becomes implemented worldwide.

The Arab world consists of over twenty countries and 200 million people; therefore, it is difficult to definitively define the culture of such a large population. Within the Arab world, there are many subcultures; however, there are several characteristics that span across sub-cultural boundaries (Hill, Loch, Straub, & El-Sheshai, 1998). Barakat (1993) found several social characteristics that traverse the entire Arab culture, such as patriarchal family structures, pyramidal class structures, and expressive social interactions. Yet, he also found conflicting values within groups of the Arab world. He acknowledges that the dominant values of the Arab world include conformity, collectivism, and obedience, among others, but subgroups within the culture are emphasizing creativity, individualism, and rebellion. The social characteristics, dominant values, and emerging contradictory values are critical in shaping the Arab culture as we know it today (Barakat, 1993). According to Hofstede's (1980) work, Arabs have a strongly collectivist culture compared to the individualist North American culture. In addition, the Arab world is characterized by large power distance and high uncertainty avoidance. Regarding the position of the Arab culture on the masculine/femininity scale, the culture is considered moderately masculine. Examining the Arab world using the works of Barakat (1993) and Hofstede (1980) yield insight into the Arab organizational culture and the response towards technology transfer.

Research has shown that culture impacts the acceptance of technology. By understanding the effect of cultural beliefs and attitudes toward technology, it can be adapted to a group's behavioral patterns (Loch et al., 2000). Kransberg and Davenport (1972) argue that "an advance in technology not only must be congruent with the surrounding technology but *must also be compatible with…existing economic and other cultural and social institutions* [italics added]."

Although culture is considered to be a powerful force in ITT, much of the research on factors affecting ITT is related to economic issues and characteristics of the organizations implementing the system (Hill, Straub, Loch, Cotterman & El-Sheshai, 1998). While the economic state of a country does affect ITT, for many countries within the Arab world, cost is not an issue. For economically developed countries, such as Saudi Arabia and Kuwait, the adoption of IT has still been slow (Straub, 2003), suggesting that other factors, such as culture, are important.

DEVELOPMENT AND IMPLEMENTATION OF SYSTEMS

System implementations are difficult across cultural boundaries. Technology designed and created in developed countries has cultural-biases embedded within the system (Escobar, 1994). Cultural-specific beliefs can inhibit ITT in developing countries and within the Arab world. Typically, when implementing systems, the national culture is not taken into account and no adaptation occurs (Hill, Straub, Loch, Cotterman, & El-Sheshai, 1994). Two examples of cultural differences that impede system implementation in the Arab world include differences in leadership style and perception of time.

Cross-cultural conflicts between the styles of western and Arab leadership affect the system development process (Straub et al., 2001). Within the Arab culture, there is a preference within organizations for face-to-face meetings and family-like environments (Straub et al., 2001). These cultural preferences lessen the impact of groupware and e-mail technology within these countries. Practical implications of these cultural-specific beliefs of ITT within the Arab world suggest that managers should ensure top management support, even more so than in developed countries, before introducing new technology (Rose & Straub, 1998). In addition, technology such as e-mail or groupware should not be used to replace face-to-face meetings, but rather serve as a supplement to information exchange (Straub et al., 2001).

Another difference between cultures that affects the implementation of systems within the Arab world is the perception of time. The task-oriented, linear view of time that is found in Western culture is monochronic. The polychronic view of time found in Latin America and in the Arab world suggests that events unfold in parallel (Straub, 2003). These differences have strong practical implications. First, the Arab culture focuses less on long-term planning and forecasting than in western cultures (Straub et al., 2001); therefore, planning systems may be less accepted within the Arab world as compared to their western counterparts (Loch et al., 2000). Second, those in monochronic cultures tend to focus on one task at a time; however, people that are part of polychronic cultures often perform several tasks in parallel (Rose, Evaristo, & Straub, 2003).

With the explosive growth of the Internet and the supporting infrastructures, companies that are marketing globally must take into account other factors than language and currency. Research by Rose et al. (2003) has found that in monochronic cultures, long download times of a Web site lead to a negative attitude about the delay in viewing the content. But they also found that polychronic Web users, such as most Arabs, had more tolerance for long download times. For Web-based companies that are targeting people from polychronic cultures, it is not as important as in monochronic cultures to have an infrastructure and Web site design that reduces download time. Lower cost servers can be installed in these countries without having a negative effect on the attitude toward the website.

Although cultural factors can influence the ability to implement systems across national and cultural boundaries, technological culturation can offset this effect (Loch et al., 2000). Acculturation is an anthropological concept in which characteristics of one culture are assimilated into another culture. People in less developed countries may implement IT systems with cultural bias themselves if they either work for international companies or have spent time in developed countries, such as the U.S., Western Europe or Japan (Straub et al., 2001). People that have become technologically culturated are often more accepting of IT (Loch et al., 2000). Straub et al. (2001) confirmed this linkage between acculturation and ITT within the Arab world, and advised that firms in developing Arab countries should expose their employees informally to high technology cultures.

IT POLICY AND DIFFUSION OF IT

In addition to the cultural differences affecting systems development and implementation, a second determinant of ITT within the Arab world is the national IT policy executed by the government. IT policies influence the development of information systems based on the level of support a government provides within a given country (Loch et al., 2000). Governmental IT policy strengthens economic development in that higher-quality information can be used by both public and private sector organizations to more effectively use national resources (Checchi et al., 2002).

The influence of national IT policy on ITT can vary dramatically between countries. Checchi et al. (2003) noted several differences in the impact of policy on the diffusion of IT. First, in the least developed countries, regional or international agencies play a more important role in creating the IT infrastructure than the national government. In less developed countries, the private sector and individuals have little involvement in the process of developing IT capabilities. In these countries, the government is the most active in influencing ITT. Table 1 summarizes the differences in policies between developed and less developed countries.

Singapore is a success story where ITT occurred more easily across segments of the population because the government played an active role and encouraged ownership and investment in technology (Loch et al., 2000). Ireland is another example of a country that has been successful at implementing national IT policies to promote economic and technological development. By educating the workforce in technology and system programming, Ireland has been able to bypass an industrial revolution and progress straight to the information age (Straub, 2003). Other countries have failed to implement effective ITT policies due to lack of training, top management support, and user acceptance of technology (Loch et al., 2000). Some countries have made the mistake of relying too strongly on foreign assistance or focus on the consumption rather than the production of IT (Checchi et al., 2003). Having a national IT policy is not enough to ensure technological advancement. Table 2 lists successful and unsuccessful factors in national IT policies.

According to Al-Abdul-Gader (1999), there are several problems that governments within the Arab world need to address to ensure information technology transfer among its people. Similar to the active role that the Singapore government played in ITT, his recommendation is that an administrative body is needed to address these problems. Some issues this administrative body would address include:

- Creating and executing a national technology plan;
- Supporting collaborations of information systems across nations and regions;

Table 1. Differences with respect to IT policy setting (Checchi et al., 2003)

Policy Intervention Component	***Developed Countries***	***Less Developed Countries***
Initiator for policy making (mostly)	Domestic government	Regional or international agencies
Attitude of the government	Proactive	Passive or reactive
Investment purpose	Invest in both research knowledge and IT infrastructure	Invest mostly in IT infrastructure
Typical capabilities of the government	Has both technical and financial capabilities	Lack of technical skills, financial limitations
Position with respect to standards	Standard setting	Standard following

Table 2. Success factors in IT policies (Checchi et al., 2003)

Successful IT Policies	*Unsuccessful IT Policies*
Long-term oriented	Short-term oriented
Capability-building	Resource-consumption
Adaptive	Less-responsive
Collaborative	Non-collaborative

- Adopting a plan to develop and implement standards for information systems;
- Promoting the use of the Arabic language in technology applications; and
- Preparing and training the workforce in technology.

Many studies have concluded that ITT is in its infancy within the Arab world, and government policy needs to be more proactive in addressing these issues. Learning from Ireland, the countries within the Arab need to better train the workforce and prepare them for the diffusion of information technology (Al-Abdul-Gader, 1999).

FUTURE TRENDS

This review of literature suggests that there are two primary factors that affect ITT to the Arab World. They are: cultural differences affecting the implementation of systems and governmental policies that encourage or discourage the diffusion of IT. While these factors are critical in transferring information technology to the Arab world, little research has been devoted to fully examining these topics. More work is needed to better understand and distinguish between the subcultures within the Arab world. While there are shared characteristics across the region, understanding the sub-cultural differences could improve ITT. Additionally, prior research on governmental policies has been almost exclusively at the national level (Checchi et al., 2003). Future work should examine the interactions between policies at the national, international, and regional levels to determine the impact on economic and technological development.

CONCLUSION

The factors influencing the transfer of information technology to the developing world and to the Arab world in particular are more complex than simply language, geographical, and currency differences. Another misconception about the lack of ITT is related solely to the monetary investments needed for infrastructure and systems development and implementation. These misconceptions divert attention from other important factors that affect the diffusion of IT in the developing world, namely, cultural differences and governmental policies. A better understanding of these two important factors and implementing the findings into practice will aid in improving ITT to the developing world in the future.

REFERENCES

Al-Abdul-Gader, A.H. (1999). *Managing computer based information systems in developing countries*. Hershey, PA: Idea Group Publishing.

Barakat, H. (1993). *The Arab world*. Berkeley, CA: University of California Press.

Checchi, R.M., Hsieh, J.J.P.-A., & Straub, D.W. (2003). Public IT policies in less developed countries: A critical assessment of the literature and a reference framework for future work. *Journal of Global Information Technology Management, 6*(4), 45-64.

Checchi, R.M., Sevcik, G.R., Loch, K.D., & Straub, D.W. (2002). An instrumentation process for measuring ICT policies and culture. Paper presented at the *Proceedings of the Information and Communications Technologies and Development Conference*, Kathmandu, Nepal.

Ein-Dor, P., Segev, E., & Orgad, M. (1992). The effect of national culture on IS: Implications for international information systems. *Journal of Global Information Management, 1*(1), 33-44.

Escobar, A. (1994). Welcome to Cyberia: Notes on the anthropology of cyberculture. *Current Anthropology, 35*, 211-231.

Hill, C., Loch, K., Straub, D., & El-Sheshai, K. (1998). A qualitative assessment of Arab culture and information technology transfer. *Journal of Global Information Management, 6*(3), 29-38.

Hill, C.E., Straub, D.W., Loch, K.D., Cotterman, W., & El-Sheshai, K. (1994). The impact of Arab culture on the diffusion of information technology: A culture-centered model. Paper presented at the *Proceedings of The Impact of Informatics on Society: Key Issues for Developing Countries, IFIP 9.4*, Havana, Cuba.

Hofstede, G. (1980). *Culture's consequences* (1st ed.). Beverly Hills, CA: Sage Publications.

Kransberg, M., & Davenport, W. (1972). *Technology and culture: An anthology*. New York: Schocken Books.

Loch, K., Straub, D., & Sevcik, G.R. (2000). *IT transfer to Egypt: A process model for developing countries* (No. 0082473): NSF Funded Proposal.

Loch, K.D., Straub, D.W., & Kamel, S. (2003). Diffusing the Internet in the Arab world: The role of social norms and technological culturation. *IEEE Transactions on Engineering Management, 50*(1), 45-63.

Pyramid Research. (2001). ASP strategies and IT markets in the Arab Middle East, from *www.pyramidresearch.com*

Rose, G., & Straub, D. (1998). Predicting general IT use: A study in Arab developing nations. *Journal of Global Information Management, 6*(3), 39-46.

Rose, G.M., Evaristo, R., & Straub, D.W. (2003). Culture and consumer responses to Web download time: A four-continent study of mono and polychronism. *IEEE Transactions on Engineering Management, 50*(1), 31-44.

Straub, D., Loch, K., & Hill, C. (2001). Transfer of information technology to developing countries: A test of cultural influence modeling in the Arab World. *Journal of Global Information Management, 9*(4), 6-28.

Straub, D.W. (2003). *Foundations of net-enhanced organizations*. New York: Wiley.

Straub, D.W., Loch, K.D., Evaristo, R., Karahanna, E., & Srite, M. (2002). Toward a theory-based measurement of culture. *Journal of Global Information Management, 10*(1), 13-23.

KEY TERMS

Acculturation: The process by which a person adopts and assimilates characteristics of another culture into one's own culture

Arab World: An area of 5.25 million square miles, with a young population of 200 million people that encompasses 21 states. The countries of the Arab world include Algeria, Bahrain, Egypt, Iraq, Jordan, Kuwait, Lebanon, Libya, Mauritania, Morocco, Oman, Palestine (including the West Bank, Gaza, and East Jerusalem), Qatar, Saudi Arabia, Sudan, Syria, Tunisia, United Arab Emirates, and Yemen.

Culture: A set of multi-layered characteristics, beliefs, and values shared by a group of people that are consistently held over time.

Information Technology Transfer (ITT): The movement of information technology from creators to users.

Lesser Developed Countries (LDCs): Classes of countries in developing, emerging markets with a least developed status; older and more politically incorrect terms for these countries include second and third world and underdeveloped/undeveloped.

National IT Policies: Technology policies that influence the development and diffusion of information systems in a country.

Western Culture: The set of beliefs and values that are found in North America and Europe, that are typically characterized by higher levels of individualism and lower levels of uncertainty avoidance.

Translation of Natural Language Patterns to Object and Process Modeling

Alexandra Galatescu
National Institute for R&D in Informatics, Romania

INTRODUCTION

In order to shorten the engineering period of the complex information systems (ISs), the *integration of the models* is needed for uniformly: (1) traversing all phases of ISs' lifecycle (analysis, design, coding, testing), and (2) representing object, functional, process and organizational models on the business domain.

The seamless integration of the four models has not yet been a satisfactory solution from the conceptual, notational, semantic and logical viewpoints. The existing tools direct the designers to object-oriented modeling, possibly combined with "use case" and "state transition" diagrams. But, the functional, process and organizational models are not well and completely integrated with the object models. Moreover, using the agent-based technology for IS implementation, the knowledge modeling and integration will be necessary.

The integration solutions proposed in important methodologies like UML (Unified Modeling Language) (OMG, 2003) and IDEF (KBSI, 2000) (for object, functional and process models) or Workflow Reference Model (WfMC, 2003) (for organizational, process and functional models) mainly merge the models. Their conceptual integration is devolved to the developers of CASE tools or to the human designers and is accomplished during the coding phase. The existing methodologies, usually relying on the symbolic notation, do not provide the seamless and explicit (outside the code) integration of the object and activity-like symbols, from the semantic and logical viewpoints.

The integration abilities of natural language (NL) appear from the observation that people describe in NL any kind of information about objects, processes, information flows, the organization of their life and work, their knowledge, belief, intention, rationale and so forth. The universality and syntactic stability of a linguistic model is supposed to facilitate the communication among distributed ISs and users.

In computational linguistics, the main objective is to solve NL ambiguities (when more than one meaning is possible in a sentence) and to correctly identify the syntactic categories. Some NL analyzers build models for knowledge representation (logical models, semantic networks, frames, conceptual dependencies, conceptual graphs, etc.).

Instead, the discourse in the conceptual models (CMs) is a priori considered unambiguous and is expected to have a sound theoretical background. The objectives of NL analysis for model integration should be: (1) a linguistic theoretical foundation to the modeling interface, and (2) a uniform translation of NL to all types of CMs, without information loss. This article will focus on the second objective.

BACKGROUND

This section gives technical reasons for the research on natural language (NL) translation to conceptual models (CMs) (basically for model integration purpose) and a review of the state of the art and of the basic concepts and problems in this domain.

The complexity of information system (IS) representation comes from the complexity and diversity of the concepts it should integrate, defined in object, functional, process and organizational models. The integration of these models should be accomplished during the IS analysis or design phase.

These models result from the abstraction of the IS requirements, expressed by the analysts in NL. A comparison between the NL model (Allen, 1995; Sag, 1999), and the existing CMs reveals the conceptual relationships between NL and conceptual modeling. The most important relationships are:

- Between *categories* in CMs and NL. For example, between objects and nouns, activities and verbs, object attributes and adjectives, activity attributes and adverbs. Also, the noun and verb determiners/ modifiers/substitutes in NL have counterparts in CMs.
- Between *semantic relationships* in CMs and NL. For example, between object aggregation/fragmentation and noun meronymy/holonymy; between object specialization/generalization and noun hyponymy/hypernymy; between functional composition/decomposition of processes and verb meronymy/ holonymy, and so forth.
- Between *syntactic roles* in NL (subject, predicate, direct/indirect/prepositional object, complement,

adverbial modifier) and primitives in CMs. For example, the predicate is represented by an activity/event; the subject in active voice is the object-like sender of the message or initiator of an activity; the direct, indirect, prepositional objects are object-like parameters in activity execution; and so forth.

- Between *structures* in NL and CMs. For example, the simple sentence in NL is represented by the activity signature (list of object-like parameters that participate in the activity execution); the complex sentences are represented by (sub)process diagrams, and so forth.

Lately, the *co-references* between concepts in two or several sentences and the *ellipses* in NL are also represented in CMs.

This similarity made the researchers think that conceptual modeling could be as powerful as NL for representing the reality. But, the translation of NL patterns into a CM that integrates the four models is still an open problem.

NL-CM translation has been tried in several research domains: linguistic interpretation of the models (mainly, entity-relationship and object-oriented models), semantic integration of the conceptual schemas, modeling the systems' dynamics, human-computer interaction, requirements engineering, organization modeling, knowledge representation, formal ontologies and their application to the search on Web, business communication modeling based on speech act theory, and so forth.

The most important results have been obtained for the translation of NL to object and event models. Among the *NL-oriented representations* of these models, the most important are the functional grammar (FG) and the semantic networks (especially, conceptual dependencies and conceptual graphs). They propose the representation of the object models by syntactic categories and rules.

Functional grammar represents the functional aspects in NL, by the description and classification of predicate frames. FG has been further used for defining CPL (conceptual prototyping language), which focuses on NL simple sentences (the intersentential relations are not explicitly revealed). Also, CPL does not approach general semantic relationships inside the lexical categories (e.g., noun/verb synonymy, antonymy, homonymy, etc). COLOR-X (Riet, 1998) can be considered the most important application of CPL and practical result for CMs' integration. It integrates static object and event models of information and communication systems, abstracted from their textual descriptions. It relies on OMT (object modeling techniques). Like in any object-oriented model, the processes merely trace events that compose scenarios, similar to use case diagrams. Lately, the textual requirements are transformed into UML-schemata, for example (Fliedl, 2000).

The *conceptual dependencies* and, lately, the *conceptual graphs* (CG) (Sowa, 2000) are other linguistic representations for CMs. The syntactic categories are suggested by their roles to each other (meaning relationships between nouns and the verb that governs them in a simple sentence, e.g., agent, patient, instrument, recipient, location, time, source, destination, etc). A similar representation is the frame description in FrameNet (Filmore, 2002). But, all these representations are data-centric (ISs' dynamic behaviour is not important as a modeling goal).

The translation of NL to organizational models (workflows) has been obtained mainly with respect to the modeling of the business communication (e.g., Steuten, 2000). The theory of communicative actions and, lately, the *speech act theory* (Johannesson, 2001) are the main linguistic representations of the communication aspects in ISs.

The globalization of the organizations has a great impact on IS representation, especially with respect to the common vocabularies and the interoperability between distributed and heterogeneous applications. In this context, the conceptual modeling must step into a new era and intersect a new field: formal *ontologies*. A first benefit from ontologies for IS representation is that they describe, categorize and constrain concepts and relationships at the development time (Guarino, 2000). Using CMs, the constraints are basically imposed at run time. Another benefit is that the ontology specification is outside the code, while many object-oriented modeling specifications (especially constraints) are implemented inside the code.

Unfortunately, for the conceptual integration and for the representation of the ISs' dynamics, the existing ontologies have the same limits as the CMs. Most of them are object-oriented, relying on OKBC (open knowledge base connectivity) specification. For building business process ontologies, PSL (process specification language) (initiated by NIST) is recommended (e.g., Gruninger, 2003), mainly because it can be logically integrated with KIF (knowledge interchange format), appropriate for object and knowledge description and exchange.

For *ontology integration*, two alternatives can be considered: (1) by an upper-level ontology, able to represent all aspects in the real world, or (2) by a translation and correlation algorithm between the concepts and rules in different ontologies. Such an algorithm is mostly encoded and the ontology integration is recommended to be accomplished at the development time. So, the first alternative appears as a better solution. For the conceptual integration of all aspects in the real world (and, implicitly, in ISs), *upper-level linguistic ontologies* have been proposed (see next section), with benefits for ontology integration as well.

MAIN THRUST OF THE ARTICLE

This section first gives the limits in the existing research on natural language (NL) translation to conceptual models (CMs) and, lately, to ontologies. Then, the representation of the linguistic meta-models or upper-level ontologies is briefly analyzed from the syntactic and semantic viewpoints.

Most important *limits* of the existing translations of NL patterns to object and process modeling and the basic *open problems* are:

- NL translation only deals with the object-oriented aspects of the real world (RW). The functional, process and organizational aspects are not retrieved or are not well integrated in the translation results. Also, the correlations between the concepts resulted from translation usually confine to the inter-object relationships. The inter-activity ones and the system's dynamic behaviour are not properly treated.
- NL translation of different aspects in RW is accomplished with different software tools and results in different NL-oriented representations (languages). For example, the object-oriented aspects are translated into functional grammar structures or semantic networks and the communication aspects are translated into communicative acts (proposed in the speech act theory). The resulted Iss' representation is still not integrated from the conceptual, notational, semantic and logical viewpoints. The only benefits from the translation are: the automated abstraction of RW from textual requirements; and, a user interface closer to human reasoning and understanding, because it is closer to NL.
- Impossible use of the concepts resulted from NL translation for the expression of coherent and unambiguous ideas on any aspect in RW and IS, mainly because these concepts are heterogeneously represented (see previous). The brainstorming is going to become important in business management (several arguments are in Galatescu & Greceanu, 2002). Today, its automation for a virtual organization confines to the exchange of ideas in NL, by the electronic mail or chat, substantially increasing the virtual traffic. The communication by structures of ideas or modeling structures, issued during brainstorming sessions, and the automated comparison of these structures, must be other objectives of NL-CMs translation.

  Other reasons that make difficult the coherent expression of the modeling ideas using NL translation results are: (a) the intersentential relations in complex sentences and many semantic relations in NL are usually not considered; (b) the resulted concepts representing objects and processes are not stable and general enough. Also, many of them are not known at the development time.
- The logical consistency of the resulted IS representation from NL translation is hardly and incompletely verified at the development time, mainly because of the varied representations and implicitly, formalizations of the different aspects in ISs. This consistency is a pre-condition for the further automatic reasoning, simultaneously on objects, processes, workflows, and so forth.

  The logical consistency of the results from both NL-object model and NL-process/workflow model translation is, generally, proved using the first or higher-order logic or their variants (especially sorted, prepositional, modal, temporal logics). The logical consistency of the resulted process models is incompletely verified at the development time, because the logic of the procedural aspects cannot be represented with these formalisms.
- The meaning of the objects/activities and of the relationships between them in the resulted models from NL translation is, usually, incomplete and is not represented outside the code. This limit impacts on the epistemological aspects of ISs that, lately, are going to be conceived as multi-agent systems. NL translation is mainly concerned with NL syntax and only partially with its semantics.

The *linguistic meta-models* or the *upper-level linguistic ontologies* appear lately a solution to the model (or ontology) integration. Two research directions are important for their representation:

- The abstraction of NL s*emantics*, using a predefined *taxonomy of universal types* of objects, processes, activities and so forth (and relationships among them). This taxonomy is supposed to allow the subsumption (from the semantic point of view) of the words found in NL expressions, belonging to any category (noun, verb, adjective, adverb).
- The abstraction of NL *syntax*, using predefined rules for building *sentence-like structures*, that stylize the NL (simple, compound or complex) sentences and comply with NL syntax. These structures are supposed to be used for the unambiguous

description of any type of object, process, activity and so forth, in the modeled reality, as well as for the representation of coherent and unambiguous ideas about them.

These two directions are complementary and should be both considered in the definition of a linguistic meta-model or ontology.

From the semantic viewpoint, there are several proposals for taxonomies of the concepts in RW. They are compared in Bateman (2003). As *linguistically oriented taxonomies,* the most important are proposed in Oltramari (2002) and Sowa (2000).

From the syntactic viewpoint, the limits (emphasized in the previous section) of the functional grammar, conceptual dependencies and conceptual graphs, used today for NL translation to CMs, impose an improvement with respect to the model integration. Based on activity-oriented conceptual graphs, a new linguistic representation is proposed in Galatescu (2001, 2002), mainly for the model integration purpose. The same representation has been used as an upper-level linguistic ontology for the integration of three ontologies (re-engineering, domain and communication ontologies) and for the ideas expression (Galatescu & Greceanu, 2002).

FUTURE TRENDS

Related to the limits and existing solutions sketched in the previous sections, the *expected results* from the future research on natural language (NL) translation to object and process modeling should be:

- the translation *from textual requirements to* a *uniform representation* of all aspects in information systems (ISs), both externally (during the analysis and design phases) and internally (during the coding phase);
- the unification of all existing proposals for NL abstraction from the semantic point of view and the standardization of a unique *universal taxonomy* for all concepts in the real world and NL;
- the *automatic subsumption* of the words found in the textual requirements (and belonging to any syntactic category) inside a universal taxonomy of the concepts in the real world;
- *ontology-driven modeling* (i.e., based on concepts in a minimal upper-level ontology) (Guarino, 2000), expected to help for the globalization of the modeling activities and for the modeling completeness at the IS analysis or design time;
- the *integration* of the object, functional, process, organizational and epistemological views on organizations by means of linguistic meta-models or upper-level ontologies;
- the linguistic *representation of the epistemological and communication aspects* in IS multi-agent architectures. This representation should be integrated with the linguistic models for objects, processes, workflows and so forth; and
- the *multi-lingual* representation of IS. In this case, a problem in the abstraction of NL syntax will arise and must be solved, due to the differences in the syntactic rules of various languages.

The expected results enumerated previously will have an important impact on the IS lifecycle and on the developers' performance.

With respect to the future *technological means and user interface* for the linguistic conceptual modeling, new aspects should be considered:

- the interfaces of the CASE tools should become natural, close to human understanding. This means a shift from the exclusive symbolic notation to a notation that abstracts NL syntax and semantics; and
- the natural language processing (NLP) technology should be more deeply involved in the resolution of the problems in the conceptual modeling, as a basic translation technology; and
- the future CASE tools should involve practical and theoretical results from other domains, at least from NLP and linguistic ontologies.

CONCLUSION

This article has motivated the research on natural language translation to conceptual models and has reviewed the main limits of the existing models, the open problems and the expected theoretical and practical results (seen as future research trends).

The presentation has focused on the need for the conceptual integration of the existing models and for their representation outside the code. In this respect, as a requirement imposed by the globalization of the organizations, a recent trend was mentioned: the use of linguistic ontologies, able to represent all aspects in the real world (in particular, in any type of organization).

The limits and expectations enumerated for both domains (linguistic modeling and linguistic ontologies) lead to the conclusion that these domains are in an incipient stage, far from results that can be standardized and used in commercial products. The convergence and the integration of research activities and results in several fields (at least, conceptual modeling, NLP, and ontologies)

T

appear necessary for obtaining, as soon as possible, the expected results.

REFERENCES

Allen, J. (1995). *Natural language understanding*. Benjamin/Cummings Publ.

Bateman, J. (2003). *Linguistic and non-linguistic ontologies: Can we find the line between them?* Computational Linguistics Colloquium. Univ. of Saarland.

Fillmore, C., & Hiroaki, S. (2002). *Transparency and building lexical dependency graphs*. 28th Annual Meeting of the Berkeley Linguistics Society.

Fliedl, G., Kop, Ch., Mayerthaler, W., Mayr, H.C., & Winkler, Ch. (2000). Linguistically based requirements engineering – The NIBA Project. *Data & Knowledge Engineering, 35*(2)

Galatescu, A. (2001). *A unifying translation of natural language patterns to object and process modeling. Information modeling in the new millennium*. Hershey, PA: Idea Group Inc.

Galatescu, A. (2002, March). Reifying model integration abilities from natural language. *Journal of Conceptual Modeling*. InConcept, USA. *http://www.inconcept.com*

Galatescu, A., & Greceanu, T. (2002). *Using ontologies in virtual brainstorming for business process reengineering. Collaborative business ecosystems and virtual enterprises*. Kluwer Academic Publishers.

Gruninger, M., Sriram, R.D., Cheng, J., & Law, K. (2003). Process specification language for project information exchange. *Intl. Journal of IT in Architecture, Engineering & Construction.*

Guarino, N., & Welty C. (2000). AAAI-2000 tutorial on conceptual modeling and ontological analysis. *http://www.cs.vassar.edu/faculty/welty/aaai-2000/*

Johannesson, P. (2001). *A language/action based approach to information modeling. Information modeling in the new millennium*. Hershey, PA: Idea Group Publishing.

KBSI. (2000). *A structured approach to enterprise modeling and analysis*. Knowledge Based Systems, Inc. (KBSI). *http://www.idef.com*

Oltramari, A., Gangemi, A., Guarino, N., & Masolo, C. (2002). DOLCE: A Descriptive Ontology for Linguistic and Cognitive Engineering. *Proc. 13th Intl. Conf. on Knowledge Engineering and Knowledge Management*, Spain.

OMG. (2003). *UML 2.0 Specifications*. Object Management Group. *http://www.omg.org/uml/*

Riet, R., Burg, H., & Dehne, F. (1998). *Linguistic instruments in information system design. Formal ontologies in information systems*. IOS Press

Sag, I., & Wasow, T. (1999). *Syntactic theory: A formal introduction*. Stanford Univ., Center for the Study of Language and Information. *http://www-csli.stanford.edu/csli/projects/cogsci9495-hpsg.html*

Sowa J.F. (2000). *Knowledge representation – Logical, philosophical and computational foundation*. Pacific Grove, CA: Brooks Cole Publishing Co.

Steuten, A., Riet, R., & Dietz, J. (2000). Linguistically based conceptual modeling of business communication. *Data & Knowledge Engineering,* 35-42.

WfMC. (2003). *Workflow Reference Model*. Workflow Management Coalition. *http://www.wfmc.org/standards/model.htm*

KEY TERMS

Business Process: The sequence of activities, the people and the technology involved in carrying out some business or achieving some desired results in an organization.

Category in NL: Noun, verb, adjective, adverb. Nouns are described by noun head, substitute, determiner, and modifier. Verbs are described by verb determiner/ modifier.

Conceptual Model: Abstraction of the real world/ domain and a mechanism for understanding and representing organizations and the information systems that support them. The most important *types of models* are:

- *object model*: describes objects by data and operations on the data. The object's identity encapsulates its state (attributes and relationships with other objects) and its behaviour (allowed operations on/with that object).
- *process model*: describes (sub)processes by the activities they involve, the activity order, decision points, pre-/post-conditions for the activity execution;
- *functional model*: describes the information flow and transformation, as well as the constraints and functional dependencies among the activities in a process;

- *organizational model*: describes the workflow (activities for the creation and movement of the documents) within an organization, the people's roles and the communication among people for performing the activities.

Object: An entity (e.g., "person") or a value (e.g., "phone number"). In object-oriented models, it is the instance of a class.

Ontology: Hierarchical structuring of knowledge about things by subcategorizing them according to their essential (or at least relevant and/or cognitive) qualities. Practically, a *vocabulary* used to describe a certain reality, plus a set of *explicit assumptions* regarding the intended meaning of the concepts in the vocabulary. The relations between concepts allow inferences (e.g., information interpretation and the derivation of new information/knowledge). The explicit axioms allow the approximation of the term meaning and the validation of the ontology specification at the development time.

Semantic Relationship in NL: Relation between the words in the same category (mainly, noun and verb category). Some relationships are bi-directional. Most important relationships are: *noun* or *verb* meronymy/ holonymy, hyponymy/ hypernymy, synonymy, antonymy, homonymy; *verb* troponymy or entailment; cause-effect relations between verbs (see the glossary of terms for WordNet, at *http://www.cogsci.princeton.edu/~wn/man/ wngloss.7WN.html*).

Sentences in NL: *Simple sentence***:** built around the static or dynamic action of the verb; *compound sentence*: joins independent simple sentences by conjunctions; *complex sentence*: composed of subordinated sentences (subclauses) correlated to a main sentence (clause).

Structure in NL: *Sentence* (simple, compound, complex); *phrase* (e.g., verb, noun phrase, adverb, prepositional, adjectival phrase); *clause* (noun, adverbial, relative, appositive).

Traversal Pattern Mining in Web Usage Data

Jenq-Foung (J.F.) Yao
Georgia College & State University, USA

Yongqiao Xiao
SAS Co., USA

INTRODUCTION

Web usage mining is designed to discover useful patterns in Web usage data, i.e., Web logs. Web logs record the user's browsing of a Web site, and the patterns provide useful information about the user's browsing behavior. Such patterns can be used for Web design, improving Web server performance, personalization, etc.

BACKGROUND

Pattern discovery applies methods and algorithms from different fields such as statistics, data mining, machine learning, etc., to the prepared data. By applying the statistical techniques to the Web usage data, some useful statistics about the users' browsing behavior can be obtained, e.g., the average view time of a page, and the most frequently accessed pages, etc. By applying clustering (a.k.a., unsupervised learning) to the usage data, groups of users which exhibit similar browsing behavior can be found. Such knowledge is useful for market segmentation and personalization. The focus here is on discovering traversal patterns from the Web usage data.

A traversal pattern is a list of pages visited by a user in one session. Several different traversal patterns and the corresponding methods of discovering them have been proposed in the literature, namely, Association Rules, Sequential Patterns, Frequent Episodes, Maximal Frequent Forward Sequences, and Maximal Frequent Sequences. The details about each type of traversal pattern are described in the following section.

TRAVERSAL PATTERNS AND DISCOVERING METHODS

Association Rules

Association rules were originally proposed for market basket data (Agrawal et al., 1993; Agrawal and Srikant, 1994). Association rules describe the associations among items bought by customers in the same transaction, e.g., 80% of customers who bought diapers also bought beer in some store.

To mine association rules from the transactions, there are two steps: first finding the frequent item sets, and then generating association rules from the frequent item sets. Since the second step is straightforward compared to the first one, the research focus is on the first step. An item set (or itemset for short) is frequent, if the support for the itemset is not less than some predefined threshold. The support for an itemset in a database of transactions is defined as the percentage of the transactions that contain the itemset.

Sequential Patterns

Sequential patterns (Srikant and Agrawal, 1997) were also originally proposed for market basket data. For example, customers buy a digital camera, then a photo printer, and then photo papers. Such sequential patterns capture the purchasing behavior of customers over time.

Sequential patterns have also been applied to Web logs (Buchner et al., 1999; see also Spiliopoulou, 2000; Pei et al., 2000). The sessions are ordered by the user id and the access time. As for association rules, the duplicate pages are discarded. Then for each user, there is a user sequence, which consists of all sessions of the user. A sequential pattern is a maximal sequence of itemsets whose support is not less than some predefined threshold. A sequence is maximal if it is not contained in any other sequence. The support of a sequence is the percentage of user sequences that contain the sequence.

Algorithm AprioriAll was proposed in Srikant and Agrawal (1995) for finding all sequential patterns given some support threshold. AprioriAll was then improved by Generalized Sequential Patterns (GSP) (Srikant and Agrawal, 1996). Traversal patterns were generalized to allow time constraints, sliding time window, and user-defined taxonomy.

Frequent Episodes

Frequent episodes were originally proposed for telecommunication alarm analysis (Mannila et al., 1997). Episodes are collections of events, which occur together within some time window. In general, they are partially ordered sets of events. There are two special types of episodes: parallel episodes and serial episodes. They differ in whether the events in the episodes are ordered. In parallel episodes the events are not ordered, while in serial episodes the events are ordered sequential. An episode is frequent if it occurs in the event sequence not less than some predefined threshold.

Frequent episodes were also applied to Web logs (Mannila et al., 1997). The clicks (pages) correspond to events. They are ordered by the access time, and usually the users need not be identified, i.e., there are no sessions.

Maximal Frequent Forward Sequences

Maximal Frequent Forward Sequences (MFFS for short) were proposed in Chen et al. (1998) Notice that MFFS was referred to as large reference sequence in Chen et al. (1998). An MFFS describes the path traversal behavior of the user in a distributed information-providing environment like World Wide Web. There are two steps to mine MFFSs from the sessions. First each session is transformed into maximal forward sequences (i.e., the backward traversals are removed). The MFFSs are then mined using level-wise algorithms (Park et al., 1995) from the maximal forward sequences.

In the raw sessions, there are often backward traversals made by the user. A backward traversal means revisiting a previously visited page in the same user session. It is assumed that such backward traversals happen only because of the structure of the Web pages, not because the user wants to do this. When a backward traversal occurs, a forward traversal path terminates. This resulting forward traversal path is called maximal forward sequence. It then backtracks to the starting point of the next forward traversal and resumes another forward traversal path.

An MFFS is a traversal sequence (consecutive subsequence of a maximal forward sequence) that appears not less than some predefined threshold in the set of maximal forward sequences. The pages in an MFFS are required to be consecutive in the maximal forward sequences, and an MFFS is also maximal, which means that it is not a subsequence of any other frequent traversal sequence.

Maximal Frequent Sequences

Maximal Frequent Sequences (MFS) were proposed in Xiao and Dunham (2001). In contrast to maximal frequent forward sequences, MFSs do not remove backward traversals from the sessions. It was argued in Xiao and Dunham (2001) that such backward traversals are useful for discovering the structures of the Web pages. For example, if a pattern <A, B, A, C> is found frequent, it may suggest that a direct link from page B to page C is needed, while the resulting maximal forward sequences <A, B> and <A, C> lose such information.

An MFS is a traversal sequence (consecutive subsequence of a session) that appears not less than some predefined threshold. Since the backward traversals are kept in the sessions, a traversal sequence may occur in a session more than once. In order to measure the actual number of occurrences of a traversal sequence, the support of an MFS is defined as the ratio of the actual number of occurrences to the total length of all sessions. The length of a session is the number of clicks in the session. The pages in an MFS are required to be consecutive in the sessions, and an MFS is also maximal.

Summary

Table 1 compares the different types of traversal patterns by the following features:

- Ordering: the pages in a traversal pattern can be ordered or not.
- Duplicates: which indicate whether backward traversals are allowed in the traversal pattern.
- Contiguity: the page references in a traversal pattern may be contiguous or not.
- Maximality: a frequent pattern is maximal if it is not contained in any other frequent pattern. A pattern could be maximal or not.

Notice that for frequent episodes, parallel episodes are not ordered, while serial episodes are ordered and the general episodes are partially ordered. Due to the different features of the traversal patterns, the support for each type of pattern is defined quite differently, which is also shown in Table 1.

These features used by different patterns can be used for different purposes. Backward traversals capture the structure information of the Web, and therefore can be used to improve the design of Web pages by adding new links to shorten future traversals. The maximality feature can reduce the number of meaningful patterns discovered. The contiguity and the ordering features could be used to predict future references and thus for prefetching and caching purposes.

These traversal patterns uncover the associations or sequences among the Web pages browsed by the user. They can be used together with other data mining techniques, such as classification and clustering, to further

Table 1: A comparison of traversal patterns

	Ordering	Duplicates	Contiguity	Maximality	Support
Association Rules	N	N	N	N	freq(X) / # sessions
Sequential Patterns	Y	N	N	Y	freq(X) / # users
Frequent Episodes	Y	N	N	N	freq(X) / # time windows
MFFS	Y	N	Y	Y	freq(X) / # maximal forward sequences
MFS	Y	Y	Y	Y	freq(X) / # clicks

facilitate Web usage mining as shown in Tan and Kumar (2002). In that paper, the authors used traversal patterns and classification to distinguish human sessions from robot sessions.

Pattern analysis is to analyze the patterns to the needs of the application. It includes filtering uninteresting patterns, visualization, etc. The Web Utilization Miner (WUM) system (Spiliopoulou and Faulstich, 1998) provides a mining language, which allows an analyst to specify characteristics of the patterns that are interesting.

Web usage mining has been applied to many applications (Srivastava et al., 2000), e.g., personalization, system improvement, Web design, business intelligence, etc. Specifically, the traversal patterns have been used for site evaluation (Spiliopoulou, 2000), prefetching (Lan et al., 1999; Pandey et al., 2001) network intrusion detection (Lee and Stolfo., 1998; Dokas et al., 2002), fraud detection (Colet, 2002), etc. Interesting results might be obtained from a comparison study among these different patterns.

FUTURE TRENDS

One interesting future research topic is to mine traversal patterns with Web document structure also captured. All the above traversal patterns ignore the document structure of the Web site, which is essentially hierarchical (a tree) or a graph. Techniques for tree mining or graph mining can then be adopted.

Other interesting work could include improving sessionizations, clustering of the traversal patterns, etc.

CONCLUSION

Five different types of traversal patterns that can be mined in the Web usage data, namely, Association Rules, Sequential Patterns, Frequent Episodes, Maximal Frequent Forward Sequences, and Maximal Frequent Sequences are described. These patterns are compared and contrasted in four aspects, i.e., ordering, duplicates, contiguity, and maximality. They can be used for different purposes.

REFERENCES

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules in large databases. In *Proceedings of the 20th International Conference on Very Large Databases,* Santiago, Chile (pp. 487-499).

Agrawal, R., Imielinski, T., & Swami, A. N. (1993). Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD international conference on management of data, Washington, DC* (pp. 207-216).

Buchner, A. G., Baumgarten, M., Anand, S. S., Mulvenna, M. D., & Hughes, J. G. (1999). Navigation pattern discovery from Internet data. In *Workshop on Web Usage Analysis and User Profiling (WEBKDD-99)*.

Chen, M. S., Park, J. S., & Yu, P. S. (1998). Efficient data mining for path traversal patterns. *IEEE Transactions on Knowledge and Data Engineering*, *10*(2), 209-221.

Colet, E. (2000). Using data mining to detect fraud in auctions. *DSStar*, *4*(26). Retrieved June 27, 2000 from *http://www.tgc.com/dsstar/00/0627/000627.html*

Dokas, P., Ertoz, L., Kumar, V., Lazarevic, A., & Srivastava, J. (2002). Data mining for network intrusion detection. In *Proceedings of the National Science Foundation Next Generation Data Mining Workshop,* Baltimore, Maryland.

Lan, B., Bressan, S., & Ooi, B. C. (1999). Making web servers pushier. In *Workshop on Web Usage Analysis and User Profiling (WEBKDD-99)*.

Lee, W., & Stolfo, S. (1998). Data mining approaches for intrusion detection. In *Proceedings of the 7th USENIX Security Symposium*. In H. Mannila, H. Toivonen, & A.

Verkamo(Eds.), Discovering frequent episodes in sequences. In *Proceedings of the 1st International Conference on Knowledge Discovery and Data Mining (KDD-95)* (pp. 210-215).

Mannila, H., Toivonen, H., & Verkamo, A. I. (1997). *Discovering frequent episodes in event sequences*. Helsinki, Finland: University of Helsinki, Department of Computer Science. (Technical Report C-1997-15).

Pandey, A., Srivastava, J., & Shekhar, S. (2001). A web intelligent prefetcher for dynamic pages using association rules - a summary of results. In *Proceedings of the SIAM Workshop on Web Mining*.

Park, J. S., Chen, M.-S., & Yu, P. S. (1995). An effective hash-based algorithm for mining association rules. In M. J. Carey & D. A. Schneider (Eds.), *Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data*, San Jose, California (pp.175-186).

Pei, J., Han, J., Mortazavi-Asl, B., & Zhu, H. (2000). Mining access patterns efficiently from web logs. In *Proceedings of the 2000 Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD-00)*.

Spiliopoulou, M. (2000). Web usage mining for web site evaluation. *Communications of the ACM*, 43(8), 127-134.

Spiliopoulou, M., & Faulstich, L. (1998). WUM: A web utilization miner. In *Proceedings of the EDBT Workshop WebDB98*. Valencia, Spain: Springer-Verlag.

Srikant, R., & Agrawal, R. (1995). Mining generalized association rules. In *Proceedings of the 21st International Conference on Very Large Databases*, Zurich, Switzerland (pp. 407-419).

Srikant, R., & Agrawal, R. (1996). Mining sequential patterns: Generalizations and performance improvements. In *Proceedings of the 5th International Conference on Extending Database Technology*, Avignon, France.

Srikant, R., Vu, Q., & Agrawal, R. (1997). Mining association rules with item constraints. In *Proceedings of the American Association for Artificial Intelligence*.

Srivastava, J., Cooley, R., Deshpande, M., & Tan, P. N. (2000). Web usage mining: Discovery and applications of usage patterns from web data. *SIGKDD Explorations*, 1(2), 12-23.

Tan, P., & Kumar, V. (2002). Discovery of web robot sessions based on their navigational patterns. *Data Mining and Knowledge Discovery*, 6(1), 9-35.

Xiao, Y., & Dunham, M. H. (2001). Efficient mining of traversal patterns. *Data and Knowledge Engineering*, 39, 191-214.

KEY TERMS

Cleansing: To filter the irrelevant entries in the Web log, such as graphics files. The HTTP protocol is stateless, which requires a separate connection for each file that is requested from the Web server. Therefore, several log entries may result from a request to view a single page, since the files for the graphics embedded in the page are automatically downloaded from the Web server. Such filtering can be done by checking the suffixes of the URI name such as jpg, gif, etc.

Extraction: To select the related fields from the Web logs. Traversal patterns typically require the three fields: IP address, access time, and the page accessed. Other fields, such as referrer and user agent, can be used in cleansing and sessionization.

Preprocessing: The tasks of preprocessing for Web usage mining include extraction, cleansing, transformation, sessionization, etc.

Server Session: A collection of user clicks to a single Web server during a user session.

Sessionization: Identifying the sessions (server sessions) from the Web logs, called sessionization and is not a trivial task. It has two subtasks: user identification and session identification. A common way is to use the IP address to identify the user, i.e., all entries with the same IP address belongs to the same user.

Transformation: To convert the fields to the format required by specific pattern discovery algorithm.

User: An individual that is accessing the Web through a browser.

User session: A delimited set of user clicks across one or more Web servers. A click corresponds to a page on the Web server, which is uniquely identified by an URI (Universal Resource Identifier).

T

Trends and Perspectives in Online Education

Bruce Rollier
University of Baltimore, USA

Fred Niederman
Saint Louis University, USA

INTRODUCTION

Although the Internet has been in existence since 1969, it was not widely used for educational purposes in its first two decades. Few students had access to e-mail, and few educators could visualize its value as a teaching tool. Programs to serve students from remote locations, often called "distance education," became popular; these were generally delivered synchronously through television broadcasts and did not involve the Internet. When the World Wide Web was created in the early 1990s (Berners-Lee, 1999) and the first browsers became available (Waldrop, 2001), the enormous potential for education began to be recognized. New global users came online at a fantastic pace, and the value of all this connectivity was increasing even more rapidly in accordance with Metcalf's Law (Gilder, 1996). Nearly all students used e-mail regularly, and college professors were putting syllabi and course assignments online and creating Web pages with increasing sophistication. Soon entire programs were offered completely via the Internet, with students from all over the globe taking courses together.

According to a major survey conducted by Allen and Seaman (2003), there were more than 1.6 million students who took at least one online course during the Fall 2002 semester and 11% of U.S. higher education students took at least one such course. These numbers were projected to increase rapidly, and most institutions considered online education as a critical long-term strategy.

BACKGROUND

At first, online courses tended to be offered sporadically, often by a few technically savvy faculty members exploring how best to use the Web. Developing the courses to fit this new medium was difficult and time-consuming, and these professors began to demand some recognition for their efforts. Soon, university administrators noticed that the online courses were very popular, that they could attract students from distant locations, and that programs in other institutions were proliferating. The need for uniform policies was recognized. What is a reasonable class size? Should development of a new Web course count toward promotion and tenure? Should Web courses be taught the same way as face-to-face classes?

It soon became clear that there were major differences between face-to-face classes and online classes in providing high quality instruction. Online, at least with the currently available technologies, must be almost completely asynchronous, whereas face-to-face is primarily synchronous. The term "virtual university" (Morrissey, 2002; Schank, 2002; Stallings, 2001) has recently come into vogue, but with varying meanings. The term is often applied to an institution with no physical campus and that is completely online; Jones International University, Cardean University, and University of Phoenix Online are examples (Mason, 2000). These are usually for-profit companies and may be spin-offs from private universities. The term can also signify an institution that has a large physical campus but that has a coordinated approach to online education and devotes significant resources to its online programs. Some of the largest of these include Penn State's World Campus and the University of Maryland's University College (Stallings, 2001). There are also a large number of Virtual University Consortiums (VUCs) that have been formed to offer online programs in multi-institutional partnerships. A recent study identified 63 of such consortia, primarily state-based (Twigg, 2003).

There are many reasons for the rapid growth of VUCs. In many countries, including the U.S., governmental funding of education has decreased in recent years, forcing universities to look for new markets and new sources of revenue (Mason, 2000). Lifelong learning has become much more critical for the workforce as skills and careers become obsolete at a rapid clip. There is high demand from adult learners for retraining for new careers or for upgrading of skills, and online programs are ideal for this.

Twigg (2003) lists several factors motivating the growth of VUCs and other online programs: increased demand for adult education and training; the educational needs of underserved communities; coping with the increasing frequency of interuniversity transfers; streamlined access to the state's institutions via a portal; providing increased variety of degrees; lowering costs, and "overcoming the possibility that the state's institutions will be left behind in the new, highly competitive online environment" (p. 5). She comments that most states have decided that "stand-alone virtual university initiatives are too expensive to initiate and sustain both fiscally and politically" (p. 7). Most VUCs do not offer degrees because that would put them in direct competition with the other state institutions. Instead, they have adopted the collaborative model, which can mean anything from a passive posting of available courses or library sharing to a fairly aggressive stance as an alternative degree path.

Twigg notes that the most successful VUCs have adopted the following policies (2003, p. 10): (1) focus on increasing access for new students that might not otherwise attend; (2) find out what students need and create a viable response rather than merely aggregating what the member institutions can offer; (3) do not get involved in irrelevant higher education policy issues; (4) create a business plan for long-term viability without reliance on state aid, and (5) use a cost-effective development and delivery model.

Morrissey (2002, pp. 460-461) cites the following conflicts which may inhibit achieving the virtual university's full potential: compensation and ownership issues; lack of recognition for course development; poor course support; push for large classes, which may result in fewer faculty positions, and possible adverse impact on research.

ISSUES OF QUALITY

The societal benefits of a first class online education program are obvious, in providing quality instruction for those in remote locations or the physically handicapped or those whose daily responsibilities do not permit them to attend on-campus classes. It is not clear that the programs are being established for such altruistic objectives. Schank (2002) is skeptical, and suspects that Web courses are developed primarily for revenue and academic prestige. He also believes that courses are being modeled to fit existing programs and are not taking advantage of the unique characteristics of the Web.

Are online courses equivalent in quality to courses in face-to-face classrooms? Quality is of course dependent on many factors, and each medium has both advantages and disadvantages. A major difference is synchrony; physical classes are largely synchronous, and current virtual courses are almost completely asynchronous. This characteristic severely limits student interaction and team activities. Chat rooms are synchronous, but quickly become confusing. On the other hand, asynchrony makes it possible to maintain availability 24 hours a day, seven days a week (Pittinsky, 2002). Schank (2002) argues that Web courses would be superior to face-to-face if the design were less dependent on our current concepts, such as performing tasks rather than listening, and length of course and material covered based on student need.

Spicer (2002) warns university presidents that they must be prepared to expend significant resources over long time periods to maintain high quality and avoid shortcuts. The view of online education as a source of revenue and profit is misguided, and costs are typically underestimated (Niederman & Rollier, 2001). Teaching Web classes is difficult and labor-intensive, and huge classes are not feasible if quality is to be preserved. The required faculty skills are not the same as in the classroom, and the necessary faculty development is often overlooked (Agee, Holisky & Muir, 2003). Procedures must be established and controls enforced to ensure that instructors are responding quickly to student needs and evaluating student learning effectively (Hall, 2002; Stein, 2001).

FUTURE TRENDS

Market share is much more important on the Web than in traditional businesses. A large firm has a major marketing advantage, especially in advertising. The largest virtual universities of the future will offer programs in almost every discipline, at every level, and will compete for students in every locality and most languages. The competition is global; note that the largest distance education institution in the world is in China (Dunn, 2000). These schools will gradually lose their identification with a particular country. Most universities will adopt a niche strategy, competing in those markets for which they perceive a competitive advantage, or markets thought to be underserved.

With continuing advances in technology it seems likely that Web teaching techniques will change greatly. Online education will continue to be basically asynchronous, but with the availability of synchronous capabilities to provide far more interaction. When a critical mass of online students has access to broadband, the instructor will be able to see the students on screen, and the students to see the instructor. Highly interactive group assignments will be possible. The instructor will be able to demonstrate concepts and skills visually and audibly rather than just describing them. Examinations can be monitored visually. Although keyboards may still be necessary, voice will be the primary means of communication, thus reducing the importance of typing skill. Simulations and virtual reality will be widely employed for user interaction and feedback (Morrison & Aldrich, 2003).

Wireless will be a major communications technology (PriceWaterhouse Coopers, 2001), greatly increasing flexibility of location. Internet2 will make digital video available, bringing large-scale indexed video archives and digital-video conferencing. As Tsichritzis (1999) points out, professors can deliver lectures to more than one university at the same time with little additional overhead. Universities may become increasingly like brokers, providing facilities for buyers and sellers of educational services to meet virtually.

Mergers, joint ventures, and strategic alliances will be increasingly common (Makri, 1999). A growing number of states are centralizing services for distance education (Morrissey, 2002), claiming that such collaborations save money and provide more choices for students. Collaborations between universities in different countries can provide obvious benefits to both. There is an increasing trend for students to take courses at more than one university, and online education seems likely to accelerate that trend. Such alliances will lead to an emphasis on standardization of curricula and on ease of transfer between institutions.

Peter Drucker (2000) has said: "Online Continuing Education is creating a new and distinct educational realm, and it is the future of education" (p. 88). If high quality can be maintained and innovative approaches are developed to take advantage of the wealth of resources available on the Web, the chief beneficiary may be society itself, and particularly the Third World societies. Quality education at reasonable cost can be made available virtually anyplace in the world. As the population ages in the wealthier countries, a steady supply of skilled workers will be needed, but an increasing proportion of younger workers will be from economic classes whose members do not traditionally attend college. At the same time, most of the fastest growing occupations require an increasingly high level of education (BLS, 2004). In Third World countries, there is a critical need for an educated workforce and for high quality instruction and effective educational materials. Online education can potentially provide all of these elements.

CONCLUSION

Online education is here to stay, and seems certain to be an increasingly important component of every university's future. It will also be increasingly competitive, and to survive in this market institutions will have to invest significant resources to achieve high quality and favorable name recognition. The ultimate arbiters of success will be the global employers of college graduates. If the online students prove to be competent employees with knowledge and skills equivalent to the graduates of traditional programs, the students will land good jobs and others will flock into the online programs. If the online graduates are seen as weak, this will have a disastrous effect on program success. Skimping on quality would be a huge mistake.

Surveys such as that of Allen and Seaman (2003) indicate that the perception of quality is high among the institutions who offer the programs, but this is weak evidence; student and employer perceptions are much more significant. Research is critically needed to determine whether the necessary quality is being achieved, whether the students are being prepared for successful careers, whether some students are better candidates for online education than others, and whether college courses are equally adaptable to online delivery.

It should not be assumed that the courses will be taught the same way in the future as they are today. Wide availability of broadband and more sophisticated software will make possible much greater interaction among students and between students and instructors. Successful programs will incorporate the best blend of synchrony and asynchrony. Staying competitive may prove to be very expensive, and less well-endowed institutions may be wise to seek out strategic partnerships for sharing of faculty and other resources.

REFERENCES

Agee, A.S., Holisky, D.A., & Muir, S.A. (2003, September/October 1-6). Faculty development: The hammer in search of a nail. *The Technology Source, Faculty and Staff Development.*

Allen, I.E., & Seaman, J. (2003). *Sizing the opportunity: The quality and extent of online education in the United States, 2002 and 2003.* Sloan Center for Online Education.

Berners-Lee, T. (1999). *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor.* San Francisco: Harper.

BLS. (2004). Employment projections. Bureau of Labor Statistics. *http://www.bls.gov/emp/home.htm*

Drucker, P.F. (2000, May 15). Putting more now into knowledge. *Forbes,* 88.

Dunn, S.L. (2000, Spring). The virtualizing of education. *The Futurist, 34*(2), 34-38.

Gilder, G. (1996). *Telecosm.* New York: Simon & Schuster.

Hall, R. (2002). Aligning learning, teaching and assessment using the Web: An evaluation of pedagogic approaches. *British Journal of Educational Technology, 33*(2), 149-158.

Makri, M. (1999). Exploring the dynamics of learning alliances. *The Academy of Management Executive, 13*(3), 113-114.

Mason, R. (2000, November). Wiring up the ivory towers. *The Unesco Courier,* 31-32.

Morrison, J.L., & Aldrich, C. (2003, September/October). Simulations and the learning revolution: An interview with Clark Aldrich. *The Technology Source,* Vision section, 1-7.

Morrissey, C.A. (2002). Rethinking the virtual university. *Communications of AIS, 9,* 456-466.

Niederman, F., & Rollier, B. (2001). How are you going to keep them in the classroom after they've seen MTV? Online education in a virtual world. In L. Chidambaram & I. Zigurs (Eds.), *Our virtual world: The transformation of work, play, and life via technology* (pp. 56-73). Hershey, PA: Idea Group Publishing.

Pittinsky, M.S. (2002). Transformation through evolution. In M. Pittinsky (Ed.), *The wired tower: Perspectives on the impact of the Internet on higher education.* Upper Saddle River, NJ: Prentice-Hall.

PriceWaterhouseCoopers. (2001). *Technology forecast: 2001-2003.* Menlo Park, CA: Pricewaterhousecoopers Technology Centre.

Schank, R. (2002). The rise of the virtual university. *The Quarterly Review of Distance Education, 3*(1), 75-90.

Spicer, D.V. (2002). Where the rubber meets the road: An on-campus perspective of a CIO. In M. Pittinsky (Ed.), *The wired tower: Perspectives on the impact of the Internet on higher education.* Upper Saddle River, NJ: Prentice-Hall.

Stallings, D. (2001). The virtual university: Organizing to survive in the 21st century. *The Journal of Academic Librarianship, 27*(1), 3-14.

Stein, B. (2001). Benchmarks for virtual learning. *NEA Today, 19*(7), 21-23.

Tscichritzis, D. (1999). Reengineering the university. *Communications of the ACM, 42*(6), 93-100.

Twigg, C.A. (2003). *Expanding access to learning: The role of virtual universities.* Monograph. Center for Academic Transformation. Rensselaer Polytechnic Institute.

Waldrop, M.M. (2001). *The dream machine.* New York: Viking.

KEY TERMS

Asynchrony: A condition whereby events occur that are not coordinated in time. In online education, asynchrony makes possible performing course tasks at the most convenient time, not tied to a schedule.

Distance Education: A somewhat broader term that includes online courses but also classes that involve broadcasting from a television-like facility, often transmitted via satellite.

Metcalf's Law: The "value" or "power" of a network increases in proportion to the square of the number of nodes on the network.

Online Education: Instruction conducted on a network, usually the Internet, for students separated geographically. Sometimes called *distance education*.

Synchrony: Simultaneous occurrence of two or more events. Synchronous features would make it possible to perform lectures, demonstrations, and group activities in real time.

Virtual: Being such in force and effect, though not actually or expressly such.

Virtual University: A loosely defined term, but usually refers to a large, coordinated, degree program offered by one institution or a consortium in which students can take classes at any time and from any global location.

Virtual University Consortium (VUC): A group of higher education institutions, usually sponsored by the state, who have formed a strategic partnership to jointly offer online programs.

Trends in Information Technology Governance

Ryan R. Peterson
Information Management Research Center, Spain

INTRODUCTION

Information technology (IT) *governance* has been a perennial item on the corporate agenda of many organizations. Ever since IT proved to be more than an administrative tool, researchers and practitioners have pondered its governance. Defined as the locus of IT decision-making authority (Brown & Magill, 1994; Sambamurthy & Zmud, 1999), discussions concerning IT governance have flourished for more than four decades across research communities and boardrooms. Posed as a question of *centralization* during the 70s, IT governance drifted towards *decentralization* in the 80s, and the recentralization of IT decision-making was a 90s trend.

Today, IT governance is experiencing yet another transformation, and persists as a complex and evolving phenomenon (Grembergen, 2003). As business environments continuously change and new technologies evolve rapidly, how to govern IT effectively remains an enduring and challenging question. This chapter discusses past developments and the present status quo of IT governance, and outlines several critical questions, which are pending future investigation.

BACKGROUND

Traditionally, three IT governance models have been distinguished (Brown & Magill, 1998; Sambamurthy & Zmud, 1999). In each model, stakeholder constituencies take different lead roles and responsibilities for IT decision-making across the *IT portfolio*. In the centralized model, corporate IT management has decision-making authority concerning *IT infrastructure* and *IT applications*. In the decentralized model, division IT management and business management have authority for IT infrastructure and IT applications. In the *federal* model, corporate IT has authority over IT infrastructure, and (either or both) division IT and business-units have authority over IT applications.

In general, it is argued that centralization provides greater efficiency, control, and standardization, while decentralization improves business ownership, flexibility, and responsiveness (Brown, 1997; Rockart, Earl, & Ross, 1996). Literature suggests that the federal model provides the benefits of both centralization and decentralization (see Table 1). Research indicates that organizations adopt a federal model when pursuing multiple

Table 1. Drivers and design of IT governance (Adapted from Hodgkinson, 1996; Peterson, O'Callaghan, & Ribbers, 2000; Sambamurthy & Zmud, 1999)

Model / Drivers	Centralized IT Governance	Decentralized IT Governance	Federal IT Governance
Synergy	+	-	+
Standardization	+	-	+
Specialization	+	-	+
Customer responsiveness	-	+	+
Business ownership	-	+	+
Flexibility	-	+	+

competing objectives involving a simultaneous focus on cost-efficiency and business-flexibility (Peterson, O'Callaghan, & Ribbers, 2000; Sambamurthy & Zmud, 1999).

MAIN THRUST

While the federal model seems to be the dominant configuration in contemporary firms (Peterson, O'Callaghan, & Ribbers, 2000; Sambamurthy & Zmud, 1999), empirical studies regarding the complexity of this configuration are sparse. Specifically, *allocation of IT decision-making authority does not resolve the need for effective coordination between corporate IT, division IT and business-unit management*. Continuous differentiation leads to fragmentation, unless a corresponding process of integration complements it. The problems reported in practice and research regarding the lack of, for example, IT prioritization, top management IT commitment, IT management business understanding, business management IT responsibility, and IT value generation, are symptomatic of this fragmentation and are typically encountered in the federal IT governance model (Peterson, 2001; Weill & Broadbent, 1998).

In order to provide direction and achieve organizational effectiveness, differentiation begets integration (Daft, 1998; Galbraith, 1994; Lawrence & Lorsch, 1967). Designing effective IT governance is dependent on both the *differentiation* and *integration* of decision-making for IT across the portfolio of business IT investments and processes (see Figure 1).

Whereas differentiation focuses on the distribution of IT decision-making rights and responsibilities among different stakeholders in the organization (i.e., the locus of IT decision-making), integration focuses on the coordination of IT decision-making/-monitoring processes and structures across stakeholder constituencies. Organizations thus need to consider and implement integration mechanisms for the effective governance of IT.

FUTURE TRENDS

Integration mechanisms for IT governance can be classified according to two dimensions (Peterson, 2003). Vertically, integration mechanisms focus either on integration structures or integration processes; whereas horizontally, a division is made between formal positions and processes, and relational networks and capabilities. Col-

Figure 1. Differentiation and integration IT decision-making (Adapted from Weill & Broadbent, 1998)

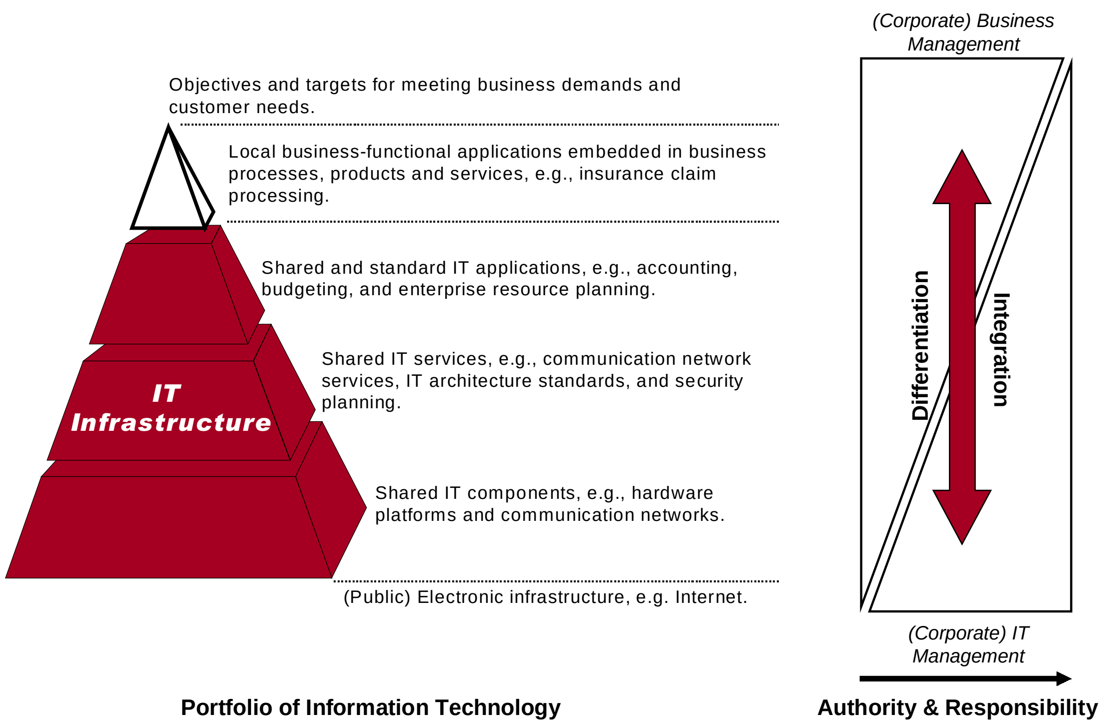
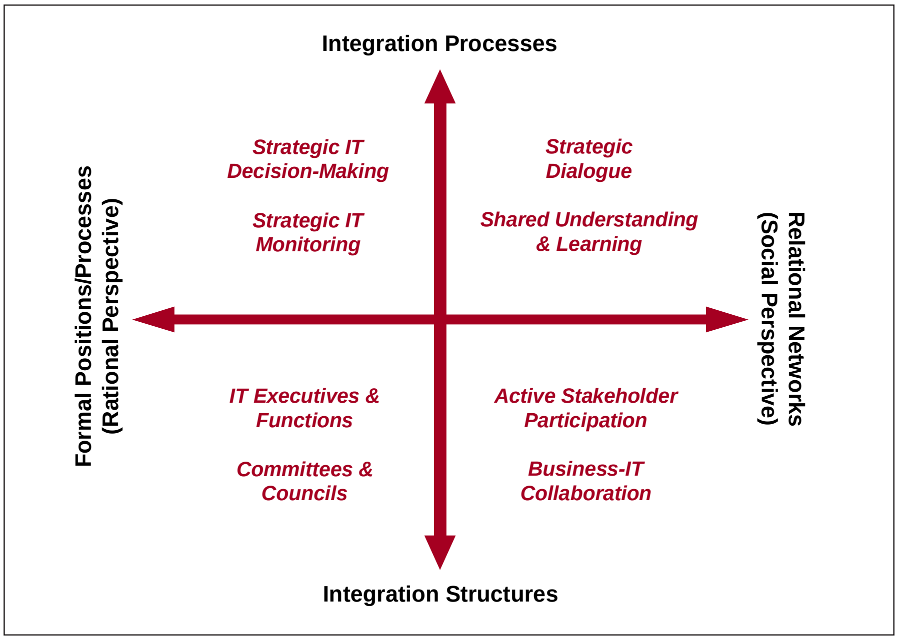


Figure 2. Generic integration mechanisms for IT governance (Adapted from Peterson, 2003)



lectively, this provides four types of generic integration mechanisms for IT governance (see Figure 2).

Formal integration structures involve appointing IT executives (e.g., CIO) and IT functions (e.g., client-account and user relationship managers), and institutionalizing special and standing IT committees and councils. Committees and/or executive teams can take the form of temporary task forces (e.g., project steering committees), or can alternatively be institutionalized as an overlay structure in the organization in the form executive and/or IT management councils. Formal integration processes describe the formalization and institutionalization of IT decision-making-/monitoring procedures and performance. Formal integration processes vary with levels of IT governance *comprehensiveness*, *formalization*, and *integration*.

Whereas the foregoing formal integration mechanisms tend to be mandatory and tangible, relational integration mechanisms are "voluntary" and "tacit" actions, which cannot be programmed and/or formalized. While formal integration mechanisms are necessary, they are insufficient for designing effective IT governance in competitive environments (Peterson, O'Callaghan, & Ribbers, 2000).

Relational integration structures involve the active *participation* and *collaboration* between corporate executives, IT management, and business management. Central to relational integration is the participative behavior of different stakeholders to clarify differences and solve problems in order to find integrative solutions. The ability to integrate relationally allows an organization to find broader solutions, and unleashes the creativity involved in joint exploration of solutions that transcend functional boundaries and define future possibilities. Relational integration processes describe strategic dialogue and *shared learning* between principle business and IT stakeholders. Strategic IT dialogue incorporates a wide range of initially unstructured business perspectives and IT views, and involves rich conversation to resolve diverging perspectives and stakeholder conflicts. Shared learning describes the co-creation of mutual understanding by members of organizational sub-units of each other's goals and objectives.

In summation, formal and relational structures and processes collectively constitute and determine the integration capability of IT governance, and underscore the importance of flexible management systems in complex and uncertain environments. The organizing logic is characterized by a collaborative network, where communication is more likely to be lateral, task definitions are more fluid and flexible (i.e., related to competencies and skills, rather than being a function of position in the organization), and where influencing of business-IT decisions is based on expertise, rather than an individual's (or group's) position (Peterson, 2003).

Table 2. Future research

Suggested Research Questions	Suggested Research Design and Methodology
What types of integration mechanisms are used in conjunction with different IT governance models? Do certain IT governance *gestalts* exist that combine IT governance differentiation and integration mechanisms?	A quantitative approach (large scale survey) to identify and validate patterns of IT governance differentiation and integration. Multiple case studies can also be used to explore and discover new patterns.
What are the most effective integration mechanisms under different IT governance models?	A quantitative approach (large scale survey) to identify effective integration mechanisms for IT governance. Multiple case studies can be of value when considering the business and industry context.
How do integration mechanisms moderate the relationship between IT governance and IT business value?	A quantitative approach (large scale survey) to measure and validate the moderating impact of integration mechanisms on IT business value realization
How do organizations design and implement integration mechanisms for IT governance?	A longitudinal research design involving field-research and multiple case studies. A single in-depth case study can also provide rich (theoretical) insights.
What are the (inter-) dependencies between integration mechanisms? Do companies follow a specific path when developing and implementing integration mechanisms (e.g., do they start with positions, the processes and finally relationships?)	A longitudinal or retrospective field-study on the development and evolution of integration mechanisms in certain companies
Are certain drivers (e.g., standardization, responsiveness, flexibility) related to the use (or non-use) of specific integration mechanisms? How do the competing drivers of efficiency and flexibility impact the adoption and usage of integration mechanisms?	Both surveys and/or case studies can be used. A (quantitative) survey would identify empirically significant relationships amongst certain drivers and integration mechanisms. Multiple case studies would help explain why and how these (competing) drivers lead to the adoption and usage of integration mechanisms

CONCLUSION

Amidst the challenges and changes of the 21st century, involving hyper-competitive market spaces, electronically-enabled global network businesses, and corporate governance reform, IT governance has become a fundamental business imperative. IT governance is a top management priority, and rightfully so, because it is the single most important determinant of IT value realization (Mata, Fuerst, & Barney, 1995; Peterson, 2001; Rockart, Earl, & Ross, 1996; Sambamurthy & Zmud, 1999; Weill & Broadbent, 1998).

More than simply assigning and allocating IT decision-making authority, IT governance is the system by which an organization's IT portfolio is directed, and describes the distribution of IT decision-making rights and responsibilities among different stakeholders in the organization, and the rules and procedures for making and monitoring decisions on strategic IT concerns. These rules and procedures address the integration mechanisms, which are fundamental to effective IT governance.

Nevertheless, research has only recently focused on the use and effectiveness of integration mechanisms for IT governance. More research is definitely and urgently required in this area (Table 2). It is only through empirical investigation of these and other related research questions that we will be able to advance the current body of knowledge and understanding in the area of IT governance.

REFERENCES

Brown, C.V. (1997). Examining the emergence of hybrid IS governance solutions: Evidence from a single case site. *Information Systems Research*, *8*(1), 69-94.

Brown, C.V., & Magill, S.L. (1994, December). Alignment of the IS functions with the enterprise: Toward a model of antecedents. *MIS Quarterly*, 371-403.

Brown, C.V., & Magill, S.L. (1998). Reconceptualizing the context-design issue for the information systems function. *Organization Science, 9*(2), 177-195.

Daft, R.L. (1998). *Organization theory and design* (6th ed.). South Western College Publishing.

Galbraith, J.R. (1994). *Competing with flexible lateral organisation*. MA: Addison-Wesley.

Grembergen, W. van (2003). Strategies for information technology governance. Hershey, PA: Idea Group Publishing.

Hodgkinson, S.T. (1996). The role of the corporate IT function in the federal IT organisation. In M.J. Earl (Ed.), *Information management: The organisational dimension*. Oxford University Press.

Lawrence, P.R., & Lorsch, J.W. (1967). *Organisation and environment. Managing differentiation and integration.* Harvard University Press.

Mata, F.J., Fuerst, W.L., & Barney, J.B. (1995). Information technology and sustained competitive advantage: A resource-based analysis. *MIS Quarterly*, *19*(4), 487-505.

Peterson, R.R. (2001). Information governance: An empirical investigation into the differentiation and integration of strategic decision-making for IT. Tilburg University, The Netherlands.

Peterson, R.R. (2003). Integration strategies and tactics for information technology governance, In W. van Grembergen (Ed.), *Strategies for information technology governance*, Hershey, PA: Idea Group Publishing.

Peterson, R.R., O'Callaghan, R., & Ribbers, P.M.A. (2000). Information technology governance by design. *Conference Proceedings of the Twenty-First International Conference on Information Systems*, Brisbane, Australia.

Rockart, J.F., Earl, M.J., & Ross, J.W. (1996, Fall). Eight imperatives for the new IT organization. *Sloan Management Review*, 43-55.

Sambamurthy, V., & Zmud, R.W. (1999). Arrangements for information technology governance: A theory of multiple contingencies. *MIS Quarterly*, *23*(2), 261-290.

Weill, P., & Broadbent, M. (1998). *Leveraging the new infrastructure: how market leaders capitalise on information technology*. Boston, MA: Harvard Business School Press.

KEY TERMS

Centralized Model: The concentration of decision-making in a single point in the organization, in which a single decision applies.

Collaboration: A close, functionally interdependent relationship, in which organizational units strive to create mutually beneficial outcomes. Collaboration involves mutual trust, the sharing of information and knowledge at multiple levels, and includes a process of sharing benefits and risks. Effective collaboration cannot be mandated.

Decentralized Model: The dispersion of decision-making, in which different independent decisions are made simultaneously.

Differentiation: The state of segmentation or division of an organizational system into subsystems, each of which tends to develop particular attributes in relation to the requirements posed by the relevant environment. This includes both the formal division, as well as, behavioral attributes of the members of organizational subsystems.

Federal Model: A hybrid configuration of centralization and decentralization, in which decision-making is differentiated across divisional and corporate units.

Integration: 1) The process of achieving unity of effort among various subsystems in the accomplishment of the organizational task (*process focus*). 2) The quality of the state of collaboration that exists among departments, which is required to achieve unity of effort by the demands of the environment (*outcome focus*).

IT Applications: Local business-functional applications embedded in business processes, activities, products and/or services.

IT Governance: 1) Locus of IT decision-making authority *(narrow definition)*. 2) The distribution of IT decision-making rights and responsibilities among different stakeholders in the organization, and the rules and procedures for making and monitoring decisions on strategic IT concerns *(comprehensive definition).*

IT Governance Comprehensiveness: Degree to which IT decision-making/-monitoring activities are systematically and exhaustively addressed.

IT Governance Formalization: Degree to which IT decision-making/-monitoring follows specified rules and standard procedures.

IT Infrastructure: The base foundation of the IT portfolio, delivered as reliable shared services throughout the organization, and centrally directed, usually by corporate IT management.

IT Governance Integration: Degree to which business and IT decisions are integrated administratively, sequentially, or reciprocally.

IT Portfolio: Portfolio of investments and activities regarding IT operations and IT developments spanning IT infrastructure (technical and organizational components) and IT applications.

Participation: Process in which influence is exercised and shared among stakeholders, regardless of their formal position or hierarchical level in the organization.

Shared Learning: The co-creation of mutual understanding by members of organizational sub-units of each other's goals and objectives.

Triangular Strategic Analysis for Hybrid E-Retailers

In Lee
Western Illinois University, USA

INTRODUCTION

For traditional retailers, the success of an e-channel lies largely in formulating and implementing a sound e-channel strategy that leverages their resource base. Numerous evidences show that poorly developed e-channels have added little value to retailers (Huang, 2003; Prencipe & McCarthy, 2002). Some of the poorly developed e-channels have had a negative impact on business performance due to an excessive investment, disappointing sales, and low margin (Nataraj & Lee, 2002). The poor performance arose from focusing on the Internet as a separate channel not affected by the activities in other existing channels (Kannan, 2001). For many traditional retailers, their costly and frequent e-channel reorganizations could have been avoided if they had adequately analyzed the strategic fit between their external environment and e-channel organization.

There has been no universally applicable business strategy for e-channels. For many retailers, the right mix of the traditional channels and e-channel is critical to their business success (Gulati & Garino, 2000). Despite the strategic value of the e-channel, there have been only a paucity of frameworks that help managers initiate and formulate the e-commerce strategy (Allen & Fjermestad, 2001; Lee, 2001). A number of strategic analysis models such as SWOT analysis, five forces model, resource-based view, and critical success factors have been applied to the e-commerce strategy development. These models attempted to formulate a business strategy from different perspectives of a business organization, but have not been fully integrated with each other. Formulating a business strategy based on the analysis and integration of multiple perspectives will result in a more competitive strategy than those of a single perspective.

Based on a number of e-channel case studies and strategic management theories, this short article presents (1) an overview of a triangular strategic analysis and (2) an example application of the triangular strategic analysis with an Office Depot case study. Data on the Office Depot's e-channel strategy and implementation were collected through secondary sources such as trade journals and Office Depot's official publications. The triangular strategic analysis consists of (1) competitive forces analysis, (2) resource base analysis, and (3) critical success factor analysis

BACKGROUND

To capture the ever-increasing B2C population, retailers have experimented with a variety of B2C business models (Gulati & Garino, 2000). Some of the widely used e-commerce models include auction models (e.g., eBay.com), reverse auction models (e.g., Priceline.com), portal models (e.g., Yahoo.com), stand-alone e-retailer models (e.g., Amazon.com), and hybrid e-retailer models (e.g., Walmart.com). While B2B e-commerce applications such as e-procurement systems and the Internet-based supply chain management have brought significant benefits to business organizations, many B2C business models have failed to generate sustainable long-term profits.

In the late 1990s, most stand-alone e-retailers of commodity type products suffered the hardest hits due to low margin, rising customer acquisition cost, and the lack of financial support of investors (Stockport, Kunnath, & Sedick, 2001). Numerous stand-alone e-retailers such as Garden.com, Boo.com, and Petopia.com were consolidated with traditional retailers or liquidated (Kujubu & Martin, 2001). These failures were attributed to the poor business plan, weak complementary resources in distribution network and customer services, lack of brand name recognition, and low entry barriers.

Evidence shows that a misdirected e-channel development leads to costly and frequent revisions of e-commerce strategies. Kmart and Wal-Mart experienced a costly revision of their e-channel strategies. Kmart initially created a spin-off entity, BlueLight.com, in December 1999 as a joint venture between Kmart and Softbank Venture Capital. After Kmart withdrew from a planned initial public offering (IPO) for BlueLight.com in 2000, it acquired all of the interests of BlueLight.com in 2001. Walmart.com is another example of the costly revision of an e-channel strategy. Walmart.com was established in January 2000 as an independent company operating as a joint venture between Wal-Mart and Accel Partners. In 2001, Wal-Mart acquired all the minority interest in Walmart.com in order to establish the tight integration between its e-channel and physical stores.

Since each organization is uniquely positioned in a market with a different set of competitive forces, critical success factors, and capabilities, no single e-channel strategy would be suitable for all organizations. The successful deployment of an e-channel requires a thorough review and analysis of all major business activities, including business strategies, processes, functions, and vendor/customer relationships. For traditional retailers, poorly deployed e-channels without cross-channel coordination and integration mechanisms in place cannot create competitive advantages. These e-channels may also have a negative impact on other channels by losing customers who value a seamless cross-channel experience. The triangular strategic analysis will provides managers with a unified view of a business organization by combining and presenting multiple organizational perspectives.

TRIANGULAR STRATEGIC ANALYSIS FOR HYBRID E-RETAILERS

The purposes of strategic analysis are to examine the current and future business environments, to identify new business opportunities and threats, and to develop strategies to counter competition and achieve strategic goals. A number of strategic analysis models have been developed with the emphasis on different perspectives of a business strategy development. Table 1 summarizes major strategic theories/models, their purposes, advantages, and disadvantages. Based on the complementarities of these models, we utilize three analysis models in an e-channel development framework. The triangular strategic analysis consists of (1) competitive forces analysis, (2) resource base analysis; and (3) critical success factor analysis. While each of these strategic analysis methods

Table 1. Summary of major strategic management theories/models

Major Theories/Models	Proponents	Characteristics
Five Forces Model of Industry Competition	Porter (1980)	His basic theory was that dynamics of five competitive forces determine the nature of competitiveness in an industry and influence the strategies available to firms in the industry. The competitive forces are: (1) threat of new entry into an industry; (2) intensity of rivalry among existing competitors; (3) pressure from substitute products; (4) bargaining power of buyers; and (5) bargaining power of suppliers.
Resource Based View (RBV)	Wernerfelt (1984)	RBV suggests that firms compete not just in terms of final products, but more fundamentally in terms of the underlying “resources” which make production and product diversification possible. From a resource-based view every firm has a unique set of resources that the firm can leverage to exploit opportunities and counter threats.
Core Competence	Prahalad and Hamel (1990)	Core competencies are the collective learning in the organization that gives the company a unique advantage over its competitors. Core competence can manifest itself in many ways. Core competence is communication, involvement, and a deep commitment to working across organizational boundaries. It is the skills of individuals who can blend their expertise with that of others in new and interesting ways.
Balanced Scorecard (BSC)	Kaplan and Norton (1996)	The Balanced Scorecard is a method for turning a company’s vision and strategy into a coherent set of performance measures distributed among four perspectives: Financial, Customer, Internal Business Processes, and Learning and Growth. The framework provides a balance between short- and long-term objectives, financial and nonfinancial measures, and external and internal performance indicators.
Critical Success Factor (CSF) Analysis	Rockart (1979)	CSF analysis is a method developed to guide businesses in creating and measuring success. CSFs are key areas where satisfactory performance is required for the organization to achieve its goals. Rockart provided the following as an example of the CSFs: new product development, good distribution, and effective advertising - factors that remain relevant today for many firms.

Table 2. Impacts of e-commerce on competitive forces' threats and firm's opportunities from retailers' perspective

Competitive Forces	Threats	Opportunities
Suppliers	Disintermediation Sell-Side Forward Auction	E-Procurement Group Purchasing Reverse Auction Internet-based EDI
Traditional Competitors	E-Channel E-Services Intranet Extranet E-Procurement	E-Channel Strategic Alliances Third-Party E-Marketplaces
New Market Entrants	E-Commerce Strategic Alliances	E-Channel Strategic Alliances, Merger/Acquisition
Customers (Corporate)	E-Procurement Third-Party Order Aggregation Price Comparison	E-Channel One-to-One Marketing Internet-based EDI Extranet
Customers (Consumers)	Price Comparison Third-Party Order Aggregation	E-Channel Web Personalization E-Services User Profiling
Substitute Products/Services	Digitized Products/Services On-Demand Delivery Services Portals	E-Channel Digitized Products/Services On-Demand Delivery Services New Services

was developed in isolation of the others, they contribute unique yet complementary perspectives on the development of the business strategy. The integration of three analyses implicitly subsumes the essence of the core competence analysis. The balanced scorecard model can be used in parallel with the triangular strategic analysis or in the subsequent planning stage, and later in evaluating the organizational performance.

Competitive Forces Analysis

Porter's Five Forces Model (1980) has been widely used in analyzing competitive forces that would shift an organization's strategic position in the industry. The components of the model include: (1) the traditional competitors, (2) the bargaining power of the customers, (3) the bargaining power of the suppliers, (4) the potential threat of new entrants, and (5) the threat of substitute products/services. The competitive forces analysis presents a picture of where the company's current strategy stands against competitive forces and a road map of where it should go for the success in the competitive environment. The emergence of e-commerce has changed the ways all competitive forces perform businesses in the industry. E-commerce provides customers with opportunities to comparison-shop and thereby raises the bargaining power of customers. E-commerce also provides suppliers with opportunities to disintermediate retailers. Traditional competitors establish e-commerce Web sites to enter into the e-marketplaces.

The competitive forces analysis provides managers with vital information on the opportunities and threats arising in the e-commerce market. Table 2 summarizes the opportunities and threats introduced by the e-commerce from the retailers' perspective. While the competitive forces analysis is effective in understanding external environment, some researchers have pointed out its limitations (Luffman, Lea, Sanderson, & Kenny, 1996). The analysis has not explicitly addressed what the critical factors are and what resources are needed for a successful execution of a business strategy. The following resource base analysis utilizes the results obtained from the competitive forces analysis and determines what resources are needed and how they are used in the execution of an e-channel strategy.

Resource Base Analysis

Many researchers have indicated that the resource-based view of a firm is the most important theory of sustainable competitive advantage (Conner, 1991). The fundamental logic of the resource-based view is that the desirable outcome of a business strategy is a sustainable competitive advantage. Since not all resources are equal in creating sustainable competitive advantage, many researchers focused on identifying advantage-creating resources. Barney (1986, 1991) suggested that the advantage-creating resources must be firm-specific, rare, and difficult to imitate. The resource-based view of a firm was applied to understand how the superior IT resources of organizations render the cost and value of IT innovations different from competitors (Bharadwaj, 2002). In a changing environment, firms must continuously invent and upgrade their resources and capabilities if they are to maintain a competitive advantage and growth. The sequential development of resources and capabilities can make a firm's advantage inimitable (Barney, 1991; Lado, Boyd, & Hanlon, 1997).

Each retailer has a different set of firm-specific resources to utilize, and an addition of an e-channel can be

Table 3. E-commerce analysis matrix of competitive forces and resource base: Office Depot's case

Competitive Forces	Suppliers		Traditional Competitors		New Market Entrants		Customers		Substitute Products/ Services	
Resource Base	Threats	Opportunities	Threats	Opportunities	Threats	Opportunities	Threats	Opportunities	Threats	Opportunities
Physical Stores							●			
Distribution Centers							●			
Database				●				●		
Integrated Information Systems	●							●		
Skills and Knowledge of Employees			●					●	●	
Organizational Culture and Trust								●		

Table 4. E-commerce analysis matrix of critical success factors and resource base: Office Depot's case

Critical Success Factors / Resource Base	Efficient Distribution Systems	Business Process Redesign	Channel Coordination and Integration	Customers' Trust	Superior Customer Service Operations	Strategic Alliances
Physical Stores	●			●	●	
Distribution Centers	●		●		●	
Database	●		●			
Integrated Information Systems	●	●	●		●	
Skills and Knowledge of Employees		●		●		●
Organizational Culture and Trust		●	●		●	●

viewed as a unique resource utilization and development process in achieving sustainable growth. The mapping between the competitive forces and the resource base provides answers to two important questions: (1) what kinds of existing resources a retail organization can leverage and (2) what resources it needs to differentiate itself from the competitors. While strong resources are leveraged to survive, weak or non-existent resources need to be critically examined for the future resource development and sustainable competitive advantage. Traditional retailers' resources include physical stores, distribution centers, patent, trademark, brand, reputation, enterprise-wide database, integrated information systems, skills and knowledge of employees, and organizational culture and trust. Some of these resources are more difficult to imitate and more valuable as a source of competitive advantage than others. In general, intangible resources would be more difficult to transfer and imitate than tangible resources.

Table 3 shows the e-commerce strategy matrix of the competitive forces and the resource bases of Office Depot. The rows list two types of resources: strong resources (sustaining resources) and weak/non-existent resources (developmental resources). Certain weights can be assigned to the competitive forces based on their importance to prioritize the resource allocation and development activities. The analysis of Office Depot suggests that the existing integrated information systems can be leveraged to counter the threat of the suppliers' e-commerce and to benefit from the growth of the customers' online purchases.

The matrix needs filling of e-commerce strategies and projects in relation to resources, threats and opportunities. Threats posed by new market entrants may force a retailer to choose the e-channel introduction as a strategic option. If a retailer has a strong resource base in support of an e-channel, then an immediate introduction of an e-channel may be feasible with little developmental resources. Overall, the e-commerce strategy matrix of the competitive forces and the resource base provides a conceptually grounded framework for assessing sustainable and developmental resources, and enables these resources to be examined in terms of the opportunities and threats for establishing sustainable competitive advantages.

Critical Success Factors Analysis

Once the previous analyses lead to an e-commerce strategy, critical success factors (CSFs) analysis decides what the most important determinants are in achieving strategic goals (Rockart, 1979). The successful execution of CSFs requires resources. Identifying a match between the firm's resources and the critical success factors in the industry is a demanding task, and the success of the match is a function of the accuracy of managerial expectations about the value of the strategy (Barney, 1986).

The e-commerce success factors for retailers include efficient business process, integrated distribution systems, multi-channel coordination and integration, customers' trust, strong brand recognition, superior customer service operations, financial stability, effective online technologies, and strategic alliances. Table 4 shows a matrix that develops matches between the Office Depot's resources and the critical success factors. For example, the business process redesign requires the integrated information systems, skills and knowledge of employees, and organizational culture and trust as critical resource bases.

The matrix helps managers determine what existing resources can be leveraged or what new resources are needed to achieve the CSFs and ultimately the business strategies. To successfully achieve e-channel strategic goals and sustain competitive advantages, retailers need to monitor the performance of these CSFs with a measurable objectives and metrics. Once CSFs and metrics are defined, a detailed e-channel plan should be developed to achieve these critical success factors and business strategies.

FUTURE TRENDS AND CONCLUSION

The rapid penetration of the World Wide Web and the explosion of e-commerce startups have changed the dynamics of the competitive forces across all industries. Due to the explosion in e-commerce competition, the e-channel has become a critical factor in the strategy development by the hybrid e-retailers. While managing the e-channel is one of the most important tasks for marketing managers, many managers are still unclear about e-channel strategies and lack core e-channel knowledge needed to analyze business environments, to develop strategies, and to evaluate alternative e-channel solutions. Empirical evidence shows that experimenting with different types of e-channels is very costly. Some retailers such as Kmart, Wal-Mart, CVS, and Staples experimented with spin-offs. These retailers later struggled to retrofit the spin-offs into their parent companies.

A number of strategic analysis models such as SWOT analysis, five forces model, resource-based view, and critical success factors have been applied to e-commerce strategy development. However, there has been little effort to integrate these models for e-channel strategy development. These models attempted to formulate a business strategy from different perspectives of a business organization, but have not been fully integrated with each other. The analysis and integration of multiple dimensions will result in a more comprehensive strategy than those of a single dimension. To develop a comprehensive e-channel strategy, the triangular strategic analysis attempted to integrate three well-known analysis models: competitive forces analysis, resource base analysis, and critical success factor analysis.

The triangular strategic analysis provides managers with a unified view of a business organization by combining and presenting multiple perspectives of an organization. The unified view will enhance managers' ability to effectively identify their present strategic position, internal resources, and critical success factors, and decide upon the most appropriate e-channel strategy. As important, the unified view will allow managers to utilize the emerging opportunities of e-commerce and to prevent threats that would be posed by carelessly developed e-channel strategies.

The triangular strategic analysis was applied to analyze Office Depot's e-channel development strategy. The evaluation of Office Depot's e-channel development strategy suggests a number of its success factors that may be useful for other retailers' e-channel development: (1) Office Depot was an early adopter of an e-commerce technology, and continued to explore different e-commerce business models over time; (2) its e-channel was not only another distribution channel but also a service channel; (3) it also leveraged its own e-channel expertise and resources in office supplies in expanding to other markets; and (4) it treated an e-channel not as an independent entity but as an internal business unit in an integrated business organization; (5) its senior management supported an e-commerce project from a strategic point of view; and (6) it pursued a resource-based e-channel IT development.

REFERENCES

Allen, E., & Fjermestad, J. (2001). E-commerce marketing strategies: An integrated framework and case analysis. *Logistics Information Management*, *14*(1/2), 14-23.

Barney, J.B. (1986). Strategic factor markets: Expectations, luck, and business strategy. *Management Science*, *32*(10), 1231-1241.

Barney, J.B. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, *17*(1), 99-120.

Bharadwaj, A.S. (2000). A resource-based perspective on information technology capability and firm performance: An empirical investigation. *MIS Quarterly*, *24*(1), 169-196.

Conner, K.R. (1991). A historical comparison of resource-based theory and five schools of thought within industrial organisation economics: Do we have a new theory of the firm. *Journal of Management*, *17*(1), 121-154.

Gulati, R., & Garino, J. (2000). Get the right mix of bricks & clicks. *Harvard Business Review*, *78*(3), 107-114.

Huang, X. (2003). Research on Australian e-tailers: Strategic issues, success factors, and challenges. *International Journal of Services Technology & Management*, *4*(4-6), 563-573.

Kannan, P.K. (2001). Introduction to the special issue: Marketing in the e-channel. *International Journal of Electronic Commerce*, *5*(3), 3-6.

Kaplan, R.S., & Norton, D. (1996). Using the balanced scorecard as a strategic management system. *Harvard Business Review*, *74*(1), 75-85.

Kujubu, L., & Martin, A. (2001), Opportunity in failure. *InfoWorld, 23*(15), 36-37.

Lado, A.A., Boyd, N.G., & Hanlon S.C. (1997). Competition, cooperation, and the search for economic rents: A syncretic model. *Academy of Management Review*, *22*(1), 110-141.

Lee, C-S. (2001). An analytical framework for evaluating e-commerce business models and strategies. *Internet Research*, *11*(4), 349-359.

Luffman, G., Lea, E., Sanderson, S., & Kenny, B. (1996). *Strategic management*. Blackwell Publishers Inc, Oxford.

Nataraj, S., & Lee, J. (2002). Dot-com companies: Are they all hype? *S.A.M. Advanced Management Journal*, *67*(3), 10-14.

Porter, M.E. (1980). *Competitive strategy*. New York: Free Press.

Prahalad, C., & Hamel, G. (1990). The core competence of the corporation. *Harvard Business Review*, *68*(3), 79-91.

Prencipe, L.W., & McCarthy, J. (2002). Battle of the shopping carts. *InfoWorld*, *24*(40), 46.

Rockart, J.F. (1979). Chief executives define their own data needs. *Harvard Business Review*, *57*(2), 81-93.

Stockport, G.J., Kunnath, G., & Sedick, R. (2001). Boo.com - The path to failure. *Journal of Interactive Marketing*, *15*(4), 56-70.

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal*, *5*(2), 170-180.

KEY TERMS

Channel Conflict: Situation in which an e-channel creates a conflict with existing channels because of real or perceived damage from inter-channel competition.

Complementarities: Products or services that provide more value together than individually. For example, hybrid e-retailers can leverage complementarities by providing offline services to online shoppers.

E-Channel: An online marketing channel where companies and customers conduct business, no matter where they are. Since the e-commerce revolution, many brick-and-mortar businesses have expanded their marketing channel to include e-channel.

Hybrid E-Retailer: A click and mortar company which conducts retailing through e-channel as well as physical stores and other distribution channels. Compared to its pure e-commerce competitors, a hybrid e-retailer can leverage existing physical stores, brand recognition, distribution network, existing customer base, and so forth.

On-Demand Delivery Services: Express delivery of products made with highly efficient transportation systems after an online order is received.

Order Aggregation: A group purchase designed to achieve a volume discount by aggregating orders placed by individual buyers.

Reverse Auction: A fixed-duration auction hosted by a single buyer in which multiple sellers compete for business.

Sell-Side Forward Auction: An auction where a seller announces the items for quick sale and buyers bid on them.

Triggers, Rules and Constraints in Databases

Juan M. Ale
Universidad de Buenos Aires, Argentina

Mauricio Minuto Espil
Universidad Católica Argentina, Argentina

INTRODUCTION

Databases are essentially large repositories of data. Since the mid-1980s up to the mid-1990s, considerable effort has been paid to incorporate reactive behavior to the data management facilities available. Reactive behavior is characterized by variants of the event–condition–action model. Applications areas include checking for integrity constraints, system alerts, materialized view maintenance (especially useful in data warehousing), replication of data for audit purposes, data sampling, workflow processing, implementation of business rules, scheduling, and many others. Practically all products offered today in the database marketplace support complex reactive behavior on the client side. Nevertheless, the reactive behavior supported by those products on the server side is poor. Recently, the topic has regained attention because of the inherent reactive nature demanded in Web applications and the necessity of migrating many of the functionalities of browsers to active Web servers (Bonifati, Braga, Campi, & Ceri, 2002).

BACKGROUND

Several applications that support reactive behavior in the electronic commerce arena appeared recently, as is the case in the following: the Active Views system, described in Abiteboul et al. (1999); the event–condition–action (ECA) rule language for XML repositories, described in Bailey, Poulovassilis, and Wood (2002); and the set of classes for remote notification within a Web service environment, described in Bonifati, Ceri, and Paraboschi (2001).

Supporting reactive behavior implies that a database management system must be viewed from a production rule system perspective (Baralis, Ceri, & Paraboschi, 1996). An active database system must support the definition of production rules. Production rules are well known nowadays, in database terminology, under the name of active rules or, simply, triggers.

Active rules and integrity constraints are related topics (Ceri, Cochrane, & Widom, 2000). Database engines do not bring a full support of declarative integrity constraints in their kernels. When a complex constraint must be enforced on data, and the constraint cannot be declared, it must be emulated by means of triggers.

From a user's point of view, reactivity is a concept related to object state evolution over time. Dynamic constraints, constraints making assertions on the evolution of object states, may be needed to control changes in the states of data objects (Sistla & Wolfson, 1995b).

ACTIVITY WITHIN DATABASES

Usually, a database system performs its actions in response to requests from users in a passive way. In some cases, it is desirable that actions be taken with no human intervention, that is, automatic response to certain events.

Traditionally, the latter behavior has been obtained by embedding it into user applications; that is, the application software recognizes some events triggered by an user and performs some actions in response.

Because of the complexity in supporting reactive behavior, it would be desirable that the active functionality be provided by the database system. A database with a capability of reacting to external or internal stimuli is called an active database. An active database system can be thought of as coupling a database management system with a rule-based programming environment (Paton & Diaz, 1999). Among the applications that use active database systems nowadays, we can mention inventory control systems, online reservation systems, and portfolio management systems, just to name a few.

Knowledge Model

A central issue in the knowledge model of active databases is the concept of active rule.

An active rule is defined throughout three dimensions: event, condition, and action. In this case, this is termed an ECA rule.

An event is something that happens at a point in time. The source of the event may be transactional (an abort, commit, or begin transaction), operational (insert, delete, or update operations), temporal (clock signaling), or external (generated by the environment). An event can be classified according to its complexity as primitive or composite.

A condition, i.e., a predicate evaluated on data states, is the second component of an active rule. Moreover, because the state of data may change, before and after the occurrence of an event, the condition should be able to refer to previous and new states.

An action consists in a sequence of operations. There are several options for possible actions: the update of the contents or structure of the database, the call of an external procedure, the abort of the current transaction, or the notification about some unexpected situation.

Execution Model

The execution model for a set of active rules determines how the rules are managed at execution time. This model is strongly dependent on the particular implementation. However, it is possible to describe it in general using a set of common activities or phases: signaling, triggering, evaluating, scheduling, and executing.

How these phases are synchronized depends on the so-called coupling modes of ECA rules. The relationship among the aforementioned activities of the rules involves the concepts shown in Table 1.

Termination and Confluence

The behavior of active rules is hard to understand and control (Baralis, Ceri, & Widom, 1993). Rule interaction is one of the most important aspects related to rule set behavior. Two important properties related to this problem are observed: termination and confluence. It is said that a rule set is guaranteed to terminate if, for any database state and initial event, rule processing cannot continue forever. A rule set is confluent if, for any database state and initial event, the final database state after rule processing is independent of the order in which activated rules are executed.

Basic methods that perform termination analysis of a rule set have been discovered. However, because of the undecidability of the problem in general (Bailey, Dong, & Ramamohanarao, 1998), we cannot always decide whether a rule firing process is guaranteed to finish.

According to the time when those methods are applied, they can be classified as static, if the rule set is analyzed at compile time, or dynamic, if the rule set behavior is analyzed at run time. Deciding whether the condition of one rule is affected by the action of other rules, and when two rule actions commute, is known as the propagation problem. Propagation has been thoroughly studied (see Baralis & Widom, 2000). Propagation is the crux of static methods to determine confluence and termination (Widom & Ceri, 1996).

In the commercial systems side, an approach consists of imposing syntactic limitations, in order to guarantee termination or confluence at run time, although in some cases, counters are used to prevent infinite execution.

ACTIVE RULES AND DECLARATIVE CONSTRAINTS

Declarative constraints are user definitions specifying restrictions that the database states must satisfy. In a SQL-1999 (Standard Query Language-1999) compliant system, four classes of declarative constraints are supported: check predicate constraints, referential constraints, assertions, and view check options. Check predicate constraints aim at validating conditions against the actual state of *one* table in the database, and they include *primary key* and *unique* definitions, *not null* column definition, and *explicit check* clauses that validate general predicates on the values of some of the columns of the table. Referential constraints aim at guaranteeing that a many-to-one relationship holds on the actual state of two

Table 1. Concepts

• Activation time
• Transition granularity
• Net effect policy
• Cycle policy
• Consumption modes
• Rule execution ordering

Table 2. Why declarative constraints and triggers must be distinguished

• Declarative constraints should be processed after all changes are effectively applied.
• Inconsistent states would lead to unpredictable behavior when firing a trigger.
• Processing constraints and triggers together should be confluent.

tables: the *referencing* or *child* table, and the *referenced* or *parent* table. A many-to-one relationship ensures that the column values of a foreign key (a list of columns of the referencing table) match the column values of a candidate key (a list of columns of the referenced table). Assertions aim at validating general predicates on rows in *different* tables. View check options deal with the problem of admitting modification operations through cascade defined views, yet retaining the natural meaning of the operations. A declarative constraint can be declared as having a deferrable or a nondeferrable activation time.

In SQL-1999, an active rule defined by the user is called a trigger, which is a schema object in a database (Ceri et al., 2000). The trigger structure is defined as follows:

CREATE TRIGGER <*trigger_name*> [BEFORE | AFTER] [<*event*> | <*events* >] ON <*table*>
REFERENCING NEW AS <*new_value*> OLD AS <*old_value*>
NEW TABLE AS <*new_table*> OLD TABLE AS <*old_table*>
FOR EACH [ROW | STATEMENT] WHEN <*condition*> < *action* >

Events can be statements INSERT, DELETE, or UPDATE <*list*>; <*table*> must be the name of a defined base table or view name; and <*list*> is a list of column names of table <*table*>. From now on, a trigger event is, therefore, a modification operation on a base table. The activation time is specified by keywords BEFORE or AFTER, thus yielding before and after triggers. Before triggers fire immediately before the operation specified as the trigger event has been issued. The after triggers fire immediately upon operation completion. The referencing clause admits defining correlation variables for transition values and transition tables, which allow the trigger to access column values of the affected rows before and after the execution of the modification operation. The transition granularity is specified by clause ON EACH and can be set to either ROW or STATEMENT. Hence, row-level and statement-level triggers can be defined. Row-level triggers fire one instantiation for each row affected by the modification operation. Provided no row is affected, a row-level trigger is never instantiated. Statement triggers fire only once per statement invocation and are evaluated even in the case the event does not happen to affect any row.

A declarative constraint includes explicitly or implicitly the specification of repairing actions. Hence, declaring a SQL constraint may be thought of as entailing the activation of internal active rules that enforce repairing actions whenever the constraint is violated. (See Ale & Minuto Espil, 2003, for details.)

Up to this point, the reader may wonder if declarative constraints and triggers are the same. Despite their similarities, declarative constraints and constraint rules must be distinguished. Table 2 summarizes some of the reasons.

The considerations above imposed the obligation of producing a precise specification on how declarative integrity constraints and constraint rules should be integrated. An actual accepted specification has been produced by the SQL-1999 standardization committee and is based on a proposal submitted by a research group in the IBM Almaden Research Center (Cochrane, Pirahesh, & Mattos, 1996).

FUTURE TRENDS

A different question is as follows: Can active rules be regarded as constraints? The answer is affirmative. If a past temporal logic formula is derived from the event and condition part of the rule, and a present temporal logic formula that fixes the new state in the case of rule activation is derived from the action part of the rule, an implication of the latter from the first defines a constraint that restricts state transition. It defines the pairs of consecutive states that are admissible in an execution and the pairs that are not (Sistla & Wolfson, 1995a).

CONCLUSION

We presented a brief survey of the interaction between active rules and integrity constraints. We discussed the current proposed techniques for dealing with situations when both declarative static constraints and triggers are defined. The perspective of our treatment has followed the SQL-1999 Standard Committee point of view, which constitutes the state of the art in that matter.

REFERENCES

Abiteboul, S., Cluet, S., Mignet, L., Amann, B., Milo, T., & Eyal, A. (1999). Active views for electronic commerce. In *Proceedings of the 25th International Conference on Very Large Databases* (pp. 138–149).

Ale, J., & Minuto Espil, M. (2003). Active rules and active databases: Concepts and applications. In S. Becker (Ed.), *Effective databases for text and document management.* Hershey, PA: IRM Press.

Bailey, J., Dong, G., & Ramamohanarao, K. (1998). Decidability and undecidability results for the termination problem of active database rules. In *Proceedings of* *the 17th ACM Symposium on Principles of Database Systems* (pp. 264–273).

Bailey, J., Poulovassilis, A., & Wood, P. (2002). An event–condition–action language for XML. In *Proceedings of the 11th ACM World Wide Web Conference* (pp. 486–495).

Baralis, E., & Widom, J. (2000). An algebraic approach to static analysis of active database rules. *ACM Transactions on Database Systems, 25*(3), 269–332.

Baralis, E., Ceri, S., & Paraboschi, S. (1996). Modularization techniques for active rules design. *ACM Transactions on Database Systems, 21*(1), 1–29.

Baralis, E., Ceri, S., & Widom, J. (1993). Better termination analysis for active databases. In *Proceedings of the First International Workshop on Rules in Database Systems* (pp. 163–179).

Bonifati, A., Braga, D., Campi, A., & Ceri, S. (2002). Active Xquery. In *Proceedings of the 18th International Conference on Data Engineering* (pp. 403–418).

Bonifati, A., Ceri, S., & Paraboschi, S. (2001). Active rules for XML: A new paradigm for e-services. *VLDB Journal, 10*(1), 39–47.

Ceri, S., Cochrane, R., & Widom, J. (2000). Practical applications of triggers and constraints: Successes and lingering issues. In *Proceedings of the 26th International Conference on Very Large Databases* (pp. 254–262).

Cochrane, R., Pirahesh, H., & Mattos, N. (1996). Integrating triggers and declarative constraints in SQL database systems. In *Proceedings of the 22nd International Conference on Very Large Databases* (pp. 567–578).

Paton, N., & Diaz, O. (1999). Active database systems. *ACM Computing Surveys, 31*(1), 63–103.

Sistla, A. P., & Wolfson, O. (1995a). Temporal triggers in active databases. *IEEE Transactions on Knowledge and Data Engineering, 7*, 471–486.

Sistla, A. P., & Wolfson, O. (1995b). Temporal conditions and integrity constraints in active databases. In *Proceedings of the 1995 ACM International Conference on Management of Data* (pp. 269–280).

Widom, J., & Ceri, S. (1996). *Active database systems: Triggers and rules for advanced database processing*. San Francisco, CA: Morgan Kaufmann Publishers.

KEY TERMS

Action: A sequence of operations.

Active Database: A database with the capability for reacting to stimula.

Condition: A predicate that evaluates a situation with respect to circumstances.

Declarative Constraint: A schema object in the database that defines whether a state of the database is consistent or not.

Event: Something that happen in a point of time.

Rule Activation Time: A concept that fixes the position of the signaling phase with respect to the event occurrence.

Rule Evaluation: An activity determining condition satisfaction.

Rule Execution: An activity running the corresponding action of an instantiated rule.

Rule Propagation: A property that determines how the condition of one rule is affected by the action of any other.

Rule Scheduling: An activity that determines the order of execution within a set of rule instances whenver they are fired simultaneously.

Rule Set Confluence: A property that indicates whether the final database state after rule processing is independent of the order of execution.

Rule Set Consumption Mode: Feature indicating how a set of rule instantiations is reduced as a consequence of execution.

Rule Set Cycle Policy: Feature indicating whether recursion must be applied when triggering a rule or rules must be instantiated only once.

Rule Set Net Effect Policy: Feature that indicates the relationship among event occurrences, rule instantiation and consumption modes.

Rule Signaling: An activity occurring when some source causes an event.

Rule Transition Granularity: A concept used in analyzing the relationship between event occurrences in rule instantiation.

Rule Triggering: Rule instantiation.

Trigger: A schema object in the database that defines the event, condition and action part of a production rule.

Trust in B2C E-Commerce for the New Zealand Mäori

Konrad J. Peszynski
Deakin University, Australia

INTRODUCTION

This study was performed to allow an understanding of some of the elements of trust that are apparent to encourage the Mäori Internet shopper to feel comfortable to shop online. Mäori arrived in New Zealand from the Pacific over 1,000 years ago. Since then, New Zealand was colonised by Europeans in the 19th century. As a result, the Mäori have become a minority (Belich, 1996). That is, their culture, language and values have become secondary to those of the dominant European culture (Liu, Wilson, McClure & Higgins, 1999). Mäori have been defined as including "all those who identify themselves as belonging to the New Zealand Mäori ethnic group, either alone or in combination with any other ethnic group" (Statistics New Zealand, 1998, p. 94).

An extensive selection of literature exists on trust and the Internet (Camp, 2000; McKnight, Cummings & Chervany, 1998). Few studies have examined the aspects of trust and the Internet for different cultures. Early studies, for example, by Jarvenpaa, Tractinsky, Saarinen and Vitale (1999), found that cultural differences do exist in Internet shopping. As a result, a gap in knowledge has been identified in terms of trust, Internet shopping and culture in New Zealand. This study has focused on Mäori, who, according to the literature have varying beliefs and values as opposed to the non-Mäori in New Zealand (Patterson, 1992).

This study will examine the issues that have helped or hindered the adoption of Internet shopping for the Mäori Internet user. From interviewee responses, a better understanding of current attitudes with regard to Internet shopping will be obtained. These responses will be linked to the various issues raised in this study. This study will therefore answer the questions set out as follows:

- What issues of trust are apparent to the Mäori Internet shopper?
- Does the current literature on trust and Internet shopping apply to the Mäori Internet shopper, and if so, how?

An attempt will be made to relate the literature reviewed and the conceptual framework developed in this study to the answers obtained via interviews in order to answer these research questions.

REVIEW OF LITERATURE

Through an investigation into the nature of business-to-consumer electronic commerce and the issues affecting the Internet customer, three key issues emerged: trust, risk and reputation.

Business-to-consumer (B2C) e-commerce is the trading and transactional relationship between an organisation's Web site and an end user (Lawrence et al., 2000; Schneider & Perry, 2000). Consumers are able to purchase goods and services such as books, computer products, and music at any time that is convenient to the consumer. Business-to-consumer e-commerce focuses on direct transactions between businesses and end consumers (Dedhia, 2001; Lawrence et al., 2000; Riggins & Rhee, 1998).

Business-to-consumer e-commerce involves smaller amounts of money in each exchange, whilst a greater volume of money is exchanged in the business-to-business (B2B) environment due to the volume of goods and services being transacted. Transactional relationships tend to be intimate for B2C e-commerce, as the objective is to satisfy customer preferences and shopping habits, but in the B2B setting, relationships are formal and impersonal (Ah-Wong et al., 2001; Lawrence et al., 2000; Riggins & Rhee, 1998).

There are factors that inhibit B2C e-commerce adoption. These factors include trust, risk and reputation. According to McKnight, Cummings and Chervany (1998, p. 459), trust is "an individual's beliefs about the extent to which a target is likely to behave in a way that is benevolent, competent, honest, or predictable in a situation." The consumer needs to believe that the Web merchant is trustworthy before he or she purchases online.

Factors that can increase the trust of an Internet consumer include assurance services such as those provided by MasterCard and TRUSTe (Cashell & Aldhizer III, 1999; Nöteberg, Christiaanse & Wallage, 1999), which permit a Web site to display their Web trust seal (a graphic) provided their "practices…[comply] with the

WebTrust principles and criteria for business-to-consumer e-commerce" (Cashell & Aldhizer III, 1999, p. 51). Web sites that show a seal are more likely to be purchased from than sites that carry no seal at all (Nöteberg et al., 1999). It made no difference as to which seal was displayed on the Web site, just as long as a seal was visible.

Another factor associated with inhibiting Internet customers from shopping online is risk. Risk is defined as the consumer's perceptions of the uncertainty and adverse consequences of engaging in an activity (Camp, 2000; Jarvenpaa et al., 1999, 2000). There is little assurance that the customer will get what he or she sees on the computer screen, in the quantity ordered.

It has been suggested that two types of risk are predominately associated with Internet shopping (Fram & Grady, 1997; Jarvenpaa et al., 1999, 2000; Lynch, Kent & Srinivasan, 2001). These risks are product category risk and financial risk. Product category risk is risk associated with the product itself and is the consumers' belief that the product will function according to their expectations. Financial risk refers to the risk associated with the Internet as a purchasing medium. Consumers are apprehensive about putting their credit card information over the Internet because it puts the consumer at risk of losing money via credit card fraud (Fram & Grady, 1997; Jarvenpaa et al., 1999, 2000).

The reputation and size of the Web merchant has also been suggested as factors that contribute to consumer trust in a seller organisation (Doney & Cannon, 1997; Ganesan, 1994). Reputation is the extent to which buyers believe that the selling organisation is honest and concerned about its customers (Doney & Cannon, 1997). Reputation requires a long-term investment of resources, effort and attention to customer relationships (Jarvenpaa et al., 1999). The better the seller's reputation, the more resources the seller has presumably committed to build that reputation. The perception of a large organisation implies that the merchant has significant resources invested in the business and has much to lose by acting in an untrustworthy way (Stewart, 1999).

MÄORI AND THE INTERNET

It is only recently that studies have been carried out on Mäori and their presence on the Internet (Kovacic, 2001; Smith, 1997; Smith & Sullivan, 1996). However, no studies examining the relationship between Mäori, trust and Internet shopping were found, indicating a gap in knowledge.

Jarvenpaa et al. (1999, 2000) and Lynch et al. (2001) found that trust varies across nationalities and cultures. Hofstede (1980) identified four constructs that have been the foundation of much cultural research: power distance, uncertainty avoidance, individualism and masculinity. Consumers coming from an individualistic culture may have greater trust and be more willing to base their trust in a Web site than consumers from a collectivistic culture (Javenpaa et al., 1999, 2000). The individualism/collectivism construct was the focus of this study.

According to Hofstede (1980), those in an individualistic society:

- Take care of him or herself and his or her immediate family;
- Have an "I" consciousness;
- Have a self-orientation; and
- Have greater economic development.

Those in a collectivistic society:

- Are born into extended families or clans, which protect them in exchange for loyalty;
- Have an emphasis on belonging to an organisation;
- Have a "we" consciousness; and
- Have less economic development (Hofstede, 1980).

A Mäori concept of community is based on collective strength, which includes sharing, nurturing, supporting and empowering of interdependent groups. The Mäori see community as a form of relatedness, in which people are interdependent (Gregory, 2001). An important method of communication for Mäori is through the *whanau* (family) structure (Light, 1999). That is, the connections are about relationships to people and to place.

Individualism tends to promote a trusting stance in an individualistic society (Jarvenpaa et al., 1999). As a result, one gets better outcomes assuming that others are reliable. Individualists are more likely to trust others until they are given some reason not to trust. By contrast, those high on collectivism are more likely to base their trust on relationships with first-hand knowledge. Members of collectivist cultures are less likely to trust someone who is not part of their in-group (Thanasankit, 1999; Yamagishi & Yamagishi, 1994).

The purpose of this study was to explore the issues of trust apparent to the interviewed Mäori Internet shoppers and see whether the current literature on trust and Internet shopping applies to the interviewed Mäori Internet shopper, and if so, how. Face-to-face interviews were performed with eight participants, all full-time professionals in either the public or private sector in Wellington, New Zealand. This allowed the researcher to identify Mäori perceptions of trust and Internet shopping. Because of the small sample size, results may only be generalised to a small population. However, Creswell (1994) and Yin (1994) claim that the result of case research is not to generalise to populations, but to generate a theory, which can be tested in future studies, or provide confirmatory evidence of existing theory. The data were analysed using a hermeneutic framework that enabled the identification of issues that arose in either the interview or the background literature review.

SOLUTIONS AND RECOMMENDATIONS

It was found that the current literature applied to the interviewed Mäori Internet shoppers to a certain extent. The use of assurance services has been noted in the literature as helping increase the trust of a consumer; however, those that were interviewed claimed to pay no attention to these Web seals. Participants claimed no knowledge of the groups or organisations supporting the Web seals or what they stand for.

In terms of risk and Internet shopping, respondents felt that financial risk is important and the same as what the literature states; however, the product category risk equally applies to the traditional "bricks and mortar" environment. Although the risk factor exists, the final purchase ultimately comes down to whether or not they want to take that risk. However, knowing the reputation of the organisation contributes to the trust in the trading relationship.

Reputation was a major theme that was apparent in most of the interviews; however, those interviewed tend to pay little attention to customer testimonials, particularly because they do not know if they are real or how the person who wrote the testimonial based their judgements. Respondents prefer to speak to someone they know that has visited that site to see how they found that site, rather than trust a total stranger.

Further questioning brought to light other factors such as the notion of anonymity; that is, no one knows they are Mäori, and as a result will not treat them any differently as opposed to non-Mäori, allowing them to shop with no discrimination. Photos of the people inside the organisation on the Web site can give the Mäori Internet shopper recognition that the organisation does exist and that they are not a faceless organisation. This also helps if the Mäori Internet shopper needs to call up for any reason, as they have a face to the name. Finally, the association of recognisable Mäori names to e-commerce sites could also make a difference, as there is that reputation or word-of-mouth notion that if other recognisable Mäori trust the site, then the Mäori Internet shopper should do so.

The framework developed in Figure 1 is an adaptation of the model created by Jarvenpaa et al. (1999, 2000). The difference is the extra step leading to reputation.

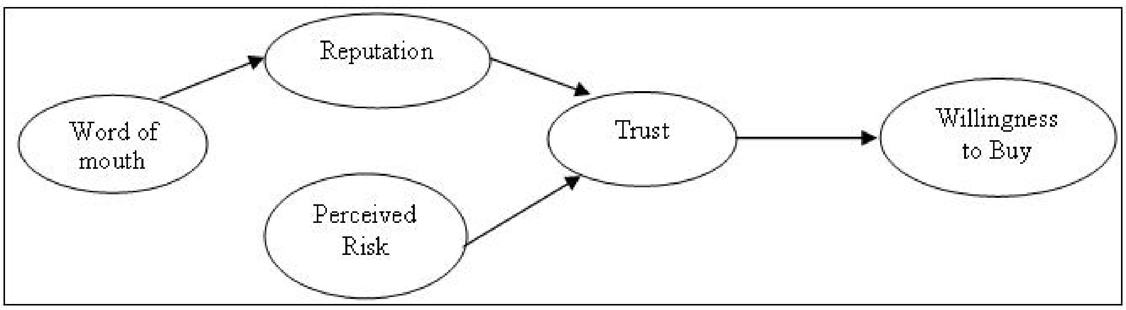
The creation of the extra parameter (word of mouth) leading to reputation has been introduced, as it has been found from the literature that because Mäori are part of a collectivist society, they are more likely to base their trust on relationships with first-hand knowledge. Mäori are also less likely to trust someone who is not part of their in-group (Thanasankit, 1999; Yamagishi & Yamagishi, 1994). That is, positive word of mouth about a Web site from someone in their in-group will increase the reputation of that Web site.

The information provided in this study can help New Zealand (and possibly international) e-commerce organisations, as it highlights some issues that Mäori have that either hinder them or assist them to purchase via the Internet. As Mäori are classed as being a collectivist society (Gregory, 2001; Light, 1999; Patterson, 2000), this information could also be applied to other collectivist societies.

One possible outcome of this study is the creation of a New Zealand Web seal. For example, a specific New Zealand organisation such as the Commerce Commission is seen as being a reputable source that those interviewed would trust. If New Zealand Web sites are registered with the Commerce Commission, then this could help increase the consumers' trust in the Web site and make them feel more comfortable to make a purchase.

The association of Mäori names with a Web site could also help the would-be Mäori Internet shopper trust the

Figure 1. Factors involved in increasing a consumer's willingness to purchase online



Web site and is linked closely to the previous suggestion of the Web seal. It is the recognition of a reputable association, and because the Mäori organisation (*iwi* (tribe), *hapu* (sub-tribe), *whanau* (family)) trusts that associated name, the Mäori Internet shopper should then trust the Web site.

Contact details should be made visible on the Web site, as one participant states: "I like the thought of it being on every page. You know, usually down the bottom they'll have all that information, the help, the contact details" (Participant Seven). Displaying contact details creates a professional image, and shows that there is the ability to contact the organisation.

FUTURE TRENDS

The study this article was based on also provides a base for future studies in the field of electronic commerce and cultural values. The first and foremost area is the New Zealand Mäori. The study was an introductory investigation into the cultural differences of Internet shopping for Mäori. As a result, possibly the first step for future studies is to investigate if and how the results of the current study compare to New Zealand Europeans.

Another extension would be to interview Mäori Internet shoppers in a New Zealand context, rather than a Wellington context. This would see if the results of this study are Wellington-specific, that is, only apply to Mäori Internet shoppers in the Wellington region. Another benefit of such an extensive study would create the ability to generalise the results New Zealand-wide, rather than concentrated on one specific city, Wellington. An investigation into other collectivist societies (Eastern countries,such as Japan, Thailand and China) could see if culture significantly influences the indigenous behaviours and attitudes of people.

CONCLUSION

The central concern of this study has been to gain some understanding of what issues of trust are apparent to make the Mäori Internet shopper feel comfortable to shop online. This study attempted to uncover some understanding of the associated factors that either help or inhibit Mäori from shopping on the Internet.

This study therefore tentatively suggests that trust for the Mäori Internet shopper in this demographic means that the Web site must have a good reputation, the ability to contact someone and the guarantee that the goods you order are actually the goods you receive in the delivery period as promised by the seller. There was also a sense of collectivism in the comment made. That is, participants want to see a contribution to Mäori society for any Mäori product they purchase.

Due to time and resource constraints, this study was restricted to a small number of interviewees. In improving on what has been done, it is suggested that conducting more interviews with the greater population of Mäori Internet shoppers within New Zealand could get more representative opinions about issues of trust and shopping on the Internet. Perhaps the adage "caveat emptor" is still applicable in this day and age when shopping via the Internet; that is, let the buyer beware.

REFERENCES

Belich, J. (1996). *The making of peoples: A history of New Zealand from Polynesian settlement to the end of the nineteenth century.* Auckland: Penguin Books.

Camp, J. (2000). *Trust and risk in Internet commerce.* Cambridge: MIT.

Cashell, J.D., & Aldhizer, G.R., III. (1999). Web trust: A seal of approval. *The Internal Auditor, 56*(3), 50-54.

Creswell, J. (1994). *Research design: Qualitative and quantitative approaches.* Beverley Hills, CA: Sage Publications.

Doney, P.M., & Cannon, J.P. (1997). An examination of the nature of trust in buyer-seller relationships. *Journal of Marketing, 61,* 35-51.

Fram, E.H., & Grady, D.B. (1997). Internet shoppers: Is there a surfer gender gap? *Direct Marketing, 59*(9), 46-50.

Ganesan, S. (1994). Determinants of long-term orientation in buyer-seller relationships. *Journal of Marketing, 58,* 1-19.

Gregory, R.J. (2001). Parallel themes: Community psychology and Mäori culture in Aotearoa. *Journal of Community Psychology, 29*(1), 19-27.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values.* Beverley Hills, CA: Sage Publications.

Jarvenpaa, S.L., Tractinsky, N., & Vitale, M. (2000). Consumer trust in an Internet store. To appear: *Information Technology and Management, 1*(1-2), 45-71.

Jarvenpaa, S.L., Tractinsky, N., Saarinen, L., & Vitale, M. (1999). Consumer trust in an Internet store: A cross-cultural validation. *Journal of Computer-Mediated Communication, 5*(2). Retrieved April 5, 2001, from *http://www.ascusc.org/jcmc/vol5/issue2/jarvenpaa.html*

Kovacic, Z.J. (2001). Positioning of Mäori Web sites in the space generated by the key concepts in Mäori culture. *2001 Informing Science Conference*. Retrieved May 12, 2001, from *http://schedule.elicohen.net/*

Lawrence, E., Corbitt, B., Fisher, J., Lawrence, J., & Tidwell, A. (2000). *Internet commerce: Digital models for business* (2nd ed.). Milton, Queensland: John Wiley and Sons.

Light, E. (1999). Market to Mäori. Key mistakes: Key remedies. *Marketing, 18*(6), 10-17.

Liu, J.H., Wilson, M.S., McClure, J., & Higgens, T.R. (1999). Social identity and the perception of history: Cultural representations of Aotearoa/New Zealand. *European Journal of Social Psychology, 29*, 1021-1047.

Lynch, P.D., Kent, R.J., & Srinivasan, S.S. (2001). The global Internet shopper: Evidence from shopping tasks in twelve countries. *Journal of Advertising Research, 41*(3), 15-23.

McKnight, D.H., Cummings, L.L., & Chervany, N.L. (1998). Initial trust formation in new organisational relationships. *Academy of Management Review, 23*, 473-490.

Nöteberg, A., Christiaanse, E., & Wallage, P. (1999). The role of trust and assurance services in electronic channels: An exploratory study. *Proceedings of the 20th International Conference on Information Systems* (pp. 472-478).

Patterson, J. (1992). *Exploring Mäori values*. Palmerston North: Dunmore Press Limited.

Patterson, J. (2000). Mana: Yin and yang. *Philosophy East & West, 50*(2), 229-240.

Schneider, G., & Perry, J. (2000). *Electronic commerce*. Cambridge, MA: Thomson Learning/Course Technology.

Smith, A.G. (1997). *Fishing with new nets: Mäori Internet information resources and implications of the Internet for indigenous peoples*. Paper presented at INET'97.

Smith, A.G., & Sullivan, R. (1996). Mäori electronic information: Issues and resources. *New Zealand Libraries, 48*(6), 111-118.

Statistics New Zealand. (1998). *New Zealand now Mäori*. Wellington, New Zealand: Statistics New Zealand.

Stewart, K.J. (1999). Transference as a means of building trust in World Wide Web sites. *Proceedings of the 20th International Conference on Information Systems* (pp. 459-464).

Thanasankit, T. (1999). *Exploring social aspects of requirements engineering – An ethnographic study of Thai software houses*. Unpublished PhD Thesis. University of Melbourne.

Yamagishi, T., & Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion, 18*, 129-165.

Yin, R. (1994). *Case study research: Design and methods*. Thousand Oaks, CA: Sage Publications.

KEY TERMS

B2C E-Commerce (Business-to-Consumer Electronic Commerce): Focuses on direct transactions between businesses and end consumers. Consumers are able to purchase goods and services such as books, computer products, music, at any time that is convenient to the consumer.

Collectivism: Concerned with group interest rather than individual interest (individualism). Collectivist societies support structures where people are born and live in extended families. The concept of collective strength includes sharing, nurturing, supporting and empowering of interdependent groups.

E-Commerce (Electronic Commerce): Usually refers to conducting business (electronically) with other businesses or consumers, but can be extended to include the inner workings within a business.

Individualism: Concerned with individual interest as opposed to group interest (collectivism). That is, everyone grows up to look after himself/herself. The relationship between superiors and subordinates is based on mutual advantage.

Reputation: The extent to which buyers believe that the selling organisation is honest and concerned about its customers.

Risk: The consumer's perceptions of the uncertainty and adverse consequences of engaging in an activity.

Trust: An individual's beliefs about the extent to which a target is likely to behave in a way that is benevolent, competent, honest, or predictable in a situation.

Trust in B2C E-Commerce Interface

Ye Diana Wang
University of Maryland Baltimore County, USA

THE NATURE OF TRUST

Electronic commerce (e-commerce) is changing the way people make business transactions, especially in the business-to-consumer (B2C) area, and it is becoming a significant global economic force. Since Internet technologies and infrastructures to support e-commerce are now in place, attention is turning to psychological factors that affect e-commerce acceptance by online users and their perceptions of online transactions. One such factor is trust, seen to be key to the proliferation of e-commerce.

Trust has existed as long as the history of humans and human social interactions, and it has been studied long before the emergence of the Internet or e-commerce. With respect to consumer behavior, studies have mainly focused on trust and trust relationships in the off-line world and have emerged from numerous disciplinary fields since the 1950s (Corritore, Kracher, & Wiedenbeck, 2001). These disciplines, including philosophy, sociology, psychology, management, marketing, ergonomics, human-computer interaction (HCI), and industrial psychology (Corritore, Kracher, & Wiedenbeck, 2003), have together contributed an extensive body of literature on trust in general, and therefore, they are important grounding points for the examination of trust in the online world. However, "trust is an extraordinarily rich concept, covering a variety of relationships, conjoining a variety of objects," as Nissenbaum (2001, p. 104) has pointed out. Due to the complex and abstract nature of trust, each discipline has its own understanding of the concept and different ways to conceptualize it according to the features of a particular context.

Even with the diverse trust research, researchers from every discipline do acknowledge the value of trust and generally observe and accept four characteristics of trust. First, there must exist two specific parties in any trusting relationship: a trusting party (trustor) and a party to be trusted (trustee). The two parties, comprised of persons, organizations, and/or products, constantly evaluate each other's behaviors. Second, trust involves vulnerability. Trust is only needed, and actually flourishes, in an environment that is uncertain and risky. Third, trust decreases complexity in a complex world and leads people to take actions, mostly risk-taking behaviors. "Without trust people would be confronted with the incomprehensible complexity of considering every possible eventuality before deciding what to do" (Grabner-Krauter & Kaluscha, 2003, p. 787). And fourth, trust is a subjective matter. It is directly related to and affected by individual differences and situational factors.

The previously mentioned characteristics of trust make it especially needed in e-commerce because people perceive economic transactions in a virtual environment as posing a higher degree of uncertainty than in traditional settings. Most e-commerce transactions are not only separated in time and space, but are also conducted via limited communication channels and impersonal interfaces, making trust a crucial facilitator for people to overcome fear, risks, and complexity. Therefore, online consumers need trust as a mental shortcut to reduce the complexity of conducting business transactions with online vendors (Luhmann, 1989). Such trust occurring in cyberspace is commonly termed "online trust," and we limit the scope to the online trust that is pertinent to B2C e-commerce, namely, the trust that occurs for an individual Internet user toward a specific e-commerce Web site or the online vendor that the Web site represents. Derived from the general definition for trust (Rousseau, Sitkin, Burt, & Camerer, 1998), online trust can be defined as follows: *an Internet user's psychological state of risk acceptance based upon the positive expectations of the intentions or behaviors of an online vendor*.

There are almost certainly many potential sources of influence that promote or hinder online trust. However, the current article focuses on the HCI or interface design perspective in inducing online trust, that is, to use what consumers can see on an e-commerce interface to affect their feelings of trust toward the online merchant that the e-commerce interface represents.

ONLINE TRUST IN THE HCI LITERATURE

Online trust is a relatively new research topic and has recently drawn great interest from researchers in HCI and human factors. There are several main themes that the majority of the existing studies can be divided into. First, some studies attempt to understand the online consumer's mind by investigating the underlying elements, antecedents, or determinants that are pertinent to the formation of online trust. For example, Gefen (2002) examined trust from a multi-dimensional perspective. According to the researcher, the specific beliefs of integrity, ability, and benevolence were seen as antecedents to overall trust. Other researchers, such as Corritore et al. (2003), also proposed that the consumer could perceive trust before, during, or after the online transaction, and they further concluded that online trust was characterized by its stage of development.

The second stream of studies focuses on conceptualizing trust into theoretical models or frameworks and dividing trust elements into various dimensions. For example, the Model of Trust for Electronic Commerce (MoTEC), is proposed by Egger (2001). The model consists of four components: the pre-interactional filters taking place before any online interaction, the interface properties of the Web site, the information content of the Web site, and relationship management. The Cheskin/Sapient Report (1999) focused on Web site interface cues and presented a model of six building blocks of online trust. These six building blocks were seals of approval, brand, navigation, fulfillment, presentation, and technology. The building blocks could be further divided into a total of 28 components to establish perceived trustworthiness. Such studies provide a theoretical account for exploring and enhancing trust in an online context and often take the effects of customer relationship management into consideration.

The third stream of studies aims to validate those conceptual frameworks or trust scales, often by analyzing data acquired directly from the consumers (e.g., Ba & Pavlov, 2002; Bhattacherjee, 2002). The main objective of these studies is to theoretically derive and empirically validate a scale that can be used to measure either individual online trust or the trustworthiness of an e-commerce Web site. In developing such an instrument, as for developing any other kind of scale, the researchers need to stress establishing its reliability, content validity, and construct validity. Factor analysis, structural equation modeling, and multiple linear regression analysis are some of the most commonly used statistical analysis methods in those efforts.

And finally, the rest of the studies suggest Web design guidelines that are intended to enhance consumer online experience and induce the feeling of trust from the consumers (Karvonen & Parkkinen, 2001; Kim & Moon, 1998; Neilsen, 2000). In other words, the main goal for the researchers of these studies is to explore Web interface design implications to maximize consumer trust or, more precisely, trust perception. A representative study of this kind is the Nielsen Norman Group Report (2000), in which explicit trust-inducing guidelines — including graphic design, surface cue, and Web usability features — are provided based on a large number of user testing observations carried out by experts.

These preceding studies provide important insights into trust in an online context. However, the research field of online trust is still far from maturity and expected to be significantly substantiated and enhanced. For example, the terms *element*, *antecedent*, *dimension*, *determinant*, and *principle* are sometimes used interchangeably due to the lack of agreement on a clear definition for each term among researchers in the field. Nevertheless, this is the current body of work from which any potential implementation is to be derived.

BUILD ONLINE TRUST BY WEB DESIGN

To initiate and build a consumer's online trust is inevitably a challenging task. Due to the nature of the Internet, people nowadays browse different e-commerce Web sites as fast as they switch TV channels. Consequently, to succeed in e-commerce, online vendors must be able to convey their trustworthiness to first-time visitors and effectively and efficiently build trust in the eyes of consumers. This requires online vendors to implement optimal electronic storefronts that can attract potential consumers and induce their trust. According to Ang & Lee (2000), "if the web site does not lead the consumer to believe that the merchant is trustworthy, no purchase decision will result" (p. 3). In other words, applying trust-inducing features to the Web sites of online vendors is the most effective method of enhancing online trust, given the current state of knowledge.

Efforts have been taken to establish a framework that classifies various trust-inducing Web design features into three broad dimensions: visual design, content design, and social-cue design (Wang & Emurian, in press). The framework is not exhaustive in the sense that it does not attempt to capture every possible trust-inducing feature that web designers can apply. It is focused on articulating the most prominent set of trust-inducing features and presenting them as an integrated entity that can be empirically evaluated and appropriately implemented in Web design. *Table 1* illustrates the framework in detail, including the explanations and design feature examples.

All the trust-inducing interface design factors that are identified in the framework have been illustrated on a synthetic e-commerce interface and evaluated by 181 survey respondents (Wang & Emurian, 2004). Along with identifying the three dimensions, the factors were found to significantly contribute to online trust ratings. This has confirmed what most HCI researchers believe — as Kim & Moon (1998) pointed out — that informative emotions such as trust can be triggered by the customer interfaces

Table 1. Framework of trust-inducing features

Dimensions	Explanations	Design Feature Examples
Visual Design	Defines the graphical design aspect and the structural organization of displayed information on the Web site.	• Use of three-dimensional and half-screen size clipart • Symmetric use of moderate pastel color of low brightness and cool tone • Use of well-chosen, good-shot photographs • Implementation of easy-to-use navigation (simplicity, consistency) • Use of accessible information (e.g., no broken links and missing pictures) • Use of navigation reinforcement (e.g., guides, tutorials, instructions, etc.) • Application of page design techniques (e.g., white space and margin, strict grouping, visual density, etc.)
Content Design	Refers to the informational components that can be included on the Web site, be they textual, graphical, etc.	• Display of brand-promoting information (e.g., prominent company logo or slogan, main selling point) • Up-front disclosure of all aspects of the customer relationship (company competence, security, privacy, financial, and legal concerns) • Display of seals of approval or third-party certificates • Use of comprehensive, correct, and current product information • Use of a relevant domain name
Social-Cue Design	Relates to embedding social and interpersonal cues, such as social presence and face-to-face interaction, into the Web interface via different communication media.	• Inclusion of a representative photograph or video clip • Use of synchronous communication media (e.g., instant messaging, chat lines, video telephony, etc.)

and further aid decision making while using e-commerce systems. It may also be concluded that the three dimensions of the framework can act together to promote online trust and reflect the different aspects of Web interface design.

The first two dimensions, which are visual design and content design, are seemingly straightforward, and they have been traditionally the focus of the research that aims to promote online trust by Web design. The last dimension, the social-cue design dimension, is a relatively new design strategy being suggested by numerous researchers (Riegelsberger, Sasse, & McCarthy, 2003; Steinbruck, Schaumburg, Duda, & Kruger, 2002; Wang, in review). This approach is aimed to remedy the prominent problem of e-commerce known as "lack of human touch" that eliminates online shopping for a considerable number of people.

At least two reasons, or disadvantageous characteristics of e-commerce, contribute to such a problem. First, e-commerce transactions are mostly separated in space and time. Second, a Web site is the only primary and direct "contact point" that online vendors can rely on to interact and communicate with their customers. While the face-to-face interaction can help to establish and stabilize consumer trust in off-line situations, the business transaction in e-commerce is deficient in the personal communication dimension. Therefore, there is need to bring e-commerce interactions closer to off-line shopping experiences by implementing social and interpersonal cues that moderate the disadvantages of an impersonal e-commerce interface and induce online trust. It is such an initiative that compels the social-cue design dimension of the framework.

The social and interpersonal cues, being investigated and embedded into e-commerce Web sites, refer to voice, gestures, appearance, and other communication cues that have been found to have a strong impact on triggering people's trust in face-to-face encounters. Using richer communication media has been seen as a valid means for facilitating the conveyance of these interpersonal cues and providing more opportunities for personal contacts between consumers and online vendors. With the advancement of technology and the increase of bandwidth, a huge collection of communication media is presently available. However, to choose suitable communication media for adding social or interpersonal cues in e-commerce Web sites, designers need to be aware of the different features of each medium, such as channel availability, synchrony, and channel symmetry (Greenspan, Goldberg, Wimer, & Basso, 2000). When implementing social cues, special care should also be taken, as advised by Riegelsberger et al. (2003), to prevent online shoppers from being disappointed by elements lacking functionality other than giving cues of social interaction.

The existing research on interpersonal cues and online trust still remains preliminary. Most research only focuses on examining the trust-inducing capacity of photography, which is the simplest form of communication media to be employed in e-commerce, and the outcomes are found to be somehow contradictory (e.g., Riegelsberger et al., 2003; Steinbruck et al., 2002). Therefore, empirical evidence and valid methodologies are in great need in this research area to address a number of intriguing research questions.

CONCLUSION AND FUTURE DIRECTIONS

As e-commerce gains widespread attention and rapidly emerges as a competitive business form, online merchants are facing an urgent challenge of building and sustaining consumer trust on the Internet. While the issue has initiated numerous investigations for valid research methods and effective solutions by researchers from multiple disciplines, this article shows the merit of an HCI approach in confronting the challenge. Adding trust-inducing interface features and interpersonal cues in Web design has been proposed as a fruitful strategy for building online trust.

Based upon the present overview, five potential areas of suggested research include (1) the effects of culture on online trust; (2) the effects of domain (e.g., .com, .edu, .org) on online trust; (3) the reasons for losing online trust and the ways to repair it; (4) the importance of civil remedies for consumers in case of violations of privacy laws; and (5) the transferability of online trust from the Internet to other activities. In addition, there is obvious need for further investigation on the effects of social and interpersonal cues on online trust, including both methodology development and experimental testing. Finally, it should be pointed out that while well-crafted Web interfaces can induce trust in those who intend to purchase online, online vendors should also pay attention to other methods, such as customer relationship management (CRM) and off-line marketing strategies, to obtain consumer trust and nurture strong business relationships (Tan, Yen, & Fang, 2002).

REFERENCES

Ang, L., & Lee, B.C. (2000). Influencing perceptions of trustworthiness in Internet commerce: A rational choice framework. In *Proceedings of Fifth CollECTer Conference on Electronic Commerce* (pp. 1-12). Brisbane.

Ba, S., & Pavlov, P.A. (2002). Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS Quarterly, 26* (3), 243-268.

Bhattacherjee, A. (2002). Individual trust in online firms: Scale development and initial test. *Journal of Management Information Systems, 19* (1), 211-241.

Cheskin Research and Studio Archetype/Sapient. (1999). *Ecommerce Trust Study*. Retrieved from *www.cheskin.com/p/ar.asp?mlid=7&arid=40&art=0.*

Corritore, C. L., Kracher, B., & Wiedenbeck, S. (2001). Trust in the online environment. In M.J. Smith, G. Salvendy, D. Harris, & R.J. Koubek (eds.), *Usability Evaluation and Interface Design: Cognitive Engineering, Intelligent Agents and Virtual Reality* (pp. 1548-1552). Mahway, NJ: Erlbaum.

Corritore, C. L., Kracher, B., & Wiedenbeck, S. (2003). Online trust: Concepts, evolving themes, and a model. *International Journal of Human-Computer Studies, 58*, 737-758.

Egger, F.N. (2001). Affective design of e-commerce user interface: How to maximize perceived trustworthiness. In *Proceedings of the International Conference on Affective Human Factors Design*. London: Asean Academic Press.

Gefen, D. (2002). Reflections on the dimensions of trust and trustworthiness among online consumers. *The DATA BASE for Advances in Information Systems, 33* (3), 38-53.

Grabner-Krauter, S., & Kaluscha, E. A. (2003). Empirical research in on-line trust: A review and critical assessment. *International Journal of Human-Computer Studies, 58*, 783-812.

Greenspan, S., Goldberg, D., Wimer, D., & Basso, A. (2000). Interpersonal trust and common ground in electronically mediated communication. In *Proceedings of the*

ACM2000 Conference on Computer Supported Cooperative Work, Philadelphia, PA.

Karvonen, K., & Parkkinen, J. (2001). Signs of trust. In *Proceedings of the 9th International Conference on HCI*. New Orleans, LA.

Kim, J., & Moon, J. Y. (1998). Designing towards emotional usability in customer interfaces: Trustworthiness of cyber-banking system interfaces. *Interacting with Computers*, *10*, 1-29.

Luhmann, N. (1989). *Vertrauen. Ein Mechanismus der Reduktion sozialer Komplexitat (3rd edition)*. Enke, Stuttgart.

Nielsen Norman Group. (2000). Trust: Design guidelines for e-commerce user experience. In *E-commerce User Experience*. Retrieved from *www.nngoup.com/reports/ecommerce.*

Neilsen, J. (2000). *Designing Web usability: The practice of simplicity*. Indianapolis: New Riders Publishing.

Nissenbaum, H. (2001). Securing trust online: Wisdom or oxymoron? *Boston University Law Review, 81*, 101-131.

Riegelsberger, J., Sasse, M. A., & McCarthy, J. D. (2003). Shiny happy people building trust? Photos on e-commerce Websites and consumer trust. In *Proceedings of CHI2003*, Ft. Lauderdale, FL.

Rousseau, D.M., Sitkin, S.B., Burt, R.S., & Camerer, C. (1998). Not so different after all: A cross disciplinary view of trust. *Academy of Management Review*, *23* (3), 393-404.

Steinbruck, U., Schaumburg, H., Duda, S., & Kruger, T. (2002). A picture says more than a thousand words: Photographs as trust builders in e-commerce Websites. In *Conference Extended Abstracts on Human Factors in Computer Systems*, Minneapolis, MN.

Tan, X., Yen, D.C., & Fang, X. (2002, Spring). Internet integrated customer relationship management: A key success factor for companies in the e-commerce arena. *Journal of Computer Information Systems*, 77-86.

Wang, Y.D., & Emurian, H. H. (in press). An overview of online trust: Concepts, elements, and implications. *Computers in Human Behavior.*

Wang, Y.D. & Emurian, H.H. (2004). Inducing consumer trust online: An empirical approach to testing e-commerce interface design features. In *Proceedings of the 15th International Conference,* New Orleans, LA.

KEY TERMS

Channel Availability: A feature of any communication medium. A communication medium's channel can be contextual, audio, visual, or any combination of the three. For example, telephone is an audio-only communication medium, while videoconferencing is an audio-visual communication medium.

Channel Symmetry: A feature of any communication medium. A communication medium affords symmetry if the recipient of a message can respond with the same type of message. For example, telephone and e-mail tools are symmetric (two-way) communication media, while television and Web sites are asymmetric (one-way) communication media.

Communication Media: The methods or tools in which information can be exchanged and communication can be facilitated. Examples include telephone, television, e-mail, Web sites, video conferencing, and instant messaging, to name a few.

Customer Relationship Management (CRM): An approach that recognizes that customers are the core of the business and that a company's success depends on effectively managing its relationship with them. CRM is about locating and attracting customers and thereby building long-term and sustainable relationships with them.

Electronic Commerce (E-Commerce): An emerging business form in which the process of buying, selling, or exchanging products, services, and information is undertaken via computer networks, including the Internet.

Online Trust: An Internet user's psychological state of risk acceptance based upon the positive expectations of the intentions or behaviors of an online vendor.

Synchrony: A feature of any communication medium. A communication medium is synchronous if the recipient of a message can respond immediately. For example, telephone and instant messaging are synchronous communication media; while e-mail and voice mail are asynchronous communication media.

Trust-inducing Design: The application of empirically verified features of a Web site to enhance a consumer's perception that the online vendor is trustworthy.

Trustworthy Web site: A Web site that reduces a consumer's perception of risk and that increases confidence in the online vendor's integrity.

Trust in Knowledge-Based Organizations

Maija-Leena Huotari
University of Helsinki, Finland

Mirja Iivonen
University of Tampere, Finland

INTRODUCTION

The knowledge-based society of the 21st century is characterized by knowledge generation as the primary source of wealth and social well-being. As partly intangible in nature, increased understanding of knowledge and information as a resource is critical (Sveiby, 1996; Teece, 1998). Such intangibles are gradually replacing traditional elements of power in states (Rosecrance, 1999), also emphasizing the role of trust in the positive aspect of economic globalization.

This development is supported by the resource and knowledge-based views of an enterprise (e.g., Barney, 1991; Grant, 1996; Penrose, 1959) and the idea of organizational learning (Argyris & Schön, 1996). A holistic view of contextual factors and means, such as trust, is necessary for managing knowledge and information processes to enhance them as a resource, whose value is difficult to estimate in economic, quantitative terms (Yates-Mercer & Bawden, 2002). To combine human, technological and structural factors in a unique manner is critical to prolonging strategic capability and sustainability. Knowledge creation occurs by combining people's distinct characteristics with a particular set of activities. Moreover, the core capabilities are created through activities consisting of values and norms, skills, managerial and physical systems (Leonard-Barton, 1995, p.25; Prahalad & Hamel, 1990). These four dimensions also relate to social capital.

In this article, we seek better understanding of knowledge-related processes in respect to trust and social capital. The foci are on trust as related to organizational culture and climate and to collaboration enhanced by appropriate organizational structures.

BACKGROUND

It is demanding to define the concept of knowledge management (KM) due to the difficulty of defining the concept of knowledge and its relation to the concept of information (see, e.g., Wilson, 2002). The perception of KM as providing knowledge representations/artefacts stored for use in information systems is close to information management (IM) (Davenport & Cronin, 2000; McInerney, 2002), though information and knowledge are distinctive concepts. The goal of KM is to transform information into learning, insight, and commitment to action, which requires turning personal knowledge into corporate knowledge to be widely shared and appropriately applied (Choo, 1998; Skyrme, 1997).

Knowledge is often understood to consist of explicit, implicit and tacit elements. Nonaka's (1994; Nonaka & Takeuchi, 1995) theory of knowledge creation, the SECI model, popularized Polanyi's (1966) identification of the tacit nature of knowledge in the mid-1990s. However, many authors argue that tacit knowledge is understood too superficially in the conceptions of KM (Yates-Mercer & Bawden, 2002) when actually referring to implicit knowledge (Nahapiet & Ghoshal, 1998, p.246; Orlikowski, 2002; Wilson, 2002). Tacit knowledge is the most intangible and very personal form in organizations and thus difficult to articulate. Manifested in organizational practices, the constructionist viewpoint that regards knowledge as a social construct (e.g., von Krogh, 1998) may prove appropriate to understand this phenomenon.

ORGANIZATIONAL KNOWING

The social nature of knowledge and information calls for the concept of organizational knowing and its management (Brown & Duguid, 2000; Choo, 1998; Choo & Bontis, 2002). Cook and Brown (1999) claim that innovation as "a generative dance" is an outcome of the interplay of knowledge and knowing. Orlikowski (2002) views cognition and action as inseparable that makes articulation of tacit knowledge unnecessary. The concept of a community of practice has evolved as knowledge appearing as a collection of processes that allow learning to occur and knowing to be internalized (McInerney, 2002, p.1012). Blackler (2002) argues that the five types of knowledge (embrained, embodied, encultured, embedded, encoded) are insufficient to account for knowledge as a social process. Boisot's (1998) contention is that the evolution of knowledge forms a social learning circle: through the codification of shared experience, personal knowledge can become proprietary knowledge or the intellectual capital of an organization. Once externally scrutinized, this knowledge becomes public while widely internalized, it turns into common sense. The assumption that knowledge exists in people's minds makes it hard to manage causing much debate about the relevance of the concept of KM. The management of people is as crucial as the management of information when aiming at the strategic management of knowledge and information as a resource (Huotari & Iivonen, 2004). Thus, the structure of the organization itself becomes critical to sharing, based on trust, and re-creation of knowledge throughout the organization.

TRUST

Trust is based on expectations of other people's willingness and ability to fulfill our needs and wishes (e.g., Fukuyama, 1996). That presupposes similar or related worldviews and shared meanings. This refers to normative trust assuming that common values provide a frame of reference for social norms creating predictability and trustworthiness. "Trust is… an intersubjective 'reality' that cannot exist… unless the symbols used to signal trustworthiness have meaning for all parties." (Hardy, Phillips & Lawrence, 1998, p.70). However, this type of trust has been challenged, too (Lane, 1998, p.8).

Trust is manifested in people's behavioral patterns, and honesty and predictability build it up (e.g., Ciancutti & Steding, 2000; Shaw, 1997). Trust has been called a communicative, sense-making process in which shared meanings develop to bridge disparate groups also in interorganizational relationships (Hardy et al., 1998). As a social phenomenon, trust is therefore a highly desirable property that also affects the well-being of the work community, produces commitment and internalised accountability, and provides a way to cope with risk and uncertainty (Huotari & Iivonen, 2004; Lane, 1998).

FUTURE TRENDS

The major managerial challenge in the globalized economy is the pooling of the intellectual capital of collaborators in a partnership. Trust is the basis for and co-evolution of social capital and through it is manifest in the concept of intellectual capital. Social capital, in particular, whose foundation is human behaviour may gain more emphasis in the future. Because personal relationships provide the basis for unique, networked organizational activities differences in types and levels of developed trust may result in different levels of resource exchange and flows (Nahapiet & Ghoshal, 1998, pp.245, 255; Tsai & Ghoshal, 1998; see also, Adler & Kwon, 2002).

Organizational culture and climate are revealed by values which form the basis for social norms and refer to social capital. Normative trust is essential for organizational knowledge involving the creation and maintenance of trust, and the resulting norms of behavior, that are important for knowledge sharing, for example, in communities of practice (Tuomi, 2002; Wenger, McDermott, & Snyder, 2002). Such facilitators as boundary spanners or roamers are an important component of the infrastructure of communities of practice contributing to the diffusion of knowledge between communities (Davenport & Hall, 2002) thereby strengthening social capital.

The development of a knowledge culture may promote learning and sharing of what is known. Trust is a core organizational value allowing people to communicate openly and without fear of unethical conduct. Dialog can build a culture for the continuing creation and sharing of knowledge. Moreover, trust is fundamental in virtual communities (McInerney, 2002, p.1014; Sonnenwald, 2004).

Collaboration relates strongly to trust and networking and enables converting individual knowledge into organizational knowledge. It is a cornerstone of social capital and necessary for innovating and accomplishing tasks in knowledge-based organizations. Collaboration and trust have a two-way relationship (Huotari & Iivonen, 2004). Co-workers learn to trust each other, but swift trust is required when collaborating without a long history of working together (e.g., Davenport & McLaughlin, 2004). Networking requires that collaboration as a behavioral model is encouraged by appropriate strategies related to the communicative and sense-making processes to cross borderlines and bridge gaps. For example, the cognitive distance of network members inhibiting collaboration can be decreased by establishing so called "epistemic communities" of shared mental categories, meanings and interpretations (Nooteboom, 2002, pp.23-29). These strategies facilitate trust and can strengthen social capital.

Organizational hierarchy may also affect knowledge sharing and the type of trust or mistrust developing. A low hierarchy and open organizational culture enhances information flows whereas a strong hierarchy has a negative effect on knowledge sharing in a multiunit organization where units compete against each other. Informal lateral relations, in turn, have a positive effect on knowledge sharing among units that compete in the market place but not when competing for internal resources. Thus external market competition in particular influences knowledge sharing and allows units to accumulate social capital (Hansen, 2002; Tsai, 2002). Social capital facilitates the creation of intellectual capital (Nahapiet & Ghoshal, 1998, p.246). Social interaction and trust increase knowledge creation through resource exchange and combination having a positive effect on product innovations and enhancing trustworthiness. A shared vision, in turn, can influence resource exchange and pooling only indirectly via its influence on trust (Tsai & Ghoshal, 1998).

CONCLUSION

This article focused on the role of trust in managing the knowledge-based, networked organization. Knowledge as organizational knowing and the importance of understanding the impact of intangible and social factors on knowledge generation were emphasized. Trust was examined within the framework of social capital, whose relation to intellectual capital was highlighted. The foci were on the role of trust in the development of organizational culture and climate, structure and collaboration. Managing intangibles may challenge the organization in the future. Therefore, their impact on knowledge generation and human behavior should be examined in theory and practice.

REFERENCES

Adler, P.S., & Kwon, S.-W. (2002). Social capital: Prospects for a new concept. *Academy of Management Review*, *27*(1), 17-40.

Argyris, C., & Schön, D.A. (1996). *Organizational learning II: Theory, method and practice*. Reading, MA: Addison-Wesley.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, *17*(1), 99-120.

Blackler, F. (2002). Knowledge, knowledge work and organizations: An overview and interpretation. In C.W. Choo & N. Bontis (Eds.), *The strategic management of intellectual capital and organizational knowledge* (pp. 47-64). New York: Oxford University Press.

Boisot, M.H. (1998). *Knowledge assets. Securing competitive advantage in the information economy*. New York: New York University Press.

Brown, J.S., & Duguid, P. (2000). *The social life of information*. Boston, MA: Harvard Business School Press.

Choo, C.W. (1998). *The knowing organization: How organizations use information to construct meaning, create knowledge, and make decisions*. New York: Oxford University Press.

Choo, C.W., & Bontis, N. (2002). *The strategic management of intellectual capital and organizational knowledge*. New York: Oxford University Press.

Ciancutti, A., & Steding T.L. (2000). *Built on trust: Gaining competitive advantage in any organizations*. Chicago, IL: Contemporary Books.

Cook, S.D.N., & Brown, J.S. (1999). Bridging epistemologies: The generative dance between organizational knowledge and organizational knowing. *Organization Science*, *10*(4), 381-400.

Davenport, E., & Cronin, B. (2000). Knowledge management. Semantic drift or conceptual shift? *Journal of Education for Library and Information Science*, *41*(4), 294-306.

Davenport, E., & Hall, H. (2002). Organizational knowledge and communities of practice. *Annual Review of Information Science and Technology (ARIST)*, *36*, 171-227.

Davenport, E., & McLaughlin, L. (2004). Interpersonal trust in online partnerships: The challenge of representation. In M.-L. Huotari & M. Iivonen (Eds.), *Trust in knowledge management and systems in organizations* (pp.107-123). Hershey, PA: Idea Group Publishing.

Fukuyama, F. (1996). *Trust: The social virtues and the creation of prosperity*. New York: The Free Press.

Grant, R.M. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal*, *17*(Winter), 109-122.

Hansen, M.T. (2002). Knowledge networks: Explaining effective knowledge sharing in multiunit companies. *Organization Science*, *13*(3), 232-248.

Hardy, C., Phillips, N., & Lawrence, T. (1998). Distinguishing trust and power in interorganizational relations: Forms and facades of trust. In C. Lane & R. Bachmann (Eds.), *Trust within and between organizations: Conceptual issues and empirical applications* (pp.64-87). New York: Oxford University Press.

Huotari, M.-L., & Iivonen, M. (2004). Managing knowledge-based organizations through trust. In M.-L. Huotari & M. Iivonen (Eds.), *Trust in knowledge management and systems in organizations* (pp. 1-29). Hershey, PA: Idea Group Publishing.

Ingwersen, P. (1992). *Information retrieval interaction*. London: Taylor Graham.

Lane, C. (1998). Introduction: Theories and issues in the study of trust. In C. Lane & R. Bachmann (Eds.), *Trust within and between organizations: Conceptual issues and empirical applications* (pp.1-30). New York: Oxford University Press.

Leonard-Barton, D. (1995). *Wellsprings of knowledge: Building and sustaining the sources of innovation*. Boston, Mass.: Harvard Business School Press.

McInerney, C. (2002). Knowledge management and the dynamic nature of knowledge. *Journal of the American Society for Information Science and Technology*, *53*(12), 1009-1018.

Nahapiet, J., & Ghoshal S. (1998). Social capital, intellectual capital and organizational advantage. *Academy of Management Review*, *23*(2), 242-266.

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science*, *5*(1), 14-37.

Nonaka, I., & Takeuchi, H. (1995). *The knowledge creating company. How Japanese companies create the dynamics of innovation*. New York: Oxford University Press.

Nooteboom, B. (2002). *Trust: Forms, foundations, functions, failures and figures*. Cheltenham: Edward Elgar.

Orlikowski, W.J. (2002). Knowing in practice: Enacting a collective capability in distributed organization. *Organization Science*, *13*(3), 249-273.

Penrose, E.T. (1959). *The theory of the growth of the firm*. New York: Wiley.

Polanyi, M. (1966). *The tacit dimension*. London: Routledge & Keagan.

Prahalad, C.K., & Hamel, G. (1990). The core competence of the corporation. *Harvard Business Review*, May-June, 79-91.

Rosecrance, R. (1999). *The rise of the virtual state: Wealth and power in the coming century*. New York: Basic Books.

Scott, J.E. (2000). Facilitating interorganizational learning with information technology. *Journal of Management Information Systems*, *17*(2), 81-144.

Shaw, R.B. (1997). *Trust in balance: Building successful organizations on results, integrity, and concer*. San Francisco, CA: Jossey-Bass Publications.

Skyrme, D.J. (1997). Knowledge management: Making sense of an oxymoron. *Management Insights*, 22. Retrieved on February 18, 2004, from *http://www.skyrme.com/insights/22km.htm*

Sonnenwald, D.H. (2004). Managing cognitive and affective trust in the conceptual R&D organization. In M.-L. Huotari & M. Iivonen (Eds.), *Trust in knowledge management and systems in organizations* (pp.82-106). Hershey, PA: Idea Group Publishing.

Sonnenwald, D.H., & L. Pierce, L. (2000). Information behavior in dynamic group work contexts: Interwoven situational awareness, dense social networks, and contested collaboration in command and control. *Information Processing & Management*, *36*(3), 461-479.

Stewart, T.A. (2001). *The wealth of knowledge: Intellectual capital and the twenty-first century organization*. London: Nicholas Brealey.

Sveiby, K.-E. (1996). What is knowledge management? Retrieved on February 18, 2004, from *http://www.sveiby.com/articles/KnowledgeManagement.html*

Teece, D.J. (1998). Capturing value from knowledge assets: The new economy of markets for know-how, and intangible assets. *California Management Review*, *40*(3), 55-79.

Tsai, W. (2002). Social structure of "coopetition" within a multiunit organization: Coordination, competition, and intraorganizational knowledge sharing. *Organization Science*, *13*(2), 179-190.

Tsai, W., & Ghoshal, S. (1998). Social capital and value creation: The role of intrafirm networks. *Academy of Management Review*, *41*(4), 464-476.

Tuomi, I. (2002). *Networks of innovation: Change and meaning in the age of the Internet*. Oxford: Oxford University Press.

von Krogh, G. (1998). Care in knowledge creation. *California Management Review*, *40*(3), 133-153.

Wenger, E., McDermott, R., & Snyder, W.M. (2002). *Cultivating communities of practice*. Boston, Mass.: Harvard Business School Press.

Wilson, T.D. (2002). The nonsense of "knowledge management". *Information Research*, *8*(1), paper no. 144. Retrieved on February 18, 2004, from http://InformationR.net/ir/8-1/paper144.html

Yates-Mercer, P., & Bawden, D. (2002). Managing the paradox: The valuation of knowledge and knowledge management. *Journal of Information Science*, *28*(1), 19-29.

KEY TERMS

Collaboration: Human behavior, sharing of meaning and completion of activities with respect to a common goal and taking place in a particular social or work setting (Sonnenwald & Pierce, 2000).

Information: A message, or data, which makes a difference. Information has meaning, and becomes knowledge when a person internalizes it.

Intellectual Capital: Knowledge that transforms raw materials (both tangible and intangible) thereby increasing their value. It can be divided into human, structural and customer capital. Human capital refers to the talent of employees. Structural capital involves intellectual property, methodologies, software, documents, and other knowledge artefacts. Customer capital refers to client relationships. By integrating organizational knowledge into intangible assets an organization aims to turn human capital into structural capital (Stewart, 2001).

Knowledge: From the cognitive viewpoint, is an individual's total understanding of himself/herself and the surrounding world at any given point in time, incorporating (sub)conscious memory, thinking and cognition, as well as emotional and intuitive properties (Ingwersen, 1992, pp.228-229). When communicated, knowledge becomes information and consequently the raw material of new knowledge.

Knowledge Management: Combines human resources management and information management, and is defined as the management of all processes concerned with the identification, acquisition, creation, storage**,** distribution, and use of information and knowledge (Huotari & Iivonen, 2004).

Organizational Culture: Embraces two levels: deep-seated ideas and beliefs, and espoused values that form part of the cultural knowledge embedded within organizations. It is reflected in the organizational climate.

The SECI Model: Assumes knowledge creation as an iterative process converting tacit and explicit knowledge through socialization, externalization, combination and internalization (Nonaka & Takeuchi, 1995).

Social Capital: The sum of the actual and potential resources embedded within, available through, and derived from the network of relationships possessed by an individual or a social unit. It includes:

- **The relational dimension** concerned with the behavioral embeddedness involving the nature of the relationships developed over time, including trust, norms, obligations and identification.
- **The structural dimension** related to ties of social interaction such as density, hierarchy, network configuration, and so forth, between actors.
- **The cognitive dimension** related to shared representations, interpretations and systems of meaning, for example, to a shared vision (Nahapiet & Ghoshal, 1998, pp.243-245).

Social capital focuses on networks and partly overlaps with intellectual capital, which covers other aspects as well.

Trust: Can be defined as follows:

- **Affect-based trust** has emotional connotations and is related to issues like care, concern, benevolence, altruism, commitment, mutual respect, and so forth (Scott, 2000).
- **Calculative trust** is based on the weights of the costs and benefits of certain actions, and on a view of man as a rational actor.
- **Cognition-based trust** is a rational view of trust and is associated with competence, ability, responsibility, integrity, credibility, reliability, and dependability.
- **Non-calculative trust** is based on values and norms (Lane, 1998; Nooteboom, 2002).
- **Normative trust** refers to common values that give a frame of reference to social norms creating predictability and trustworthiness.
- **Swift trust** coincides with risk and uncertainty. It is needed in circumstances where the history of working together is limited to learn trust.

Trust in Technology Partnerships

Kirsimarja Blomqvist
Lappeenranta University of Technology, Finland

INTRODUCTION

Technology partnerships are seen as a key element both in knowledge creation for technological development itself and in leveraging the value of technological knowledge (Bidault & Cummings, 1994; Teece, 1987; Ford, 1998). Complementary knowledge is generally regarded as a source for competitiveness (Doz & Hamel, 1998; Powell, 1998; Nahapiet & Ghoshal, 2000). Potentially, the focus on core competencies enables relatively stronger competitiveness (cumulative learning, focused use of critical resources) and ability to gain synergistic benefits and scale by leveraging different knowledge bases and networks (Miles, 2000; Blomqvist, 2002).

In the converging ICT sector, the technological discontinuities like the Internet bring new innovative players, which try to break the rules of competition and seek the niches not noticed by the incumbent players. For incumbents the discontinuities pose a threat that may be turned into an opportunity through cooperating with innovative players and learning from them. In technology partnerships, complementary companies are able to focus on their core competencies and simultaneously leverage external knowledge and resources to complement their knowledge and resource base.

According to a recent study, 94% of the technology executives believed that alliances and technology partnerships were becoming more critical to their strategy, yet every second alliance turns out to become a failure (Kelly, Schaan & Joncas, 2002). In several studies,, empirical trust has been identified as a key success factor for technology partnerships (Forrest & Martin, 1992; Bruce, 1995). The aim of this article is to increase the understanding of trust as a critical factor in technology partnerships. The empirical illustrations in this article are based on research on asymmetric technology partnerships, in which eight small software firms' partnerships with five large firms were empirically analyzed (Blomqvist, 2002; Blomqvist & Ståhle, 2003). The knowledge creation in technology partnerships between large and small firms, as well as the converging ICT market as an example of complex and dynamic business environment, make the context especially interesting from the point of view of trust. Perceived or assumed dissimilarities in complementary actors' values, goals, time-horizon, decision-making processes, culture, and logic of strategy can form barriers to cooperation (see Doz, 1988; Blomqvist, 2002). Rapid changes and high risks concerning technological success and economic rewards are typical.

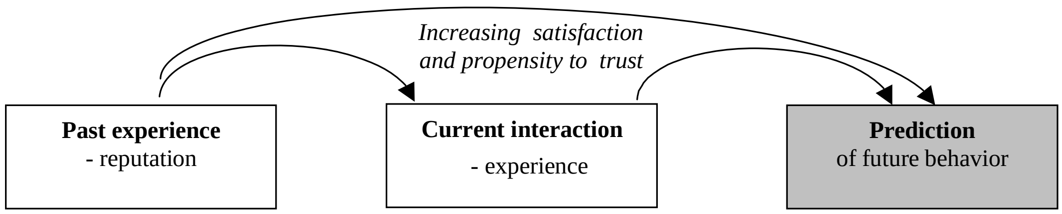
TRUST AND INTER-FIRM COLLABORATION

Trust has been identified as an important issue in partnerships and alliances (Håkanson, 1993; Parkhe, 1993, 1998; Das & Teng, 1998; Lazarec, 1998; Ariño, de la Torre & Ring, 2000) and supplier relations (Sako, 1998). It has been proposed that some *cooperative threshold amount of trust* is needed for cooperation to evolve (Dibben, 2001). For the context of asymmetric technology partnerships, based on an extensive literature analysis and conceptualization, Blomqvist (1997) has defined trust as the "*actor's expectation of the other party's competence and goodwill.*" This definition includes the dimensions of competence (i.e., technological capabilities, skills, and know-how) and the more abstract goodwill, which implies moral responsibility and positive intentions toward the other. Signs of *goodwill* and the related expectation of the partner's positive intentions are necessary for a partner to be able to accept risk and a potentially vulnerable position. It could be assumed that *competence* is a necessary antecedent for trust in the business context, especially in technology partnerships, where complementary technological knowledge and competencies are a key motivation for partnership formation (Blomqvist, 2002). Also Sako (1998) has included competence in his definition of trust.

There is a strong temporal dimension in trusting. Trust is seen as evolving from past experience and current interaction (Deutch, 1973). It is seen as an outcome of a process, that is, trust relationships develop gradually (see Figure 1).

Trust between partners can be said to be a bridge between past experiences and the anticipated future. In general, trust is believed to evolve slowly, through repeated interactions of increasing satisfaction (e.g., Blau, 1964) and through incremental investments and experiences. Reputation and experienced *similarity both in character and values* enhance the experience of trust

Figure 1. Trust as a prediction resulting from satisfactory interaction and past experience (Blomqvist, 2002)



(Zucker, 1987; Gulati, 1995; Jones & George, 1998) through the ability to predict the other's behavior.

THE CRITICAL ROLE OF TRUST IN TECHNOLOGY PARTNERSHIPS

Blomqvist (2002) has used interviews of business practitioners from large and small technology firms representing operative and strategic organizational levels in order to look deeper into the role of trust in technology partnerships. The components of trust were formed through interviews based on open questions: How do you evaluate your partner? What characteristics do you find important? What do you tell and highlight about you and your company when presenting yourself to a potential partner? The quotes in this article are drawn from these in-depth interviews.

The interviewed large firm managers, small firm managing directors, and other key persons very clearly saw trust to be critical for technology partnering. The interviews confirmed the two components of trust—capability and goodwill—that were brought up also in the literature review. However, the interviews included many viewpoints that could not be labeled under competence or goodwill, but dealt often with the other party's behavior, for example, what really happened in the cooperation. Already at the very first meetings, the behavioral dimension of trust is present in signs and signals, for example, what information is revealed and in which manner. Signs of goodwill are necessary for the trusting party to be able to accept a potentially vulnerable position. Through the partnering process (a long time), the actual behavior (e.g., kept promises) becomes more visible and easier to evaluate. This was highlighted for example as follows:

"I evaluate a potential partner with feelings anyway. It is the first couple of minutes...The first impression is important... how they present themselves, whether they are really interested, and whether they have a vision. I also look whether they are able to communicate their vision and have self-respect. It is the feeling…Yes, I don't know if I should have, but I don't have any systematic way of evaluating them. You listen to their story and evaluate whether it is credible or not." (Partner Director, Large ICT Company B)

"Communication is very critical…trust is created between individuals, not organizations. Individuals act like representatives of their organizations…You cannot trust a large firm's organization more than the small firm's organization." (Technological Expert, Internet Technology Net)

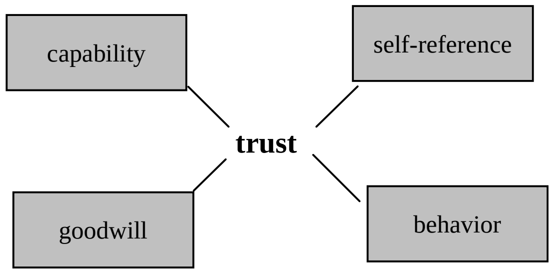
The interviews of large and small firm managers produced empirical evidence about the role of self-reference in asymmetric technology partnership formation. In technology partnerships the individuals and organizations can be referred to as self-referential, if they are aware of their identity, capabilities, and values. A self-referential actor has an ability of building relationships and is more willing to accept interdependency (double contingency). Some of the interviewed managers discussed self-reference as follows:

"First (when introducing your company to a potential partner) you need to have focused your business so that the other party understands that and sees it clearly. Then they can see my role, my firm's role, and how we and they will work together…and what they need and can get from my company." (Managing Director, Small Software Company Alpha)

"The ability to build trust comes from the ability to be humble. If you have been able to dictate what will be done…It will be a long drop and major change in the mental mode to actually listen to what the other person is after and what they want and to think of them as equals." (Partner Director, Large ICT Company)

As a result of the analyses of all the perspectives of the interviewees, two new components of trust—*behavior* and *self-reference*—were included. Subsequently the

Figure 2. Components of trust



concept of trust can be seen to include four components: capability, goodwill, behavior, and self-reference (see Figure 2).

Some factors related to each component were identified. In the technology partnership formation context, *capability* consists of *technological capability*, *business capability*, and *meta-capability to cooperate*. Capability is a necessary but passive component for trust in a technology partnership formation. Capability component can be explicitly emphasized and analyzed when negotiating for technology partnerships. Capability can be communicated as references of difficult projects, prizes for outstanding technology, first-to-market technologies, prominent members of the board, venture capital, and symbols like certificates and employee education.

Goodwill implies a more abstract, yet a very important component of trust in technology partnerships. It was defined earlier as the "partner's moral responsibility and positive intentions toward the other" (Blomqvist, 1997). Also Tyler (1996) and Sako (1998) have highlighted the trustee's intentions as a source of experienced trustworthiness. Experienced goodwill is a necessary and active component for trust in technology partnership formation. The goodwill component is more difficult to discern explicitly, but negotiating parties can both stress goodwill in their behavior and attempt to evaluate goodwill in other parties' behavior. Reciprocity, care, and concern, as well as social and characteristic similarity, may enhance feelings of goodwill. Expressed norms such as keeping promises and active consideration for mutual norms also build trust through goodwill (Blomqvist, 2002).

As the relationship is developing, the actual *behavio*r (e.g., that the trustee fulfills the positive intentions) adds to the trustworthiness (Bidault, 2000; Lazarec & Lorenz, 1998). Thus the capability and goodwill dimensions of trust become visible in behavioral signals of trustworthiness. Open communication, adaptation, and commitment are critical behavioral factors building trust (Blomqvist, 2002).

Clear self-reference enables the individual or organization to connect to other actors and to cooperate (Luhmann, 1979). At the corporate level, self-reference is the basis for strategy and business plans. At the individual level, the individual's clear identity and values form a basis for self-reference. Without clear self-reference, recognizing the right partners (strategic and cultural compatibility) and establishing partnerships is difficult. Self-reference enables the actor to trust, to be trusted, and to cooperate. It is proposed that an individual or organization with a strong self-reference is able to recognize, maintain, and develop the heterogeneous strength in its identity, yet connect and cooperate at an equal level with diverse and complementary actors.

Therefore, it can be summarized that in technology partnerships, *"Trust is an actor's expectation on the other party's capability, goodwill, and self-reference, which needs to be confirmed by experience."* Thus trust is increased by—and decreased by the lack of—these components in parties' actual behavior and communication.

CONCLUSION

In this chapter the concept and phenomenon of trust, with special focus on technology partnerships, has been examined. The basic conceptualization of trust was based on Blomqvist's research (2002) on trust in asymmetric technology partnerships, in which the conceptual analyses with the four components—capability, goodwill, self-reference, and behavior—was based on both literature and empirical data. Some empirical illustrations were shown to support conceptualization.

It has been shown that trust is a critical issue in technology partnerships. However, on the way toward methods and tools for trust building in partnership formation, we are only in the beginning (Blomqvist & Ståhle, 2003). It is believed that this research area with its multidisciplinary nature will be quintessential for technology firms competing in the dynamic business environment of today and the future.

REFERENCES

Ariño, A., de la Torre, J. & Ring, P.S. (1999). The role of trust in inter-organizational alliances: Relational quality and partner behavior. *Ciber Working Chapter Series*, No. 99-22, The Anderson School at UCLA, Los Angeles, California.

Bidault, F. & Cummings, T. (1994). Innovating through alliances: Expectations and limitations. *R&D Management*, *24*(1), 33-45.

Blau, P.M. (1964). *Exchange and power in social life*. New York: John Wiley & Sons.

Blomqvist, K. (1997). The many faces of trust. *Scandinavian Journal of Management, 13*(3), 271-286.

Blomqvist, K. (2002). *Partnering in the dynamic environment: The role of trust in asymmetric technology partnership formation* (p. 122). Acta Universitatis Lappeenrantaensis.

Blomqvist, K. & Ståhle, P. (2003). Trust in technology partnerships. In M.-L. Huotari & M. Iivonen (Eds.), *Trust in knowledge management and systems in organizations*. Hershey, PA: Idea Group Publishing.

Bruce, M., Leveric, F. & Wilson, D. (1995). Success factors for collaborative product development: A study of supplier of information and communication technology. *R&D Management*, *25*(1), 535-552.

Das, T.K & Teng, B.-S. (1998). Between trust and control: Developing confidence in partner cooperation in alliances. *Academy of Management Review*, *23*(3), 491-512.

Deutch, M. (1973). *The resolution of conflict: Constructive and destructive processes*. New Haven, CT: Yale University Press.

Dibben, M.R. (2000). *Exploring interpersonal trust in the entrepreneurial venture*. Basingstoke: Macmillan Press.

Dodgson, M. (1993). Learning, trust, and technological collaboration. *Human Relations*, *46*(1), 77-94.

Doz Yves, L. (1988). Technology partnerships between larger and smaller firms: Some critical issues in cooperative strategies in international business. In F.J. Contractor & P. Lorange (Eds.), (pp. 317-338). Boston: L-e-x-i-n-g-ton Books.

Doz, Y.L. & Hamel, G. (1998). *Alliance advantage: The art of creating value through partnering*. Boston: Harvard Business School Press.

Ford, D. (1998). *Managing business relationships*. Chichester: John Wiley & Sons.

Forrest, J.E. & Martin, M.J.C. (1992). Strategic alliances between large and small research intensive organizations: Experiences in the biotechnology industry. *R&D Management, 22*(1), 41-53.

Gulati, R. (1995). Does familiarity breed trust? The implications of repeated ties for contractual choice in alliances. *Academy of Management Journal*, *38*(1), 85-112.

Håkanson, L. (1993). Managing cooperative research and development: Partner selection and contract design. *R&D Management*, *23*(4), 273-285.

Jones, G.R. & George, J.M. (1998). The experience and evolution of trust: Implications for cooperation and teamwork. *Academy of Management Review*, *23*(3), 531-546.

Kelley, M.J., Schaan, J.-L. & Joncas, H. (2002). Managing alliance relationships: key challenges in the early stages of collaboration. *R&D Management*, *32*(1), 11-22.

Lazarec, N. & Lorenz, E. (1998). Trust and organisational learning during inter-firm cooperation. In N. Lazaric & E. Lorenz (Eds.), *Trust and economic learning* (pp. 209-226). Cheltenham: Edward Elgar Publishing.

Luhmann, N. (1979). *Trust and power*. Chichester: John Wiley & Sons.

Miles, R.E., Snow, C.C. & Miles, G. (2000). TheFuture.org. *Long Range Planning*, *33*(3), 297-474.

Nahapiet, J. & Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage. *Academy of Management Review*, *23*(2), 242-266.

Paija, L. (2000). *ICT cluster—the engine of knowledge-driven growth in Finland*. Discussion Paper No. 733, The Research Institute of the Finnish Economy, Helsinki, Finland.

Parkhe, A. (1993, October). Trust in international joint ventures. *Proceedings of the Academy of International Business Meetings,* Hawaii, USA.

Parkhe, A. (1998). Understanding trust in international alliances. *Journal of World Business, 33*(3), 219-240.

Powell, W.M. (1998). Learning from collaboration: Knowledge and networks in the biotechnology and pharmaceutical industries. *California Management Review*, *40*(3), 228-240.

Sako, M. (1998). The information requirements of trust in supplier relations: Evidence from Japan, Europe and the United States. In N. Lazaric & E. Lorenz (Eds.), *Trust and economic learning* (pp. 23-37). Cheltenham: Edward Elgar Publishing.

Ståhle, P. (1998). *Supporting a system's capacity for self-renewal*. Research Report 90, Department of Teacher Education. University of Helsinki, Finland.

Teece, D.J. (1987). Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. In D.J. Teece (Ed.), *The competitive challenge* (pp. 187-219). Berkeley, CA: University of California.

Tyler, T.R. & Kramer, R.M. (1996). Whither trust? In R.M. Kramer & T.R. Tyler (Eds.), *Trust in organizations, frontiers of theory and research* (pp. 1-15). London: Sage Publications.

Zucker, L.G. (1986). Production of trust: Institutional sources of economic structure, 1840-1920. *Research in Organizational Behavior*, *8*, 53-111.

KEY TERMS

Asymmetry: Here we use asymmetry to mean that parties are heterogeneous and possess diverse knowledge bases. Asymmetry manifests also in corporate culture and management. As a result of Blomqvist's study (2002), asymmetry will be defined as "difference in resources, capabilities, and power, as well as management and culture of actors."

Convergence: In the ICT industry convergence at the *firm level* means that firms cooperate to combine their different knowledge bases. Also industries, such as communications and media industries, can be said to converge if focal actors leverage knowledge from two previously separated industries.

Information and Communications Technology (ICT): Depending on the author, ICT equipment, network operation and network services, and digital content provision can be listed as *key industries*. *Related industries* (e.g., entertainment and traditional media or consumer electronics) are getting closer to the key industries by acquisitions, joint ventures, and established new business units. It is difficult to describe industry boundaries as clear-cut, as boundaries are blurring and actors are searching the best possible position in the value network.

Technology Partnership: In a technology partnership, the technology and a common objective are included in inter-firm collaboration. R&D collaboration is a typical example.

Trust: An actor's expectation on the other party's capability, goodwill, and self-reference, which needs to be confirmed by experience. In the context of technology partnership, formation capability consists of technological capability, business capability, and the meta-capability to cooperate.

Trust in Virtual Enterprises

T.T. Wong
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hong Kong

INTRODUCTION

For a virtual enterprise(VE) to be competitive and successful in a dynamic environment characterized by constantly changing customer demands and technological innovations, it must be capable of rapid adjustment in order to reduce the time and cost needed to deliver to the customer a quality product. The success in the agile competition naturally depends on the effective operation of the VE.

Trust has been recently regarded as the foundation of the digital economy (Keen, 2000; Stewart et al., 2002). The e-business environment is characterized by:

a) the impersonal nature of the online environment;
b) the extensive use of communication technology as opposed to face-to-face transactions;
c) the implicit uncertainty of using an open technological infrastructure for transactions; and
d) the newness of the transaction medium.

Given these attributes, trust development in VEs presents significant challenges because it is difficult to assess partners' trustworthiness without ever having met them (McDonough et al., 2001). Moreover, as the life of many virtual teams is relatively limited, trust must quickly develop (Jarvenpaa & Leidner, 1999). Yet, trust development is deemed crucial for the successful completion of virtual team projects (Sarker et al., 2001). The objective of this chapter is to identify the essential conditions for trust building in VE and to evaluate the likely impact of trust on the operations and management of VE. The study raises a number of issues to be explored by future research.

BACKGROUND

Trust encompasses constructs as diverse as ethics, morals, emotions, values, and natural attitudes. Further, trust spans interdisciplinary fields, including philosophy, computer science, economics, and organizational behavior. Consequently, there is a myriad of definitions of trust. Since the dyadic view of familiarity-based trust does not readily correspond to the reality of the VE environment, which involves a large number of potential partners, the author proposes to define trust as the expectation by one member of ethical behavior on the part of the other member in a joint VE project. This definition emphasizes the importance of trust in organizational relationships and includes the idea of a joint undertaking, implying that there is a level of understanding of shared business goals and practices between the partners. Another factor that is implicit in the definition is the role of ethical behavior. Any change in ethics will cause a change in actions and thus influence trust. Finally, the definition implies that the VEs will undertake jointly to contribute to the final outcome.

According to traditional studies, trust builds incrementally and accumulates over time. VE business relationships, however, are characterized by project-oriented relationships that may entail no past history nor any plan for future association. In these temporary relationships, time is a vital but often elusive component in the trust building process. This does not mean, however, that trust cannot be apparent in temporary groups. On the contrary, McKnigh et al. (1998) have shown that trust in initial relationships can often be high. Further, Jarvenpaa & Leidner (1999) argue that trust is maximally important in new and temporary organizations, because it acts as a substitute for the traditional mechanisms of control and coordination.

Creating a VE takes more than just information technology. A study on issues of information technology and management concluded that there is no evidence that IT provides options with long-term sustainable competitive advantage. The real benefits of IT derive from the constructive combination of IT with organization culture, supporting the trend towards new, more flexible forms of organization. Information technology's power is not in how it changes the organization, but the potential it provides for allowing people to change themselves. Creating these changes, however, presents a new set of human issues. Among the biggest of these challenges is the issue of trust between partner organizations in the VE (Wong & Lau, 2002). The following section identifies the essential conditions for trust development in a VE.

ESSENTIAL CONDITIONS FOR TRUST DEVELOPMENT

From the literature, the essential conditions for trust development in VE may be summarized as follows:

a) Common Business Understanding

Researchers such as Wigand et al. (1997) and Fuehrer & Ashkanasy (2001) note that an important element in any business cooperation is the establishment of common business understanding. An earlier work suggests that there are three specifications necessary for the establishment of a common business understanding in the virtual context. The first is a clear product specification: the design, quality, and functionality of the product or service. The second is a specification of the level of cooperation, which requires agreement about deadlines, liability, prices, profit allocation, and staff and resource input. The third is a formal specification of agreements between the virtual partners. In a virtual organization, these specifications need to be communicated clearly between the partners in order to achieve a common business understanding. There is always varying uncertainty between members, however. Therefore, the need to guard against opportunistic behavior varies between the partners (Wicks et al., 1999). This depends on the risk that the member is prepared to sustain as a potential loss, and also upon the partner's fear of opportunistic exploitation and the uncertainty of their behavior.

The three specifications (production, cooperation, and agreements between partners) can be achieved by negotiating relational contracts that guide the formation, operation, and dissolution of the virtual organization, thereby facilitating an increase in the level of collaboration-enabling trust. VEs, like other organizations, create fiscal and legal issues that must be clarified, but they lack a formalized legal framework (Fuehrer & Ashkanasy, 1998). Therefore, it is incumbent on the organization's members to develop their own guidelines for the operation of the enterprise. Such agreements may include clarification of members' tasks and responsibilities, agreement on contracts, allocation of funds, potential liability, and how members will contribute their expertise. In this sense, clear guidelines, spelled out in an early stage of the partnership, serve to reduce misperceptions and to foster the establishment of trust.

Other mechanisms to establish a common business understanding in VEs include development of an organization handbook, design of a mutual Internet site, chat room technology, and the use of team addresses for e-mail. A specific example is Livelink, a software selected by Siemens to enable the creation of a common business understanding through a standard computer interface.

The concept of common business understanding, therefore, shares similarities with organizational identity, which may be described as a set of distinctive and enduring traits that members associate with their organization. Scott and Lane (2000) have proposed further that identity is determined in part by the nature of stakeholder networks. Common business understanding, however, is more akin to Barney's broader concept of identity: "the theory organizational members have about who they are" (Barney et al., 1998, p. 103). In this respect, the author agreed with Gioia, Schultz, and Corley (2000) that organizational identity is not necessarily a stable phenomenon, but mutates to suit the prevailing environment. In the virtual context, therefore, a common business understanding may be defined as a transient understanding between network partners as to what they stand for, the nature of the business transactions that they engage in, and the outcomes that they expect—their "vision."

Scott and Lane (2000) emphasize that a common business understanding requires the creation of a shared vision, together with communication of mutual aims through a clear definition of the roles and expectations within the team, especially in the early stages of the partnership. In this respect, the process is typically initiated by agreement on a symbolic logo or design for a product or service, since understanding each member's role, together with group identification, determines critical behaviors such as willingness to cooperate with others, and willingness to engage in mutual goal setting. The VE partners thus rapidly need to establish group identity and an awareness of mutual needs and expectations, along with the clarification of tasks and responsibilities. In traditional partnerships, awareness and identity are in part shaped by the legal framework that regulates organizational relationships, as well as by networks, artifacts, and the organization chart (Scott & Lane, 2000). In the case of the VEs, however, mechanisms outside of the domain of traditional organizations need to be put in place to establish a common business understanding, which constitutes an important precursor of trust formation (Jarvenpaa & Leidner, 1999).

These examples illustrate how the creation of a sense of shared meaning, member identification, and mission identity, especially in an early stage of the partnership, facilitates collaboration at an individual level and the operation and productivity of the VE as a whole. As such, a common business understanding provides an essential condition for the development of trust within the organization. In effect, a common business understanding provides the virtual organization's members with an opportunity to share their perceptions of the organization's defined features, and creates a feeling of ownership and trust.

b) High Ethical Standards

Three factors uniquely characterize the virtual organization's position in regard to business ethics. First, VEs are rarely guided by pre-existing codified laws in which values and standards are written into legal systems enforceable in court. Since the organization's partners are not usually legally bound to the organization, any negative outcomes or perceptions attributed to poor business ethics could result in the organization's reputation suffering (Fichman, 1997). Second, because VEs like those created in our opening example are intrinsically temporary (Fuehrer & Ashkanasy, 1998), corporate ethics are difficult to develop because members will typically finish one virtual collaboration and enter into another. Third, VEs are intrinsically boundary spanning in nature, so that they must incorporate a diversity of culturally based values and morals.

Researchers focused on the notion of advances in ICT and the related effects on social behavior agree that unethical behavior in the virtual context is predominantly caused by technological changes and by the "inside keepers of the information systems" (Pearson et al., 1997, p. 94). They also agree that social behavior needs more than "new laws and modified edicts" (Johnson, 1997, p. 60) and that ethical issues will become increasingly important to enable business transactions to be carried out safely and securely. Although technology has been largely secured by advancing software and technology for virus detection, as well as encryption or decryption of information to ensure the security of business processes, Johnson (1997) notes that technology can never be sufficient to control all aspects of social behavior. Consequently, online behavior is predicated on an awareness and acceptance of ethical norms and behaviors. This can best be achieved through specification and clarification of the members' tasks, responsibilities, and agreed sanctions for proscribed behavior.

Johnson posits further that the "only hope to control online behavior is for individuals to internalize norms of behavior" (1997, p. 60) and suggests three rules for online ethics: (1) know and follow the rules of the forums; (2) respect the privacy and property rights of others (if there is any doubt, assume the user's desire for privacy and ownership); and (3) respect interacting partners by not deceiving, defaming, or harassing them. Not surprisingly, these rules for online behavior are essentially identical to rules for off-line behavior. Indeed, there is no reason why the same ethical guidelines that apply to regular behavior should not be employed in online behavior.

Pearson et al. (1997), reported on ethical standards for the IS profession that were proposed by three major professional associations in this field. These associations share an agreed set of behavioral obligations to society, colleagues, and professional organizations. The standards aim to promote the principle that individuals within the professions act in an ethical and responsible manner in order to influence the success of their organizations (Pearson et al., 1997). Clearly, similar standards can be developed for the operation of individual VEs specifying, for instance, the obligation to virtual organization members and clients.

Other possible mechanisms to promote ethical behavior in VEs include formal codes of ethics that comprise statements of prescribed and proscribed values or behaviors, and thus provide a strategic tool within organizations to inculcate and demonstrate ethical standards. Ethical standards also fulfill a strategic external role through recognition by government agencies and insurance companies. In the case of VEs, informal rules known as "netiquette" are usually in place, but a lack of a formal legal infrastructure means that a code of ethics is simultaneously both imperative and difficult to achieve. This is further compounded by different ethical standards and regulations between countries. Nevertheless, trust in interorganizational VEs clearly cannot be established until all members recognize that ethical standards are in place and are made aware of what the standards are.

c) Mutual Forbearance Between VE Partners

Some researchers approach the issue of trust by defining cooperation as coordination effected through mutual forbearance. Forbearance is refraining from cheating. Cheating may take a weak form (failing to perform a beneficial act for the other party) or a strong form (committing a damaging act). The incentives for forbearance arise from the possibility of reciprocity, leading to mutual forbearance. Parties that are observed to forbear may gain a reputation for this behavior, which makes them potentially attractive partners for others. The parties to a successful agreement may develop a commitment to mutual forbearance, which cements the partnership, and, in this way, mutual trust is created, which alters the preferences of the parties towards a mutually cooperative mode. Thus, short-term, self-interested behavior converts to cooperative trusting behavior.

d) Ability of VE Partners

In a VE, participants will be more willing to share knowledge when they trust others' abilities. It is only natural that participants want to converse with others who have the knowledge and skills regarding the topic at hand, since VEs almost always center around a common theme.

e) Effective Communication and Interaction between VE Partners

By communicating with people, we calibrate them, we get a better sense of them, and we understand their priorities. Members of VE therefore can increase the trust they are giving and the amount they will trust others by actively seeking opportunities to communicate with other members.

OPERATIONAL BENEFITS

Assuming that all of the above-mentioned conditions are in place, the likely impact on the functioning and management of VEs would include the following:

a) Favorable Trading Environment

In VEs, the risk that the Internet creates through identity and product uncertainty, physical separation, and the newness of the medium, has been attributed as an important barrier to online transactions (Bakos, 1998). Thus, perceived risk reduction is an important element that trust should influence. In addition, the success of online B2B marketplaces depends on transaction volume; hence, continuity is another crucial variable. Finally, satisfaction is an overall construct that describes the quality of the exchange relationship (Cannon & Perreault, 1999; Geyskens et al., 1998). In short, satisfaction, perceived risk, and continuity are posited as important success elements in online B2B marketplaces that could be influenced by interorganizational trust.

A recent research examines how institution-based trust develops in online B2B marketplaces to facilitate buyers' trust in sellers (Pavlou, 2002). This study proposes how specific institution-based structures help engender interorganizational trust and indirectly influence transaction success in B2B marketplaces. To assess the influence of the two dimensions of interorganizational trust—credibility and benevolence—in buyer-seller relationships, three trust outcomes are examined: perceived risk, continuity, and satisfaction.

Risk is a fundamental element of an organization's strategy because it has implications for performance. Since risk is difficult to capture as an objective reality, research has addressed the notion of perceived risk, which may be defined as the subjective probability of suffering a loss in pursuit of a desired outcome. Perceived risk has been negatively associated with transaction intentions (Jarvenpaa, 1999), interfirm partnerships, and joint ventures (Leverick & Cooper, 1998). Buyers' perceived risk in VEs mostly stems from the sellers' behaviors and not from the marketplace context. Most buyer-supplier transactions are characterized by information asymmetry, since the seller usually possesses more information than the buyer. Information asymmetry may give scope to opportunistic behavior and uncertainty. Although risk is inevitable in every transaction, trust reduces the expectations of opportunistic behavior and reduces perceptions of risk. Trust has also been shown to reduce the perceived risk of being taken advantage of by the other party. Following the trust literature, it is expected that both dimensions of trust should reduce perceived risk.

Continuity is defined as the perception of a buyer's expectation of future transactions with sellers in a B2B marketplace. There is significant evidence in the literature to suggest a strong association between trust and continuity; hence, it could be positively correlated with both dimensions of trust. Credibility allows buyers to concentrate on future transactions by reducing fears of short-term opportunism. Similarly, benevolence helps buyers to establish a long-term horizon, since they are confident that marketplace sellers will seek future transactions with mutual gain in the future.

The importance of satisfaction on business success has long been recognized. Cannon and Perreault (1999) posit that satisfaction with the relationship represents an important outcome of any business exchange. Satisfaction is the outcome of a trust-based relationship, since trust enhances satisfaction by reducing conflict (Geyskens et al., 1998). When buyers have faith in the credibility and benevolence of marketplace sellers, they are likely to be satisfied with their transactions, since conflict and fears of opportunism are diminished.

These findings provide evidence on how specific institutional mechanisms build trust in online B2B marketplaces, which in turn help to build a trustworthy trading environment for VE.

b) Reduction of Transactional Costs

The distinctive characteristic of a VE—shared ownership—is also its key problem. Shared ownership or control implies two or more companies deciding the strategic direction and operational issues of the joint subsidiary. Shared control brings with it increased transaction costs.

Transaction costs have both ex ante and ex post elements. Ex ante transaction costs include drafting, negotiating, and safeguarding an agreement. These costs can be quantified by examining actual costs (mainly management time) of the agreement process and examining the insurance costs of default. Ex post transaction costs include (1) the adaptation costs incurred when transactions drift out of alignment with the terms of the agreement; (2) the haggling costs incurred if bilateral efforts are made to correct ex post misalignments; (3) the setup costs and running costs associated with the governance structures (either in the courts or outside them) to which any disputes are referred; and (4) the bonding costs of securing partner commitments. Ex post costs are not easily quantifiable prior to the implementation of the agreement, because they include an element of uncertainty (How likely is it that the partner will default on the agreement?). This implicit probability exists in the minds of the executives entering into the agreement. It is difficult for the executives to articulate the complexity of these issues, and hence, "The transaction costs that are really there, in the sense that they determine the outcome are those transaction costs that are perceived by the managers who make the decisions" (Buckley & Chapman, 1997).

The impact of trust can now be seen in clear focus. Trust is a transaction cost-reducing mechanism that lowers the subjective risk of entering into an agreement. Both ex ante and ex post elements are reduced by trust. Costs of negotiating are reduced (perhaps legal trappings are replaced by "a gentleman's agreement"), and the subjective probability of ex post transactions failure declines when the parties trust one another.

c) Continuous Innovation in VE Products and Services

Rapid technological improvement means that continuous innovation in VE products and services is required to prevent users from switching to another system. To meet this challenge, VEs must possess state-of-the-art technological expertise. Trust is particularly important in a networked organization that requires constant and close attention to shared commitments, to safety and reliability, and to a shared willingness to learn and adapt. It has been suggested that trust permits a virtual organization to focus on its mission, unfettered by doubts about other members' roles, responsibilities, and resources, and that with trust, synergistic efforts in interorganizational missions are possible.

Trust plays an important synthesis role, as well. VE, with its flexible organizational structures, can leverage the ability and willingness to learn, thereby enhancing performance and attention to reliability over time. VEs with high levels of trust among their members can effectively utilize interactions and communication processes at their interfaces so members can learn together and develop shared mental models of reliability and a shared culture of safety.

d) Better Relationship Management

The strategic management literature often cites interorganizational systems (IOS)—automated information systems shared by two or more organizations—as examples of information technology used for competitive advantage. However, there is increasing awareness that IOS are becoming a strategic necessity and that their organizational impact demands attention. Examples show that IOS offer opportunities for enhanced organizational design. At the same time, however, IOS lead to mutual dependence between participating organizations, which potentially limits their adaptability. This makes management of relationships between IOS participants a key issue in establishing and managing successful IOS. To gain a better understanding of this issue, it is important to go beyond existing functional typologies of IOS. A framework of different contexts of IOS usage presented by Meier (1995) helps to assess the importance of relationship management between different groups of IOS participants. The framework classifies IOS contexts along two dimensions. First, does the system give the user a competitive advantage, or is it a strategic necessity for participating in the industry? Second, is the system an adjunct to the IOS provider's primary product, or does it constitute a stand-alone business? The author concludes that trust between IOS participants is essential for successful IOS. Analyzing IOS from the perspective of relationship management yields potential win-win outcomes amid all the rhetoric about the competitive battlegrounds.

e) Better Information Exchange

In the VE environment, one would expect that increased trust would result in increased information sharing and acceptance between members. It is important to note that information sharing occurs in both directions—giving information and getting information. These two are distinctly different behaviors, as getting information is the act of querying the members of the VE or searching the VE's conversation for items of interest, while giving information involves answering others' queries or volunteering comments or observations to the community. Additionally, giving information generally involves exposing oneself to a greater degree than just inquiring.

Trust in the benevolence and integrity of other members would increase the desire to get information, because the value of such information depends on the honesty of the person providing it and the person's willingness to help. It is hypothesized that this trust would make an individual more inclined to ask for information in the VE because the individual would know the provision of information is likely to be governed by these principles. In the same manner, members would only desire to give information when they trust in others' benevolence and integrity; otherwise, they would expose themselves to possible opportunistic behavior. Belief in the benevolence and integrity of others will also influence information exchange due to the reciprocal nature of communication that it implies. The VE will not survive if reciprocity does not exist; all must contribute reciprocal rewards and have a desire to do good to others. Individuals will be less inclined to share knowledge in the VE if they feel this adherence to benevolence norms is lacking. There are many indications in the literature that highlight the centrality of the reciprocal nature of communication in virtual enterprises.

f) Ownership Culture

An ownership culture exists in a VE where member organizations think and act like the core company. Building an ownership culture requires that member companies change the way they behave and move from conventional business patterns to challenging new roles. Members of the VE must trust one another, and, in particular, frontline partners must trust the core company before they will risk change. Trust is a prerequisite to building an ownership culture in a VE. It is commonly assumed that ownership primarily is an incentive that aligns financial interests of member organizations with the interests of the core company. To a great extent, this is true; this incentive plays an essential role in motivating partner companies because it gives each partner an individual profit motive to promote the success of the entire VE. The financial incentive is necessary, but the real power of ownership results from the trust effect. The trust effect is a deep connection to the VE, a relationship based on more than money. Ownership can give partners a reason to belong to the VE. The trust effect is the result of psychological ownership and only exists in VEs that actively nurture a sense of ownership throughout organizations in the network. Studies show that ownership and significant participation make it possible for a VE typically to have an advantage unavailable to their competitors.

FUTURE TRENDS

Current research on the impact of trust in VE has limitations. Most work was based on a limited number of case studies. These case studies cannot be considered representative of all VE because of their industrial and cultural biases. Because most of these models are longitudinal, for generalization purposes it is necessary to test them against the behavior of VE over time—a difficult, costly, and time-consuming exercise. A possible next step is to confront the above models with a richer, more widely dispersed set of cases, with more cultural and structural variety in the VE analysed in order to investigate its degree of robustness.

While some studies (Pavlou, 2002) posit positive relationships between trust and its consequences, these relationships are non-linear. Following Castelfranchi and Falcone (2001), buyers would only engage in a transaction if their level of trust exceeded their organization’s threshold, which depends on the importance of the transaction, the level of risk, and their individual propensity. Given a minimum threshold for trust to become effective, it is important to recognize this non-linearity.

Currently, a popular experimental paradigm employed by Human-Computer Interaction(HCI) researchers to assess trust between people interacting via computer-mediated communication, covers social dilemma games based on the Prisoner’s Dilemma(PD). HCI researchers employing this experimental paradigm currently interpret the rate of cooperation—measured in the form of collective payoff—as the level of trust the technology allows its users to develop. Some researchers argue that this interpretation is problematic, since the game’s synchronous nature models only very specific trust situations (Reigelsberger et al, 2003). Furthermore, experiments that are based on PD games cannot model the complexity of how trust is formed in the real world, since they neglect factors such as ability and benevolence.

Additionally, it is noted from the literature that little theoretical explanation exists in order to understand the impact of trust in various forms of VE relationships. It has been found that firms in horizontal alliances would display a lower level of organizational trust and a weaker relationship between interfirm cooperation compared to firms in vertical integration of alliances, and that trust is unrelated to cooperation in horizontal alliances (Rindfleisch & Moorman, 2001). It was suggested that this different impact of trust could be due to higher opportunism, lower interdependency, and stronger institutional linkages among horizontal collaborators compared to their vertical counterparts. If this finding is substantiated by future empirical research in the VE domain, researchers may need to reconsider the popular notion that trust is an essential component of all types of relationship exchanges in the VE.

CONCLUSION

Cooperative teamwork offers a greater chance for VE to be successful in today’s’ agile competition. Trust models based on traditional familiarity would not meet the special needs of the VE, which is having a short life cycle and involves a large number of partners who have never met before. Consequently, the VE partners must realize the need to effectuate this paradigm shift. This chapter identified five essential conditions needed for effective trust development in a virtual environment: a common business understanding, high ethical standards, mutual forbearance between partners, ability of partners, effective communication, and interaction within the VE. It is anticipated that the likely impact of trust would include a favorable trading environment, reduced transactional costs, continuous innovation in VE products and services, better partnership management, better information exchange among VE partners, and the development of an ownership culture.

REFERENCES

Bakos, J.Y. (1998). The emerging role of electronic marketplaces on the Internet. *Communications of the ACM, 41*(8), 35–42.

Barney, J.B. et al. (1998). A stratgey conversation on the topic of organizational identity. In D.A. Whetton & P.C. Godfrey (Eds.), Identity in organizations: Building theory through conversations (pp. 99-168). Thousand Oaks, CA: Sage.

Buckley, P., & Chapman, M.K. (1997). The perception and measurement of transaction costs. *Cambridge J Econ, 21*(2), 127–145.

Cannon, J.P., & Perreault, W.D. (1999). Buyer-seller relationships in business markets. *Journal of Marketing Research, 36*, 439-460.

Castelfranchi, C., & Falcone, R. (2001). Social trust: A cognitive approach. In C. Castelfranchi & Yao-Hua Tan (Eds.), *Trust and deception in virtual societies* (pp. 55-90). Kluwer Academic Publishers.

Fichman, M. (1997). A multilevel analysis of trust in inter-organizational customer-supplier ties [Report]. *Annual Academy of Management Meeting*, Boston, MA.

Fuehrer, E.C., & Ashkanasy, N.M. (1998). The virtual organization: Defining a Weberian ideal type from the inter-organizational perspective [Report]. *Annual Meeting of the Academy of Management*, San Diego, CA.

Fuehrer, E.C., & Ashkanasy, N.M. (2001). Communicating trustworthiness and building trust in interorganizational virtual organizations. *Jr. of Management, 27*(3), 235-254.

Geyskens et al., (1998). Generalizations about trust in marketing channel relationships using meta-analysis. *International Journal in Marketing, 15*, 223-248.

Gioia, D.A., Schultz, M., & Corley, K.G. (2000). Organizational identity, image, and adaptable instability. *Academy of Management Review, 25*, 63-81.

Jarvenpaa, S.L., & Leidner, D.E., (1999). Communication and trust in global virtual teams. *Organ Sci.*, *10*(6), 791-815.

Johnson, D. (1997). Ethics online. *Communications of the ACM, 40*,(1), 60-65.

Keen, P.G.W. (2000). Ensuring e-trust. *Computerworld*, *34*(11), 46.

Lefebvre, L.A., & Lefebvre, E. (2002). E-commerce and virtual enterprises: Issues and challenges for transition economies. *Technovation*, *22*, 313-323.

Leverick, F., & Cooper, R. (1998). Partnerships in the motor industry: Opportunities and risks for suppliers. *Long Range Planning*, *31*(1), 72-81.

McDonough, E., Kahn, K., & Barczak, G. (2001). An investigation of the use of global, virtual, and collocated new product development teams. *The Journal of Product Innovation Management*, *18*(2), 110-120.

McKnight, D.H., Cummings, L.L., & Chervany, N.L. (1998). Initial trust formation in new organizational relationships. *Academy of Management Review,* 23(3), 473-490.

Pavlou, P.A. (2002). Trustworthiness as a source of competitive advantage in online auction markets. *Proceedings of the Academy of Management Conference*, Denver, CO.

Pearson, J.M., Crosby, L., & Shim, J.P. (1997). Measuring the importance of ethical behavior criteria. *Communications of the ACM, 40*(9), 94-100.

Riegelsberger, J., Sasse, M. A., & McCarthy, J. (2003, April 20-25). Shiny happy people building trust? Photos on e-commerce Web sites and consumer trust. *Proceedings of CHI 2003*, Ft. Lauderdale, FL.

Rindfleisch, A., & Moorman, C. (2001). The acquisition and utilization of information in new product alliances: A strength-of-ties perspective. *Journal of Marketing, 65*, 1-18.

Sarker, S., Lau, F., & Sahay, S. (2001). Using an adapted grounded theory approach for inductive theory building about virtual team development. *Database for Advances in Information Systems, 32*(1), 38-56.

Scott, S.C., & Lane, V.R. (2000). A stakeholder approach to organizational identity. *Academy of Management Review***,** *25*, 43-62.

Stewart, D.W., Pavlou, P.A., & Ward, S. (2002). Media influences on marketing communications. In J.B.A.D. Zillmann (Ed.), *Media effects: Advances in theory and research*. Hillsdale, NJ: Erlbaum.

Wicks, A.C., Berman, S.L., & Jones, T.M. (1999). The structure of optimal trust: Moral and strategic implications. *Academy of Management Review 29*, 99-116.

Wigand, R., Picot, A., & Reichwald, R. (1997). *Information, organization and management: Expanding markets and corporate boundaries*. Chichester, UK: Wiley.

Wong, T. T., & Lau, H.C.W. (2002). The impact of trust in virtual enterprises. In A. Gunasekaran (Ed.), *Knowledge and information technology management in the 21st century organizations: Human and social perspectives*. Hershey, PA: Idea Group Publishing.

Zaheer, A., McEvily, B., & Perrone, V. (1998). Does trust matter? Exploring the effects of interorganizational and interpersonal trust on performance. *Organization Science, 9*(2), 141-159.

KEY TERMS

Agile: Being agile means to be proficient at change, which allows an organization to do anything it wants, whenever it wants. Since virtual enterprises do not own significant capital resources of their own, it helps if they are agile, so they can be formed and changed very rapidly.

Benevolence: An act intending or showing goodwill and kindness.

Credibility: The quality of being believable or trustworthy.

Ethical: Conforming to standards of professional or social behavior agreed to by all members of a virtual enterprise.

Information Asymmetry: Condition in which at least some relevant information is known to some, but not all, parties involved. Information asymmetry causes markets to become inefficient, since all market participants do not have access to the information they need for their decision-making processes.

Ownership: Ownership of a business by the people who work for it (e.g., partners of a virtual enterprise).

Risk: A possibility of incurring financial loss.

Virtual Enterprise: A temporary business organization set up between trading partners operating from geographically dispersed sites, for the duration of a common project. The design and manufacture of new products or services frequently require the talents of many specialists. When many corporations combine their specialties to create a product or service, the result can be called a virtual enterprise. A virtual enterprise must be able to form quickly in response to new opportunities and dissolve just as quickly when the need ceases.

Traditional Familiarity: Traditional familiarity combines an assumption of continuity with the past experience of a partner. Traditional trust, therefore, relies on the fact that VE partners who could be observed as trustworthy in the past will display the same kind of behavior in the future.

Trustworthiness: The trait of deserving trust and confidence.

Trust Placement Process in Metateam Projects

Walter D. Fernández
Australian National University, Australia

INTRODUCTION

Metateams are temporary confederations of dislocated teams from different firms working on a single Information Systems (IS) development project. These teams (or groups) are linked by interdependencies and commercial agreements and use information and telecommunication technologies as the main media for their communication activities. Within these socio-technical systems, teams are members of *virtual teams of teams*, where key teams belong to different firms, each performing well-defined functions in accordance with its contractual role, with the objective of executing a single overarching project.

Metateams can bring expertise from multiple firms to the project while information and communication technologies facilitate the collaboration of their teams. However, managing metateams presents unique difficulties, as achieving effective metateam collaborations is both critical and difficult. Metateams are particularly exposed to the lack of common understanding of prime objectives and deficient pre-project arrangements observed in traditional IS project teams (Jiang, Klein, & Means, 2000); identity issues of autostereotype (how groups perceive themselves) and heterostereotype (how groups perceive other groups), arising from encounters of groups exhibiting organizational or national cultural differences (Hofstede, 1997); difficulties in successfully applying "foreign" management techniques to culturally heterogeneous groups (Trompenaars & Hampden-Turner, 1998); and, goal incongruence as a product of organizational fragmentation resulting from deregulation, privatization, or outsourcing (Berggren, Soderlund, & Anderson, 2001).

This overview, based on a theory-building empirical study, suggests that the effectiveness of the trust placement process—and not just exhibiting specific levels of trust— significantly impacts on project success. We focus on trusting behaviors, what we *do* when we trust or distrust others, and how our actions impact on the *quality* and *cost* of the metateam project.1

STUDY BACKGROUND: THE PROJECT AND ITS KEY PLAYERS

The SUN Project involved three key organizations – RedCorp, ITSP, and OSC – and dislocated teams from three countries (Figure 1). RedCorp was in charge of the total project. RedCorp and ITSP were linked by an information technology (IT) outsourcing agreement in which RedCorp was the client. OSC was working in partnership with ITSP on the SUN project.

SUN was a strategic multimillion-dollar IT development and implementation project. Due to its high priority, magnitude, and impact within the client company, SUN was highly visible at the top management level in all participating firms.

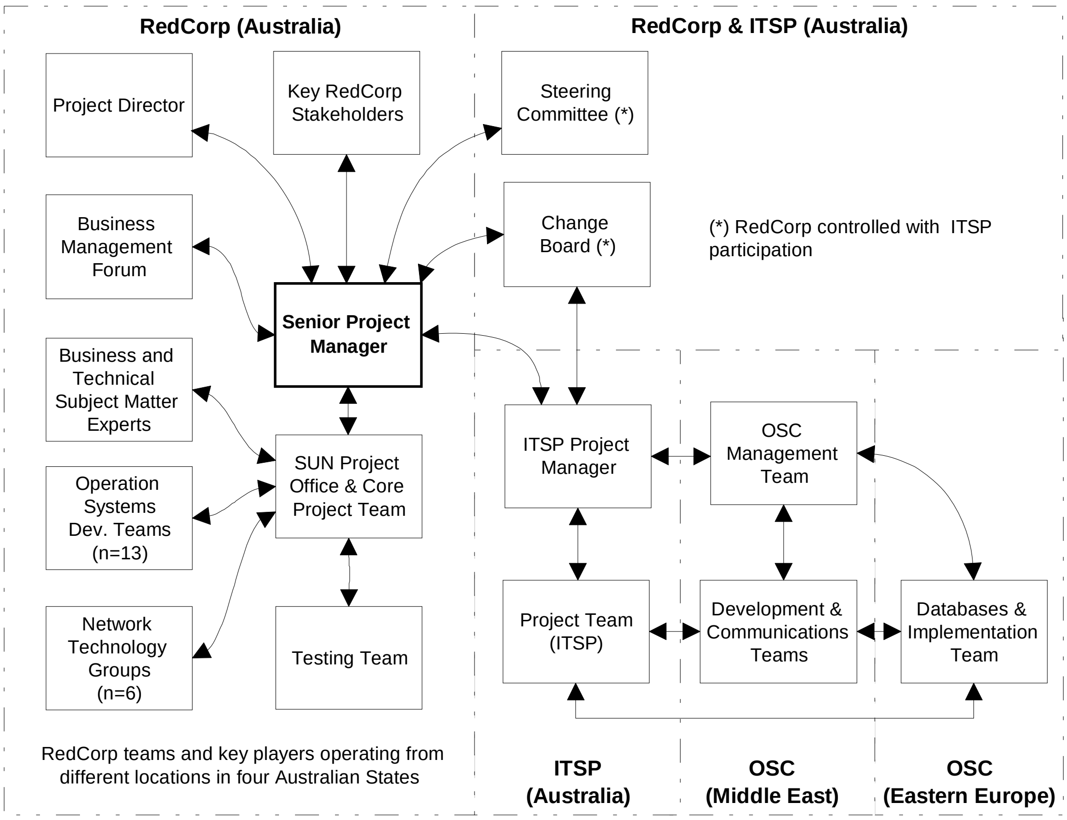
The case data included semi-structured interviews, observations, and access to project documents and electronic correspondence. The SUN Project provided rich documentary evidence; more than 4,000 e-mails and 800 project documents were available for the study. Grounded Theory Methodology (Glaser & Strauss, 1967 Glaser, 1978, 2001) was used to develop a conceptual account while the software package ATLAS.ti facilitated text analysis and management of research memos.

CONTEXTUAL ISSUES

Contextual issues are critical to understanding socio-technical organizations such as metateams; among others, the following aspects need to be briefly described:

- Metateams are organizationally fragmented systems where managers: (a) do not have the full range of options regarding project strategies and control mechanisms available to more traditional project organizations; (b) deal with teams from multiple firms having multiple mental models of what reality is or should be; and (c) may confront resentment as a consequence of anti-outsourcing sentiment.
- Metateams may be subject to unrealistic demands and expectations. Modern organizations, pushed by markets or regulators and competing in a fast-changing environment, may resort to metateams expecting to achieve the fastest project cycles and the lowest possible costs. In our study, the combination of high demands for rapid delivery and poor understanding of project complexity resulted in inadequate inter-firm agreements, which enforced rigidity when flexibility was required.
- Communication effectiveness and efficiency face multiple obstacles (e.g., miscommunication due to

Figure 1. SUN Project key players, their firms, locations, and formal communication channels.



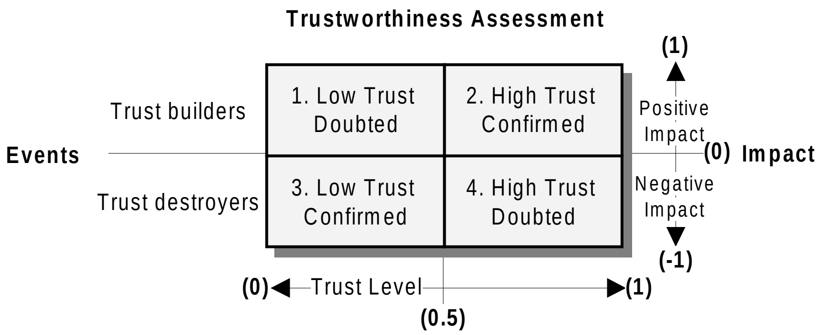
language barriers, message distortion; dogmatic approach; remoteness, inappropriate levels of trust). Additionally, it was observed that the quality of the inter-team communication was impeded, not only because of the "virtual" nature of their communication (as it is in the case of virtual teams), but also during face-to-face events due to diverging interests, urgencies, and priorities.

- Achieving a cohesive metateam is restricted by: multiple senses of identity; the magnification of the negative aspects of eliteness; the multiplicity of meaning regarding the "team product" and ownership of products and sub-products; and the often stressful nature of forced interrelations under low levels of trust.
- Goal incongruence can create different worldviews affecting teams and actors. Incongruent needs, wants, and perceptions can potentially lead to sub-goal strategies that are detrimental to successfully accomplishing the end goal. Goal incongruence is a default state of metateams, a consequence of imperfect contracting and imperfect cognition.
- While temporality is a condition of all project work, this condition is exacerbated in metateam. In traditional project teams, some level of expectation exists regarding the probability of team continuity; this probability is very low in metateams, reducing the perceived value of getting along with each and therefore adding more weight to the task-oriented, less personalized relationship.
- Although the higher level of experience required of team members is perceived as a critical asset, it also brings issues such as: allegiance to referential constituency versus project team allegiance; issues of replacement as top-level experts are more difficult to replace; and, skepticism inhibiting the beneficial effects of swift trust in virtual team formation.

TRUSTING OTHERS IN CONDITIONS OF UNCERTAINTY

Trust has been viewed as a lubricant to cooperation (Arrow, 1974; Misztal, 1996), as a product of cooperation (Axelrod, 1984) and as a way of dealing with imperfect knowledge and uncertainty about others' behavior (Gambetta, 2000). Therefore, we can expect trust to be relevant to those virtual team and metateam contexts exhibiting a high need for cooperation amid conditions of

Figure 2. Effect of project events on perception of trustworthiness (source: SUN Project-Data analysis)



uncertainty. Trust facilitates open communication, cooperative efforts, reduction of uncertainty, resolution of conflicts, common understanding, and control of transaction costs.

In our perspective, trust and mistrust are polar opposites in a conceptual continuum (a view supported by Gambetta, 2000; Giddens, 1990; and Rotter, 1967, among others). Therefore, trust is a threshold value between zero (complete mistrust) and one (complete trust). This threshold is a *subjective probability*, with which, according to Gambetta (2000), an agent assesses that another agent or group of agents will perform a particular action, both *before* he can monitor such action (or independently of his capacity ever to be able to monitor it) *and* in a context in which it affects *his own* action. p. 217

In trusting others, we take a risk because we expect certain outcomes with neither complete control of others' actions nor complete knowledge of their intentions, and because we are subject to the unforeseeable events of social reality (Lorenz, 2000; Misztal, 1996). Therefore, *trusting behavior* increases our vulnerability to another party and could make us regret our actions if that party abuses our trust (Lorenz, 2000).

In a metateam project, trust per se is neither good nor bad. Trusting behaviors, *when well placed*, have a lowering (positive) effect on the cost of inter-team transactions. Yet, when trust is misplaced, the effect on the project cost is detrimental due to the cost of opportunism. Thus, in metateams, trust is an economic asset that, when well placed, reduces the need for and, thus, the cost of controls. The importance of trust as an asset is given by the level of uncertainty in the metateam environment; where neither complete trust nor complete controls are possible due to issues such as organizational fragmentation, imperfect commercial arrangements, and diverging profit motives. In the following sections, we discuss how familiarity-building processes and the role of familiarity, exchange and reciprocity contribute to improve the quality of our trust placement, that is, economic decisions based on perceptions of others' trustworthiness.

Building Familiarity

Building familiarity is an interaction-based process that improves the quality of trust placement. By building familiarity, team members develop a grounded appreciation of others' trustworthiness and can reduce attribution errors (Cramton, 2002). Observable events such as project meetings, responses to deadlines, acceptance of deliverables, detected contract breaches, or joint conflict management and resolution efforts all provide opportunities to build familiarity and form the basis for *ex post* assessment of the trust placement's correctness. When significant, these events can modify the trust threshold, resulting in a new psychological state that is used to predict future events; using the lessons learned to look ahead, perceive possible hazards, and take *ex ante* inter-team hazard-mitigating actions (Williamson, 1985, 1996). Therefore, these events can either build or destroy trust. Figure 2 illustrates the effect of project events on trustworthiness perception.

In this evolving psychological process, project managers continuously conduct post-event evaluations. Trustworthiness evaluations involve an element of equivocality and non-rationality because they depend on availability of information and the belief that predictability is somehow possible.

Achieving greater familiarity in the SUN Project facilitated the actors' perception of hazards and the development of hazard-mitigating strategies. These strategies impact on transaction costs (immediate and/or future) and have potential long-term effects on the relationship between the client and the vendor. Yet, building familiarity in metateams can be difficult, as the next section explains.

Familiarity, Exchange, and Reciprocity in Metateams

The notion of building familiarity as a process based on metateam interactions aligns with Simmel's view of society as a number of individuals connected by exchange-based interaction, and that exchange and reciprocity are basic constitutive forms of relationships (Misztal, 1996). Exchange in this context is defined as "sacrifice in return for a gain," and it is "one of the functions that creates an inner bond between people – a society, in place of a mere collection of individuals" (Simmel, Bottomore, & Frisby, 1978).

A positive exchange occurs when the risks actors take in trusting their partners (sacrifice) return the benefits of effective and open collaboration (gain). Conversely, the outcome of the exchange will be negative when the expected gain is not realised (or is perceived as such by the actors). Unrealistic expectations, poor understanding of structural and situational drivers, opportunism, and lack of required skills or resources are some of the factors contributing to unrealised exchanges.

Because the metateam relationship is primarily task-based, the strength of the bond between teams depends on how these teams perceive the outcomes of their task-based exchanges. To expect an exchange, people in a particular team need to feel that other teams are (a) capable of doing the tasks they are contracted to do, and (b) able to understand why these tasks are important in a wider context (not just for their immediate goals), and to act according to that understanding. Therefore, expectations of exchange involve trust, as without trust an exchange cannot be rationally expected.

In the SUN Project, expectations of trustworthiness primarily resulted from evaluating other teams' contractual trust, competence trust, and goodwill trust, confirming Sako's (1998) assertions regarding the nature of trust in virtual teams. In other words, actors were concerned with questions such as: Are *they* going to respect the contract? Do *they* have the competence to deliver as promised? and Are *they* prepared to go beyond the letter of the contract to deliver on their promises?

THE HIGH PERFORMING METATEAM

We propose that effective and efficient metateam systems will (a) require appropriate controls, (b) allow flexibility, (c) foster cooperative conflict management, and (d) moderate extreme values of the trust-distrust continuum. To help in fulfilling these requirements, we need to improve the effectiveness of trust placement processes by, first, increasing awareness of the metateam's challenges among the team members and their organizations and, second, addressing key environmental issues.

Furthermore, we suggest that effective metateam controls need to be entrenched in *appropriate:*

- pre-project agreements and project agreements, and
- trust, congruency, and conflict-related processes.

By "appropriate," we mean agreements and controls *contributing* to the implementation of processes that are critical to the effective management of trust, congruency and conflict. Two key concepts—cognitive conflict embracement and trust processes and awareness—are briefly discussed next.

Cognitive Conflict Embracement

To embrace cognitive conflicts (rather than affective conflicts), we have to see them as opportunities to deliver what our customers want, to build relationship and to develop mutual knowledge. The continuous process of discovery and resolution of conflicts allows the project to move forward by building a shared mental model of the project that includes a more realistic view of its opportunities and risks.

Trust Processes and Awareness

Trusting other teams to conduct their work as expected considerably reduces the *perception* of need for interteam communication and controls. When high trust is justified, the project benefits. However, high trust is a double-edged sword that could save time and money but also result in costly disappointment when high-trust is placed on opportunistic partners, something that can only be known *ex post*. In metateams, project managers must balance the need to capitalize on the cost-saving properties of trust and the risk of trusting others in conditions of uncertainty. As trust shapes perception of relationship risks in metateams, the *trust placement process* has an important role in implementing effective hazard-mitigating strategies. To enhance the effectiveness of the trust placement process, organizations must:

1. Develop awareness of the trust placement process, of its nature and consequences. This awareness should help to assess trustworthiness and to generate conditions for trusting relationships.
2. Implement strategies to help metateam project managers to improve *ex ante* trust-based decisions by structuring their project in a manner conducive to building familiarity as early in the project as possible, i.e., using mini-milestone techniques allowing for early and frequent exchanges.

CONCLUSION

Trust plays an important role in the management of metateams. In deciding how much others can be trusted, actors make choices affecting the project's transaction cost, and thus they take risks. In this article, we presented contextual issues of metateams that must be considered when assessing project risks and strategies.

We perceive trust and distrust as opposite ends of the trust continuum. Trust has the power to shape perceptions of reality and therefore to influence the hazard-mitigating strategies actors adopt to achieve their objectives. We argued that trust is relevant to project success and that it affects the project's cost. Without trust, the interteam collaboration becomes difficult and carries a higher transaction cost due to (a) the need to implement additional controlling processes, and (b) less efficient conflict management processes requiring frequent escalation and causing delays. However, we also warned of the dangers of total trust in conditions of uncertainty and organizational fragmentation, and proposed that a balanced system of metateams will necessarily have to alternate between tendencies toward either end of the trust continuum.

Therefore, on the one hand, we believe that trust facilitates the relationship while, conversely, we are cautious in deriving a positive correlation between trust based behaviors and the success of the collaborative endeavor because *misplaced* trust is indeed detrimental to the success of the metateam collaboration. While trust is often seen as a lubricant for cooperation (Arrow, 1974), it can also be a cause of failure in major collaborative IT projects enacted by multiple teams. As our study shows, metateams exhibit both the need for and the risks of trust. Acknowledging the key role this ambivalence plays is critical to our understanding of metateams. Consequently, the key issue discussed in this article is the risk involved in placing trust; that is, trusting others and then discovering that the trust was misplaced or, in opposition, not trusting when we should and therefore introducing costly friction. As the effectiveness of the trust placement process increases the chances for project success, we argue that more effort (in research and practice) is required to improve this process.

We took an economist's view of trust, where the efficient building of a relationship through familiarity, exchange, and reciprocity serves the utilitarian purposes of cost minimization, effectiveness, and efficiency in major IT projects. We maintain that the economic value of creating trusting relationships outstrips the cost of doing so; consequently, we propose several areas to be considered when assembling and conducting metateam projects, including appropriate project agreements that align with critical needs for congruence building, conflict embracement, and effective trust placement.

Finally, metateams and virtual teams represent a revolution in the way we do IT projects and organize work. As the global business environment continues to demand innovation, flexibility, and responsiveness, the number of major IT projects using metateams will increase. The same demands for flexibility and responsiveness will further highlight the criticality of achieving project success; therefore, client and vendor organizations able to evolve and achieve effective metateam systems will thrive at the expense of those more inflexible and slow to adapt.

REFERENCES

Arrow, K. J. (1974). *The limits of organization* (1st ed.). New York: Norton.

Axelrod, R. M. (1984). *The evolution of cooperation*. New York: Basic Books.

Berggren, C., Soderlund, J., & Anderson, C. (2001). Clients, contractors, and consultants: The consequences of organizational fragmentation in contemporary project environments. *Project Management Journal, 32*(3), 39-48.

Cramton, C. D. (2002). Attribution in distributed work groups. In P. Hinds & S. Kiesler (Eds.), *Distributed Work: New ways of working across distance and technology* (pp. 191-212). Cambridge, MA: MIT Press.

Fernández, W. D. (2003). Trust and the trust placement process in metateam projects. In D. Pauleen (Ed.), *Virtual teams: Projects, protocols and processes* (pp. 40-69). Hershey, PA: Idea Group Publishing.

Gambetta, D. (2000). Can we trust trust? In D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations (electronic edition,* pp. 213-237). Oxford, UK: Oxford University Press.

Giddens, A. (1990). *The consequences of modernity*. Cambridge, UK: Polity Press in association with Basil Blackwell, Oxford UK.

Glaser, B. G. (1978). *Theoretical sensitivity: Advances in the methodology of grounded theory*. Mill Valley, CA: Sociology Press.

Glaser, B. G. (2001). *The grounded theory perspective: Conceptualization cContrasted with description*. Mill Valley, CA: Sociology Press.

Glaser, B. G. & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New York: Aldine Publishing Company.

Hofstede, G. (1997). *Cultures and organizations: Software of the mind* (Rev.ed.). New York: McGraw-Hill.

Jehn, K. A. (1995). A multimethod examination of the benefits and detriments of intragroup conflict. *Administrative Science Quarterly, 40*(2), 256-282.

Jiang, J. J., Klein, G., & Means, T. L. (2000). Project risk impact on software development team performance. *Project Management Journal, 31*(4), 19-26.

Lorenz, E. H. (2000). Neither friends nor strangers: Informal networks of subcontracting in French industry. In D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations (electronic edition*, pp. 194-210). Oxford, UK: Oxford University Press.

Misztal, B. A. (1996). *Trust in modern societies: The search for the bases of social order*. Cambridge, MA: Polity Press.

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality, 35,* 651-665.

Sako, M. (1998). Does trust improve business performance? In C. Lane & R. Bachmann (Eds.), *Trust within and between organizations: Conceptual issues and empirical applications* (pp. 88-117). New York: Oxford University Press.

Simmel, G., Bottomore, T. B., & Frisby, D. (1978). *The philosophy of money*. London: Routledge and Kegan Paul.

Townsend, A. M., DeMarie, S. M., & Hendrickson, A. R. (1998). Virtual teams: Technology and the workplace of the future. *Academy of Management Executive, 12*(3), 17-29.

Trompenaars, A., & Hampden-Turner, C. (1998). *Riding the waves of culture: Understanding cultural diversity in global business (2nd. ed.)* New York: McGraw-Hill.

Williamson, O. E. (1985). *The economic institutions of capitalism: Firms, markets, relational contracting*. New York, London: Free Press; Collier Macmillan.

Williamson, O. E. (1996). Economic organization: The case for candor. *Academy of Management Executive, 21*(1), 48-57.

KEY TERMS

Affective or Relationship Conflict: "Incompatibilities among group members, which typically includes tension, animosity, and annoyance among members within a group" (Jehn, 1995, p.256).

Cognitive Conflict: Overt discrepancy among intervening parties about how a particular task or issue should be resolved and their subsequent engagement in action to redress the situation.

Major Projects: Multimillion-dollar IT development and implementation projects often defined as "*strategic*" due to their high priority and organisational impact.

Metateams: Temporary groups composed of two or more geographically and inter-organisationally dispersed teams, commercially linked by project-specific agreements and primarily enabled by electronic means of communication.

Virtual Teams: "Groups of geographically and/or organizationally dispersed co-workers that are assembled using a combination of telecommunication and information technologies to accomplish an organizational task" (Townsend, DeMarie, & Hendrickson, 1998, p.18).

Trust: A subjective *ex-ante* assessment of another party's ability and willingness to perform a particular action on which the assessor depends.

Trust Placement Process: The decision making process and the consequential actions resulting from trusting others under conditions of uncertainty and risk.

ENDNOTE

[1] This overview is derived from Fernández (2003).

U.S. Disabilities Legislation Affecting Electronic and Information Technology

Deborah Bursa
Georgia Institute of Technology, USA

Lorraine Justice
Georgia Institute of Technology, USA

Mimi Kessler
Georgia Institute of Technology, USA

INTRODUCTION

The Americans with Disabilities Act (ADA) is the cornerstone legislation to address the civil rights of people with disabilities, including making products, services, and physical environments accessible to them. Almost everyone in the U.S. is familiar with the ADA, but designers of technology products and services need to be aware of accessibility standards that go beyond the ADA: specifically, Section 508 of the Rehabilitation Act and Section 255 of the Communications Act. These laws define accessibility standards and guidelines that impact the design of electronic, information, and telecommunication technologies, and they are intended to promote products and services that are as accessible to persons with disabilities as those without (Section 508, 1998). Furthermore, with the aging of the U.S. and world populations (Forrester, 2004), the number of people who want to use technology but cannot, because of disabilities, is on the rise. Designers of technology need to understand how modifications in traditional design will make products more marketable and usable by a wider range of customers. This article reviews important aspects of Sections 508 and 255, assistive technology and accessible design, and additional sources of information and training.

BACKGROUND

Key Implications of the Legislation

Section 508 of the Rehabilitation Act states that "each Federal department or agency, including the United States Postal Service…when developing, procuring, maintaining, or using electronic and information technology (EIT)…shall ensure…that individuals with disabilities…have comparable access" (Section 508, 1998). Section 508 required the Architectural and Transportation Barriers Compliance Board (also known as the Access Board) to develop accessibility standards for EIT, and it stipulated that the law applied to procurements after June 25, 2001 (Federal Acquisition Regulations, 2001).

The technology addressed in the Section 508 technical standards cover: "software applications; operating systems; Web-based intranet and Internet information and applications; telecommunications products; video and multimedia products; self-contained, closed products; and desktop and portable computers" (Federal Acquisition Regulations, 2001). The standards also address information and documentation including "product support in alternative formats, descriptions of accessibility and compatibility features in alternative formats, and product support services in alternative communications modes" (Access Board, 2001). Federal departments and agencies can be exempt from compliance only if they can show that compliance is an "undue burden" as a result of "significant difficulty or expense." Prior to the Section 508 amendments, Section 501 (federal employment) and Section 504 (federally funded programs and activities) of the Rehabilitation Act addressed accommodation of individuals. The amendments to Section 508 addressed the technology itself, that is, making it accessible to everyone right "out of the box" (although this does not necessarily eliminate the need for individual accommodation). It was thought that by making accessibility requirements a part of the federal procurement process, there would be financial incentives for companies to design products that meet these standards.

The Section 508 standards apply specifically to the United States government when purchasing, developing, maintaining, or using electronic and information technology. In addition, an increasing number of states purchase electronic products and services that conform to the Section 508 standards (or similar state-developed regulations or standards). Due to the complexity of the regulations and range of requirements, creators and vendors of technology products need information and training to ensure their products and services adhere to the 508 standards.

The Telecommunications Act of 1996, which was the first major overhaul of American telecommunications policy in nearly 62 years, added *Section 255* to the Communications Act of 1934. Section 255 requires telecommunications manufacturers and providers of telecommunications services to make their products and services accessible to and usable by people with disabilities if "readily achievable." The Federal Communications Commission (FCC) makes readily achievable determinations on a case-by-case basis, but generally, companies with more resources need to do more to make their products and services accessible to people with disabilities. When it is not possible to provide direct access, Section 255 requires manufacturers and providers to make their devices and services compatible with peripheral devices and specialized customer premises equipment (CPE) that are commonly used by people with disabilities, if readily achievable. Examples of specialized CPE include teletypewriters (TTYs) and assistive listening devices.

Section 255 also requires companies that develop telecommunications products and services to include the following activities as business practices:

- When the company conducts market research, product design, testing, pilot demonstrations, and product trials, it should include individuals with disabilities in target populations of such activities.
- Companies should work cooperatively with disability-related organizations.
- Companies should undertake reasonable efforts to test access solutions with people with disabilities. (Federal Communications Commission, 2002)

Unlike Section 508, Section 255 is not restricted to just the federal marketplace; it applies to telecommunications products and services purchased by anyone in the U.S. Section 508 covers a wide variety of disabilities: people who are deaf or hard of hearing, who have mobility or dexterity impairments, who have speech impairments, and those who have low vision or who are blind. Section 255 includes all of these impairments as well as cognitive disabilities. While the standards, guidelines, and directives associated with Section 508 and Section 255 may appear to complicate the product design process, often these challenges bring about innovation and new product ideas. In most cases, these innovations lead to product features that are desired by a customer market that is much larger than the disability community.

Disabilities and Applications

People with disabilities need different product and service features so they can access information and communicate with others at a level that is equal to those without disabilities. Often an assistive technology is needed by someone with a disability to overcome access barriers. For example, a text-to-speech software program, commonly called a "screen reader," is an assistive technology that allows someone who is blind to access electronic information, such as Web pages on the Internet. Designers of Web sites and software applications need to understand accessibility requirements so they can make their content accessible to users of screen readers. For example, Web sites and other software programs need to have text "tags" that describe every non-trivial image used in their applications. The screen reader reads these descriptions aloud so the user with low or no vision can understand the information being conveyed. These tags benefit sighted people too because the tag's text appears whenever they "mouse over" the image, which can help the users identify the function of an icon.

People who are deaf need visual assistance to access information that is typically delivered aurally. As a result of the Television Decoder Circuitry Act of 1990, televisions must now be manufactured with the circuitry necessary to show captioning. In this case, the viewer requires no assistive technology, and these caption-ready TVs are examples of making technology accessible "out of the box." Continuing the example related to Web site design, when sound is used to communicate information, there should be text to notify the user of the presence of sound and to describe the sound itself. Often the solution is as simple as a caption that says "music."

Another assistive technology is voice-to-text software, commonly called "voice recognition software," which assists those with motor or dexterity limitations to transmit information through speech. While this assistive technology was initially developed for people with disabilities, voice recognition is a popular feature for many technology products and is in high demand by the mass market.

Training for Accessible Design

Accessibility training is available to designers in several places around the country, and there is a wealth of information on the Internet. Web sites listed in the resources section of this article provide contacts for online training, courses, and conferences. Many of these courses are "hands on" so designers and information specialists can see the assistive technology software packages in action. Online training sessions are available through the national centers listed at the end of this article. Several centers have training materials that can be downloaded or requested through the U.S. Mail. Professional societies, such as the Human-Computer Interaction (HCI) group of the Association for Computing Machinery (ACM), American Institute of Graphic Arts (AIGA), and the Industrial Design Society of America (IDSA) include peer-reviewed papers and research on related topics.

FUTURE TRENDS

It is important for information providers and designers to know the requirements and guidelines of Sections 508 and 255 before they begin a project. If they include people with disabilities early in their user-centered design processes, they can avoid the cost of retrofitting later. In addition, increased software and hardware capabilities will ease compliance through software that easily converts information to text, or sound to image.

CONCLUSION

Section 508 of the Rehabilitation Act and Section 255 of the Communications Act are the two primary federal laws that directly affect the design of accessible electronic, information, and telecommunication technology in the U.S. These laws go beyond the Americans with Disabilities Act by defining specific standards and guidelines that address the accessibility needs of people with disabilities. Designers of technology, therefore, should understand and apply these principles when designing electronic and information technology that will be sold to the federal government (Section 508), or when developing telecommunications products and services that will be sold to the public (Section 255).

It would be a mistake, however, to assume that federal legislation is the only factor driving the need for more accessible technology. The U.S. population is aging, and with the years come age-related disabilities such as visual, hearing, and mobility impairments. The appeal of accessible technology, therefore, extends to a larger portion of the overall market, and accessibility requirements should become an essential consideration in the technology design process.

REFERENCES

Access Board. (2000). *36 CFR, SubPart 1194 text.*

Federal Acquisition Regulations (FAR). (2001). *Federal Register*, (April 25). Retrieved from *www.section508.gov/index.cfm?FuseAction=Content&ID=13*

Forrester. (2004). Study commissioned by Microsoft. Retrieved from *www.microsoft.com/presspass/press/2004/feb04/02-02AdultUserBenefitsPR.asp*

Section 508. (1998). *Section 508 of the Rehabilitation Act (29 U.S.C. 794d), as amended by the Workforce Investment Act of 1998* (P.L. 105-220), August 7, 1998.

Vanderheiden, G.C. & Tobias, J. (2000). *Universal design of consumer products: Current industry practice*. Madison, WI: Trace Research and Development Center.

RESOURCES

Architectural and Transportation Barriers Compliance Board

- Also known as the Access Board: www.access-board.gov
- ADA Accessibility Guidelines for Buildings and Facilities: www.access-board.gov/adaag/html/adaag.htm
- Section 508 standards: www.section508.gov/index.cfm?FuseAction=Content&ID=12
- Section 255 guidelines: www.access-board.gov/telecomm/html/telfinl2.htm

Section 508 of the Rehabilitation Act

- Also known as Section 508: www.Section508.gov

Federal Communications Commission

- Also known as the FCC: www.FCC.gov
- Section 255 information page: www.fcc.gov/cgb/dro/section255.html

Department of Justice

- Also known as the DOJ: www.usdoj.gov
- Guide to Disability Rights Laws: www.usdoj.gov/crt/ada/cguide.htm
- ADA Title III Technical Assistance Manual: www.usdoj.gov/crt/ada/taman3.html

Information Technology Technical Assistance and Training Center

- Also known as ITTATC: www.ittatc.org
- Accessibility in the user-centered design process:

www.ittatc.org/technical/access-ucd/
- Web accessibility online course: www.ittatc.org/training/Webcourse/
- Product accessibility evaluation Webcast: www.tvworldwide.com/event_020314_ittatc.cfm
- Other Webcasts about accessibility and Sections 508 and 255: www.ittatc.org/training/Web_training.cfm
- Speak Out! about inaccessible information and telecommunication technology: www.ittatc.org/technical/speakout/index.cfm

The TRACE Research Center at the University of Wisconsin–Madison

- Also known as the TRACE Center: www.trace.wisc.edu
- Collation of Access Board's 508 Final Rule and Guides: trace.wisc.edu/docs/508-collation/index.shtml?style=default
- Product design ideas browser: trace.wisc.edu/docs/browser/index.html

National Institute on Disability and Rehabilitation Research
- Also known as NIDRR: www.ed.gov/about/offices/list/osers/nidrr/index.html?src=mr
- List of NIDRR-funded programs and projects: www.ed.gov/rschstat/research/pubs/programs.html

Center for Assistive Technology and Environmental Access
- Also known as CATEA: www.catea.org
- List of CATEA projects: www.catea.org/projects.html

KEY TERMS

Accessible Technology: "[T]echnology that can be used by people with a wide range of abilities and disabilities. It incorporates the principles of universal design. Each user is able to interact with the technology in ways that work best for him or her. Accessible technology is either directly accessible—in other words, it is usable without assistive technology—or it is compatible with standard assistive technology" (Knowledgebase entry from University of Washington: www.washington.edu/accessit/articles?110).

Assistive Technology: Any item, piece of equipment, or system that is commonly used to increase, maintain, or improve functional capabilities of individuals with disabilities. Examples include screen readers, teletypewriters (TTYs), and Braille keyboards.

Disability: Under the ADA, an individual with a disability is a person who: (1) has a physical or mental impairment that substantially limits one or more major life activities; (2) has a record of such an impairment; or (3) is regarded as having such an impairment.

Rehabilitation Act: A federal law that was created to empower individuals with disabilities to maximize employment, economic self-sufficiency, independence, and inclusion and integration into society. It also was enacted to ensure that the federal government plays a leadership role in promoting the employment of individuals with disabilities. You can read the text of the act at www.ed.gov/policy/speced/leg/rehabact.doc.

Section 255: A part of the Communications Act. Section 255 requires telecommunications manufacturers and providers of telecommunications services to make their products and services accessible to and usable by individuals with disabilities, if readily achievable. For more information, see www.access-board.gov/telecomm/html/telfinal.htm.

Section 508: A part of the Rehabilitation Act. Under Section 508, agencies must give employees with disabilities and members of the public access to information that is comparable to the access available to others. The law applies to all federal agencies when they develop, procure, maintain, or use electronic and information technology. For more information, see www.section508.gov.

Telecommunications Act: The first major overhaul of American telecommunications policy in nearly 62 years. This Act added Section 255 to the Communications Act of 1934. See www.fcc.gov/telecom.html for more information.

Universal Design: "A process of creating products (devices, environments, systems, and processes) which are usable by people with the widest possible range of abilities, operating within the widest possible range of situations (environments, conditions, and circumstances), as is commercially practical" (Vanderheiden & Tobias, 2000).

ENDNOTES

1 Forrester, 2004. Study commissioned by Microsoft: http://www.microsoft.com/presspass/press/2004/feb04/02-02AdultUserBenefitsPR.asp

2 Section 508, 1998. Section 508 of the Rehabilitation Act (29 U.S.C. 794d), as amended by the Workforce

Investment Act of 1998 (P.L. 105-220), August 7, 1998 Full text of Section 508

3 Access Board, 2000 36 CFR, SubPart 1194 text

4 Federal Acquisition Regulations (FAR), As published in the Federal Register April 25, 2001 http://www.section508.gov/index.cfm?FuseAction=Content&ID=13

Unified Modeling Language

Peter Fettke
Johannes Gutenberg-University Mainz, Germany

INTRODUCTION

Mature engineering disciplines are generally characterized by accepted methodical standards for describing all relevant artifacts of their subject matter. Such standards not only enable practitioners to collaborate, but they also contribute to the development of the whole discipline. In 1994, Grady Booch, Jim Rumbaugh, and Ivar Jacobson joined together to unify the plethora of existing object-oriented systems engineering approaches at semantic and notation level (Booch, 2002; Fowler, 2004; Rumbaugh, Jacobson, & Booch, 1998). Their effort led to the Unified Modeling Language (UML), a well-known, general-purpose, tool-supported, process-independent, and industry-standardized modeling language for visualizing, describing, specifying, and documenting systems artifacts. *Table 1* depicts the origin and descent of UML.

UML is applicable to software and non-software domains, including software architecture (Medvidovic, Rosenblum, Redmiles, & Robbins, 2002), real-time and embedded systems (Douglass, 1998), business applications (Eriksson & Penker, 2000), manufacturing systems (Bruccoleri, Dieaga, & Perrone, 2003), electronic commerce systems (Saleh, 2002), data warehousing (Dolk, 2000), bioinformatics (Bornberg-Bauer & Paton, 2002) and others. The language uses multiple views to specify system's structure and behavior. The recent version UML 1.5 supports nine different diagram types. *Table 2* and *Figure 1* overview the main concepts of each diagram, a more detailed description is given below. For a full description of all semantics see Fowler (2004), OMG (2003a) and Rumbaugh et al. (1998).

The specification of the UML is publicly available and maintained by the Object Management Group (OMG). OMG's standardization process is formalized and consists of several proposal, revision, and final implementation activities (Kobryn, 1999, p. 31f.). Modeling tools supporting the development of UML diagrams are available from a number of commercial vendors and the open source community (OMG, 2004; Robbins & Redmiles, 2000).

BACKGROUND

There is a great deal of terminological confusion in the modeling literature. A modeling language or grammar provides a set of constructs and rules that specify how to combine the constructs to model a system (Wand & Weber, 2002, p. 364). It can be distinguished between an abstract syntax and a concrete syntax or notation of a language. While the abstract syntax specifies conceptual relationships between the constructs of the language, the concrete notation defines symbols representing the abstract constructs. In contrast, a modeling method provides procedures by which a language can be used. A consistent and suited set of modeling methods is called a methodology. A model is a description of a domain using a particular modeling language.

The UML specification provides an abstract syntax and a concrete notation for all UML diagrams as well as an informal description of the constructs' semantics. The UML's language specification is independent of but strongly related to other OMG standards such as Common Data Warehouse Model, XML Metadata Interchange or Meta Object Facility. A modeling method or a modeling methodology is not defined by the UML standard. Hence, the language is process-neutral and can be used with different software development processes.

Conceptual modeling has a long history. Other modeling approaches that are to a certain degree accepted in practice, for instance the Entity-Relationship Model or flow charts, have a much more limited scope than UML. These approaches address just some aspects of systems' specification, namely data and process view. In contrast, UML supports the specification of static as well as dynamic aspects. Other approaches with a similar scope, for example, Open Modeling Language (Firesmith, Henderson-Sellers, & Graham, 1998), are not widely accepted in practice.

STRUCTURAL DIAGRAMS

Structural or static diagrams describe the objects of a system in terms of classes, attributes, operations, relationships, and interfaces.

(1) *Class diagram*. A class diagram can be viewed as a graph of several elements connected by static relationships. The main element is a class. Classes represent concepts within the system being modeled and are descriptors for a set of objects with similar structure, behavior, and relationships. An

Table 1. History of UML (Fowler, 2004, pp. 151-159; Kobryn, 1999, p. 30)

Version	Year	Comments
0.8	1995	Origin of UML, so-called “Unified Method”
0.9	1996	Refined proposal
1.0	1997	Initial submission to OMG
1.1	1997	Final submission to OMG
1.2	1998	Editorial revision with no significant technical changes
1.3	1999	New use case relationships, revised activity diagram semantics
1.4	2001	Minor revisions, addition of profiles
1.5	2003	Adding action semantics
2.0	2004 (?)	Planned major revision, deep changes to meta-model, new diagram types

Table 2. UML diagram types

Focus	Diagram	Purpose	Main Concepts
Static diagrams	Class	Object structure	Class, features, relationships
	Object	Example configuration of instances	Object, link
Dynamic diagrams	Use case	User interaction with system	Use case, actor
	Sequence	Interaction between objects emphasizing sequences	Interaction, message
	Collaboration	Interaction between objects emphasizing collaborations	Collaboration, interaction, message
	Statechart	Change of events during object’s lifetime	State, transition, event, action
	Activity	Procedural and parallel behavior	State, activity, completion, transition, fork, join
Implementation diagrams	Component	Structure and connections of components	Component, interface, dependency
	Deployment	Deployment of components to nodes	Node, component, dependency

object represents a particular instance of a class. Each class has a unique name among other classes within a specific scope (usually a UML package). A class can hold several attributes and operations. Attributes have names and belong to particular types that can be simple data types such as integer, string, and Boolean, as well as complex types (e.g., other classes). Operations are services offered by an instance of the class and may be requested by other objects during run-time. Different relationships between classes can be defined.

Figure 2 depicts a class diagram for banking systems. An account is described by the attributes “number” and “balance.” The operations “deposit,” “withdrawal,” and “freeze” are offered by an account. Each account is kept by a “branch” and is assigned to a “holder.” The classes “deposit account” and “current account” reuse the structure and behavior of the class “account” (inheritance relationship). In addition, the specialized account classes define further feature; for example, an object of the class “current account” is described by the property “overdraft facility” and offers an operation calculating the current debit balance.

(2) *Object diagram*. An object diagram is an instance of a class diagram and depicts the state of the system at a point in time (e.g., a particular configuration of several objects). It contains objects including their actual values of attributes and links describing object references.

BEHAVIORAL DIAGRAMS

Behavioral diagrams describe the dynamics between objects of a system in terms of interactions, collaborations, and state histories.

Figure 1. UML diagram examples

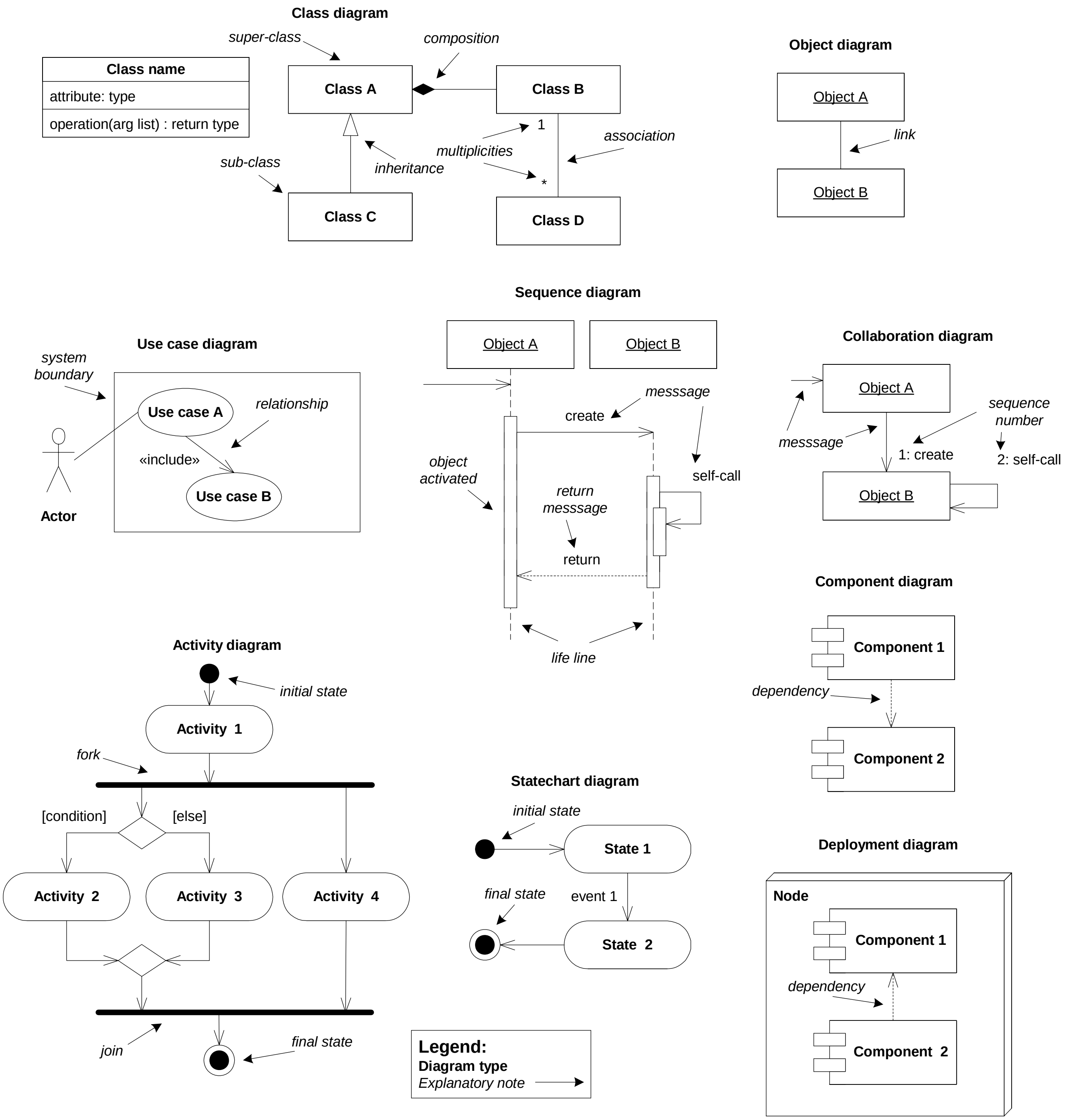
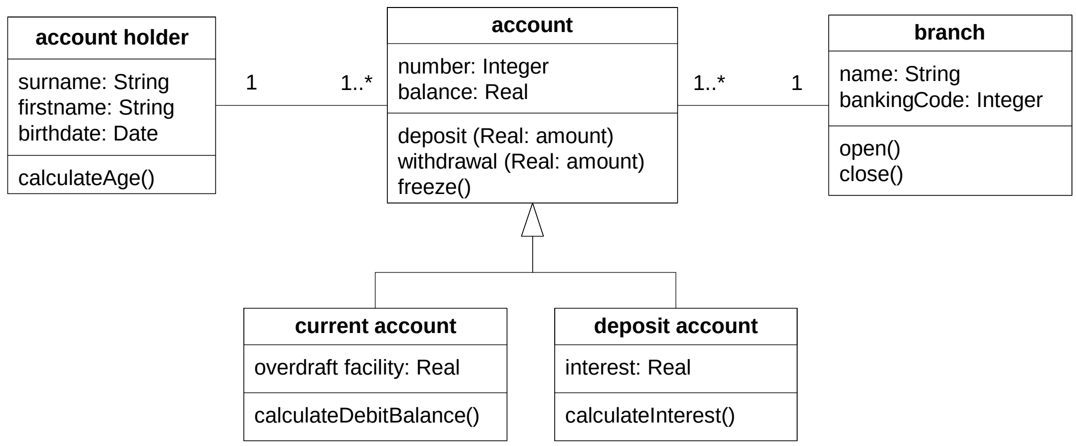


Figure 2. Class diagram for banking systems



(1) *Use case diagram*. A use case specifies a complete set of events within a system to fulfill tasks or transactions in an application from a user's point of view. In a use case diagram, a set of use cases, actors, and relationships between these elements are depicted. Several use cases may optionally be enclosed by a rectangle that represents the boundary of the containing system. An actor describes a particular role of a human or non-human user of the system being modeled.

(2) *Sequence diagram*. Sequence diagrams describe interactions between different objects. An interaction consists of a partially ordered set of messages that are exchanged by the participants of that interaction. Sequence diagrams have two dimensions: the horizontal dimension represents the participants of the interaction; the vertical dimension represents the flow of time (usually time proceeds from up to down).

(3) *Collaboration diagram*. Collaboration and sequence diagrams use the same underlying information and can easily be transformed into each other. While sequence diagrams emphasize the sequence of communication between objects, collaboration diagrams show the roles of the participants of an interaction and their relationships. A sequence number specifies the flow of messages in an interaction, so no time dimension is needed in this diagram. Simple communication patterns can be depicted by collaboration diagrams; sequence diagram can better specify complex message exchanges or requirements for real-time systems.

(4) *Statechart diagram*. Object behavior is represented by statechart diagrams that can specify the behavior of an entire object or a single method. A state describes a condition during the lifetime of an object. Transitions are relationships between two states describing that an object's state can change from the first to the second state. The change of a state is triggered by an event that occurs in the modeled system. There are two special types of states: an initial state identifies the point at which behavior starts when an object is created, a final state identifies the point at which behavior ends (end of object's lifetime).

(5) *Activity diagram*. While statechart diagrams are used to specify the behavior of a single object, activity diagrams can describe behavior that crosses object boundaries. They are analogous to traditional flowcharts and are often used to document (business) processes or the dynamics inside a use case. So-called fork bars are used to describe activities that can be executed in parallel. Parallel activities get synchronized by so-called join bars. Guards are used to specify conditional forks that are only executed if particular conditions hold.

IMPLEMENTATION DIAGRAMS

Implementation diagrams capture the physical structure of a software system during build- and run-time.

(1) *Component diagram*. Components in UML are physical elements such as source, binary or executable modules and files respectively. Simple components can be aggregated to complex components to specify physical containment relations. Directed relationships between components specify that one component relies or refines the other. Such relationships are called dependencies.

(2) *Deployment diagram*. While component diagrams primary show build-time dependencies of components, deployment diagrams show a run-time configuration of the system's components. In addition, a deployment diagram uses nodes representing a processing resource, for instance a server or workstation that can execute system operations during run-time.

ADVANCED TOPICS

Object Constraint Language (OCL). There is often a need to capture unambiguously a system's semantics in a precise and rigorous way. OCL is used for that purpose. It is a formal textual language inspired by the "Design by Contract" concept and provides concepts for the definition of constraints such as invariants, pre- and post-conditions (Cengarle & Knapp, 2004; Warmer & Kleppe, 2003).

Language specification and meta-model. The UML itself is specified using textual descriptions and a four-layered meta-modeling approach (Atkinson & Kühne, 2002, pp. 291-296). In this approach, the semantic constructs at each layer are recursively refined. The top layer, the meta-meta-model (M3), provides a so-called Meta Object Facility (MOF) to specify meta-models on the next lower layer. The MOF is used on the meta-model (M2) layer to specify the concepts of UML diagrams,for example, class diagram and etcetera. The model (M1) and object (M0) layer are user-defined. The former specifies concrete UML models, the later instances of the former.

Extension mechanisms. So-called heavyweight extensions are supported by MOF and carried out on the meta-model (M2) layer. Such extensions have great impact on the language and are not performed by a particular modeler. User extensions are usually lightweight extensions that are built-in mechanisms of the UML. Lightweight extensions comprise constraints (OCL expressions), tagged values (attached additional information to model elements), and stereotypes [most powerful lightweight mechanism ranging from concrete syntax modifications to semantics redefinitions (Berner, Glinz, & Joos, 1999)].

FUTURE TRENDS

The forthcoming UML Version 2.0 (UML 2) was at first planned for 2001 (Kobryn, 1999, p. 30) but is until now (early 2004) not fully completed. In the meantime, a strong discussion about what UML 2 should and should not be evolved (Engels, Heckel, & Sauer, 2001; Miller, 2002). Currently, the UML 2 standard is voted to recommend by the OMG's technical board (OMG, 2003b) and is in its finalization phase. It consists of four separate documents (Kobryn, 2002). The UML 2 Infrastructure Specification is concerned with core language features. Advanced topics such as component and activity modeling are specified in the UML Superstructure Specification. The OCL and the Diagram Interchange Specification are two further separate UML 2 specification documents. This major revision mainly focuses on language extensibility, language specification, language precision and expressiveness. Although the complete language specification is almost fully rewritten, this revision will be primary an internal reorganization with just minor consequences for the end user. For example, the new diagrams mainly clarify and resemble existing diagram types.

Further trends include:

- *Model Driven Architecture* (MDA). MDA promotes modeling through the whole system's life-cycle (Frankel, 2003). Its objective is to fully automate the system's development process.
- *Executable UML*. Executable UML enriches modeling concepts with execution semantics (Mellor & Balcer, 2002). This opens the possibility of software development without "classical" programming.
- *Model libraries*. UML is used to standardize domain-specific models fostering the usage of reference models (Fettke & Loos, 2003a). Known reference models, for example, OMG's Business Enterprise Integration or Finance Domain Task Forces, they support model reuse.
- *Ontological analysis and semantics*. This research line evaluates UML from an ontological point of view and incorporates real-world semantics to UML constructs (Opdahl & Henderson-Sellers, 2002). The aim of an ontological evaluation is to examine if all constructs of an ontology can be mapped onto the constructs of UML and vice versa.
- *Component-based development*. UML is primary an object-oriented language. To fully support component-based development, some enhancements are needed (Dahanayake, 2003; Fettke & Loos, 2003b; Kobryn, 2000). Particularly, component descriptions must include dependencies on other components, quality specifications for needed and offered services, and domain-specific semantics.

CONCLUSION

Although almost everyone acknowledges the practical benefits of a standardized modeling language [e.g., protection of investments in technology, easier model exchange and reuse, better professional training (Frank, 1997, p. 13)], there are important opportunities that have to be challenged. UML's size (UML 2 has approximately 1,000+ pages) and complexity is overwhelming compared with other languages (Siau & Cao, 2001). Therefore users have difficulties in writing and reading diagrams (Agarwal & Sinha, 2003; Laitenberger, Atkinson, Schlich, & Emam, 2000) and tool venders have problems to fully support the UML standard. Furthermore, the maintenance of the standard is very expensive and error-prone, for example, Fuentes, Quintana, Llorens, Génova, & Prieto-Díaz (2003) identified several hundred errors in UML's meta-model. Other authors criticize UML for its semantic inconsistency, construct ambiguity, notation inadequacy, and cognitive misdirection (Champeaux, 2003; Frank, 1998; Henderson-Sellers, 2002; McLeod, Halpin, Kangassalo, & Siau, 2001; Shen & Siau, 2003; Thomas, 2002; Wang, 2001).

On the other hand, UML is the de-facto standard for object-oriented modeling and an important milestone in software engineering. Modeling of software systems increases the degree of abstraction during system development tremendously. This change is similar to the replacement of assembly languages by high-level languages in the 1960s and 1970s. Today, high-level languages are not used in all but most domains. We predict that, in the future, UML has an analogous position as high-level languages have today. Hence, UML continues to play a major role in systems development.

REFERENCES

Agarwal, R., & Sinha, A. P. (2003). Object-oriented modeling with UML: A study of developers' perceptions. *Communications of the ACM, 46* (9), 248-256.

Atkinson, C., & Kühne, T. (2002). Rearchitecting the UML infrastructure. *ACM Transactions on Modeling and Computer Simulation, 12* (4), 290-321.

Berner, S., Glinz, M., & Joos, S. (1999). A classification of stereotypes for object-oriented modeling languages. In R. France & B. Rumpe (eds.), *UML '99 - The Unified Modeling Language - Beyond the Standard* (Vol. 1723) (pp. 249-264). *Second International Conference, Fort Collins, CO, October 28-30, 1999*,). Berlin: Springer.

Booch, G. (2002). Growing the UML. *Software and Systems Modeling, 1*, 157-160.

Bornberg-Bauer, E., & Paton, N.W. (2002). Conceptual data modelling for bioinformatics. *Briefings in Bioinformatics, 3* (2), 165-180.

Bruccoleri, M., Dieaga, S.N.L., & Perrone, G. (2003). An object-oriented approach for flexible manufacturing control systems analysis and design using the unified modeling language. *The International Journal of Flexible Manufacturing Systems, 15*, 195-216.

Cengarle, M. V., & Knapp, A. (2004). OCL 1.4/5 vs. 2.0 Expressions: Formal semantics and expressiveness. *Software and Systems Modeling, Online First Issue*. Retrieved from www4.in.tum.de/lehre/seminare/hs/WS0405/uml/CK04b.pdf.

Champeaux, D.D. (2003). Extending and shrinking UML. *Communications of the ACM, 46* (3), 11-12.

Dahanayake, A. (2003). Methodology evaluation framework for component-based system development. *Journal of Database Management, 14* (1), 1-26.

Dolk, D.R. (2000). Integrated model management in the data warehouse era. *European Journal of Operational Research, 122*, 199-218.

Douglass, B.P. (1998). *Real-time UML: Developing efficient objects for embedded systems*. Reading, MA: Addison-Wesley.

Engels, G., Heckel, R., & Sauer, S. (2001). UML: A Universal Modeling Language? In M. Nielsen & D. Simpson (eds.), *Application and Theory of Petri Nets 2000: 21st International Conference* (pp. 24-38). *ICATPN 2000, June 2000, Aarhus, Denmark*. Berlin: Springer.

Eriksson, H.E., & Penker, M. (2000). *Business modeling with UML: Business patterns at work*. New York: John Wiley & Sons.

Fettke, P., & Loos, P. (2003a). Classification of reference models: A methodology and its application. *Information Systems and e-Business Management, 1* (1), 35-53.

Fettke, P., & Loos, P. (2003b). Specification of business components. In M. Aksit, M. Mezini, & R. Unland (eds.), *Objects, Components, Architectures, Services, and Applications for a Networked World* (Vol. 2591) (pp. 62-75). *International Conference NetObjectDays, NODe 2002, Erfurt, Germany, October 7-10, 2002, Revised Papers*. Berlin: Springer.

Firesmith, D., Henderson-Sellers, B., & Graham, I. (1998). *The OPEN Modeling Language (OML) reference manual*. Cambridge: Cambridge University Press.

Fowler, M. (2004). *UML distilled: A brief guide to the standard object modeling language* (3rd ed.). Boston: Addison-Wesley.

Frank, U. (1997). *Towards a standardization of object-oriented modelling languages?* (Working Paper No. 3). Koblenz, Germany: Institut für Wirtschaftsinformatik der Universität Koblenz Landau.

Frank, U. (1998). Object-oriented modelling languages: State of the art and open research questions. In M. Schader & A. Korthaus (eds.), *The Unified Modeling Language: Technical Aspects and Applications* (pp. 14-31). Heidelberg: Physica.

Frankel, D.S. (2003). *Model driven architecture: Applying MDA to enterprise computing*. Indianapolis: Wiley.

Fuentes, J. M., Quintana, V., Llorens, J., Génova, G., & Prieto-Díaz, R. (2003). Errors in the UML metamodel? *ACM SIGSOFT Software Engineering Notes, 28* (6), 1-13.

Henderson-Sellers, B. (2002). The use of subtypes and stereotypes in the UML model. *Journal of Database Management, 13* (2), 43-50.

Kobryn, C. (1999). UML 2001: A standardization odyssey. *Communications of the ACM, 42* (10), 29-37.

Kobryn, C. (2000). Modeling components and frameworks with UML. *Communications of the ACM, 43* (10), 31-38.

Kobryn, C. (2002). Will UML 2.0 be agile or awkward? *Communications of the ACM, 45* (1), 107-110.

Laitenberger, O., Atkinson, C., Schlich, M., & Emam, K.E. (2000). An experimental comparison of reading techniques for defect detection in UML design documents. *The Journal of Systems and Software, 53* (2000), 183-204.

McLeod, G., Halpin, T., Kangassalo, H., & Siau, K. (2001). *UML: A critical evaluation and suggested future*. Paper presented at the 34th Hawaii International Conference on System Sciences, Hawaii.

Medvidovic, N., Rosenblum, D.S., Redmiles, D.F., & Robbins, J.E. (2002). Modeling software architectures in the Unified Modeling Language. *ACM Transactions on Software Engineering and Methodology, 11* (1), 2-57.

Mellor, S.J., & Balcer, M.J. (2002). *Executable UML: A foundation for model-driven architecture*. Boston: Addison Wesley.

Miller, J. (2002). What UML should be. *Communications of the ACM, 45* (11), 67-69.

OMG. (2003a). *OMG Unified Modeling Language Specification: Version 1.5, formal/03-03-01*. Needham, MA.

OMG. (2003b). *UML 2.0 standard officially adopted at OMG technical meeting in Paris*. Retrieved March 1, 2004, from *www.omg.org/news/releases/pr2003/6-12-032.htm*

OMG. (2004). *UML tools*. Retrieved March 1, 2004, from *www.omg.org/uml*

Opdahl, A.L., & Henderson-Sellers, B. (2002). Ontological evaluation of the UML using the Bunge-Wand-Weber Model. *Software and Systems Modeling, 1* (1), 43-67.

Robbins, J.E., & Redmiles, D.F. (2000). Cognitive support, UML adherence, and XMI interchange in Argo/UML. *Information and Software Technology, 42*, 79-89.

Rumbaugh, J., Jacobson, I., & Booch, G. (1998). *The Unified Modeling Language reference manual*. Boston: Addison-Wesley.

Saleh, K. (2002). Documenting electronic commerce systems and software using the unified modeling language. *Information and Software Technology, 44* (2002), 303-311.

Shen, Z., & Siau, K. (2003). *An empirical evaluation of UML notational elements using a concept mapping approach*. Paper presented at the Twenty-Fourth International Conference on Information Systems, Seattle, Washington, USA.

Siau, K., & Cao, Q. (2001). Unified Modeling Language (UML): A complexity analysis. *Journal of Database Management, 12* (1), 26-34.

Thomas, D. (2002). UML: Unified or universal modeling language? *Journal of Object Technology, 2* (1), 7-12.

Wand, Y., & Weber, R. (2002). Research commentary: Information systems and conceptual modeling: A research agenda. *Information Systems Research, 13* (4), 363-377.

Wang, S. (2001). *Experiences with the Unified Modeling Language (UML)*. Paper presented at the Seventh Americas Conference on Information Systems (AMCIS) 2001.

Warmer, J.V., & Kleppe, A. (2003). *The object constraint language: Getting your models ready for MDA* (2nd ed.). Boston: Addison-Wesley.

KEY TERMS

Conceptual Modeling: An action describing a domain with the help of some artificial or formalized language.

Meta-Model: A meta-model is a model of model.

Methodology: A consistent and suited set of modeling methods providing procedures to apply the constructs of a modeling language.

Model: A model is a particular product of conceptual modeling. It is a description of a domain using a particular language.

Object-Oriented Analysis and Design (OOA & OOD): Software engineering approach to constructing software systems by building object-oriented models that abstract key aspects of the target system.

Object-Oriented Programming (OOP): Object-oriented programming emphasizes the hiding or encapsulation of the inner state of objects and the specification of these objects by an interface. OOP languages support objects, classes and inheritance.

Reference Model: A reference model is a model representing a class of domains, for example, a reference model for production planning and control systems. It is a conceptual framework or blueprint for system's development.

Usability and Learnability Evaluation of Web-Based ODL Programs

Karoulis Athanasis
Aristotle University of Thessaloniki, Greece

Pombortsis Andreas
Aristotle University of Thessaloniki, Greece

INTRODUCTION

The rapid establishment of third generation distance learning environments, the so-called Web-based or tele-teaching environments, brought some problems with it. The main means for the delivery of the new educational approach is the World Wide Web, and there are some good reasons for it: It is easily accessible by many groups of learners. It supports multiple representations of educational material and various ways of storing and structuring this information. It is powerful and easy to use as a publishing medium. Additionally, it has been widely accepted that the hyper-medial structure of the Web can support learning. Some researchers characterize the Web as an active learning environment that supports creativity. In addition to this, the Web encourages exploration of knowledge and browsing, behaviors that are strongly related to learning. The associative organization of information in the Web is similar to that of human memory, and the process of information retrieval from the Web presents similarities to human cognitive activities (Tselios, Avouris, Dimitracopoulou, & Daskalaki, 2001). However, a hyper-medial space, like the Web, cannot be considered, only by these features, as an effective tutoring environment. It is rather more appropriate to think of the Web as a powerful tool that can support learning, if used in an appropriate way. This is because learning is a process (Duchastel, 2001) that depends on other features, such as learner's motivation, previous experience and learning strategies that the individual has been supported to develop, and so forth. Effectiveness of any educational environment cannot be considered independently of these aspects.

USABILITY VS. LEARNABILITY

To define the notion of "learnability", we must first answer the question, "what makes the (instructional) content of an environment easily learned?" The answer to this question defines in general the learnability of the environment. A more formal definition attempt would be the statement that it is *the set of the methods, the modes and the tools of the environment, which facilitate the acquisition of knowledge*. This work assumes that the notion of learnability embeds de facto the notion of the communication channel; it is impossible for someone to gain knowledge if the environment lacks the means to offer it. So, an evaluation of the learnability of the environment, evaluates hence the success rate of the established communication channel. Continuing on, the next question that emerges is the relation between usability and learnability and in how far one affects the other, which is a central question in this chapter. In more detail, we are interested if a usable environment facilitates the acquisition of knowledge, or, the opposite, if a less usable environment sets certain obstacles. And finally, this question transits to the question if one can heuristically assess the learnability of an educational environment. Heuristic evaluation, as described initially by Nielsen and Molich (1990) and following by Nielsen (1993), is a well-established, expert-based interface evaluation method for the evaluation of the usability of a particular interface. A set of interface experts is asked to judge the interface, based on a set of a few criteria, the heuristics.

What do we mean by the term "usability"? According to ISO-9241 (ISO, 1998) standard, we have the following definition:

Usability of a system is its ability to function effectively and efficiently, while providing subjective satisfaction to its users.

Usability of an interface is usually associated with five parameters (ISO, 1998; Nielsen, 1993), derived directly from this definition:

1. *Easy to learn*: The user can get work done quickly with the system;
2. *Efficient to use*: Once the user has learned the system, a high level of productivity is possible;

3. *Easy to remember*: The casual user is able to return to using the system after some period without having to learn everything all over again;
4. *Few errors*: Users do not make many errors during the use of the system or if they do so they can easily recover them; and
5. *Pleasant to use*: Users are subjectively satisfied by using the system; they like it.

To conclude, when the synergy between usability and learnability occurs, the use of the software can be thought of as "integrated", in that a seamless union develops between the use of the software and the learning process (Squires & Preece, 1996).

THE EDUCATIONAL EVALUATION

The evaluation methodologies applied in the field usually utilize questionnaires in the classroom, however most questionnaires embody closed-type questions, so they lack the opportunity to clarify some other aspects that could be of interest, have an impact on the environment and on the involvement of the software on learning. Moreover, closed-type questions do not take into consideration the individual differences of the students in learning. In general, quantitative approaches to evaluate an educational environment have been strongly debated as monosemantic approaches that must be supplemented by qualitative ones, which focus on *how* and *what* the student learns.

Other studies in the research field of the evaluation of a distance learning environment are the studies of Holmberg (1977), Saba and Shearer (1994), and Garrison (1993), however none of them deals absolutely in the field of evaluation of Web-based environments, as it is in the case of Makrakis, Retalis, Koutoumanos, and Skordalakis (1998) and Koutoumanos, Papaspyrou, Retalis, Maurer, and Skordalakis (1996). These are all user-based evaluation approaches, since they utilize more or less questionnaires that have to be answered by users.

THE HEURISTIC EVALUATION

Concerning the expert-based approaches, Jacob Nielsen and Rolf Molich (1990) started their research in 1988, and in 1990 they presented the "heuristic evaluation". The basic point was the reduction of the set criteria to just a few, at the same time being broadly applicable and generally agreed; simultaneously augmenting the evaluators' expertise, and consequently their reliability. These "heuristic rules" or "heuristics" derived from studies, criteria lists, field observations and prior experience of the domain.

The core point to evaluate in the initial approach is the usability of the interface. Based on the ISO principles about usability (ISO, 1998), Nielsen (1993) stated the following heuristics, slightly modified and reorganized by us:

1. Simple and natural dialog and aesthetic and minimalistic design;
2. Visibility of the system status – provide feedback;
3. Speak the users' language: match between system and real world;
4. Minimize the users' cognitive load: recognition rather than recall;
5. Consistency and standards;
6. Flexibility and efficiency of use – provide shortcuts;
7. Support users' control and freedom;
8. Prevent errors;
9. Help users recognize, diagnose and recover from errors with constructive error messages; and
10. Help and documentation.

The method refers mainly to traditional formative human-computer interface evaluation, yet a number of studies (e.g., Instone, 2000; Levi & Conrad, 1996) have proven its easy adaptability to the evaluation of Web sites as well.

LEARNABILITY HEURISTIC LIST

The next step one must perform is the construction of the heuristics for learnability. A good starting point provides the socio-constructivist view of instruction. Some studies in the field (e.g., Kordaki, Avouris, & Tselios, 2000) argue that an expert evaluator cannot predict the students' performance, although he/she can assess heuristically the learnability of the environment, however with mediocre results. The authors base their claims on the constructivist approach for open learning environments, also known sometimes as microworlds.

Squires and Preece (1999) proceed one step further: They do not make a combination, but a fusion of the Nielsen's heuristics with the five socio-constructivist learning criteria (credibility, complexity, ownership, collaboration and curriculum) providing thus a new list, which they claim to be a versatile tool to predictively evaluate educational pieces by their usability and simultaneous learnability.

1. Match between designers' and learners' mental models;

2. Ensure navigational fidelity;
3. Provide appropriate levels of learner control;
4. Prevent peripheral cognitive errors;
5. Understandable and meaningful symbolic representation;
6. Support personally significant approaches to learning;
7. Build strategies for the cognitive error recognition, diagnosis and recovery cycle; and
8. Match with the curriculum.

A PROPOSAL FOR A NEW LIST

Two studies performed by the authors (Karoulis & Pombortsis, 2001a, 2001b) based on the aforementioned approaches concerning usability and learnability. They concluded to a combined list of 10 heuristics axes to assess the usability and the learnability of an open and distance learning (ODL) environment. So, the proposed final heuristic list for usability/learnability expert-based evaluation of Web-based ODL environments is as follows.

1. *Content*: Is the quantity, quality and value of the content acceptable? Can one characterize it as "student centered" and "open"?
2. *ODL adaptation and integration*: Does the program provide spatial freedom and temporal flexibility? Are the principles of ODL concerned?
3. *User interface*: Are Nielsen's usability heuristics concerned?
4. *Use of technologies*: Does the environment make full use of the potential of the underlying system, without hampering the students' performance in any way?
5. *Interactivity with the instructional material*: Is there navigational fidelity and correct multimedia elements? Is the environment adaptable, and does it support collaborative work?
6. *Students' support*: Is there adequate guidance and support for the student? Are the supporting elements easily accessible?
7. *Communication channel*: Is synchronous and asynchronous communication possible? Are there tutorial sessions and a tutor?
8. *Acquisition of knowledge*: Easily acquired knowledge and support of personal styles and cognitive levels. Support of authentication of the knowledge.
9. *Projects and "learning by doing"*: Are there enough exercises and hands-on practice? Is exploratory learning supported?
10. *Assessment and self-assessment*: Is the student assessed according to the principles of ODL? Are there self-assessment tools?

DISCUSSION

The general impression from the literature survey we presented in this article and the studies we have conducted is that the method is applicable and provides some great advantages as well, which are advantages of the heuristic evaluation in general: it is cheap, fast, easy to apply; the experts, despite the difficulty in locating them, are more easily brought together than the users; and it is very efficient, according to the problems it discovers in relation to the effort and the resources needed (Nielsen, 1993; Levi & Conrad, 1996).

However, there are some important concerns regarding the applicability of the method one has to keep in mind when working with it.

Maybe the most important concern is that the heuristic approach may sometimes not allow for usability issues that occur for specific tasks and users groups. It is also not able to identify and to take into consideration socially relevant issues (Hess & Azuma, 1991), such as the individual characteristics of ODL students, financial categorization, minorities and cultural differentiations, or the well-known problem of the harmonization of the ODL class, due to the observed student diversity. These issues could lead the evaluation of the usability and the learnability of the system to the edge, in the sense that any opinion of an evaluator could be rejected by an opposing statement of a second one; the paradox being in this case that both are right. This problem occurs in empirical evaluations as well; we presented it briefly as we mentioned the antagonistic studies published up to now on the domain. So, it is obvious that one has to make an assumption on the harmonization of the ODL class, a common profile of the students, which however has to be broad and flexible more than usual, to enclose as many of the characteristics of the "common ODL student" as possible, even if some researchers advocate that "the common user" doesn't exist.

FUTURE TRENDS

Concerning the delivery of information over the Web and the collaboration of many people over a network, socially relevant issues arise as well, such as the preservation of privacy (Bellotti & Sellen, 1993). Information technology can store, transmit and manipulate vast quantities and varieties of information. ODL environments are extremely prone to intrusions, due to their structure, and a clear way to be protected against intrusions has not yet been found. Future research about evaluation methodologies in a holistic manner, should take this point of concern into consideration as well.

CONCLUSION

We claim that expert-based methodologies in general and the heuristic evaluation in particular can be successfully applied in the evaluation of Web-based open and distance learning environments. Generally, many problems of the site under evaluation can be discovered, and the general impression is that with the results of the heuristic evaluation, one could propose many concrete improvements to the site under consideration, which finally is the aim of every evaluation.

Levi and Conrad (1996) argue that the final assessment of every system must be made according to its usability. Myers and Rosson (1992) calculated that 48% of the total code of every software concerns the interface; Avouris (2000) augments this number to 70% while Myers (1992) claims that in many cases the interface is the system itself. We do not entirely agree with this position; we rather believe that the heuristic methodology provides a powerful tool to evaluate the usability and the learnability of a particular interface; however, in the case of complex environments, like ODL environments, it must be combined with in-field evaluations. There is, more or less, a coincidence of the researchers in the field over this issue (Kordaki et al., 2000; Preece, Rogers, Sharp, Benyon, Holland, & Carey, 1994; Squires & Preece, 1999). So, we consider the heuristic approach as a powerful tool for the formative evaluation of the first design stages of the environment; however, as the system matures, it should fade out in favor of more user-centered approaches, which can provide an in-depth analysis of the system's usability and learnability. Yet, one must be aware of the constraints of the empirical evaluations, and, most of all, the problem in finding *real users* to evaluate the system under *real circumstances*.

REFERENCES

Avouris, N. (2000). *Introduction to human-computer interaction*. Diavlos, Athens (in Greek).

Bellotti, V., & Sellen, A. (1993). Design for privacy in ubiquitous computing environments. In *Proceedings of ECSCW'93, The 3rd European Conference on Computer Supported Cooperative Work*, Milano, Italy, 13-17 September, Kluwer.

Duchastel, Ph. (2001). Learnability. Preprint article. Retrieved on September 15, 2001 from the Internet at *http://home.earthlink.net/~castelnet/info/learnability.htm*

Garrison, D.R. (1993). Quality and access in distance education: Theoretical considerations. In D. Keegan (Ed.), *Theoretical principles of distance education* (pp. 9-21). London: Routledge.

Hess, R.D., & Azuma, M. (1991). Cultural support for schooling: Contrasts between Japan and the United States. *Educational Researcher*, *20*(9), 2-8.

Holmberg, B. (1977). *Distance education. A survey and bibliography*. London: Kogan Page.

Instone, K. (2000). Usability heuristics for the Web. Retrieved on October 10, 1997 from the Internet at *http://www.webreview.com/1997/10_10/strategists/10_10_97_2.shtml*

ISO (1998). ISO 9241 - International Standarization Organization. Ergonomic requirements for office work with visual display terminals (VDT's), Part 10, Dialogue Principles.

Karoulis, A., & Pombortsis, A. (2001a). Heuristic evaluation of Web sites: The evaluators' expertise and the heuristic list. *Web-Net Conference*. Orlando, Florida, October 23-27, Charlottesville, VA: AACE publ.

Karoulis, A., & Pombortsis, A. (2001b). Heuristicaly evaluating distance learning Web-based environments. *EDEN 10th Anniversary Conference*, June 10-13, Stockholm.

Kordaki, M., Avouris, N., & Tselios, N. (2000). Tools and methodologies for evaluation of open learning environments. *Proceedings-CD of 2nd Panhellenic Conference with International Participation «Information & Communication Technologies in Education»* (pp. 371-381), Patras, October (in Greek).

Koutoumanos, A., Papaspyrou, N., Retalis, S., Maurer, H., & Skordalakis, E. (1996). Towards a novel networked learning environment. In *Proceedings-CD of World Conference of Web Society (Web Net '96)* (pp. 267-272), San Francisco, AACE.

Levi, M.D., & Conrad, F.G. (1996, July/August). A heuristic evaluation of a World Wide Web prototype, *Interactions Magazine*, *III.4*, 50-61, ACM Publ. Available at *http://www.bls.gov/ore/htm_papers/st960160.htm*

Lewis, C., & Rieman, J. (1994) *Task-centered user interface design - A practical introduction*.

Makrakis, V., Retalis, S., Koutoumanos, A., & Skordalakis, E. (1998). Evaluating the effectiveness of a Web-based open and distance learning system: A case study. *Journal of Universal Computing Science*, *4*(3), 259-272. Berlin: Springer Verlag.

Myers, B.A. (1992, August). Demonstrational interfaces: A step beyond direct manipulation. *IEEE Computer*, *25*(8), 61-73.

Myers, B.A., & Rosson, M.B. (1992). Survey on user interface programming. *Proceedings of ACM CHI '92 Conference* (pp. 195-202). Monterey, CA, May 3-7.

Nielsen, J. (1993). *Usability engineering*. San Diego, CA: Academic Press.

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces, *Proeedings. of Computer-Human Interaction Conference* (CHI) (pp. 249-256), Seattle, WA, April 1-5.

Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., & Carey, T., (1994). *Human-computer interaction*. Reading, Mass.: Addison-Wesley Publishing Company.

Saba, E., & Shearer, R. (1994). Verifying theoretical concepts in a dynamic model of distance education, *American Journal of Distance Education, 8*(1), 36-59.

Squires, D., & Preece, J. (1996). *Usability and learning: Evaluating the potential of educational software.* Computers and Education, 27*(1), 15-22.*

Squires, D., & Preece, J. (1999). *Predicting quality in educational software: Evaluating for learning, usability, and the synergy between them.* Interacting with Computers, The Interdisciplinary Journal of Human-Computer Interaction, 11*(5), 467-483.*

Tselios, N., Avouris, N., Dimitracopoulou, A., & Daskalaki, S. (2001). *Evaluation of distance-learning environments: Impact of usability on student performance.* International Journal of Educational Telecommunications, 7*(4), 355-378.*

KEY TERMS

U

Constructivism: An educational theory, arguing that the student "constructs" his/her own knowledge on the domain, rather than "acquiring" certain behaviors on how to interact with it.

Educational Evaluation: Assessment of the educational value of an educational software piece.

Empirical Evaluation: Evaluation methodology which employs users to interact with the system.

Expert-Based Evaluation: Evaluation methodology which employs experts, mostly from different cognitive domain, to assess certain system aspects.

Heuristics: Rules or criteria which derive directly from practice. Usually, they are not "proven" in a scientific way, yet they are broadly acceptable.

Learnability: The degree to which a system can transfer knowledge in a "natural" way. "Natural" in this context is "the way the user is used to acquire the knowledge."

Usability: The ability of a system to interact "natural" and "transparent" with its users. "Natural" in this context is "the way the user is used to act" and "transparent" means "not interfering with the performed task".

Usability Evaluation of Online Learning Programs

Bernard Blandin
CESI SAS, France

INTRODUCTION

Human activities are social activities, in essence. Evaluating or using a piece of software is a social activity, as well as a learning activity. Nonetheless, these activities are generally considered as relying upon universal and culturally neutral criteria or ways of achieving them. This chapter makes the case for an opposite viewpoint and aims to support the idea that "usability of online learning software" has to be considered as situated.

In its broad sense, "usability" addresses the relationship between tools and their users. From a designer's viewpoint, usability is seen as a relationship between a human being and an artifact that measures the productivity of a user using the artifact (Nielsen, 1994). Some authors consider an extended notion of usability that distinguishes clearly between two usability-related concepts, "ease of use" and "usefulness," but takes them both into account (Notess, 2001). This extended notion of "usability," which is the one supported by this chapter, has a strong dependence on social and cultural aspects, because the use of any object is a social activity, which implicitly implies social relations belonging to different registers (Blandin, 2002).

Therefore, the standpoint presented here is the following: usability evaluation should not be limited to consider "simply ease of use of a tool" criteria, but should also take into account all the conditions that lead a user to actually use a tool. When "learning" is the aim of the activity, underpinned by the use of the said tool, social, cultural, and context-dependent factors are driving forces that shape the action and that need to be taken into account within the evaluation process.

BACKGROUND

Usability, in its first sense in the field of software design and engineering, was synonymous with "ergonomical issues," though the sense of the word can now be broader and encompass "usefulness" or "utility" issues.

In its first, "historical" sense, the word "usability" intended to define rules to design the application to match users' habits, various types of equipments, various languages, etc. The basic principles of a good design for usability, still in use today (Notess, 2001), were, and are, those described in Norman (1988):

1. Provide a good conceptual model
2. Make things visible
3. Map the controls, their movements, and their results in the real world
4. Provide feedback for any action

Such principles apply to "usability engineering" (Notess, 2001) and are considered as good heuristics (Nielsen, n.d.) to design learning software. Their application may, in some conditions, facilitate the use of learning software.

A "good conceptual model" cannot be universal—it is related to a particular culture. If "learning" is concerned, even in a given culture, different "learning styles" and "learning cultures" exist that generate different appreciation of what a "good conceptual model" might be (Blandin, 2003).

It can also be argued that the same tools are used differently within different communities (Brown, Collins, & Duguid, 1989), which reinforces the need to take into account social and cultural aspects belonging to the world of the user.

Sociology of uses, which now relies on a large "corpus" of surveys published on both sides of the Atlantic (Martin, 1997), points out the fact that "usefulness" or "utility" is a precondition to use a tool. Research made to understand why tools or facilities, such as video recorder, a personal computer, or pay-TV, are not used lead to this evidence: "no need" is the main reason given to explain rejection of the tool (Punie, 1997). A recent two-year survey (Rouet, 2003) on the development of uses related to the Internet among a sample population representative of a French provincial area showed that the uses developed after two years are based on the applications that were considered useful at the beginning (e-mail, browser, search engine, etc.), and those not considered useful (chat, Webcam) remain unused. Utility does not spontaneously evolve.

This means that the first criterion of usability evaluation of an online learning program is that it has to be considered as useful by the user. Again, this is context dependent. As a result, the use of online learning programs appears dependent on what I called "social schemata of uses" (Blandin, 2002), which determine and format conditions for using tools as well as the ways the tools are used.

MAIN THRUST OF THE ARTICLE

For these reasons, factors describing the environment of uses need to be encompassed in any usability assessment. To take this into account, Blandin (2003) proposed the notion of "situated usability" to name a set of heuristics that account for describing how the environment impacts usability factors. These heuristics, as stated in Blandin (2003) are as follows:

1. **Social Schemata of Uses:** to be used within a given community, a tool should embed common Social Schemata of Uses of this community;
2. **Type of Action:** to be used in a given situation, a tool should correspond to user's need and purposes in this situation, and allow him/her to perform a given action;
3. **Culture of the Users:** to be used within a given community, a tool should convey representations and practices which are considered as "common sense knowledge" by the user;
4. **Culture of the Environment:** to be used within a given community, a tool should convey representations and practices which are considered as "common sense knowledge" by the community;
5. **Tool Efficiency in a Given Situation:** to be used in a given situation, a tool should have proved efficiency in such a situation;
6. **Ability of the User to Use the Tool:** to be used in a given situation, a tool must be mastered, to some extent, by the user; which also means that an object does not become a tool at hand;
7. **Motivation of the User to Use the Tool:** to be used in a given situation, a tool should interest enough the user to perform his/her action using this tool rather than in any other manner (pp. 319–320).

These heuristics are general and may apply to any type of tools. They could be refined in the case of online learning programs by identifying the "learning culture" of the user (Brown, Collins, & Duguid, 1989).

Following the principle used by Lakoff and Johnson (1999), four families of "learning cultures" can be considered, based on four different "learning metaphors," or four different "learning paradigms," built upon different learning theories throughout history (Table 1). These learning cultures still correspond to different communities of thinking within the educational world.

Some tools originated in, or better fit, with a particular learning culture. For example, mail and other written communication tools fit better with the Socratic paradigm, because they allow discussion; computer-based training programs originated from the Plato system, based on the work of renowned behaviourists, such B. F. Skinner (Dooijes, n.d.).

A typology of "formal learning systems," as presented in Blandin (2000) or in TTNET Dossier Number 4 (2001), will help to classify online learning programs according to the learning cultures. It is presented in Table 2.

It can be noticed that these four learning cultures do not allow for the implementation of all the "situated usability" heuristics presented above in the same manner, and some of the tools corresponding to Platonism, objectivism, and behaviourism might not always fulfill the requirements stated by heuristics (2), (3), (6), and (7).

One of the reasons why online learning programs are not used is because there is a "trench" (Brown, Collins, & Duguid, 1989) between the learning culture that they implicitly embed and the learning culture of the learner, or of the learning system in which they are inserted. This trench generates learning situations that are ill-formed, and thus inefficient, or that are rejected by one of the actors, either the teacher or the learner.

From current research, it appears that the learning culture might influence the use of online learning programs in the following ways (Blandin, 2003):

- There could be a mismatch between a teacher's learning culture and off-the-shelf programs, which

Table 1. Four learning cultures

The Learning Paradigm	The metaphor underpinning the learning culture
Platonism (Socratic)	Learning is remembering
Objectivism (Aristotelian)	Learning is recording
Behaviourism	Learning is training
Constructivism	Learning is building

Table 2. Four learning cultures and online learning programs (Source: Blandin 2003).

The Learning Paradigm	Type of online learning programs fitting with the culture
Platonism (Socratic)	Programs based on collaborative tools or communication tools (virtual communities, virtual campuses, etc.)
Objectivism (Aristotelian)	Poorly interactive and broadcast programs (information resources, Webcast courses and conferences), virtual classroom (audio or video real-time instruction)
Behaviourism	Interactive learning systems (computer-based training, Web-based training, computer-assisted instruction, etc.)
Constructivism	Interactive learning systems (simulators, virtual labs, micro-worlds, etc.), problem-based learning systems

leads to "no-use recommendations."

- Most current programs provided by online learning providers are of Aristotelian or Behaviourist style and, therefore, will not be used or recommended by teachers implementing the Constructivist paradigm.
- The use of online learning programs might be facilitated when the conditions to develop and promote self-directed learning are fulfilled. This supposes both to use the Constructivist paradigm and to implement the "seven pillars of self-directed learning" as described by Carré (2002, 2003).

From these findings, more specific usability criteria related to the use of online learning programs can be derived, such as the following (Blandin, 2003):

- Adequacy between the learning paradigm embedded within the learning program and the learning paradigm in use within the school or the university
- Adequacy between the content of the learning program and the course curriculum in which it is supposed to be embedded
- Adequacy between the content of the learning program and the way this content is taught by the teacher in charge of the course
- Adequacy between the learning paradigm embedded within the online learning program and the learning paradigm in which tutoring support is rooted
- Adequacy between the learning paradigm embedded within the online learning program and efficient learning strategies for such content for the learner
- Adequacy between the use of a given type of learning material, the learner's motivation, and his or her self-directed learning readiness
- Adequacy between the learner's ranking in the self-directed learning readiness scale (Guglielmino, 1977) and the kind of support provided
- Adequacy between the learning environment culture and supportiveness and the expected motivation of learners

These criteria correlate with the findings of recent research and can explain their results (Masie, 2001; Paxton, 2000; Quinn, 2001).

FUTURE TRENDS

From a broader viewpoint, this topic is a subpart of sociology of uses. Research in sociology of uses traditionally focused on uses "as they were" at a given time among a given population.

A more recent trend is illustrated by the "*100 fenêtres sur Internet*" survey in France: 100 people were given a computer with cheap Internet access over 2 years and, in counterpart, accepted to be questioned and observed four times, at the beginning and then every 6 months (Rouet, 2003). This research looked upon the evolution of uses over a 2-year period: uses are no longer considered "static."

But this is not enough. If we consider that uses have their own dynamics, it is important to start research on how uses are built and how they emerge. This will allow us to understand how a technology is accepted and how its "usability" is created.

Because this topic is a particular application domain, the processes that might result from the research described above will have to be specified to learning software and to educational contexts. Further research on learning cultures and their impact on the use of learning tools is needed in order to better understand such phenomena and to produce more criteria that will facilitate the implementation of "e-learning motivators" adapted to a given context.

CONCLUSION

The epistemological stance of the teacher or of the trainer; the motivation of learners; the learning culture of the school, of the university, or of the company; and context-related factors interact together to produce the conditions that will determine how a given type of learning material and, in particular, online learning programs, are used in the school, the university, or the company.

This is why usability has to be considered as "situated," and why a sociological standpoint is required to understand how a user uses a tool in a given context, and what factors are involved in influencing the relationship between the user, the tool, and the community to which they belong. Because the use of a tool is deeply rooted in users' cultures, and may be learned, usability has to take into account the social and material environments of the users.

REFERENCES

Blandin, B. (2000). Open and distance learning within the world of vocational training and lifelong learning, Part 1: Open and distance learning, an overall survey at the beginning of 2000. In A. R. Trindade (Ed.), *New learning* (pp. 104–141). Lisbon, Portugal: Universidade Aberta.

Blandin, B. (2002). *La construction du social par les objets*. Paris, France: Presses Universitaires de France.

Blandin, B. (2003). Usability evaluation of online learning programmes. A sociological standpoint. In C. Ghaoui (Ed.), *Usability evaluation of online learning programmes* (pp. 313–330). Hershey, PA: Idea Group Publishing.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32–42.

Carré, P. (2002). Accompagner l'autoformation en APP. In *Actualité de la formation permanente n°176* (pp. 100–106). Paris, France: Centre Inffo.

Carré, P. (2003). L'autoformation accompagnée en APP ou les 7 piliers revisités. In P. Carré & M. Tetart (Ed.), *Les ateliers de pédagogie personnalisée ou l'autoformation accompagnée en actes* (pp. 125–148). Paris, France: L'Harmattan.

Dooijes, E. H. (no date). *The Plato IV system for computer aided instruction*. Retrieved August 25, 2001, from *http://www.science.uva.nl/faculteit/museum/PLATO.html*

Guglielmino, L. (1977). *Development of the self-directed learning readiness scale*. Doctoral dissertation, University of Georgia, Athens.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh. The embodied mind and its challenge to Western thought*. New York: Basic Books.

Martin, M. (1997). Variations sur le thème "usages." In *Proceedings of the First International Conference* "*Imagining* Uses" (pp. 177-190), Bordeaux, France, May 27-29. Pessac, France: ADERA.

Masie, E. (2001). *E-learning: If we build it, will they come?* Alexandria, VA: American Society for Training and Development.

Nielsen, J. (1994). *Usability engineering*. San Francisco, CA: Morgan Kaufmann Publishers.

Nielsen, J. (1999). *Designing Web usability: The practice of simplicity*. Indianapolis, IN: New Riders Publishing.

Nielsen, J. (no date). *Ten usability heuristics*. Retrieved August 25, 2001, from *http://www.useit.com/papers/heuristic/heuristic_list.html*

Norman, D. (1988). *The psychology of everyday things*. New York: Basic Books.

Notess, M. (2001). Usability, user experience, and learner experience. *E-Learn Magazine In-Depth Tutorials*. Retrieved August 25, 2001, from *http: // www. elearnmag org/subpage/sub_page.cfm?s ection= 4&list_ item=2 &page=1*

Paxton, S. L. (2000). *When success matters: How to make e-learning work?* Boulder, CO: Frontline-Group. Retrieved August 25, 2001, from http://www.frontline-group.com/e_news_brief/wpaper.pdf

Punie, Y. (1997). Imagining "non-uses." Rejections of ICTs in Flemish households. In *Proceedings of the First International Conference* "*Imagining Uses*" (pp. 165-176), Bordeaux, France, May 27-29. Pessac, France: ADERA.

Quinn, A. (2001). *Why people can't use e-learning. What the e-learning sector needs to learn about usability*. Retrieved August 25, 2001, from *http://infocentre.frontend. com/servlet/Infocentre?access=no&page=article&rows=5&id=163*

Rouet, J. -F. (coord.). (2003). *Cent fenêtres sur Internet. Rapport de fin de contrat*. CNRS–Université de Poitiers. Retrieved April 5, 2004, from *http://www.mshs.univ-poitiers.fr/laco/Pages_perso/Rouet/Textes.htm*

TTnet Dossier No. 4. (2001). *Open and distance learning and the professionalisation of trainers*. Thessaloniki, Greece: CEDEFOP.

KEY TERMS

Formal Learning System: This is a situation that is specially designed to facilitate learning. This situation takes place at a given time in a particular physical environment equipped with dedicated furniture and technical facilities.

Learning Culture: This is a particular way of considering the learning process, how it occurs, in what conditions, and so forth. It relies on an explicit or an implicit theory that provides more or less accurate representations of the learning process.

Learning Metaphor: This is an image representing the way the brain is supposed to process information to produce knowledge. This image is built according to the implicit theory in use or to the explicit theory espoused.

Learning Style: This is the preferred way a person collects, decodes, and processes information to produce knowledge.

Online Learning Program: This is a piece of software, running on an electronic network, that was explicitly designed to support a learning situation.

Sociology of Uses: This is a domain of Sociology that studies how a given population uses technical objects. Initiated by telecom operators in the second part of the 20^{th} century, the field of research has been enlarged to encompass the use of any type of technical object.

Usability: This is the capacity of an object to be used within a given course of action. It includes the "ease of use" of the object for the action and its "usefulness" or "utility" for the user to achieve the action.

Uses: These are the different ways a tool is used. Uses are highly cultural and depend upon the context of use and the culture of the social surroundings.

Utility: This measures the degree of usefulness of an object. Classical economical theories consider that the utility of an object and the available quantity of this object determine its price.

Usability of Online Learning Systems and Course Materials

Elizabeth Furtado
Universidade de Fortaleza, Brazil

INTRODUCTION

Human-computer interaction (HCI) aims to design and develop high-usability interactive systems (ISs) focusing on users' needs and ergonomic principles, among others. The usability of an IS refers to how easy it is to use and to learn. Similarly, software engineering (SE) aims to design and develop high-quality ISs focusing on schedule, budget, communication, and productivity. The quality of an IS refers to how satisfied the system clients and/or users are, verifying whether the system is performing exactly what was requested.

In order to achieve both IS usability and quality, it is necessary to integrate HCI concepts into an IS development method. HCI concepts can be characteristics of users (such as their preferences, language, culture, and system experience) and of their context of use (such as easy accessibility and good luminosity of the environment). In the online learning context, it is necessary to integrate HCI concepts into an online learning system development method. The pedagogic usability of an online learning system is related to how easy and effective it is for a student to learn something using the system. For these reasons, it is important not only to think about the IS quality, but about its usability as well. In this text, an online learning system on the Web is composed of a virtual learning environment (VLE), with tools to support a collaborative learning and online course materials available for the users through this environment. So, it is important not only to think about the VLE usability, but also about the online course material usability.

We have identified some problems to achieve a successful deployment of online learning systems (Furtado, Mattos, Furtado & Vanderdonckt, 2003):

- *Lack of learning quality:* Many academic staffs are not worried about the design of online course materials. The material of a face-to-face course is hardly ever adapted to online course material. Whenever a course is to be published on the Internet, it is important to envision the virtual course as the software. This way, it is expected that a course is to be developed with the same severity as the software is planned.
- *Lack of adaptive tools and guidelines:* Learning systems are very useful, but most of them are not adaptive and neither is the user model predefined (Gomes & Viccari, 1999). In addition, user interfaces of such systems are generally specified without taking into account guidelines (Eleuterio & Eberspacher, 1999).
- *Lack of training in collaborative technologies and methods:* Any academic staff (such as a teacher), as part of his/her professional development, needs continuous training. Such training is often carried out without using technologies that can deal with adaptive and collaborative processes. It is necessary to fulfill these needs by adopting an integrated pedagogical-technological content (Perrenoud, 2001).

All of these issues have a critical impact on the usability and quality of online learning systems. Thus, we developed a general architecture for such systems, which aims to show the concepts that must be considered to increase the quality of the learning process and to increase their user interface (UI) usability.

The remainder of this article is structured as follows: in the next section, we explain the main concepts that helped us to develop such general architecture. Then, we provide the best practices used in a development cycle of an IS, focusing on the usability issue. Finally, we summarize the main points of this text.

VLE AND ONLINE COURSE MATERIAL BACKGROUND

As we have mentioned before, an online learning system is composed of a VLE and online course instructional materials.

A VLE has to provide students with spatial freedom and time flexibility. It has to be flexible enough so that every student may profit from his/her own skills and abilities, use his/her previously developed idiosyncratic characteristics (cognitive, social, or emotional), and apply his/her previously gained experience and expertise

(Karoulis & Pombortsis, 2003). Some tools available in a VLE are the following: links to tutorials and course materials, collaborative tools (as discussion forums, chats), evaluation tools, and administrative tools.

The main focus in instructional material is on: content, exercises and solutions, and project and lecture notes. The online course material needs ad hoc preparation: target and expected results must be stated, keywords must be provided, and a review must be present at the beginning and at the end of each chapter. Some authoring tools allow teachers to develop their own instructional materials. Other tools, such as those for specific programming languages (HTML, FLASH, SVG), are only used by specialized teams.

BASIC CONCEPTS RELATED TO USABILITY IN ONLINE LEARNING SYSTEMS

The general architecture proposed here (see Figure 1) aims at the development of VLE and online course instructional materials, taking into account some concepts studied in different areas (human-computer interaction, cognitive sciences, ergonomic, artificial intelligence, and pedagogy).

According to this figure, an online learning system's usability can be assured when its components have been built with quality and when users' needs have been taken into account. Quality of a component means: (i) quality in the application corresponds to content, which refers to the information and knowledge involved in the system. Information (such as learning stories and objects) are related to the development of instructional materials, and knowledge (such as cases) are especially related to the collaborative practice in forums, for instance; (ii) usability in the UI, which refers to a good specification of the interactive information of the system (its windows, its buttons, etc.); and (iii) usability through interaction devices, which makes the interaction with different media (sound, text, image) possible through devices as cameras, microphones, and so on. The quality of the user refers to his/her ability to use new interaction devices and technologies, experience with computer-based systems, and acquaintance of the domain in question.

The concepts related to usability in an online learning system are the following:

- Utilization of ontology to assure the flexibility in modeling learning applications. The ontology notion comes from the artificial intelligence area where it is identified as the set of formal terms with one knowledge representation, since the representation completely determines what "exists" in the system (Guarino, 1995). During an application modeling, models (such as the user model), knowledge (such as cases studies), and learning stories and their learning objects associated to an instructional material can be represented using ontology. The advantage of using this representation is that the ontology can be defined once and used as many times as necessary (evolutionary approach). In addition, the ontology is useful to create learning objects and reuse them when a new course is initialized.
- Utilization of guidelines and human factors to assure learnability (effective interaction and maximum performance) and flexibility (multiplicity of ways the user and the system exchange information). Human factors, such as the teachers' beliefs, and guidelines related to graphic aspects and characteristics of the users and their context of use, must be considered. Guidelines are suggestions about the ergonomic aspects of the interfaces, such as showing only the necessary information or letting the user control the system dialog (Bastien & Scapin,

Figure 1. General architecture proposed

<table>
<tr><td rowspan="2">Quality of the User

(eg. Allowing the user to collaborate and communicate with other users)</td><td>Usability through Interaction Devices
Video and microphone</td><td>Usability of the User Interface
Guidelines
Multimedia
Adaptive User Interaction</td><td>Quality of the Learning Application
Learning stories and objects
Ontology
Cases studies
User model</td></tr>
<tr><td colspan="3">Usability of the VLE</td></tr>
<tr><td colspan="4">Usability of the Overall On-line Learning System</td></tr>
</table>

1993; Bodart & Vanderdonckt, 1993). Taking into account guidelines during the interface design of a system allows the designer to determine the best way in which the information is to be provided to users and to ensure optimal accessibility of the system.

- Utilization of multimedia resources to improve interaction. The use of different interaction devices (such as microphone and video camera) makes it possible to provide a multimedia environment. Specify a multimedia-supporting system derives from the worldwide desire for non-textual information. Non-textual information characterizes systems that use different channels, and codes in their interactions with the user.
- Utilization of collaboration and communication mechanisms to assure the quality of the user and the continuous usability of the learning system. Some VLEs implement the collaborative aspect by allowing users to share an application (such as TELE (Neto, 2001)) and/or by using forums and chats. It allows users to share knowledge, when they are motivated to collaborate through principles of participation, for instance, in problem definition practical situations. There is a need to continuously assure usability of a system due to technological changes and to the evolution of users' needs. We believe instructional materials of a course should evolve from human interactions between students and teachers occurring within an online discussion forum (Mattos, Maia & Furtado, 2003). Hence, we believe in a collaborative process between users and designers to adjust accessibility, acceptability, and usability criteria of a system.

USABILITY OF ONLINE LEARNING SYSTEMS: THE BEST PRACTICES OF REQUIREMENTS ENGINEERING

In this section, the concept of usability (learnability, flexibility, and robustness) is related to some best practices of the requirements modeling and validation of an IS (see Table 1).

An IS must be developed with the participation of users throughout the development process, because it is easier for developers to define and evaluate the functional and non-functional (usability) requirements. The functional requirements of a VLE are related to the tasks that the user wants to perform (user tasks), for instance, to interact with other students and tutors, and to access the rules and regulations of the course. Usability requirements are related to users' satisfaction and the performance of the system. These requirements directly influence aspects of the system quality of use (e.g., never lose sight of navigational functions).

Sutcliffe (2002) gives practical guidance for requirements modeling and validation based on scenarios. A scenario represents a story or example of events taken from real-world experience. These stories are close to the common sense use of the word and may include details of the context of use for a system. These representations help users think about how the system would support their tasks.

During the material design, students must be able to inform their learning requirements, which are related to content (students must perform the study tasks for a particular unit). These requirements are usually represented in storyboard sketch or animated sequences. A storyboard represents a future vision of a designed material with sequences of behavior and possibly contextual description. These representations help teachers focus on the pedagogical functions of the material.

However, a single scenario or storyboard shows only some possible sequences of events among many possible sequences permitted during the interaction of a user with an IS. Interactive prototype is an interactive medium that allows the users to explore all alternative paths, and it gives a look and feel overview of the IS. Before doing prototypes, Constantine, Windls, Nolbe, and Lockwood (2003) suggest completion of an abstract content model, or an abstract prototype, because it facilitates creative thinking, leading to more innovative solutions instead of thinking over the real interfaces.

During a session of requirements validation, it is necessary to actively engage users in checking that the designed IS actually does what they want. If the IS developed is accepted by the user, we can say that its development process was user centered, according to their needs and suggestions. As we mentioned, the usability of an online learning system is not just the UI. The focus must also be on the support that the VLE or instructional material provides for its online students.

As requirements are so volatile, it is important to examine how they can be specified so software can evolve and adapt to the users´ needs. The evolutionary process involves continuous requirement adjustments in two points of view: (i) of the user—when using the system, his/her practices and working methods can be adapted to satisfy the evolving needs individually and/or collectively; and (ii) of the system—the system's behavior can be adapted. The evolution of users' needs can involve modifying a system's characteristics related to its design options, for instance. In order for an IS to be considered evolutionary, its quality must be continuously verified. In VLEs, this means that its functionalities must be changed (for instance, to realize that students would like to publish their work to be viewed by any of their tutors other than themselves). In course materials, the changes can require a definition of a new learning object from an existing one.

To develop an adaptive IS, the UI designer must consider guidelines that must conform to usability requirements defined previously and users' characteristics and their context of use. It is usual to gather users sharing the same value for a given set of characteristics into stereotype. The problem is there is no predefined information on the users to ensure that an IS has high quality of interaction to a stereotype of all users (Furtado et al., 2001).

Ontology can be used to represent a variety of parameters that are not necessarily identified nor truly considered in the requirements analysis. The notion of ontology allows the definition of the meta-models, which define the specification language with which any model can be specified. This resource makes it easier to consider new information in models (Vanderdonckt et al., 2004). In online adaptive learning systems, a flexible user modeling approach is very important. The ontology of a user model can be updated accordingly to consider more information.

So the success of an IS will depend on a complex trade-off between the classic view of requirements being satisfied by a design, but evolving, and the desired degree of satisfaction in acquiring the desired product (Sutcliffe, 2002).

These practices and the HCI concepts described here are more detailed in the description of the lifecycle called CONE (Furtado & Sousa, 2003). It provides the developers and users a new spatial organization, including the activities that must be performed according to these practices and using these HCI concepts and artifacts. This organization aims to facilitate the participation of users, clients, and development team members in monitoring any iterative and incremental process.

FUTURE TRENDS

The multidisciplinary dimension of this work is characterized by the studies done in diverse areas of knowledge, from human-computer interaction and software engineering to pedagogy. Distance learning adheres to a vision where online learning applications are developed for the widest population of users in the most varied contexts of use by taking into account individual differences. This population of users is being called "online community". To create an online community-centered design will be the best practice of the requirements engineering for developing the applications, which will be able to better support the collaborative learning. In addition, it will be possible to develop applications with adaptive user interfaces. These interfaces must be able to *adapt* themselves to the community and contexts of use characteristics.

CONCLUSION

From our experience with the community of students and teachers in the online learning systems and with the development of interactive systems, it was possible to establish the following conclusion about how to obtain more usable ISs: the main question for the success of learning processes for users through the Internet does not lie exclusively on the choice of pedagogical methodology and techniques. It lies, fundamentally, on three factors: i) the understanding of the needs of both teachers and students though a participatory design, ii) the transformation of such needs in a consistent UI and adaptive functions of a VLE and its available course materials; and iii) the need to continuously assure usability of a system through an extensible representation of requirements.

Table 1. Summary of requirements engineering practices to have usability

Usability of VLEs	**Usability of Course Instructional Materials**
Participatory design of users	Participatory design of students and teachers
User-centered design	Student-centered design
Focuses on the UI and pedagogical functions of the VLE	Focuses on the UI and pedagogical content of the material
Requirements are represented in scenarios and prototypes	Requirements are represented in storyboards and prototypes
Evolutionary and adaptive VLE	Evolutionary and adaptive material

U

REFERENCES

Bastien, J.M.C., & Scapin, D.L. (1993). Ergonomic criteria for the evaluation of user interfaces. *INRIA, 156.*

Bodart, F., & Vanderdonckt, J. (1993). Expressing guidelines into an ergonomical style-guide for highly interactive applications. In S. Ashlundn, K. Mullet, A. Henderson, E.L. Hollnagel & T. White (Eds.), *Proceedings of InterCHI'93* (pp. 35-36).

Constantine, L., Windls, H., Nolbe, J., & Lockwood, L. (2003). *From abstraction to realization in user interface designs: Abstract prototype based on canonical abstract components.* Working Paper on Tutorial Usage-Centered Software Engineering. ICSE'03, Portland, Oregon.

Eleuterio, M., & Eberspacher, H. (1999). A knowledge management approach to virtual learning environments. *Proceedings of the International Workshop on Virtual Education* (WISE 1999) (pp. 55-61). Fortaleza. Demócrito Rocha Editora.

Furtado, E., Mattos, F.L., Furtado, J.J.V., & Vanderdonckt, J. (2003). Improving usability of an online learning system. In C. Ghaoui (Ed.), *Usability evaluation of online learning programs* (pp. 69-86).

Furtado, E., Furtado, V., Bezerra, W., William, D., Taddeo, L., Limbourg, Q., & Vanderdonckt, J. (2001). An ontology-based method for universal design of user interfaces. *Proceedings of the Workshop on Multiple User Interfaces over the Internet: Engineering and Applications Trends*, France. Retrieved July, 10, 2001, from *www.cs.concordia.ca/%7Efaculty/seffah/ihm2001/program.html.*

Furtado, E., & Sousa, K. (2003). A user interface generation process based on the RUP and represented in a new spatial organization. Retrieved March 1, 2004, from ead.unifor.br

Gomes, F., & Viccari, R. (1999). Uso de heurísticas no projeto de interfaces inteligentes. *Proceedings of the International Workshop on Virtual Education* (WISE 1999) (pp. 103-109). Fortaleza. Demócrito Rocha Editora.

Guarino, N. (1995). Formal ontology, conceptual analysis and knowledge representation: The role of formal ontology in information technology. *International Journal of Human-Computer Studies, 43*(5-6), 623-640.

Karoulis, A. & Pombortsis, A. (2003). Heuristic evaluation of Web-based ODL programs. In C. Ghaoui (Ed.), *Usability evaluation of online learning programs* (pp. 89-109).

Mattos, F.L., Maia, M. &, Furtado, E.S. (2003). Formação docente em processos colaborativos online: Em direção a novos "círculos de cultura"? *Proceedings of the Workshop em Informática na Educação* (WIE).

Neto, H., Raimir H., Bezerra W., & Sarquis O. (2000). Especificando o tele-ambiente no contexto da educação a distância. *Proceedings of the Simpósio Brasileiro de Informática Educativa* (SBIE'2000) (pp. 120-132). Alagoas. Universidade Federal de Alagoas Editora.

Perrenoud, P. (2001). *Formando professores profissionais: Quais estratégias? Quais competências?* Porto Alegre: Artmed.

Sutcliffe, A. (2002). *User-centred requirements engineering. Theory and practice.* London: Springer-Verlag.

Vanderdonckt, J., Furtado, E., Furtado, V., Limbourg, Q., Bezerra, W., William, D., & Taddeo, L.. (2004). *Multi-model and multi-layer development of user interfaces in multiple user interfaces*. London: Ahmed Seffah and Homa Javahery, John Wiley & Sons.

KEY TERMS

CONE: A new life cycle in which development process occurs in iteration cycles, each one having many activities grouped together in phases.

Evolutionary System: Involves continuous adjustments of its functionalities and UI according to the user and/or technological changes.

Extensible Representations of Requirements: Ways to represent easy requirements that were not necessarily identified nor truly considered in the requirements analysis.

Pedagogic Usability of an Online Learning System: Related to how easy and effective it is for a student to learn something using the system.

Requirements Engineering: The human acts of identifying and understanding what people want from an IS.

Usability of an IS: Refers to how easy it is to use and learn the system.

Usability Requirements: Related to users' satisfaction and the performance of the system.

Usable M-Commerce Systems

John Krogstie
NTNU and SINTEF, Norway

INTRODUCTION

Today, the PC is only one of many ways to access information resources. On one hand, traditional computing technology is becoming more mobile and ubiquitous and, on the other hand, traditional mass media are becoming richer. Whereas information services related to interactive TV and ubiquitous computing are projected to become prominent in a few years, mobile computing is the most important current trend within information and communication technology.

According to Siau, Lim and Shen (2001), the essence of m-commerce and mobile information systems is to reach customers, suppliers and employees regardless of where they are located and to deliver the right information to the right person(s) at the right time. So far, relatively little attention has been paid to the user side of m-commerce systems. The ability to develop and evolve usable m-commerce systems may become an even more critical success factor for enterprises in the next few years than is their ability to develop and evolve usable e-commerce systems today.

BACKGROUND

M-commerce systems differ from more traditional information systems along several dimensions (Hirsch, Coratella, Felder, & Rodriguez, 2001; Krogstie, Brandtzæg, Heim, & Opdahl, 2003b; Siau et al., 2001). We have grouped the differences within the following three areas:

- User-orientation and personalization;
- Technological aspects including convergence and need for multi-channel support; and
- Methodology for development, evolution and operations to ensure organizational returns.

User-Orientation and Personalization

Mobile information systems often address a wide user-group, which means that user-interface aspects should feature prominently and early in the design process and often need to be very simple. Input and output facilities may be severely restricted (no keyboard, small screen-size, etc.) or based on new modalities (speech-recognition and synthesis, etc.). This means that individualization of mobile information systems becomes increasingly important, both at the individual level where user-interface details are tailored to personal preferences and hardware, and at the work-group level, where functions are tailored to fit the user's preferred work processes.

Individualization means information systems that both automatically adapt themselves to the preferences of the user and that can be explicitly tailored by users through a specific user-interface. The main goal is to achieve usability of the applications on all possible interfaces, based on adaptation to the different physical devices. This calls for intelligent, adaptive and self-configuring services that enable automatic context-sensitivity, user profiling and personalization.

Technological Aspects including Convergence and Multi-Channel Support

Mobile devices have limited processing, memory and communication capacities compared to other kinds of computers. Performance considerations, therefore, become increasingly important. Analytically-based predictive methods are necessary in order to assess a large number of alternatives during the design. Mobile information systems also pose new challenges to achieving information systems dependability. The new mobile devices provide integration and convergence of technologies into a wide range of innovative mobile and multimodal applications. Mobile and other new technologies provide many different ways to offer the same or similar services to customers. Novel approaches for the development and evolution of applications on and across different mobile and traditional platforms are thus needed.

Methodology for Development and Operations to Ensure Organizational Return

Mobile information systems are often radical and, therefore, reward an increased focus on idea generation early in the requirements and design process. Understanding the mobile users requirements for new services is thus of large importance. One would need both to be able to develop these systems and to address the major hurdles for the deployment of applications and services. Another effect of the radically new approaches is that the introduction of mobile information systems often spawns several other initiatives for changing other information systems within an organization. It is important to focus on the interoperability of services and seamless access to corporate and government resources from the mobile devices.

STATUS FOR MODEL-DRIVEN DEVELOPMENT AND EVOLUTION OF USABLE M-COMMERCE SYSTEMS

When speaking about *model*-driven system development, we refer to models developed in languages that have the following characteristics:

- The languages are diagrammatic, with a limited vocabulary (states, classes, processes, etc).
- The languages utilize powerful abstraction mechanisms.
- The languages have a formal syntax and semantics. The formal semantics is either operational enabling, e.g., generation of other models including executable programs or mathematical enabling advanced analyses.
- The languages are meant to have general applicability across problem domains.

Although most software developers are aware of model-driven methodologies, they are seldom followed in great detail in practice, and mostly only in initial development stages.

In general, a model-driven approach to information systems development has been found to provide the following advantages (Krogstie & Sølvberg, 2003b):

- Explicit representation of goals, organizations and roles, people and skills, processes and systems;
- An efficient vehicle for communication and analysis;
- Basis for design and implementation; and
- Readily available documentation as a basis for extensions and personalization.

One striking aspect in connection to contemporary information systems development and evolution is that there is an increasing demand for shorter development time for new products and services (Pries-Heie & Baskerville, 2001). This is specifically important for m-commerce systems, where the convergence of different platforms continuously creates opportunities for new functionality. Some would argue that this highly dynamic situation would make model-based approaches impractical. To the contrary, we claim that idea generation should not be limited by currently available technologies and that systems must be developed for change.

We can identify the following areas for potentially increased utility of techniques earlier developed as part of model-driven development:

User-Orientation and Personalization

Traditionally, support for workers performing nomadic processes has not been provided. Functions of the mobile information system should be tailored to fit the user's preferred work processes, which typically involve other persons. To support teamwork, raising awareness of the status of knowledge resources is increasingly important in a mobile setting. To enhance social mobility, organizations and industries need to develop "social ontologies", which define the significance of social roles, associated behaviors and context (Lyytinen & Yoo, 2002). Given that knowledge resources include both individuals and technology that can be mobile, one should look into interactive systems to improve group performance. Peter Wegner's interaction framework (Wegner, 1997) was triggered by the realization that machines that must interact with users in the problem solving process can solve a larger class of problems than algorithmic systems computing in isolation. The main characteristic of an interaction machine is that it can pose questions to human actors (users) during its computation. The problem solving process is no longer just a user providing input to the machine, which then processes the request and provides an answer (output), but rather is a multi-step conversation between the user and the machine, each being able to take initiative. A major research question in this area is how to specify and utilize interaction machines on a multi-channel platform. Process support technologies are a natural choice for enabling interaction machines. Such technologies are typically based on process models, which need to be available in some form for people to alter them to support their emerging goals. Thus, interactive models should be supported (Jørgensen, 2001). The outset for this thinking is that models can be useful tools in a usage situation, even if the models are changing and incomplete. The user is included as an interpreter and changer of the models. Emergent workflow systems (Colombo, Francalanci, Mecella, Pernici, & Plebani, 2002; Jørgensen, 2001) represent a different approach to static and adaptive workflow systems with respect to their use of models. They target very different kinds of processes - unique, knowledge-intensive processes where the structure emerges or processes with strong requirements for adaptation to the context during process execution

Technological Aspects Including Convergence and Multi-Channel Support

There is a multitude of competing technologies available for providing the underlying infrastructure and access devices for distributed and mobile applications. A central element when addressing this is the development of model-based specification techniques that are powerful enough to be used as a basis for the development of systems on a large number of technical platforms, but still general enough to represent the commonalties at one place only. A major initiative within Object Management Group (OMG) is on model-driven architectures (MDAs) where both platform independent and platform specific modeling notations including refinement techniques and mappings between platform independent and platform specific notations are specified. Meta-modeling techniques and domain-specific modeling (DSM) have found a special application for the design of mobile phone software (Kelly & Tolvanen, 2001). It could be argued that mobile information systems are a particularly good climate for using domain-specific modeling:

- The software (on the client side) is partly embedded and thus needs a higher reliability than traditional software and can be supported by restricting choices through the addition of modeling rules and code-generation.
- You need many very similar variants of the same application.
- There are a number of standards to adhere to, and the technology and standards change rapidly.

Model-based development can be used to support dependability analyses, that is, the use of methods, techniques and tools for improving and estimating dependability, such as risk analyses, probabilistic safety assessment, testing, formal walkthrough, simulation, animation, exhaustive exploration, and formal verification.

The new separation between content and medium found in mobile information systems serves as a major challenge. Generally, the context should be explicitly modeled to maintain an overview of, analyze, and simulate the multitude of possibilities open for adapting to the context and the use of context traces. For the more general use of mobile applications, it is also important to be able to adapt these systems to the user at hand, thus making a case for simple user-models to guide the adaptation. Banavar and Bernstein (2002) highlight the importance of semantic modeling in this respect.

Recently, work within user interface modeling has focused increasingly on mobile user interfaces (Eisenstein, Vanderdonckt, & Puerta, 2001). This is often done to facilitate some level of common models for both the mobile and more traditional user interfaces. A central element in this is the development of model-based approaches that are powerful enough to form the basis for the development of user-interfaces on the multitude of platforms needed, but still general enough to represent the commonalities in a single place. One approach is to define user-interface patterns with general usability principles as powerful building blocks (Nilsson, 2002).

Methodology for Development to Ensure Organizational Return

Siau et al. (2001) highlight as an important application-oriented research area the development of m-commerce business models. In order for m-commerce to succeed, it is vital to ensure that all of the related applications and services can be accessed with ease and at little cost. Thus, in addition to externalizing the business models in the manner of independent computing, it is important to integrate these models with the internal enterprise models and enterprise architecture in order to pinpoint the links to, for example, internal systems for the efficient billing of services provided

Mobile information systems are often radical and, therefore, reward an increased focus on idea generation early on in the development phase. This also means that services or situations in which to anchor problem analysis efforts such as using as-is analysis as a starting point for to-be design (Rolland & Prakash, 2000) do not always exist. Technology in the field still develops rapidly, which means that idea generation should not be limited by currently available technologies. One needs to enhance the techniques for modeling of scenarios to take this into account. Applications of the new technology call for highly distributed systems that comprise new user-interface systems on the client side, new and existing back-end systems, as well as new bridging systems (which port information between other systems). The new technologies, therefore, highlight the need for principled, long-term IS-architecture management and for integrating architecture management with software development methodologies. Often, there is a need to interface with existing enterprise systems and architectures to enable the new workflow. Another aspect is how to integrate the user-interface models discussed previously with other parts of the requirements and design model, for instance, the entity model, process model and goal model. On both the process and the user-interface side, the challenges can be attacked by extending existing approaches to modeling, although research is needed to investigate the techniques that should be extended and how they could be best adapted to the new problem areas. Looking at the architecture for the general existing mobile information systems framework and solutions (e.g., Celesta, 2001), we notice that the generic architecture is geared towards the modeling of data, business processes and tasks, events and behavior, rules, user interfaces and general context information

FUTURE TRENDS

The large-scale application of m-commerce is in its infancy and, unsurprisingly, limited work has so far been done on usable m-commerce systems or on model-based development and evolution of usable m-commerce systems. On the other hand, the upcoming, 3G- (UMTS-) infrastructure provides higher bandwidth and constant connection to the network from virtually everywhere, and the number of m-commerce applications is therefore predicted to explode as will the number of users of complex applications on a mobile platform. Thus, the need for evolving best practice within systems development will be increasingly important.

CONCLUSION

In this article, we have highlighted the need for adapting and extending modeling-based approaches to ensure usable m-commerce systems for the future. Obviously, we are not starting this work from scratch, it is possible to build on existing work within the usability, user interface and modeling fields, specifically on techniques for modeling of functional and non-functional requirements, process modeling, user-centered design, model-driven architectures, requirements specifications of web-applications, domain-specific modeling and dependability analysis.

REFERENCES

Banavar, G. & Bernstein, A. (2002). Software infrastructure and design challenges for ubiquitous computing. *Communications of the ACM, 45*(12), 92-96.

Celesta (2001, June 6). Universal mBusiness Platform [Online]. Available at *http://www.celesta.com/pdf/products/mBusiness_Platform.pdf*

Colombo, E., Francalanci, M., Mecella, C., Pernici, B., & Plebani, P. (2002, December). Cooperative information systems in virtual districts: The VISPO approach. *Data Engineering*.

Eisenstein, J., Vanderdonckt, J., & Puerta, A. (2001). Applying model-based techniques to the development of UIs for mobile computers. In *Proceedings of ACM Conference on Intelligent User Interfaces IU 2001*.

Hirsch, R., Coratella, A., Felder, M., & Rodriguez, E. (2001). A framework for analyzing mobile transaction models. *Journal of Database Management*, *12*(3), July-September.

Jørgensen, H.D. (2001). Interaction as a framework for flexible workflow modeling. In *Proceedings of GROUP 2001*, Boulder, Colorado, October 2001.

Kelly, S. & Tolvanen, J-P. (2001). *Visual domain-specific modeling: Benefits and experiences of using metacase tools*. Metacase Consulting.

Krogstie, J. & Sølvberg, A.(2003a). *Information systems engineering - Conceptual modeling in a quality perspective*. Trondheim, Norway: Kompendiumforlaget.

Krogstie, J., Brandtzæg, P.B., Heim, J., & Opdahl, A.L. (2003b). Requirements construction for usable m-commerce systems. In K. Siau & E.-P. Lim (Eds.), *Advances in mobile commerce technologies* (pp.190-204), Hershey, PA: Idea Group Publishing.

Lyytinen, K. & Yoo, Y. (2002, April). The next wave of nomadic computing: A research agenda for information systems research, *Information Systems Research*.

Nilsson, E.G. (2002). Combining compound conceptual user interface components with modeling patterns – A promising direction for model-based cross-platform user interface development. *Proceedings of 9th International Workshop on the Design, Specification and Verification of Interactive Systems*, Rostock, Germany June 12–14.

Pries-Heie, H. & Baskerville, R. (2001). eMethodology. In *Proceedings of the IFIP TC 8 Conference on Developing a Dynamic, Integrative, Multi-disciplinary Research Agenda in E-commerce/E-Business*, Salzburg, June 22-23.

Rolland, C. & Prakash, C. (2000). Bridging the gap between organizational needs and ERP functionality. *RE Journal*, *5*(3), 180-193, Springer.

Siau, K., Lim, Ee-P., & Shen, Z. (2001). Mobile commerce: Promises, challenges, and research agenda. *Journal of Database Management, 12*(3) July-September.

Wegner, P. (1997). Why interaction is more powerful than algorithms. *Communications of the ACM*, *40*(5).

KEY TERMS

M-Commerce: A technological approach to reach customers, suppliers and employees regardless of where they are located and to deliver the right information to the right person(s) at the right time.

Mobile Computing: The capability to physically move computing services with us.

Mobile Information Systems: Information systems that include end-user terminals that are easily movable in space, are operable independent of location and typically have wireless access to information resources and services.

Multi-Channel Information Systems: Information systems that are to be used by different types of end-user equipment such as traditional PC, PDA, and a mobile phone in an integrated manner.

Nomadic Computing: The use of computers while on the move.

Pervasive Computing: An environment where computers have the capability to obtain information from the environment in which it is embedded and utilize it dynamically.

Ubiquitous Computing: An environment where computers are embedded in our natural movements and interactions with our environments. Combines mobile and pervasive computing.

U

Use Cases and the UML

Brian Dobing
University of Lethbridge, Canada

Jeffrey Parsons
Memorial University of Newfoundland, Canada

INTRODUCTION

The Unified Modeling Language, or the UML (Booch, Jacobson & Rumbaugh, 1999), has rapidly emerged as a standard language and notation for object-oriented modeling in systems development, while the accompanying Unified Software Development Process (Jacobson, Booch & Rumbaugh, 1999) has been developed to provide methodological support for application of the UML in software development. The UML is a non-proprietary modeling language managed by the Object Management Group, a not-for-profit consortium, which also manages several related modeling specifications. The UML has evolved from its initial version, with UML 2.0 formally adopted by the OMG in June 2003. This article is based on the UML 1.5 specifications (OMG, 2003), as those for UML 2.0 have not been finalized. However, the role of use cases appears to be essentially unaffected by the changes proposed for UML 2.0.

The term "use case" was introduced by Jacobson (1987) to refer to text document that outlines "a complete course of events in the system, seen from a user's perspective" (Jacobson, Christerson, Jonsson & Overgaard, 1992, p. 157). The concept resembles others being introduced around the same time. Rumbaugh, Blaha, Premerlani, Eddy, and Lorensen (1991), Wirfs-Brock, Wilkerson, and Wiener (1990), and Rubin and Goldberg (1992) use the terms "scenario" or "script" in a similar way. While use cases were initially proposed for use in object-oriented analysis and are now part of the UML, they are not inherently object-oriented and can be used with other methodologies.

BACKGROUND

A use case should have a clear goal and describe what should typically happen (but not how it should happen). Common examples would include a customer renting a video, purchasing an item, withdrawing funds from an account, and so forth. The use case also identifies the main "actors" involved, which, in these examples, could include the customer, employees (e.g., rental clerk), a device (bank machine), and so forth.

The use case differs from typical structured requirements analysis tools that preceded it in two important ways. First, the use case is largely text-based. Some refer to them as use case narratives to clearly distinguish them from use case diagrams, which provide an overview of the use cases and actors in the system. Together, they form the use case model. Structured analysis emphasized the importance of graphical tools, such as work flow and data flow diagrams. The UML has not abandoned diagrams; 13 are now included. The class, activity, communication (previously collaboration), sequence, state machine (previously statechart) and use case diagrams play important roles. But use cases are written in the customer's language, so that "users and customers no longer have to learn complex notation" (Jacobson et al., 1999, p. 38).

Second, use cases focus on complete transactions from the user's perspective. In particular, use cases have a goal, which comes from the goals of those who will be using the system (Cockburn, 2001). This also helps facilitate communication with the system's intended users. In UML terminology, a use case is initiated by an actor, usually a person in a particular role (e.g., cashier), but sometimes an external system. Other actors may be involved as well (e.g., customer). The use case provides a complete view of the transaction, from initiation to achievement of the defined goal.

Consistent with an object-oriented approach, use cases can also have generalizations and include/extend relationships. As with classes, generalization allows a child use case to override the behavior of its parent use case in certain situations. An include relationship is generally used when the same steps are required by several use cases, in order to avoid repetition. An included use case is dependent on base use cases and "never stands alone" (Booch et al., 1999, p. 221), although not everyone follows this convention. An extend relationship exists when a base use case incorporates another use case depending on certain conditions.

As discussed extensively in the next section, the content and format of use cases vary widely among published books and articles. In addition to the basic narrative, use cases may contain initial sections that specify assumptions, preconditions (that must be in place before the use case is applicable), and triggers that initiate the use case. At the conclusion, there may be specified postconditions (that must be true when the use case ends). Exceptions (or alternative paths) may also be documented along with relevant business rules that govern behavior within the use case. None of these are part of the UML 1.5 specifications (OMG, 2003), which contains no sample use cases or suggestions on format, but some or all of them may be useful in different situations.

ROLE OF USE CASES

Use cases have been all but universally embraced in object-oriented systems analysis and development books written since Jacobson et al. (1992). Despite this strong endorsement, there are many variations on Jacobson's original theme. First, there is a difference in content. Use cases, at least during the analysis phase, are generally viewed as a conceptual tool. The use case should emphasize "what" and not "how" (Jacobson et al., 1994, p. 146). This suggests use cases should not mention technology (e.g., Evans, 1999).

A review of use case examples shows that determining when the "what" ends and the "how" begins is not always easy. Brown (2002) interprets "what" to mean what the system will do rather than the internal implementation. Thus, his use cases include references to screen designs. So do those of Satzinger and Orvik (1996, p. 126). Others, such as Harmon and Watson (1998, p. 121) refer to specific technology (salesperson's laptop). And even Jacobson et al. (1992, p. 162) refer to a display "panel," "receipt button" and "printer" in one of their examples. Some use cases also include more detail on business rules. For example, the IBM Object-Oriented Technology Center (1997, p. 489) video store example includes the condition that customers who are not members pay a deposit of $60.00. In contrast, Kulak and Guiney (2000, p. 23, emphasis at source) state that use cases "should show the *what* exclusively," and their examples seem to follow this philosophy. However, as Larman (2002, p. 75) notes, use cases are not tied to object-oriented methodologies and thus are technology-independent in that sense.

Second, there are several variations proposed for use case formats. While the first use cases in Jacobson et al. (1992) were written as a paragraph of text, most others have adopted numbered steps. Soon after, Jacobson et al. (1994, p. 109) did so as well.

Third, the granularity of use cases also varies from coarse (few use cases) to fine (many). Most take a minimalist approach. Jacobson et al. (1994, p. 105) suggest that use cases should offer "measurable value to an individual actor". MacMaster (1997) argues that use cases be used only for main system functions. But White (1994, p. 7) states that "the collected Use Cases specify the complete functionality of the system". While Dewitz (1996) uses 11 use cases in her video store example, the IBM object-oriented technology center (1997) has 24. Kulak and Guiney (2000, p. 37) suggest that "most systems would have perhaps 20 to 50 Use Cases and some small systems even fewer". But, as they later point out (p. 88), "there are no metrics established to determine correct granularity". In contrast, Armour and Miller (2001, p. 244) claim that large systems may have hundreds of use cases.

Fourth, the level of detail within each use case also varies. For example, both Kulak and Guiney (2000, p. 125) and Armour and Miller (2001, p. 125) recommend limiting the length of the flow of events to two pages of text, but the latter also note that some practitioners prefer a few longer use cases to many short ones. Constantine and Lockwood (2000) distinguish between "essential" use cases, containing few if any references to technology and user interface implementation, and "concrete" use cases that specify the actual interactions.

Jacobson et al. (1999) advocate an iterative development approach in which both the number of uses cases and their level of detail increase as the life cycle progresses. They suggest that only the most critical use cases (less than 10%) be detailed in the first (inception) phase. As analysis progresses and requirements become firmer, additional use cases can be added and each can be expanded to include considerably more detail. The analyst could move toward concrete use cases or simply expand the detail within essential use cases. Some authors have become quite specific in describing the different levels. For example, Kulak and Guiney (2000) have identified four levels. However, knowing where to start, how far to go at each phase, and when to stop are clearly critical issues not easily resolved.

To further complicate the issue, some of those who favor fewer or less detailed use cases supplement them with "scenarios". Booch et al. (1999, p. 225) define scenarios as "basically one instance of a use case". "Add a customer" is a use case. Adding a specified customer with a particular name, address, and so forth is a scenario. Some references use scenarios to provide further detail on exception handling and other special cases, for example, customers with missing, improbable, or unusual data (Bennett, Skelton & Lunn, 2001; Lockheed Martin, 1996). However, the UML defines a scenario as "a specific sequence of actions that illustrates behaviors" (OMG,

Table 1. Summary of a research framework for studying the need for, and effectiveness of, use cases in the UML

Research Question	**Primary Independent Variable**	**Primary Dependent Variable**	**Methodology**
Do design/implementation details in use cases impede process redesign efforts?	Use case structure	Process innovation	Experiment; Case study
Can class diagrams be effectively extracted from use cases?	Use cases	Class diagram completeness	Case study; Developer surveys
Do use cases facilitate communication between developers and users?	Communication medium (use cases or class diagrams)	User understanding Domain coverage	Experiments; User surveys

2003, Glossary p. 13) with use case variants ("alternative sequences, exceptional behavior, error handling, etc.") handling other situations (pp. 2-136). While scenarios are commonly recommended for testing the completed system (e.g., Booch et al., 1999, p. 221), they can also be used to test the use cases themselves (Armour & Miller, 2001; Kulak & Guiney, 2000). How many scenarios, alternate paths and exception paths should be developed, and what their role should be in developing class and object diagrams, is not clear. A minimalist approach to use cases combined with extensive scenarios and paths may still result in a large and very detailed set of specifications.

While the general consensus seems to be in favor of a smaller set with relatively brief descriptions, "Use Case modeling concepts may be applied very informally and loosely or very rigorously and formally" (Armour & Miller, 2001, p. 70). The difficulty is determining when each is appropriate. Stories about organizations mired in hundreds or even thousands of use cases suggest that some limits need to be applied. Users will not read, or at least not properly understand, long and complex use cases. But a smaller set may be insufficient to develop the class diagram. Thus, the two key roles of use cases, to gather requirements and support development of the class diagram, may conflict somewhat and the ideal set of use cases for each role are perhaps quite different.

Fifth, and perhaps most important, the role of use cases varies among methodologies. Earlier work on the UML focused on the language itself, and was largely agnostic on issues of methodology. More recently, the UML specifications (OMG, 2003, p. 1-7) state: "The UML authors promote a development process that is use-case driven, architecture centric, and iterative and incremental." In particular, use cases provide "major input when finding and specifying the classes, subsystems and interfaces" (Jacobson et al., 1999, p. 34). In addition, they drive the development of subsequent behavioral models, including activity, sequence, and statechart diagrams (Masciaszek, 2001, pp. 133-150).

Schneider and Winters (2001), however, precede use case development by preparing a problem description and risk analysis, while Booch et al. (1999) include a vision statement (p. 344) containing a feature list and some risk identification. Why this document is not considered part of the UML is unclear, although it may be because projects have different starting points that are often outside the control of the analyst (Booch et al., 1999, p. 113). Armour and Miller (2001, pp. 48, 84) start with a vision document and business case, and then create a "system glossary" prior to use case development. Rosenberg and Scott (1999) suggest that "domain modeling" precede use case development. Blaha and Premerlani (1998, p. 49) state, "Once you have a sound object model, you should specify Use Cases." Schmuller (1999) starts with class diagrams and then moves to writing use cases. We have also seen organizations begin with class diagrams for pragmatic reasons, such as when the data structure is already largely defined by existing systems. Regardless of which is used, some clear understanding of the problem seems essential before use case writing begins.

FUTURE TRENDS

Use cases have generally been well received and are not only being used as part of the UML but also with traditional structured methodologies. While no major changes to the basic concept are foreseen, some resolution of the issues raised previously is likely as both researchers and system developers become more familiar with them.

Better tools may also change the ways in which use cases are written and used. Many developers are still relying largely on word processors to develop use cases. A closer integration with other UML diagrams would help ensure consistent specifications across all UML components and greatly facilitate changing the specifications when modifications are required.

There is considerable opportunity for researchers to examine the issues identified, particularly format, granularity, level of detail and usage. Perhaps the most fundamental issues deal with the effectiveness of use cases in their two key roles, communicating with users and supporting development of UML diagrams and particularly the class diagram. How effective are use cases in helping users find incomplete or incorrect specifications? More importantly, how effective are they at helping find better ways to do things? Do use cases support creative problem solving?

CONCLUSION

In summary, review of the literature shows extensive differences in how use cases are defined and used. This is not surprising, given the UML's relatively short history. But these differences can also be attributed to the lack of a theoretical foundation. The UML is promoted as the "culmination of best practices in practical object-oriented modeling" (OMG, 2003, p. xxv). This is a notable achievement, but there is little research to support claims of "best practices" beyond the claims of successful practitioners and the review contained here demonstrates that recommended practices are sometimes contradictory.

REFERENCES

Armour, F., & Miller, G. (2001). *Advanced use case modeling.* Boston: Addison-Wesley.

Bennett, S., Skelton, J., & Lunn, K. (2001). *Schaum's outline of UML.* New York: McGraw-Hill.

Blaha, M., & Premerlani, W. (1998). *Object-oriented modeling and design for database applications.* Upper Saddle River, NJ: Prentice Hall.

Booch, G., Jacobson, I., & Rumbaugh, J. (1999). *The Unified Modeling Language user guide.* Reading, MA: Addison-Wesley.

Brown, D. (2002). *An introduction to object-oriented analysis: Objects in plain English* (2nd ed.). New York: John Wiley & Sons.

Cockburn, A. (2001). *Writing effective use cases.* Boston: Addison-Wesley.

Constantine, L.L., & Lockwood, L.A.D. (2000). Structure and style in use cases for user interface design. In M. Van Harmelen & S. Wilson (Eds.), *Object modeling user interface design.* Reading, MA: Addison-Wesley.

Dewitz, S. (1996). *Systems analysis and design and the transition to objects.* New York: McGraw-Hill.

Evans, G. (1999). Why are use cases so painful? *Thinking Objects, 1*(2). *http://evanetics.com/articles/Modeling/UCPainful.htm*

Harmon, P., & Watson, M. (1998). *Understanding UML: The developer's guide.* San Francisco: Morgan Kaufmann.

IBM Object-Oriented Technology Center. (1997). *Developing object-oriented software.* Upper Saddle River, NJ: Prentice Hall.

Jacobson, I. (1987). Object-oriented development in an industrial environment. *OOPSLA'87 Conference Proceedings, SIGPLAN Notices, 22*(12), 183-191.

Jacobson, I., Booch, G., & Rumbaugh, J. (1999). *The unified software development process.* Reading, MA: Addison-Wesley.

Jacobson, I., Christerson, M., Jonsson, P., & Overgaard G. (1992). *Object-oriented software engineering: A use case driven approach.* Reading, MA: Addison-Wesley.

Jacobson, I., Ericsson, M., & Jacobson, A. (1994). *The object advantage: Business process reengineering with object technology.* Reading, MA: Addison-Wesley.

Kulak, D., & Guiney, E. (2000). *Use cases: Requirements in context.* New York: ACM Press.

Larman, C. (2002). *Applying UML and patterns: An introduction to object-oriented analysis and design* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Lockheed Martin Advanced Concepts Center and Rational Software Group. (1996). *Succeeding with the Booch and OMT methods: A practical approach.* Menlo Park, CA: Addison-Wesley.

Maciaszek, L. (2001). *Requirements analysis and system design: Developing information systems with UML.* New York: Addison-Wesley.

MacMaster, B. (1997). Saving time with 'use cases.' *Computing Canada, 23*(21), 52.

OMG. (2003). OMG Unified Modeling Language specification, version 1.5. *http://www.omg.org/cgi-bin/doc?formal/03-03-01*

Rosenberg, D., & Scott, K. (1999). *Use case driven object modeling with UML.* Reading, MA: Addison-Wesley.

Rubin, K., & Goldberg, A. (1992). Object behavior analysis. *Communications of the ACM, 35*(9), 48.

Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F., & Lorensen, W. (1991). *Object-oriented modeling and design*. Englewood Cliffs, NJ: Prentice Hall.

Satzinger, J., & Orvik, T. (1996). *Object-oriented approach: Concepts, modeling, and system development*. Danvers, MA: Boyd & Fraser.

Schmuller, J. (1999). *Teach yourself UML in 24 hours*. Indianapolis, IN: Sams.

Schneider, G., & Winters, J.P. (2001). *Applying use cases: A practical guide* (2nd ed.). Boston, MA: Addison-Wesley.

White, I. (1994). *Rational rose essentials: Using the Booch method*. Redwood City, CA: Benjamin/Cummings.

Wirfs-Brock, R., Wilkerson, B., & Wiener, L. (1990). *Designing object-oriented software*. Englewood Cliffs, NJ: Prentice Hall.

KEY TERMS

Actor: An actor plays one or more roles in relation to a set of use cases. An actor could correspond to a job title (e.g., purchasing agent, sales clerk) or can be non-human (e.g., another system or database). Each actor in a use case must be directly involved at some point and is not merely a stakeholder (someone or something that is affected by the success or failure of a particular transaction).

Use Case: Use cases are used during system analysis to help specify the behavior of a proposed system. A use case is a largely text-based description of a complete transaction as seen from the user's perspective. A use case should emphasize what the system will do, not how it will do it, and be written in the user's language. Thus, use cases are an important communication tool between developers of systems and the intended users.

Use Case Diagram: The use case diagram shows a set of use cases (by title) and the actors involved in them. This provides an overview of the use case structure and also shows how each actor is involved in a system.

Use Case Scenario: A scenario, as specified in the UML, is an instance of a use case that can help illustrate its use. For example, a use case (rent video) might have different scenarios for renting a video to a child, a new customer, an existing customer with overdue videos, and so forth. However, the term is also used in other ways outside the UML.

User Experiences of the E-Commerce Site with Standard User Interface

Shin-Ping Liu
JCC, USA

Dennis Tucker
Yaasuo Consulting Company, USA

INTRODUCTION

The future of economic competitiveness for most enterprises relies on entrance and active participation in the e-commerce market (Vestal, 1999). E-commerce offers speed, convenience and often cost-effectiveness for today's busy shopper, but most e-commerce sites are still too hard to use. Zona Research (1999) found that 62% of online shoppers had given up at least once while looking for products, and 42% had turned to traditional channels to make their purchases. These statistics are astounding. In addition, according to Nielsen and Norman (2000), users fail when they try to purchase products on an e-commerce site about a third of the time. In reality, what happens is not just that the user fails, but that the site fails and does not sell a thing. Is "lost-in-hyperspace" primarily a psychological or an engineering problem? In other words, is "lost-in-hyperspace" a problem for users, or is it a symptom of poor design, which itself may be a psychological problem for authors of e-commerce sites?

Traditional services design user interfaces to treat users as suppliants who will simply have to learn to ask the right questions to get the right answers (Gonzalez, 1999). This approach will not work for the world of e-commerce. Much like driving a car or talking on the telephone, e-commerce follows the same rules in that the users are not concerned with the inner workings, but with the end result. E-commerce will make significant advances as long as the interface is kept simple and can be used in the same context as telephones are used today, according to Gonzalez (1999).

The problem with e-commerce sites is that the controls and organization are different for each site. This study investigates how an individual's perceptions of key beliefs surrounding the use of the e-commerce site with the standard user interface would influence the individual's decision to accept these e-commerce sites for online shopping purposes. This study applies the technology acceptance model (TAM) identifying components of usefulness and ease of use that predict user attitude toward the usage of the standard user interface. The research uses a Web-based survey and employs TAM with path analysis to identify features of the standard user interface that might contribute to its usefulness and ease of use.

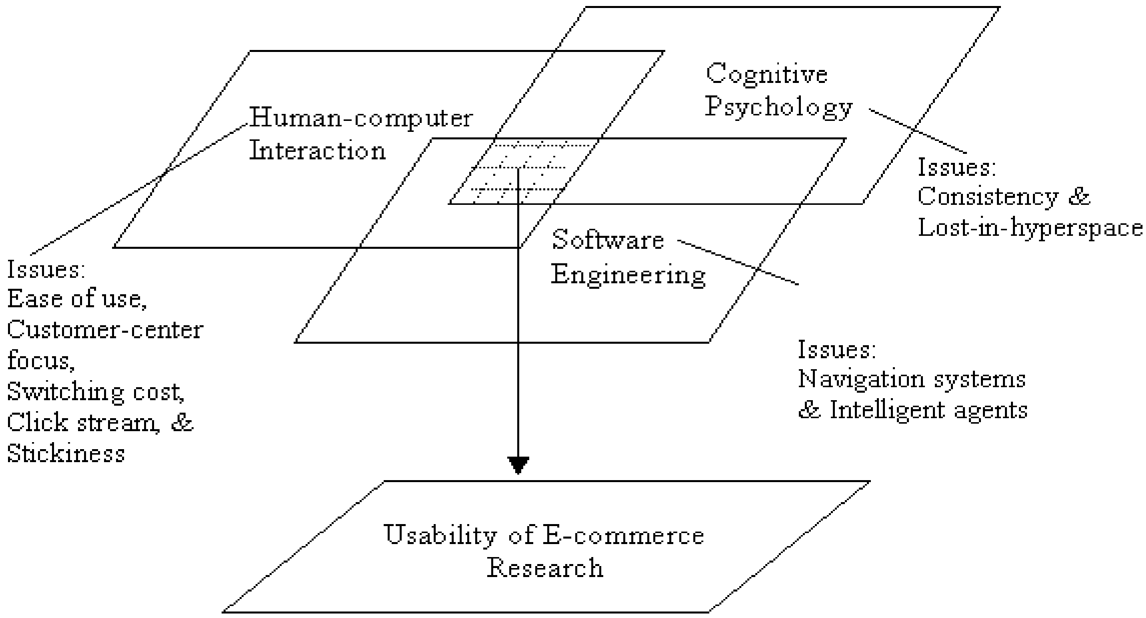
BACKGROUND

This study provides a review of the pertinent literature via an interdisciplinary approach integrating different disciplines into the usability of e-commerce research. In addition, to study user acceptance of e-commerce sites from an interdisciplinary perspective, TAM used by Davis (Davis, 1993; Davis, Bagozzi & Warshaw, 1989) is adopted to explain the usage of information technology.

Interdisciplinary Approach to E-Commerce Usability

An interdisciplinary approach integrates three disciplines of human-computer interaction, cognitive psychology and software engineering into the usability of e-commerce research. Theng and Thimbleby (1998) state that the human-computer interaction is typically grounded in empirical data from usability studies such as ease of use, customer-center focus, switching cost, click-stream, and stickiness. Cognitive psychology is typically grounded in models of human behavior and performance such as consistency and lost-in-hyperspace. Software engineering is typically grounded in sound engineering practice such as navigation systems and intelligent agents. Academic research may wish to promote a discipline-specific well without seeing the wider picture; for example, few software engineers are seriously concerned with usability (Theng & Thimbleby, 1998). This study is concerned with how the concepts, values, methods, and procedures from these disciplines can be integrated into the usability of e-commerce research; therefore, this study explores closely related literature in the areas of increasing the usability of e-commerce.

Figure 1. Interdisciplinary approach to the usability of e-commerce: integrating the disciplines of human-computer interaction, software engineering, and cognitive psychology into the usability of e-commerce research



Technology Acceptance Model (TAM)

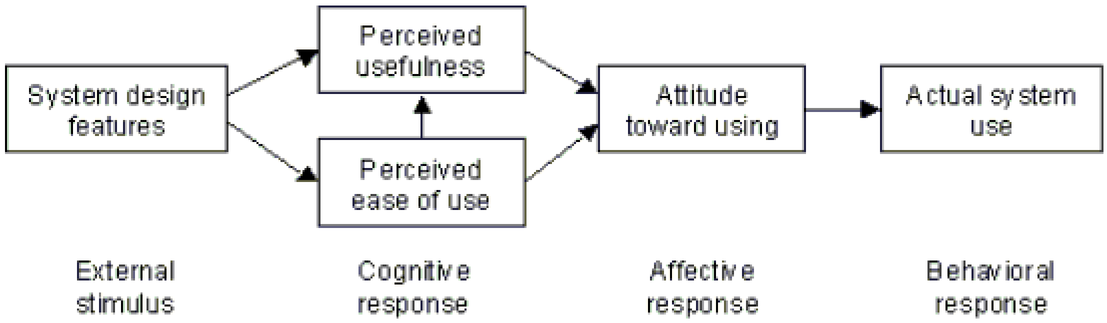
Several researchers have validated TAM using different applications including e-mail, voice mail, word processing, micro-computers, automated teller, spreadsheet, calculator, Web pages development software, among others (Adams, Nelson & Todd, 1992; Bagozzi, Davis & Warshaw, 1992; Chau, 1996; Davis, Bagozzi & Warshaw, 1989; Fulk, Schmitz & Ryu, 1995; Hendrickson & Collins, 1996; Igbaria, Guimaraes & Davis, 1995; Szajna, 1996; Thompson, 1998). Other researchers have recommended TAM for the investigation of Web user behavior (Shaw, Gardner & Thomas, 1997). TAM provides a foundation for research on why users accept or reject information technology and how to increase user acceptance by judicious choice of system design features (Davis, 1993). Davis's proposed TAM is shown in Figure 2.

A prospective user's overall attitude toward a given system is hypothesized to be a major determinant of whether the user actually uses it or not. The perception of the stimuli creates cognitive beliefs, which initiate an affective response. The affective response has an influence on consumer behavior. Attitude is determined by cognitive beliefs. Attitude toward use, in turn, is a function of two beliefs: perceived usefulness and perceived ease of use. Perceived ease of use has a causal effect on perceived usefulness. System design features directly influence perceived usefulness and perceived ease of use. System design features have an indirect effect on attitude toward use and actual usage behavior through their direct effect on perceived usefulness and perceived ease of use. Perceived usefulness refers to the degree to which a person believes that a particular information system would enhance his or her job performance by reducing the time to accomplish a task or providing timely information (Davis, Bagozzi & Warshaw, 1989; Lederer, Maupin, Sena & Zhuang, 1999). Perceived ease of use refers to the degree to which a person believes that using a particular system would be free of effort (Davis, Bagozzi & Warshaw, 1989).

METHODOLOGY

A Web-based survey study with two specific research objectives was conducted. The first objective was to explore the current usability problems of e-commerce sites and employ an artificial intelligence user interface

Figure 2. Technology acceptance model by Davis (1993)



agent to create a standard user interface for increasing the usability of e-commerce. The second objective was to examine the user acceptance of the e-commerce site with the standard user interface. Employing TAM, the study was designed to identify features of the standard user interface that might contribute to its usefulness and ease of use.

The Web-based survey instrument contains instructions asking respondents to compare an e-commerce site without a standard user interface to the same site with the standard user interface. Altogether the instrument contains 33 items, including seven demographic questions. A letter that solicits participation in the study was sent out via e-mail to students and faculty members listed on several mailing lists at a major university. All together a total of 2,946 potential participants were contacted.

Research Hypotheses

The hypotheses are adopted from the Davis study (1993). TAM is represented by the following structural equations

$$\text{Ease of Use} = \beta_{11}\ (\text{System}) + e_1$$

$$\text{Usefulness} = \beta_{21}\ (\text{System}) + \beta_{22}\ (\text{Ease of Use}) + e_3$$

$$\text{Attitude} = \beta_{31}\ (\text{Ease of Use}) + \beta_{32}\ (\text{Usefulness}) + e_3$$

$$\text{Usage} = \beta_{41}\ (\text{Attitude}) + e_4$$

H_1: Attitude toward use will have a significant positive effect on actual standard navigation usage.

H_2: Perceived usefulness will have a significant positive effect on attitude toward use, controlling for perceived ease of use.

H_3: Perceived ease of use will have a significant positive effect on attitude toward use, controlling for perceived usefulness.

H_4: Perceived ease of use will have a significant positive effect on perceived usefulness, controlling for system.

H_5: System will have a significant positive effect on perceived ease of use and perceived usefulness.

H_6: Perceived usefulness, perceived ease of use, and system will not have significant direct effects on actual standard navigation usage, controlling for attitude toward use.

H_7: System will not have a significant direct effect on attitude toward use, controlling for perceived usefulness and perceived ease of use.

Figure 3. Standard navigation in action

(Duncan, 1975). While "System" is a dummy variable taking on value 0 for an e-commerce site without the standard user interface and 1 for an e-commerce site with the standard user interface, "Usage" refers to intensity of the actual standard navigation usage, "Attitude" refers to attitude toward use, "Usefulness" refers to perceived usefulness, and "Ease of use" refers to perceived ease of use. Ordinary least-squares (OLS) regression is used to test this structural equation model.

The statistical significance of the proposed TAM relationships, represented in the following hypotheses, was assessed using the t-statistic corresponding to each estimated parameter. The hypotheses for this study are as follows: H_1-H_5. Furthermore, two more hypotheses (H_6 and H_7) are postulated to test whether the causal relationships implicitly hypothesized to be indirect have no significant direct effect or not. Hierarchical regression and associated F-tests of the significance of the increase in R^2 due to the additional variables were used for these hypotheses. In addition to testing for the significance of the hypothesized relationships, the data to estimate the magnitudes of the causal parameters are also analyzed. The estimates are the standardized regression coefficients, expressed as both point and confidence interval estimates.

Standard User Interface in E-Commerce Sites

One hundred and two e-commerce sites are evaluated that had been recognized as most popular with consumers by several independent polling organizations for several design criteria that may affect customer experience. From this evaluation of the 102 sites, 42 different features commonly used are identified and compiled to assist customers who browse e-commerce sites. Some of the most frequently used such features include contact information, privacy statement, company information, customer service/help, search engine, shopping cart, and security statement. Based on the list of features identified from the 102 e-commerce sites, a standard user interface that allows users to navigate was constructed. The idea behind the standard user interface is that all e-commerce sites should have some type of standard navigation to improve usability and functionality. Unfortunately, most e-commerce sites lack such standard navigational systems (Allum, 1998; Booker, 1997; Bort, 1999; Hoffman, 1996). In addition, based on the findings from the study of the 102 e-commerce sites, a menu bar as shown in Figure 3 was developed as the standard user interface to be used by study participants.

Then three example sites were selected from the list of 102 e-commerce sites to be used in this study. For each of the three e-commerce sites, two versions of the site are created—one with the standard user interface and another without. Figure 4 compares the two versions of one of the sites used in the study.

DATA ANALYSIS AND RESULTS

The survey produced 127 usable responses out of 2,946 students and faculty members at UNT for a 4.4% response rate. This response rate may be low in comparison to conventional paper-based postal surveys. However, the total number of subjects suffices for the analysis described in the following section. The testing of the TAM relationships then used a multiple regression model and a simple linear regression model to test the mean differences among various conditions. To examine the construct validity of the instrument, factor analysis was performed to test the factor loadings of the items on these constructs: usefulness, ease of use, and attitude toward use. The Cronbach model of reliability was used to assess internal consistency, based on the average inter-item correlation. Cronbach's alpha coefficients were 0.95 for the items used to measure of perceived usefulness, 0.85 for perceived ease of use, 0.83 for attitude toward use, and 0.87 for actual standard navigation usage.

Findings of the Study

The TAM motivational variables of perceived usefulness and perceived ease of use mediate the effect of system design features on the attitude toward use. The system characteristics appear to influence behavior through these motivational variables. The most significant finding is the direct effect of usefulness on attitude toward use. How-

Figure 4. E-commerce site without the standard user interfaces (left) and with the standard user interfaces (right)

ever, the fact that perceived usefulness exerts almost the same as the system design features do direct and indirect influence on attitude toward use. In addition, perceived usefulness exerts more than two times as much direct effect on attitude toward use as does perceived ease of use. Compared to perceived usefulness, perceived ease of use has a direct effect on attitude toward use, much smaller to affecting attitude indirectly via its relatively influence on perceived usefulness. In addition, the previous discussion has emphasized the importance of perceived usefulness, arguing the ease of use performs via this variable. Thus, the model views the standard navigation usage behavior to be intrinsically motivated; being driven by concern performance gains more than associated rewards. That is, people use systems in part because users who have more system familiarity or experience enjoy the process of using them per se, not just, because they are extrinsically rewarded for the performance impacts of usage.

FUTURE TRENDS

The future of economic competitiveness for most enterprises relies on entrance and active participation in the e-commerce market (Vestal, 1999). Bonisteel (2000) claims that, to be successful, sites have to be extremely intuitive and simple to use. The site's providers need to make navigation systems dramatically simpler. Bonisteel's (2000) advice for online retailers is to remove the obstacles consumers most often encounter while shopping online by improving search engines, information entry systems, and site navigation. Designers must start studying the trends of real people. Gonzalez (1999) states that answers to real trends are solutions that must be implemented onto the user interface. The designers no longer have full control, and the users are the ones placing demands on how things are to be done. Although shopping online by simply clicking a button is extreme convenience, users become more demanding as time passes and e-commerce becomes more and more popular. The success of e-commerce depends on the ability to attract non-computer-literate people through ease of use (Gonzalez, 1999).

CONCLUSION

Many designers believe that the key barrier to user acceptance is the lack of user friendliness and that adding a user-friendly interface to increase usability is the key to success (Branscomb & Thomas, 1985; Davis, 1993). Yet, the current results indicate that, although ease of use is clearly important, the usefulness of the system is even more important. Users who are familiar with the system may be willing to tolerate a difficult interface in order to access functionality that helps them on their online shopping, while no amount of ease of use can compensate for a system that does not do a useful task. Thus, the intention of this study is to test the behavioral constructs of perceived usefulness and perceived ease of use as predictors of usage acceptance of the standard user interface in e-commerce sites. The results show a strong relationship between users' perceptions of usefulness and ease of use about the standard user interface.

REFERENCES

Adams, D.A., Nelson, R.R., & Todd, P.A. (1992). Perceived usefulness, ease of use, and usage of information technology: A replication. *MIS Quarterly*, *16*(2), 227-247.

Allum, M. (1998). Building dynamic menu bars. Retrieved March 26, 2000, from http://webreview.com/wr/pub/98/07/24/coder/index.html

Bagozzi, R.P., Davis, F.D., & Warshaw, P.R. (1992). Development and test of a theory of technological learning and usage. *Human Relations*, *45*(7), 659-686.

Bonisteel, S. (2000, March). More online shopping: More unhappy shoppers - Study. *Newsbytes*. Retrieved May 7, 2000, from *http://www.andovernews.com/cgi-bin/news_story.pl?151228,topstories*

Booker, E. (1997, July). Defining a strong Web interface usability. *Web Week*. Retrieved May 7, 2000, from http://*www.internetworld.com/print/1997/07/07/undercon/19970707-strong.html*

Bort, J. (1999). Navigation: An art for e-commerce sites. Retrieved April 4, 2000, from *http://www.microtimes.com/201/ecombort201a.html*

Branscomb, L.M., & Thomas, J.C. (1985). Ease of use: A system design challenge. *IBM Systems Journal*, *23*, 224-235.

Chau, P.Y.K. (1996). An empirical assessment of a modified technology acceptance model. *Journal of MIS*, *13*(2), 185-204.

Davis, F.D. (1993). User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *International Journal of Man-Machine Studies*, *38*, 475–487.

Davis, F.D., Bagozzi, R.P., & Warshaw, P.R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, *35*(8), 982-1003.

Duncan, O.D. (1975). *Introduction to structural equation models*. New York: Academic Press.

Fulk, J., Schmitz, J., & Ryu, D. (1995). Cognitive elements in the social construction of communication technology. *Management Communication Quarterly*, *8*(3), 259-288.

Gonzalez, M. (1999). Developing the e-commerce user interface. Retrieved April 15, 2000, from *http://eies.njit.edu/~turoff/coursenotes/CIS732/samplepro/ecom.html*

Hendrickson, A.R., & Collins, M.R. (1996). An assessment of structure and causation of IS usage. *The DATA BASE for Advances in Information Systems*, *27*(3), 61-67.

Hoffman, M. (1996). Enabling extremely rapid navigation in your Web or document. Retrieved April 15, 2000, from *http://www.pdrinterleaf.com/infoaxcs.htm*

Igbaria, M., Guimaraes, T., & Davis, G.B. (1995). Testing the determinants of microcomputer usage via a structural equation model. *Journal of Management Information Systems*, *11*(4), 87-114.

Lederer A.L., Maupin, D.J., Sena, M.P., & Zhuang, Y. (1999). The technology acceptance model and the World Wide Web. *KIKM Research Paper*, 125.

Nielsen, J., & Norman, D.A. (2000, January). Web-site usability: Usability on the Web isn't a luxury. Retrieved March 28, 2000, from http://www.informationweek.com/773/web.htm

Shaw, M.J., Gardner, D.M., & Thomas, H. (1997). Research opportunities in electronic commerce. *Decision Support Systems*, *21*, 149-156.

Szajna, B. (1996). Empirical evaluation of the revised technology acceptance model. *Management Science*, *42*(1), 85-92.

Theng, Y.L., & Thimbleby, H. (1998). Addressing design and usability issues in hypertext and on the World Wide Web by re-examining the "lost in hyperspace" problem. *Journal of Universal Computer Science*, *4*(11), 839-855.

Thompson, R. (1998). Extending the technology acceptance model with motivation and social factors. *Proceedings of Association for Information Systems Annual Conference* (pp. 757-759).

Vestal, C. (1999). *Electronic commerce environments: Corporate obstacles and opportunities to competitiveness*. Masters thesis. Bowie State University.

Zona Research. (1999). Zona Research's online shopping report. Retrieved April 4, 2000, from *http://www.zonaresearch.com/info/press/preleases99.htm*

KEY TERMS

E-Commerce: Financial business transaction that occurs over an electronic network such as the Internet.

Navigation: The process by which a user explores all the levels of interactivity, moving forward, backward, and through the content and interface screens.

Perceived Ease of Use: The degree to which a person believes that using a particular system would be free of effort.

Perceived Usefulness: The degree to which a person believes that a particular information system would enhance his or her job performance by reducing the time to accomplish a task or providing timely information.

Standard User Interface: Even as implementation technology changes at any given time, usability is improved by following standards and conforming to user expectations, but over time, these standards and expectations have to evolve as the technology improves and new interface ideas are invented.

Stickiness: "Stickiness" of a site refers to the likelihood that visitors will stay on the site for a while and be inclined to return again and again.

Switching Cost: Refers to how much cost it takes to switch from one vendor to another.

Technology Acceptance Model: Describes the relationship between perceived qualities of a system's usage, affective attitude, and behavioral responses to the system.

Usability: How easy an interface design is to understand and use. A user-friendly document will let the user read or play any content at will; it will have unambiguous interactive controls and a clear navigational scheme.

User Interface: Controls how data and instructions are entered and how information displays on a screen.

User Perceptions and Groupware Use

Gina Green
Baylor University, USA

John Day
Ohio University, USA

Hao Lou
Ohio University, USA

Craig Van Slyke
University of Central Florida, USA

INTRODUCTION

The increasingly complex, global nature of work tasks has led to increased interest in virtual teams that interact across space, time, and organizational boundaries to achieve organizational objectives (Bell & Kozlowski, 2002; Roebuck & Britt, 2002). One of the factors thought to contribute to the popularity of virtual teams is the availability of groupware systems (Townsend, DeMarie & Hendrickson, 1998). While some groupware systems, such as electronic mail, have become almost ubiquitous in many organizations, other groupware applications have not enjoyed similar acceptance (Briggs, Adkins, Mittleman, Kruse, Miller & Nunamaker, 1999; Orlikowski, 1993). Because of the importance of groupware to the success of virtual teams, it is important to understand *why* this innovation has not successfully diffused.

This article uses a diffusion of innovation (DOI) perspective (Rogers, 1995) to understand factors that impact intentions to use groupware technology. The Rogers' DOI perspective gives us a much richer set of factors than other technology adoption models, and should therefore better aid in understanding groupware adoption (Plouffe, Hulland & Vendenbosch, 2001). We surveyed 186 college students and found that intentions to use groupware technology are impacted by perceptions of: relative advantage gained from use of the groupware, amount of complexity in groupware use, compatibility with work practices, and demonstrable results. Suggestions for positively influencing these factors are offered in order to ensure more successful groupware implementations.

BACKGROUND

Group Support Systems

Groupware technology facilitates the work of groups by providing electronic means to communicate, cooperate, coordinate, solve problems, compete, or negotiate. While traditional technologies such as the telephone qualify as groupware, the term is ordinarily used to refer to a specific class of technologies relying on modern computer networks.

The origins of groupware technology are often traced back to the early 1980s, when academic researchers at the University of Arizona, University of Minnesota, and Southern Methodist University developed group "decision rooms" supported by group decision-making software (Power, 2003). With advances in telecommunications over the last two decades, groupware applications have expanded to include e-mail, audio/video/data conferencing, instant messaging, electronic meeting systems, and a host of Web-based collaboration tools. With approximately 130 million workers worldwide expected to telework in 2003, the integration of groupware into organizations is expected to grow rapidly (Roebuck & Britt, 2002).

The growth in virtual teams also reflects this change in work habits, as employees may be located anywhere around the world at any point in time (Townsend et al., 1998). Virtual teams use groupware to span geographic, temporal, and organizational boundaries. The sophisticated communication facilities of groupware facilitate

Figure 1. Groupware classification (source: Johansen, 1988)

	Same time "synchronous"	**Different time** "asynchronous"
Same place "co-located"	Group decision support systems, Voting, presentation support	Shared computers
Different place "distance"	Videophones, chat, instant messaging	Discussions, e-mail, workflow

frequent communication among team members, which is an important factor in creating a sense of identity in virtual teams (Kezsbom, 2000). In particular, asynchronous groupware helps overcome time-related barriers to distributed work (Kelly & Jones, 2001).

Groupware technologies are typically categorized along two dimensions, time and place (Johansen, 1988), as shown in Figure 1. Based on the time dimension, users of the groupware can work together at the same time or different times. On the other hand, the place dimension indicates that groupware users can work together in the same place or in different places.

Diffusion of Innovations

Diffusion of innovation (DOI) research is concerned with how use of an innovation spreads throughout a social system (Mahajan, Mueller & Bass, 1990). Diffusion theory has been applied to a wide range of technologies, including information and communication technologies such as groupware. Diffusion theory states that potential adopters' *perceptions* of an innovation's characteristics, rather than an objective assessment of how an innovation rates on these characteristics, impact the diffusion rate (Rogers, 1995). Rogers (1995) identifies five perceived characteristics of an innovation that influence its adoption: relative advantage, compatibility, complexity, trialability, and observability. Moore and Benbasat (1991) provide empirical support for breaking observability into two constructs: result demonstrability and visibility. Additionally, because use of innovations may be optional in some settings, the degree to which potential users feel that innovation use is voluntary has been found to be important in understanding innovation use (Agarwal & Prasad, 1997).

Other models such as the technology acceptance model, or TAM (Davis, 1989), have proposed characteristics that may influence adoption, including perceived usefulness and perceived ease of use, which are conceptually similar to relative advantage and complexity (Moore & Benbasat, 1991). However, this research uses Rogers' diffusion theory with the additional constructs noted previously. Reasons for this are two-fold: first, we seek a better understanding of the groupware technology diffusion process. As such, Rogers' model offers a richer set of potential factors than does the more parsimonious TAM model (Plouffe et al., 2001); second, we hope to provide guidance to managers to ensure successful groupware implementations. Again, Rogers' model provides us with more areas that can be influenced by management to create environments conducive to groupware adoption.

MODEL OF GROUPWARE TECHNOLOGY DIFFUSION

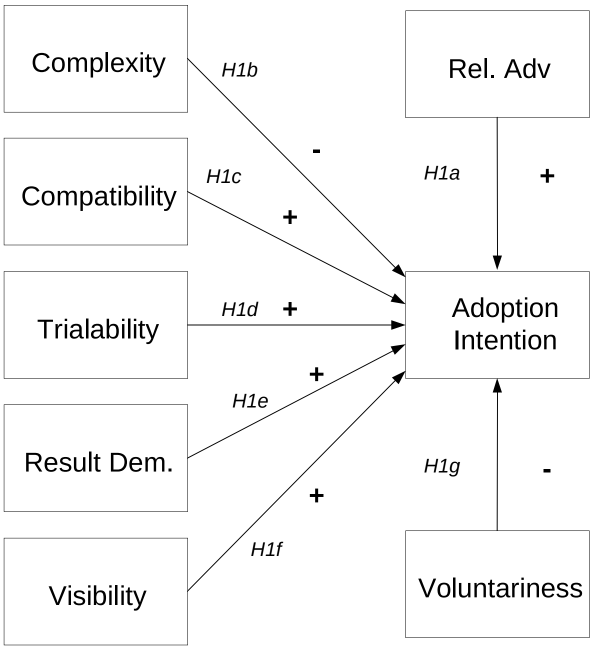
Based on the previous research, we propose the model in Figure 2 for understanding factors important to groupware technology diffusion. In general, the model suggests that users' perceptions of a groupware system influence their subsequent intentions to use the groupware system. Research indicates that these intentions are highly correlated with actual future use (Davis, 1989).

Validating the Model Using Domino

As a test of our model, we surveyed students at a major midwestern university. The students, primarily college seniors, were enrolled in courses where a groupware system, Lotus Domino, was made available to them for optional use in the course. The Lotus Domino Discussion database is an asynchronous groupware product designed to be used "any time and any place," placing it into the lower right-hand portion of the grid in Figure 1. The Domino groupware system was chosen because it supports student-centered, project-based courses, with faculty as facilitators of student learning, as opposed to providers of information. In our experience, Domino does help in achieving learning goals in such courses (Day, Lou & Van Slyke, 2004).

Users access Domino discussion databases and participate in group discussions over the Internet using a Web browser. Users have the ability to browse through or participate in discussion topics and responses contributed by others. The history of any discussion is pre-

Figure 2. Model for diffusion of groupware technology



served in the discussion database, and is presented as a discussion thread. Figure 3 illustrates a threaded discussion.

Findings

We used a regression analysis to test our data against the research model. Sixty-five percent of the variance in use intention is explained by the model variables (adjusted R^2 = 0.656). Results, summarized in Table 1, show that all but three perceived groupware innovation characteristics are important in predicting groupware use. Examining the significance values for each variable suggests that perceptions of relative advantage and compatibility are better predictors of subsequent use than are perceptions of result demonstrability and complexity. While examining the significance values allows us to detect probable phenomenon, an examination of subjects' comments regarding Domino confirms these findings and confirms the statistical results.

Several subjects mentioned the value of the frequent and timely instructor feedback via Domino, as well as the value of being able to post assignments to Domino, rather than having to travel to campus to hand them in. Others noted that Domino facilitated interaction with team members. These comments confirm the relative advantages of using Domino.

Also interesting are comments related to compatibility. Many subjects either praised the ability to work from home, or lamented the necessity to travel to a campus computer lab to use Domino. Further, while one subject stated, "It was easier to 'talk' about things without feeling put on the spot the way I sometimes feel in a classroom," others commented that they would rather interact face-to-face than through Domino. While these comments seem contradictory, they demonstrate the importance of groupware being compatible with the work and communication styles of the intended users.

Implications

Our study validated that groupware use is greatly influenced by users' perceptions of the technology. Thus it is important that efforts be taken to ensure potential users have positive perceptions of groupware. One aspect that

Figure 3. Discussion thread in a Domino discussion database

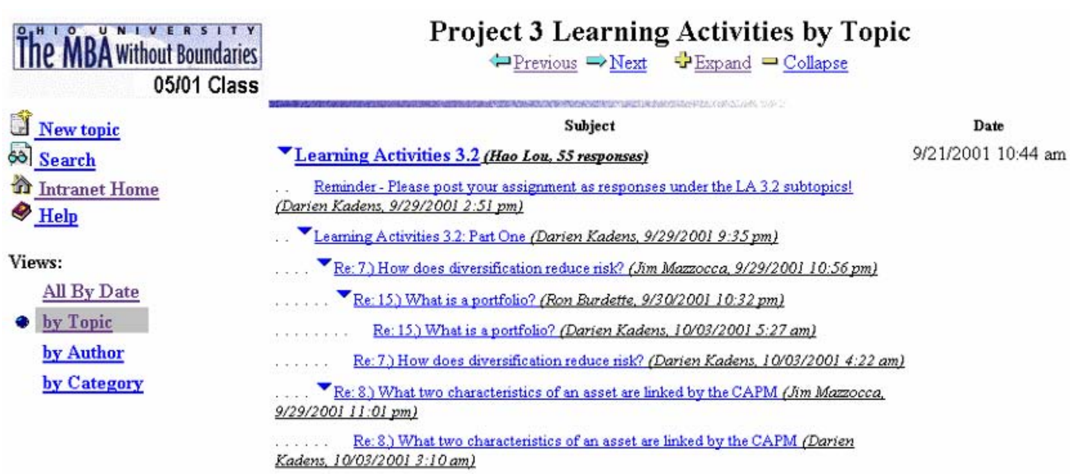


Table 1. Regression analysis results (a < 0.05) (Note: italics indicate statistical significance).

Construct	beta	Significance
Relative advantage	*0.383*	*0.001*
Complexity	*-0.170*	*0.035*
Compatibility	*0.335*	*0.001*
Trialability	0.012	0.824
Result demonstrability	*0.207*	*0.014*
Visibility	-0.027	0.601
Voluntariness	0.044	0.475

can have a negative impact on perceptions is a poorly functioning, unreliable groupware system. Having a sound technology infrastructure in place for the groupware system is critical. Adequately-performing servers and a fast, reliable network should be viewed as requisite for the use of groupware. If users have difficulty accessing the groupware system, it is likely that perceptions of relative advantage and complexity will be impacted in a manner that negatively influences use intentions.

In educational settings, it is important that instructors pay close attention to the quality and frequency of early, class-related discussion database postings by students. Encouraging students to make frequent, proper use of the groupware system is likely to increase perceived relative advantage for most users. The same holds, we believe, in non-educational settings. Organizational managers should encourage early and frequent use of groupware for team discussions. Offering incentives and showcasing teams that frequently and productively use groupware can serve to motivate others to increase their groupware use. If only a handful of users utilize the system, the value of this utilization is less than if a larger number of users provide meaningful postings.

Another factor that change agents should consider is their own use of groupware. For example, in an educational setting, it is important for instructors to make full use of the groupware system. In non-educational settings, it is likewise important for managers to model desired behavior by using groupware to, for example, solicit staff feedback on organizational issues, and perhaps supplement or replace regular departmental face-to-face meetings. If these efforts are not undertaken by change agents, a message is sent to other users that the system is not valuable, which is likely to have a negative impact on perceptions of relative advantage. If the instructor or manager makes use of the system, the opposite message is sent, increasing perceptions of relative advantage.

Providing users with an introduction to the groupware system may impact perceptions of complexity. For example, posting interesting, provocative topics to a discussion database may engross users in a discussion and help overcome any trepidation some users may have about their ability to use the system. Such playful use of an application may lead to learning that can be applied to other, more serious uses (Belanger & Van Slyke, 2000). Of course, providing training and technical assistance to users can also impact users' perceptions of complexity.

Perceptions of compatibility are also important determinants of use intentions. A consistent look and feel of the groupware system should be maintained for different applications. If a consistent look and feel is maintained across the applications, users with experience using one application of groupware are likely to feel that a new application is compatible with their past experience. Even something as simple as changing font styles or colors can have a negative impact on perceived compatibility.

Our study has confirmed that perceived innovation characteristics are important indicators of one's intention to use groupware technology. Further, our results also validate Moore and Benbasat's (1991) contention that the traditional innovation characteristic of observability is actually two constructs since result demonstrability is highly significant (p = 0.014), while visibility is clearly non-significant (p=0.601).

Interestingly, while we found perceived complexity to be a significant factor in predicting use, it had the weakest association with adoption intention. In the case of Domino, perceptions of complexity are less important than those of relative advantage, compatibility and result demonstrability. This result may be surprising to many who feel that user friendliness is *the* critical factor in determining use. The finding is particularly interesting in light of the much-studied technology acceptance model (TAM) (Davis, 1989). TAM in its basic form suggests that ease of use and usefulness determine intention to use. The results of this study suggest that the TAM may not be sufficiently detailed for some innovations. TAM does not consider other factors such as compatibility and result demonstrability. Using the richer diffusion of innovation theory as a framework allows researchers to better understand factors that influence groupware use.

FUTURE TRENDS

This research points to the need to manage the perceptions of potential groupware users to ensure successful groupware implementations. To this end, suggestions for influencing the perceptions of potential groupware users were offered in the previous section. This research also suggests other organizational factors that warrant examination in future groupware studies. For example, the degree to which the groupware software is available, reliable, and secure can have ramifications on the degree to which groupware technology is adopted. Future studies should examine the degree to which these infrastructure requirements impact the use of groupware systems.

Future studies should also explore the training aspects of groupware in more detail. While our research has clearly pointed to the need for technical training in order to decrease the perceptions of complexity of groupware use, other research has suggested that the *nature* of training in groupware situations should go beyond the technical aspects, and extend to "human" areas such as communications and diversity (Roebuck & Britt, 2002). These aspects of training should be explored further to see what impacts they have in groupware use, and in groupware effectiveness.

CONCLUSION

This article shows that groupware technology, while necessary for the success of virtual teams, needs to be appropriately implemented in organizations in order for intended users to adopt the technology. By focusing on improving team members' perceptions of the technology characteristics we have identified, more groupware systems can achieve widespread success in organizations.

While our study focused on the adoption of groupware in an educational setting, we feel the results of the study can be generalized to other settings where group communications are required or desired. In this research, student subjects were surveyed about a topic that had direct relevance to them (i.e., their use of Domino was directly related to their coursework). They were not asked to project themselves into an artificial role; they were asked their perceptions as students. Thus, it is reasonable to expect that any associations found between their perceptions and their adoption intentions are valid. Further, it should also be noted that many of the students in this study had work experience; the mean full-time work experience for the sample was 5.5 years. Yet statistical analysis indicated that the amount of work experience had little significance on their intentions to use the Domino groupware. Again, this lends support for the notion that the significant relationships found in this study between individuals' perceptions and intentions to use groupware would apply to those in actual working settings as well as educational settings.

In the future, the exact form and function-set of groupware is likely to change. We are already seeing this with the emergence of instant messaging software use in organizations. Whatever form the groupware of the future takes, one fact is likely to remain. Users' perceptions of the characteristics of groupware significantly impact its use. Those interested in promoting the use of groupware are well advised to carefully consider how the design and implementation of their systems might impact users' perceptions. Failing to do so may well lead to systems that go unused.

REFERENCES

Agarwal, R., & Prasad, J. (1997). Role of innovation characteristics and perceived voluntariness in the acceptance of information technologies. *Decision Sciences, 28*(3), 557-582.

Belanger, F., & Van Slyke, C. (2000). End-user learning through application play. *Information Technology, Learning and Performance Journal, 18*(1), 61-70.

Bell, B., & Kozlowski, S. (2002). A typology of virtual teams: Implications for effective leadership. *Group & Organization Management, 27*(1), 14-49.

Briggs, R., Adkins, M., Mittleman, D., Kruse, J., Miller, S., & Nunamaker, J. (1999). A technology transition model ferived from field investigation of GSS use aboard the U.S.S. Coronado. *Journal of Management Information Systems, 15*(3), 151-195.

Davis, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-340.

Day, J., Lou, H., & Van Slyke, C. (2004). Using Lotus Domino discussion databases to support collaborative, project-based learning. *International Journal of Distance Education Technologies* (forthcoming).

Johansen, R. (1988). *Groupware: Computer support for business teams*. New York: The Free Press.

Kelly, S., & Jones, M. (2001). Groupware and the social infrastructure of communication. *Communications of the ACM, 44*(12), 77-79.

Kezsbom, D. (2000). Creating teamwork in virtual teams. *Cost Engineering, 42*(10), 33-36.

Mahajan, V., Muller, E., & Bass, F.M. (1990). New product diffusion models in marketing: A review and directions for research. *Journal of Marketing, 54*(1), 1-26.

Moore, G., & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research, 2*(3), 192-222.

Orlikowski, W. (1993). Learning from Notes: Organizational issues in groupware implementation. *Information Society, 9*(3), 237-250.

Plouffe, C., Hulland, J., & Vendenbosch , M. (2001). Research report: Richness versus parsimony in modeling technology adoption decisions—Understanding merchant adoption of a smart card-based payment system. *Information Systems Research, 12*(2), 208-222.

Power, D. (2003). A brief history of decision support systems. *http://DSSResources.com/history/dsshistory.html*

Roebuck, D., & Britt, A. (2002). Virtual teaming has come to stay—Guidelines and strategies for success. *Southern Business Review, 28*(1), 29-39.

Rogers, E. (1995). *Diffusion of innovations.* New York: The Free Press.

Townsend, A.M., DeMarie, S., & Hendrickson, A.R. (1998) Virtual teams: Technology and the workplace of the future. *Academy of Management Executive, 12*(3), 17-28.

KEY TERMS

Adoption: The stage of technology diffusion in which an individual or organization decides to select a technology for use.

Compatibility: Degree to which an innovation is seen to be compatible with existing values, beliefs, experiences and needs of adopters.

Complexity: Degree to which an innovation is seen by the potential adopter as being relatively difficult to use and understand.

Diffusion: The spread of an innovation through a social system.

Groupware: Class of technologies that allow groups to communicate and coordinate activities. Typically network-driven software.

Relative Advantage: Degree to which an innovation is seen as being superior to its predecessor.

Result Demonstrability: Degree to which the results of using an innovation are perceived to be tangible.

Trialability: Perceptions of the degree to which an innovation can be used on a trial basis before confirmation of the adoption must occur.

Visibility: The perception of the actual visibility of the innovation itself as opposed to the visibility of outputs.

Voluntariness: Degree to which use of an innovation is perceived as being of free will.

User Spreadsheet Systems Development

Anders Avdic
Örebro University, Sweden

INTRODUCTION

In the early days of computers, expertise was needed not only to develop systems, but also to use them. As IT tools have become more powerful and user friendly, more and more people have been able to use computers and programs as tools when carrying out working tasks. Nowadays it is even possible for people without special IT training to develop information systems that only IT specialists could have done some years ago.

BACKGROUND

In this article the conditions and effects of user systems development (USD) using a spreadsheet program (SP) are discussed. USD is characterized as a sub-area of end-user computing. USD is performed by a user-developer, a person who acts both as a user and a systems developer. A typical feature of a user-developer is that he has a good knowledge of the business and the work related to the information system (IS) in question, which is called the user-developed application (UDA).

To a large extent USD is a question about learning. User-developed applications are often developed in order to learn and understand. In Figure 1 the difference between traditional systems development (TSD) (1) and USD (2) is outlined in order to demonstrate the nature of USD in contrast to TSD, since TSD is familiar to the IS community. To the IT specialist, knowledge about IS development tools (e.g., methods, program languages) (1a) is in primary focus when developing TISs (1c). This is the core of the user-developer's professional knowledge. Knowledge about business (1b) is of course essential, but not primary. To the user-developer knowledge about business (2a) is in primary focus and knowledge about IS development tools (2b) is just a means to accomplish business-oriented tasks, eventually by developing UDAs (2c). The IT specialist has access to knowledge about IS development tools that is hard to access for non-professionals. Some business knowledge is hard to access for the IT specialist, since this knowledge is not in the professional knowledge domain of the IT specialist. The user-developer on the other hand is the expert on business knowledge. His professionalism depends on his knowledge about business. No one can replace him in this matter. In order to perform USD, the user-developer needs some knowledge about IS development tools. It is not possible though to have access to as much knowledge about IS development tools as the IT specialist has.

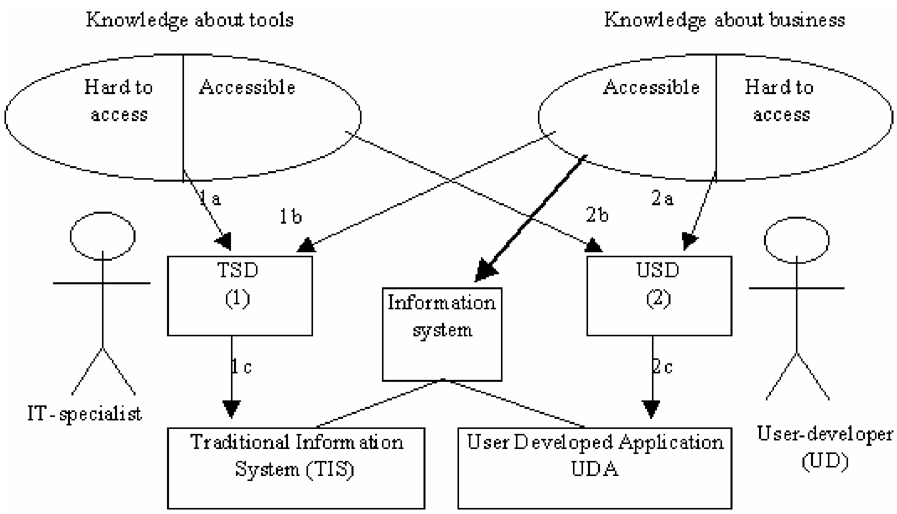
To both the IT specialist and the user-developer, both kinds of knowledge are to some degree necessary. In order to make an information system, the most important kind of knowledge is in general knowledge about business, since the information system is about the business. The thick arrow in Figure 1 demonstrates this circumstance.

In order to develop information systems, knowledge about business has to be transferred from business specialists to IT specialists. This transfer is problematic since people have different frames of references (Yourdon, 1989; Alter, 1996). The whole intention of the sender can therefore not be transferred to the IT specialist. The IT specialist cannot, on the other hand, fulfill the requirements since he cannot completely understand the business specialist. Complex systems development tasks still have to be performed through TSD, but as more powerful systems development tools are at hand, the possibilities to perform USD are enhanced from year to year. Spreadsheet programs have properties that give the user-developer access to IS development features without being an IT specialist. Of course there are other ways to overcome this gap, for example, by performing systems development with a participative approach like RAD (Tudhope, Beynon-Davies, Mackay & Slack, 2001). The systems discussed in this article are often small and local, and thereby often are not suitable for traditional systems development projects.

CONDITIONS AND EFFECTS OF USER SYSTEMS DEVELOPMENT

As a framework model, a modified version of the model of generic practice (the ToP model) (Goldkuhl & Röstlinger, 1999) is used to systemize empirical findings and related theory. The model can be used to specify the conditions and result of a specific practice, such as a controller practice or an IT specialist practice. The modified model consists of a set of conditional categories—*knowledge*, *norms*, and *tools*. The categories that express the specific practice are named *producers* (the user-developer) and their *actions* (user systems development). The last cat-

Figure 1. Relation between knowledge and development (Avdic, 1999)



egory is the *result* of the practice (the application). When a user-developer develops UDAs, he acts in at least two types of practices, the primary (e.g., controller) practice and the secondary (developer's) practice. Each practice is related to a profession, such as a controller and an IT specialist profession. The model makes it possible to separate the conditions of the different practices. It also makes it possible to discuss which parts of the developer's practice can improve the main practice without consulting an IT specialist. The use of the model makes it possible to show how different practices exchange conditions and effects. The result of user developer practice might, for example, be a condition of the controller practice. The model is described in Figure 2.

The ToP model is slightly related to the Work Systems model (Alter, 2002), in that the model focuses on practice without specific references to IT artifacts. The ToP model emphasizes knowledge aspects more explicitly, which makes it especially suitable to analyze the practice of user systems development. The nature of ToP model categories is described below.

Information Systems (Result)

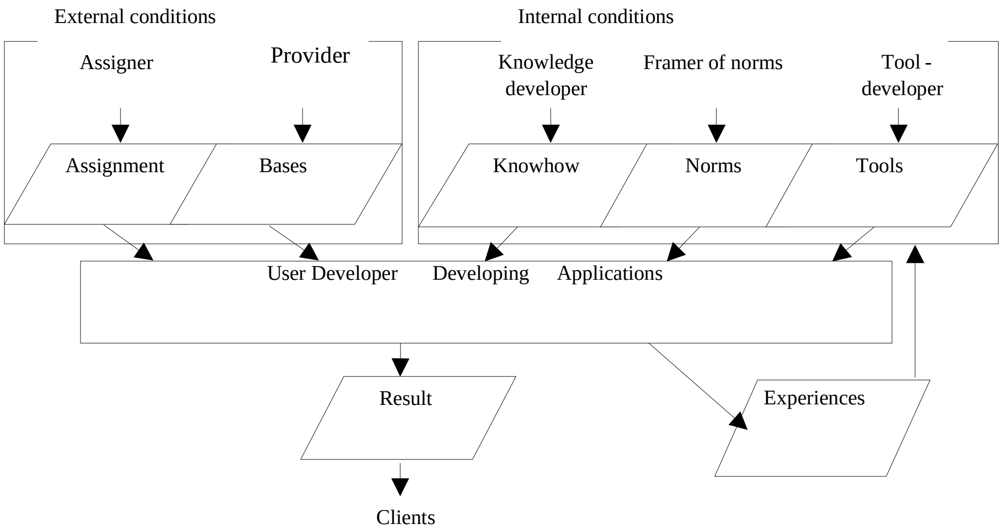
A UDA is an information system, and an information system is a result of systems development. The difference between a traditional information system (TIS) and a UDA is mainly a question of how it is built. UDAs are built by user-developers with a good knowledge of the business, while TISs are built by IT specialists.

User Systems Development (Actions)

Traditional systems development can be characterized by the notion of the 'Life Cycle', where tasks are specialized and activities are separated and systemized. USD and TSD are profoundly different in many ways. USD actions are seldom organized nor planned (Avdic, 1999). Specific work-related tasks or problems make the user-developer aware of some information need. USD is looked upon as work rather than systems development by the user-developer. From the user-developer's point of view, any tool is useful that might help him solve work-related problems. Compared to TSD, USD is characterized by integration rather than specialization. Where TSD professionals get specialized in programming, analysis, or database design, the user-developer integrates skills and performs the entire life cycle by himself.

Success factors of USD have been discussed in the scientific community for more than two decades. The reasons why USD is successfully adapted in an organization have been claimed to depend on the presence of informal channels of communication and how common training on USD tools is (Kruck, Maher & Barkhi, 2003; Brancheau & Brown, 1993). Basic conditions (suitable tasks, equipment, knowledge, and certain independence) must be fulfilled to make USD possible (Carlsson, 1993). If business and information needs are dynamic, USD can be justified. USD is appropriate when user-developers also have access to well-organized data, and get support from management and the IT department (Auer, 1998). Perceived importance is also claimed to be vital (Blili, Raymond & Rivard, 1998).

Figure 2. ToP model (Goldkuhl & Röstlinger, 1999)



When discussing how to manage and control USD, advocates of high control recommend (strict) organization of USD activities (e.g., Andersen, 1994). Advocates of low control consider USD as time saving and appropriate because of the lack of detailed monitoring (e.g., Speier & Brown, 1997).

The discussion of what factors determine successful USD is implicitly aimed at organizing USD activities with a certain degree of control. This discussion is not really relevant since the user-developers are professional in their respective professions. They use IT tools when they find it relevant in relation to their work tasks. Since they do not separate USD from running work, they have the same quality demands on the USD result as on the rest of their work. Contrary to some research (e.g., Teo & Tan, 1999), we claim that the risk of poor quality in UDA information output should be related to the user-developer's professionalism rather than to design methods or tool properties.

User-Developers (Producers)

A user-developer is a person with a good knowledge of the business who develops UDAs that support the user-developer in his work. The user-developer is primarily a professional (e.g., a controller) who integrates to some extent the role of one or more IT specialists when performing USD. The user-developer could have good knowledge about IS development tools. This does not disqualify him as a user-developer; it rather makes him even more efficient.

User-developers can be categorized in different ways. Commonly, categorization is done by knowledge of computing technologies. Rockart and Flannery (1983) and Govindarajulu (2003) are using the following types:

- Non-programming end-users, who are developing very simple applications and who are using a very limited set of tool functions.
- Command-level users, who are developing applications using declarative functions (e.g., formulas and functions).
- End-user programmers, who are developing applications that can be more complex (e.g., multi-user) applications using declarative as well as procedural functions.

With regard to the second and third types, the combination of deep knowledge about business and knowledge about IT tools makes it possible for the user-developer to perform systems development without engaging IT specialists.

Knowledge

When performing USD, knowledge is divided between the user-developer and the tool (SP). Certain kinds of (not too complex) knowledge are formalized into the SP and can be used in the SP-UDA. Other kinds can be formalized by the user-developer into the SP-UDA. Some kinds of knowledge (e.g., of critical evaluation of the relevance of formulas) cannot be formalized at all. Still, this kind of not easily formalized (sometimes tacit) knowledge can be taken into consideration when using the UDA, since the user-developer (with business knowledge) is the user of the system. We also claim that goals, not easily formalized, can be taken into consideration when performing USD.

Knowledge about tools can be used to deepen knowledge about business. User-developers can make tacit knowledge explicit when developing USD, which in turn makes it possible for others to evaluate and criticize the UDA and its output (Adam, Fahy & Murphy, 1998). The user-developers are conscious that an ongoing change in the company/authority's environment makes it important to develop not-yet-known knowledge about conditions and circumstances of their work. One important aim of the user-developer is to articulate knowledge about business and that the UDA is an important means to do this.

Norms

Norms and knowledge are closely related and sometimes hard to keep apart. One set of norms that are central is professional ethics. Professional ethics are crucial to the user-developer since the professionals' activities are monitored not by procedures, but by professional and business ethics. Professional ethics as well as professional tacit knowledge (see above) cannot easily be transferred to IT specialists in systems development projects. Therefore when USD is performed by user-developers, professional ethics and tacit knowledge can be taken into consideration in a way not possible in TSD. Ongoing questioning of business using UDAs can implicitly or explicitly challenge existing models as well as their norms.

Tools

USD tools are closely related to norms and knowledge, since norms and knowledge are implemented in tools. The main tool when performing SP-USD is of course the SP. The SP integrates functions for input, output, storage, processing, and presentation. This integration results in interactive development and use. The open nature of the SP can cause different kinds of errors (Panko & Sprague, 1998; Kreie, Cronan, Pendley & Renwick, 2000). Knowledge of business, tools, and design can prevent some of these errors. Lack of experience and training can also be a problem related to the tool (McGill, 2002).

FUTURE TRENDS

Since IT is becoming more and more common in different environments, it is likely that USD will also become more common. Rapid change in most aspects of everyday life and business contexts will make it more important to be able to deal with situations, questions, and specific information needs that are not anticipated in standard packages and systems developed years ago. The increased need of learning tools, eventually Web based, will therefore increase the need of flexible tools aimed at providing user-developers with learning possibilities. This way end-user computing and knowledge management will merge together and inspire future research.

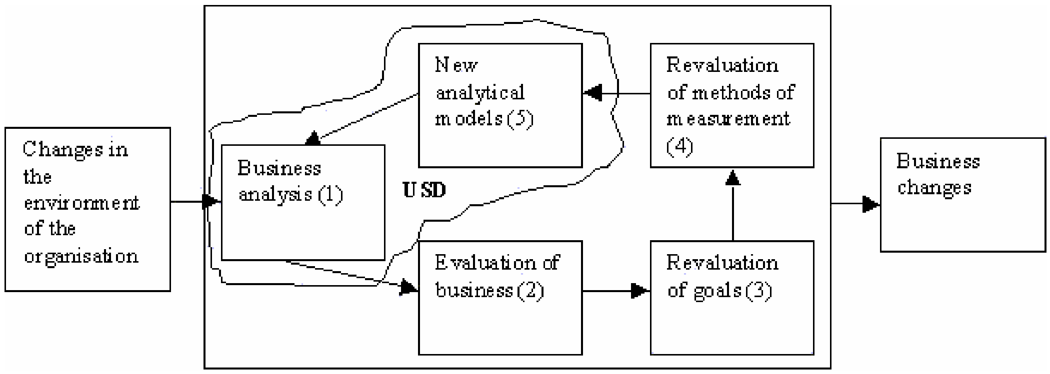
CONCLUSION

SP-USD is characterized by integration, interactivity, and capacity of questioning. The notion of integration can be looked upon in several dimensions: 1) aspects of ISs (integration of collecting, storing, processing, and distribution of information), 2) roles (integration of developer, user, and manager roles), 3) roles of actors in systems development (integration of analyst, programmer, database, and designer roles), and 4) integration of processing functions of the IS. The integrated nature of USD results in interactivity. Interactivity means that the user-developer can change quickly between developing and using the SP-UDA. During the USD process the user-developer knowledge of the business and USD increases. This is actually the goal of the user-developer. Since the user-developer's knowledge of the business increases when performing USD, the user-developer can analyze and also question aspects of business (e.g., production measuring methods). The questioning aspect makes it possible to improve business.

SP-USD can be used as a means of controlling continuous changes in the environment of the organization by changing business with the help of USD (see Figure 3). A business analysis (1) can result in a revaluation of the business (2), which can result in a revaluation of its goals (3) (and norms), which can result in a revaluation of methods of measurement (4), which can result in new analytical models (5) (UDA), which can lead to a new business analysis (1), and so on. USD is discussed as one way to meet change as a permanent business condition, which differs from traditional methods for systems development.

This way of reevaluating organizational goals can be related to double-loop learning, as presented by Argyris and Schön (1996). This includes not only changes in behavior or strategies. It means that norms of the organization can be changed. The ongoing questioning of business practice that is performed through USD can imply this form of norm changing.

Figure 3. Continuous change and user systems development



U

REFERENCES

Adam, F., Fahy, M. & Murphy, C. (1998). A framework for the classification of DSS usage in organizations. *Decision Support Systems*, *22*(1), 1-13.

Alter, S. (1996). *Information systems—a management perspective.* Menlo Park, CA: Benjamin/Cummings.

Alter, S. (2002). The work system method for understanding information systems and information systems research. *Communication of the Association for Information Systems*, *9*, 90-104.

Andersen, E.S. (1994). *Systemutveckling—principer, metoder och tekniker* (in Swedish). Lund: Studentlitteratur.

Argyris, C. & Schön, D.A. (1996). *Organizational learning II—theory, method, and practice.* Reading, MA: Addison-Wesley.

Auer, T. (1998). *Factors affecting end-user computing: Skills.* TUCS Technical Report No. 159, Department of Computer Science, University of Turku, Finland.

Avdic, A. (1999). *Användare och utvecklare—om anveckling med kalkylprogram* (in Swedish). Dissertation, Linköping University, Sweden.

Blili, S., Raymond, L. & Rivard, S. (1998). Impact of task uncertainty, end-user involvement, and competence on the success of end-user computing. *Information & Management*, *33*, 137-153.

Brancheau, J.C. & Brown, C.V. (1993). The management of end user computing: Status and directions. *ACM Computing Surveys*, *25*(4), 437-482.

Burnett, M., Cook, C., Pendse, O., Rothermel, G., Summet, J. & Wallace, C. (2003). End user software engineering with assertions in the spreadsheet paradigm.

Carlsson, S.A. (1993). *A longitudinal study of user-developed decision support systems.* Dissertation, Department of Informatics, Lunds University.

Goldkuhl, G. & Röstlinger, A. (1999). *Expanding the scope—from language action to generic practice.* CMTO Research Paper No. 1999:06, Linköping University, Sweden.

Govindarajulu, C. (2003). End users: Who are they? *Communications of the ACM*, *46*(9), 152-159.

Kreie, J., Cronan, T.P., Pendley, J. & Renwick, J.S. (2000). Applications development by end-users: Can quality be improved? *Decision Support Systems*, *29*, 143-152.

Kruck, S.E., Maher, J.J. & Barkhi, R. (2003). Framework for cognitive skill acquisition and spreadsheet training. *Journal of Organizational and End User Computing*, *15*(1), 20-37.

McGill, T.J. (2002). User developed applications: Can users assess quality? *Journal of End User Computing*, *14*(3), 1-15.

Panko, R.R. & Sprague Jr., R.H. (1998). Hitting the wall: Errors in developing and code inspecting a 'simple' spreadsheet model. *Decision Support Systems*, *22*, 337-353.

Rockart, J.F. & Flannery, L.S. (1983). The management of end-user computing. *Communications of the ACM*, *26*(10), 776-784.

Speier, C. & Brown, C. (1997). Differences in end-user computing support and control across user departments. *Information & Management*, *32,* 85-99.

Teo, T.S.H. & Tan, M. (1999). Spreadsheet development and 'what-if' analysis: Quantitative versus qualitative errors. *Accounting, Management and Information Technologies*, *9,* 141-160.

Tudhope, D., Beynon-Davies, P., Mackay, H. & Slack, R. (2001). Time and representational devices in Rapid Application Development. *Interacting with Computers*, *13,* 447-466.

Yourdon, E. (1989). *Modern structured analysis*. Englewood Cliffs, NJ: Prentice-Hall.

KEY TERMS

End-User Computing: "…the adoption and use of information technology by personnel outside the information systems department to develop software applications in support of organizational tasks" (Brancheau & Brown, 1993, p. 437).

Information System: A system consisting of functions for input, processing, storing, output, and presentation of information.

Spreadsheet Program: The most common standard package next to word processors, suitable for user systems development.

Systems Development: A process where an information system is developed.

Traditional Systems Development: Systems development carried out according to the 'life cycle' model.

User Systems Development: Systems development initiated and performed by user-developers, who have good knowledge about and who are (partly) responsible for (part of) the organization the system is to serve.

User-Developed Application: An information system developed by a user-developer. The system is often small and dedicated to a specific task in the user-developer's working environment.

User-Developer: A person who develops UDAs that support the user-developer in his work. The user-developer has deep (often tacit) knowledge about, and is often (partly) responsible for (part of) the organization the system is to serve.

Uses and Gratifications for the World Wide Web

Thomas F. Stafford
University of Memphis, USA

INTRODUCTION

It has always been important to understand what motivates consumers to use the Internet, since network use is a critical precursor to electronic commerce (Eighmey & McCord, 1998; Lohse & Spiller, 1998; Schonberg, Cofino, Hoch, Podlaseck & Spraragen, 2000). Without Internet use, there is no e-commerce use; this is much the same sort of issue that retailers study when trying to determine issues that motivate store visits. Internet use motivations, then, serve an important purpose in the understanding of e-commerce activities (Stafford, 2003b).

Technically speaking, the Internet has always been a communication medium, so we understand much of its functionality from the standpoint of its telecommunications utility. As we strive to understand motivations for its use, it is important to realize that the Internet is far more than a computer network; as the Internet evolves to become not only a telecommunications network, but also a consumer entertainment source, it gains the potential uses of familiar entertainment and communications media such as telephones, radio and television. The Internet remains a network for the distribution of information and telecommunications services, but it also becomes a channel for the delivery of other, richer media. In this sense, it has become a medium of media, or *meta-medium* (Stafford, Stafford & Shaw, 2002). Thus, where understanding the motivations related to computer use once was sufficient to diagnose Internet user motivations, we must now consider a wider range of potential uses and motivating gratifications arising from use of this complex and rich medium.

Early work on the uses and gratifications for Internet use centered on the process of using the network and the gratifications related to the content that the network provided. However, recent work indicates that additional motivations exist for Internet use, and these are important for site and service operators to understand, if they wish to successfully motivate customer use of and loyalty to their resource. These new motivations are potential differentiators between operators *within* the Internet medium as well as *between* the Internet and conventional promotional media (Stafford, Stafford & Shaw, 2002).

BACKGROUND

It has been known for decades that individuals are motivated in their use of media; it is not a random or undirected activity (Katz, 1959), though it may occasionally seem so when one encounters the typical “couch potato” channel surfer or aimless “Web surfer.” Audiences are not passive consumers of media; they get involved (Katz, Blumler & Gurevitch, 1974; Rubin, 1981), which means that understanding the motivations that lead to media use also provides the basis for engineering more involved media use – an outcome that is surely beneficial to media operators and their commercial sponsors.

Media choice, being motivated by individual uses and related goals – something we characterize as “gratifications” (Lin, 1977) – is much like any other product or service that might be marketed to users. Consumers make choices about what to use, and in their media choices, users are both active and selective in choice (Levy & Windahl, 1984). This means that media operators dare not assume a captive audience, and this assumption applies equally well in the mass media world of television and radio as it does in the online world of the Internet and its rich media sites. For that reason, the robust theoretical perspective of uses and gratifications, which has been used to study media since before the advent of commercial television, is useful in investigating the Internet since our media knowledge in one context is often useful in another (Eighmey, 1997; McDonald, 1997; Newhagen & Rafaeli, 1996).

Uses and gratifications theory (U&G) has to do with what people do with media and why they do it. Classic applications of U&G theory have consistently identified only two key areas of motivation for media use: media content uses and motivations for use (*gratifications*, in U&G parlance), and media usage process gratifications. Content gratifications concern the *messages* carried by the medium (which could be informative, or entertaining), and process gratifications concern *actual use* of the medium itself (Cutler & Danowski, 1980). The modern analogies would be the Web surfer, who is clearly motivated by the process of using the Internet, versus the highly focused online researcher, who is engaged in searches for very specific message content to support information needs (Stafford & Stafford, 2000).

U&G UP-TO-DATE

In research that has spanned the course of the past five years, an emerging stream of literature documenting new Internet-specific U&G studies are now reaching publication. Previous Internet-related U&G research was useful, but authors had generally adapted measurement scales directly from television studies of uses and gratifications (cf., Eighmey, 1997; Eighmey & McCord, 1998; Rafaeli, 1988). The emerging stream of research began with the premise that motivations for Internet use might be different from the motivations that drive the use of other media, and that a new set of measures ought to be developed (e.g., Stafford & Stafford, 1998). This article documents some of the more prominent findings in the process of developing these new Internet-specific U&G dimensions.

The Initial Factors of Internet U&G

Stafford and Stafford (2001a) leveraged a qualitative study with HotWired site users into an AOL user survey to investigate the dimensionality of uses and gratifications for the Internet. Five Internet-specific U&G factors were identified: Searching, Cognitive, New and Unique, Social, and Entertainment. While searching certainly speaks for itself as an Internet activity, the cognitive factor was interesting, since it was characterized by descriptors related to learning: education, information, learning and research. The "new and unique" factor was characterized by descriptors such as ideas, interesting, new, progressive and relaxing; clearly a gratification factor, as opposed to a specific use. The social factor (chatting, friends, interaction, newsgroups and people) identified an exciting new possibility in U&G applications for the Internet, since previous U&G research had never identified a social motivation for media use. The entertainment factor (entertainment, fun and games) was clearly a content-based gratification related to having fun with Internet site content.

Applications of New Internet U&G Factors

With the interpersonal social factor identified as an aspect of Internet usage motivation, the Internet can be considered as both an interpersonal and a mass exposure medium, with simultaneous commercial and noncommercial opportunities to users (Stafford & Stafford, 2001a). Stafford's (2001) confirmatory analysis of emerging Internet U&G issues produced U&G factors that were applied in a study of Internet use and diffusion among the consumer market (Stafford, 2003b), examining differences between early and late Internet adopters; there were significant differences between each adoption category on specific Internet U&G factors. Internet laggards exhibited the lowest social gratification for online services, while Internet "innovators" (or, early adopters) exhibited the highest social gratifications. Innovators also produced significantly higher content gratification scores, as compared to laggards.

It began to appear, as a process of developing and applying specific U&G dimensions for Internet users, that the impact of social gratifications might be largely a function of user experience. A general expectation emerged which specified that heavy Internet service users are more motivated by social gratifications than light users (Stafford, Stafford & Shaw, 2002). To the extent that heavy users are desirable consumer targets for ISPs, verification of this supposition could have valuable implications for practice.

This presumption was tested in another AOL survey (Stafford, 2003a), and for all of the variables comprising both process and social Internet usage gratifications, heavy users scored higher than light users. Hence, heavy users can be expected to be more motivated in their Internet use specifically on the social dimension – a finding that resonates with unrelated Internet research on user demographics (e.g., Emmannouildes & Hammond, 2000) – but also any other U&G dimension, as well. Heavier users have more specific uses and find more enjoyment in the Internet usage process than do light users, it seems.

Internet U&G factors have also been applied to understand Web use in the distance education classroom (Stafford, 2003c; Stafford & Stafford, 2003). In this study of student motivations for use of distance course Internet resources, social gratifications were dominant, accounting for 70% of the variance. It is generally agreed that students on the remote end of a distance education teleconference feel somewhat socially removed and isolated from colleagues in the live origination section of the class (cf., Berger & Topol, 2001; Hamer, 2001). Internet technology has been shown to ameliorate the inherent social isolation of a distance course, and that consideration seems to be confirmed in the strong showing of the social factor in this specific distance education application (Stafford, 2003c).

Since much of the initial development of U&G dimensions for Internet use came from samples of AOL users, one recent study made an effort to understand specific characterizations of AOL users for their Internet activities. AOL users appear to be highly gratified by activities characterized by Web browsing and the guided search for information, and AOL users also appreciate communications-related Internet uses, and have clear social gratifications for Internet use. Interestingly, among all the potential indicators related to U&G factors examined in analysis, shopping as a specific activity does *not* appear to be highly gratifying to AOL users (Stafford & Gonier, 2004).

Further, while AOL users are certainly socially motivated in their use of the Internet, mean scores for the "chatting" component of the social gratification dimension were below the scale midpoint, suggesting that the explicit social activity related to Instant Messenger, chatting, was not prized as highly as other social activities available online. Light AOL users produced a "chatting" mean of 3.2 on a 7-point scale, while heavy users produced a mean of 3.9 (Stafford, 2003a), which suggests that AOL users in general are not overly enamored with chatting as a gratification for Internet use. One interesting aspect of the AOL user study was the potential dimensionality of social gratifications; in performing an oblique factor rotation, the list of social variables factored into a distinct communication factor, aside from the standard socialization factor that had been seen in previous analyses (Stafford & Gonier, 2004).

The communication factor referenced terms related to using the Internet as an interpersonal communications channel: access, answers, communication and e-mail. The reduced socialization factor was distinguished by what would appear to be the targeted use of the Internet as a communication channel (that is, the focus on the message receiver): chatting, friends, fun, interaction, newsgroups, people, relaxing. Analysis of variance was performed on the salient variables for each usage factor, and results show that heavy users score higher than light users on each of these variables (Stafford & Gonier, 2004).

Relationships between Internet U&G Dimensions

In a measurement study seeking to establish trait validity for U&G factors (Stafford, 2001), process and content gratifications were found to be highly related ($\Phi = .72$), but there were much weaker relationships between the social factor and the more traditional process and content gratifications, which tends to reinforce the multi-modal conceptualization of the Internet as providing both personal entertainment and interpersonal communications channels in one meta-channel. The process dimension was only moderately related to social gratifications ($\Phi = .38$), and the content dimension was also only moderately related to social gratifications ($\Phi = .34$), so social gratifications are distinct from (and not closely related to) standard process and content gratifications for Internet use.

The social gratification construct did display excellent measurement qualities, indicated by measurement model fit indices (GFI = .97, AGFI = .95, RMSR = .11, SRMR = .043, NFI = .96, CFI = .97), so social gratifications appear to be a distinct trait and a valid area of motivation for Internet use, if not directly related to the better understood process and content gratifications.

CONCLUSION

Given that they are derived and validated from a broad base of users, Internet social gratifications are worth considering as a key benefit sought by consumers using the Internet. Unlike other media studied in the past, where sought content and enjoyment of media usage processes were the defining factors of motivation, the Internet provides standard media gratifications along with interpersonal connectivity, giving it a multi-modal appeal and influence in human life.

Interpersonal communication and interaction with other people over the Internet seems to characterize the social Internet gratification. Site operators as well as Internet service providers can enhance this experience for users through the inclusion of Webmail, chat utilities, message boards and discussion groups. AOL already makes great use of this social dimension of Internet use in the way it leverages its Instant Messenger utility as a promotional tool for building the membership base and increasing online time among existing member (Stafford, 2001).

Regardless of the use or gratification a business may wish to impact in its provision of Internet service and utility, the important point is to realize that users actively interact with the medium, and that uses and gratifications is one of the best ways that exists in which to study active audience motivations for media use. Determining what audience members seek to do, and the benefit they expect to accrue from activity, is the important step in marketing and supporting Internet offerings.

A highly practical implication of what we know about U&G, in a business context, is that the Internet service providers will have more difficulties recruiting and motivating late adopters. Heavy users and early adopters are the *most motivated* of users, so perhaps efforts in the ISP business should focus on customer retention and current customer satisfaction, as opposed to new customer recruitment – to the extent that many of the new customers to be had in the current environment are late adopters and likely light users, if they are not already making use of the Internet and its functionality. For the purposes of e-commerce businesses, the implications that heavy Internet users are more motivated by social and process/content gratifications means that e-commerce sites must be both easy to use and appeal to the heavy user population. But it also means that heavy users can respond well to social utilities provided by Internet sites, so chat rooms and message boards dedicated to product users at a commerce site may have some attraction for the heavy Internet users of commerce sites.

U&G also reminds us to be customer-centric in our design and provision of products and services, keeping in mind that users want specific things because these things bring them enjoyment. Goal oriented activity is the lynchpin of motivation, and understanding the goals sought by our Internet customers will allow us to design and provide more compelling and satisfying offerings. Inevitably, business objectives will be more fully attained in a customer-centric approach to the Internet industry.

REFERENCES

Berger, K.A., & Topol, M.T. (2001). Technology to enhance learning: Use of a Web site platform in traditional classes and distance learning. *Marketing Education Review, 11*(3), 15-26.

Cutler, N.E., & Danowski, J.A. (1980, Summer). Process gratification in aging cohorts. *Journalism Quarterly, 57,* 269-277.

Eighmey, J. (1997, May/June). Profiling user responses to commercial Web sites. *Journal of Advertising Research, 37,* 59-66.

Eighmey, J., & McCord, L. (1998). Adding value in the information age: Uses and gratifications of sites on the World Wide Web. *Journal of Business Research, 41,* 187-194.

Emmannouildes, C., & Hammond, K. (2000). Internet usage: Predictors of active users and frequency of use. *Journal of Interactive Marketing, 14,* 17-32.

Hamer, L.O. (2001). Distance learning technologies as facilitators of learning and learning-related student activities. *Marketing Education Review, 11*(3), 55-67.

Katz, E. (1959). Mass communication research and the study of popular culture: An editorial note on a possible future for this journal. *Studies in Public Communication, 2,* 1-6.

Katz, E., Blumler, J.G., & Gurevitch, M. (1974). Uses of mass communication by the individual. In W.P. Davison & F.T.C. Yu (Eds.), *Mass communication research: Major issues and future directions.* New York: Praeger.

Levy, M.R., & Windahl, S. (1984, January). Audience activity and gratifications: A conceptual clarification and exploration. *Communication Research, 11,* 51-78.

Lin, N. (1977). Communication effects: Review and commentary. In B. Rubin (Ed.), *Communication yearbook 1.* New Brunswick, NJ: Transaction Books.

Lohse, G.L., & Spiller, P. (1998). Electronic shopping. *Communications of the ACM, 41,* 81-87.

McDonald, S.C. (1997). The once and future Web: Scenarios for advertisers. *Journal of Advertising Research, 37*(2), 21-28.

McGuire, W.J. (1974). Psychological motives and communication gratifications. In J. Blumler & E. Katz (Eds.), *The uses of mass communications: Current practices on gratifications research.* Beverly Hills, CA: Sage Publications.

Newhagen, J., & Rafaeli, S. (1996). Why communication researchers should study the Internet: A dialogue. *Journal of Communication, 46*(1), 4-13.

Rafaeli, S. (1988). Interactivity: From new media to communication. In R. Hawkins, J. Wieman & S. Pingree (Eds.), *Advancing communication science: Merging mass and interpersonal processes.* Newberry Park, CA: Sage Publications.

Rubin, A.M. (1981, April). An examination of television viewing motivations. *Communication Research, 8,* 141-165.

Schonberg, E., Cofino, T., Hoch, R., Podlaseck, M., & Spraragen, S.L. (2000). Measuring success. *Communications of the ACM, 43,* 53-57.

Stafford, M.R., & Stafford, T.F. (2000). Identifying the uses and gratifications of Web use. *Proceedings of the 2000 American Academy of Advertising Conference,* Newport, RI.

Stafford, T.F. (1999). Consumer motivations for commercial Web site use: Antecedents to electronic commerce. *Proceedings of the Association for Information Systems 1999 Americas Conference on Information Systems,* Milwaukee, WI.

Stafford, T.F. (2001). *Motivations related to consumer use of online services.* Unpublished doctoral dissertation. University of Texas, Arlington.

Stafford, T.F. (2003a). Social and usage process motivations for Internet use: Differences between light and heavy users. *Proceedings of the 2003 Americas Conference for Information Systems,* San Diego, CA.

Stafford, T. F. (2003b). Differentiating between innovators and laggards in the uses and gratifications for Internet services. *IEEE Transactions on Engineering Management, 50,* 427- 435.

Stafford, T.F. (2003c). Understanding motivations for Internet use in distance education. *IEEE Transactions on Education*, forthcoming.

Stafford, T.F., & Gonier, D. (2004). Gratifications for Internet use: What Americans like about being online. *Communications of the ACM*, forthcoming.

Stafford, T.F., & Stafford, M.R. (1998). Uses and gratifications of the World Wide Web: A preliminary study. *Proceedings of the 1998 American Academy of Advertising Conference*, Lexington, KY.

Stafford, T.F., & Stafford, M.R. (2000). Consumer motivations to engage in electronic commerce: Uses and gratifications of the World Wide Web. In S. Rahman & M. Raisinghani (Eds.), *Electronic commerce: Opportunities and challenges*. Hershey, PA: Idea Group Publishing.

Stafford, T.F., & Stafford, M.R. (2001a, January-March). Identifying motivations for the use of commercial Web sites. *Information Resources Management Journal*, *14*, 22-30.

Stafford, T.F., & Stafford, M.R. (2001b). Investigating social motivations for Internet use. In O. Lee (Ed.), *Internet marketing research: Theory and practice*. Hershey, PA: Idea Group Publishing.

Stafford, T.F., & Stafford, M.R. (2003). Uses and gratifications for Internet use in the distance education classroom. *Proceedings of the 2003 American Marketing Association Winter Educators Conference*, Orlando, FL.

Stafford, T.F., Stafford, M.R., & Shaw, N. (2002). Motivations and perceptions related to the acceptance of convergent media delivered thought the World Wide Web. In M. Khosrow-Pour (Ed.), *Advanced topics in information resources management*. Hershey, PA: Idea Group Publishing.

KEY TERMS

Active Audience: Uses and gratifications theory presumes media users are actively involved in selection and use of media, and are not passive recipients. This implies the need to specifically target media offerings to perceived user needs.

Content Gratification: Enjoyment of message specifics. Content can mean information, and often does, though it also includes entertainment in the form of medium-carried programming.

Gratifications: What people derive from use…the "why" of media use motivations.

Meta-Medium: A channel of channels, such as the Internet. This term conveys the sense of rich and complex media transmission across a multiplexed channel of conveyance. Numerous motivations for use could arise related to such a complex media venue.

Process Gratification: Enjoyment of media use, as distinguished from enjoyment of specific message content. This is much the same as channel surfing or Web browsing with no goal other than entertainment through engaging in the activity.

Uses: Things people do *with* media…the "how" of media use motivations.

Uses and Gratifications: Customer activities and the enjoyment that derives from such activities, particularly in a mass media context.

Using Geographic Information Systems to Solve Community Problems

Jane L. Garb
Baystate Medical Center, USA

Richard B. Wait
Baystate Medical Center, USA

INTRODUCTION

This article describes how the technology of geographic information systems (GIS) can be used as a tool to integrate various types of community-level data to address local problems. The purpose of the article is to present an approach that can be replicated by others. This approach is based on community-wide collaborative sharing of resources, data, and research applications with an aim to enhance the health and well-being of the local area population. Although the example used relates to health, the approach can be used to deal with any "event" or series of events in the community.

BACKGROUND

A GIS is a set of hardware and software for inputting, storing, managing, displaying, and analyzing geographic or spatial data or any information that can be linked to geographic location, such as events, people, or environmental characteristics (Bailey & Gatrell, 1995; Burrough & McDonnell, 1998). The advantages of GIS as an information system pertain to its ability to handle spatial data; integrate data from many sources; uncover spatial patterns and relationships by superimposing different data layers and viewing data at different levels of aggregation; and conduct spatial analyses to test hypotheses and make predictions. Significant advances in the field of GIS and spatial statistics in the last 20 years have enabled researchers to undertake a more extensive examination of the spatial component of a wide variety of applications. The visualization and analytic capabilities of a GIS enable the user to examine and model the interrelationship between spatial and nonspatial etiologic factors in a variety of events ranging from house sales to crime activity to breast cancer.

The American Community Survey (ACS; 2002) is an ongoing survey conducted by the U.S. Census Bureau. Data are collected monthly and are used to provide annually adjusted estimates of the current population based on an approximate 2.5% sample, with oversampling for small governmental units, such as American Indian Reservations. Full implementation of the survey will begin in 2004 (Census, 2003). ACS data forms the source of population data for the current study.

Despite the fact that ACS does not provide complete population counts, it can be used to derive community profiles, emphasizing relative proportions in each population subgroup. It can also be used as a mechanism for measuring changes and trends in the population in the interval between decennial census counts.

The Health Geographics Program at Baystate Medical Center in Springfield, Massachusetts, takes a comprehensive approach to the implementation of services, research, and community applications utilizing GIS. Because of its collaboration, giving access to community and hospital data, the program was contracted by the U.S. Census Bureau to undertake two case studies to demonstrate the utility of annually adjusted data from the American Community Survey.

The case study described here utilizes the ACS population and housing data in a GIS to improve breast cancer intervention programs. Previous research has clearly established that by lowering the rate of late-stage disease with increased mammography screening, breast cancer mortality can be reduced (Feig, 1988; Marchant, 1994). Furthermore, there is evidence that socioeconomic and cultural disparities in breast cancer screening exist (Katz, Zemencuk, & Hofer, 2000; O'Malley, Kerner, Johnson, & Mandelblatt, 1999; Phillips, Kerlikowske, Baker, Chang, & Brown, 1998). Several investigators have applied GIS and spatial analysis in the past to identify etiologic factors in breast cancer and late-stage breast cancer (Brody et al., 1997; Gardner, Joyce, & Melly, 1999; Kulldorff, Feuer, Miller, & Freedman, 1997; Lewis-Michl et al., 1996; Melly, Joyce, Maxwell, & Brody, 1997; Roche, Skinner, & Weinstein, 2002; Selvin, Merrill, Erdmann, White, & Ragland, 1998; Sheehan et al., 2000; Timander & McLafferty, 1998).

THE GIS APPROACH

The main functions of a GIS are *data integration*, *visualization*, *exploration*, *statistical analysis*, and *modeling* (Bailey & Gatrell, 1995). These functions can be combined in a systematic approach to solve problems. This approach can be outlined as follows:

1. **Integrate** data from multiple sources
2. **Visualize** the data with maps
3. **Explore** patterns further with spatial statistics
4. **Generate** hypotheses
5. **Test** hypotheses with mathematical modeling

We will describe how this approach was applied to the case study referred to above.

CASE STUDY: INVESTIGATING LATE-STAGE BREAST CANCER

The aim of the study was to create a profile of communities in Springfield in need of increased breast cancer screening. Specifically, we wanted to identify parts of the city with high rates of late-stage disease as well as identify socioeconomic and demographic factors in late-stage disease. This information would aid resource allocation by focusing intervention efforts on high-risk areas. Furthermore, it would allow the design of "culturally appropriate" (Healthy People, 2000) screening programs.

Applying the GIS approach, the first step was to gather and *integrate data* from three different sources. Geographic data were obtained from the City of Springfield Planning Department. This consisted of geographic boundaries and street locations that would be used in the geocoding process described later. ACS housing and population data provided aggregate information on demographic and socioeconomic characteristics of women over 40 by police sector (Table 1). Police sectors were used for this case study, because this was the smallest geographic unit for which ACS data were available. There are nine police sectors in Springfield. Breast cancer case data from the Springfield's two hospital oncology registries gave information on the dates and stages at diagnosis and home addresses for all patients diagnosed at these two hospitals. Together, these registries capture 95% of all cases of breast cancer in the city.

Cases were staged according to the American Joint Cancer Committee (AJCC; AJCC, 1997). Cases were defined as "late stage" if they were Stage 2 or greater. This definition captures all cases that should have been detected earlier had mammography been performed. A total of 891 breast cancer cases were diagnosed during 1995–1999, with 194 of these defined as late stage.

All data for the study were converted to dBASE IV format to be read into the GIS. Geographic data was in the form of ArcView (ESRI, 1999) shapefiles. Geographic and tabular data were imported into the GIS using ArcView. ArcView was used for all GIS functions. Geocoding of case locations was based on the patient's street address, using the City of Springfield streets shapefile as the reference database. Mapping of case locations was based on patient's home address. Mapping was done at a small enough scale so that individual patient addresses could not be determined from the map, in order to preserve patient confidentiality.

The next step in the GIS approach was to visualize the data. This was done by mapping the data in ArcView. We

Table 1. ACS variables

Variable	Universe
Proportion for each race (white, black, etc.)	Women > age 40
Proportion Hispanic	
U. S.-born, foreign-born, Puerto Rican born	
Naturalized citizens, noncitizens	
Linguistically isolated	
Married, unmarried	
Unemployed	
Employing public transportation to work	
High school diploma	
Below poverty level (12.5K)	
Using food stamps	
Receiving public assistance	
Median income	Households
Vacant	Housing
Median value	

Figure 1. Dot map depicting the locations of late-stage breast cancer cases and mammography facilities in Springfield, Massachusetts

chose a dot map to display the geographic distribution of cases of late-stage disease and indicate areas of concentration. This would show where raw numbers of cases were greatest and more resources were needed, e.g., where mobile mammogram units and educational or other intervention programs would be likely to have the highest yield. Visual examination of the map in Figure 1 reveals no apparent clustering.

Clustering can be formally tested in the next step, *exploring patterns* further with spatial statistics. In this study, we used the spatial scan statistic (Kulldorff et al., 1997; Kulldorff, Rand, Gherman, Williams, & DeFrancesco, 1998). This technique employs a systematic, iterative "searching" of space conceptually with overlapping circles, counting cases and people, and revealing areas of unusually high rates, or clusters. The software used to perform this analysis was SATScan, a free, publicly domain program (Kulldorff, Rand, Gherman, & DeFrancesco, 1998). Preliminary findings from this study showed no statistically significant spatial clustering of cases. From the results of this visual and statistical examination, we concluded that rather than concentrating resources in specific parts of the city, a more global approach to intervention seemed warranted.

We were also interested in whether the location of mammogram facilities influenced the rate of late-stage disease in the different police sectors. This has been shown to be a factor in screening (Athas & Amir-Fazli, 2000). Using ArcView, we overlaid a map of their locations onto the dot density map of cases (Figure 1). Although case location did not appear related to mammography facility location, we would later test this relationship with formal statistical modeling.

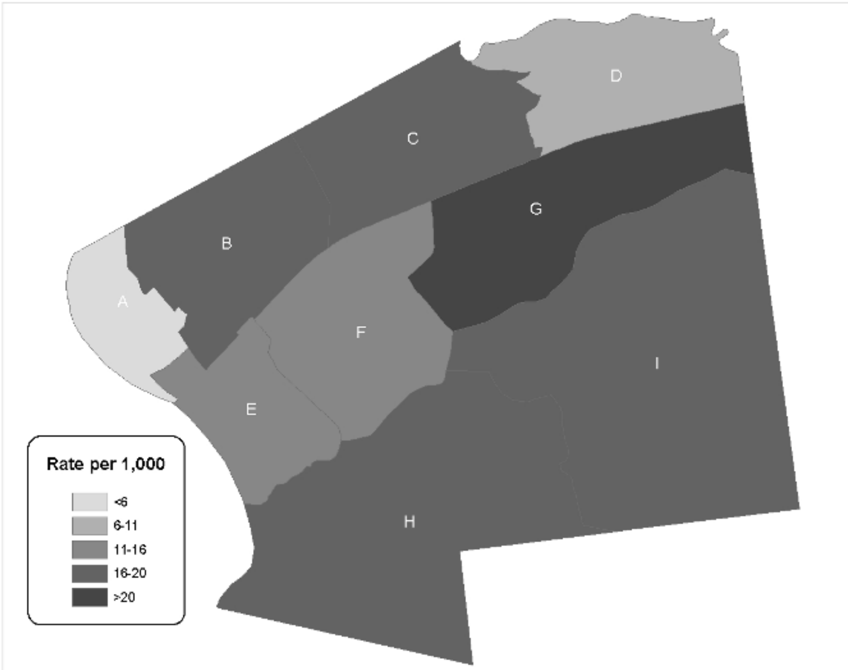
To estimate the risk of getting late-stage disease, case data were aggregated or summed to the police-sector level for which population estimates were available. The total number of cases of late-stage breast cancer for each sector were divided by the ACS population estimates for the total number of women over 40 (the population at risk for breast cancer) for that sector. This represented the prevalence rate of late-stage disease.

We mapped these rates in the GIS using choropleth maps to look for areas with unusually high rates. A choropleth map is used to display quantities for various geographic areas. Figure 2 shows the rate of late-stage breast cancer per 1,000 women over 40 in each police sector. Darker shades indicate police sectors with higher rates of late-stage disease. According to this map, Sector G has the highest rate. We needed to investigate what characteristics of this sector make women at higher risk for late-stage disease.

To *generate hypotheses* about etiologic factors in late-stage disease, the next step in the GIS approach, we compared the choropleth map of late-stage disease by

U

Figure 2. Choropleth map showing the rate of late-stage breast cancer per 1,000 women over 40 (darker shades indicate higher rates)



sector, with maps showing the distribution of various demographic and socioeconomic sector characteristics. Pattern similarities would indicate a possible relationship between these characteristics and the risk of late-stage breast cancer. Figure 3 shows the proportion of women over 40 who are married, receive public assistance, have a high school diploma, or are African American. Darker shades on the maps represent higher concentrations of these populations, and vise versa. Similar maps (not shown) were generated for other ACS demographic and socioeconomic variables listed in Table 1. The sector with the highest rate of late-stage disease (Sector G) had among the highest rates of African American or married women or those with a high school diploma or receiving public assistance. This generated the hypothesis that socioeconomic and demographic factors are related to late-stage disease.

The final step in the GIS approach was to use *mathematical modeling to test hypotheses* about cause–effect relationships. To accomplish this, a spatial regression was conducted to identify sector demographic and socioeconomic characteristics that made women at particularly high or low risk for late-stage disease. Spatial regression was conducted using the S-PLUS for ArcView extension (Insightful, 2000). This technique is helpful in sorting out the effect of various factors not necessarily apparent by mapping, and identifying characteristics of high-risk population subgroups. The unit of analysis for the regression analysis was police sector. The dependent variable was the rate of late-stage disease per thousand women over 40. Independent variables were the ACS estimates for sector population and housing characteristics from Table 1, as well as the number of mammogram facilities.

Four factors were predictive of late-stage disease: African American race, presence of a high school diploma, and foreign born or married status in women over 40. These factors were all positively related to the rate of late-stage disease: the risk of a woman's getting late-stage disease was greater in sectors with higher rates of African American, married, or foreign-born women, or those with a high school diploma. The number of mammogram facilities in a sector did not affect the rate of late-stage disease in that sector.

The results of this study can help in the allocation of resources and the design of culturally appropriate intervention programs. Areas of the city with high concentrations of cases or populations at highest risk of late-stage disease (e.g., African Americans or foreign born) can be targeted for increased resources for screening intervention such as education programs or mobile mammogram units. The fact that high school graduation was a risk factor in late-stage disease indicates that program materials should be geared to a higher educational level so as not to discourage more educated women from participat-

Figure 3. Rates of women over age 40 in Springfield, Massachusetts, according to various socioeconomic and demographic factors from the American Community Survey (darker shades indicate higher rates).

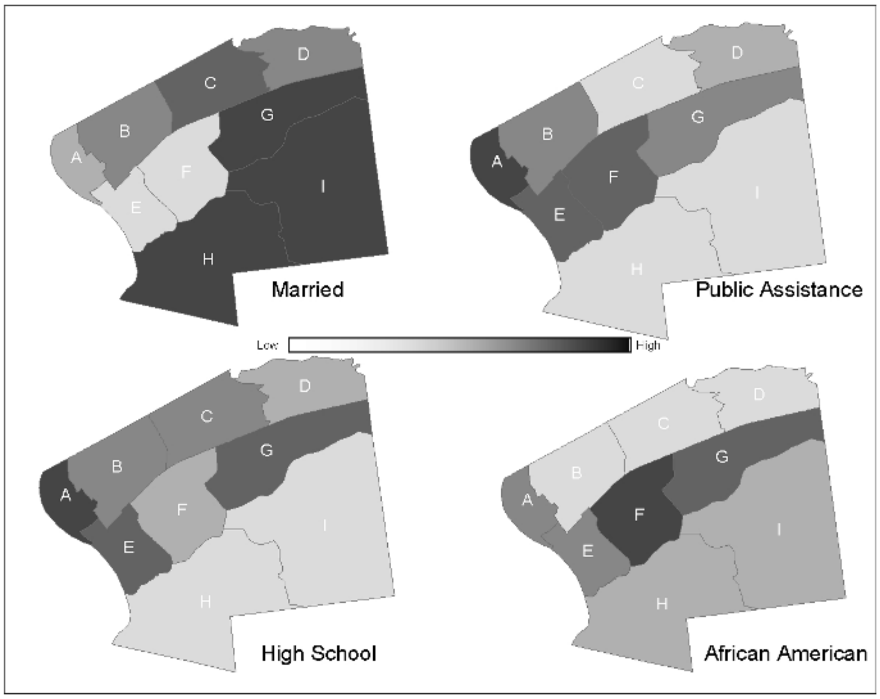
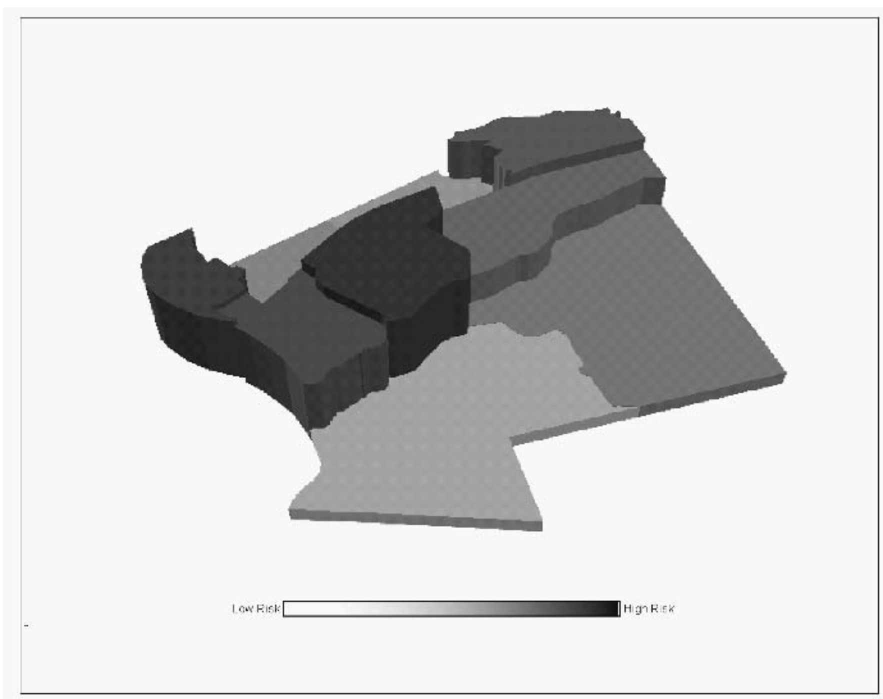


Figure 4. Three-dimensional map of locations of high-risk populations for late-stage breast cancer in Springfield, Massachusetts (darker and taller sectors are higher risk)



ing. The importance of foreign birth suggests potential avenues for reaching high-risk women through various national cultural organizations, such as the Greek Cultural Council in Springfield. The increased risk for African American women suggests working with African American organizations, many of which already target specific health problems.

Although the ACS data used are only preliminary estimates based on one year of survey data on a relatively small sample of the population, they demonstrate the importance of socioeconomic factors in the geographic distribution of late-stage disease, which is relevant to the design of intervention programs.

FUTURE TRENDS

With future advances in GIS technology, many of the functions of database management (such as importing data, joining tables, creating new variables, calculating new fields) as well as visualization (mapping) and analytic functions can be automated, and the entire process can be saved in the GIS for future use with updated or entirely different data sets. The ability to model the process described in this paper, then edit, save, and recall it is a new feature in the latest version of ArcGIS (ESRI, 2004).

Although not yet fully implemented throughout the country, the American Community Survey holds the most promise as a source of up-to-date U.S. population estimates for the type of application we have described. Data used in the present study were a special tabulation provided by the American Community Survey as part of our contract. However, the first data products for the smallest areas and populations throughout the United States will be available publicly in 2009. Once data products are produced for a population group or area, they will be updated each year (Census, 2003). The ACS will completely replace the Long Form of the next U.S. Decennial Census in 2010.

CONCLUSION

This article demonstrates how each of the main GIS functions outlined by Bailey and Gatrell can be implemented in an approach to address a particular community health problem. Geographic and population data can be used in a GIS to create a visual demographic profile of communities by providing denominators for the calculation and visualization of risks (incidence or prevalence rates), identifying the size and location of high-risk populations, and providing information on socioeconomic and demographic characteristics that can be used in analysis to identify risk factors for late-stage disease. It facilitates the design of culturally appropriate intervention and prevention programs. It enables the creation of community demographic profiles by identifying the size and geographic location of high-risk population subgroups. This is a tremendous aid for public health or other community agency planning and resource allocation.

The approach described here can be applied with any census geography (tracts, block groups, etc.) for which ACS or other population estimates (e.g., the Decennial Census) are available. This enhances its usefulness for a number of purposes and geographic scales. The broad utility of the ACS in providing accurate and timely data on the population and the economic environment in which it exists make it an essential tool in health care as well as other current community issues, such as homeland security. As ACS implementation continues to expand throughout the country, these benefits will become widely available to state and local public officials and agencies. Hopefully, this article will stimulate them to utilize these data in significant ways to improve and protect the health and well-being of our population.

REFERENCES

AJCC. (1997). Breast. In *AJCC Cancer Staging Manual* (5th ed., pp. 171–180). Philadelphia, PA: Lippincott-Raven Publishers.

Athas, W. F., & Amir-Fazli, A. (2000, March 17–19). *Geographic variation in breast cancer stage of disease at diagnosis.* Paper presented at the Second International Health Geographics Conference, Chevy Chase, MD.

Bailey, T., & Gatrell, A. C. (1995). *Interactive spatial data analysis*. New York: Crown Books.

Brody, J., Foord, K., Maxwell, N., Melly, S., Polk, R., Rudel, R., et al. (1997). *Cape Cod breast cancer and environment study: Results of the first 3 years of study*. Boston, MA: Boston University School of Public Health.

Burrough, P. A., & McDonnell, R. A. (1998). *Principles of geographic information systems*. Oxford: Oxford University Press.

Census. (2002). *Meeting 21st century demographic data needs implementing the American Community Survey: May 2002*. Washington, D.C.: U.S. Census Bureau.

Census. (2003). *American Community Survey Operations Plan*. Washington, D.C.: U.S. Census Bureau.

ESRI. (1999). ArcView GIS (Version 3.2). Redlands, CA: Environmental Systems Research Institute, Inc.

ESRI. (2004). ArcGIS (Version 9.0). Redlands, CA: ESRI.

Feig, S. (1988). Decreased breast cancer mortality through mammographic screening: Results of clinical trials. *Radiology, 167*, 659–665.

Gardner, J., Joyce, Y., & Melly, S. (1999). GIS as a tool for a breast cancer and environmental study on Cape Cod, Massachusetts.

HHS. (2000). *Healthy People 2010*. Washington, D.C.: U.S. Department of Health and Human Services.

Insightful. (2000). S-Plus 2000 [Windows]. Seattle, WA: Insightful Corporation.

Katz, S. J., Zemencuk, J. K., & Hofer, T. P. (2000). Breast cancer screening in the United States and Canada, 1994: Socioeconomic gradients persist. *Am. J. Public Health, 90*(5), 799–803.

Kulldorff, M., Feuer, E., Miller, B., & Freedman, L. (1997). Breast cancer clusters in the Northeast United States: A geographic analysis. *Am. J. of Epidem., 146*(2), 161–170.

Kulldorff, M., Rand, K., Gherman, G., & DeFrancesco, D. (1998). SaTScan: Software for the spatial and space–time scan statistics (Version 2.1). Bethesda, MD: National Cancer Institute.

Kulldorff, M., Rand, K., Gherman, G., Williams, G., & DeFrancesco, D. (1998). SaTScan: Software for the spatial and space–time scan statistics. (Version 2.1). Bethesda, MD: National Cancer Institute.

Lewis-Michl, E. L., Melius, J. M., Kallenbach, L. R., Ju, C. L., Talbot, T. O., Orr, M. F., et al. (1996). Breast cancer risk and residence near industry or traffic in Nassau and Suffolk Counties, Long Island, New York. *Arch. Environ. Health, 51*(4), 255–265.

Marchant, D. J. (1994). Contemporary management of breast cancer. *Obstet. Gynecol. Clin. North Am., 21*(4), 555–560.

Melly, S., Joyce, Y., Maxwell, N., & Brody, J. (1997). *Investigating breast cancer and the environment using a geographic information system*. Massachusetts Department of Public Health.

O'Malley, A. S., Kerner, J., Johnson, A. E., & Mandelblatt, J. (1999). Acculturation and breast cancer screening among Hispanic women in New York City. *Am. J. Public Health, 89*(2), 219–227.

Phillips, K. A., Kerlikowske, K., Baker, L. C., Chang, S. W., & Brown, M. L. (1998). Factors associated with women's adherence to mammography screening guidelines. *Health Serv. Res., 33*(1), 29–53.

Roche, L., Skinner, R., & Weinstein, R. (2002). Use of geographic information system to identify and characterize areas with high proportions of distant stage breast cancer. *J. Public Health Management Practice*.

Selvin, S., Merrill, D. W., Erdmann, C., White, M., & Ragland, K. (1998). Breast cancer detection: maps of two San Francisco Bay area counties. *Am. J. Public Health, 88*(8), 1186–1192.

Sheehan, T. J., Gershman, S. T., MacDougall, L. A., Danly, R. A., Mroszczyk, M., Sorensen, A. M., et al. (2000). Geographic assessment of breast cancer screening by towns, zip codes, and census tracts. *J. Public Health Management Practice, 6*(6), 48–57.

Timander, L. M., & McLafferty, S. (1998). Breast cancer in West Islip, NY: A spatial clustering analysis with covariates. *Social Science and Medicine, 46*(12), 1623–1635.

KEY TERMS

ACS (American Community Survey): An ongoing survey conducted by the U.S. Census Bureau that collects detailed demographic and socioeconomic information on a sample of the population.

Choropleth Map: A color-coded map, also called a "thematic" map, in which geographic areas are portrayed in different hues or intensities according to their values on some quantities.

Culturally Appropriate: Refers to an unbiased attitude in organizational policy that values cultural diversity in the population served. Reflects an understanding of the diverse attitudes, beliefs, behaviors, practices, and communication patterns that could be due to race, ethnicity, religion, socioeconomic status, historical and social context, physical or mental ability, age, gender, sexual orientation, or generational and acculturation status. Includes awareness that cultural differences may affect health, and includes the effectiveness of health care delivery. Knowledge of disease prevalence in specific cultural populations, whether defined by race, ethnicity, socioeconomic status, physical or mental ability, gender, sexual orientation, age, disability, or habits (Healthy People, 2000).

Dot Map: A map in which the geographic locations of events, people, or other entities are depicted as points or dots.

Geocoding: A function of the GIS through which the geographic location of an address is given a set of geographic coordinates by reference to a standard geographically referenced database. These coordinates are then used for mapping.

Geographic Information System (GIS): A computer-based set of tools for capturing (collecting), editing, storing, integrating, analyzing, and displaying spatially referenced data (Bailey & Gatrell, 1995).

Late-Stage Breast Cancer: Breast cancer in an advanced stage, usually defined as involvement of regional lymph nodes or size larger than 2 cm.

Spatial Regression: A spatial analytic technique modeling the relationship of various factors to the geographical distribution of some attribute measured on a continuous scale.

Spatial Scan Statistic: A cluster detection statistic that uses a moving window to compare the number of events or case locations inside versus outside the window. This statistic can identify geographic clustering of disease cases or geographic areas with unusually high or low rates of disease.

Video Content-Based Retrieval Techniques

Waleed E. Farag
Zagazig University, Egypt

INTRODUCTION

Recently, multimedia applications are undergoing explosive growth due to the monotonic increase in the available processing power and bandwidth. This incurs the generation of large amounts of media data that need to be effectively and efficiently organized and stored. While these applications generate and use vast amounts of multimedia data, the technologies for organizing and searching them are still in their infancy. These data are usually stored in multimedia archives utilizing search engines to enable users to retrieve the required information.

Searching a repository of data is a well-known important task whose effectiveness determines, in general, the success or failure in obtaining the required information. A valuable experience that has been gained by the explosion of the Web is that the usefulness of vast repositories of digital information is limited by the effectiveness of the access methods. In a nutshell, the above statement emphasizes the great importance of providing effective search techniques. For alphanumeric databases, many portals (Baldwin, 2000) such as *google, yahoo, msn,* and *excite* have become widely accessible via the Web. These search engines provide their users keyword-based search models in order to access the stored information, but the inaccurate search results of these search engines is a known drawback.

For multimedia data, describing unstructured information (such as video) using textual terms is not an effective solution because the information cannot be uniquely described by a number of statements. That is mainly due to the fact that human opinions vary from one person to another (Ahanger & Little, 1996), so that two persons may describe a single image with totally different statements. Therefore, the highly unstructured nature of multimedia data renders keyword-based search techniques inadequate. Video streams are considered the most complex form of multimedia data because they contain almost all other forms, such as images and audio, in addition to their inherent temporal dimension.

One promising solution that enables searching multimedia data, in general, and video data in particular is the concept of content-based search and retrieval (Deb, 2004). The basic idea is to access video data by their contents, for example, using one of the visual content features. Realizing the importance of content-based searching, researchers have started investigating the issue and proposing creative solutions. Most of the proposed video indexing and retrieval prototypes have the following two major phases (Flinkner et al., 1995):

- **Database population phase** consisting of the following steps:
- **Shot boundary detection.** The purpose of this step is to partition a video stream into a set of meaningful and manageable segments (Idris & Panchanathan, 1997), which then serve as the basic units for indexing.
- **Key frames selection.** This step attempts to summarize the information in each shot by selecting representative frames that capture the salient characteristics of that shot.
- **Extracting low-level features from key frames.** During this step, some of the low-level spatial features (color, texture, etc.) are extracted in order to be used as indices to key frames and hence to shots. Temporal features (e.g., object motion) are used too.
- **The retrieval phase.** In this stage, a query is presented to the system that in turns performs similarity matching operations and returns similar data (if found) back to the user.

In this article, each of the these stages will be reviewed and expounded. Moreover, background, current research directions, and outstanding problems will be discussed.

VIDEO SHOT BOUNDARY DETECTION

The first step in indexing video databases (to facilitate efficient access) is to analyze the stored video streams. Video analysis can be classified into two stages: shot boundary detection and key frames extraction (Rui, Huang & Mchrotra, 1998a). The purpose of the first stage is to partition a video stream into a set of meaningful and manageable segments, whereas the second stage aims to abstract each shot using one or more representative frames.

In general, successive frames (still pictures) in motion pictures bear great similarity among themselves, but this generalization is not true at boundaries of shots. A shot is a series of frames taken by using one camera. A frame at a boundary point of a shot differs in background and content from its successive frame that belongs to the next shot. In a nutshell, two frames at a boundary point will differ significantly as a result of switching from one camera to another, and this is the basic principle that most automatic algorithms for detecting scene changes depend upon.

Due to the huge amount of data contained in video streams, almost all of them are transmitted and stored in compressed format. While there are large numbers of algorithms for compressing digital video, the MPEG format (Mitchell, Pennebaker, Fogg & LeGall, 1997) is the most famous one and the current international standard. In MPEG, spatial compression is achieved through the use of a DCT-based (Discrete Cosine Transform-based) algorithm similar to the one used in the JPEG standard. In this algorithm, each frame is divided into a number of blocks (8x8 pixel), then the DCT transformation is applied to these blocks. The produced coefficients are then quantized and entropy encoded, a technique that achieves the actual compression of the data. On the other side, temporal compression is accomplished using a motion compensation technique that depends on the similarity between successive frames on video streams. Basically, this technique codes the first picture of a video stream (I frame) without reference to neighboring frames, while successive pictures (P or B frames) are generally coded as differences to that reference frame(s). Considering the large amount of processing power required in the manipulation of raw digital video, it becomes a real advantage to work directly upon compressed data and avoid the need to decompress video streams before manipulating them.

A number of research techniques was proposed to perform the shot segmentation task such as template matching, histogram comparison, block-based comparison, statistical models, knowledge-based approach, the use of AC coefficients, the use of motion vectors, and the use of supervised learning systems (Farag & Abdel-Wahab, 2001a, 2001c).

KEY FRAMES SELECTION

The second stage in most video analysis systems is the process of KFs (Key Frames) selection (Rui, Huang & Mchrotra, 1998) that aims to abstract the whole shot using one frame or more. Ideally, we need to select the minimal set of KFs that can faithfully represent each shot. KFs are the most important frames in a shot since they may be used to represent the shot in the browsing system, as well as be used as access points. Moreover, one advantage of representing each shot by a set of frames is the reduction in the computation burden required by any content analysis system to perform similarity matching on a frame-by-frame basis, as will be discussed later. KFs selection is one of the active areas of research in visual information retrieval, and a quick review of some proposed approaches follows.

Clustering algorithms are proposed to divide a shot into M clusters, then choose the frame that is closest to the cluster centroid as a KF. An illumination invariant approach is proposed that applies the color constancy feature to KFs production using hierarchical clustering. The VCR system (Farag & Abdel-Wahab, 2001b, 2001c) uses two algorithms to select KFs (AFS and ALD). The AFS is a dynamic adapted algorithm that uses two levels of threshold adaptation—one based on the input dimension, and the second relying upon a shot activity criterion to further improve the performance and reliability of the selection. AFS employs the accumulated frame summation of luminance differences of DC frames. The second algorithm, ALD, uses absolute luminance difference and employs a statistical criterion for the shot-by-shot adaptation level, the second one.

FEATURE EXTRACTION

To facilitate access to large video databases, the stored data need to be organized; a straightforward way to do such organization is the use of index structures. In case of video databases we even need multi-dimension index structures to account for the multiple features used in indexing. Moreover, we are in need of tools to automatically or semi-automatically extract these indices for proper annotation of video content. Bearing in mind that each type of video has its own characteristics, we also need to use multiple descriptive criteria in order to capture all of these characteristics.

The task of the feature extraction stage is to derive descriptive indexes from selected key frames in order to represent them, then use the indexes as metadata. Any further similarity matching operations will be performed over these indexes and not over the original key frames data. Ideally, content-based retrieval (CBR) of video should be accomplished based on automatic extraction of content semantics that is very difficult. Thus, most of the current techniques only check the presence of semantic primitives or calculate low-level visual features. There are mainly two major trends in the research community to extract indices for proper video indexing and annotation. The first one tries to automatically extract these indices, while the second trend performs iconic annotation of video by manually (with human help) associating icons to parts of the video stream. One example of the latter trend uses a multi-layered representation to perform the annotation task, where each layer represents a different view of video content. On the other hand, works on the first trend, automatic extraction of content indices, can be divided into three categories:

- *Deriving indices for visual elements using image-indexing techniques.* For example, using the color and texture as low-level indices.
- *Extracting indices for camera motion* (panning, zooming, etc.). Generally, optical flow is used in such techniques.
- *Deriving indices for region/object motion.* One system detects major objects/regions within the frames using optical flow techniques.

Color and texture are commonly used indexing features in most of the above systems. Color feature extraction can work directly on the original decoded video frame or on its DC form. One technique converts the color space of DC video frames (YCbCr in case of MPEG) to the traditional RGB color space, then derives color histograms from the RGB space. Deriving the histogram can be done in many ways. An efficient technique uses some of the most significant bits of each color component to form a codeword. For instance, the most significant two bits of each color component are selected and concatenated to form a 6-bit codeword. This codeword forms a 64-bin color histogram that is used as the color feature vector. This histogram is a good compromise between computational efficiency and representation accuracy.

Many techniques are proposed in the literature to perform texture feature extraction. Some of them use auto-regression and stochastic models, while others use power spectrum and wavelet transform (Bimbo, 1999). The main disadvantage of these techniques is that they are computationally expensive.

THE RETRIEVAL SYSTEM

The basic objective of any automated video indexing system is to provide the user with easy-to-use and effective mechanisms to access the required information. For that reason, the success of a content-based video access system is mainly measured by the effectiveness of its retrieval phase. The general query model adapted by almost all multimedia retrieval systems is the QBE (Query By Example) (Yoshitaka & Ichikawa, 1999). In this model, the user submits a query in the form of an image or a video clip (in case of a video retrieval system) and asks the system to retrieve similar data. QBE is considered to be a promising technique, since it provides the user with an intuitive way of query presentation. In addition, the form of expressing a query condition is close to that of the data to be evaluated.

Upon the reception of the submitted query, the retrieval stage analyzes it to extract a set of features, then performs the task of similarity matching. In the latter task, the query-extracted features are compared to the features stored into the metadata, then matches are sorted and displayed back to the user based on how close a hit is to the input query. A central issue here is how the similarity matching operations are performed and based on what criteria (Farag & Abdel-Wahab, 2003d). This central theme has a crucial impact on the effectiveness and applicability of the retrieval system.

Many techniques have been proposed by various researchers in order to improve the quality, efficiency, and robustness of the retrieval system. Some of these techniques are listed below.

- *Relevance feedback.* In this technique the user can associate a score to each of the returned clips, and this score is used to direct the following search phase and improve its results.
- *Clustering of stored data.* Media data are grouped into a number of clusters in order to improve the performance of the similarity matching.
- *Use of linear constraints.* To come up with better formal definitions of multimedia data similarity, some researchers proposed the use of linear constraints that are based upon the instant-based-point-formalism.
- *Improving browsing capabilities.* Using KFs and mosaic pictures to allow easier and more effective browsing.
- *Use of time alignment constraints.* The application of this technique can reduce the task of measuring video similarity to finding the path with minimum cost in a lattice. The latter task can be accomplished using dynamic programming techniques.
- *Optimizing similarity measure.* By defining optimized formulas to measure video similarity instead of using the exhaustive similarity technique in which every frame in the query is compared with all the frames into the database (a computationally prohibitive technique).
- *Human-based similarity criteria.* This technique tries to implement some factors that humans most probably use to measure the similarity of video data.

Most of the above techniques end up with calculating the similarity between two frames. Below is one example of how similarity between the colors of two frames (*I* and *M*) represented by their color histograms can be calculated using the normalized histogram intersection. If *Sim* approaches zero, the frames are dissimilar, and if it is near one, the frames are similar in color.

$$Sim(I, M) = \frac{\sum_{i=1}^{n} \min(I_i, M_i)}{\sum_{j=1}^{n} M_j}$$

FUTURE TRENDS

Some of the important issues under investigation are surveyed briefly in this section. Efficient algorithms need to be proposed to parse video streams containing gradual transition effects. Detecting semantic objects inside video frames is also an open challenge that needs to be addressed. Moreover, multi-dimension indexing structure is one of the active areas of research that can be further explored. Determining the similarity of video data is another area of research that requires more exploration to come up with more approaches that are effective.

CONCLUSION

In this article, we briefly review the need and significance of video content-based retrieval systems and explain their four basic building stages. Current research topics and future work directions are covered too. The first stage, the shot boundary detection, divides video streams into their constituent shots. Each shot is then represented by one or more key frame(s) in a process known as key frames selection. The feature extraction stage, the third one, derives descriptive indexes such as color, texture, shapes, and so forth from selected key frames, and stores these feature vectors as metadata. Finally, the retrieval system accepts a user query, compares indexes derived from the submitted query with those stored into the metadata, and then returns search results sorted according to the degree of similarity to the query. At the end, we need to emphasize the importance of content-based video retrieval systems and assert that there is still a considerably large number of open issues that require further research to achieve more efficient and robust indexing and retrieval systems.

REFERENCES

Ahanger, G., & Little, T. (1996). A survey of technologies for parsing and indexing digital video. *Journal of Visual Communication and Image Representation*, *7*(1), 28-43.

Baldwin, R. (2000). Portals on edge. *IEEE Multimedia*, *7*(1), 10-11.

Bimbo, A. (1999). *Visual information retrieval*. San Francisco: Morgan Kaufmann.

Deb, S. (2004). *Multimedia systems and content-based retrieval*. Hershey, PA: Idea Group Publishing.

Farag, W., & Abdel-Wahab, H. (2001). A new paradigm for detecting scene changes on MPEG compressed videos. *Proceedings of the IEEE International Symposium on Signal Processing and Information Technology* (pp. 153-158).

Farag, W., & Abdel-Wahab, H. (2002). A new paradigm for analysis of MPEG compressed videos. *Journal of Network and Computer Applications*, *25*(2), 109-127.

Farag, W., & Abdel-Wahab, H. (2002). Adaptive key frames selection algorithms for summarizing video data. *Proceedings of the 6th Joint Conference on Information Sciences* (pp. 1017-1020).

Farag, W., & Abdel-Wahab, H. (2003). A human-based technique for measuring video data similarity. *Proceedings of the 8th IEEE International Symposium on Computers and Communications* (ISCC'2003) (pp. 769-774).

Flinkner, M. et al. (1995). Query by image and video content: The QBIC system. *IEEE Computer*, *28*(9), 23-32.

Idris, F., & Panchanathan, S. (1997). Review of image and video indexing techniques. *Journal of Visual Communication and Image Representation*, *8*(2), 146-166.

Mitchell, J., Pennebaker, W., Fogg, C., & LeGall, D. (1997). *MPEG video: Compression standard*. Chapman and Hall.

Rui, Y., Huang, T., & Mchrotra, S. (1998). Browsing and retrieving video content in a unified framework. *Proceedings of the IEEE Workshop on Multimedia Signal Processing* (pp. 9-14).

Yoshitaka, A., & Ichikawa, T. (1999). A survey on content-based retrieval for multimedia databases. *IEEE Transactions on Knowledge and Data Engineering*, *11*(1), 81-93.

KEY TERMS

Content-Based Access: A technique that enables searching multimedia databases based on the content of the medium itself and not based on keywords description.

Key Frames Selection: The selection of a set of representative frames to abstract video shots. Key frames (KFs) are the most important frames in a shot, so that they can represent the shot in both browsing and similarity matching operations, as well as be used as access points.

Query by Example: A technique to query multimedia databases where the user submits a sample query and asks the system for similar items.

Retrieval Stage: The last stage in a content-based retrieval system where the extracted features from the query are compared to those stored in the metadata and matches are returned to the user.

Shot Boundary Detection: A process with the objective of partitioning a video stream into a set of meaningful and manageable segments.

Video Indexing: The selection of indices derived from the content of the video to help organize video data and metadata that represents the original video stream.

Video Shot: A sequence of contiguous video frames taken using the same camera.

V

Virtual Communities of Practice

Chris Kimble
University of York, UK

Paul Hildreth
K-Now International Ltd., UK

INTRODUCTION

When knowledge management (KM) began to emerge in the 1990s it was seen as an innovative solution to the problems of managing knowledge in a competitive and increasingly internationalised business environment. However, in practice it was often little more than information management re-badged (Wilson, 2002). More recently, there has been recognition of the importance of more subtle, softer types of knowledge that need to be shared. This raises the question as to how this sort of knowledge might be managed. Communities of practice (CoPs) have been identified as means by which this type of knowledge can be nurtured, shared and sustained (Hildreth & Kimble, 2002). Do CoPs offer a means of managing the softer aspects of knowledge and, if they do, are they applicable to today's increasingly "virtual" world?

BACKGROUND TO COMMUNITIES OF PRACTICE

The term communities of practice (CoPs) was coined in 1991 when Jean Lave and Etienne Wenger used it in their exploration of situated learning (Lave & Wenger, 1991). Although the examples they used (non-drinking alcoholics, Goa tailors, quartermasters, butchers and Yucatan midwives) were all based on what might be broadly termed an apprenticeship model, the concept of a CoP is not restricted to this form of learning.

Lave and Wenger (1991) saw the acquisition of knowledge as a social process in which people participated in communal learning at different levels depending on their authority or seniority in the group, that is, whether they were a newcomer to the group or had been an active member for some time. The process by which a newcomer learns by being situated in the group was central to their notion of a CoP; they termed this process legitimate peripheral participation (LPP).

LPP is both complex and composite; legitimation, peripherality and participation are each indispensable in defining the other. Legitimation is concerned with power and authority relations in the community but is not necessarily formalised. Peripherality is not a physical concept or a measure of acquired knowledge, but concerned with the degree of engagement with the community. Participation is engagement in an activity where the participants have a shared understanding of what it means in their lives.

For Lave and Wenger (1991), the community and participation in it were inseparable from the practice. Being a member of a CoP implied participation in an activity where participants have a common understanding about what was being done and what it meant for their lives and their community. Thus, it would appear that CoPs with their concentration on situated learning and the exchange of understanding might be well suited to the management of the softer aspects of knowledge: but can this idea be applied to the business world?

EXTENSIONS TO THE COMMUNITY OF PRACTICE CONCEPT

Interest in CoPs continued to grow throughout the 1990s and several attempts were made to re-define Lave and Wenger's (1991) original model to encompass new areas such as communities of circumstance, communities of interest and communities of purpose. In particular, several attempts were made to re-define CoPs in a way that was more relevant to the commercial environment (e.g., Seely Brown & Duguid 1991, 1996; Stewart 1996). One of the most popular work related definitions of a CoP was offered by John Seely Brown and Estee Solomon Gray in their 1995 article called "The People Are the Company":

"At the simplest level, they are a small group of people … who've worked together over a period of time. Not a team not a task force not necessarily an authorised or identified group … they are peers in the execution of "real work". What holds them together is a common sense of purpose and a real need to know what each other knows" (Brown & Gray, 1995).

In 1998, Wenger (1998) published the results of an ethnographic study of a claims processing unit in a large insurance company that described how employees exchanged knowledge during meetings and by the passing of handwritten notes. He proposed a view of the company not as a single community, but as a constellation of interrelated CoPs. CoPs arise out of the need to accomplish particular tasks and can provide learning avenues that exist within, between and outside organisations. CoPs are formed through mutual engagement in a joint enterprise and will share a repertoire of common resources (e.g., routines, procedures, artefacts, vocabulary) that members develop over time.

Thus, according to Wenger (1998) a CoP becomes defined in terms of:

- What it is about:

The particular area of activity/body of knowledge around which it has organized itself. It is a joint enterprise in as much as it is understood and continually renegotiated by its members.

- How it functions:

People become members of a CoP through shared practices; they are linked to each other through their involvement in certain common activities. It is this mutual engagement that binds its members together in a single social entity.

- What it produces:

The members of a CoP build up a "shared repertoire" of communal resources over time. Written files are a more explicit aspect of this, although less tangible aspects such as procedures, policies, rituals and idioms can also included.

Wenger (1998) also identified two key processes at work in CoPs: participation and reification. He described participation as:

"... the social experience of living in the world in terms of membership in social communities and active involvement in social enterprises" (Wenger, 1998, p. 55)

and reification as:

"... the process of giving form to our experience by producing objects that congeal this experience into thingness" (Wenger, 1998, p. 58)

Wenger emphasises that like LPP, participation and reification are analytically separable, but are inseparable in reality. Participation is the process through which people become active participants in the practice of a community and reification gives concrete form to the community's experience by producing artefacts. One is meaningless without the other and vice versa. In day-to-day work, people both negotiate meaning through participation in shared activities and project that meaning onto the external world through the production of artefacts.

Wenger's (1998) work with CoPs shows that the concept can be applied in a business setting. Since then, several other authors have identified the business benefits of CoPs (e.g., Fontaine & Millen, 2004: Lesser & Storck, 2001). However, almost all of the previous work on CoPs has described co-located communities. With the increasing globalisation of business and the heavy reliance on information and communication technology (ICT), the next question is "Can CoPs continue to operate in a modern business environment?"; that is, "Can a CoP be virtual?"

FUTURE TRENDS

Concerning the future of CoPs, and virtual CoPs in particular, two main issues must be considered. The first concerns the relationship between a CoP and its wider (electronic) environment; the second concerns the nature of the "work" that CoPs do; that is, do processes in a virtual CoP differ from one that is co-located?

CoPs in an Electronic Environment

Internet-based networking technologies, which can provide a single platform for groups or networks of groups to form within larger organisations, have led to the development of various forms of virtual groups and communities. Seely Brown and Duguid (2000) coined the phrase "networks of practice" (NoPs) to describe one type of virtual group. NoPs are composed of people who are geographically separate and may never even get to know each other, but who share similar work or interests. Thus, NoPs are organised more at the individual level and based on personal social networks than CoPs with their notions of mutuality and the collective social will of the community.

In a study of job seeking activity, Granovetter (1973) introduced the notion of strong and weak social ties. In terms of the previous description, CoPs are characterised by strong social ties, whereas NoPs are characterised by weak social ties. Within a wider network consisting of weak ties, an individual may act as a "local bridge" or broker that enables the network to react more quickly and provide a coordinated response. Nevertheless, within a network there is also a need for strong ties to encourage local cohesion and avoid fragmentation that would make knowledge sharing and the adoption of innovation more difficult.

CoPs can be seen in the role of hub for the wider network, providing a more tightly knit sub-network that serves as knowledge generating centres for the larger NoPs. CoPs can act as bridges drawing together different groups and combining knowledge in new ways. They can also provide the access points for individuals to engage with the wider network and to establish a local identity within the larger organisation. Previous research has shown that the most common distributed form of a "virtual" CoP has a co-located active core (Hildreth, Kimble & Wright, 1998), which tends to support this view of distributed working.

A more recent example was provided by Lundkvist's (2004) study of customer networks as sources of innovation. This case study was generated from a long-term study of the Cisco Systems newsgroup, which identified user networks as peripheral and yet vital sites of innovation. In this case, the co-located core of the network was provided by a group of university technicians.

Work in Virtual CoPs

How might the balance between reification and participation be maintained in virtual working? This issue was addressed in an earlier paper (Kimble, Hildreth & Wright, 2000) where we described how a geographically distributed CoP managed both hard (reified) and soft (social) knowledge. In this case, the CoP was made up of four members co-located in the UK, a group of five members in the USA and one member in Japan.

In this situation, it might have been expected that sustaining participation would be more difficult and therefore reification would play a greater role. However, the findings of the case study showed that this was not necessarily the case. Shared artefacts, such as a planning document, did play an important role but the importance of social relationships remained paramount. While the group was able to sustain itself using e-media, it was still dependent on the development of relationships in the physical environment through face-to-face meetings.

It is interesting to observe how artefacts such as a planning document (reification) were used not only as ways of projecting knowledge from within the CoP but were also instrumental in the process of creating it (participation). The document stimulated discussion, problem solving, innovation and further participation. It was used both to drive meetings and as the focus of meetings. During discussions based on the document, new and innovative ideas would be triggered that could form the basis for new projects. Thus, as well as acting as a stimulus for innovation, the document acted as a catalyst leading to further participation.

A similar account can be found in Bradshaw, Powell and Terrell (2004) that describes how a team of remote workers developed into a CoP. They describe not only how the group deploys a variety of technologies to maintain contact but also the efforts that went into building commitment, ownership, engagement and focus in the group. In this case, the members of the group were all engaged in collaborative research. Writing about their work and presenting papers for peer- review was seen as a key factor in maintaining cohesion and developing the community's shared understanding of goals, development of knowledge and sense of belonging.

CONCLUSION

Reporting on a recent case study of how CoPs translate to a geographically distributed international environment, Hildreth (2003) throws further light on a number of these issues. The study examines the work of an internationally distributed CoP that spans three continents. In particular, it highlights the role that shared artefacts play in the process of creating, sharing and sustaining both types of knowledge and highlights the role that the creation of artefacts plays in enabling and sustaining participation in CoPs. Hildreth (2003) observes that the process of creating the artefact and regular (although not necessarily frequent) face-to-face contact are instrumental in maintaining the relationships that allow a CoP to function successfully in a virtual environment. Thus, paradoxically, it appears that one of the keys to a successful virtual CoP is an occasional, non-virtual, face-to-face meeting.

However, the changes that are sweeping the corporate infrastructure mean that increasingly workers find themselves forced into one or another form of virtual working. Instead of inhabiting a world of fixed roles with easy access to co-located resources, today's workers are increasingly based in a world of weak ties where resources are only obtained through personal and individual relationships. Rather than being embraced by a collective CoP, workers often find themselves functioning as individuals and building up networks, one contact at a time. Again, paradoxically, as social networks such as NoPs become more important, the fundamental unit for many examples of virtual working is not the group but the individual. This is not to say that collective groups such as CoPs and teams have ceased to exist but simply that the difficulty of building and maintaining the strong social ties needed to build a sense of community in a virtual environment should not be underestimated.

REFERENCES

Bradshaw, P., Powell, S., & Terrell, I. (2004). Building a community of practice: Technological and social implications for a distributed team. In P. Hildreth & C. Kimble (Eds), *Knowledge networks: Innovation through communities of practice* (pp. 184-201). Hershey, PA: Idea Group Publishing.

Fontaine, M.A., & Millen, D.R. (2004). Understanding the benefits and impact of communities of practice. In P. Hildreth & C. Kimble (Eds), *Knowledge networks: Innovation through communities of practice* (pp. 1-13). Hershey, PA: Idea Group Publishing.

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology, 78*(6), 1360-1380.

Hildreth, P. (2003). *Going virtual: Distributed communities of practice.* Hershey, PA: Idea Group Publishing.

Hildreth, P., & Kimble, C. (2002). The duality of knowledge. *Information Research, 8*(1). Paper no. 142. *http://InformationR.net/ir/8-1/paper142.html*

Hildreth, P., Kimble, C., & Wright, P. (1998, March). Computer mediated communications and communities of practice. *Proceedings of Ethicomp'98,* Erasmus University, the Netherlands (pp. 275 – 286).

Kimble, C., Hildreth, P., & Wright, P. (2000). Communities of practice: Going virtual. In Y. Malhotra (Ed.), *Knowledge management and business model innovation* (pp. 220-234). Hershey, PA: Idea Group Publishing.

Lave, J., & Wenger, E. (1991). *Situated learning. Legitimate peripheral participation.* Cambridge: Cambridge University Press.

Lesser, E.L., & Storck, J. (2001). Communities of practice and organizational performance. *IBM Systems Journal, 40*(4), 831 – 841. Retrieved from *http://researchweb.watson.ibm.com/journal/sj/404/lesser.pdf*

Lundkvist, A. (2004). User networks as sources of innovation. In P. Hildreth & C. Kimble (Eds.), *Knowledge networks: Innovation through communities of practice* (pp. 96-105). Hershey, PA: Idea Group Publishing.

Seely Brown, J., & Duguid, P. (1991). Organizational learning and communities of practice. *Organization Science, 2*(1), 40-57.

Seely Brown, J., & Duguid, P. (1996). Universities in the digital age. *Change*, 11-19.

Seely Brown, J., & Duguid, P. (2000). *The social life of information.* Boston, MA: Harvard Business School Press.

Seely Brown, J., & Solomon Gray, E. (1995). The people are the company. *Fast Company.* Retrieved February 10, 2004, from *http://www.fastcompany.com/online/01/people.html*

Stewart, T.A. (1996). The invisible key to success. *Fortune Online.* Retrieved October 4, 1996, from *http://pathfnder.com/@@V3AagAUAZyqOEYKS/fortune/magazine/1996.960805/edg.html*

Wenger, E. (1998). *Communities of practice. Learning, meaning and identity.* CUP.

Wilson, T.D. (2002). The nonsense of 'knowledge management'. *Information Research, 8*(1). Paper no. 144. Retrieved from *http://InformationR.net/ir/8-1/paper144.html*

KEY TERMS

Artefact: An artefact in the context of CoPs indicates objects, articles, and "things" which have been created by the CoP to assist the members in their work and which may have some of the community's knowledge embedded in them. Artefacts do not have to be concrete – a process or procedure may be an artefact.

Communities of Circumstance: Communities of circumstance are driven by position, circumstance or life experiences. Communities of circumstance are distinguished from CoPs in that they tend to be personally focused and are often built around "life stages," such as teenagehood, university, marriage or parenthood.

Communities of Interest: Communities of interest are groups of people who share a common interest. Members exchange ideas and thoughts about the given interest, but may know little about each other outside of this area. Participation in a community of interest can be compelling and entertaining but is not focussed on learning in the same way as a CoP.

Communities of Practice: Communities of practice are groups of people who have a common goal and who are internally motivated to reach the goal. The members have some form of common background and shared language.

Communities of Purpose: Communities of purpose form around people who are to achieve a similar objective. Such communities only serve a functional purpose. Members of the community can assist each other by sharing experiences, suggesting strategies and exchanging information on the process in hand.

Hard Knowledge: Hard knowledge is unambiguous and unequivocal, can be clearly and fully expressed, can be formalised and structured, can be "owned" without being used and is both abstract and static: it is about, but not in, the world.

Knowledge Management: Knowledge management is the means whereby an organisation "manages" and leverages its knowledge resources. This can include reports, databases and patents; it also includes people – identifying experts, sharing knowledge, and helping people learn.

Legitimate Peripheral Participation: LPP is the process by which a newcomer gradually works his/her way towards full participation in the community. Lave and Wenger's (1991) examples were based on the apprenticeship model, where a newcomer (the apprentice) was allowed to undertake basic tasks. As they became more experienced, they were given more complicated tasks until they could fully participate in the practice of the community and became old-timers.

Network of Practice: People who are not directly connected to each other but still engage in similar kinds of activities are said to belong to a network of practice (NoP). NoPs link local communities whose members have similar interests and give a minimal coherence to the network.

Soft Knowledge: Soft knowledge is implicit and unstructured, cannot be articulated, can be understood without being openly expressed, is associated with action and cannot be possessed, is about what we do and is acquired through experience.

Virtual Organization in the Human Mind

Jinyoul Lee
State University of New York at Binghamton, USA

INTRODUCTION

Virtual organization has been well documented as both a tool for organizations to seek further profitability through the removal of traditional barriers as well as a method to extend the provision of services to clientele in a manner previously achievable by only large, multinational corporations (Markus, 2000). The widespread implementation of information technology and its many applications in modern business have moved the act of management toward a virtual focus, where managers are able to complete tasks through the use of teams in varying physical locations, with members that may or may not be employees of that firm, sharing a wide variety of data and information. With so many companies now employing virtual organization techniques, referring to a company as "virtual" or to its components as possessing "virtuality" lacks the clarity and specificity needed when using these firms as examples for others. The variety of methods through which a firm can achieve virtuality represents a span nearly as wide as the business community itself.

BACKGROUND

The earliest definitions of a virtual organization appeared when the concept of virtuality was applied to studies of management before information technology existed in a refined state to support the theory. Giuliano (1982) saw that with the addition of telecommunications and networking technology, there was little need for work teams to assemble at the same time or even a contiguous location. A structured concept of virtual organization was formed by Abbe Mowshowitz (1994, 2002). He defined a virtual organization in previous work as a group of employees communicating and being managed electronically through *metamanagement*. This concept defines the way in which a *virtual task* is managed and further categorizes a virtual organization as a series of virtual tasks completed by virtual teams in strategic global locations. As each team has a certain commitment to the parent organization, the similarity in purpose and communication style allows for clear distribution of work across multiple groups of organizational members.

As with Net-enabled organizations, the concept of virtual organizations has gained prominence among researchers and practitioners. As shown by the recent work of Schultze and Orlikowski (2001), virtuality can be understood through the perception of time and space. This paper extends the scope of the virtual organization in terms of *virtual space*, a metaphor used in time and space (beyond the constraints of the actual location we belong to) dimensions (Allcorn, 1997). As opposed to the virtual organization, time and space dimensions are constrained in traditional or "real" organizations. Time constraints occur in real organizations due to the operational time dimension of such organizations, while the space dimension occurs due to constraints of location.

It is true that a virtual organization inherits the attributes of virtual dimensions—a newly defined concept of time and space. In other words, a virtual organization does not exist in our time and space, but rather exists only in virtual space (perceptual world), which is only a metaphor of our consciousness and not reality. A virtual organization, in this sense, is the metaphor of our designed and structured consciousness that exists in virtual space to perform the intended actions of interest. However, the most important thing in a virtual organization is to identify the role of human actors who get involved in both the physical and the perceptual world. We attempt to explain the relationships between the human actors, the real and virtual organizations, and our perceptions of these concepts.

DUALITY OF HUMAN IDENTITIES

Metaphors play a very powerful role in structuring virtual organizations because terms like virtual space and virtual organization originate from symbolic languages (Faucheux, 1997). These metaphors provide the meaning of existence, thus we can treat the organization like a real organization in virtual space. Continuous analogical processes between virtual and real organizations explain the existence of virtual organizations because there exist similarities and discrepancies in them (Ahuja & Carley, 1999). A virtual organization, operating within virtual-space imagery, exists in our consciousness while an actual organization physically exists in various forms (more tangible or definable manner) such as culture, politics, resources, and so forth (Morgan, 1986). Although a virtual organization exists in our consciousness, it is associated with its physical counterpart in the real world. Allcorn (1997) described this counterpart as a parallel virtual organization and bureaucratic, hierarchical organization counterpart. However, there is a possibility that in the near future, a real organization will exist only when its virtual counterpart exists in virtual space. Mowshowitz (1994, 2002) described this as "a dominant paradigm" of virtual organization due to its unique advantages in the efficiency, cost, and effectiveness of goal-oriented activity. Surprisingly, human actors manage to control these two opposing ideas of real and virtual worlds; thus, it becomes obvious that humans possess duality of existence in both the real and the virtual world.

This paper discloses the social aspects of a virtual organization and identifies the role of human actors in a virtual organization (or consciousness in Faucheux, 1997). This consciousness exists in the perceptual world that we create beyond the limits of time and space (Allcorn, 1997). However, its counterparts exist in various forms (entities) in the real world. To bridge the gaps between the consciousness and the entities, there exists a need for human interveners who possess dual identities in both virtual and real worlds. This research provides the meaning of virtual organization, and proceeds to explain the relationship between the consciousness (virtual organizations) and entities (real organizations) with human intervention (human players).

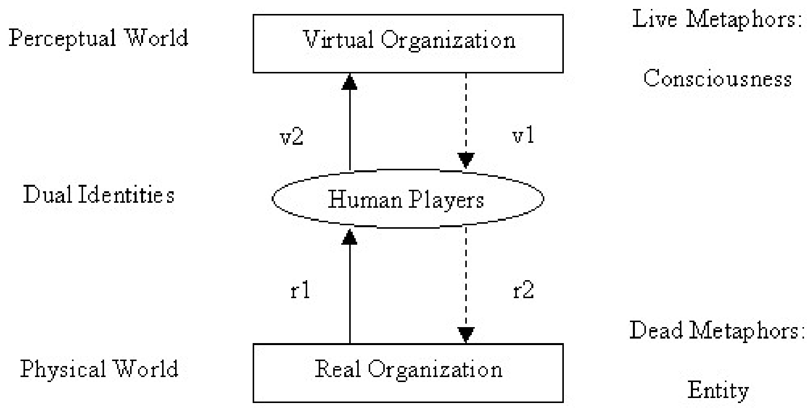
Schultze and Orlikowski (2001) examine rhetorical oppositions between real organizations and virtual organizations, and in doing so, apply metaphors to the discourse. The visions or views of two opposing elements are not divergent or dichotomous; rather, they offer substitutes for the opposition through a process referred to as dualism. As Orlikowski (1991) proposed in her earlier paper, "The Duality of Technology," this dualism is not mutually exclusive. The dualism originated from the admirable work by Giddens (1984), *The Constitution of Society*. Giddens' structuration theory integrated two main streams of sociology: objectivism and subjectivism. It appears that the structuration theory adopts the notion of phenomenology, as it seeks to make explicit the implicit structure and meaning in human experiences (Sanders, 1982). Phenomenology searches for the essence of what an experience *essentially is* and is the intentional analysis between objective appearance and subjective apprehension. Structuration theory (the process of structuration of an organization) seeks a complementary essence in the structure of organization science and in the process of struggles between objectivism and subjectivism. Interestingly, the conflict of objectivism and subjectivism was reflected in metaphors, as Lakoff and Johnson (1980, p. 189) stated:

Objectivism and subjectivism need each other in order to exist. Each defines itself in opposition to the other and sees the other as the enemy. Objectivism takes as its allies: scientific truth, rationality, precision, fairness, and impartiality. Subjectivism takes as its allies: the emotions, intuitive insight, imagination, humaneness, art, and a "higher" truth.… They coexist, but in separate domains. Each of us has a realm in his [or her] life where it is appropriate to be objective and a realm where it is appropriate to be subjective.

Human players have very important roles in both phenomenology and metaphors due to their valuable experience. The key differentiator between objectivism and subjectivism is always human experience. Another important fact (usually overlooked by researchers) is that the use of metaphors appears in both the physical world and in the perceptual world (Harrington, 1991) because the term organization itself results from "dead" metaphors. Tsoukas (1991) describes the process in which metaphors "have become so familiar and so habitual that we have ceased to be aware of their metaphorical nature and use them as literal terms." It implies that the metaphors of virtual organizations are "live" metaphors (Tsoukas) as we know "that these words are substitutes for literal utterances" that use dead metaphors (organization per se). Therefore, live metaphors are used to describe virtual organizations in another dimension, where we can do things that are not possible in the real world because the virtual world operates without the constraints of time and space, unlike the real world.

The process of structuration involves the reciprocal interaction of human players and institutional properties of organizations (Giddens, 1984); as Orlikowski (1991) pointed out, "the theory of structuration recognizes that human actions are enabled and constrained by structures, yet these structures are the result of previous actions." Because we live in both real and virtual worlds, we have both objective and subjective understandings of each world—dual identities. Figure 1 shows the relationship between real organizations and virtual organizations in the presence of human interveners. Both real and virtual organizations consist of rule resource sets that are implicated in their institutional articulation, thus these rule resource sets act as structures of the organizations (both virtual and real), where a structure is the medium and the outcome of organizational conduct. The structural properties do not exist outside of the actions of the organizational actors. Therefore, structural properties, related to space and time, are implicated in the production and reproduction of the organizations. In other words, both real and virtual organizations undergo structuration across the different sets of dimensions of time and space based on the perspectives of each human player.

Figure 1. Dual identities of human players in both real and virtual organizations



The above model, which is adopted from the duality of technology of Orlikowski (1991), depicts four processes that operate continuously and simultaneously in the interaction between human players and both real and virtual organizations. These processes include (a) institutional properties, represented by arrow r1 (objective appearance of the real organization) and arrow v1 (objective appearance of the virtual organization), which are the medium of human players; (b) structures, represented by arrow r2 (subjective construction of the real organization) and arrow v2 (subjective construction of the real organization), which are the product of human players; (c) the interaction of human players in both worlds and the resultant influences on the social contexts of the real organization within which it is built and used (the direction of arrow r1 and v2); and (d) how the virtual organization is built and used within particular social contexts in a real organization (the direction of arrow v1 and r2).

In Figure 1, there are two structurations from human players: one for the real organization and the other for the virtual organization. The realms of virtual organization and real organization are objective while the consciousness of the human player is subjective. The process of structuration (Figure 1) involves the reciprocal interaction of human players and institutional properties of organizations (Giddens, 1984). As Orlikowski (1991) pointed out, "the theory of structuration recognizes that human actions are enabled and constrained by structures, yet these structures are the result of previous actions." We have both objective and subjective understandings of each world (dual identities) because we live in both real and virtual worlds.

The maturity phase of the real organization (with its established tradition) and the fledgling state of the virtual organization (with its newly emerging phenomena) indicate that the objective appearance of the real organization (arrow r1) and subjective construction of the real organization (arrow v2) dominate the structuration process in modern organizations. Many observations show that the knowledge and experiences accumulated in a real organization enforce the formulation of the virtual organization in the abstraction process of efficient and effective goal-oriented activity (Mowshowitz, 1994). It is partly true that the creation of the virtual organization is only for the representation of the real organization in virtual space. A considerable amount of explanation arises from rethinking the basic assumptions of time and space. The real organization, whether tangible or intangible, is bound in time and space while the virtual organization, an imaginative concept established in computer hardware and software, is free from the constraints of time and space.

Barley and Tolbert (1997) defined an institution as "shared rules and typifications that identify categories of social actors and their appropriate activities or relationships." As they explained their recursive model (institutions and actions), institutionalization involves the behavior of revision or replication of organizational abstracts (work procedures) and entails objectification and externalization of behaviors. In this sense, the successful functioning of a virtual organization is reaching institutionalization in virtual space. Through this process, the virtual organization becomes stable and helps serve as the constitution where human players can follow their activities. Upon further inference, institutions from business processes of real organizations constrain human actors (constitutive nature, r1) who in turn construct institutions of virtual organization (constituted role, v2), and/or vice versa (from v1 to r2).

The above arguments provide complementary insights into the social process explained by the structuration theory (Giddens, 1984). In this theory, actions and institutions continuously interact, thereby determining the structure. The structuration theory lacks the explanation of how these interactions (revising and reproducing an institution or structure) are processed, although this is arguable as Giddens explains the role of reflection, interaction, and so forth. However, Barley and Tolbert (1997) clearly stated that their work, "the aim of institutional theory," is "to develop the implications of structuration theory for the interplay between actions and institutions and to address the practical problem of how to study institutional maintenance and change in organizations."

The authors believe that the results of this study are compatible with the belief of Barley and Tolbert (1997) that "the institutional perspective must come to grips with institutionalization as a process if it is to fulfill its promise in organization studies." The focus of this study is the explanation of what is going on in a virtual organization. The result revealed by this research is a rich description of theoretical induction. A limitation of this process is that it only reflects one part of the recursive model of institutional theory (Barley & Tolbert).

The approach of this paper is that organizational transformation is "the ongoing practices of organizational actors, and emerges out of their (tacit and not so tacit) accommodations to and experiments with the everyday contingencies, breakdowns, exceptions, opportunities, and unintended consequences that they encounter" (Orlikowski, 1996). The above statement is identical to the theoretical framework of this paper in that users of the system continuously interact with the system through producing, reproducing, and transforming work practices (Giddens, 1984).

Mowshowitz's (1997) depiction of a virtual organization is limited to a computer with a communication tool or computer network that increases the efficiency and effectiveness of organization performance (Mowshowitz, 1994). This paper complements the virtual organization as a social system giving the new meanings of time and space. This study rethinks the philosophy of the virtual organization, providing insight into the concept of the duality of human identity. It is not only a lens for understanding virtual organizations, but also a sociotechnical understanding of virtual organizations through structuration.

CONCLUSION

Organizations today are usually faced with a turbulent environment that requires flexible and dynamic responses to changing business needs. Many organizations have responded by adopting decentralized, team-based, and distributed structures. Advances in communication technologies have enabled organizations to acquire and retain such distributed structures by supporting coordination among people working from different locations (Abuja & Carley, 1998). Thus, virtuality plays an important role to achieve above goals of current business environments. The virtuality of a virtual organization can be seen as an "emptying" of the organization, where emptying of information and knowledge has occurred in current communication technology (Giddens, 1984, 1990). IT (or information systems) generates and stores information that helps in the emptying (separation) of information from organizations. Knowledge, a supposedly higher format of information, is managed by knowledge-management systems (KMSs), another evidence of the emptying of knowledge from organizations. Because data, information, and knowledge of organizations are emptying from their organizations (i.e., can be stored and manipulated in information systems), the separation of the organization from its four-dimensional entity (time factor and three location factors: latitude, longitude, altitude) can be achieved in the form of a virtual organization (Giddens) if the rest of the components of the organization are implemented in network.

REFERENCES

Ahuja, M., & Carley, K. (1999). Network structure in virtual organization. *Organization Science, 10*(6), 741-757.

Allcorn, S. (1997). Parallel virtual organizations: Managing and working in the virtual workplace. *Administration & Society, 29*(4), 412-439.

Barley, S. R., & Tolbert, P. S. (1997). Institutionalization and structuration: Studying the links between action and institution. *Organization Studies, 18*(1), 93-117.

Faucheux, C. (1997). How virtual organizing is transforming management science. *Communications of the ACM, 40*(9), 50-55.

Giddens, A. (1984). *The constitution of society*. Berkeley, CA: University of California Press.

Giddens, A. (1990). *The consequences of modernity*. Stanford, CA: Stanford University Press.

Harrington, J. (1991). *Organizational structure and information technology*. Hertfordshire, UK: Prentice Hall International.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Markus, L. M. (2000). What makes a virtual organization work? *MIT Sloan Management Review, 42*(1), 13-26.

Morgan, G. (1986). *Images of organization.* Beverly Hills, CA: Sage Publications.

Mowshowitz, A. (1994). Virtual organization: A vision of management in the information age. *The Information Society, 10*, 267-288.

Mowshowitz, A. (1997). Virtual organization. *Communications of the ACM, 40*(9), 30-37.

Mowshowitz, A. (2002). *Virtual organization: Toward a theory of societal transformation stimulated by information technology.* CT: Quorum Books.

Orlikowski, W. J. (1991). The duality of technology: Rethinking the concept of technology in organizations. *Organization Science, 3*(3), 398-427.

Orlikowski, W. (1996). Improvising organizational transformation over time: A situated change perspective. *Information Systems Research, 7*(1), 63-92.

Sanders, P. (1982). Phenomenology: A new way of viewing organizational research. *Academy of Management Review, 7*(3), 353-360.

Schultze, U., & Orlikowski, W. J. (2001). Metaphors of virtuality: Shaping an emergent reality. *Information and Organization, 11*(1), 45-77.

Tsoukas, H. (1991). The missing link: A transformational view of metaphors in organizational science. *Academy of Management Review, 16*(3), 566-585.

KEY TERMS

Collaborative Culture: By their nature, virtual organizations foster camaraderie between members even in the absence of face-to-face communications. Since the built-in communications tools are so easy to access and use, relationships form between members who have not even met. A corporate culture forms out of friendship that produces a highly collaborative nature unlike traditional organizations where such extensive communicating is not required.

Complementary Core Competencies/The Pooling of Resources: The ease with which two members of a virtual organization can communicate allows them to pool their resources, even with members not directly involved in a specific project. Separate entities can quickly be called upon to provide secondary service or consult on a project via virtual channels.

Customer-Based/Customized Products: A virtual organization provides the unique opportunity to provide their customers with highly specialized products as per their specific needs. This can be accomplished through outsourcing work to a separate organization or through the use of a virtually connected interorganizational node located closer to the customer. Either way, it becomes simple to add a function based on the customer's request and seamlessly integrate that function into the existing framework.

Electronic Communication: A vital concept to the virtual organization is the ability to communicate through purely electronic means, eliminating the need for physical contact and allowing the geographical dispersion of organization members. Online collaboration via e-mail, discussion boards, chat, and other methods, as well as telephone and facsimile communications, are primary contributors to the removal of time and space in this new organizational concept.

Explicit Goals: Similar to metamanagement, each member of the organization is charged with an explicit task to complete as it relates to the overall function of the organization. Often times, after this single goal is completed, the link between the organization and the entity is dissolved until a further need for it is realized. At this point, the link is reestablished.

Flexibility: Virtual organizations are, by their nature, flexible. Traditional organizational structures are rooted in the physical world and rely on structures, unalterable networks, and specific locations to function properly. Because of this, when it becomes necessary to introduce change into a specific organization, a barrier is reached where further alteration requires physical, costly modifications. A virtual organization is unhindered by these problems. These structures are designed so that they can operate regardless of time or place, independent of existing physical realities.

Functionally or Culturally Diverse: The nature of global diversity and the ability to locate organizational functions across the globe creates a diverse environment for the entire organization. Since members are all in different locations and charged with different tasks, diversity exists that is only found in the very largest multinational corporations.

Geographical Dispersion: The combination of virtual organization with IT allows groups of employees to make progress on one project while working in tandem with another group in a distant physical location. Because information can be shared and meetings can be held with the use of high-speed networks and computers, tasks can be carried out in the location that is most appropriate and germane to that function.

IT: Information technology is the crucial component of a modern virtual organization. Without advances in technology, many of the realities of today's virtual companies would be merely science fiction. These components include the Internet, LANs (local area networks) and WANs (wide area networks) for business, e-mail and online chat and bulletin boards, and real-time video conferencing. These technologies allow smaller work groups as part of a larger company operate independently of each other across a room or the globe.

Open Communication: The foundation of a virtual organization is its communications components that exist in absence of face-to-face exchanges. A virtual organization can only survive if its members communicate freely through the provided channels between them, be they based on the Internet or more traditional telephone technologies. The organization cannot continue to function unless it is aware of what all its members are currently completing and often times, when communication is more closed, work that is being completed in tandem by more than one member can be hindered or brought to a halt.

Participant Equality: Each individual facet of a virtual organization is expected to contribute an equal amount of work toward a given goal, if appropriate. While the equality may not be measured best in quantity, it can be restated as effort and the successful completion of all tasks assigned to it, be they large or small. Since every task is considered to be essential as a part of the project, the equality comes in the addition of that piece to a larger puzzle.

Sharing of Knowledge: Members of a virtual organization collaborate to share their knowledge gained from individual activities performed. Since collaboration is facilitated through the communications channels that are afforded through the virtual organization, it is common to find "knowledge bases" or other database systems that contain information and documents pertaining to past experience.

Switching: The switching principle is a fundamental advantage that a virtual organization has over a traditional one. Because the links between organizational functions are largely electronic and nonphysical, it is easy to replace a weak component with a stronger one. Where this activity could be considerably expensive if the item in question was a physical supply chain, it may only be a change of suppliers for the virtual organization and can be made with a phone call and database edit, as opposed to a building project.

Temporary: Virtual organizations are often formed to fill temporary needs, only extending to the end of the specific project that is charged to them. In a manufacturing project, a virtual organization may be formed between the engineers who design the project, suppliers who provide the raw materials, and the factory who processes those into finished goods. At the end of that particular project, those alliances are dissolved as they are no longer necessary to benefit the three independent groups.

Trust: The lack of physical interaction places a higher regard to the trust that exists between each entity involved in the organization. Since fewer "checks and balances" can be placed on appropriate departments, management and other entities trust that they will complete the appropriate work on time or be straightforward about delays or problems. If two entities working on a project together separated by thousands of miles are unwilling to trust each other, the work slows and suffers to a critical point.

Vague/Fluid/Permeable Boundaries: As a continuation of flexibility, the virtual organization is characterized by vague boundaries as to the extent of its use and purpose. Since small tweaks can easily and largely affect the overall organization, it is quite possible to extend the boundaries of an organization so that they encompass new purpose, people, or control.

Virtual Schools

Glenn Russell
Monash University, Australia

INTRODUCTION: THE EMERGENCE OF THE VIRTUAL SCHOOL

Until recent times, schools have been characterised by the physical presence of teachers and students together. Usually, a building is used for instruction, and teaching materials such as books or blackboards are often in evidence. In the 20th century, alternatives to what may be called "bricks-and-mortar" schools emerged. These were forms of distance education, where children could learn without attending classes on a regular basis. The technologies used included mail, for correspondence schools, and the 20th century technologies of radio and television.

Virtual schools can be seen as a variant of distance education. Russell (2004) argued that they emerged in the closing years of the 20th century and can be understood as a form of schooling that uses online computers to provide some or all of a student's education. Typically, spatial and temporal distancing is employed, and this results in students being able to use their computers at convenient times in their homes or elsewhere, rather than being subject to meeting at an agreed upon time in a school building.

The concept of a virtual school is agreed upon only in broad terms, as there are a number of variants. Some virtual schools insist on an agreed upon minimum of face-to-face contact, while others are so organized that a student might never set foot in a classroom. It is possible for a virtual school to have no physical presence for students to visit, and an office building in one state or country can be used to deliver virtual school services to interstate or international students.

One way of categorizing virtual schools is by imagining where they might be placed on a scale of face-to-face contact between students and teachers. At the conservative end of this scale, there would be conventional schools, where students use online computers in classrooms or labs for some of their lessons. A trained teacher in the same subject area might be available to help students, or other teachers, volunteers, or parents could supervise them.

Toward the middle of such a scale would be mixed-mode examples, where some subjects are offered in virtual mode, but students are asked to visit the school on a regular basis to monitor their progress or to participate in other face-to-face subjects, such as sport, drama, or art. At the other end of the scale are virtual schools where the student and teacher never meet, and there is no requirement for the student to enter a school building for the duration of the course. One example of such a virtual school is Florida High School, where, as the Florida High School Evaluation (2002) noted, there is no Florida High School building, and students and teachers can be anywhere in the world.

FACTORS PROMOTING THE RISE OF VIRTUAL SCHOOLS

The principal factors that account for the growth of virtual schools include globalisation, technological change, availability of information technology (IT), economic rationalism, the model provided by higher education, perceptions about traditional schools, and the vested interests of those involved in them.

The first of these factors, globalisation, refers to a process in which traditional geographic boundaries are bypassed by international businesses that use IT for globally oriented companies. It is now possible for curriculum to be delivered remotely from across state and national borders. Educational administrators can purchase online units of work for their school, and parents in developed countries can sometimes choose between a traditional school and its virtual counterpart.

As IT continues to develop, there is a correspondingly increased capacity to deliver relevant curricula online. As broadband connections become more common, students will be less likely to encounter prolonged delays while Web pages load or other information is downloaded. Advances in computers and software design have led to developments such as full-motion video clips, animations, desktop videoconferencing, and online music. Collectively, what is referred to as the Internet is already very different from the simple slow-loading Web pages of the early 1990s.

Economic rationalism also drives the spread of virtual schools, because the application of economic rationalism is associated with productivity. For education, as Rutherford (1993) suggested, the collective or government provision of goods and services is a disincentive to private provision, and deregulation and commercialisation should be encouraged. Consistent with this understanding is the idea that schools, as we know them, are inefficient and should be radically changed. Perelman (1992) argued that schools are remnants of an earlier industrial age that ought to be replaced with technology.

The ways in which higher education has adopted online teaching provide an example of how online education can be accepted as an alternative. The online courses provided by universities in recent years have proliferated (Russell & Russell, 2001). As increasing numbers of parents complete an online tertiary course, there is a corresponding growth in the conceptual understanding that virtual schooling may also be a viable alternative.

Those convinced that existing schools are unsatisfactory can see virtual schools as one alternative. Criticism of schools for not adequately meeting student needs, for providing inadequate skills required for employment, or not preparing students for examinations and entrance tests, are continuing themes that can be identified in a number of educational systems. Discussions related to school reform can include funding, resourcing, teacher supply, curriculum change, and pedagogy, but they can also include more radical alternatives, such as virtual schooling.

PROBLEMS OF VIRTUAL SCHOOLS AND THEIR SOLUTIONS

Virtual schools face a number of challenges related to the way that teaching and learning are implemented in online environments. While similar problems can also be identified in conventional schools, the different natures of virtual schools serve to highlight these concerns. These problems include authenticity, interactivity, socialzsation, experiential learning, responsibility and accountability, teacher training, certification, class sizes, accreditation, student suitability, and equity.

The first of these problems, authenticity, relates to the verification of the student as the person who has completed the corresponding assignments and tests from a virtual school. Virtual schools may assign students a secure password to use over the Internet, but this procedure would not preclude students from giving their passwords to a parent or tutor who completed the work on their behalf. A possible solution that may have to be considered is to independently test students to confirm that they have the understanding, knowledge, and skills suggested by their submitted work.

Interactivity describes the relationship between the learner and the educational environment. For virtual school students, there is an interactive relationship involving the multimedia, the online materials used, and the teacher. Students would typically access materials on the World Wide Web, respond to them, and send completed work electronically to their teachers. The preferred way for students to become involved in online learning is to have an active engagement involving a response. If a student is directed to a static Web page containing a teacher's lecture notes, learning may be less effective, unless other teaching methods are used to supplement it. The solution to this problem will be found in both the increased capability of students' online computers to operate in a rich multimedia environment, and the recognition by course designers that virtual schools should take advantage of advances in learning theory and technological capability.

Socialization continues to be a problem with virtual schools, because there is an expectation in conventional schooling that students will learn how to work cooperatively with others and will internalize the norms and values necessary for living in a civilized community. Moll (1998) is concerned with disruption to the tradition of public education as the primary vehicle for the transference of national narratives and humanistic and democratic values. Clearly, socialization will still occur if students use online learning supplemented by some contact with teachers and opportunities for organized sports. However, students' ability to relate to others in society is likely to change. Despite this concern, a type of virtual school that routinely insists on organized face-to-face learning and social situations, with peers, teachers, and other adults, will reduce the problems that otherwise are likely to arise

A related concern to that of socialization is the belief that Web culture is inherently isolating, and that by encouraging students to pursue their education with a virtual school, an existing trend toward loss of community may be exacerbated. Kraut et al. (1998) originally suggested that Internet use could be associated with declines in participants' communication with family members in the household, declines in the size of their social circles, and increases in depression and loneliness. However, more recent research (Kraut, Kiesler, Boneva, Cummings, Helgeson, & Crawford, 2002) found that negative effects had largely dissipated.

There are some teaching activities in conventional schools referred to as experiential. These usually involve some form of hands-on activity or physical interaction with others. Typically, a teacher will provide a demonstration, explanation, or modeling of what is to be learned, and activities that follow provide opportunity to correct errors. While virtual schools commonly offer subjects such as mathematics and social studies, the study of physical education, drama, art, and the laboratory component of science is more problematic. Sometimes the problem does not arise, because students will enroll only for subjects that they missed or that they need for credit toward a qualification.

A common solution to these problems is for the virtual school to provide online or print-based teaching materials, as with other subjects in the range to be offered. Students complete the activities and send evidence of the completed work to the school. The Open School (2002) in British Columbia, Canada, offers art in both elementary and secondary school levels. At the Fraser Valley Distance Education Centre (2002), students are invited to participate in a science fair by sending in digital pictures and a digital video clip of their project to the supervising teacher.

Changing notions of responsibility, accountability, and student discipline are also likely to arise in virtual school environments (Russell, 2002). In a traditional school, teachers accept responsibility for the students in their charge, including the prevention of physical injury, and accountability for using appropriate teaching techniques. When there is a spatial and temporal distance between teacher and student, teachers are unable to exercise some of their accustomed responsibilities. While there is still a requirement to act ethically, and to ensure that appropriate teaching materials and methods are used, much of the responsibility shifts to parents, students, and to the suppliers of the online materials.

Teacher training is also emerging as an area of concern. Virtual teachers will find that some new skills are required, while others are less important. Class management skills in a face-to-face environment will differ from their online equivalents, as will many of the teaching practices. Salmon (2002) identified a number of skills that will be required by online teachers in the future. It is clear that there will be an ongoing need to use technological skills and to apply these skills to an appropriate educational context. However, it is unlikely that many teachers' colleges and other providers of trained teachers have modified their courses to reflect these changes, as mainstream teacher education is still focused on conventional school education. There are, nevertheless, some hopeful signs. The California Virtual School Report (2002) provided evidence of the use of online modules for teachers at Durham Virtual High School, in Canada, and a 15-week teacher-training program in Fairfax County School District.

Parents would normally expect that the virtual teacher working with their child would be a competent online teacher and be certified or registered with the corresponding school system. Where a student is working from home, and the principal contact with the teacher is by e-mail, the anonymity of the communication mode could conceivably cover the use of unqualified teachers. The necessity for demonstrating that a high-quality educational experience is being supplied is, however, likely to reduce this possibility. Florida Virtual High School uses only certified classroom teachers (Schnitz & Young, 2002, p. 4). As the online environment becomes more competitive, it is likely that virtual schools will provide evidence of their teachers' certifications.

With conventional schools, the issue of class sizes is a perennial problem. The diversity of virtual schools means that it is not easy to determine corresponding workloads. The evaluation of Virtual High School (VHS; Kozma et al., 2000) revealed that some of the teachers involved in the case study had to complete their VHS work at home in addition to their normal teaching load during the day. When teachers are asked to take responsibility for large groups of students, the time available for individual attention is likely to be reduced, and the quality of the educational service provided may be less satisfactory. There are indications that some virtual schools have recognised this problem. Louisiana Virtual School (2002), for example, is limited to 20 students per course.

Accreditation of courses across geographic regions will also become an increasing problem. Palloff and Pratt (2001) noted concerns with the quality of online high school programs as early as 2001. Varying standards can mean that a course in one area is not recognized in another. Students will increasingly be able to choose programs across state and even national borders and complete their schoolwork by sitting at home with their computers.

An important item relating to the quality of a student's educational experience in a virtual school is the recognition that not all students are suited to online learning. Already, some virtual schools try to determine whether the prospective student is suited to online learning by using questionnaires. Typically, these questionnaires ask students about their independent learning skills, motivation, time management abilities, and comfort with technology.

If virtual schools are perceived to be advantageous for those enrolled in them, there are also concerns as to when the access to them is seen as inequitable. Bikson and Paris (1999) found that there were "highly significant differences in household computer access based on income" (p. 9), in the United States. It is reasonable to assume that households with children will have less access to computers to use in a virtual school if they are part of a disadvantaged group. Unless there is careful planning, the use of technology-mediated education is likely, in the short term, to further entrench those inequalities that exist in society.

FUTURE TRENDS IN VIRTUAL SCHOOLS

Two broad trends can be identified in the growth of virtual schools. These are the continued expansion in the number of virtual schools, and the trend from virtual high schools to virtual K–12 schools. Research by Clark (2001, p. 3) indicated that more virtual schools began their operations in the United States during the period 2000–2001 (43%) than in the previous 4 years combined. Fifty-one percent of virtual schools surveyed offered junior high and middle school courses as well as high school courses, and about one in four schools offered courses across the whole K–12 spectrum (Clark, 2001, p. 4). In Canada, there is also evidence of growing demand for virtual schools. The 2-year cumulative growth rate for Alberta virtual schools was 125% (SAEE, 2002).

Collectively, the implication of these trends is that there will be increased attention devoted to those problems that arise from virtual schooling across the K–12 range. When virtual schools made their first appearance, it would have been possible for some educators to dismiss them because they were experimental, or ignore their existence because they catered only to a niche market of high school students. In some cases, this suggestion may still be valid, but support for virtual schooling is increasing, rather decreasing, and the nature of what is offered is becoming more comprehensive.

CONCLUSION

Virtual schools continue the tradition whereby students learn at a distance from their teachers. The availability of online courses through the Internet has simultaneously reduced the emphasis given to older forms of distance education, while it increased the opportunities for students to explore alternatives to traditional school education. It is likely that there will be an increase in the number of virtual schools, and that they will continue to attract students. The expected increase in the number and type of virtual schools is likely to provide both exciting possibilities and daunting challenges.

REFERENCES

A national survey of virtual education practice and policy with recommendations for the state of California. Available online: *VHS_Report_lowres.pdf*

Bikson, T. K., & Paris, C. W. A. (1999). Citizens, computers and connectivity: A review of trends. Available from *http://www.rand.org/publications/MR/MR1109/mr1109.pdf*

California Virtual School Report. (2000). The California Virtual High School report.

Clark, T. (2001). Virtual schools: Trends and issues—A study of virtual schools in the United States. Available from *http://www.WestEd.org/online_pubs/virtualschools.pdf*

Florida High School Evaluation. (2002). The Florida High School Evaluation: 1999–2000 year-end report for the Orange County School Board. Tallahassee, FL: Center for the Study of Teaching and Learning, Florida State University. Available from *http://www.flvs.net/_about_us/pdf_au/fhseval_99-00.pdf*

Fraser Valley Distance Education Centre. (2002). Available from *http://www.fvrcs.gov.bc.ca/*

Kozma, R., Zucker, A., Espinoza, C., McGee, R., Yarnell, L., Zalles, D., & Lewis, A.(2000). The online course experience: Evaluation of the Virtual High School's third year of implementation, 1999–2000. Available from *http://www.sri.com/policy/ctl/html/vhs.html*

Kraut, R., Kiesler, S., Boneva, B., Cummings, J., Helgeson, V., & Crawford, C. (2002). Internet paradox revisited. *Journal of Social Issues*, *58*(1), 49–74.

Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukopadhyay, T., & Scherlis, W.(1998). Internet paradox: A social technology that reduces social involvement and psychological well-being? *American Psychologist, 53*(9), 1017–1031.

Louisiana Virtual School. (2002). Available from *http://www.icet.doc.state.la.us/distance*

Moll, M. (1998). No more teachers, no more schools: Information technology and the "deschooled" society. *Technology in Society, 20*, 357–369.

Open School. (2002). Open School in British Columbia, Canada. Available from *http://openschool.bc.ca*

Palloff, R. M., & Pratt, K. (2001). *Lessons from the cyberspace classroom: The realities of online teaching*. San Francisco: Jossey-Bass.

Perelman, L. (1992). *School's out: Hyperlearning, the new technology and the end of education*. New York: William Morrow and Company.

Russell, G. (2002). Responsibility for school education in an online globalised world. Focus paper presented to Technology Colleges Trust Vision 2020 Online Conference (United Kingdom).

Russell, G. (2004). Virtual schools: A critical view. In C. Cavanaugh (Ed.), *Development and management of virtual schools: Issues and trends* (pp. 1–25). Hershey, PA: Information Science Publishing.

Russell, G., & Russell, N. (2001). Virtualisation and the late age of schools. *Melbourne Studies in Education, 42*(1), 25–44.

Rutherford, T. (1993). Democracy, markets and Australian schools. In C. James, C. Jones, & A. Norton (Eds.), *A defence of economic rationalism* (pp. 151–159). St Leonards: Allen and Unwin.

SAEE. (2002). Executive summary of e-learning: Studying Canada's virtual secondary schools. Available from *http://www.saee.bc.ca/vschoolsum.html*

Salmon, G. (2000). *E-moderating: The key to teaching and learning online*. London: Kogan Page.

Schnitz, J., & Young, J. E. (2002). Models of virtual schooling. Available from *http://www.can.ibm.com/k12/pdf/Virtualschool.pdf*

KEY TERMS

Bricks-and-Mortar Schools: These are traditional schools, where students attend at a physical school building.

Distance Education: A generic term referring to education where teachers and students are geographically separate. Modes employed include print and nonprint technologies.

Experiential Learning: Learning based on direct and unmediated instruction or on physical interaction with people and materials.

Globalization: The bypassing of traditional geographic borders using information technology to enable global orientation of business and remote curriculum delivery.

Interactivity: The relationship between the learner and the educational environment.

Socialization: The process by which students internalize the norms and values necessary for living in a civilized community.

Virtual School: A form of schooling that uses online computers for part or all of a student's education.

V

Virtual Teams as Sociotechnical Systems

Haydee M. Cuevas
University of Central Florida, USA

Stephen M. Fiore
University of Central Florida, USA

Eduardo Salas
University of Central Florida, USA

Clint A. Bowers
University of Central Florida, USA

INTRODUCTION

A virtual team can be described as an organizational unit unconstrained by geographical, temporal, organizational, and/or national boundaries (Townsend, DeMarie & Hendrickson, 1998). Despite their rising popularity, numerous issues exist surrounding how virtual teams can productively coordinate their resources, activities, and information, often in dynamic and uncertain task environments (Fiore, Salas, Cuevas & Bowers, 2003; Townsend et al., 1998). With organizational structure increasing in complexity to include both co-located and virtual team members, explicit linkages between theory and practice are critically needed to mitigate the negative effects that technology-mediated interaction may have on virtual team productivity. Our goal here is to demonstrate how classic and current theories and principles from organizational psychology can be effectively integrated within a sociotechnical systems framework to address the unique challenges faced by this subset of teams. Specifically, we analyze the effects that collaborative information technology and lack of co-location may have on virtual team members. We conclude with suggested interventions for organizational practice.

BACKGROUND

Open Sociotechnical Systems

Radical changes in organizational structure brought about through advances in technology represent a critical challenge for the appropriate application of theoretically-based principles in system design. Researchers and practitioners need to focus on system design issues not only at the individual or task level, but also at the team and organizational level. This involves a system-level analysis (Hendrick, 1997) of the following sociotechnical factors that interact to shape organizational outcomes and may hinder attainment of organizational goals:

(1) *Personnel subsystem:* comprised of the organizational unit's members
(2) *Technological subsystem:* representing the technology available to the organizational unit
(3) *External environmental variables:* which act upon the organizational unit

Taken as a whole, these subsystems collectively represent the organizational unit as a *sociotechnical system*. Because the organizational unit both acts on and is acted upon by external forces, it would be considered an *open* sociotechnical system (Emery & Trist, 1960). Thus, the organizational unit can be viewed as a complex set of dynamically intertwined and interconnected elements, including inputs, processes (throughputs), outputs, feedback loops, and the environment in which it operates and interacts (Katz & Kahn, 1966).

Team Opacity in Distributed Environments

The technological component, in particular, plays a key mediating role by setting limits upon the system's actions as well as by creating new demands that must be reflected in the internal structure and goals of the organizational unit (Emery & Trist, 1960). In distributed environments, the technological subsystem may have a potentially greater effect on team member interactions than would be expected in traditional co-located task environments. Virtual teams rely primarily on electronic communication processes to work together both synchronously (e.g., video conferencing, Internet chat rooms) and asynchronously (e.g., electronic mail, bulletin boards) to accom-

Table 1. Factors influencing virtual team productivity

Factor	Definition/Description	Impact on Virtual Team Productivity
INPUT FACTORS		

Resources	Personnel subsystem:
	• Individual member knowledge, skills, attitudes
	• Team size, composition
		• May be differentially affected by technological subsystem limitations (e.g., media richness, information synchrony)
		• With low media richness, team opacity may:
		  • filter out critical paralinguistic cues
		  • hinder development of mutual trust (Avolio et al., 2001; Fiore, Salas & Cannon-Bowers, 2001)
		• Conversely, lack of visual cues may lead to:
		  • greater focus on task-relevant member attributes (e.g., skills)
		  • less focus on task-irrelevant stereotypical attributes (e.g., race)
		  • selection of leaders that more closely embody team values, ideals, and goals (McKenna & Green, 2002)

Task Demands	Nature of task (e.g., task complexity) and other work structure factors (e.g., communication channels) that form technological subsystem and external environment	• Team opacity may lead to:
		  • limited ability to monitor task-relevant cues provided by geographically dispersed teammates
		  • over-reliance on explicit communication strategies, resulting in poor task performance under conditions of high workload and task complexity (Entin & Serfaty, 1999)
		  • additional cognitive workload upon memory processes, inducing greater occurrence of memory failures (Fiore, Cuevas, Schooler & Salas, in press)

THROUGHPUT FACTORS		

Combination Processes	Various *implicit* (i.e., tacit) and *explicit* (i.e., overt) team processes and/or behaviors (e.g., communication, coordination, decision making) necessary to accomplish team's goals and/or task objectives	• With low media richness (e.g., electronic mail), team opacity may limit or altogether eliminate use of:
		  • nonverbal, paralinguistic cues (e.g., hand gestures) when conveying information crucial to completion of complex tasks (Tang, 1991)
		  • beneficial information yielded in *process* artifacts emerging from collaborative work (Tang, 1991)
		  • implicit communication and/or coordination strategies (Fiore et al., 2003)

Process Losses	Loss in team productivity resulting from poor coordination among members (i.e., lack of simultaneity of effort) and/or decreased social motivation (Steiner, 1972)	• Due to lack of nonverbal cues, ambiguous nature of distributed interaction, and subsequent over-reliance on explicit strategies (Fiore et al., 2003), team opacity may:
		  • negatively impact execution of combination processes needed to attain desired outcomes
		  • impede evolution of mutual trust, collective efficacy, and group cohesion, leading to poorly developed team attitudes and decreased social motivation

Motivation	Intrinsic and extrinsic factors that lead an individual to engage in a particular behavior or choose one course of action over another (Locke & Latham, 2002); motivation theories (e.g., goal-setting, self-regulation) focus on the underlying behaviors necessary to accomplish set goals (Bandura, 1986; Locke & Latham, 2002)	• Team opacity may negatively impact:
		  • *goal commitment* due to impoverished nature of interaction and lack of motivating influence of paralinguistic cues inherent in face-to-face interactions (Teasley, Covi, Krishnan & Olson, 2000)
		  • development of common, engaging *direction* for virtual team, depending upon information flow (i.e., synchronous or asynchronous), resulting in poor motivation to meet training and/or performance objectives (Fussell et al., 1998)
		  • *collective efficacy* (i.e., members' belief in their team's ability or competence to attain desired outcomes; Bandura, 1986) due to limited opportunities for monitoring and evaluating other members' performance (Fiore et al., 2003)

Shared Mental Models (SMMs)	Shared understanding among team members of each member's roles and responsibilities, task demands, and team-level interactions required to meet these demands (Cannon-Bowers, Salas & Converse, 1993); SMMs improve team's ability to coordinate efforts, adapt to changing demands, anticipate needs of task and other members, and foster development of mutual trust (Avolio et al., 2001; Entin & Serfaty, 1999)	• Team opacity hinders SMM development due to decreased awareness of team member actions and expectations, resulting in:
		  • increased communication and coordination overhead due to over-reliance on explicit strategies (Entin & Serfaty, 1999)
		  • uncoordinated efforts, low team productivity, and unsuccessful attainment of organizational goals (Espinosa, Lerch & Kraut, 2004)
		  • diminished identification of role knowledge and poor source monitoring (Durso, Hackworth, Barile, Dougherty & Ohrt, 1998)
		  • poorly developed team attitudes (Fiore et al., 2001)

plish their tasks (Avolio, Kahai, Dumdum & Sivasubramanium, 2001). However, such technology-mediated interactions may increase the level of abstraction forced upon teams—a phenomenon referred to as *team opacity* (Fiore et al., 2003). Team opacity describes the experience of increased ambiguity and artificiality (i.e., unnatural quality) associated with distributed interaction. This decreased awareness of team member actions, resulting from a distributed organizational structure, creates an environment lacking in the rich visual, auditory, and social array of cues normally experienced in co-located team interaction, thereby limiting the use of implicit (i.e., tacit) coordination and communication strategies, and impeding the development of positive team attitudes and successful team evolution (Fiore et al., 2003).

FACTORS INFLUENCING VIRTUAL TEAM PRODUCTIVITY

Following a sociotechnical systems approach, we next discuss how the structure of information flow (i.e., synchronous or asynchronous) and format (e.g., text, audio, video) associated with distributed environments may contribute to team opacity and negatively impact productivity (for a full discussion, see Fiore et al., 2003). As with traditional co-located teams, virtual team productivity is critically dependent upon the *resources* available to the team, the *task demands*, and the *combination processes* enacted by the team which dictate how these resources are used to meet task demands (Steiner, 1972). Whereas the personnel subsystem comprises the virtual team's principle resources, the technological subsystem and environmental constraints drive the task demands. The degree to which combination processes (e.g., coordination, communication, decision making) are effectively executed is especially susceptible to the impact of team opacity on the occurrence of *process losses*, the team's *motivation* to achieve task objectives, and the development of a *shared mental model*. Table 1 summarizes these important input and throughput factors, highlighting the potential effect of team opacity.

FUTURE TRENDS

In distributed environments, the technological subsystem sets significant limits upon the personnel subsystem's actions and creates new demands for optimal team productivity that must be addressed through training interventions and system design. Moreover, a virtual team is a *dynamic* organizational unit, developing as activities strengthen the quality of team member interactions (Morgan, Salas & Glickman, 1993). Fiore et al. (2003) described how these issues can be conceptualized within a distributed coordination space where virtual teams interact, evolve, and mature over *time* and *space*. Specifically, these activities, so critical to team development, occur not only during *in-process* interaction, but also during *pre-* and *post-process* interaction. Whereas in-process interaction occurs during actual task execution, pre-process interactions involve preparatory pre-task behaviors (e.g., planning session) where initial shared expectations are created in anticipation of team interaction (Fiore et al., 2001). Similarly, post-process interactions (e.g., after-action review) include post-task reflection on performance (Smith-Jentsch, Zeisig, Acton & McPherson, 1998). Such antecedent and/or consequent behaviors are critical to the team evolution and maturation cycle. Table 2 summarizes several organizational practices that can be employed at each of these stages to foster team development and the successful execution of team processes. Future research is clearly warranted to empirically evaluate the effectiveness of these interventions on virtual team productivity. Our goal in this article was to demonstrate how organizational psychology theories could be effectively integrated within a sociotechnical systems framework to address the unique challenges faced by virtual teams. Though much work still remains to be done and many questions still remain unanswered, we hope the issues set forth here will inspire others to explore further ways to achieve this goal.

CONCLUSION

As the prevalence of virtual teams increases, researchers must continue to address issues surrounding their design, implementation, and management. We must identify the sociotechnical factors that both help and hinder effective virtual team productivity so as to maximize their potential while mitigating the occurrence of process losses. Adopting a sociotechnical systems approach to investigate how team opacity interacts with these unique task demands and situational constraints to alter team processes and products will enable organizations to effectively utilize the technological subsystem's capabilities to support virtual team productivity. Similarly, a better understanding of the distinct forms of group dynamics that may emerge in virtual teams will advance the design of appropriate training interventions. Unquestionably, the future success of virtual teams within organizations will depend upon the joint optimization of the personnel and technological subsystems comprising this unique sociotechnical system.

Table 2. Pre-, in-, and post-process interventions to promote virtual team productivity

Intervention	**Rationale and Relevant Theoretical/Empirical Work**
PRE-PROCESS	

Pre-Task Briefing	• Specifying clear, challenging yet attainable goals may lead to:
	  • increased effort on task, better use of strategies, and commitment to team (Locke & Latham, 2002)
	  • better SMM of task demands and team-level interactions required to meet these demands (Cannon-Bowers et al., 1993)

Pre-Task Interaction	• Initial face-to-face or technology-mediated (e.g., video conferencing) pre-task interactions may facilitate development of:
	  • trust and cooperation among team members (Rocco, 1998; Zheng, Bos, Olson & Olson, 2001)
	  • team's social identity and positive team attitudes (Bos, Olson, Gergle, Olson & Wright, 2002; Zheng et al., 2001)

Team Building Exercises	• Distinct training interventions aimed at improving the effectiveness of team processes and operations by prompting members to evaluate their behaviors and relationships (Tannenbaum, Beard & Salas, 1992) • Valuable as both pre-process and in-process interventions

Role Identification (RI) & Interpersonal Relations (IR)	• Most beneficial for fostering positive team attitudes and commitment to team
	  • RI: identifies each member's roles and responsibilities to minimize any difficulties arising from role conflict or role ambiguity
	  • IR: improves relations among team members

Goal-Setting (GS) & Problem-Solving (PS)	• Well suited to enhance processes by which virtual teams use their resources to meet task demands
	  • GS: assists teams in setting individual and group goals, and determining strategies to meet these objectives
	  • PS: guides team members in developing requisite skills for identifying relevant elements in a problem (e.g., givens, goals, obstacles/constraints) and employing effective problem-solving strategies

IN-PROCESS	

Technological Support Tools	• Designed to increase awareness of member actions and provide feedback on performance to enhance combination processes (e.g., coordination) and foster development of positive collective efficacy (Steinfeld, Jang & Pfaff, 1999)
	  • *Team-level*: inform distributed workgroups of member actions and important changes in within-team and external information, without diverting their attention away from central tasks (Cadiz, Fussell, Kraut, Lerch & Scherlis, 1998)
	  • *Individual-level*: provide powerful and flexible tools with which to interact with shared workspace and its artifacts (Gutwin & Greenberg, 1998)

Cross-Training	• Particularly effective in fostering SMM development (Cannon-Bowers et al., 1993) • Helps members better understand each other's roles and responsibilities • Enables members to more accurately predict and anticipate each other's behavior and make greater use of implicit team processes (Entin & Serfaty, 1999)
Team Adaptation and Coordination Training (Entin & Serfaty, 1999)	• Focuses on building team's SMM • Enhances teamwork behaviors and coordination strategies by increasing quality and quantity of cues utilized by teams, leading to improved decision-making performance • Adaptive to varying levels of stress and workload
Cue-Recognition Training	• Increases saliency of critical aspects of task environment (Salas, Cannon-Bowers, Fiore & Stout, 2001) • Decreases occurrence of memory failures and increases probability that team members recognize when critical cues have been missed
POST-PROCESS	

Debriefing Sessions & After-Action Reviews	• Involves careful, well-structured dissemination of feedback information (Cannon-Bowers et al., 1993)
	• Promotes positive collective efficacy by prompting guided team self-correction and involving team members in self-regulation of their performance (Smith-Jenstch et al., 1998)
	• Strengthens team's SMM by:
	  • fostering shared knowledge regarding expectations and specific preferences of team members and effective teamwork processes (Smith-Jenstch et al., 1998)
	  • increasing source knowledge of member expertise (Libby, Trotman & Zimmer, 1987)

ACKNOWLEDGEMENTS

The views herein are those of the authors and do not necessarily reflect those of the organizations with which the authors are affiliated. This research was funded by Grant Number F49620-01-1-0214, from the Air Force Office of Scientific Research to Eduardo Salas, Stephen M. Fiore, and Clint A. Bowers. Portions of this article were excerpted from: Cuevas, H.M., Fiore, S.M., Salas, E. & Bowers, C.A. (2004). Virtual teams as sociotechnical systems. In S.H. Godar & S.P. Ferris (Eds.), *Virtual and collaborative teams: Process, technologies, and practice* (pp. 1-19). Hershey, PA: Idea Group Publishing. Address correspondence to Dr. Stephen M. Fiore, Institute for Simulation and Training, 3280 Progress Drive, Orlando, FL 32826; e-mail sfiore@ist.ucf.edu.

REFERENCES

Avolio, B.J., Kahai, S., Dumdum, R. & Sivasubramanium, N. (2001). Virtual teams: Implications for e-leadership and team development. In M. London (Ed.), *How people evaluate others in organizations* (pp. 337-358). Mahwah, NJ: Lawrence Erlbaum.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bos, N., Olson, J.S., Gergle, D., Olson, G.M. & Wright, Z. (2002). Effects of four computer-mediated communications channels on trust development. *Proceedings of CHI 2002*. New York: ACM Press.

Cadiz, J.J., Fussell, S.R., Kraut, R.E., Lerch, F.J. & Scherlis, W.L. (1998, November). The Awareness Monitor: A coordination tool for asynchronous, distributed work teams. Unpublished manuscript. Demonstrated at the *1998 ACM Conference on Computer Supported Cooperative Work*, Seattle, WA.

Cannon-Bowers, J.A., Salas, E. & Converse, S. (1993). Shared mental models in expert team decision making. In N.J. Castellan, Jr. (Ed.), *Individual and group decision making* (pp. 221-246). Hillsdale, NJ: Lawrence Erlbaum.

Durso, F.T., Hackworth, C.A., Barile, A.L., Dougherty, M.R.P. & Ohrt, D.D. (1998). Source monitoring in face-to-face and computer-mediated environments. *Cognitive Technology, 3*, 32-38.

Emery, F.E. & Trist, E.L. (1960). Socio-technical systems. In C.W. Churchman & M. Verhulst (Eds.), *Management sciences: Models and techniques* (Volume 2, pp. 83-97). New York: Pergamon.

Entin, E.E. & Serfaty, D. (1999). Adaptive team coordination. *Human Factors, 41*, 312-325.

Espinosa, J.A., Lerch, F.J. & Kraut, R.E. (2004). Explicit versus implicit coordination mechanisms and task dependencies: One size does not fit all. In E. Salas & S.M. Fiore (Eds.), *Team cognition: Process and performance at the inter- and intra-individual level* (pp. 107-129). Washington, DC: American Psychological Association.

Fiore, S.M., Cuevas, H.M., Schooler, J.W. & Salas, E. (in press). Cognition, teams, and team cognition: Memory actions and memory failures in distributed team environments. To appear in C.A. Bowers & E. Salas (Eds.), *Teams and technology*. Washington, DC: American Psychological Association.

Fiore, S.M., Salas, E. & Cannon-Bowers, J.A. (2001). Group dynamics and shared mental model development. In M. London (Ed.), *How people evaluate others in organizations* (pp. 309-336). Mahwah, NJ: Lawrence Erlbaum.

Fiore, S.M., Salas, E., Cuevas, H.M. & Bowers, C.A. (2003). Distributed coordination space: Toward a theory of distributed team process and performance. *Theoretical Issues in Ergonomics Science, 4*, 340-363.

Fussell, S.R., Kraut, R.E., Lerch, F.J., Scherlis, W.L., McNally, M.M. & Cadiz, J.J. (1998). Coordination, overload, and team performance: Effects of team communication strategies. *Proceedings of CSCW 1998* (pp. 275-284). New York: ACM Press.

Gutwin, C. & Greenberg, S. (1998). Design for individuals, design for groups: Tradeoffs between power and workspace awareness. *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (pp. 207-216). New York: ACM Press.

Hendrick, H.W. (1997). Organizational design and macroergonomics. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 594-636). New York: John Wiley & Sons.

Katz, D. & Kahn, R.L. (1966). *The social psychology of organizations*. New York: John Wiley & Sons.

Libby, R., Trotman, K.T. & Zimmer, I. (1987). Member variation, recognition of expertise, and group performance. *Journal of Applied Psychology, 72*, 81-87.

Locke, E.A. & Latham, G.P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist, 57*, 705-717.

McKenna, K.Y.A. & Green, A.S. (2002). Virtual group dynamics. *Group Dynamics: Theory, Research, and Practice, 6*, 116-127.

Morgan, B.B. Jr., Salas, E. & Glickman, A.S. (1993). An analysis of team evolution and maturation. *Journal of General Psychology, 120*, 277-291.

Rocco, E. (1998). Trust breaks down in electronic contexts but can be repaired by some initial face-to-face contact. *Proceedings of CSCW 1998* (pp. 501-529). New York: ACM Press.

Salas, E., Cannon-Bowers, J.A., Fiore, S.M. & Stout, R.J. (2001). Cue-recognition training to enhance team situation awareness. In M. McNeese, E. Salas & M. Endsley, (Eds.), *New trends in collaborative activities: Understanding system dynamics in complex environments* (pp. 169-190). Santa Monica, CA: Human Factors and Ergonomics Society.

Smith-Jentsch, K.A., Zeisig, R.L., Acton, B. & McPherson, J.A. (1998). A strategy for guided team self-correction. In J.A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 271-297). Washington, DC: American Psychological Association.

Steiner, I.D. (1972). *Group process and productivity*. New York: Academic Press.

Steinfeld, C., Jang, C. & Pfaff, B. (1999, November). Supporting virtual team collaboration: The TeamSCOPE System. *Proceedings of GROUP '99: International ACM SIGGROUP Conference on Supporting Group Work,* Phoenix, AZ.

Tang, J.C. (1991). Findings from observational studies of collaborative work. *International Journal of Man-Machine Studies, 34*, 143-160.

Tannenbaum, S.I., Beard, R.L. & Salas, E. (1992). Team building and its influence on team effectiveness: An examination of conceptual and empirical developments. In K. Kelley (Ed.), *Issues, theory, and research in industrial/organizational psychology* (pp. 117-153). Amsterdam: Elsevier.

Teasley, S., Covi, L., Krishnan, M.S. & Olson, J.S. (2000). How does radical collocation help a team succeed? *Proceedings of CSCW 2000* (pp. 339-346). New York: ACM Press.

Townsend, A.M., DeMarie, S.M. & Hendrickson, A.R. (1998). Virtual teams: Technology and the workplace of the future. *Academy of Management Executive, 12*, 17-29.

Zheng, J., Bos, N., Olson, J.S. & Olson, G.M. (2001). Trust with*out* touch: Jump-start trust with social chat. *Proceedings of CHI 2001*. New York: ACM Press.

KEY TERMS

Explicit Team Processes: Openly articulated, overt communication and coordination behaviors. *Example*: Member A directly requests task-relevant information from Member B.

External Environment: Relevant work structure variables (e.g., task complexity) and other external forces (e.g., organizational climate) within the sociotechnical system that both act on and are acted upon by the organizational unit.

Implicit Team Processes: Largely unverbalized, tacit communication and coordination behaviors, founded on shared understanding of members' roles/abilities and task/situational demands. *Example*: Member A provides Member B with backup assistance without being asked.

Open Sociotechnical System: A complex set of dynamically intertwined and interconnected elements, including inputs, processes (throughputs), outputs, feedback loops, and the environment in which it operates and interacts.

Personnel Subsystem: Sociotechnical system component comprised of the organizational unit's members, including individual knowledge, skills, and attitudes, and team size and composition.

Team Opacity: Increased level of abstraction forced upon virtual teams due to the over-reliance on technology-mediated interactions; involves an experience of increased ambiguity and artificiality (i.e., unnatural quality) associated with distributed interaction, which inherently lacks the rich visual, auditory, and social array of cues traditionally available to co-located teams.

Team Productivity: Degree to which a group of two or more individuals can successfully coordinate their efforts to perform a task.

Technological Subsystem: Sociotechnical system component representing the technology (e.g., collaborative information technology) available to the organizational unit, influencing the nature of information flow (synchronous or asynchronous) and format (e.g., text, audio, video).

Virtual Team: An organizational unit that is unconstrained by geographical, temporal, organizational, and/or national boundaries, and whose members rely upon electronic communication processes to work together both synchronously (e.g., video conferencing) and asynchronously (e.g., electronic mail) to accomplish their tasks.

Virtual Work Research Agenda

France Bélanger
Virginia Polytechnic Institute and State University, USA

INTRODUCTION

The paper by Bélanger, Watson-Manheim, and Jordan (2002) addresses the gap between research conducted and practitioner concerns in virtual work. One of the key difficulties in conducting research in this area is the overlap between terms used (McCloskey & Igbaria, 1998; Pinsonneault & Boisvert, 2001). While there are other distributed work arrangements such as hotelling, neighborhood work centers and flextime, most of the previous literature has focused on telecommuting (telework) and virtual teams/organizations. In this article, the term virtual work represents work environments where individuals spend some time working in a non-face-to-face (FTF) mode, using information and communication technologies to perform work activities.

Virtual work environments are increasingly employed by organizations. While there is increased complexity and potential for problems, virtual work strategies allow organizations a great deal of flexibility to compete in a rapidly changing business environment. While existing research provides insights into such environments, it does not clearly deal with major concerns faced by managers (referred to as the "gap" between research and practice). One of the potential reasons for this gap is that practicing managers are concerned with current challenges in their own work setting while academics are concerned with developing more generalizable rules and understanding.

This article addresses these issues, with three particular objectives:

1. examine the gap between research and practice in virtual work;
2. investigate factors leading to the gap; and,
3. identify a research agenda that addresses emerging issues and concerns relevant to practice in virtual work.

BACKGROUND

To explore the gap between virtual work research and practice, the authors first review previous literature, which they then compare to concerns raised by practitioners in two organizations. To identify relevant academic research, the authors searched for articles in mainstream IS journals. They then used the "snowball" technique, mining citations in articles for further references. They did not include the large number of conference papers and studies of home-workers, entrepreneurs, or supplemental work at home. They focused on empirical and/or theoretically grounded studies.

Literature Review

Their review of recent literature (1998 to 2001) revealed six literature reviews and 35 empirical studies. In the original paper, a table including methodology details and key concepts was provided but is not included here for brevity, although sample references are provided. Overall, literature addresses the following questions:

- Who are the virtual workers? There are two types of studies that discuss who virtual workers are. The first type is descriptive, usually presenting demographics and other characteristics of virtual workers based on general surveys or public records (e.g., Johnson, 2001). The second type investigates characteristics of telecommuters (e.g., Bélanger, 1999b).
- How is communication influenced by virtual work? This area has been the most researched in recent years. The published work comprises studies of communication patterns (e.g., Bélanger, 1999a) and studies of choices of communication modes (e.g., Wiesenfeld, Raghuram, & Garud, 1999).
- What technologies are used and how do they influence virtual work outcomes? There were few studies prior to 1998 focusing on technologies in virtual work. Recent studies look at computer-mediated communication systems, the Web, the role of technologies in affecting productivity of teleworkers (e.g., Bélanger, Collins, & Cheney, 2001), and usage patterns of technology in virtual teams (e.g., Majchrzak, Rice, Malhotra, King, & Ba, 2000).
- What is the nature of the work-family conflict in virtual work? There are some recent studies looking at stress but most studies were published prior to 1998 (e.g., Duxbury, Higgins, & Mills, 1998), or in non-IS mainstream journals.

- What are the outcomes of virtual work environments? Most hypothesis-driven studies used outcomes of virtual work, such as productivity, satisfaction, or level of communication, as dependent measures (e.g., McCloskey, 2001). Potential outcomes were also discussed extensively in pre-1998 literature.
- What happens in virtual group work? Studies investigate trust, development processes, and performance in virtual teams, and perceptions in virtual groups (e.g., Maznevski & Chudoba, 2000).
- What are the key issues in managing remote group workers? Studies typically look at issues with managing teleworkers (e.g., Staples, 2001).

Overall, the review showed that a number of barriers, enablers, and outcomes of virtual work have been studied. The samples have often been limited, for example, one organization, which can limit the generalizability of the findings. However, sample size has increased in recent years. In general, given the complexity of organizations, the current research still seems to be narrowly focused.

Case Narratives

To investigate the gap, interviews were conducted in two organizations with distributed workers but with quite different levels of worker distribution. The first organization is Booz Allen Hamilton. It is a global management and technology consulting firm. For their IT practice of the Worldwide Technology Business (WTB), they rely on distributed teams comprised of geographically dispersed employees. The teams are classified as functional, delivery, development, and external teams. Their flexible matrix allows members to participate on multiple teams.

The organization has had success using technology support for team communication. The technologies available include project management, collaboration, and knowledge management tools. Management challenges, however, do occur with practical issues such as the needs for breadth of multi-disciplinary domains, collaboration tools, and training. In managing the breadth of multi-disciplinary domain, identifying the right "mix" of team members with requisite skill sets is a challenge. While electronic collaboration tools are available, managers are not sure whether and how using particular tools makes positive outcomes more likely. Computer and collaboration technology training is left up to each consultant. Management wonders whether project managers are effective because of interpersonal qualities or because of automated tools usage.

The second case is a Fortune 100 telecommunications company headquartered in Southeast USA. It services residential and business telephone customers in nine states. The narrative focused on the management of 700 network service technicians responsible for installation and repair of telephone services within one district. Technicians complete four to five work orders per day assigned by a centralized provisioning center. They are evaluated based on efficiency in completing orders and quality of the work performed. Some teams are staffed in shifts 24 hours/day, seven days/week, while others have eight-hour days with frequent overtime.

Technicians and supervisors use information and communication technologies extensively, including cell phones, pagers, and wireless laptops. The first work order is loaded on the technician's laptop before the work day begins. Technicians update the work order, and completed orders are updated in the system as soon as a worker establishes a connection. A new work order is then assigned. Such a system allows dynamic assignment of work orders based on changing priorities throughout the day.

Managers are responsible for overseeing eight to 15 technicians, including visiting and inspecting the site where work was completed. Supervisory duties also include providing training for technicians (formal and informal). They must respond to individual questions which arise due to unique field conditions or changes in technology. In addition, they must conduct performance evaluations, counsel technicians on career development, and mentor technicians new to the job.

Such responsibilities pose challenges for managers. Management challenges include work activity coordination, measurement tools, training, and information sharing for team building. Coordinating activities of distributed field workers who face process changes, often due to field conditions, is challenging. Supervisors are not able to observe the work performed, which causes difficulty in providing feedback to technicians on the quality of their work. Training in a timely and consistent manner is difficult with staggered schedules of field workers. In addition, it is difficult for distributed workers to develop relationships with team members. Thus, inability to share information with team members, especially for newer people, poses a challenge for managers.

MAIN TRUST OF THE ARTICLE

A number of managerial themes emerged from the interviews. The themes are summarized in Table 1, and only a few sample questions are provided in the following text for each theme. Please refer to Table 1 for further research questions.

Team building is an issue due to the complexity of lateral communication among team members in virtual work environments. The lack of physical interaction causes

Table 1. Gaps and overlap in virtual work research and practice

Practitioner Issues	Examples of questions of interest	Overlap*	Examples of questions/areas researched	Research in Virtual Work
Team building	What factors are critical for building and maintaining a solid team culture and effective communications process in distributed environments?		What is the effect of virtual work on communication patterns and structures in teams and work groups?	How is communication influenced by virtual work?
	What communication structures and mechanisms can be used to distribute information and coordinate tasks among team members in timely manner?		What effects does telework have on group communication structures?	How is communication influenced by virtual work?
Organizational/ management structure	Does management role change in distributed environments?		What factors lead to more successful management of remote workers?	What are key issues in managing remote workers?
	What role does distance play in determining organizational structure and job design?			
	How are resources best allocated in distributed environments?			
Information sharing & distribution	How can information sharing be facilitated in distributed environments?		How is communication and coordination performed in virtual work?	How is communication influenced by virtual work?
	How is time-sensitive information best distributed?			
	What are effective interpersonal networking techniques used by successful distributed workers?		How is communication in teams affected by virtual work?	How is communication influenced by virtual work?
Employee assessment & development	What are the best methods and metrics for employee assessment in virtual work?		Are demographics of teleworkers linked to success or other outcomes?	Who are the virtual workers?
	What is the role and appropriateness of employee monitoring in this environment?			
	How is the critical management function of employee development best performed in distributed work?			
	In distributed work teams that do not meet daily, are there differences between the factors considered in team members' evaluations and those of co-located teams?		How can trust be developed between distributed team members?	What happens in virtual group work?
	Should some performance evaluation factors be weighted more heavily than others in virtual work? Which?			
Work process training	What skills need to be developed for different types of virtual work environments?		What are the characteristics of individuals performing virtual work, including skills?	Who are the virtual workers?
	How can we train distributed workers for optimum job knowledge sharing and work in virtual teams?			
IT training and readiness	What IT training is needed and how is it most effectively employed in distributed work?			
	At what stage are employees (and org.) in their acceptance of and readiness to use information and communication tools?			
Communication tools & technology choice	How does management assess the appropriateness of available communication technologies and applications?		Which communication tools lead to greater success in telework?	What technologies are used and how do they influence virtual work outcomes?
	What tools are available, and which are best to support collaborative distributed work?		Which communications tools are available and used in virtual work?	What technologies are used and how do they influence outcomes?
	How does management assess the effectiveness of collaborative tools, where they are successful, and under what conditions?			
				What is the nature of work-family conflicts in virtual work?
				What are outcomes of virtual work?

* Filled cells indicate overlap. White cells indicate limited or no overlap.

difficulty in creating relationships. An example of a question practitioners need answered is: What factors are critical for building and maintaining a solid team culture and effective communication process in distributed work?

Hierarchical communication between employees and managers also becomes more complex in virtual work. Coordination of unpredictable tasks is difficult for supervisors. In relation to organizational management structure, practitioners need answers to issues like: How are resources best allocated in the distributed environment?

Communicating organizational information in a timely manner is a challenge. Information sharing among team members may be difficult since work teams are physically distributed. Timely updates are difficult without methods for sharing information such as posting system messages and alerts for fellow workers. A sample question of interest to practitioners could be: What are effective interpersonal networking techniques used by the most successful teleworkers?

Measuring and monitoring employees' work is complex in distributed work. Better measurement criteria on work patterns or processes would enable management to more fairly assess and provide effective feedback to employees. This leads to questions like: What are the best methods and metrics for employee assessment in virtual work?

Development and training of employees for future assignments can be difficult, particularly for new employees in distributed settings. A sample question that needs further investigation is: How are the critical management functions of employee development best performed in the distributed environment?

Ensuring that employees are performing work activities most effectively is difficult in distributed work. Tension occurs between organizational consistency and employee independence. This area has just started to be researched by academics (Cooper & Kurland, 2002; Kurland & Cooper, 2002). A sample question needing further investigation in this area is: How can we train distributed workers for optimum job knowledge sharing and work in virtual teams?

The use of information and communication technologies is critical to effective performance in distributed settings. However, the complexity and rapid change of technologies adds to the challenge of providing training at a distance. An example of a question of interest is: How is necessary IT training most effectively conducted in distributed settings?

Choosing the best communication technology is critical. Management in both organizations interviewed, however, had little guidance in how to make the choice. A typical question of interest could be: What tools are available, and which are best to support collaborative work in distributed settings?

FUTURE TRENDS

A summary of the issues raised in the case narratives as compared to the literature is presented in Table 1. It provides a high level view of potential gaps between virtual work research and practice. In general, the literature does not always adequately capture the complexity of virtual work environments, which creates a gap between managerial concerns and academic research. In addition, topics addressed by research are not always addressed in as much depth as what is needed by practitioners. For example, researchers look at coordination and communication in general while practitioners are interested in how communication within virtual work can be better used for information sharing, information distribution, performance feedback and/or relationship development. Some gaps, however, may be justifiable. For example, work-family conflict issues should be studied by researchers, but do not seem to be major concerns of practitioners. The need to address longer-term issues about societal effects of virtual work may justify research in this area. One overlap is team building where issues of trust and communication in virtual teams are researched and are seen as practical concerns by managers. Another overlap is evaluation of tools and technologies for virtual work. Research has focused on e-mail and the Web, also important to managers. In addition, managers are interested in groupware and knowledge management tools.

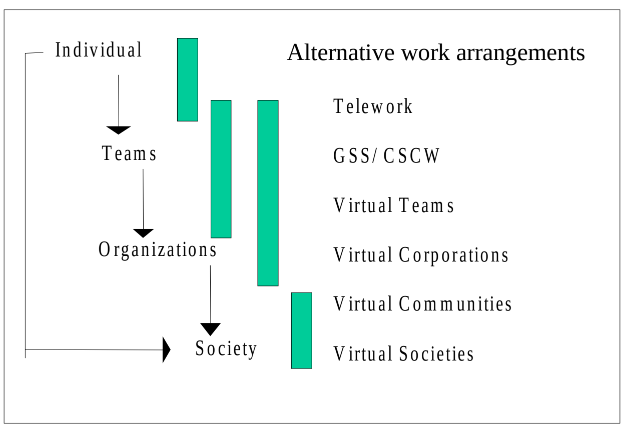
While exploring the gap, several factors that might cause a disparity between research on virtual work and concerns of practitioners became apparent. Possible reasons for the gap include multidisciplinary nature of managerial concerns, time-intensive requirements for research methodologies, and lack of proper definition of the unit of analysis.

Virtual work research is fragmented by areas while problems faced by managers are multi-disciplinary in nature. Business organizations require a more systemic and holistic approach to studying virtual work. For example, issues of technology and organizational communication cannot be separated in virtual work. Understanding interpersonal relations is critical to understand how relationships are formed and maintained in an environment where cooperating individuals are working in different contexts with different technologies.

Longitudinal case studies or multiple case studies are the most appropriate research methodologies to study virtual work. The type of research needed requires substantial time investments from the researchers' perspective. It is, therefore, more difficult to accomplish this research and get the appropriate rewards of timely publications.

The unit of analysis needs to be appropriate to the research question, and the research question should be relevant to the "unit" being studied. A proposed view of the appropriate unit of analysis in virtual work research adapted from Agres, Edberg, and Igbaria (1998) is shown in Figure 1. For example, the organizational level unit of analysis could include studies of telecommuting, GSS/CSCW, virtual teams or virtual corporations.

Figure 1. Proposed virtual work units of analysis



CONCLUSION

Through an in-depth review of virtual work literature and insights from two organizations, a gap between research and practice was identified since practitioners are faced with issues and challenges in virtual work environments for which research does not always capture the complexity. Although possible reasons provided for a gap between research and practice is not exhaustive, it is apparent that a need exists to have better communication between practitioners and researchers on issues of importance and on how each can benefit from one another's work. As academics, we should consider these as opportunities to perform research of importance to both the academe and practitioners.

REFERENCES

Agres, C., Edberg, D., & Igbaria, M. (1998). Transformation to virtual societies: Forces and issues. *Inform Soc, 14*, 71-82.

Bélanger, F. (1999a). Communication patterns in distributed work groups: A network analysis. *IEEE Transactions on Professional Communication, 42*(4), 261-275.

Bélanger, F. (1999b). Workers' propensity to telecommute: An empirical study. *Information and Management, 35*(3), 139-153.

Bélanger, F., Collins, R., & Cheney, P.H. (2001). Technology requirements and work group communication for telecommuters. *Information Systems Research*.

Bélanger, F., Watson-Manheim, M.B., & Jordan, D.H. (2002). Aligning IS research and practice: A research agenda for virtual work. *Information Resources Management Journal, 15*(3), 48-70.

Cooper, C.D., & Kurland, N.B. (2002). Telecommuting, professional isolation, and employee development in public and private organizations. *Journal of Organizational Behavior, 23*, 511-532.

Duxbury, L.E., Higgins, C.A., & Mills, S. (1998). After-hours telecommuting and work family conflict: A comparative analysis. *Information Systems Research, 3*(2), 173-190.

Johnson, N.J. (2001). Case study of the St.Paul Companies' virtual office for the risk control division. In N.J. Johnson (Ed.), *Telecommuting and virtual offices: Issues and opportunities* (pp. 148-161). Hershey, PA: Idea Group Publishing.

Kurland, N., & Cooper, C. (2002). Manager control and employee isolation in telecommuting environments. *Journal of High Technology Management Research, 13*, 107-126.

Majchrzak, A., Rice, R.E., Malhotra, A., King, N., & Ba, S. (2000). Technology adaptation: The case of computer-supported inter-organizational virtual team. *MIS Quarterly, 24*(4), 569-600.

Maznevski, M.L., & Chudoba, K.M. (2000). Bridging space over time: Global virtual team dynamics and effectiveness. *Organization Science, 11*(5), 473-492.

McCloskey, D.W. (2001). Telecommuting experiences and outcomes myths and realities. In N.J. Johnson (Ed.), *Telecommuting and virtual offices: Issues and opportunities* (pp. 231-246). Hershey, PA: Idea Group Publishing.

McCloskey, D.W., & Igbaria, M. (1998). A review of the empirical research on telecommuting and directions for future research. In M. Igbaria & M. Tan (Eds.), *The virtual workplace* (pp. 338-358). Hershey, PA: Idea Group Publishing.

Pinsonneault, A., & Boisvert, M. (2001). The impacts of telecommuting on organizations and individuals: A review of the literature. In N.J. Johnson (Ed.), *Telecommuting and virtual offices: Issues and opportunities* (pp. 163-185). Hershey, PA: Idea Group Publishing.

Staples, D.A. (2001). Making remote workers effective. In N.J. Johnson (Ed.), *Telecommuting and virtual offices: Issues and opportunities* (pp. 186-212). Hershey, PA: Idea Group Publishing.

Wiesenfeld, B.M., Raghuram, S., & Garud, R. (1999). Communications patterns as determinants of organizational identification in a virtual organization. *Organization Science, 10*(6), 777-790.

KEY TERMS

Computer Supported Collaborative Work (CSCW): Research area that focuses on investigations and development of technologies that can be used for collaborative work in distributed settings.

Group Support Systems (GSS): A set of technologies used to help groups in their decision making processes.

Hotelling (Neighborhood Work Center): Organizational facility for employees to work at but where they do not have a permanently assigned desk. They must "check-in" every time they come to work there.

Longitudinal Case Studies: Research method that involves looking at particular cases over a longer period of time, with repeated measures to observe a phenomenon as it evolves.

Telecommuting (Telework): Work arrangement that allows employees to work at home during regular work hours.

Virtual Teams/Organizations: Teams and/or organizations where some or all of the members work from different physical locations.

Virtual Work: Work environments where individuals spend some time working in a non-face-to-face (FTF) mode, using information and communication technologies to perform work activities.

Virtual Work, Trust and Rationality

Peter Murphy
Monash University, Australia

INTRODUCTION

Since the development of the Internet—and the emergence of computer networking as a mass medium in the mid-1990s—many organizations and institutions have experimented with Internet protocol (IP)-based communications to coordinate work and activities across geographical distance. This has been in response to growing needs to coordinate business and projects between different offices, firms, regions, and states. Rather than organizations flying people to meet face-to-face, network technology presents opportunities for persons located apart to work together. It offers the potential for cheap and efficient collaborations across distance. Yet, while economic pragmatics drive organizations to adopt virtual work methods, virtual working is difficult to implement. This is because it strains many conventional assumptions about work behaviour and the cognitive and emotional foundations of collaboration.

BACKGROUND

Since the 1970s, there has been a general trend worldwide for organizations to move from being closed systems to open systems. This has involved growing pressures on organizations to interact with their environment rather than trying to internalize their environment. The most visible consequences of this have been the escalating tendency of organizations to contract out functions, to relocate parts of their operations across the world, and to grow the number of strategic collaborations with other organizations. The result is more and more organizational actors working with persons—often persons they do not know—in other locations. Working with people at a distance means working virtually (Duarte & Snyder, 1999; Franke, 2002; Igbaria & Tan, 1998; Jackson, 1999; Kisielnicki, 2002; Lipnack & Stamps, 2000; Mowshowitz, 2002; O'Hara-Devereaux & Eccles, 1994). Virtual collaborators (teams and partners) have no shared physical presence. Collaborators may see one another only rarely if at all.

The technologies of virtual collaboration are relatively straightforward: e-mail, ftp, collaborative groupware, and audio-video conferencing. Groupware and IP-based conferencing is still relatively under-utilized. Third-party hosted groupware offers solutions to high-level collaboration across firewalls. IP-based conferencing provides opportunities to enrich interactions with sound and visuals. Groupware to date, however, does little more than make conventional file storage and threaded discussion available to persons working in multiple locations across organizational boundaries. Conferencing software is only beginning to be able to deliver quality audio across low bandwidth connections. Typically, high-quality video and the sharing of complex software applications still require high network bandwidth, and are often unavailable from roaming and non-institutional locations.

While technology shapes the possibilities of virtual interactions, psychology is a more powerful factor in determining the viability of such interactions. A basic condition of virtual collaboration is the ability to work with others without seeing them, knowing them, or meeting them in person. While technology can enable such work, to effectively leverage these technological possibilities, organizations have to adapt themselves to different ways of working, and in some cases they have to re-invent themselves. Working virtually at the micro-level of teams, groups, and pairs is only effective where the larger organizational environment lends itself to virtual interaction.

There are three basic types of organization: social, procedural, and the virtual or self-organizing (Miller, 2002). Social organizations are the most common type. These are based on face-to-face interactions and on character norms such as loyalty, dedicated service, and "keeping your word". Procedural organizations are built on impersonal roles and rules. Virtual organizations are structured around more abstract patterns and forms. The family firm and the relationship-driven Japanese corporation are examples of the social organization (Fukuyama, 1995). The Fordist-type American corporation typifies the procedural kind (Chandler, 1977). In contrast, production and distribution reliant on intangible or intellectual capital, such as licensing, patents, or correspondence, encourages forms of virtual collaboration based on high degrees on self-organization (Barley, Freeman & Hybels, 1992).

In order to be effective, any organized human activity must be rational. Rationality is another word for continuity, identity, and stability of expectation. Organizational behaviours deteriorate or collapse if the members of an organization cannot see that these behaviours are for the most part rational. The emotional correlate of rationality is trust. What is felt to be reliable, and worthy of trust, is also that which is recognized to be rational. Any organization where people trust one other is more effective than an organization where persons are suspicious of each other (Kramer & Tyler, 1996).

In social organizations, people "with character" are generally recognized as rational actors. These might be persons who are dependable, loyal, and unwavering in their treatment of each other. Through demonstrating that they are good at following social norms, such agents generate trust (Fukuyama, 1995; Handy, 1995). With the development of equity corporations and modern management in the late nineteenth century, social organizations in many places were replaced at least in part by procedural or bureaucratic organizations (Chandler, 1977; Yates, 1989). These developed around rules, roles defined by rules, procedures, work demarcations, impersonal written communication, and file management. Knowledge of rules rather than of people provided organizational continuity, identity, and stability. Thus persons who were consistent at following and applying rules acquired reputations for trustworthiness. Predictability in decision-making and task execution became the primary source of trust in bureaucratic organizations—complementing and often superseding the loyalty and patronage work cultures of social organizations.

Virtual work does not follow the logics of either social or procedural organizations. Without face-to-face interaction, character norms cannot be the basis of organized action. At the same time, procedural rules are difficult to agree on, to follow, or to enforce because virtual collaborators do not share the same office, organization, or manager. Virtual actors have to deal with multiple rule sets across diverse institutions, geographies, and cultures. Under these conditions, rules become ambiguous, conflicted, and uncertain. One party's rationality becomes another's irrationality. Such conflicting expectations breed distrust. Thus, under virtual conditions, rationality and trust have to be generated by other means (Murphy, 2003).

CRITICAL ISSUES

Because there is not the same history of working virtually as there is of working socially or working procedurally, identification of the means by which virtual partners and teams generate rationality and trust is less developed. If virtual collaborators cannot rely on personal moral character or on impersonal "rules and roles" to facilitate their interaction, then what can they rely on? The simplest answer is that, in the absence of social cues or clear-cut procedural direction, persons working have to be self-organizing. The key to successful self-organization is the sense of pattern or designing intelligence. Where self-directed activity (Ray & Bronstein, 1995) dominates cooperative and peer interaction, design intelligence and pattern rationality function as the coordinating medium of organized activity and group behaviour. If not, collective cohesion readily collapses.

Human beings have a strong design sense. They pick up exceptionally quickly on design characteristics such as rhythm, harmony, and proportion. Pattern recognition is central to brain processing (Davies, 1992). For instance, we use our pattern sense to make judgments about regular sentences, trustworthy buildings, and reliable machines (Alexander, 1977; Fodor & Pylyshyn, 1988; Gelernter, 1998). Such pattern rationality is also conducive to building trust. Patterns generate feelings of surety, satisfaction, and reliability. This applies as much to work environments as to cities, machines, or sentences. To create patterns, organizations employ tacit forms of aesthetic cognition (Calas & Smircich, 1996). Aesthetic cognition uses beauty, elegance and economy rather than rules or roles to achieve its ends.

Successful virtual work is conducted like a design process (Murphy, 2003). It relies less on the passing around of overt messages, and more on the ability of collaborators to understand through the exercise of imagination where their part "fits" into the overall design of the workflow. "Fit" is achieved by thinking in aesthetic terms of proportionality, rhythm, and harmony rather than in terms of rules or roles. The rationality of a virtual organization is not the rationality of character or procedure but of design. Much of this "acting by design" is intuitive or unspoken. It rests on imaginative cognition. Persons who work virtually by necessity cannot talk a lot or interact a lot with each other—so they need to imagine a lot. They need to be good at projective or anticipatory thinking. This projective thinking is not the same as the anticipatory thinking involved in either relationship empathy or in Gantt chart style project management. Rather, it is much more figurative in nature. The virtual collaborator who uses imagination is good at "seeing the shape of things" in lieu of dense social relationships or strong procedural guidance.

Virtual team or partnership work relies heavily on imaginative visualization and intuition. This is a kind of tacit knowledge. It is tacit in the sense that it involves picture thinking and pattern cognition rather than verbalization. It requires the cognitive-psychological capacity to "figure" things out (Mintzberg & Westley, 2001). Such cognitive figurative methods are closer in kind to processes of creative design than they are to processes of social recognition. In this context tacit does not mean the implicit understanding we derive from the warm handshake or the disapproving stare of another person. The tacit nature of figurative work methods thus are different in nature from the tacit knowledge that we draw from the bodily presence of collocated work partners. In the case of the imagination, tacit refers to high levels of picture-like abstraction. At the same time, however, because many aspects of this design intelligence operate non-discursively, the imaginative abstraction that is required in virtual working is quite unlike the explicit rules of procedural organizations or the rule-driven inferential reasoning typical of procedural rationality.

Virtual work elides socio-emotive contents and makes conventional discursive (mile-stone) office planning difficult to implement. For virtual work to be successful, even on a micro-scale, it must draw heavily on the faculty of imaginative and figurative thinking. To be proficient at cooperation and interaction, virtual workers must be able to picture what is absent and organize that picture "aesthetically".

In virtual collaborations, designing intelligence generates shared integrative schemas—such as asynchronous rhythms of interaction or proportionate distributions of task load. Where virtual teams and collaborators "find their rhythm," they will—more likely than not—produce good work. Correspondingly, it is the act of producing good work that builds trust amongst virtual collaborators. Trust arises where collaborators know that each will "do their part". "Doing their part" does not mean sending lots of social messages, nor does it mean showing procedural fluency. Virtual trust is generated not through relationships or procedures but rather through the aura of reliability that arises from the visible effects of the "invisible" qualities of beauty, elegance and economy that virtual actors employ when they cooperate at a distance to produce things in imaginative ways. This means producing well-designed objects—be they physical goods, processes, systems, or learning objects—in a "choreographed" manner. The choreography of the virtual team rests not on social gestures or on rules but on a good sense of rhythmic or harmonic "fit". The sense of satisfaction and surety derived from such "fit" provides the emotional trust that is the counterpart of aesthetic rationality.

FUTURE TRENDS/CONCLUSION

As business and government are pressed to operate over increasingly large scales and long distances, and as inter-organizational, inter-agency, and inter-state activity becomes more common, the need for distance communications and virtual teams is growing and a class of virtual workers is gradually emerging.

All the while, established paradigms of work remain entrenched, meaning that there is a latent propensity to try and build virtual teams and partnerships around social communication and procedural norms. This creates conflict between the intrinsic nature of virtual cooperation and the extrinsic goal extending the reach of traditional organizational structures. As organizations at the micro-level of team and peer relations continue to expand their geographical scope, and as the boundaries of organizations become increasingly fluid as a result, a major challenge for the future will be increasing the understanding of the role of aesthetic rationality and designing trust in the formation of productive relations between actors who are separated by distance and time.

REFERENCES

Alexander, C. (1977). *A pattern language*. Oxford: Oxford University.

Barley, S.R., Freeman, J., & Hybels, R.L. (1992). Strategic alliances in commercial biotechnology. In N. Nohria & R.G. Eccles (Eds.), *Networks and organizations: Structure, form, and action*. Boston: Harvard Business School Press.

Calas, M., & Smircich, M. (1996). Essays on aesthetics and organization. *Organization*, *3*(2).

Chandler, A.D. (1977). *The visible hand: The managerial revolution in American business*. Cambridge, MA: Harvard University Press.

Davies, P. (1992). *The mind of God: The scientific basis for a rational world*. New York: Simon & Schuster.

Duarte, D., & Snyder, N. (1999). *Mastering virtual teams*. San Francisco: Jossey-Bass.

Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*. Lausanne: Elsevier.

Franke, U. (2002). *Managing virtual Web organizations in the 21st century*. Hershey, PA: Idea Group Publishing.

Fukuyama, F. (1995). *Trust: The social virtues and the creation of prosperity*. New York: Free Press.

Gelernter, D. (1998). *Machine beauty: Elegance and the heart of technology*. New York: Basic Books.

Handy, C. (1995). Trust and the virtual organization. *Harvard Business Review, 73,* 3.

Igbaria, M., & Tan, M. (1998). *The virtual workplace.* Hershey, PA: Idea Group Publishing.

Jackson, P. (1999). *Virtual working.* New York: Routledge.

Kisielnicki, J. (2002). *Modern organizations in virtual communities.* Hershey, PA: IRM Press.

Kramer, R.M., & Tyler, T.R. (1996). *Trust in organizations.* Thousand Oaks, CA: Sage.

Lipnack, J., & Stamps, J. (2000). *Virtual teams: People working across boundaries with technology.* New York: John Wiley.

Miller, K. (2002). *Organizational communication.* Belmont, CA: Wadsworth.

Mintzberg, H., & Westley, F. (2001). Decision making: It's not what you think. *MIT Sloan Management Review, 42,* 3.

Mowshowitz, A. (2002). *Virtual organization.* Westport, CN: Quorum.

Murphy, P. (2003). Trust, rationality and the virtual team. In D. Pauleen (Ed.), *Virtual teams: Projects, protocols and processes.* Hershey, PA: Idea Group Publishing.

O'Hara-Devereaux, M., & Johansen, R. (1994). *Global work: Bridging distance, culture, and time.* San Francisco, CA: Jossey-Bass.

Ray, D., & Bronstein, H. (1995). *Teaming up: Making the transition to a self-directed, team-based organization.* New York: McGraw-Hill.

Yates, J. (1989). *Control through communication: The rise of system in American management.* Baltimore: Johns Hopkins University Press.

KEY TERMS

Design: The structured composition of an object, process or activity.

Distance Communication: Communication under conditions of geographic separation that minimize the possibility of face-to-face and synchronous interactions.

IP-Based: Network technologies based on Internet protocols.

Open System: Any system without strong boundaries, where information and other goods flow to and from the system's environment.

Organization: The deliberate integration of persons in order to achieve a goal or outcome.

Procedure: A rule that governs a formal organizational process.

Rationality: The ability to infer with relative certainty from existing or past behaviour and statements future behaviour and statements.

Trust: Confidence that an object, process, institution, or another person's actions can be relied upon to produce some good.

Virtual Interaction: Computer-mediated interaction between persons who do not occupy the same physical space.

Virtualization and Its Role in Business

Jerzy Kisielnicki
Warsaw University, Poland

INTRODUCTION

A new management trend of the global information technology (IT) application—virtualization—has appeared in contemporary management. Virtualization is a process of enterprise transformation (using IT) that allows breaking through various limitations of organizational constrains. Virtualization changes dramatically the image of business, especially of small and medium enterprises (SMEs); by adopting the concept of virtualization, they can become fully competitive and may effectively operate in the global market. Barriers of scale between SMEs and large organizations disappear. This new type of organization is often called in the literature a *modern organization* or *virtual organization*. Organizations of this type have an effective decision-making process and function based on economic criteria. Consequently, their opportunities to grow and compete in the global market are greater than for traditional SMEs. Hence, the thesis: virtualization allows individual organizations to enter strategic cooperative alliances with other similar businesses. Such virtual organizations have a competitive position in the global market.

In the literature, there are many terms used to define the virtual organization: "network organizations" (Drucker, 1988, p. 9); "organizations after re-engineering" (Hammmer &. Champy, 1993, pp. 77-79; Peters, 1994, pp. 5, 7); "intelligent enterprise" (Quinn, 1992, p. 3).

BACKGROUND

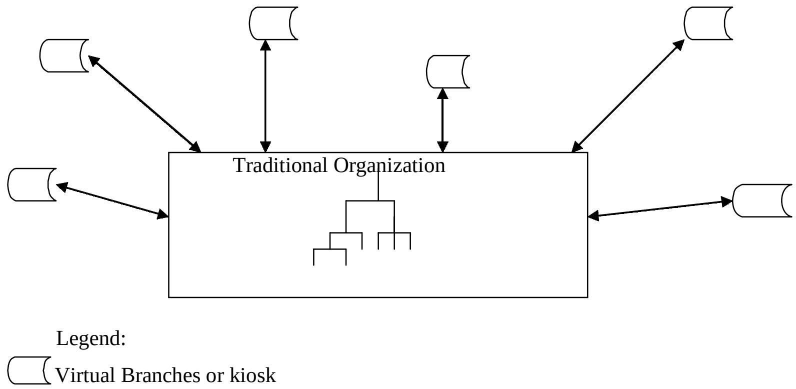
Virtualization, defined as a process of continuous transformation, is a herald of a new direction in the science of organization management. In the context of this analysis, this process may assume such form that will allow them to become competitive in the global market. The process of transformation consists of quick adjustments of the enterprise to new requirements (Hendberg, Dahlgren, Hansson, & Olive, 2000). This is done through changes in the organizational structure as well as in the portfolio of products and services. These changes are possible due to development in the IT sector, particularly Internet applications (Keeny & Marshall, 2000).

From the theoretical perspective, we can separate the following forms of virtualization:

1. Functional extension (i.e., a vertical development). This occurs when the enterprise either wishes to be closer to the customer and it does not have adequate resources, or when the establishment of a traditional branch is not profitable. The enterprise creates for this purpose virtual branches or *kiosks*. Sometimes it enables its customers to use its services via computer or mobile phone. Examples are Internet banks, bookshops (best-known is amazon.com), department stores, and travel agencies. Large companies also commonly extend their scope through such vertical development. It ensures increased competitiveness with a simultaneous control over the whole organization. SMEs apply such a strategy to a limited extent, most often for the purpose of marketing their presence in the Internet.
2. Creation of the virtual organization, or horizontal development. Such a development occurs through a virtual incorporation of other organizations. The literature lacks a unanimous definition of this concept (Hendberg, 2000; Kisielnicki, 1998; Quinn, 1992; Scholzch, 1996). For the purpose of this analysis, we assume that:

   • A virtual organization is created when its members voluntarily enter in relations of various types to achieve their common goal. Every member who creates this organization defines the duration of the relation. The member who first admits that the existence of that relation is unfavorable makes the decision on its liquidation and withdraws. The virtual organization operates in *cyberspace* and requires the existence of the Internet and global IT infrastructure.
   • Large-scale enterprises (LSEs) use IT to strengthen their competitive position in relation to other enterprises. As Hammer and Stanton (1999) rightfully notice, IT becomes—for a certain class of organizations—a "wall" that divides them from other enterprises. Large enterprises are described as "castles" (Hammer & Stanton). LSEs build these "castles" to protect themselves from competition. SMEs do not have such a sheath. They are more flexible than LSEs at a price: They are more prone to infiltration. In general however, the more experienced and knowl-

Figure 1. Virtual organization based off traditional structure



edgeable the SMEs are, the more attractive they are to other enterprises seeking their share in the virtual enterprise.

MAIN THRUST OF THE ARTICLE

Virtualization in Traditional Organizations

Virtualization allows for organizational development at a much lower cost than through the traditional process. Figure 1 presents a virtual organization based off a traditional structure and connected with its virtual elements. These elements are flexible, allowing for quick adjustment to the changing environment by the organization.

The virtualization may be developed as follows:

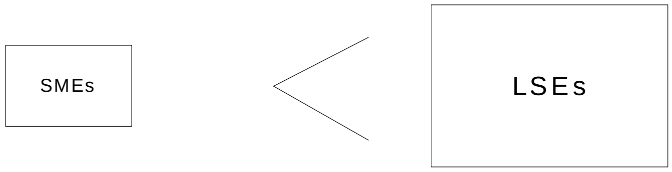
- The organization creates virtual kiosks or shops. For example, an organization that sells furniture using the Internet may present its products on a computer screen. It may also receive orders from customers located outside its traditional market. Other organizations (tourist, real estate, bookshops, stock exchange, etc.) may operate in the same manner.
- The organization places information about its activity on the Internet (Porter, 2001). It is available to its clients 24/7 rather than being limited to office hours. Internet banking services are a good example. In addition, the expenses connected with the services of an organization are shifted to the client (the client pays for the terminal, access to the Internet, etc.). The organization covers the cost of the development and maintenance of an application.
- The organization creates a possibility of working from home (called telework). It is a good way of increasing professional activity in those regions where it is difficult to find "traditional" employment. Also, through telework, a local organization may become a global one.

Virtualization allows traditional organizations to have a wider range of influence. Society is better informed of the organization's activities both by the organization itself and by its clients. The restrictions on this development are varied. The most common include: available financial resources for IT infrastructure, language, and a necessity to have global, reliable computer networks. It should also be stressed that, unfortunately, virtualization enables organizations that are socially unacceptable (pornography, terrorism) to operate freely.

Based on the analysis of organizations that use virtualization, it is estimated that their operations require five times lower investment outlays. Generally, minimum savings obtained as a result of virtualization exceeded 60%. Only in a few cases was this proportion less favorable.

In the organizations using telework, the proportions are difficult to calculate. The analysis of organizations using telework for outsourcing their services to developing countries confirms the effectiveness of virtualization. A good example is software development. The companies from highly developed countries (USA, Great Britain, etc.) employ programmers from India, China, or Pakistan. This situation is beneficial for both the company and the countries providing resources.

Figure 2. Comparison of individual SMEs and LSEs (LSEs have a competitive advantage over individual SMEs)



A different situation occurs when telework is connected with professional activation of the disabled or unemployed. Direct costs are higher, as we deal with the poorer part of the society. Thus additional costs have to be incurred for training, hardware, and software. Unfortunately, there is no data available to make a precise estimate. In many countries, the cost of training and equipment is covered by special social programs. It is also difficult to estimate advantages. It may be said that social effect (i.e., decreased unemployment) and - in case of the disabled - the ability to live a normal life are the most important ones. This is possible only through virtualization.

Generally, telework reduces operating costs (parking space, office space) and provides timesaving (i.e., no travelling to work) to employees.

Virtual Organization as a Chance for SMEs

Virtualization allows more and more SMEs to leave local markets and become global enterprises. Thus, the global market, for many years available only for LSEs, opens up to a wide range of enterprises. As G. Yip (2004, p. 27) states, "the skill of defining and implementation a global strategy is a true test of leadership and management skills." The contemporary global market most often exists as an e-market as well as an e-business (Turban, King, Lee, Warkentin, & Chung, 2002). The enterprise, irrespective of its size, that wants to make its existence in the global market should meet a number of conditions, such as:

1. Possessing a well-known and reputable brand
2. Built-up distribution and service network
3. Product or services that are unique and in demand
4. Management team able to support the global enterprise

To meet the conditions specified above, the enterprise must have at its disposal adequate funds as well as material and human resources. Comparing SMEs with LSEs, the former are in a very difficult situation. The enterprise, to make its appearance in the global market, must incur a definite outlay (a break-even point) to enter the global market. This break-even point is determined by:

1. Advertising and promotional campaign
2. Volume of production or services, allowing for selling price below the current market price
3. Quality of product or service meeting international standards, such as ISO
4. Access to the distribution channels
5. Formal barriers (legal, duties, taxation, etc.)

Operating in the global market is much more difficult than in local markets as it demands application of advanced IT solutions, access to the international data warehouses, etc. Only with the use of IT and access to information can a global company reduce the risk of failure.

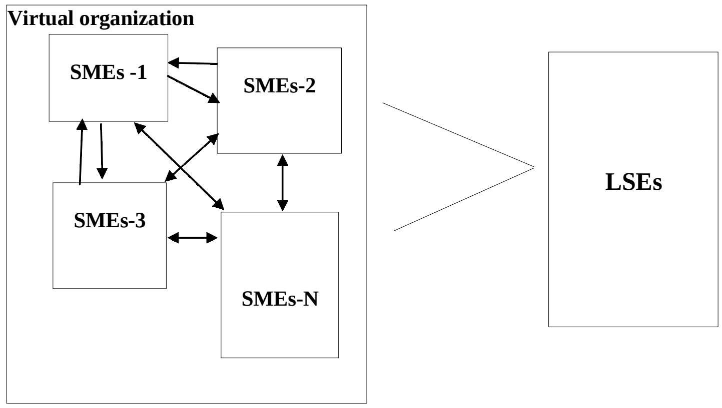
Operational costs of the enterprise are still larger if it operates in the global e-market (Elstrom, 2001). The surveys (Reichheld & Schefter, 2000; Reichheld, 2001) have shown that the cost of gaining a customer in this market is significantly higher than in the traditional market.

For example, in the fashion industry, the cost of gaining a customer in the e-market is up to 40% higher. The profits in the following years, however, grow much faster. In the near future, organizations that cannot succeed in this market will not be able to succeed at all.

Identification of the break-even point requires empirical surveys. It depends on the industry and the degree of globalization of the enterprise (there are not many enterprises that have been operating on all continents). To identify the break-even point, we may apply the methods of strategic analysis and especially the analysis of Porter's five forces. For example, the break-even point may be identified by the investment required to realize 2% of revenue from the global market. Sometimes, we can consider as an organization of such a type the one that has more than 50% of revenue from the global market. Because single SMEs do not have such resources to surpass the break-even point, they create strategic alliances in the form of the virtual enterprise.

Theoretically, development of the virtual organization may be significant. Practically, a degree of its development is connected with technical barriers and with the

Figure 3. Comparison of a virtual organization and LSEs (virtual organization as a set of SMEs is more competitive than LSEs)



necessity to achieve the goals set up for the virtual SMEs. Modification of the goal causes a fast modification of the enterprise.

FUTURE TRENDS

My research suggests that a significant number of students feels the need to learn business administration (Kisielnicki, 2001). At the same time, learning should be as close to the reality as possible. Virtualization is the discipline of studies which may cause the distance between theory and practice to diminish. One may even risk a statement that it is the exact direction of virtualization which, in the nearest future, shall have the most significant impact on society. Virtualization increases the effectiveness of the teaching process in the widest possible sense by supporting traditional teaching methods. Virtualization may be applied in such areas as lowering cost and decreasing time of training jet pilots and also in improvement of military command, operating on the stock exchange, or cognitive analysis of genetic processes. In science, there are many examples of big discoveries which were first tried on computer-simulated models. It is virtualization which allows for business simulation of both the decision-making process and the analysis of complex technical or sociological processes. In virtualization of teaching, there are two basic directions.

The first one is the direction of common education where everybody can, using IT, gain a given knowledge. A classic example is a virtual stock exchange. In many countries, a lot of people want to learn how to operate on the stock exchange before they actually start using it to make money. They can get the necessary experience using appropriate software. They can acquire the necessary skills in the virtual world. There are also numerous games available through the Internet. These games not only provide a pleasant time spent on playing but also teach foreign languages or how to drive a car.

Another direction is dedicated teaching. There are the following activities where virtualization can help:

- Self-evaluation, using special software to assess the level of language knowledge.
- Learning assistance, includes, amongst others, enterprise laboratories, business games, and special simulators that teach how to use specific technical equipment, e.g., flying simulators.
- Distant learning; a student who has a proper terminal and software may participate in classes from a distance. This direction of virtualization is similar to the previously presented distance work.

CONCLUSION

The virtual organization operates as a *transparent organization* in the sense that all of its initiatives are known to the competition. This enforces responsibility. The image of a virtual organization is influenced by the activities of its individual participants. Virtual organizations do not have common administration, offices, and real estate. In addition, virtual organizations do not have a joint executive team, middle management, or coordinators. Since the influence of each individual virtual organization is limited, the key success factor is mutual trust.

Virtual organizations give new insight into business management. They may trigger increased entrepreneurship and competitiveness. They also introduce new management methods different than traditional ones. Virtual organizations are an interesting alternative for current organizations. They are especially attractive for developing countries that want to operate in the international environment. It is fascinating how virtual organizations—without a formal reporting structure and control—can achieve high operational performance and thus have a competitive advantage over their traditional counterparts.

Regardless of speculation on future solutions, we can now safely define the areas where virtual organizations can easily outperform traditional organizations. They are: trade, tourism, and services. In these areas, the benefits of virtual organizations are as follows:

1. Operational flexibility is much higher, especially when a quick reaction to the emerging market niche is required
2. A transaction life cycle is much shorter (especially the closure)
3. Use of cyberspace to close some transactions (despite legal and organizational barriers)
4. Lower operational costs
5. Lower start-up costs
6. Ability to cooperate with organizations which LSEs cannot accept (for political, geographical, racial, religious reasons)

Depending on each individual opportunity, virtual organizations can be successful in other industries as well. Very rarely, traditional organizations can outperform the virtual organizations.

The economic metrics supporting the benefits of virtual organizations are generally available for individual cases. However, for illustration purposes, I include the results of research recently carried out in Poland:

1. Virtual travel agencies achieved 20% higher profit per transaction that traditional organizations.
2. Transaction life cycle was 100% faster in the Internet bank than in a brick-and-mortar bank.
3. The average transaction cost on the Internet bank was $0.10-0.15 compared to $1.10 in the brick-and-mortar bank.

We can put forward a hypothesis that progress in IT will create a snowball of virtual organizations. Don Tapscott (1998), introducing the 12 rules of the new economic deal, writes that they lead towards "virtual reality."

REFERENCES

Barrenchea, M. (2001). E-business.

Byrne, J.A., & Brandt, R. (1993, August 02). The virtual corporation. *Business Week*.

Drucker, P. (1988). The coming of the new organisation. *Harvard Business Review*, 1-2, 9.

Elstrom, E. (2001, July). E-money. *Business Week*, 63.

Hammer, M., & Champy, J. (1993). Reengineering the corporation. *HarperBusiness*, 77-79.

Hammer, M., & Stanton, S. (1999, November-December). How process enterprises really work. *Harvard Business Review, 108*.

Hendberg, B., Dahlgren, G., Hansson, J., & Olive, N. (2000). Virtual organizations and beyond: Discovering imaginary systems.

Kenny, D., & Marshall, J. F. (2000, November-December). The real business of the Internet. *Harvard Business Review*, 119.

Kisielnicki, J. (1998). Virtual organization as a product of information society. *Informatica*, *22*, 3.

Kisielnicki, J. (2001). Virtual organization as a chance for enterprise development. In M. Khosrowpour (Ed.), *Managing information technology in a global economy* (p. 349). Hershey, PA: Idea Group Publishing.

Kisielnicki, J. (Ed.). (2002). *Virtual organization in modern society*. Hershey, PA: Idea Group Publishing.

Peters, T. (1994). *Crazy times call for crazy organisations*. Ton Peters Seminar.

Porter, M. E. (2001, March). Strategy and the Internet. *Harvard Business Review*, 62.

Quinn, J. B. (1992). The intelligent enterprise. *The Free Press N-Y*, 3 .

Reichheld, F. F. (2001, July-August). Lead for the loyalty. *Harvard Business Review*, 76.

Reichheld, F. F., & Schefter, P. (2000, July-August). E-loyalty: Your secret weapon on the Web. *Harvard Business Review*, 105.

Scholzch, C. (1996). Virtuelle Unternehmen—Organisatorische Revolution mit Strategischer Implikation, Amnagement & Computer, *2*, 16.

Tapscott, D. (1998). *Digital economy*. Warsaw: Business Press.

Turban, E., King, D., Lee, J., Warkentin, M., & Chung H. M. (2002). Electronic commerce: Managerial perspective.

Yip, G. S. (2004). Global strategy. PWE, Warsawza, 27.

KEY TERMS

Cyberspace: An "online world of computer networks." This definition can be augmented by the following characteristics: The network consists of various globally distributed computers that can send and receive information using common protocols. In addition, this network does not have physically defined measured boundaries. Examples of the network types are: electronic mail (e-mail), World Wide Web (WWW), electronic data interchange (EDI), business-to-business (B2B) applications, business-to-customer (B2C) applications, peer-to-peer (P2P) applications (Defined by the Merriam-Webster Online Dictionary).

Virtual Organization (defined by Kisielnicki, 2002. p. 12)**:** A virtual organization is created voluntarily; its members create this organization to accomplish a certain goal. A virtual organization is created anytime a manager realizes that he or she needs the cooperation of other organizations to accomplish a goal. The duration of a virtual organization is defined by each of its members; a virtual organization operates in cyberspace, which means that its duration can be very short; so short, in fact, that it would be impossible to cooperate with other organizations using traditional methods. The decision to restructure or fold the organization can be made anytime by any of its members. A virtual organization does not have one president or general manager. The virtualization process is a process of transformation from traditional organization into virtual using informational technology. This process can result in two forms of virtual organization: traditional organization with virtual divisions or an association of many organizations, as depicted in Figure 3 (Kisielnicki, 2002, p. 102).

Virtual Organization (defined by Schozch)**:** An artificial organizational structure where individual organizations provide base competencies. The integration and coordinations of these base competencies allow for effective execution of chain process to deliver a product and satisfy a customer. This integration and coordination does not incur additional costs and maintains the customer focus (Scholzch, 1996).

Virtual Organization: A temporary network of independent enterprises—suppliers, customers, ever former competitors—linked through IT and working together to share their knowledge and costs to enter a new market (Byrne & Brandt, 1993).

VRML-Based System for a 3D Virtual Museum

José-Fernando Díez-Higuera
University of Valladolid, Spain

Francisco-Javier Díaz-Pernas
University of Valladolid, Spain

INTRODUCTION

In the last few years, because of the increasing growth of the Internet, general-purpose clients have achieved a high level of popularity for static consultation of text and pictures. This is the case of the World Wide Web (i.e., the Web browsers). Using a hypertext system, Web users can select and read in their computers information from all around the world, with no other requirement than an Internet connection and a navigation program. For a long time, the information available on the Internet has been series of written texts and 2D pictures (i.e., static information). This sort of information suited many publications, but it was highly unsatisfactory for others, like those related to objects of art, where real volume, and interactivity with the user, are of great importance. Here, the possibility of including 3D information in Web pages makes real sense.

As we become an increasingly visual society, a way to maintain heritage is to adapt museums to new times. The possibility of not only visiting and knowing the museums nearby but also enabling anybody to visit the building from their homes could be enabled. This would imply the incorporation of the virtual reality (Isdale, 1998), although today few museums allow this kind of visit via Internet. In virtual reality, human actions and experiences that interact with the real world are emulated although, obviously, with some limitations. With virtual reality the user could walk, examine and interact with the environment, in contrast to traditional media like television that present excellent graphics but lack interactivity. Although this is not a new idea, it is achieving a wider expression due to the availability of the Virtual Reality Modeling Language (VRML) (Carey, Bell, & Marrin, 1997), a widespread language for the description of 3D scenes and WWW hyperlinks (an analogy of the HTML for virtual reality).

VRML is, perhaps, most interesting for Internet users eager to discover new interesting sites in Internet, and for the people that use it like hobby, but VRML could also allow us to see a 3D artifact from any angle and perspective, to turn it in any way, and so on. In sum, we can manipulate it — something totally forbidden in a real museum.

This work deals with the design of a system which allows this interactive Web access to works of art in 3D, as a step in a research project dealing with the design and implementation of a virtual and interactive museum in 3D on the Web. Also, all the associated information like history, architectural data, archaeological data, and culture will be available at the click of a mouse.

BACKGROUND

Several museums around the world are already committed to a strong Web presence and many others will adopt one very soon. Dynamic museum leaders understood that the increasing number of internauts requires special attention from museums: Internet — and CD-ROM's — represent new media that will challenge museum communication strategies.

According to Proença, Brito, Ramalho, and Regalo (1998):

Two distinct Web approaches are being adopted by the museums. Some regard their presence on the Web as another way to publicize the museum and to promote their activities; others use the Web as a powerful resource to achieve their purposes: to conserve, to study and to display.

The most common attitude is to consider the Web as a simple sum of the different kinds of information already in use by museums — specially printed information — but gathered in a global structured way. These data include a museum description and a list of activities and collections, where a typical Web page structure contains: collections and exhibitions, visit planning and conditions, new acquisitions, projects and activities, museum organizational scheme and educational programs. Several museums on the Web follow this approach. Among them it may be worth a visit to Museo Arqueológico Nacional of Madrid (http://www.man.es/), On-line Picasso Project (www.tamu.edu/mocl/picasso/), Museo de Cera de Madrid (www.museoceramadrid.com), Asian Art Museum of San Francisco (www.asianart.org), The Museum of Modern

Art (www.moma.org) and Library of Congress Vatican Exhibit (www.ibiblio.org/expo/ vatican.exhibit/exhibit/ Main_Hall.html); this site has a good image quality, but with a traditional structure to present the exhibition themes.

Some museums demonstrate greater innovation in their Web presences: They have temporary exhibitions online, promote virtual visits and access to their databases, present technical information for museums professionals and researchers, keep available information about previous activities and exhibitions, and organize links to related sites. For these museums, the Web is also an exhibition and a presentation medium that must be integrated in the communication policy of the museum. Among them, it may be worth a visit to Musée des Beaux Arts de Montréal (www.mbam.qc.ca/ index.html), The Museum of Anthropology at University of British Columbia (www.moa.ubc.ca/Collect/moaview.html), and Museo del Prado (museoprado.mcu.es/prado/html/home.html).

Latest advances are becoming popular 3D (plus color) scanners, which allow the measurement of 3D artifacts such as art works (Gómez, Díaz & López, 1997; Rocchini, Cignoni, Montani, Pingi & Scopigno, 2001). After measuring, a 3D plus color model from the real object can be obtained. 3D scanning technology has been adopted in a number of recent projects in the framework of cultural heritage. Just to give some examples, we may cite the Digital Michelangelo Project of the Stanford University (Levoy et al., 2000) , the acquisition of a Michelangelo's Pietá by a team of the IBM T.J. Watson Research Center (Rushmeier, Bernardini, Mittleman & Taubin, 1998) , or the acquisition of a section of the Coliseum in Rome (Gaiani et al., 2000) . Unfortunately, a detailed 3D (plus color) model of a free form object usually requires a great amount of data. This data can hardly pass through the Web, even when using compression. Therefore, additional reduction of transmission requirements is desirable.

Recently developed image-based modeling and rendering techniques (Chen, 1995; Chen & Williams, 1993) have made it possible to simulate photo-realistic environments. Two of the most popular image-based modeling and rendering techniques are Apple's QuickTime VR and the Virtual Reality Modeling Language (VRML). QuickTime VR (Chen, 1995) has its roots in branching movies, for example, the movie-map (Lippman, 1980), the Digital Video Interactive (DVI) (Riply, 1989), and the "Virtual Museum" (Miller et al., 1992). QuickTime VR uses cylindrical panoramic images to compose a virtual environment, therefore provides users an immersive experience. However, it only allows panoramic views at separate positions. Recently, the Virtual Reality Modeling Language (VRML) (Carey et al., 1997) became a standard file format for the delivery of 3D models over the Internet. Subsequently, many efforts have been made to effectively compress and progressively transmit the VRML files over the Internet (Deering, 1995; Li & Kuo, 1998; Matsuba & Roehl, 1999; Taubin, Horn, Lazarus, & Rossignac, 1998).

Using these techniques, some systems allow to see art works in 3D (Cignoni, Montani, Rocchini, & Scopigno, 2001), while others allow a virtual walk through the rooms of some real building as The Virtual Living Kinka Kuji Tempers (Refsland, Ojika, & Berry, 2000), some reconstructed scenario as the Historic Villages of Shirakawa-go (Hirayu, Ojika, & Kijima, 2000) or some imaginary buildings as Virtual Museum of Helsinki (www.virtualhelsinki.net/museum) and Virtual Museum of Art (www.elpais.com.uy/muva).

The main feature of our system is that users may walk through a three-dimensional (3D) representation of the whole Fabio Neri's Palace, the building where Museum of Valladolid is located, viewing its collections, and seeing pictures in 2D and archaeological objects in 3D, together with information about them. To allow all of this, an architecture of interactive dynamic Web pages has been designed (Díez-Higuera & Díaz-Pernas, 2002). In order to capture 3D information, we have used the Laser Acquisition System developed by the Industrial Telematic Group of Telecommunications Engineering School of Valladolid (Gómez et al., 1997). These data, together with 2D images and information files, are compressed and stored in a remote server, and can be retrieved over the Internet. Rather than transmitting a high-resolution object, our system at the client end allows users to selectively retrieve images at specific resolutions. This selective retrieval is achieved by implementing a client-server communication protocol. Information is accessed through intuitive exploration of the site, and therefore, each session varies depending on both the participant and the path chosen. In this manner, the visitor becomes familiar with the virtual museum, in much the same way as they would become familiar with the physical museum. User may identify particular areas of interest, which may be revisited using familiar routes or accessed via browsing.

DESCRIPTION OF THE SYSTEM

Figure 1 shows the general architecture of the system. It has two main parts: the Dynamic Web pages system, based on the *Microsoft Internet Information Server*, which embraces the virtual visit to the 3D Museum and the access to data and its visualization in 3D; and the platform of telematic services, which implements the server-client architecture, allowing the transmission of three-dimensional and colorimetric data of the objects in several resolutions.

Client-Server Architecture

The server is used to store the huge amount of scene data that is being requested by the user, while the client is designed to interact with the user and retrieve the necessary scene data from the server. Java programming language has been chosen for the implementation of the server-client architecture. The reason is that Java allows the introduction of executable codes in Web pages, and, therefore, giving the desired portability, security and interactivity to the system.

Dynamic Web Pages System

A dynamic Web pages system has been implemented to give access to the database so the information required by the user about any object can be shown. This system has also been used for a 3D virtual walk through the museum.

Each section, or room, implemented in the museum becomes a specific VRML file (sometimes, even a single room is divided in several VRML files). Using smart nodes from VRML language, which activate any element in a 3D universe when the user clicks on it, the file stored in the URL of the VRML filecode shall be loaded, interconnect these files. Once the user is in a VRML file (containing any of the rooms where both 2D objects — pictures — and 3D ones are displayed) (s)he can walk around the room at will: Approach any object to have a first impression and, if wanted, click on it to acquire information about it (author, period, technique, etc.) as well as to visualize it in 3D (or with a higher resolution still image if the object is a picture). When the user clicks on any object, the dynamic Web pages system starts to work, giving access to the database and bringing the required information.

Format of the Data Used by the System

Virtual Reality Modeling Language (VRML) is the working format for the 3D display of the objects, as it is standard and commonly used in the Web. First of all, 3D plus color data must be acquired from art works. The corresponding object model, obtained from the acquired data, could then be directly expressed in VRML format. However, the raw format of the data in our database is not VRML. Instead, it is the specific one given by the Laser Acquisition System developed by the Industrial Telematic Group of the Telecommunications Engineering School of Valladolid (Gómez et al., 1997). This system, starting from a real object, gives a *.pie* file with the three-dimensional data, plus another three files (*.r, .g, and .b*) with its colorimetric information. Those files are converted into ASCII files, using software developed by the Industrial Telematic Group; these new files are *.pts*, with the Cartesian coordinates of all the points as read from the object, plus its colors, and *.tri*, with the triangular distribution of points for a certain resolution (which allows the construction of faces for the display of the object).

The server stores the last couple of files: one *.pts* file for each object, and several *.tri* (one for each different resolution). From them, and given a certain resolution, the server obtains the files needed by the client to reconstruct the VRML file. It is important to notice that there is no duplication of information in the server, as points and colors appear only once in the *.pts* file and the set required by the client at each moment is obtained from it. There are only several files of triangles, as the distribution of points on the faces is different for each resolution. Moreover, three-dimensional and colorimetric information is sent only once to the client. The result is that this design, as a whole, improves the efficiency of the system.

Also, we have developed an alternate solution which allows lower transmission requirements. In short, we allow art works to be requested in different levels of detail (LOD). User begins with the lowest LOD model, and requests a set of progressively increasing LOD models as his interest on the object increases. We benefit from this by building the different LOD models in particular way: Each LOD model is obtained from the immediately higher LOD model, by just picking some of its points. In this way, when the users ask for a higher LOD model, the whole model transmission is no more required. Instead, the new model can be reconstructed by adequately merging new points into the previously existing model. Unfortunately this strategy is not implemented in actual VRML; so we have implemented it by using a dedicated Java client. Basically, when the user asks for a superior LOD model, only additional 3D points, and a new faces description, are sent. This faces description must then be translated by the client. This is done by means of a *Look-Up Table* (LUT), which indicates actual position of the new points in a local VRML file.

User Interface

The user interface module is based on Web page template in which the virtual world is represented on a part of the screen, while on the rest, the name of the content plus the options that the user can select are left. We have a basic system in which the user can move along the building and rooms and, already inside these rooms, the user can select a picture or a 3D artifact in order to get information about it.

Different perspectives of the museum are shown in Figures 2, 3, and 4.

Figure 1. Global architecture of the proposed system

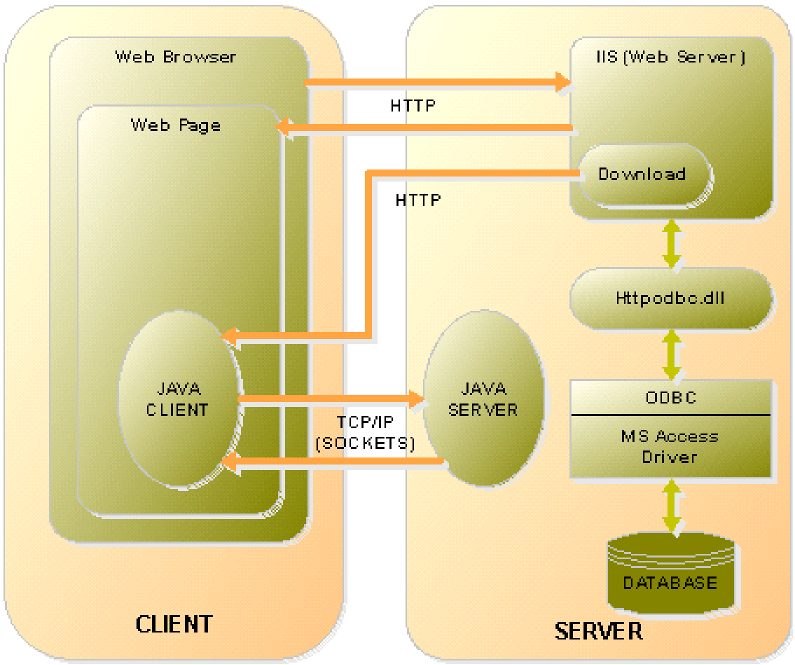


Figure 2. Initial Web page of the virtual museum: View of Fabio Neri's palace

Figure 3. General view from the patio

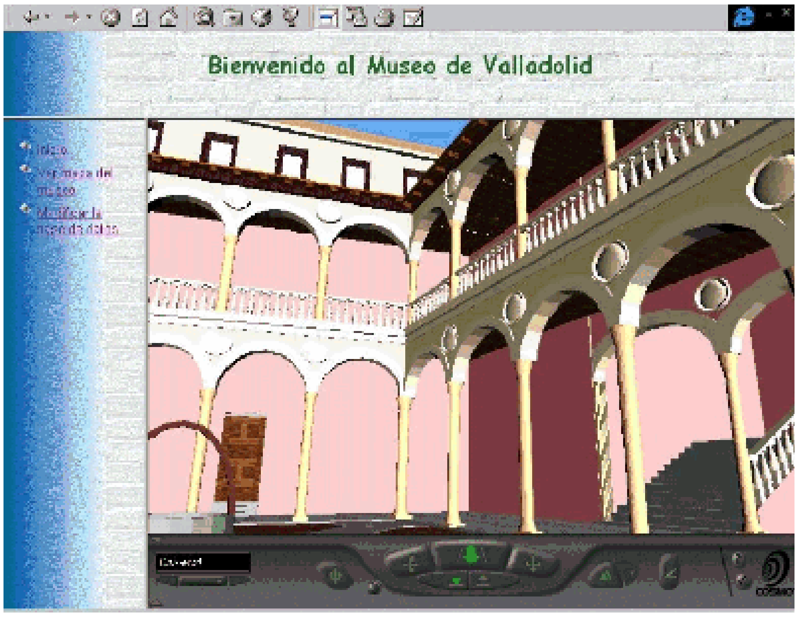
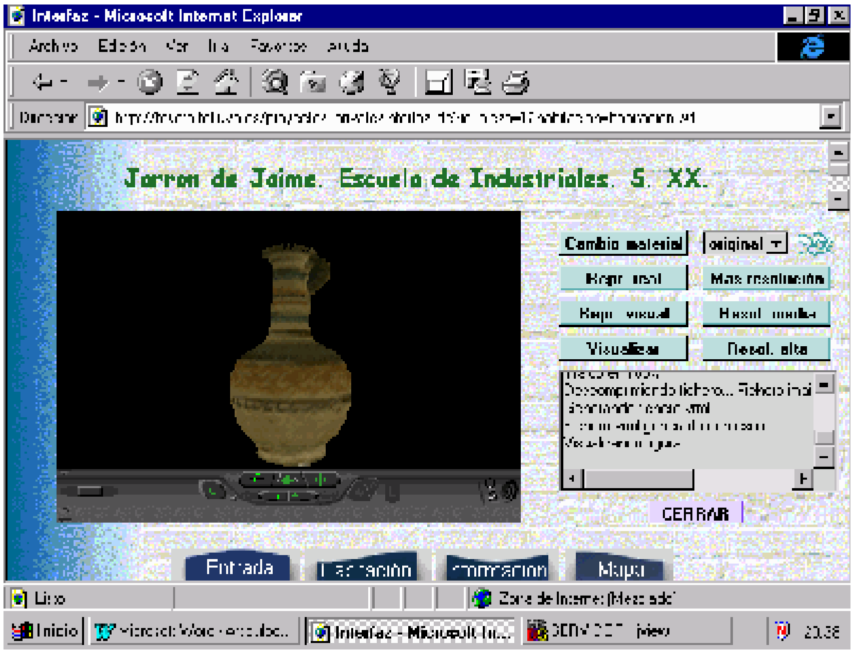


Figure 4. Initial view of the virtual room

Figure 5. An example of 3D artifact



One of the features of the proposed system is that it allows the 3D display of an artifact in VRML format in different levels of resolution (see Figure 5), which implies that the user can manipulate it as he pleases.

FUTURE TRENDS

Our system marks the advent of a new and interactive form of documenting, referencing, and archiving heritage structures. Future work involves advanced development of this system as follows:

- Include dynamic elements like 3D surround sound, voice-overs for history, culture, and other information along with traditional audio pieces to make a complete user-friendly interactive interface.
- Increase the number of works of art and information in texts in order to give a more detailed presentation of archaeology in Spain.
- Achieve better frame rates by using clever culling techniques and modeling techniques for a smooth walk-through.
- Incorporate accurate details by using the latest photogrammetric techniques.
- Depict “as was”, “as is”, and “as it could be” transitions for studies specific to architecture, conservation, games for children and light simulations.

CONCLUSION

Our project resulted in the following observations:

- We achieved a fairly accurate 3D model of a complex heritage structure using archaeological orthographic projections as reference material.
- Construction of a virtual museum is possible, allowing users to examine, conduct research, or navigate any part of the building at their own convenience where they not only can see photos, and 3D objects, and even have the opportunity to play with them.
- The viability of the system has been demonstrated, as well as its correct operation in the net in the particular case of the Museum of Valladolid.

REFERENCES

Carey, R., Bell, G., & Marrin, C. (1997, April). The virtual reality modeling language. ISO/IEC DIS 14772-1.

Chen, S.E. (1995, August). QuickTime VR – An image-based approach to virtual environment navigation. In *Computer Graphics Proceedings, Annual Conference Series, Proc. SIGGRAPH 1995 (Los Angeles)* (pp. 29-38).

Chen, S.E., & Williams, L. (1993, August). View interpolation for image synthesis. In *Computer Graphics Proceed-*

ings, Annual Conference Series, , Proc. SIGGRAPH 1993 (California) (pp. 279-288).

Cignoni, P., Montani, C., Rocchini, C., & Scopigno, R. (2001, July). Acquisition and management of digital 3D models of statues. In *Proceedings 3rd International Congress on Science and Technology for the Safeguard of Cultural Heritage in the Mediterranean Basin, Alcalá de Henares (Spain)* (pp.1159-1164).

Deering, M. (1995, August). Geometry compression. In *Computer Graphics Proceedings, Annual Conference Series (Proc. SIGGRAPH)* (pp.13-20).

Díez Higuera, J.F., & Díaz Pernas, F.J. (2002). VRML-based system for a three-dimensional virtual museum. In T.K. Shih (Ed.), *Distributed multimedia databases: Techniques & applications* (pp. 306-317). Hershey, PA: Idea Group Publishing.

Gómez García-Bermejo, J., Díaz Pernas, F.J., & López Coronado, J. (1997). Industrial painting inspection using specular sharpness. In *Proc. of the International Conference on Recent Advances in 3-D Digital Imaging and Processing (IEEE Press)*, April (pp.335-338). Ottawa, Canada,

Hirayu, H., Ojika, T., & Kijima, R. (2000). Constructing the historic villages of Shirakawa-go in virtual reality. *IEEE Multimedia,* ***7*(2),** 61-64.

Isdale, J. (1998, September). What is virtual reality. A Web-based introduction Version 4, Draft 1. Retrieved from the World Wide Web at http://vr.isdale.com/WhatIsVR.html

Levoy, M., Pulli, K., Curless, B., Rusinkiewicz, S., et al.(2000). The digital Michelangelo project: 3D scanning of large statues. In *Comp. Graph. Proc., Annual Conf. Series (Siggraph '00), ACM SIGGRAPH*, July 24-28 (pp.131-144). Addison Wesley.

Li, J.K., & Kuo, C.J. (1998, June). Progressive coding of 3D graphics models. *Proceedings of IEEE, 86*(6), 1052-1063.

Lippman, A. (1980). Movie maps: An application of the optical videodisc to computer graphics. In *Computer Graphics (Proc. SIGGRAPH)*, (pp. 32-43).

Matsuba, S.N., & Roehl, B. (1999, Spring). Bottom, thou art translated: The making of VRML dream. *IEEE Computer Graphics and Applications, 19*(2), 45-51.

Miller, G., Hoffert, E., Chen, S.E., Patterson, E., Blackketter, D., Rubin, S., Applin, S.A., Yim, D., & Hanan, J.(1992). The virtual museum: Interactive 3D navigation of a multimedia database. *The Journal of Visualization and Computer Animation, 3*, 183-197.

Proença, A., Brito, M., Ramalho, T., & Regalo, H. (1998). Using the Web to give life to museums [CD-ROM]. In *Proceedings of the Museums and the Web*, Toronto, Canada.

Refsland, S.T., Ojika, T., & Berry, Jr., R.(2000). The living virtual Kinka Kuji Temple: A dynamic environment. *IEEE Multimedia Magazine, 7*(2), 65-67.

Riply, D.G. (1989). DVI- A digital multimedia technology, *Communications of the ACM, 32*(7), 811-822.

Rocchini, C., Cignoni, P., Montani, C., Pingi, P., & Scopigno, R. (2001). A low cost scanner based on structured light. In *Computer Graphics Forum. Eurographics 2001 Conference Proceedings, 20*(3), (pp.299-308).

Rushmeier, H., Bernardini, F., Mittleman, J., & Taubin, G. (1998). Acquiring input for rendering at appropriate levels of detail: Digitizing a piet'a. In G. Drettakis, & N. Max (Eds.), *Rendering techniques '98* (pp. 81-92). Ed. Springer Wien.

Taubin, G., Horn, W.P., Lazarus, F., & Rossignac, J. (1998, June). Geometry coding and VRML. *Proceedings of IEEE, 86*(6), 1228-1243.

KEY TERMS

Client-Server Architecture: A network architecture in which each computer or process on the network is either a client or a server. Servers are powerful computers or processes dedicated to managing disk drives (file servers), printers (print servers), or network traffic (network servers). Clients are PCs or workstations on which users run applications. Clients rely on servers for resources, such as files, devices, and even processing power.

Dynamic HTML: A collective term for a combination of new HTML tags and options, style sheets and programming, which enable you to create web pages that are more interactive and faster to download.

Java: A platform-independent programming language, produced by Sun Microsystems. Java is built as a method to provide services over the WWW. With Java, a Web site provides a Java application (called an applet) which is downloaded by the client and executed on the client machine. Java is specifically built so that an application can be run on any kind of system, so a separate Mac, Windows, Sun, and so forth, version is not needed. Java also has some security features built in, to make it more difficult for destructive applets to be written.

QuickTime VR: An enhanced version of the standard developed by for displaying content (animation, audio, and) on computers. This enhanced version adds the ability to display and rotate objects in three dimensions.

URL: The address of a computer or a document on the Internet that consists of a communications protocol followed by a colon and two slashes (as http://), the identifier of a computer (as www.m-w.com) and usually a path through a directory to a file " called also *uniform resource locator, universal resource locator*.

Virtual Museum: A collection of digitally recorded images, sound files, text documents, and other data of historical, scientific, or cultural interest that are accessed through electronic media. A virtual museum does not house actual objects and therefore lacks the permanence and unique qualities of a museum in the institutional definition of the term.

Virtual Reality (VR): The use of computer modeling and simulation to enable a person to interact with an artificial three-dimensional visual or other sensory environment. VR applications immerse the user in a computer-generated environment that simulates reality through the use of interactive devices, which send and receive information and are worn as goggles, headsets, gloves, and so forth.

Virtual Reality Modeling Language (VRML): A programming language for the creation of virtual worlds. Using a VRML viewer, you can take a virtual tour of a 3D model building, or manipulate animations of 3D objects. Hyperlinks to other sites and files can be embedded in the world you visit.

Web Access by Older Adult Users

Shirley Ann Becker
Florida Institute of Technology, USA

INTRODUCTION

The older adult population in the U.S. continues to increase at a rapid pace due to aging baby boomers and increased life expectancy. Older Americans, 60 years and older, will comprise about 20% of the total population by 2030, which is more than twice the number of aging adults than in 2000 (Administration on Aging, 2002).

The Web offers an unprecedented opportunity for older adults to access a wealth of online resources. Increasingly, older adults are using the Web for information on self-diagnosis, treatment, and prevention of health problems (Preidt, 2003). They are taking advantage of electronic government resources to vote, file taxes, obtain social services, voice their opinions to officials, and search historical records. Older adults are also using the Web to stay socially active in terms of communicating with other users via lists and chat rooms (Czaja, Guerrier, Nair & Landauer, 1993; Kerschner & Hart, 1984).

Older adults are getting online by an estimated growth rate of 15% per year (Coulson, 2000). They log over eight hours of online time per week and visit more Web sites than persons in younger age groups when they are online (Morrell, Dailey, Feldman, Holt, Mayhorn & Echt, 2002). Their use of the Internet is predicted to increase as much as 358%, from 3.7 million users in 2001 to 17.3 million in 2005 (Scanlon, 2001).

Unfortunately, older adults may have trouble accessing a Web site because of design issues that impede its use. Barriers may be encountered due to a site's color scheme, font size and type, navigation, vertical screen length, image mouseovers, and sentence complexity, among others. In this information-rich society, many older adults will remain "information have-nots" unless these barriers are removed.

BACKGROUND

In the U.S., the Internet has emerged as a major communications medium with the potential to disseminate information to all citizens including older adults. Yet, there is an ongoing concern that the opportunities associated with Internet access may not be readily available to all citizens. This concern has been expressed in a recent Pew Foundation study (Lenhart, Horrigan, Rainie, Allen, Boyce, Madden & O'Grady, 2003, p. 6):

"Internet non-users will have less power as consumers and fewer economic opportunities, less access to high-quality health information, fewer options for dealing with government agencies, no chance to learn about their world from the millions of organizations and learning centers that have posted their material on the Web, and less opportunity to interact with others through email and instant messaging."

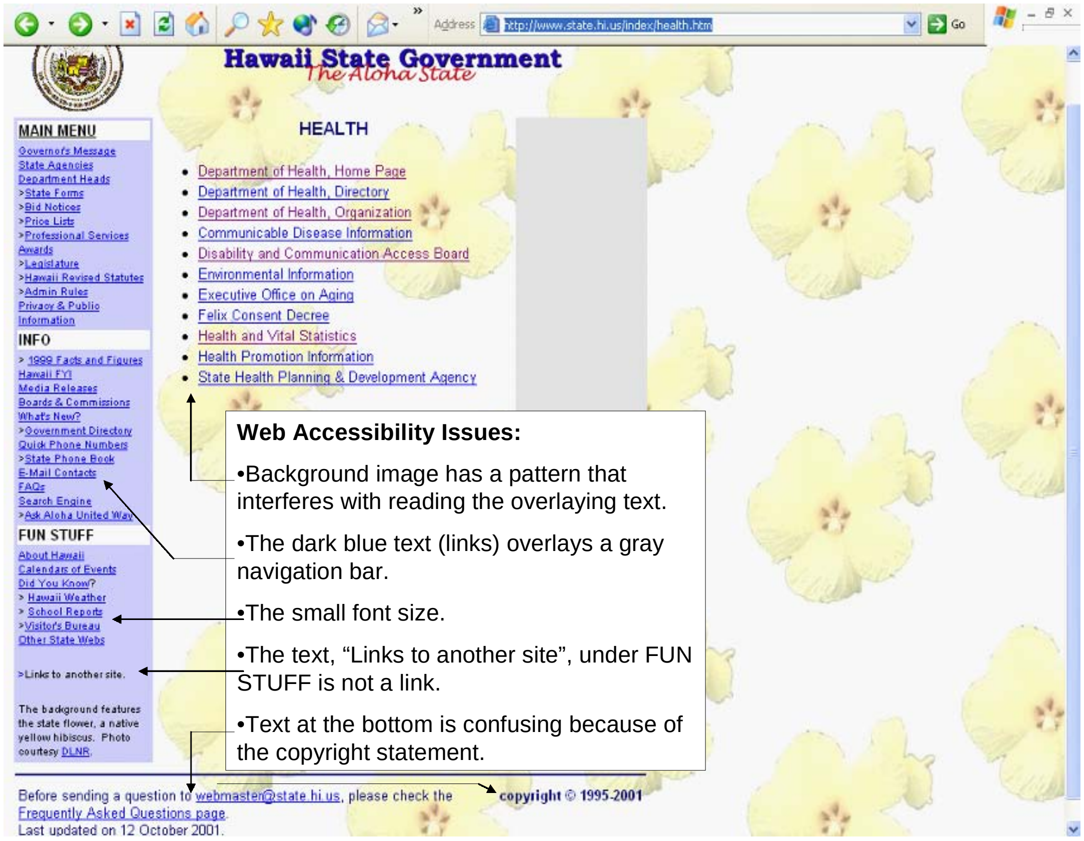
Older adults in particular may encounter Web accessibility barriers due to vision, cognition, and physical changes that are associated with the normal aging process. Reading complexity may also become a barrier when literacy skills of older adults are not taken into account in the design of Web content.

Vision

The aging eye has a reduced ability to focus on close objects due to a reduction in the elasticity in the lens. Other vision changes due to the normal aging process include: a decline in visual acuity, impacting the ability to see objects clearly, yellowing and thickening of the lens, impacting color perception, decreased light sensitivity, impacting adaptation to changes in light levels, increased sensitivity to glare from light reflecting or shining into the eye, and reduced depth perception, making it more difficult to judge the distance of an object (American Foundation for the Blind, 1999). These vision changes impact the use of the Web in terms of the legibility of written content on the page. They also impact searches, navigation, and reading speed and comprehension (Echt, 2002).

Figure 1 illustrates readability issues associated with a state government Web page when taking into account aging vision. The patterned background image may negatively impact readability, especially given the small size of the foreground text[1]. The format of the text at the bottom of the page also impacts readability given that the sentence "Before sending a question to <webmaster>, please check the Frequently Asked Questions page" breaks to

Figure 1. Illustration of Web barriers and aging vision (http://www.state.hi.us/index/health.htm)



accommodate the copyright statement appearing on the right.

The use of color can also impact the readability of information content on a Web page due to aging vision (refer to Becker, 2004a for a discussion on color and Web accessibility for older adults). For many older adults, foreground and background color combinations may render a Web page visually inaccessible. In Figure 1, the contrast between the text and background colors in the navigation bar may be insufficient for older adult readers. Figure 2 shows the New Mexico state government homepage with saturated colors for both the foreground and background. The edges of the text tend to blur when bright or neon colors are used in combination with saturated colors (e.g., bright yellow text displayed on a red background), thus reducing legibility of the text for many older adult users.

Cognition

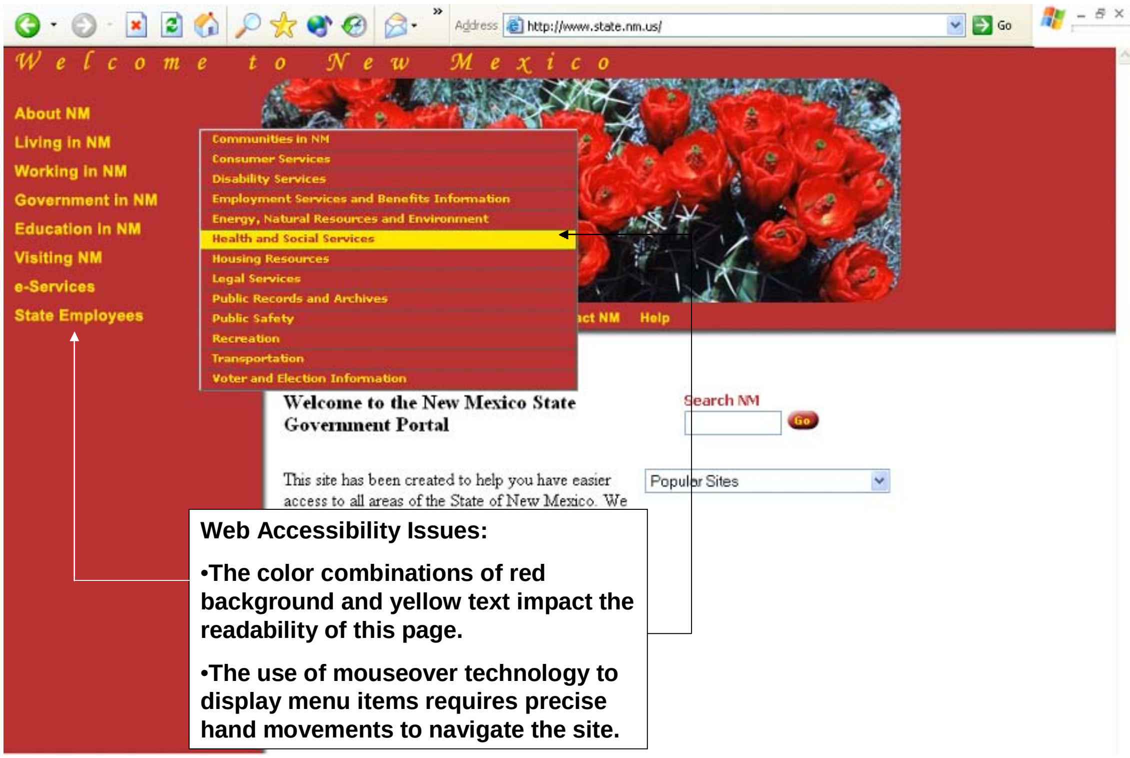
Studies show that an older adult's working and spatial memory task performance declines with age (Holt & Morrell, 2002). As a result, an older adult may not be able to discern details in the presence of distracting information. In addition, complex navigational schemes, nonintuitive searches and cluttered pages may impede use of a Web site because of declines in working and spatial memory.

Of 40 U.S. state government sites assessed, over 60% required traversing three or more screen pages to navigate directly to resources for older adults (Becker, 2004b). Less than 8% of these sites had a descriptive link on the homepage linking to senior resources. Those Web sites having nondescript links required trial and error searches for senior resources. This type of navigational complexity impedes the use of electronic government by many older adult users.

Physical Impairments

Older adults experience a decrease in motor coordination, and as such, may have difficulty with cursor positioning, precise mouse movement, and clicking on links (Chaparro, Bohan, Fernandez, Choi & Kattel, 1999; Ellis & Kurniawan, 2000; Hawthorne, 2000). Figure 2 shows a Web page that requires the precise use of mouseover technology in order

Figure 2. Illustration of Web barriers due to color and mouseovers (http://www.state.nm.us)



to navigate the Web site. This site becomes virtually inaccessible to those users who cannot precisely click on a link and move the mouse to the pop-up menu.

Literacy

The Joint Committee on National Health Education Standards (1995) defines health literacy as the capacity of an individual to obtain, interpret, and understand basic health information and services and the competence to use such information and services in ways that are health-enhancing. This includes the ability to understand instructions on prescription drug bottles, appointment slips, medical education brochures, doctor's directions and consent forms, and the ability to negotiate complex health care systems. In today's wired world, health literacy is expanded to include the comprehension of online health information.

The Web is transforming health care by providing online health information to patients, clinicians, caregivers, and family members. There are hundreds of health-related Web sites offering access to unprecedented amounts of health information. Because of their heavy use of health services, older adults could be among the major beneficiaries of Web-accessible health information. But, when the literacy requirements of these pages are misaligned with the literacy skills of older adults, the results could be devastating.

The Web offers an extraordinary opportunity to disseminate timely and much needed health care information to older adults. This opportunity is expanding further with technological advances in wireless communication and increasing availability of community computing resources through institutions such as nursing homes, schools, and libraries.

Older adults in general have a lower education level than younger adults according to the U.S. Census (www.census.gov). The National Adult Literacy Survey found that 71% of older adults performed in the lowest two levels of literacy defined in the survey (Kirsch, Yamamato, Norris, Rock, Jungeblut & O'Reilly, 2001). Approximately 67% of older adults appeared to have difficulty with finding and processing quantitative information when compared to younger adults (Brown, Prisuta, Jacobs & Campbell, 1996).

The reading complexity of content targeting older adults is illustrated by the following sentence appearing on the Alabama state Web site (http://www.adss.state.al.us/seniorrx.htm):

Table 1. NIA/NLM guidelines for making senior-friendly Web sites

Guideline	**Description**
Sans serif typeface	Sans serif font types should be used to display information content because they are not condensed.
12 point or greater font size	The use of a large font size improves legibility of information content such that text body, buttons, links, images, and other textual objects are readily seen by an older adult.
Mixed case letters in text body	The text body should be in mixed case text to improve readability. Upper case text should be reserved for headlines on a page.
Left justification	Text should be left justified because spacing between letters is consistently the same.
Plain background images	Patterned background images should be removed from a Web page because they reduce the legibility of text overlaying them.
Text effects only in headlines	Text effects including underlining, italics, bold, or strikethrough should not be used in the body of the text.

"In case you're not familiar with the program, SenioRx is a partnership of state agencies and community organizations designed to assist senior citizens (ages 60 and older) with chronic medical conditions who have no prescription insurance coverage and limited financial means (living below 200% of the poverty level) with applying for drug assistance programs provided by pharmaceutical manufacturers."

The reading grade level associated with this sentence is far beyond a 12th grade level because it is composed of 58 words. In addition, about 33% of the sentence is composed of three or more syllable words, adding to its complexity. Though there is only one sentence in this sample, it illustrates the reading comprehension barriers that older adults may encounter while surfing the Web.

WEB ACCESSIBILITY INITIATIVES

The National Institute on Aging in conjunction with the National Library of Medicine has developed Web accessibility guidelines for making sites senior-friendly (NIA & NLM, 2001). These guidelines are based on scientific findings from research in aging and cognition and human factors (Morrell et al., 2002). They provide information on how to improve the design, navigation, and information content of Web sites to remove accessibility barriers for older adult users. Table 1 lists several of these guidelines for promoting the development of accessible Web sites. Note that several of these guidelines, font size, and patterned background images were not followed in the design of the government Web page previously shown in Figure 1.

Another initiative is the National Cancer Society's usability.gov Web site, which provides information about making Web content more usable, accessible, and useful (*www.usability.gov*). It provides research and practitioner-based guidelines on design layout, navigation, information content, and other usability aspects of Web design targeting the general population of computer users. Though it does not specifically target the usability needs of older adults, many of these resources will improve Web site usability for this user group.

Nonprofit groups, including the SPRY foundation (*www.spry.org*), Seniornet (*www.seniornet.org*), and AARP (*www.aarp.org*), provide online resources and support research in Web accessibility for older adults, health literacy, and related areas. Much of the research conducted by these groups provides feedback on the digital divide, health literacy issues, community resources, and the status of older adults getting and staying online.

From a broader user perspective, there have been initiatives in promoting Web accessibility for those with disabilities. Microsoft, IBM, Apple, Watchfire and other corporations have provided resources, tools, and techniques for improved accessibility of Web designs. The Trace Research and Development Center at the University of Wisconsin (*www.tracecenter.org*) is one of several centers that focus on universal usability in order to make technology accessible to anyone, anytime, and anyplace. The ACM SIGCAPH (Special Interest Group on Computers and the Physically Handicapped) Web site provides a full listing of Web accessibility resources (*http://www.hcibib.org/accessibility/#ORGANIZATIONS*).

Section 508, an amendment to the 1973 Rehabilitation Act, was enacted by the U.S. government in order to eliminate information barriers for those persons with disabilities. It requires that individuals with or without disabilities have equal access to information provided by federal agencies (www.Section508.gov). Web content guidelines have been put forth by the World Wide Web Consortium (www.w3c.org/WAI/) in order to eliminate accessibility barriers. Though Section 508 does not specifically address the barriers to Web use due to the normal aging process, many guidelines on design layout, information content, navigation, and design consistency improve usability from an older adult perspective.

CONCLUSION

Though significant strides have been made to promote Web accessibility for older adults, there are still barriers to overcome. Web accessibility and online literacy research requires further study in determining an optimal presentation of content to meet the literacy needs of older adult users. The NIA/NLM guidelines recommend active versus passive voice sentence structures, short sentences, and appropriate reading grade levels. Because these guidelines are rather vague, it may be difficult to enforce them during the design of Web content. For example, what is a sufficiently "short" sentence in terms of word count and syllable complexity? Cultural diversity and English proficiency of the older adult population also require further study in terms of potential barriers to Web use.

REFERENCES

Administration on Aging. (2002). A profile of older Americans: 2002. Administration on Aging, U.S. Department of Health and Human Services, http://www.aoa.gov/ prof/ Statistics/profile/2002profile.pdf

American Foundation for the Blind. (1999). Normal changes in the aging eye fact sheet. Retrieved June 20, 2003, from http://www.afb.org/info_document_view.asp?documentid=203

Becker, S.A. (2004a). E-government visual accessibility for older adult users. *Social Science Computer Review*, *22*, 1.

Becker, S.A. (2004b). Architectural accessibility and reading complexity of U.S. state e-government for older adult users. Forthcoming in *Electronic Government.*

Brown, H. Prisuta, R. Jacobs, B., & Campbell, A. (1996). *Literacy of older adults in America: Results from the national adult literacy survey*. U.S. Department of Education, National Center for Education Statistics, NCES 97-576, Washington DC.

Chaparro, A., Bohan, M., Fernandez, J.E., Choi, S.D., & Kattel, B. (1999). The impact of age on computer input device use: Psychophysical and physiological measures. *International Journal of Industrial Ergonomics*, *24*, 503-513.

Coulson, I. (2000). Introduction: Technological challenges for gerontologists in the 21st century. *Educational Gerontology*, *26*, 307 - 315.

Czaja, S.J., Guerrier, J.H., Nair, S.N., & Landauer, T.K. (1993). Computer communications as an aid to independence for older adults. *Behaviour and Information Technology*, *12*, 197 - 207.

Echt, K.V. (2002). Designing Web-based health information for older adults: Visual considerations and design directives. In R.W. Morrell (Ed.), *Older adults, health information, and the World Wide Web* (pp. 61 - 88). Mahwah, NJ: Lawrence Erlbaum Associates.

Ellis, R.D., & Kurnaiwan, S.H. (2000). Increasing the usability of on-line information for older users. A case study in participatory design. *International Journal of Human-Computer Interaction*, *12*, 263-276.

Hawthorne, D. (2000). Possible implications of aging for interface designers. *Interacting with Computers*, *12*, 507-528.

Holt, B.J., & Morrell, R.W. (2002). Guidelines for Web site design for older adults: The ultimate influence of cognitive factors. In R.W. Morrell (Ed.), *Older adults, health information, and the World Wide Web* (pp. 109-129). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Joint Committee on National Health Education Standards. (1995). *National health education standards.* Available from the American School Health Association, P.O. Box 708, 7263 State Route 43, Kent, OH 44240.

Kerschner, P.A., & Chelsvig Hart, K.C. (1984). The aged user and technology. In R.E. Dunkle, M.R. Haug & M. Rosenberg (Eds.), *Communications technology and the elderly: Issues and forecasts* (pp. 135-144). New York: Springer.

Kirsch, I., Yamamato, K., Norris, N., Rock, D., Jungeblut, A., & O'Reilly, P. (2001). *Technical report and data file users manual for the 1992 national adult literacy survey*. Washington DC: National Center for Education Statistics, U.S. Department of Education, NCES 2001-457.

Lenhart, A., Horrigan, J., Rainie, L., Allen, K., Boyce, A., Madden, M., & O'Grady, E. (2003). *The ever-shifting*

Internet population. Washington DC: The Pew Internet and American Life Project.

Morrell, R.W., Dailey, S.R., Feldman, C., Mayhorn, C.B., & Echt, K.V. (2002). *Older adults and information technology: A compendium of scientific research and Web site accessibility guidelines.* Bethesda, MD: National Institute on Aging.

NIA & NLM. (2001). *Making your Web site senior friendly: A checklist.* National Institute on Aging and National Library of Medicine. Retrieved July 30, 2003, from *http://www.nlm.nih.gov/pubs/checklist.pdf*

Preidt, R. (2003). Seniors turning to Internet for health help. Retrieved September 23, 2003, from *http://www.healthscout.com/template.asp?page=newsdetail& ap=1&id=511832*

Scanlon, B. (2001, October). The future of the net: Surf's up for seniors. Retrieved August 21, 2003, from *http://www.eweek.com/article2/0%2C3959%2C950787%2C00. asp*

KEY TERMS

Digital Divide: The digital divide is a term used to describe the disparity between persons who have access to information and computing technology and those who do not. Often, it is used to describe the lack of Internet accessibility to those living in rural or remote areas or who lack computing knowledge and skills.

Electronic Government: Electronic government (e-government) refers to the use of information and computing technologies by government agencies to deliver services, information, and resources to citizens, businesses, and other organizations.

Health Literacy: Health literacy is the capacity of an individual to obtain, interpret, and understand basic health information and services and the competence to use such information and services in ways that are health-enhancing (refer to text box 1).

Older Adults: An older adult is defined as a person who is 60 years or older in the National Institute on Aging's guidelines on making senior-friendly sites.

Visual Aging: Visual aging takes into account age-related changes in vision that have consequences on daily activities. In this article, the consequences are related to using the Web.

Web Accessibility: Web accessibility means that any person, regardless of disabilities, is able to use Web technology without encountering any barriers.

Web Usability: Web usability refers to the user satisfaction associated with a Web site. It typically includes the effectiveness of the site in meeting the needs of the user. It also includes the site's performance, reliability, and overall efficiency in supporting specified user goals.

ENDNOTES

[1] The impact on readability is minimized when displaying in black and white the patterned background image and its overlaying text and foreground and background color combinations.

The National Science Foundation under Grant No. 0203409 supports this work. Any opinions, findings and conclusions or recommendations expressed in this content are those of the authors and do not necessarily reflect the views of the National Science Foundation. Web accessibility for older adult research efforts can be found at: http://www.cba.nau.edu/facstaff/becker-a/Accessibility/main.html

Web Accessibility and the Law

Holly Yu
California State University, Los Angeles, USA

INTRODUCTION

In the past three decades, the method of receiving and conveying information has shifted from paper-based, typewriter-generated, hand-edited, and printing-press-produced publications to more technology-mediated, intelligent, WYSIWYG software-generated forms. Consequently, the concept of access to information has changed to reflect this phenomenon. The new forms of access to information that have made it easier for non-disabled people have often created barriers for people with disabilities.

The notion of access to information involving the civil rights of people with or without disabilities arises from the fact that access to information through technology has increasingly become a necessary tool for success and the source of opportunity in education and employment. With the advent of information technology and the unprecedented opportunities they created for people with and without disabilities, it has become apparent that information technologies have a tremendous potential for allowing people with disabilities to participate in mainstream activities and to support their ability to live independently.

The disability rights movement in the United States originated during the post-World War II era when large numbers of veterans who were disabled in the war joined the efforts of parents seeking education and independent living options for their children with disabilities (Slatin & Rush, 2003). Recently, we have seen a growing body of significant laws, regulations, and standards concerning Web accessibility that impact people with disabilities and their ability to fully overcome digital barriers and participate in the Web environment. In defining Web accessibility, Section 508 of the Rehabilitation Act of 1973 as amended in 1998 documents that "Web sites are accessible when individuals with disabilities can access and use them as effectively as people who don't have disabilities" (Section 508, 1998). A person with a disability is defined in the Americans with Disabilities Act (ADA) as "someone who has a physical or mental impairment that substantially limits one or more major life activities, a person who has a record of such impairment, or a person who is regarded as having such impairment" (ADA, 1990). The legal foundation for protecting the civil rights of persons with disabilities has been established through a series of federal and state laws, and court decisions. These laws provide a legal ground on Web accessibility implementation.

LAWS, REGUALTIONS, STANDARDS, AND GUIDELINES

Under the provisions of laws, some of the legal milestones that have direct impact on Web accessibility are Section 504 of the Rehabilitation Act of 1973, Americans with Disabilities Act (ADA) of 1990, and Section 508 of the Rehabilitation Act of 1973, as amended in 1998.

Section 504, Rehabilitation Act, 1973

Signed on October 1, 1973, Section 504 of the Rehabilitation Act is regarded as landmark legislation and the first civil rights law prohibiting recipients of federal funds from discriminatory practices on the basis of disability.

Core areas of the legislation consist of the prohibition of such activities as discriminatory employment practices, and discrimination in the delivery of educational offerings; health, welfare, and social services; or any other type of programs benefit or service supported in whole or in part by federal funds.

Section 504 is currently applied to all entities that receive federal government funds, including states, counties, cities, towns, villages, and their political subdivisions, public and private institutions, public and private agencies, and other entities that receive federal money. Each federal agency has its own set of Section 504 regulations that guide its own programs. Over the years, the Rehabilitation Act has been amended several times to address the constant changes in technology and its impact on society. The amendments most relevant to the access to information technology are those made to Section 508. The significance of the Section 504 lies not only in that it was the first statute applying civil rights protections to people with disabilities, but that it also "furnished the model for major subsequent enactments, including the ADA" (NCD, 2001). Section 504 was legislated too early to specifically addresses the issue of access to services and programs provided over the Web.

Americans with Disabilities Act (ADA), 1990

Passed on July 26, 1990, the ADA establishes a clear and comprehensive prohibition of discrimination on the basis of disability. While Section 504 applies to federal government agencies and those that receive federal funds, the ADA extends the rights of equal treatment for people with disabilities to the private area, to all places of public accommodation, employers, and entities that deliver government services. The core sections of the law are found in the first three titles: employment, state and local government activities, and public accommodation. Title II of the ADA requires that state and local governments give people with disabilities an equal opportunity to benefit from all of their programs, services, and activities, such as: public education, employment, transportation, recreation, health care, social services, courts, voting, and town meetings. Section 202, Title II indicates that:

"…no qualified individual with a disability shall, by reason of such disability, be excluded from participation in or be denied the benefits of the services, programs, or activities of a public entity, or be subjected to discrimination by such entity." (ADA, 1990)

Title II recognizes the special importance of communication, which includes access to information in its implementing regulation at 28 CFR Section 35.160(a). The regulation requires that a public entity must take appropriate steps to ensure that communications with persons with disabilities are as effective as communications with persons without disabilities. The ADA mandates for "effective communication, reasonable accommodations, and auxiliary aides and services" (ADA, 1990).

However, Web accessibility did not become prominent until 1996 when the Department of Justice (DOJ) responded to Senator Tom Harkin (D-Iowa), the author of the ADA, when he inquired on behalf of one of his constituents on Web page compatibility for the blind and other people with disabilities. In response, DOJ stated that ADA Title II and III do indeed require covered entities to provide "effective communication" regardless of the media used, and that information offered through digital media must be offered through "accessible means" as well. The Internet is an excellent source of information and, of course, people with disabilities should have access to it as effective as people without disabilities (DOJ, 1996). This response involves understanding to what extent the ADA requires Web pages to be accessible to people with disabilities. The DOJ's ruling explains how the mandate for "effective communication" in ADA should apply to Web pages and Web design.

Section 508 Rehabilitation Act, 1998

Signed into law on August 7, 1998, as part of the Workforce Investment Act, Congress revised Section 508, an amendment to the Rehabilitation Act of 1973. The core area of this amendment is to ensure that all Americans have access to information technology. The law applied specifically to all U.S. government agencies, but it also affects anyone who works with the U.S. government. The law requires that when developing, procuring, maintaining, or using electronic and information technology (EIT), federal departments or agencies should ensure that the EIT allows federal employees with disabilities to have access to and use of information and data that is comparable to the access to and use of information and data by other federal employees (Section 508, 1998).

Section 508 charges the Architectural and Transportation Barriers Compliance Board (Access Board) with the task of producing accessibility standards for all electronic and information technologies used, produced, purchased, or maintained by the federal government. On December 21, 2000, the Access Board finalized and issued the standards. With its publication, the federal government for the first time set specific access standards for information presented over the Web. According to the law, federal agencies were permitted a six-month period after the standards were issued to demonstrate that their information technology systems met the standards. Federal agencies became subject to civil rights litigation for noncompliance under Section 508 after June 21, 2001. These standards have been folded into the federal government's procurement regulations, namely the Federal Acquisition Regulations (FAR) to which most agencies are subject. Agencies not covered by the FAR should incorporate the Access Board's Section 508 Standards into their own procurement regulations. The standards define means of disseminating information, including computers, software, and electronic office equipment. They provide criteria that make these products accessible to people with disabilities, including those with vision, hearing, and mobility impairments. The scope of Section 508 and the Access Board's standards are limited to the federal sector. It does not explicitly apply to the private sector or to state and local governments.

The statements issued by the U.S. Department of Education and the Access Board support a concept of state government obligations wherein they abide by Section 508 by virtue of their linkage with funding supplied to state governments through the Assistive Technology Act (29 USC 3001). There is a statement in the U.S. Department of Education's enforcing regulations to the Assistive Technology Act at 34 CFR Part 345, requiring "compliance with Section 508 of the Rehabilitation Act of 1973" (34 CFR 345.31).

Web Content Accessibility Guidelines (WCAG)

Founded in 1994, the World Wide Web Consortium (W3C) was originally organized by the European Center for Nuclear Research (CERN) and by the Massachusetts Institute of Technology (MIT). The W3C launched the Web Accessibility Initiative (WAI) in April 1997 specifically to address the question of how to expand access to the Web for people with disabilities. In February 2000, the WAI published the Web Content Accessibility Guidelines 1.0 (WCAG) and Authoring Tool Accessibility Guidelines 1.0 (ATAG), and the User Agent Accessibility Guideline 1.0 (UAAG).

The WCAG establishes three priority checkpoints for Web content developers to meet. The guidelines include 64 checkpoints arranged under 14 separate guidelines within the three priorities.

- *Priority I.* Developers MUST satisfy Priority I checkpoints in order to attain a minimum degree of accessibility. Otherwise, one or more groups will find it impossible to access some part of the information.
- *Priority II.* Developers SHOULD satisfy Priority II checkpoints for a higher degree of accessibility.
- *Priority III.* Developers MAY address Priority III checkpoints for maximum accessibility and usability (WAI, 2000).

Topics of the guidelines include images, programming scripts, navigation, multimedia, forms, frames, and tables. The guidelines are detailed and prioritized with an associated checklist *(*www.w3.org/TR/WAI-WEBCONTENT/full-checklist.html*)*. WCAG 1.0 has become the basis for accessibility standards adopted by the international community.

Laws, Regulations, and Policies in Other Countries

The World Wide Web has no boundaries, therefore, "nations everywhere are developing policies and standards to enhance the accessibility of electronics and information technology" (Paciello, 2000). These nations include European countries (Belgium, France, Greece, Norway, Portugal, Sweden, and United Kingdom), Canada, Australia, and Japan. As a founding contributor, the European Community provides financial support to the WAI. Among these countries, Australia and Canada are two counties that have enacted legislation. Prominent Australian legal standards include the Disability Discrimination Act (DDA) of 1992 and the Anti-Discrimination Act of New South Wales of 1977. Australia's Human Rights and Equal Opportunity Commission, a government agency, is responsible for creating and establishing legal decrees mandating equality of access for the disabled. In Canada, the Equity and Diversity Directorate of the Public Service Commission of Canada (PSC) "was the first government institution in any country to create a series of Web accessibility guidelines used to evaluate Web pages" (Paciello, 2000). The Common Look and Feel Working Group, created in 1998 by the Treasury Board Secretariat's Internet Advisory Committee, is charged to establish standards and guidelines as a base for official government policy.

ISSUES RELATED TO WEB ACCESSIBILITY

The current problems encountered in the implementation of accessible Web design do not only stem from design issues, nor are they because of a lack of laws, regulations, and government-related standards, although there are unanswered questions in that laws create difficulty in the implementation process. The current problems we are facing are the absence of an obligation to fulfill legal requirements and by not being aware that ensuring resource accessibility is a legal mandate. Many Web developers have not focused on the extent to which their Web pages are accessible to people with disabilities. This situation reveals that institutions are only just beginning to apply accessible Web design techniques and are not aware of the urgency behind designing accessible Web pages.

A primary issue that constitutes a barrier for implementation is that the current legal framework for electronic and information technology accessibility is actually a patchwork of laws covering certain categories of technology in some settings, other categories in other settings, but nowhere reflecting an overview or comprehensive assessment of either the issues or the solutions (NCD, 2001). The absence of obligations in fulfilling legal requirements has further complicated the implementation process. In many cases, the absence of obligations is due to unfamiliarity with the legal responsibility for creating accessible Web sites. Compounding these problems have been the conflicting views between minimum compliance and maximum accessibility. There is a tendency that most Web designers are either unaware of the laws and regulations, or ignore the accessibility requirements, or only pay attention to them as part of legal requirements.

Part of problem created by today's networked environment is that stand-alone workstations utilizing assistive technology software solutions are no longer sufficient because in the Web environment, the linkage between an individual and the Internet community as a whole is addressed, and access to the Web cannot be handled on a case-by-case basis using workstations with assistive technology. When approaching the issue of accessibility, John Mueller, the author of *Accessibility for Everyone: Understanding the Section 508 Accessibility Requirements*, points out that the issue should be approached from the viewpoint of universal design: "The application you create should work equally well with any hardware the user might have attached to the machine" (2003). Assistive technology alone cannot overcome the barriers that are created at a more basic level, the format in which content is presented (Schmetzke, 2001). For example, when electronic forms are used, the form shall allow people with disabilities to fill out the required information and to complete the online transaction. If the instruction for where to enter the appropriate information is not given, even the smartest screen reader would not know what to do with the input box, therefore the user with visual impairment or blindness would not know what to do either. Access barriers created by inaccessible design cannot be overcome even with the most sophisticated assistive technology. Further, assistive technology may also give rise to issues of incompatibility. Therefore, the Web environment mandates that potential barriers in Web design going beyond the computer workstation be examined.

IMPLEMENTATION

The benefits resulting from the implementation of universal Web design principles, which allow pages to be accessible to the largest number of the population, have been emerging. Functionalities found in assistive technology can streamline our digital architecture, and the very functionality required in Web design by people with disabilities can meet dynamic requirements for Web-based transactions. A universal Web design will greatly reduce the cost for assistive technology geared specifically to individual computer workstations, and allow universal access for users from anywhere at anytime. In a broad sense, a universal Web design allows the disability community to benefit as a whole rather than achieving accessibility through a segregated, compartmentalized, and ad hoc approach.

The overwhelming benefits of creating universally accessible Web pages can hardly be disputed. However, accessibility barriers are usually systematic. There are practical and legal reasons for addressing accessibility issues in our education, policy, and design. As the rapid development of new Web applications continues, it is necessary to ensure that barriers themselves will not continue to expand. We have seen tremendous efforts underway in the implementation process.

Education to raise awareness of Web accessibility issues, understanding the laws related to accessibility, and the consequences of major disparities in our society are a first step toward solutions. It is crucial that people with disabilities be involved in the procurement and/or development of accessibility solutions. Thus, issues of compatibility and accessibility can be anticipated and addressed during the strategic planning stage. Identifying problems and implementing repairs are two integral elements of Web site accessibility assessment. One way to ensure the assessment can be carried out is to recommend its inclusion into the institution's Web planning and design process. Another is for the institution to develop its own checklists or follow the most up-to-date guidelines published by the W3C. The assessment should indicate the areas of violations and content that are inaccessible. Based on the assessment, a plan should be made to integrate the findings and recommendations for change. Developing tools to facilitate accessible Web design is a key to the ultimate success of an accessible Web site.

Principles of universal design should be achieved, and the view of depending heavily on the tool of the assistive devices to mediate and decode should be utilized as a last result. Conducting usability studies with the understanding that the accessibility is an integral part of usability will assist in promoting inclusion.

FUTURE TRENDS

The ultimate goal for Web sites is universal design, or pervasive accessibility-interfaces that adapt to the user, regardless of ability. It is so often that with each release of new technology, there will be fixes to accessibility problems. "Subsequently, that technology that increases access to people with disabilities is termed assistive and adaptive as opposed to being a technology that is in inherently accessible" (Paciello, 2000). As we observed, while much of today's technology is designed lacking accessibility, recent awareness of laws and guidelines, and advances in promoting accessibility give rise to hope for achieving the ultimate goal for accessibility. The benefits resulting from the implementation of universal Web design principles, which allow pages to be accessible to the largest number of segments of the population, have been emerging.

Discussions about the "digital divide" problem, needs for accessible Web design, and practical tips for designing barrier-free Web sites found in literature in recent years have demonstrated awareness-raising efforts (Casey, 1999; Rouse, 1999; Jobe, 2000; Valenza, 2000).

Awareness of the laws has recently increased the development of computers utilizing voice recognition and speech output systems. With technology advancement, personal computers with neural interfaces using brain and sensory technology that enable an individual to interact with a computer will allow people who have otherwise no use of their hands to interact with computers by sensing head and brain signals and translating those to computer input actions. The combination of Web and television services provides a promising future for the disabled community, particularly potential solutions for educaion.

CONCLUSION

The concept of accessible design or universal design is increasingly becoming an important component of Web design. Today, public and private entities are subject to the requirements of the ADA, Section 504, and Section 508, whether or not they receive federal funds. According to the definition of the law, denial of access resulting from the inaccessibility of mainstream information technology is considered to be discrimination. Barriers mostly resulting from the absence of obligations to fulfill legal requirements should be eliminated as the awareness of the importance of accessibility increases. Functionalities found in assistive technology can streamline our digital architecture, and the very functionality required in Web design by people with disabilities can meet dynamic requirements for Web-based transactions. A universal Web design will greatly reduce the cost for assistive technology geared specifically to individual computer workstations, and allow universal access for users from anywhere at anytime. The benefits and value for overcoming these barriers for the community of people with disabilities cannot be disputed, and maximum accessibility is a goal for us to achieve.

REFERENCES

Americans with Disabilities Act of 1990 (ADA). (1991). Pub.L. No. 101-336, §2,104 Stat. 328.

Casey, C.A. (1999). Accessibility in the virtual library: Creating equal opportunity Web sites. *Information Technology and Libraries*, *18*(1), 22-25.

Department of Justice (DOJ). (2000). *Information technology and people with disabilities: The current state of federal accessibility*. Washington, DC: DOJ. Retrieved June 6, 2001, from *www.usdoj.gov/crt/508/report/content.htm*

Hricko, M. (2003). *Design and implementation of Web-enabled teaching tools*. Hershey, PA: Information Science Publishing.

Jobe, M.M. (1999). Guidelines on Web accessibility for the disabled. *Colorado Libraries*, *25*(3).

Mueller, J. (2003). *Accessibility for everyone: Understanding the Section 508 accessibility requirements*. Berkeley, CA: Apress.

National Council on Disability (NCD). (2001). *The accessible future*. Washington, DC: NCD. Retrieved July 10, 2001, from *www.ncd.gov/newsroom/publications/accessiblefuture.html*

Paciello, M.G. (2000). *Web accessibility for people with disabilities*. Lawrence, KS: CMP Books.

Rouse, V. (1999). Making the Web accessible. *Computers in Libraries*, *19*(6).

Schmetzke, A. (2001, March 15-18). Distance education, Web-resources design, and compliance with the Americans with Disabilities Act. *Proceedings of the 10th National Conference of College and Research Libraries* (pp. 137-142), Denver, Colorado, USA. Chicago: Association of College and Research Libraries. Retrieved October 20, 2003, from *www.ala.org/Content/NavigationMenu/ACRL/Events_and_Conferences/schmetzke.pdf*

Section 504 of 1973. (1994). Pub.L. No. 93-112, § 504, 87 Stat. 355, 394 (codified as amended at 29 U.S.C. § 794.

Section 508 of 1998. Pub. L. No. 105-220, §112, Stat. 936 (codified at 29 U.S. C. § 798).

Slatin, J.M. & Rush, S. (2003). *Maximum accessibility: Making your Web site more usable for everyone*. New York: Addison-Wesley.

State Grants Program for Technology-Related Assistance for Individuals with Disabilities. (2002). 34 C.F.R. § 345.

Valenza, J.K. (2000). Surfing blind. Library journal fall supplement. *Net Connect*, *125*(14), 34+.

Web Accessibility Initiative (WAI). (2000). Web content accessibility guidelines (WCAG). Retrieved October 20, 2003, from *www.w3.org/TR/WAI-WEBCONTENT/*

KEY TERMS

ALT-Text: Stands for Alternative Text, primarily used to render graphics when the image is not being displayed.

Americans with Disabilities Act (ADA): U.S. public law enacted in 1990 ensuring rights for people with disabilities. This legislation mandates reasonable accommodation and effective communication.

Bobby: A Web page accessibility validation service provided by the Center for Applied Special Technologies (CAST); uses the W3C's Web Content Accessibility Guidelines, and evaluates Web pages for possible errors and ranks them in order of priority.

JAWS: A Windows screen reader from Henter-Joyce; stands for Job Access for Windows. JAWS works with PCs to provide access to software applications and the Internet. With its internal software speech synthesizer and the computer's sound card, information from the screen is read aloud for visually impaired and blind people.

Section 508: Section 508 of the Rehabilitation Act is U.S. legislation that establishes requirements for electronic and information technology developed, maintained, procured, or used by the federal government.

W3C: The World Wide Web Consortium (W3C) develops interoperable technologies such as: specifications, guidelines, software, and tools to lead the Web to its full potential. W3C is a forum for information, commerce, communication, and collective understanding.

Web Accessibility Initiative (WAI): Established by W3C and the Yuri Rubinsky Insight Foundation in 1997, the WAI works with organizations around the world to promote Web accessibility in five key areas: technology, guidelines, tools, education and outreach, and research and development.

Web Content Accessibility Guidelines (WCAG 1.0): The official set of guidelines published by the W3C to assist Web content developers in the creation of accessible Web sites. The guidelines established three priority levels with 64 checkpoints for Web developers to meet.

Web Caching

Antonios Danalis
University of Delaware, USA

INTRODUCTION

The popularity of the World Wide Web has led to an exponential increase of the traffic generated by its users for over a decade. Such a growth, over such a long period of time, would have saturated both the content providers and the network links had Web caching not been efficiently deployed. Web caching can improve the overall performance of the World Wide Web in several ways, depending on the decisions made regarding the deployment of the corresponding caches. By placing caches in strategic positions, the core network traffic can be reduced, the load of a content provider can be scaled down, and the quality of service, as the users perceive it, can be improved. In this article we present an overview of the major design and implementation challenges in Web caching, as well as their solutions.

BACKGROUND

A Web cache can be placed at different positions on the Web and yield different benefits. Namely, it can be placed next to a content provider, in strategic positions inside the network, at the boundaries of local area networks, or inside the host that runs the requesting browser.

Most users tend to access particular pages quite often. An example could be the "home page" that has been set on their browser, or Web sites that contain documentation they are interested in. Furthermore, during browsing, many users visit the same pages multiple times in short periods of time, by using the history of their browsers. To exploit these access patterns, most browsers keep local copies of the frequently or recently accessed pages. This way, the user is served in near zero time, and the network traffic is reduced. Although this technique can have a positive effect on the browsing experience of a single user, it has a minimal effect on the network traffic. This is mostly due to the small disk space used by this type of caching and the lack of sharing between different user caches.

To yield significant results, dedicated computer systems, called proxies, are installed at the edges of local or wide area networks. These systems can achieve significant reduction in the network traffic and improvement in the user perceived quality of service, by filtering and serving the Web requests generated inside the entire network they serve. If a user has defined in the browser's settings a particular proxy to be used, every time he or she requests a Web page the browser will send this request to the proxy. If the proxy happens to have the page, the user will be served promptly without the original content provider being contacted. If the proxy cannot serve the request, it will fetch the appropriate Web objects (such as the text documents, images, applets) from the original server, and possibly keep a copy for later use. Transparent proxies are a special kind of proxy that end users do not explicitly specify in their browser's settings. Rather, the gateway of the network identifies Web requests and forwards them to the proxy. This model is more efficient, and easier to administrate, since all the users of the local network will be using the proxy, and they need not know about possible changes concerning the proxy.

To improve the performance of a content provider, a cache can be deployed on the server side. These caches, also called server accelerators, or reverse proxies, are usually deployed in front of clusters of Web servers (Challenger, Iyengar & Dantzig, 1999) and cache the most popular documents in their main memory. Since a few highly popular documents are in most cases responsible for a large percentage of requests, by using a server-side cache, a Web server can decrease its disk I/O load (Abrams et al., 1995; Markatos, 1996; Tatarinov, Rousskov & Soloviev, 1997) and significantly improve the request serving quality of service. Nevertheless, a reverse proxy benefits only a particular content provider and does not reduce the network traffic.

In addition to client and server-side caching, there exists the approach of adaptive caching (Michel et al., 1998). In this scheme caches can exist at different points inside the network, and be configured to cooperate with one another, in order to improve the overall performance and balance the load. In this model, different caches, potentially belonging to different organizations, are organized into dynamic groups, which each one can join or leave depending on content demand. The communication needed for a group to form and for documents to be located is done through the Cache Group Management Protocol (CGMP) and the Content Routing Protocol (CRP), respectively. Adaptive caching can deal very well with request surges, occurring when some documents become highly popular in a short period of time (usually articles in online newspapers), but assumes that cooperation across administrative boundaries is not an issue.

CRITICAL ISSUES OF WEB CACHING

To improve their effectiveness, caches can be combined in groups to serve requests cooperatively. Such groups may be organized in meshes, as in the case of adaptive caching, or in hierarchical formations, as in the case of the Harvest Cache project (Chankhunthod et al., 1996). In hierarchical designs, caches are usually organized in tree-like structures, and are configured such that child nodes can query their parents and their siblings, but never their children. Although grouping several caches can improve the scalability of the caching system, control messages between a large number of nodes could saturate the parent nodes and the network links (Baentsch, 1997). To make the communication between the caches efficient, several special purpose protocols have been introduced.

The most popular inter-cache communication protocols are ICP, CRP, CARP, cache digest, and WCCP (Melve, 1999). The Internet Cache Protocol (ICP) was developed for the communication of the caches in the Harvest project and was refined within Squid (SQUID). It is a lightweight message format, used by the caches for issuing queries and replies, to exchange information among one another regarding the optimal location from where a certain object can be retrieved. The large number of messages exchanged in ICP, when the number of caches is high, has been shown to impede scalability (Baentsch, 1997; Fan et al., 2000). To avoid similar effects, the Content Routing Protocol (CRP) uses multicast to query cache meshes. CRP exploits the overlapping that exists in cache group formations to propagate queries or popular objects between the groups. To further improve on the performance of the inter-cache communication protocol, cache digests can be used (Fan et al., 2000; Rousskov & Wessels, 1998). Cache digests compact (using a lossy algorithm) the information about the contents of a cache, and make it available to its neighbors. By checking the digests of its neighbors, a cache can identify, with some uncertainty, which neighbors are likely to have a given document. In addition to these protocols, there exist two proprietary ones, CARP and WCCP, designed by Microsoft and Cisco respectively. Unlike the protocols mentioned before, the Cache Array Routing Protocol (CARP) implements routing in a deterministic way. In particular, it uses hashing to identify which member of the proxy array has the requested document. This way it avoids the scalability problems that appear in ICP due to the large number of control messages. The Web Cache Communication Protocol (WCCP) is designed to support transparent caching. The idea is that a router (such as Cisco Cache Engine) that can recognize Web traffic will intercept user requests and redirect them to a cache.

The protocol used for the communication between the Web clients, proxies, and servers is the Hyper-Text Transfer Protocol (HTTP). HTTP runs on top of TCP/IP, which is the most common protocol used in the Internet for reliable transfers, flow control and congestion avoidance. To initialize a connection, TCP sends a special packet (SYN) from the client to the server, and the server responds with an acknowledgment. After initializing the connection, in order to request a page, the client has to send an additional message to the server, and wait for the reply. In the first version of HTTP, this four-step procedure had to take place for all individual objects of a page, such as images, or applets, separately. To reduce the number of round-trips needed, persistent connections, that is, connections that remain open after the retrieval of a file, were proposed (Caceres et al., 1998; Heidemann, Obraczka & Touch, 1997; Mogul, 1995). Persistent connections allow a client to fetch all the components of a Web page by facing the connection startup latency only once. In addition to the multiple steps handshake, slow-start is another feature of TCP that has negative effects in the context of Web transfers. As the name implies, slow starts demand that newly created TCP connections transmit data very slowly and they increase the throughput as they transmit more and more. In the general case, where flow control and congestion avoidance is necessary, this technique is very efficient. In the case of Web transfers though, where the majority is short-lived since most documents are a few kilobytes long (Arlitt & Williamson, 1996; Cunha, Bestavros & Crovella, 1995; Shriver et al., 2001), slow start translates to slow transfer. In order to deal with these issues, HTTP/1.1 (Fielding et al., 1997) was introduced. This new version, among other improvements, supports persistent connections and has been shown to dramatically outperform its predecessor (Nielsen et al., 1997).

In heavily loaded Web caching systems, disk I/O can become a significant bottleneck (Markatos et al., 1999; Mogul, 1999; Rousskov & Soloviev, 1998), because disks are significantly slower than the main memory, and traditional file systems are not optimized to handle Web traffic. To reduce this overhead, Gabber and Shriver (2000) proposed the use of a special purpose file system and Markatos et al. (1999) and Maltzahn, Richardson and Grunwald (1999) proposed the use of special purpose storing policies that reduce the overhead of file creation, deletion, and access. Such policies take into account Web access patterns when they make decisions regarding file placement. For example, in most cases, the HTML text of a Web page and the embedded objects (such as images) of the page will be requested one after the other. By placing these objects in files located close to each other on the disk (Maltzahn, Richardson & Grunwald, 1999), or in neighboring locations within the same file (Markatos et al., 1999), both reading and writing will access almost sequential disk blocks, yielding a significant performance improvement. Additionally, a specialized file system (Gabber & Shriver, 2000) could avoid keeping metadata to improve performance, or could use in-memory data structures to quickly identify cached objects without performing any disk I/O (Tomlinson, Major & Lee, 1999).

Regardless of the method used to access the disk, and no matter how large the storage space is, it is not possible to store the whole Internet. Therefore, there must exist some techniques to decide which documents to keep in the cache and which to replace when the cache fills up. Several algorithms have been proposed, each trying to keep the most valuable documents in the cache. To assign a value to a document, most algorithms take into account one or more parameters such as recency of access, frequency of access, document size, and fetching latency, and try to evict the documents with the lowest value. The effectiveness of a replacement policy is measured by Hit Rate (HR) and Byte Hit Rate (BHR). HR is the percentage of requests that are served from the proxy, and BHR is the percentage of bytes served from the cache. The Least Recently Used (*LRU*) algorithm is a replacement policy already known and used in virtual memory paging systems. Based on the fact that the possibility of a file to be requested again drops dramatically as the time since the previous request increases (Cao & Irano, 1997; Rizzo & Vicisano, 2000), it tries to replace the least recently used document, when space is needed for a new one. The *SIZE* algorithm (Rizzo & Vicisano, 2000) considers the large files to be the least valuable, and thus removes the largest document to make room for several smaller. Due to the high popularity of small files (Shriver et al., 2001), *SIZE* is among the best replacement algorithms. Capitalizing on frequency of use, the Least Frequently Used (LFU) algorithm evicts the documents that have been used the least amount of times. Although this algorithm can perform similarly to LRU for many access patterns, its main disadvantage is that it does not evict documents that were popular once, but not any more (such as old newspaper articles). To deal with this issue, Arlitt et al. (1999) presented LFU-DA, which is LFU with Dynamic Aging. Their variation takes into account the age of a document and thus manages to perform better than most existing algorithms in terms of both HR and BHR. To achieve even higher efficiency, researchers have introduced algorithms that combine the different dimensions. Lowest Relative Value (LRV) (Rizzo & Vicisano, 2000), for example, assigns a value to each document, calculated based on its age, popularity and size, and replaces the one with the lowest value. Hybrid (Wooster & Abrams, 1997) takes into account even more parameters, such as bandwidth to the content provider, and time to connect, for calculating the value of a document, and has been shown to outperform most traditional algorithms in terms of both HR and latency reduction. Finally, Cao and Irani (1997) proposed GreedyDual-Size, which consists of a family of replacement policies, all working under the same basic idea, but optimized for different goals. Therefore, depending on the goals a cache wants to achieve, GreedyDual-Size can be configured to maximize the HR, or minimize the average download latency, or reduce the network traffic.

Besides evicting documents when space needs to be made for new ones, a proxy needs to delete documents that are not valid any more, to preserve the consistency of its cache. The most common cache consistency mechanisms are time-to-live (TTL) fields, client polling, and invalidation protocols (Gwertzman & Seltzer, 1996). TTL is actually an estimation of how long the current copy of a document will be valid for. Although this approach can be very useful in cases where this information is known, as in the case of periodically changing news articles, it can lead to waste of bandwidth if the estimation is too conservative, or cache inconsistency if the value of TTL is too large. Adaptive TTL tries to find a better estimation by taking into account the document's age (Gwertzman & Seltzer, 1996). In client polling, every time a document is requested, the cache sends an if-modified-since (IMS) request to the original server. According to the result, the file is either proven valid and served by the cache, or the new version is fetched. This approach generates a large number of control messages, since an IMS is sent to the server for every document requested. To gain the benefits of both techniques, squid (SQUID) uses an adaptive TTL policy to decide when to send an IMS validation request to the server. Invalidation callbacks are an approach where the original server keeps track of the proxies that have a copy of its documents, and contacts them when the documents change. Callbacks provide strong consistency without saturating the network with control messages (Cao & Liu, 1998), but raise privacy and security concerns.

Dynamic pages, either created according to the preferences of each individual user, or updated dynamically to meet real-time changing data, such as the stock market or sports scores, raise important issues in regard to cache consistency. If a dynamic document is cached and served by a traditional proxy, it is very likely to be invalid even before the very next request. On the other hand, if a proxy decides not to cache dynamic objects, it will lose an important percentage of Web traffic (Caceres et al., 1998; Wolman et al., 1999). Markatos (2001) suggests caching of search engine results for hours or days, based on the fact that query results are already several days old.

Several approaches deal with dynamic pages by caching static, sub-document objects (Douglis, Haro & Rabinovich, 1997; Meira et al., 1999; Shi et al., 2003). Furthermore, Mohapatra and Chen (2002) and Challenger et al. (1999) construct a graph of the objects that constitute a Web page and upon a change, graph traversal reveals the objects that need to be updated for the page to be reconstructed. The transfer of "deltas", that is, the differences between the cached and the original document, is another technique shown to work well (Mogul et al., 1997; Savant & Suel, 2003). Finally Cao, Zhang and Beach (1998) suggested an approach where proxies can fetch from the server a special purpose program (a CacheApplet) attached to the data, to deal with reconstructing or refetching the page the way the original content provider wishes.

CONCLUSION AND FUTURE TRENDS

Web caching has been extensively deployed during the last years to support the needs of the exponentially growing World Wide Web. Depending on their placement, Web caches can accelerate the performance of a content provider, improve content availability, reduce network traffic and download latency, and improve the overall user experience of the Web. In this article we have presented the most important design and implementation issues concerning Web caches, as well as their solutions. Some issues, such as dynamic document handling, remain open and new ideas are still being proposed. In particular, the issues of caching dynamic documents (Rhea, Liang & Brewer, 2003; Yuan, Chen & Zhang, 2003; Yuan, Hua & Zhang, 2003) and ensuring cache consistency (Mikhailov & Wills, 2003; Pandey, Ramamritham & Chakrabarti, 2003) in the presence of changing, non-static Web pages, are expected to attract considerable scientific interest in the near future. Regardless of the final decisions to prevail though, it is most likely that in the following years, caching will keep being a key solution for the performance of the World Wide Web.

REFERENCES

Abrams, M., Standbridge, C.R., Abdula, Williams, S., & Fox, E.A. (1995, December). *Caching proxies: Limitations and potentials.* WWW-4, Boston Conference.

Arlitt, M.F., & Williamson, C.L. (1996, May). Web server workload characterization: The search for invariants. *Proceedings of the ACM SIGMETRICS* (pp. 126-137).

Arlitt, M.F., Cherkasova, L., Dilley, J., Friedrich, R., & Jin, T. (1999, May). Evaluating content management techniques for Web proxy caches. *Proceedings of the Second Workshop on Internet Server Performance (WISP '99).*

Baentsch, M., Baum, L., Molter, G., Rothkugel, S., & Sturm, P. (1997, June). World-Wide Web caching – the application level view of the Internet. *IEEE Communications Magazine, 35*(6).

Caceres, R., Douglis, F., Feldmann, A., Glass, G., & Rabinovich, M. (1998, June). Web proxy caching: The devil is in the details. *ACM SIGMETRICS Workshop on Internet Server Performance.*

Cao, P., & Irani, S. (1997, December). Cost-aware WWW proxy caching algorithms. *Proceedings of the 1997 USENIX Symposium on Internet Technologies and Systems (USITS99)* (pp. 193-206).

Cao, P., & Liu, C. (1998). Maintaining strong cache consistency in the World Wide Web. *Proceedings of 3rd International Conference on Web Caching.*

Cao, P., Zhang, J., & Beach, K. (1998, September). Active cache: Caching dynamic contents (objects) on the Web. *Proceedings of the IFIP International Conference on Distributed Systems Platforms and Open Distributed Processing (Middleware '98).*

Challenger, J., Iyengar, A., & Dantzig, P. (1999, March). A scalable system for consistently caching dynamic Web data. *Proceedings of the IEEE Infocom '99 Conference.*

Chankhunthod, A., Danzig, P., Neerdaels, C., Schwartz, M., & Worrell, K. (1996). A hierarchical Internet object cache. *Proceedings of the USENIX Technical Conference.*

Cunha, C., Bestavros, A., & Crovella, M. (1995, April). *Characteristics of WWW client-based traces.* Technical Report 95-010. Boston University.

Douglis, F., Haro, A., & Rabinovich, M. (1997, December). HPP: HTML macropre -processing to support dynamic document caching. *Proceedings of the 1st USENIX Symposium on Internet Technologies and Systems (USITS97).*

Fan, L., Cao, P., Almeida, J., & Broder, A. (2000). Summary cache: A scalable wide-area Web cache sharing protocol. *IEEE/ACM Transactions on Networking, 8*(3), 281-293.

Fielding, R., Gettys, J., Mogul, J., Frystyk, H., & Berners-Lee, T. (1997, January). *Hypertext transfer protocol – HTTP/1.1. RFC2068.*

Gabber, E., & Shriver, E. (2000, September.). Let's put NetApp and CacheFlow out of business. *Proceedings of the SIGOPS European Workshop.*

Gwertzman, J., & Seltzer, M. (1996, January.). World-Wide Web cache consistency. *Proceedings of 1996 USENIX Technical Conference* (pp. 141-151).

Heidemann, J., Obraczka, K., & Touch, J. (1997). Modeling the performance of HTTP over several transport protocols. *IEEE/ACM Transactions on Networking, 5*(5), 616-630.

Maltzahn, C., Richardson, K.J., & Grunwald, D. (1999, June). Reducing the disk I/O of Web proxy server caches. *Proceedings of the 1999 USENIX Annual Technical Conference.*

Markatos, E.P. (1996, May). Main memory caching of Web documents. *Fifth International WWW Conference*.

Markatos, E.P. (2001, February). On caching search engine query results. *Computer Communications, 24*(2).

Markatos, E.P, Katevenis, M.G., Pnevmatikatos, D., & Flouris, M. (1999). Secondary storage management for Web proxies. *Proceedings of the 2nd USENIX Symposium on Internet Technologies and Systems (USITS99).*

Meira, W., Cesario, M., Fonseca, R., & Ziv, N. (1999). Integrating WWW caches and search engines. *Proceedings of the IEEE 1999 Global Telecommunications Internet Mini-Conference.*

Melve, I. (1999). *Inter-cache communication protocols.* IETF WREC Working group draft.

Michel, S., Nguyen, K., Rosenstein, A., Zhang, L., Floyd, S., & Jacobson, V. (1998). Adaptive Web caching: Towards a new global caching architecture. *Computer Networks & ISDN Systems, 30*(22-23), 2169-2177.

Mikhailov, M., & Wills, C. (2003, May). Evaluating a new approach to strong Web cache consistency with snapshots of collected content. *12th Int'l World Wide Web Conference (WWW 2003),* Hungary.

Mogul, J.C. (1995, May). *The case for persistent-connection HTTP.* Western Research Laboratory. 95.4-Research Report 95/4.

Mogul, J.C. (1999). Speedier squid: A case study of an Internet server performance problem. *Login: The USENIX Association Magazine, 24*(1), 50-58.

Mogul, J.C., Douglis, F., Feldmann, A., & Krishnamurthy, B. (1997). Potential benefits of delta-encoding and data compression for HTTP. *Proceedings of the ACM SIGCOMM 97.*

Mohapatra, P., & Chen, H. (2002) WebGraph: A framework for managing and improving performance of dynamic Web content. *Special Issue of Proxy Servers in the IEEE Journal of Selected Areas in Communications.*

Nielsen, H.F., Gettys, J., Baird-Smith, A., Prud'hommeaux, E., Lie, H.W., & Lilley, C. (1997, September). Network performance effects of HTTP/1.1, CSS1 and PNG. *Proceedings of ACM SIGCOMM '97.*

Pandey, S., Ramamritham, K., & Chakrabarti, S. (2003, May). Monitoring the dynamic Web to respond to continuous queries. *12th Int'l World Wide Web Conference (WWW 2003),* Hungary.

Rhea, S., Liang, K., & Brewer, E. (200, May). Value-based Web caching. *12th Int'l World Wide Web Conference (WWW 2003),* Hungary.

Rizzo, L., & Vicisano, L. (2000). Replacement policies for a proxy cache. *IEEE/ACM Transactions on Networking, 8*(2), 158-170.

Rousskov, A., & Soloviev, V. (1998, June). On performance of caching proxies. *Proceedings of the 1998 ACM SIGMETRICS Conference.*

Rousskov, A., & Wessels, D. (1998). Cache digests. *Computer Networks & ISDN Systems, 30*(22-23), 2155-2168.

Savant, A., & Suel, T. (2003, September). Server-friendly delta compression for efficient Web access. *Proceedings of the 8th International Workshop on Web Content Caching and Distribution.*

Shi, W., Collins, E., & Karamcheti, V. (2003, September). Modeling object characteristics of dynamic Web content. *Journal of Parallel and Distributed Computing (JPDC), special issue on Scalable Internet Services and Architecture.*

Shriver, E., Gabber, E., Huang, L., & Stein, C. (2001, June). Storage management for Web proxies. *Proceedings of the 2001 USENIX Annual Technical Conference.*

SQUID proxy. *http://www.squid-cache.org*

Tatarinov, I., Rousskov, A., & Soloviev, V. (1997, September.). Static caching in Web servers. *Proceedings of the 6th IEEE Conference on Computer Communications and Networks.*

Tomlinson, G., Major, D., & Lee, R. (1999). High-capacity Internet middleware: Internet caching system architectural overview. *Second Workshop on Internet Server Performance.*

Wolman, A., Voelker, G., Sharma, N., Cardwell, N., Brown, M., Landray, T., Pinnel, D., Karlin, A., & Levy, H. (1999). Organization-based analysis of Web-object sharing and caching. *Proceedings of the 2nd USENIX Symposium on Internet Technologies and Systems (USITS99).*

Wooster, R., & Abrams, M. (1997, April). Proxy caching that estimates page load delays. *6th International World Wide Web Conference.*

Yuan, C., Chen, Y., & Zhang, Z. (2003, May). Evaluation of edge caching/offloading for dynamic content delivery. *12th Int'l World Wide Web Conference (WWW 2003)*, Hungary.

Yuan, C., Hua, Z., & Zhang, Z. (2003, September). Proxy+: Simple proxy augmentation for dynamic content processing. *8th International Web Caching Workshop and Content Delivery Workshop (WCW'03).*

KEY TERMS

Byte Hit Rate: The ratio of bytes served by the cache over the total number of bytes requested by the clients. BHR can be significantly different from HR in a case where only few, but large files are being served by the cache.

Hit Rate: The ratio of requests served by the cache (hits) over the total number of requests made by the clients.

HTTP: Hyper-Text Transfer Protocol. The protocol used for most of the communication on the Web, between the clients, the proxies, and the servers.

Layer 4 Switch: A switch that can retrieve from the network packets information about the port number they are using, and thus the application that generated them.

Proxy Cache: A machine dedicated to serving client requests for Web documents. Usually it is installed at the edge of local or wide area networks to provide fast responses to the users and reduce the traffic exiting the network.

Server Accelerator: A machine dedicated to improving the performance of a Web server by caching popular Web objects. Usually it is installed in front of a single or a farm of Web servers in order to improve their performance and load the balance among them.

Transparent Proxy: A proxy that closely cooperates with either a router or a Layer 4 switch, to intercept Web requests while invisible to the users. Other than being invisible it works as a regular proxy cache.

Web Initiatives and E-Commerce Strategy

Ron Craig
Wilfrid Laurier University, Canada

INTRODUCTION

Business use of modern computers started in the 1950s. Since then, IT has progressed through a series of hardware, software and applications improvements, delivering more value to individuals, firms, and organizations for lower cost. Productivity improvement (efficiency), at both the individual and organizational levels, has been a major focus. As well, firms have learned how to gain competitive advantage from IT.

While historical roots of the Web go back several decades, it was only in the last two that business really started to embrace the Internet, and in the last one that commercial opportunities on the Web grew rapidly. Business use has gone from simple operational efficiencies (e-mail on the Internet, replacement of private EDI networks, etc.) to effectiveness, enhanced services, and virtual products. Information and information products available in digital form, and the ability to quickly transfer these from one party to another, have led to a paradigm shift in the way many organizations operate. While the transition has followed the automate, infomate, transformate progression of historical IT, the pace has been unprecedented. There have been successes and failures, with fortunes made and lost.

Paralleling the improvement in IT and the Internet has been a series of economic shifts including globalization, flattening of hierarchical organizations, increasing emphasis on knowledge work (contrasted with manual labour), plus growth in the service sector and information economy. IT has both hastened these economic shifts, and provided a welcome means of addressing the accompanying pressures. E-commerce, which now inevitably includes Web initiatives, has most recently been at the forefront among these economic shifts.

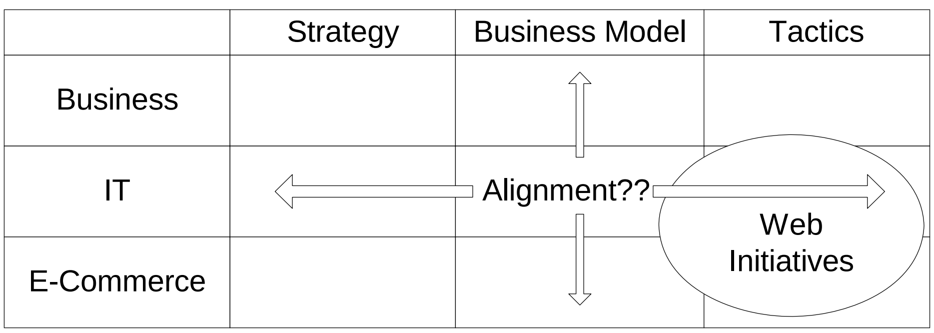
To consider Web initiatives and e-commerce (EC) strategy, one needs to first understand strategy, and then extend this to the organization's business model and tactics. A firm's general business strategy includes, but is not limited to, its IT strategy (Figure 1). Similarly, EC strategy is a subset of IT strategy. Strategy should drive actions (tactics) through an appropriate business model. When strategy (business, IT, and EC) and tactics are closely aligned, and tactics are successfully executed, desirable results are obtained.

In addition to commercial use of the Web, there are many non-commercial uses and non-commercial users (governments, educational institutions, medical organizations, etc.). The term e-business is often used to include both commercial and non-commercial activity on the Internet. In this article, the focus is on commercial activities.

BACKGROUND: BUSINESS STRATEGY, IT STRATEGY, AND WEB INITIATIVES

Business strategy and IT strategy have been extensively studied. The strategic alignment model of Henderson and Venkatraman (1993) identifies four domains of strategic

Figure 1. Strategic alignment



choice: business strategy, IT strategy, organizational infrastructure and processes, and IT infrastructure and processes. This model recognizes that a firm's IT operates within, and supports, a larger environment. As well, a firm's IT strategy can lead, lag, be independent of, or be aligned with a firm's business strategy. When alignment exists, there are significant payoffs (Tallon & Kraemer, 2003).

On the business strategy side, Porter provides several frameworks to guide firms in selecting their strategy and business model. His five-forces model, value chain network, and generic strategies (Porter, 1996) are useful frameworks when considering both business and IT strategies. In response to the question of whether or not the Internet renders established rules of strategy obsolete (as some have proposed), Porter answers that it makes strategy more vital than ever (Porter, 2001). He shows how the Internet has both positive and negative effects on industry structure, and identifies six principles of strategic positioning: (1) start with the right goal – superior long-term return on investment; (2) a firm's strategy enables it to deliver a value proposition, or set of benefits, that differentiates itself from competitors; (3) a firm's strategy is reflected in a distinctive value chain; (4) effective strategies require trade-offs; (5) strategy defines how all the elements of what a company does fit together; and (6) strategy involves continuity of direction. Porter concludes, "In our quest to see how the Internet is different, we have failed to see how the Internet is the same."

An extension to Porter's value chain is the virtual value chain (Rayport & Sviokla, 1995). Just as the physical value chain identifies the value-adding stages through which physical goods flow, the virtual value chain identifies the value-adding steps for information (gathering, organizing, selecting, synthesizing, and distributing). For virtual products and services, EC strategy and Web initiatives are especially important.

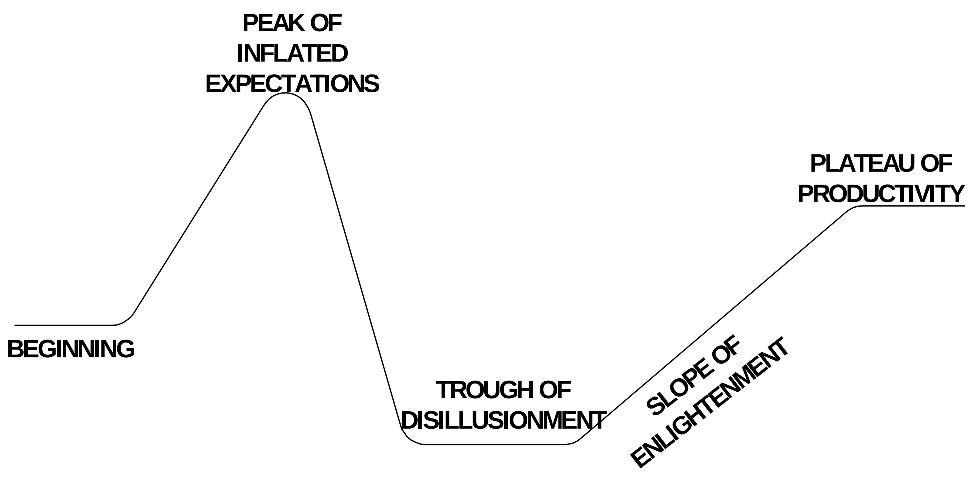
Hence, firms need to consider their IT and EC strategies as part of their business strategy. None of these should be developed or implemented independently. Decisions within any firm are made in the context of a particular business strategy, a particular set of experience and skills, a particular culture and organizational structure, and a particular set of technology and data capabilities (Hammer, 2001). Web initiatives, as a subset of EC and IT strategy, lie at the execution end (as shown in Figure 1).

E-COMMERCE STRATEGY

During the rampant optimism of the mid to late '90s, there seemed to be much more hype than reality concerning e-business. Statements were made that business was different now, that the Internet and Web changes everything, and that new e-business models were needed. The feeding frenzy among venture capitalists, eager to fund almost any start-up, allowed incomplete and ill-conceived concepts to be financed. It did not take long before reality took hold again, as the dot.com boom became the dot.com bust. The pendulum has now shifted from an overemphasis on "e" to a more balanced perspective on both "e" and "c". The Gartner Group Hype Cycle (Figure 2) provides a somewhat light-hearted, yet still realistic, view of this technology lifecycle. EC has gone through the first half of this cycle, and is now maturing.

Understanding an organization's strategic grid position (Figure 3) is critical for developing an appropriate IT

Figure 2. Gartner Group Technology Hype Cycle



and EC strategy and determining the requisite level of resources to commit. EC is not strategic to all firms, nor is all EC strategic. As Carr (2003) argues, much of IT today is a commodity-like service for many organizations, and can be managed as such. Yet, EC and Web initiatives can be strategic. There are usually significant resource implications for firms transitioning from one quadrant to another within this grid (i.e., changing their EC strategy).

An important component of EC strategy is the business model used by a firm. As shown in the previous section, strategy is about making decisions and choices. For some firms, there will be much greater emphasis on the virtual side of their business; for others it will be the opposite. However, all firms need to consider the needs of their customers and the strategies (both business and IT) of their competitors, and be able to deliver required goods/services in a sustainable manner. Hence, the firm's business model must align with the strategy selected (be it EC strategy, IT strategy, or business strategy). Magretta (2002) identifies two tests for a powerful business model – the narrative test, and the numbers test. The first test requires a logical, defendable explanation of who one's customers are, what they value, and how the firm will be profitable by providing that value. The second test requires evidence of ongoing financial viability, based on realistic assumptions and a financial analysis. Online auction giant eBay passed both these tests, as did Amazon. In contrast, most online grocery models failed because of false assumptions about what customers valued, and overly optimistic estimates of marketing, technology and delivery cost. Another retailer, Lands End, utilized Web initiatives and other IT to successfully pioneer mass customization for apparel (Ives & Piccoli, 2003).

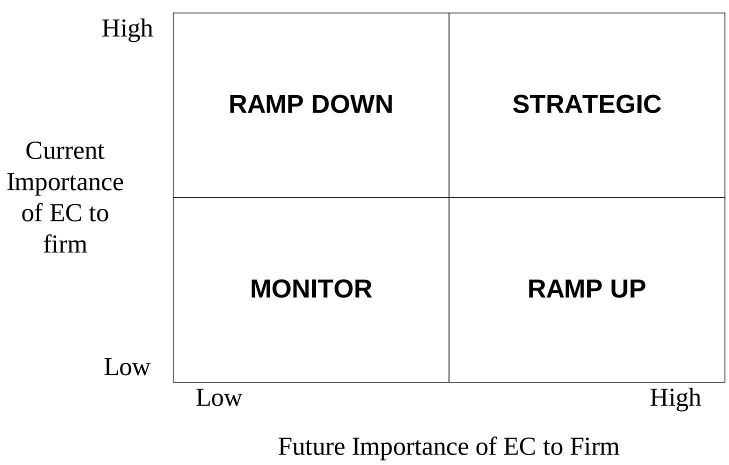
A business model is about much more than simply technology. In the earlier days of the dot.com boom, there was an overemphasis on executing Web initiatives while ignoring the rest of the business model. Now, as the EC field matures, firms have a better understanding of the opportunities, the costs/benefits, risks, and the technology and applications to use. While technology is an important component, so are organizational characteristics (culture, interorganizational relationships, leadership, reward and control systems, staffing, structure), resources (capabilities, financial, fixed assets, human, marketing, relationships, reputation and brand awareness, technology), and managerial preferences (beliefs and values, personality and needs, job experience and context, leadership style, political elements) (Crossan et al., 2002).

One question facing firms is when to use Web initiatives. Andal-Ancion et al. (2003) identify 10 drivers of new information technology (NIT), which also apply to Web initiatives. These different drivers determine the competitive advantages of deploying NIT, and fit under three general classifications: (1) Inherent characteristics of the product or service (electronic delivery, information intensity, customizability, and aggregation effects), (2) Interactions between a company and its customers (search costs, real-time interface, and contracting risk), and (3) Interactions between a company and its partners and competitors (network effects, standardization benefits, and missing competencies).

In their paper, they apply these drivers to various industries and firms, explaining the basis for disintermediation, remediation, or network-based mediation within an industry. These drivers should also be considered when a firm develops its EC strategy and considers Web initiatives.

A particular Internet benefit comes from its ability to integrate business processes across organizations, facili-

Figure 3. EC importance strategic grid



tated by the sharing of information between the various partners. This is seen in B2B exchanges (whether private or public) and portals. After a rocky start with involvement in public exchanges (which tended to overemphasize price), many firms are now participating in, or hosting, private by-invitation-only exchanges (Hoffman et al., 2002). Once again, it was lack of understanding of a sustainable B-model that resulted in so many failures. The original plan was to use the Internet's power and reach to form a more efficient marketplace. Most exchanges failed once their venture capital financing dried up. The strongest surviving exchanges are those that added value for all participants (appropriate alignment, per Figure 1). Examples of successful industry exchanges include Covisint LLC (automotive), Trade-Ranger Inc (energy/chemical), and Global Healthcare Exchange LLC (health care).

In the area of supply chain management (SCM), the sharing of information allows efficiency improvements all along the chain – firms can reduce inventory levels, increase the accuracy of their forecasting, and supply parts in a timely and cost-efficient manner. With the move from EDI to the Internet, even small firms can easily participate.

Michael Hammer (2001) identifies nine things that firms must do to thrive in the customer economy of today. Four of these items (make yourself easy to do business with; sell through, not to, your distribution channels; push past your boundaries in pursuit of efficiency; and lose your identify in an extended enterprise) are directly facilitated by the Internet and Web initiatives.

When developing an EC or Web initiative there are many things to consider. A partial list includes: the source and target (business, consumer, government), the focus (internal or external or both), whether the objective is efficiency or effectiveness, go it alone versus partnership, proactive versus reactive approach, targeting one-time versus ongoing customers, physical versus virtual goods, and single good/service versus package.

From an ROI perspective, any investment in Web and e-commerce should provide a reasonable return. Yet, traditional investment analysis techniques may not be appropriate. Kohli et al. (2003) discuss the unique challenges that e-business environments pose to the measurement of IT payoff. These include the productivity paradox, level of measurement, choice of metrics, and the measurement process. The authors identify four general areas for future research (appropriate metrics, the e-business environment, technology, and business process change).

FUTURE TRENDS

There is considerable uncertainty about predicting the future, particularly with IT and EC. Yet, there are current observable trends.

On the technology side, there is the gradual convergence of technical standards/protocols (although some areas, such as Web service standards, are currently described as a battle ground – Koch, 2003). A benefit of collaborative consortiums is that a single protocol can be agreed upon by all participants. Certainly, the number of technologies and software capabilities will increase, and the move towards commoditization will continue, with the vendor community battling the Open Source community. For most firms, the problem will continue to be which technology to implement for what purpose. Overall, firms will continue to have more difficulty implementing technical initiatives because of non-technical issues. Usability of Web sites will continue to improve, as firms apply human factors analysis and other design procedures to implement better Web sites (see the Nielsen Norman group at www.useit.com). Research opportunities in this area will include empirical studies of how different types of organizations are using new technology, along with the study of implementation methods and challenges, and normative studies of how organizations should decide on which technologies to implement, when and how.

Customer trends are also noticeable. Today's customers have heightened expectations. One approach to meeting these is through self-service applications (a "win/win" for both customers and the firm) for which Web initiatives are ideal. The development and use of intra-organizational applications will continue to grow. Growing use of analytics will assist firms in segmenting their customers, quickly becoming aware of demand shifts, and providing increased value at lower cost. Retail customers will have fewer Web site security concerns and continue to become more comfortable with Web purchasing of goods and services. Research opportunities include developing better analytics, and studying the behaviour of early adopters, mainstreamers, and laggards.

At the firm level, several things are noticeable. As firms have gained experience with EC, many problems (such as channel conflict) have been identified and dealt with. SCM has led in consolidating firms into value chain networks. An increased understanding of the business levers for effective e-commerce strategies is leading to better decisions and standardization on best practices (which results in operational efficiency, but not necessarily in competitive advantage). In addition, there will be continued digitization of business processes; current estimates are that only 20%-25% are currently digitized (Kalakotra & Robinson, 2003). Smaller firms, facing resource scarcity, are likely to continue to lag in EC initiatives (Craig, 2002).

Shaw (1999) identified seven issues for EC research: (1) scope, components and potential impacts, (2) B2C EC, (3) strategy for developing online business and digital services, (4) B2B EC, (5) security, privacy and legal concerns, (6) technology and infrastructure for EC, (7) strategy for coordinating channel partners and streamlining their processes. While much has been learned about these areas, as many new questions have appeared; so these seven issues remain as attractive EC research areas.

CONCLUSION

Business success comes from making difficult decisions about business, IT and EC strategy, then implementing these successfully, using appropriate resources. While spending on IT and EC will continue, it is possible to under-spend, as well as overspend. Firms need to set priorities for investments, and ensure their business strategy aligns with their IT strategy (Farrell et al., 2003) and EC strategy. Sometimes this will result in being an IT and EC leader within an industry segment, and sometimes it will mean staying within the mainstream group or even lagging (Carr, 2003). This is an importance choice for firms and should be made consciously, rather than by default.

Web initiatives, when undertaken as part of a firm's EC strategy, will only be successfully implemented when adequate resources are provided for each project, and appropriate project management processes followed. For leading edge (bleeding edge?) projects, significant contingency allowances are important, and firms must be prepared for occasional failures (one could call these "learning experiences").

Finally, firms should not overly focus on technology (Barua et al., 2001; Davenport et al., 2001) when it comes to EC and Web initiatives. Good, appropriate technology (not necessarily the best), successfully implemented, will bring significant benefits at a reasonable cost.

REFERENCES

Andal-Ancion, A., Cartwright, P.A., & Yip, G.S. (2003). The digital transformation of traditional businesses. *MIT Sloan Management Review, 44*(4), 34-41.

Barua, A., Konana, P., Whinston, A.B., & Yin, F. (2001). Driving e-business excellence. *MIT Sloan Management Review, 43*(1), 36-44.

Carr, N.G. (2003). IT doesn't matter. *Harvard Business Review, 81*(5), 41-49.

Craig, R. (2002). Web initiatives & e-commerce strategy: How do Canadian manufacturing SMEs compare? In S. Burgess (Ed.), *Managing information technology in small business* (pp. 193-208). Hershey, PA: Idea Group Publishing.

Crossan, M., Fry, J., & Killing J. (2002). *Strategic analysis and action* (5th ed.). Toronto: Prentice Hall.

Davenport, T.H., Harris, J.G., De Long, D.W., & Jacobson, A.L. (2001). Data to knowledge to results: Building an analytic capability. *California Management Review, 43*(2), 117-138.

Farrell, D., Terwilliger, T., & Webb, A.P. (2003). Getting IT spending right this time. *The McKinsey Quarterly, 2003*(2).

Hammer, M. (2001). *The agenda: What every business must do to dominate the decade.* New York, NY: Crown Publishing.

Henderson, J.C., & Venkatraman, N. (1993). Strategic alignment: Leveraging information technology for transforming organizations. *IBM Systems Journal, 32*(1), 4-16.

Hoffman, W., Keedy, J., & Roberts, K. (2002). The unexpected return of B2B. *The McKinsey Quarterly, 2002*(3).

Ives, B., & Piccoli, G. (2003). Custom-made apparel at Lands' End. *Communications of the Association for Information Systems, 11*(3), 79-93.

Kalakota, R., & Robinson, M. (2003). *Services blueprint: Roadmap for execution.* Reading, MA: Addison-Wesley.

Koch, C. (2003). The battle for Web services. *CIO Magazine, 16*(19).

Kohli, R., Sherer, S.A., & Baron, A. (2003). Editorial – IT investment payoff in e-business environments: Research issues. *Information Systems Frontiers, 5*(3), 239-247.

Magretta, J. (2002). Why business models matter. *Harvard Business Review, 80*(5), 86-91.

Porter, M.E. (1996). What is strategy? *Harvard Business Review, 74*(6), 61-78.

Porter, M.E. (2001). Strategy and the Internet. *Harvard Business Review, 79*(3), 62-78.

Rayport, J.F., & Sviokla, J.J. (1995). Exploiting the virtual value chain. *Harvard Business Review, 73*(6), 75-85.

Shaw, M.J. (1999). Electronic commerce: Review of critical research issues. *Information Systems Frontiers, 1*(1), 96-106.

Tallon, P.P., & Kraemer, K.L. (2003). Investigating the relationship between strategic alignment and IT business value: The discovery of a paradox. In N. Shin (Ed.), *Creating business value with information technology: Challenges and solutions* (pp. 1-22). Hershey, PA: Idea Group Publishing.

KEY TERMS

Business Model: A specific arrangement of organizational strategies, goals, processes, resources (technologies, finances, people, etc.), structures, products and services that enable a firm to successfully compete in the marketplace. Many EC researchers have taken a narrower view, based on organizations involved (i.e., B2B, B2C, B2G, etc.), or specific framework used (i.e., hierarchy, hub, or intermediary for e-markets). While there is not yet a consensus about what makes up a business model, the trend is away from a narrower view.

E-Business Model: That subset of the general business model that supports e-business.

E-Commerce: Commercial activities taking place over electronic networks (primarily the Internet); e-commerce is a subset of general commerce.

E-Commerce Strategy: A subset of general business and information technology strategy, focusing on Web-based commercial opportunities. It may dominate general strategy in some firms.

M-Commerce: Mobile commerce, with wireless access to the Web.

Protocol: A set of rules and procedures that govern transmission between the components of an electronic network.

Strategy: The determination of the basic long-term goals and objectives of an organization, and the adoption of courses of action and allocation of resources necessary for achieving these goals; major components of strategy include: goals, product/market focus, business system focus, and competitive premise.

Web Initiative: Any use of the World Wide Web for a specific purpose.

Web Personalization: Customizing Web content, in real time, to a specific user.

Web Search via Learning from Relevance Feedback

Xiannong Meng
Bucknell University, USA

Zhixiang Chen
University of Texas-Pan Americana, USA

INTRODUCTION

Recently, three general approaches have been taken to increase Web search accuracy and performance. One is the development of meta-search engines (e.g., MetaCrawler, *www.metacrawler.com*) that forward user queries to multiple search engines at the same time in order to increase the coverage and hope to include what the user wants in a short list of top-ranked results. Another approach is the development of topic-specific search engines that are specialized in particular topics. These topics range from vacation guides (*www.vocations.com*) to kids' health (*www.kidshealth.com*). The third approach is to use some group or personal profiles to personalize the Web search. Examples of such efforts include *GroupLens* (Konstan et al., 1997).

Those meta-search engines suffer to a certain extent the inherited problem of information overflow that it is difficult for users to pin down specific information for which they are searching. Specialized search engines typically contain much more accurate and narrowly focused information. However, it is not easy for a novice user to know where and which specialized engine to use. Most personalized Web search projects reported so far involve collecting users' behavior at a centralized server or a proxy server. While it is effective for the purpose of e-commerce, where vendors can collectively learn consumer behaviors, this approach does present the privacy problem.

The clustering, user profiling and other advanced techniques used by those search engines and other projects such as Bollacker et al. (1998) are *static* in the sense that they are built before the search begins. They cannot be changed dynamically during the real-time search process. Thus they do not reflect the changing interests of the user at different time, at different location or on different subjects. Intelligent Web search systems that dynamically learn the users' information needs in real-time must be built to advance the state of the art in Web search. Machine learning techniques can be used to improve Web search, because machine learning algorithms are able to adjust the search process dynamically so as to satisfy users' information needs.

BACKGROUND

There have been great research efforts on applications of machine learning to automatic extraction, clustering and classification of information from the Web. Some earlier research includes *WebWatcher* (Armstrong et al., 1995), which interactively helps users locate desired information by employing learned knowledge about which hyperlinks are likely to lead to the target information; *Syskill and Webert* (Pazzani et al., 1996), a system that uses a Bayesian classifier to learn about interesting Web pages for the user; and *NewsWeeder* (Lang, 1995), a news-filtering system that allows the users to rate each news article being read and learns a user profile based on those ratings. Some research is aimed at providing adaptive Web service through learning. For example, *Ahoy! The Homepage Finder* (Shakes et al., 1997) performs dynamic reference shifting; and *Adaptive Web Sites* (Perkowitz & Etzioni, 2000) automatically improves their organization and presentation based on user access data.

A series of work in Chen et al. (1999, 2000, 2001, 2002) and Meng and Chen (2004) study intelligent Web search as an adaptive learning process, where the search engine acts as a learner and the user as a teacher. The user sends a query to the engine, and the engine uses the query to search the index database and returns a list of URLs that are ranked according to a ranking function. Then the user provides the engine relevance feedback, and the engine uses the feedback to improve its next search and returns a refined list of URLs. The learning (or search) process ends when the engine finds the desired documents for the user. Conceptually, a query entered by the user can be understood as the logical expression of the collection of the documents wanted by the user. A list of URLs returned by the engine can be interpreted as an approximation to the collection of the desired documents.

ADAPTIVE LEARNING FOR WEB SEARCH

Let ***X*** denote the set of all index keywords for the whole Web (or, practically, a portion of the whole Web). Given any Web document *d*, let *I(d)* denote the set of all index keywords in ***X*** that are used to index *d* with non-zero values. Then, the following two properties hold. (1) The size of *I(d)* is substantially smaller than the size of ***X***. Practically, *I(d)* can be bounded by a constant. The rationale behind this is that in the simplest case only a few of the keywords in *d* are needed to index it. (2) For any search process related to the search query *q*, let *D(q)* denote the collection of all the documents that match *q*; then the set of index keywords relevant to *q*, denoted by *F(q)*, is $F(q) = \cup_{d \in D(q)} I(d)$. Although the size of *F(q)* varies from different queries, it is still substantially smaller than the size of *X*, and might be bounded by a few hundred or a few thousand in practice.

Definition 1 - *Given any search query q, F(q), which is given in the previous paragraph, is defined as the set of dynamic features relevant to the search query q.*

Definition 2 - *Given any search query q, the dynamic vector space V(q) relevant to q is defined as the vector space that is constructed with all the documents in D(q) such that each of those documents is indexed by the dynamic features in F(q).*

Adaptive learning for intelligent Web search as studied in Chen et al. (1999, 2000, 2001, 2002) and Meng and Chen (2004) is formulated as follows. Let *S* be a Web search system. For any query *q*, *S* first finds the set of documents *D(q)* that match the query *q*. It finds *D(q)* with the help of a general-purpose search strategy through searching its internal database, or through external search engine such as AltaVista (*www.altavista.com*) when no matches are found within its internal database. It then finds the set of dynamic features *F(q)*, and later constructs the dynamic vector space *V(q)*. Once *D(q)*, *F(q)* and *V(q)* have been found, *S* starts its adaptive learning process with the help of the learning algorithm that is to be presented in the following subsections. More precisely, let $F(q) = \{K_1, \ldots, K_n\}$ such that each K_i denotes a dynamic feature (i.e., an index keyword). *S* maintains a common weight vector $w = (w_1, \ldots, w_n)$ for dynamic features in *F(q)*. The components of *w* have non-negative real values. The learning algorithm uses *w* to extract and learn the most relevant features and to classify documents in *D(q)* as relevant or irrelevant.

Practically efficient adaptive learning algorithms such as TW2 have been developed in (Chen et al. (1999, 2000,

Figure 1. Algorithm TW2

Step 0: Set $w_0 = (w_{01}, \ldots, w_{0n}) = (0, \ldots, 0)$.

Step $i > 0$: Classify documents in $D(q)$ with $\sum_{j=1}^{n} w_{ij} x_j > \theta$.

While (user judged a document $x = (x_1, \ldots, x_n)$) do
    If (x is relevant) do //promotion
        For $(j = 1;\ j \le n;\ j{+}{+})$
            If $(x_j = 0)$ set $w_{i+1,j} = w_{ij}$
            If $(x_j \ne 0\ \&\ w_{ij} = 0)$ set $w_{i+1,j} = \alpha$
            If $(x_j \ne 0\ \&\ w_{ij} \ne 0)$ set $w_{i+1,j} = \alpha w_{ij}$
    If (x is irrelevant) do //demotion
        For $(j = 1;\ j \le n;\ j{+}{+})$
            Set $w_{i+1,j} = w_{ij} / \alpha$
If (user has not judged any document) stop
Else set $i = i + 1$ and go to step $i + 1$.

2001, 2002) and Meng and Chen (2004). For example, for each query *q* entered by the user, algorithm *TW2* uses a common weight vector *w* and a real-valued threshold *q* to classify documents in *D(q)*. Let *a > 1* be the promotion and demotion factor. *TW2* classifies documents whose vectors $x = (x_1, \ldots, x_n)$ satisfy $\sum_{i=1}^{n} w_i x_i > \theta$ as relevant, and all others as irrelevant. When the user responds with a document that may or may not contradict to the current classification, TW2 updates the weights through promotion or demotion. The detailed description of TW2 is given in Figure 1.

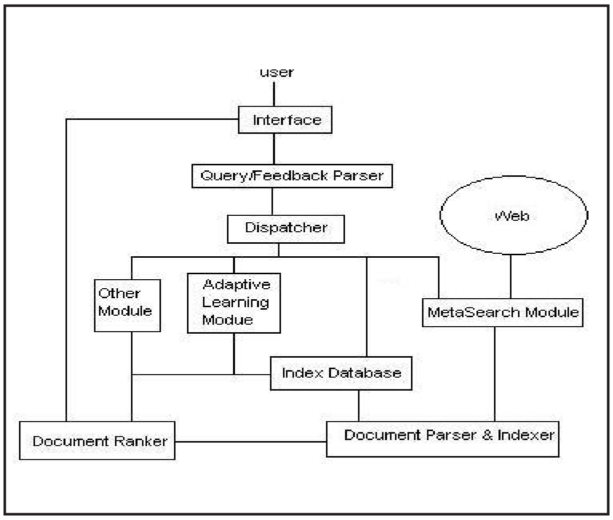
In contrast to the linear lower bounds proved for Rocchio's similarity-based relevance feedback algorithm (Chen & Zhu, 2002), *TW2* has a very small mistake bounds for learning any collection of documents represented by a disjunction of a small number of relevant features. As pointed out in Chen et al. (2002) that to learn a collection of documents represented by a disjunction of at most *k* relevant features (or index keywords) over the *n*-dimensional Boolean vector space, *TW2* makes at most $\frac{\alpha^2 A}{(\alpha-1)\theta} + (\alpha+1)k \ln_\alpha \theta - \alpha$ mistakes, where A is the number of dynamic features occurring in the learning process.

Other multiplicative adaptive algorithms for user preference retrieval were designed in Chen (2001) and Meng and Chen (2004). It was shown that those two algorithms have substantially better performances than Rocchio's similarity-based relevance feedback and the gradient descent procedure (Wong et al., 1988).

SYSTEM IMPLEMENTATION

Several experimental intelligent Web search systems have been implemented and analyzed in Chen et al. (1999, 2000, 2001, 2002) Meng and Chen (2004). An exemplary architecture of those systems is shown in Figure 2. In general, such a system has a graphic user interface to allow the user to enter his/her query and to specify the number of the top matched document URLs to be returned. It may or may not maintain an internal index database of documents. It has a meta-search module to query a set of general-purpose search engines such as AltaVista whenever needed. When the user enters a query and starts a search process, it first searches its internal index database. If no relevant documents can be found within its index database then it receives a list of top matched documents externally with the help of its meta-search module.

Figure 2. System architecture



At each iteration, the system provides the user a short list of documents so that the user can judge whether a document is relevant or not. After the user finishes judging documents as relevance feedback to the system, the query/feedback parser parses out the feedback and sends it to the adaptive learning module for the underlying learning algorithms to learn the user's preference. At the end of the current iteration of learning, the system re-ranks the documents with its document ranker and displays a short list of documents to the user. At each iteration, the dispatcher parses query or relevance feedback information from the interface and decides which of the following components should be invoked to continue the search process: adaptive learning module, or index database, meta search module, or other module if there is one. The document parser and indexer can parse and index documents returned by the meta search module, and send those documents to the index database or document ranker. The other module can have additional functions for clustering, profiling, personalizing or feature extracting, to enhance the capability of the adaptive learning module.

The experimental systems in Chen et al. (1999, 2000, 2001, 2002) and Meng and Chen (2004) are evaluated with performance metrics of relative recall, precision, response time, and/or ranking accuracy. For example, relative recall and precision are defined as follows. For any query *q*, the relative recall and precision are

$$Rrecall = \frac{|Rm|}{|R|}, \quad Rprecision = \frac{|Rm|}{m},$$

where *R* is the total number of relevant documents among the set of the retrieved documents, and Rm is the number of relevant documents ranked among the top *m* positions in the final search result of the search engine. Empirical analytical results in Chen et al. (1999, 2000, 2001, 2002) and Meng and Chen (2004) all show that adaptive learning algorithms significantly increase Web search performance and accuracy in comparison to general-purpose search engines such as AltaVista.

FUTURE TRENDS

With an overwhelming amount of information available on the Web, using an adaptive learning approach to refine Web search results will become the natural next step of Web search. Users can interact with search engines to refine their queries to narrow down the collection of search results. In this refinement process, the search engine is *learning* the user's information needs to provide more accurate search results. During this interactive process, the search engines can also pick up other information automatically or manually regarding this particular search from the user. Such information may include past search history, and user preferences, among others. With the capability of interacting with users, the search engines will become more intelligent and a true assistant in serving users' information needs.

CONCLUSION

This article presents an overview of research on intelligent Web search through adaptive learning from relevance feedback (Chen et al., 1999, 2000, 2001, 2002; Meng & Chen, 2004).

The authors believe that other techniques such as Web mining and clustering shall be used to further enhance the performance of the adaptive learning algorithms. Adaptive learning approach to intelligence Web search requires the user to interact with the search engine. Through such interaction the user provides relevance feedback to the adaptive learning algorithm used by the search engine to improve search results. Adaptive learning algorithms cannot work well without users' interaction of relevance feedback. On other hand, Web mining and clustering can uncover hidden patterns of user access behaviors and associations of these patterns with groups of users from large Web access logs and user profiles. These patterns and associations can be used to predict user interests or information needs in order to help the search engine to filter out undesired Web pages and to recommend desired ones. Web mining and clustering are essentially retrospective techniques, and cannot adaptively interact with the user. Once incorporated with an adaptive learning algorithm, these techniques can help the learning algorithm to build the initial search result for the user before the interaction starts. They can also help the adaptive learning algorithm to include, or exclude, some of the Web pages, when a new search result is computed with respect to the user's recent relevance feedback.

REFERENCES

Armstrong, R., Freitag, D., Joachims, T., & Mitchell, T. (1995). Webwatcher: A learning apprentice for the World Wide Web. *Working Notes of the AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments* (pp. 6-12). AAAI Press.

Balabanovi, M. (1997). An adaptive Web page recommendation service. *Proceeding of the 1st International Conference on Autonomous Agents* (pp. 378-387). New York: ACM Press.

Bollacker, K., Lawrence, S., & Giles, C.L. (1998). Citeseer: An autonomous Web agent for automatic retrieval and identification of interesting publications. *Proceedings of the 2nd International Conference on Autonomous Agents* (pp. 116-113). New York: ACM Press.

Chen, X., Zhu, B., & Meng, X. (2002). Intelligent Web search through adaptive learning from relevance feedback. In N. Shi & V.K. Murthy (Eds.), *Architectural issues of Web-enabled electronic business* (pp. 139-153). Hershey, PA: Idea Group Publishing.

Chen, Z. (2001). Multiplicative adaptive algorithms for user preference retrieval. *Proceedings of the 7th International Computing & Combinatorics Conference, Lecture Notes in Computer Science, 2108*, 540-549. Heidelberg, Germany: Springer-Verlag.

Chen, Z., & Zhu, B. (2002). Some formal analysis of the Rocchio's similarity-based relevance feedback algorithm. *Information Retrieval, 5*(1), 61-86.

Chen, Z., & Meng, X. (2000). Yarrow: A real-time client-side meta-search learner. *Proceedings of the AAAI 2000 Workshop on Artificial Intelligence for Web* (pp. 12-17). AAAI Press.

Chen, Z., Meng, X., & Fowler, R.H. (1999). Searching the Web with queries. *Knowledge and Information Sysems, 1*, 69-375.

Chen, Z., Meng, X., Fowler, R., & Zhu, B. (2001). FEATURES: Real-time adaptive feature andmmdocument learning for Web search. *Journal of the American Society for Information Science, 52*(8), 655-665.

Chen, Z., Meng, X., Zhu, B., & Fowler, R. (2000). WebSail: From on-line learning to Web search. *Knowledge and Information Science*, *4*(2), 219-227.

Konstan, J., Miller, B., Maltz, D., Herlocker, J., Gordon, L., & Riedl, J. (1997). GroupLens: Applying collaborative filtering to Usernet news. *Communications of ACM*, *40*(3), 77-87.

Lang, K. (1995). Newsweeder: Learning to filter news. *Proceedings of the 12th International Conference on Machine Learning* (pp. 331-339).

Meng, X., & Chen, Z. (2004). MARS: Multiplicative adaptive refinement Web search. In A. Scime (Ed.), *Web mining: Applications and techniques*. Hershey, PA: Idea Group Publishing.

Pazzani, M., Muramatsu, J., & Billus, D. (1996). Syskill & Webert: Identifying interesting Web sites. *Proceedings of the 13th National Conference on Artificial Intelligence* (pp. 54-61).

Perkowitz, M., & Etzioni, O. (2000). Adaptive Web sites: Concept and case study. *Artificial Intelligence, 118*, 245-275.

Rocchio, J. (1971). Relevance feedback in information retrieval. In G. Salton (Ed.), *The smart retrieval system - experiments in automatic document processing* (pp. 313-323). NJ: Prentice-Hall, Inc.

Shakes, J., Langheinrich, M., & Etzioni, O. (1997). Dynamic reference sifting: A case study in the homepage domain. *Proceedings of the 6th International World Wide Web Conference* (pp. 189-200).

Wong, S.K.M., Yao, Y.Y., & Bollmann, P. (1988). Linear structures in information retrieval. *Proceedings of the 12th Annual International ACM-SIGIR Conference on Information Retrieval* (pp. 219-232). New York: ACM Press.

KEY TERMS

Adaptive Learning: An algorithm that can learn a hidden concept from interactive feedback provided by a user or the underlying environment.

Document Ranking: A function that scores documents in a collection according to their relevance to a given query so that the more relevant a function is, the higher score it has.

Intelligent Web Search: A Web search system that learns a user's information preference.

Precision: Precision is the percentage of relevant document in a given query in the set of documents that are returned by an information retrieval system.

Recall: Recall is the percentage of relevant documents in a given query in a collection that are returned by an information retrieval system.

Relevance Feedback: The information that is judged by a user and provided to an information system.

Vector Space: A model that represents every document in a collection with an n-dimensional vector for some constant n.

Web Technologies and Data Warehousing Synergies

John M. Artz
The George Washington University, USA

INTRODUCTION

Data warehousing is an emerging technology that greatly extends the capabilities of relational databases specifically in the analysis of very large sets of time-oriented data. The emergence of data warehousing has been somewhat eclipsed over the past decade by the simultaneous emergence of Web technologies. However, Web technologies and data warehousing have some natural synergies that are not immediately obvious. First, Web technologies make data warehouse data more easily available to a much wider variety of users. Second, data warehouse technologies can be used to analyze traffic to a Web site in order to gain a much better understanding of the visitors to the Web site. It is this second synergy that is the focus of this article.

DATA WAREHOUSE

A data warehouse is a repository of nonvolatile temporal data used in the analysis and tracking of key business processes. Temporal or time varying is the most important characteristic that distinguishes a data warehouse from a traditional relational database, which represents the state of an organization at a point in time. A relational database is a snapshot of the organization, whereas the data warehouse is a collection of longitudinal data.

One could argue that it should be possible to store longitudinal data in a relational database, and this claim is true. However, relational databases, which model data as entities, create sever limitations in data exploitation. First, although standard SQL does provide a DateTime data type, it is very limited in its handling of dates and times. If an analyst wanted to compare summary data on weekends versus weekdays or holidays versus non-holidays, it would be difficult if not impossible using standard SQL. Second, analysis involving drill down or roll up operations becomes extremely awkward using standard SQL against entities as represented in relational tables.

Data warehousing technology overcomes these deficiencies in the relational model by representing data in a dimensional model. A dimensional model consists of a fact table (see Figure 1) and the associated dimensions. The fact table contains measures of the business process being tracked and the dimensional tables contain information on factors that may influence those measures. More specifically, the fact table contains dependent variables while the dimension tables contain independent variables. Online analytical processing (OLAP) tools provide a means of summarizing the measures in the fact table according to the dimensions provided in the dimension table toward the end of determining what factors influence the business process being modeled. Typically OLAP tools provide a means of easily producing higher levels of summary (roll-up) or greater levels of detail (drill-down).

WEB LOG

A visitor to a Web site requests a page by typing in the address of the page in a Web browser, or by clicking on a link that automatically requests that page. A message is

Figure 1. A dimensional model

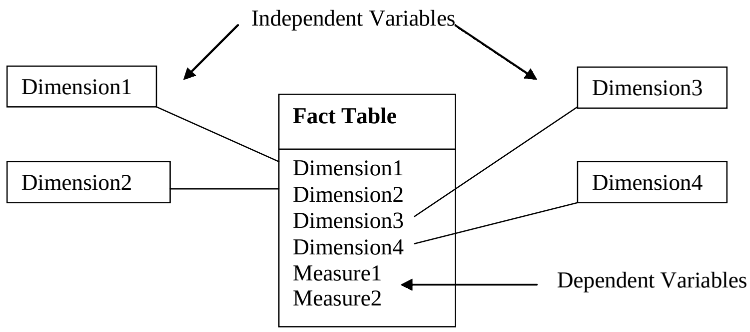
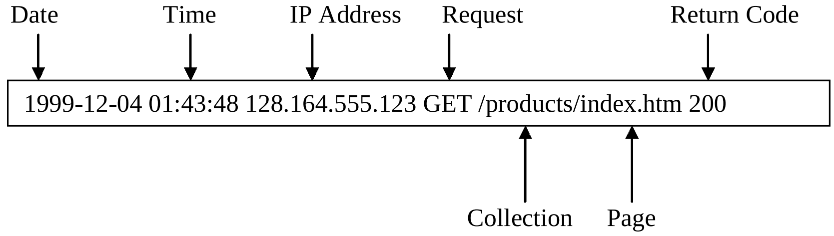


Figure 2. A typical Web log record



sent to the Web server, at that address, and the Web server responds by returning the page. Each request that is processed by the Web server is recorded in a file called the Web log, which contains a record of all activity on the Web site. The record typically contains the date and time, the IP address of the requestor, the page requested and the result. A typical record in standard format is shown in Figure 2.

From this simple log record we can determine quite a bit about the traffic coming to the Web site. For example, we can determine peak times for visitors by date or time and we can determine if Web site usage is cyclical or has other temporal patterns. Further, we can determine which pages or collections are most heavily visited and if their usage also reflects a temporal pattern. Answers to these questions are useful for site management and maintenance and for determining the effectiveness of design decisions or the behavior of the visitors to the site.

SYNERGY

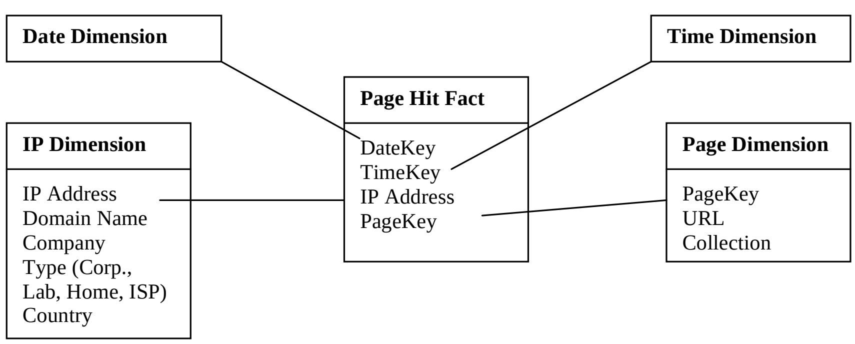
A lot of valuable information can be derived from the Web log. But that is only the beginning. By viewing the Web log as a data source for a data warehouse, it becomes an even richer source of information about the Web site and its visitors. Consider the dimensional model in Figure 3. From this we can derive relationships between visitors and the pages they visit. Web log records can be summarized to produce dwell time (the time a visitor spends viewing a page) and visit length (the time a visitor spends at the site). Further, if the site is used for sales, the IP dimension can be converted to a customer dimension and the page dimension can be converted into a product dimension allowing analysis of customer purchasing behavior. For a more in depth examination of the process of evolving a Web log into a customer data warehouse, see *From Web Log to Data Warehouse: An Evolving Example*, listed in the references.

FUTURE TRENDS

According to information navigators, there are approximately 72 million hosts on the Web. Many of these hosts do not support Web sites, but many others (such as those owned by ISPs) have multiple Web sites. If only 10 million of these Web sites attract as few as 100 visitors a minute, then 1 billion records are generated every minute. This becomes 60 billion records per hour, or 1.4 trillion records per day. This is an enormous amount of information providing great insight into the behavior of consumers and information seekers, among others. This huge volume of temporal data cannot be exploited effectively without data warehouse technology. Hence, the growth of the Web may well push the growth of data warehousing.

At the same time, data warehousing concepts continue to evolve and the technology continues to improve.

Figure 3. A dimensional model based on the Web log



OLAP tools are making it easier to analyze dimensional data through visual interfaces and data mining tools are making it easier to find useful patterns in large volumes of data. In today's tools there are gaps between SQL and OLAP, and again between OLAP and data mining. These disparate approaches will eventually become integrated in seamless tool sets that provide a variety of analytical techniques along with some guidance on which approaches to use.

The collection and analysis of all these data is, in many ways, a double-edged sword. From a productivity perspective, making white-collar workers more productive is clearly a good thing. However, measuring and monitoring have a very dark downside. There has been much concern since the industrial revolution that the mechanization of manufacturing could lead to undesirable working conditions for factory workers. Similarly, the mechanization of white-collar work could have similar impacts.

Further, the amassing of huge amounts of personal information is a sword that swings both ways also. On the positive side, understanding our needs allows product producers to meet those needs more fully. When you go to a shopping Web site you are often prompted with teasers that say "If you liked product X you should try product Y". If product Y is something that you think you might want, this is good. However, many people feel that the collection of such large amounts of personal data could have serious implications for personal privacy. Indeed, it may. It all depends on how the data are used and on how that use is perceived.

CONCLUSION

Data warehousing technologies and Web technologies are highly synergistic. Web technologies make data warehouse data more accessible to a much wider variety of users on both the Internet and the Intranet. Since the value of data is determined by their usefulness in supporting decision-making, Web technologies help increase the value of the data in the warehouse. Perhaps, more importantly, as shown in this article, data warehouse technologies can be used to analyze the enormous volumes of data generated by a Web site. As Web sites generate large volumes of data for data warehouses and in turn provide easy access to those data, it is conceivable that these two highly synergistic technologies may become indistinguishable in the future.

REFERENCES

Adriaans, P., & Zantinge, D. (1997). *Data mining.* Addison-Wesley.

Artz, J. (2001). From Web log to data warehouse: An evolving example. In S. Dasgupta (Ed.), *Managing Internet and intranet technologies in organizations: Challenges & opportunities.* Hershey, PA: Idea Group Publishing.

Han, J., & Kamber, M. (2001). *Data mining: Concepts and techniques.* Morgan Kaufman.

Information Navigators Internet Growth Charts. *http://navigators.com/statall.gif*

Inmon, W. (2002). *Building the data warehouse* (2^{nd} ed.). John Wiley and Sons, Inc.

Kimball, R. (2002). *The data warehouse toolkit.* John Wiley & Sons, Inc.

Kimball, R. et al. (1998). *The data warehouse lifecycle toolkit.* John Wiley & Sons, Inc.

Kimball, R., & Merz, R. (2000). *The data Webhouse toolkit.* John Wiley & Sons, Inc.

Thomsen, E. (2002). *OLAP solutions* (2^{nd} ed.). John Wiley & Sons.

Westphal, C., & Blaxton, T. (1998). *Data mining solutions.* John Wiley & Sons, Inc.

KEY TERMS

Data Mining: The use of computer algorithms to discover relationships between data items than are hidden in the complexity of the data.

Data Warehouse: A repository of nonvolatile temporal data used in the analysis and tracking of key business processes.

Dimensional Model: A data model that represents measures of a key business process in the form of facts and the independent variables that affect those measurements.

Key Business Process: A collection of intentional business activities that is both measurable and worthy of improvement or optimization.

OLAP: On Line Analytical Process usually distinguished from On Line Transaction Processing, which is the model associated with transaction-oriented databases.

Temporal Data: Time oriented data.

Web Log: A file in which a Web server records requests for pages.

Web Tools for Molecular Biological Data Analysis

Denise Fukumi Tsunoda
Centro Federal de Educação Tecnológica do Paraná, CEFET/PR, Brazil

Heitor Silvério Lopes
Centro Federal de Educação Tecnológica do Paraná, CEFET/PR, Brazil

Ana Tereza Vasconcelos
Laboratório Nacional de Computação Científica, LNCC, Brazil

INTRODUCTION

Bioinformatics means solving problems arising from biology using methods from computer science. The National Center for Biotechnology Information (*www.ncbi.nih.gov*) defines bioinformatics as:

"…the field of science in which biology, computer science, and information technology merge into a single discipline...There are three important sub-disciplines within bioinformatics: the development of new algorithms and statistics with which to access relationships among members of large data sets; the analysis and interpretation of various types of data including nucleotide and amino acid sequences, protein domains, and protein structures; and the development and implementation of tools that enable efficient access and management of different types of information."

There are many sub-areas in bioinformatics: data comparison, data analysis, DNA assembly, protein structure prediction, data visualization, protein alignment, phylogenetic analysis, drug design, and others.

The biological data (sequences and structures) are naturally very large. In addition, the number of records in the biological databases is increasing every year because of intensive research in the field of molecular biology. Analysis of this overwhelming amount of data requires intelligent bioinformatics tools in order to manage these data efficiently.

During the past two decades, the world has witnessed a technological evolution that has provided an unprecedented new medium of communications to mankind. By means of the World Wide Web, information in all forms has been disseminated throughout the world. Since the beginning, research in bioinformatics primarily used the Internet due to the fast information dissemination it allows, at essentially no cost.

This article aims to discuss some bioinformatics Web tools, but given the accelerated growth of the Web and the instability of the URLs (Uniform Resource Locators), an Internet search engine should be used to identify the current URL.

BACKGROUND

Living organisms possess entities called *genes* that are the basic inherited units of biological function and structure. An organism inherits its genes from its parents, and relays its own genes to its offspring.

Molecular biologists, in the later half of the 20th century, determined that the gene is made of DNA (deoxyribonucleic acid)—that is, DNA is the heredity material of all species. More than 50 years ago, Crick and Watson (1953) discovered the *double helix* structure of DNA and concluded that this specific form is fundamental to DNA's function.

Each strand of the DNA double helix is a polymer (a compound made up of small simple molecules) consisting of four elements called *nucleotides*: *A*, *T*, *C*, and *G* (for adenine, thymine, cytosine, and guanine). The two strands of DNA are complementary: when a *T* resides on one strand, an *A* occupies the corresponding position on the other strand; when there is a *G* on one strand, a *C* occupies the corresponding position on the other. The sequence of nucleotides encodes the "instructions" for forming all other cellular components and provides a template for the production of an identical second strand in a process called replication.

From a computer scientist's point of view, the DNA is information storage and a transmission system. Like the binary alphabet {0,1} used in computers, the alphabet of DNA {*A*, *T*, *C*, *G*} can encode messages of arbitrary complexity when encoded into long sequences.

The decoding of the genetic information is carried out through intermediary RNA (ribonucleic acid) molecules that are transcribed from specific regions of the DNA. RNA molecules use the base uracile (U) instead of a thymine. RNA is then translated into a protein—a chain assembled from the 20 different simple amino acids. Each consecutive triplet of DNA elements specifies one amino acid in a protein chain. Once synthesized, the protein chain folds—according to the laws of chemistry/physics—into a specific shape, based on the properties and order of its amino acids. The structures of a protein can be viewed hierarchically (Lehninger, Nelson & Cox, 2000): primary (linear amino acid sequence), secondary (local sequence elements with well-determined regular shape like á-helices and â-strands), tertiary (formed by packing secondary structures into one or several compact globular units), and quaternary (combination of tertiary structures).

BIOINFORMATICS WEB TOOLS

Sequence Analysis

There is a known relationship between sequence and structure of proteins, since proteins with similar sequences tend to have similar three-dimensional structures and functions. Sequence alignment methods are useful when it is necessary to predict the structure (or function) of a new protein whose sequence has just been determined. Therefore, alignment provides a powerful tool to compare two (or more) sequences and could reflect a common evolutionary origin.

ClustalW is a general purpose multiple sequence alignment program for DNA or proteins (Higgins et al., 1994). ClustalW currently supports seven multiple sequence formats that are detailed (including examples) in the *ClustalW Services Help Menu*: NBRF/PIR, EMBL/SwissProt, FASTA, GDE, ALN/ClustalW, GCG/MSF, and RSF. It produces biologically meaningful multiple sequence alignments of divergent sequences and calculates the best match for the selected sequences, considering individual weights, amino acid substitution matrices—like PAM (Altschul, Gish, Miller, Myers & Lipman, 1991) or Blosum (Henikoff & Henikoff, 1992)—and gap penalties (Apostolico & Giancarlo, 1998). After the identities, similarities and differences can be seen. ClustalW is freely available on the Internet, either as a Web-based tool or for downloading.

Another tool, T-Coffee (Notredame, Higgins & Heringa, 2000) is more accurate than ClustalW for sequences with less than 30% identity, but much slower. The T-Coffee input must have from 3 to 30 sequences (or 10,000 characters) in the FASTA format. The submission form is simple, but it does not allow user-selected options.

Cinema—Colour INteractive Editor for Multiple Alignments—is a program for sequence alignment that allows visualization and manipulation of both protein and DNA sequences (Parry-Smith, Payne, Michie & Attwood, 1997). It is a complete package in Java, locally installed, that runs on most platforms. This tool allows upload of an alignment from a local computer to the Cinema server. The input file must be in a PIR format and may then be imported into Cinema via the *Load Alignment File* option.

Structural Analysis

The Dali—Distance mAtrix aLIgnment—server (Holm & Sander, 1994) is a network service for comparing three-dimensional (3D) protein structures. Once the coordinates of a query protein structure is submitted, Dali compares them against those in the Protein Data Bank (PDB). The input file must be in the PDB format (Berman et al., 2000) and can be submitted by e-mail or interactively from the Web. The input options are *disabled*. The results are mailed back to the user. In favorable cases, comparing 3D structures may reveal biologically interesting similarities that are not detectable in primary sequences. There is a Dali database built based on exhaustive all-against-all 3D structure comparison of protein structures currently in the Protein Data Bank (PDB). The classification and alignments are continuously updated using the Dali search engine.

The Macromolecular Structure Database tool (MSD) (Golovin et al., 2004) allows one to search the active site database based on ligand or active site information. The PDB contains a significant number of protein structures that have ligands bound which are often more highly conserved across a functional family than the overall structure and fold of the macromolecule. The target of the search can be based on an uploaded file. It is possible to limit the scope of a search using restrictions based on author, keywords, experiment, resolution, and release date. Results of the search are presented in a list of PDB ID codes that can be analyzed further or viewed within a structure viewer like Rasmol (Sayle & Milner-White, 1995)—a program that intends the visualization of proteins, nucleic acids, and small molecules.

Swiss-Model (Schwede, Kopp, Guex & Peitsch, 2003) is a server for automated comparative modeling of 3D protein structures, and provides several levels of user interaction using a Web-based interface: in the first approach mode, only an amino acid sequence is submitted to build a 3D model. It could also be accessible from the program DeepView—an integrated sequence-to-structure workbench. All models are mailed back with a detailed modeling report. Template selection, alignment, and model building are automatically done by the server. The Swiss-Model alignment interface allows the submission of multiple sequence alignments in the following formats: FASTA, MSF, ClustalW, PFAM, and SELEX. The alignment must contain at least the target sequence and one template from the ExPDB template library, because the modeling process is based on this user-defined template.

Homology and Similarity Tools

The extent of similarity between two sequences can be based on percent sequence identity and/or conservation. In BLAST—**B**asic **L**ocal **A**lignment **S**earch **T**ool (Altschul et al., 1990)—similarity refers to a positive matrix score. A homology class is a class whose members have been inferred to have evolved from a common ancestor.

The BLAST input file could be a sequence in FASTA format, lines of sequence data, or NCBI sequence identifiers, as explained at the *Blast Search Format* form. BLAST emphasizes regions of local alignment to detect relationships among sequences which share only isolated regions of similarity. Results (many formats available) can be seen by using the browser or e-mailed to the user. Since the BLAST algorithm can detect both local and global alignments, regions of similarity in unrelated proteins can be detected, such that the discovery of similarities may provide important clues to the function of uncharacterized proteins.

FASTA (Pearson, 2000) provides sequence similarity and homology searching among a protein sequence to another protein sequence or to a protein database, or a DNA sequence to another DNA sequence or a DNA library. It can be very specific when identifying long regions of low similarity, especially for highly diverged sequences. Release 3.x of the FASTA package provides a modular set of sequence comparison programs that can run on conventional single processor computers or in parallel on multiprocessor computers.

BLAST and FASTA are used for database searching because they are fast. They use slightly different approaches to discover similar sequences, but both make refinements during searching process to increase the searching speed.

Protein Function Analysis

InterProScan (Zdobnov & Apweiler, 2001) is a tool that combines different protein signature recognition methods into a single resource. The number of signature databases and their associated scanning tools, as well as the further refinement procedures, increase the complexity of the problem. InterProScan performs considerable data look-ups from databases and program outputs. This Web tool allows the input of protein sequences, either in single or multiple files. The input file format for protein sequence(s) are free text, FASTA, or UniProt; for nucleotide sequence, the GenBank format is also accepted.

GeneQuiz (Scharf et al., 1994) is an integrated system for large-scale biological sequence analysis that goes from a protein sequence to a biochemical function, using a variety of search and analysis methods and up-to-date protein and DNA databases. The input file must be in the FASTA format; the maximum number of sequences that can be uploaded per day is 12, and the maximum number of amino acids is 18,000. GeneQuiz automatically runs a series of sequence analysis tools, including BLAST and FASTA. The results are displayed as structured text. The server is freely provided to the biological research community.

STING Millennium Suite—SMS (Neshich et al., 2003)—is a Web-based suite of programs and databases providing visualization and analysis of molecular sequence and structure for the PDB data. SMS operates with a huge collection of data (PDB, HSSP, Prosite). STING Millennium is both a didactic and a research tool. The interface is user friendly, and there are many options available for macromolecular studies.

ExPASy—Expert Protein Analysis System (Gasteiger et al., 2003)—provides access to a variety of databases and analytical tools dedicated to proteins and proteomics, including Swiss-Prot and TrEMBL, Swiss-2Dpage, PROSITE, ENZYME, and the Swiss-Model repository. There is also the UniPRot (Universal Protein Resource) (Apweiler et al., 2004), a catalog of protein information, a central repository of protein sequence and function created by joining Swiss-Prot, TrEMBL, and PIR analysis tools. Others tools are also available at ExPaSy: pattern and profile searches; topology prediction; primary, secondary, and tertiary structure analysis; sequence alignment; and others.

PSORT (Gardy et al., 2003) is a set of computational methods that make predictions for the protein sites in a cell, examining a given protein sequence for amino acid composition, similarity to proteins of known localization, presence of a signal peptide, transmembrane alpha-helices, and motifs corresponding to specific localizations. A version of PSORT-B for Linux platforms has been recently released. PSORT is recommended for bacterial/plant sequences, but the PSORT-B currently accepts only protein sequences from Gram-negative bacteria. The input file format is FASTA, and the output format can be selected by the user from among formats described in the PSORT documentation.

RNA Analysis

The RNAsoft suite of programs (Andronescu, Aguirre-Hernandez, Condon & Hoos, 2003) provides tools for predicting the secondary structure of a pair of DNA or RNA molecules, testing that combinatorial tag sets of DNA and RNA molecules have no unwanted secondary structure and designing RNA strands that fold to a given input secondary structure. The tools are based on standard thermodynamic models of RNA secondary structure formation and can be used for prediction as well as design of molecular structures. RNAsoft online access is freely available; however, some restrictions have been imposed on the size of the input data, in order to not overload the server limits.

The Vienna RNA package consists of a portable ISO C code library and several programs for the prediction and comparison of RNA secondary structures. This tool has a Web interface to the RNAfold program (Hofacker, 2003) and can predict secondary structures of single-stranded RNA or DNA sequences. The input file is a string of letters, consisting of A, U, G, and C. The result page shows the optimal structure, the energy of the optimal structure, and the ensemble free energy, if requested by user.

Others

A set of tools for clustering, analysis, and visualization of gene expression and other genomic data can be found in Expression Profiler—EP (Vilo, Kapushesky, Kemmeren, Sarkans & Brazma, 2003). Besides, EP allows searching gene ontology categories, generating sequence logos, extracting regulatory sequences, and studying protein interactions. It also links analysis results to external tools and databases.

ESTAnnotator (Hotz-Wagenblatt et al., 2003) is a tool for the throughput annotation of expressed sequence tags (ESTs) by automatically running a collection of bioinformatics methods. There are mainly four steps: a repeated quality check is performed, low-quality sequences are masked, successive steps of database searching and EST clustering are performed, already known transcripts that are present within mRNA and genomic DNA reference databases are identified, and finally, tools for the clustering of anonymous ESTs and for further database searches at the protein level are applied. The outputs are presented in a descriptive summary. ESTAnnotator was already successfully applied to the systematic identification and characterization of novel human genes involved in cartilage/bone formation, growth, differentiation, and homeostasis.

FUTURE TRENDS AND CONCLUSION

The WWW has a vast collection of bioinformatics tools that offer imminent possibilities for sharing, researching, and disseminating biological information. As presented, the bioinformatics research community has used Web-based application platforms intensively as the main suite for biological data analysis. Unfortunately, some application platforms use tools such as ClustalW, BLAST, or FASTA to analyze and search data from different databanks. This requires extra programs or software components for data format conversion of programs' output data. This not only complicates the software development process, but also sometimes distracts from the main research intention. For example, programs such as ClustalW output their results in a particular format, and this format cannot be easily parsed. A small change of the ClustalW source, such as an extra or missing field in the output, could break the program. The same problem happens with biological databases such as Swiss-Prot, PDB, and others.

In order to organize this area, *Nucleic Acids Research*, one of the most important magazines of this field, has devoted its first issue each year, over the last several years, to documenting the availability and features of the specialized databases.

There is also a XML (eXtensible Markup Language) framework (Shui, Wong, Graham, Lee & Church, 2003) proposal to integrate different biological databanks into a unified XML framework. The proposed framework has been implemented with the emphasis of reusing the existing bioinformatics data and tools.

As the Web continues to expand, newer tools will arise, and new challenges will be proposed to researchers. In fact, a safely prediction can be made: Web-based tools are invaluable ones for daily use of those working in this exciting area of bioinformatics, and this availability and use will continue to grow.

REFERENCES

Altschul, S.F. (1991). Amino acid substitutions matrices from an information theoretic perspective. *Journal of Molecular Biology, 219,* 555-665.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215,* 403-410.

Andronescu, M., Aguirre-Hernandez, R., Condon, A. &

Hoos, H.H. (2003). RNAsoft: A suite of RNA secondary structure prediction and design software tools. *Nucleic Acids Research*, *31*, 3416-3422.

Apostolico, A. & Giancarlo, R. (1998). Sequence alignment in molecular biology. *Journal of Comp. Biology*, 5, 173-196.

Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M.J., Natale, D.A., O'Donovan, C., Redaschi, N. & Yeh, L.S. (2004). UniProt: The Universal Protein Knowledgebase. *Nucleic Acids Research*, 32, D115-D119.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000). The Protein Data Bank. *Nucleic Acids Research*, ***28***, 235-242.

Crick, F.H. & Watson, J.D. (1953). Molecular structure of nucleic acids. *Nature*, *171*, 737-738.

Gardy, J.L., Spencer, C., Wang, K., Ester, M., Tusnády, G.E., Simon, I., Hua, S., deFays, K., Lambert, C., Nakai, K. & Brinkman, F.S. (2003). PSORT-B: Improving protein subcellular localization prediction for Gram-negative bacteria. *Nucleic Acids Research, 31*, 3613-3617.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D. & Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, *31*, 3784-3788.

Golovin, A., Oldfield, T.J., Tate, J.G., Velankar, S., Barton, G.J., Boutselakis, H., Dimitropoulos, D., Fillon, J., Hussain, A., Ionides, J.M.C. et al. (2004). E-MSD: An integrated resource for bioinformatics. *Nucleic Acids Research, 32*, D211-D214.

Henikoff, S. & Henikoff, J. (1992). Amino acid substitution matrices from protein blocks. *Proceedings of the National Acad. Science*, *89*, 10915-10919.

Higgins, D., Thompson, J., Gibson, T., Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and weight matrix choice. *Nucleic Acids Research*, *22*, 4673-4680.

Hofacker, I.L. (2003). Vienna RNA secondary structure server. *Nucleic Acids Research*, *31*, 3429 3431.

Holm, L. & Sander, C. (1994). The FSSP database of structurally aligned protein fold families. *Nucleic Acids Research, 22*, 3600-3609.

Hotz-Wagenblatt, A., Hankeln, T., Ernst, P., Glatting, K., Schmidt, E.R. & Suhai, S. (2003). ESTAnnotator: A tool for high throughput EST annotation**.** *Nucleic Acids Research*, *31*, 3716-3719.

Lehninger, A.L., Nelson, D.L. & COX, M.M. (2000). *Principles of biochemistry* (2nd edition). New York: Worth Publishers.

Neshich, G., Togawa, R.C., Mancini, A.L., Kuser, P.R., Yamagishi, M.E., Pappas, G., Torres, W.V., Fonseca, C.T., Ferreira, L.L., Luna, F.M. et al. (2003). STING Millennium: A Web-based suite of programs for comprehensive and simultaneous analysis of protein structure and sequence. *Nucleic Acids Research*, *31*, 3386-3392.

Notredame, C., Higgins, D. & Heringa, J. (2000). T-Coffee: A novel method for multiple sequence alignments. *Journal of Molecular Biology*, 302, 205-217.

Parry-Smith, D.J., Payne, A.W.R, Michie, A.D. & Attwood, T.K. (1997). Cinema—a novel Colour INteractive Editor for Multiple Alignments. *Gene*, *211*, GC45-56.

Pearson, W.R. (2000). Flexible sequence similarity searching with the FASTA3 program package. *Methods of Molecular Biology*, *132*, 185-219.

Sayle, R.A. & Milner-White, E.J. (1995). RasMol: Biomolecular graphics for all. *Trends in Biochemical Science*, 20, 374-376.

Scharf, M., Schneider, R., Casari, G., Bork, P., Valencia, A., Ouzounis, C. & Sander, C. (1994). GeneQuiz: A workbench for sequence analysis. *Proceedings of the 2nd Conference on Intelligent Systems for Molecular Biology* (pp. 348-353), Menlo Park, California, USA.

Schwede, T., Kopp, J., Guex, N. & Peitsch, M.C. (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Research, 31***,** 3381-3385.

Shui, W., Wong, R.K., Graham, S., Lee, L. & Church, B. (2003). Integrating, managing and analyzing protein structures with XML databases. *Proceedings of the 8th International Conference on Database Systems for Advanced Applications* (pp. 26-28), Kyoto, Japan.

Vilo, J., Kapushesky, M., Kemmeren, P., Sarkans, U. & Brazma, A. (2003). Expression profiler. In G. Parmigiani, E.S. Garrett, R. Irizarry & S.L. Zeger (Eds.), *The analysis of gene expression data: Methods and software*. New York: Springer-Verlag.

Zdobnov, E.M. & Apweiler, R. (2001). InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*, *17*, 847-848.

KEY TERMS

Alignment: Explicit mapping of characters of a sequence to characters of one or more other sequence(s).

Alpha-Helix: A helical conformation of a polypeptide chain, once of the most common secondary structure in proteins.

Base Pair: Two nucleotides in nucleic acid chains are paired by hydrogen bonding of their bases; for example A with T or U, and G with C.

DNA (deoxyribonucleic acid): A specific sequence of deoxyribonucleotide units covalently joined through phosphodiester bonds.

Domain: Combines several secondary structure elements and motifs; has a specific function.

Genome: The genetic information of an organism.

Homology: Relationship by evolutionary descent from a common ancestral precursor.

Motif: Combines a few secondary structure elements with a specific geometric arrangement.

Protein: A macromolecule composed of one or more polypeptide chains, each with a characteristic sequence of amino acids linked by peptide bonds.

RNA (ribonucleic acid): A specific sequence linked by successive phosphodiester bonds.

Similarity: Maximum degree of match between two aligned sequences as indicated by some (arbitrarily chosen) scoring function, for example, percent identity.

Web Usability

Shirley Ann Becker
Florida Institute of Technology, USA

INTRODUCTION

The study of computing technology and user interfaces was initiated during the 1970s when industrial research laboratories began to focus on human-computer interaction (HCI) (Badre, 2002). In the 1980s, the personal computer was introduced, thus expanding the need for designing effective user interfaces. HCI became a discipline during this time, and the Association for Computing Machinery (ACM) established the Special Interest Group in Computer Human Interaction. One of the first textbooks on HCI, *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (Schneiderman, 19891), was published. Shortly thereafter, HCI became part of the ACM curriculum promoting the development of effective user interfaces. Software tools were developed in order to assist in designing usable interfaces while employing usability engineering methods. Many of these methods focused on usability from the perspective of ease of use, ease of learning, user satisfaction, and zero defects (Nielsen, 1993).

The World Wide Web (Web) became an integral part of HCI research in the 1990s, as organizations rushed to deploy a corporate Web site. Many of these Web sites took advantage of cutting-edge technology, including graphics and animation, with little regard for the impact on the user. As a result, users became disgruntled by lengthy download times, complex navigation schemes, nonintuitive search mechanisms, and disorganized content.

While others were predicting a "Y2K meltdown," Jakob Nielsen (1999a) correctly predicted a "Web meltdown" due to the number of poorly designed Web sites that cluttered the Internet. Numerous studies showed that users were frustrated with glitzy Web sites that had too many usability barriers. A Forrester report estimated a 50% loss of potential online sales due to users not finding a product or service on the Web site (Manning, McCarthy & Souza, 1998). As importantly, 40% of users did not return to a site when their initial visit was a negative one.

Shortly after 2000, electronic commerce sites (dot coms) began to fail at an increasing rate. A Deloitte and Touche report found that many retailers had developed online sites to "test the waters" for consumer demand with no clearly articulated strategy for success (Speigel, 2000). The demise of many dot coms has been attributed to unfriendly user interfaces that negatively impacted the online experience.

BACKGROUND

Many researchers and practitioners alike have studied usability in order to develop Web sites that are navigable, consistent, appealing, clear, simple, and forgiving of user mistakes (Murray & Costanza, 1999). Existing user interface design recommendations were extended to include user interfaces for the Web (Lynch & Horton, 1999; Schneiderman, 1998). Those experienced in designing user interfaces provided heuristics and guidelines for designing Web pages, often by identifying design layout, navigation, and performance issues associated with particular Web sites (Flanders & Willis, 1998; Hurst, 1999; Spool, Scanlon, Schroeder, Snyder & DeAngelo, 1999). Jakob Nielsen, a well-known usability expert, provided much needed guidance on Web usability through featured online articles (www.useit.com/alertbox) and published guidelines (Nielsen, 1999b; Nielsen & Tahir, 2002).

Web usability has been defined as the measure of the quality of the user's online experience. There are several factors that are commonly used as a means of measuring this experience. These factors include (www.usability.gov):

- Learnability – A measure of the user's learning time for accomplishing basic tasks given that the user interface has not previously been used (or used infrequently).
- Efficiency – A measure of the user's time and error rate for task completion.
- Effectiveness – A measure of user productivity in performing a task.
- Satisfaction – A measure of the attitude, perceptions, and feelings about the site.
- Memorability – A measure of user recall such that a previously visited site can be used effectively with no new learning curve.

It is commonly accepted that the usability of a Web site is impacted by the user's online goal, the user's profile, and his or her computing environment. A user, for example, would have some tolerance for lengthy download times when searching for medical information with graphic illustrations. This tolerance level is greatly reduced when searching for information on the cost of an airline ticket. The user profile, including age, gender, income, education, computer skills, and other factors, influences the online experience. Web content written at a high reading grade level, for example, may be difficult to comprehend for users with low English proficiency. The use of color to convey meaning on a Web site may impede its use by those who have color-deficient sight. Small font size, patterned background images, and pastel colors may become Web barriers to older adults experiencing vision degradation due to aging (Morrell, 2002). The user's computing environment also has an impact on Web usability. Environmental factors, such as hardware, software, browsers, connectivity, and bandwidth, impede the use of a Web site when it is cluttered with graphics, animation, and other objects adding little value to the online experience.

Since 1998, much has been accomplished in promoting Web usability for persons with disabilities. Section 508 of the 1973 Rehabilitation Act was enacted to eliminate information technology barriers in order to provide those with disabilities equal access. The law applies to all federal agencies when they develop, procure, maintain, or use electronic and information technology (http://www.Section508.gov). As a result of this initiative, significant strides have been made to electronic government access by enforcing the Web content guidelines put forth by the World Wide Web Consortium. Though not mandated by law, many commercial and nonprofit Web sites have implemented Section 508 in order to provide access to a broad user base.

WEB USABILITY ASSESSMENT METHODS

There are several popular methods that have been employed to effectively study Web usability. The inquiry approach makes use of field observation, interviews, self-reporting logs and online sessions. The inspection approach utilizes heuristic evaluations, walkthroughs, and checklists. Usability testing may also be used in conjunction with the other approaches to gather feedback during and after Web site design (Hom, 1998).

- Field Observation – The user is observed while surfing a Web site in order to gather usability data in a real-world setting.
- Interviews, Surveys, and Questionnaires – The objective of these methods is typically to gather feedback about the user's perspective of usability. In terms of data gathering, the interview is a formal, structured process, whereas the survey is an informal, interactive process. Interviews and surveys may involve one or more users in a focus group setting. The questionnaire provides the means to obtain written responses regarding a user's online experience.
- Session and Self-Reporting Logs – The user records his or her actions and makes observations during an online session. Software is often used during a session to automatically record data about the user's online experience. The self-reporting log requires

Table 1. Web usability online resources

Resource	Description
http://www.usability.gov	National Cancer Institute summarizes research activities on Web usability. It also provides links to usability resources.
http://www.itl.nist.gov/iad/vvrg/index.html	National Institute of Standards and Technology provides resources and tools for usability testing.
http://www.useit.com	Jakob Nielsen and colleagues provide alert box articles, summaries of usability studies, and other usability resources.
http://www.acm.org/sigchi/	ACM Special Interest Group on Computer-Human Interaction provides a bibliography of usability research.
http://www.w3.org/WAI/	The World Wide Web consortium (W3C) Web initiative provides resources on making sites accessible to those with disabilities.
http://www.usabilitynews.org	The *Software Usability Research Laboratory (SURL)* specializes in software and Web site user interface design research, human-computer interaction research, and usability testing and research.

the user to manually record data while surfing the Web.

- Heuristic Evaluation – A usability expert (or group of experts) assesses a user interface to determine whether the Web design follows established usability practices (heuristics).
- Walkthrough – A usability expert (or group of experts) evaluates online experiences by constructing scenarios of Web use and then role-playing the targeted user.
- Usability Inspection – A usability expert (or group of experts) conducts usability inspections of a user interface in order to uncover usability problems in the design.
- Checklists – A usability expert (or group of experts) uses a checklist often in conjunction with an inspection to ensure that established usability practices are evaluated.
- Usability Testing – Experiments are conducted regarding usability aspects of a Web design. The objective is to gather data about an online experience in order to draw conclusions about the usability of the site. Though usability testing can involve sophisticated technology including usability labs, videotaping, and eye-tracking, this does not have to be the case (Murray & Costanzo, 1999). Often, usability test cases can be generated from an existing Web design without the use of sophisticated technology.

WEB RESOURCES

There are valuable online resources promoting usable Web designs. Many of these sites offer links for usability design and testing, good practices and lessons learned in the field. Some of the more popular sources for Web usability guidance are listed in Table 1. Though too numerous to list, there are many educational sites that offer resources on research activities and course materials. The Trace Research and Development Center at the University of Wisconsin-Madision, in particular, offers usability resources and links promoting universal usability (*http://trace.wisc.edu/world/web/*).

FUTURE TRENDS

Murray and Costanzo (1999) point out that from a HCI perspective, Web development is significantly different from software development. As such, there are challenges facing developers who are pursuing usable Web designs. These challenges include the following:

- The demographic diversity of online customers makes it difficult to develop user-friendly interfaces to meet all needs. For example, older adult users (60 years and older) may have trouble seeing Web content based on the use of color, font size, font type, and patterned background images (Becker, 2004; Morrell, Dailey, Feldman, Mayhorn & Echt, 2002). These design elements may have no usability impact on a younger adult for whom aging vision changes have not yet occurred. Web site images or textual references to religious holidays (e.g., Valentine's Day), as another example, may be offensive in certain global regions due to local religious or cultural beliefs (Becker, 2002).
- There is significant diversity among hardware, software, and network components being used to surf the Web. The usability of mobile technology, for example, must take into account the tiny screen in which Web content is displayed (Russell & Chaparro, 2002).
- Slower network access speeds impact usability due to performance degradation for a Web page with graphics and animation. Usability is also impacted by the browser version being used to surf the Web. Web content may display differently in older browser versions of Netscape© and Internet Explorer© than newer versions.
- The internationalization of many Web sites must account for culture, religion, and language in designing localized, user-friendly interfaces. Too often, organizations develop localized versions that do not meet the needs of regional customers (Marcus & Gould, 2000). The localized site may still have design aspects of the country of origin such as: English content, clichés, acronyms, and abbreviations (Becker & Mottay, 2001). Graphics may become a usability issue when cultural and religious beliefs are not taken into account during Web design (e.g., scantily clad figure on a homepage).
- Unlike software, users do not have a vested interest in a particular site. Often times, a user has purchased software and therefore is willing to accept usability barriers associated with it. Since there is no personal investment in a Web site, a user is more likely to leave and not return to a Web site that is perceived as unusable.

CONCLUSION

Web usability remains an important consideration in the design of effective user interfaces. There has been significant research on Web usability in terms of design layout, performance, navigation, and searches, among other areas. This initial work has been broadened to include usable Web designs for all users regardless of age, skills, education, culture, language, or religion. The Web accessibility initiative has promoted Web designs that take into account users with vision, physical, and cognitive disabilities. The internationalization of Web sites has promoted Web designs that meet the needs of a particular locale taking into account the customs, culture, religion, education, and other factors. Though much has been accomplished in developing usable Web sites, there is still work to be done. New technologies and expanded marketplaces pose unique challenges for universally usable Web designs.

REFERENCES

Badre. (2002). Shaping Web usability: Interaction design in context. *Ubiquity*. Retrieved October 1, 2003, from *http://www.acm.org/ubiquity/book/a_badre_1.html*

Becker, S.A. (2002). An exploratory study on Web usability and the internationalization of U.S. electronic businesses. *The Journal of Electronic Commerce Research, 3*(4), 265-278.

Becker, S.A. (2004). E-government visual accessibility for older adult users. Forthcoming in *Social Science Computer Review*.

Becker, S.A., & Mottay, F. (2001). A global perspective of Web usability for online business applications. *IEEE Software, 18*(1), 54-61.

Flanders, V., & Willis, M. (1998). *Web pages that suck*. San Francisco, CA: SYBEX.

Hom, J. (1998, June). The usability methods toolbox. Retrieved October 1, 2003, from *http://jthom.best.vwh.net/usability/usable.htm*

Hurst, M. (1999, September). Holiday '99 e-commerce. *Research Report*. Creative Good, Inc. Retrieved October 1, 2003, from *http://www.creativegood.com*

Lynch, P.L., & Horton, S. (1999). *Web style guide: Basic design principles for creating Web stes*. New Haven, CT: Yale University Press.

Manning, H., McCarthy, J.C., & Souza, R.K. (1998). *Why most Web sites fail*. *Interactive Technology Series, 3*(7). Forrester Research.

Marcus, A., & Gould, W.E. (2000). Crosscurrents: Cultural dimensions and global Web user-interface design. *ACM Interactions, 7*(4), 32-46.

Morrell, R.W. (Ed.). (2002). *Older adults, health information, and the World Wide Web*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Morrell, R.W., Dailey, S.R., Feldman, C., Mayhorn, C.B., & Echt, K.V. (2002). *Older adults and information technology: A compendium of scientific research and Web site accessibility guidelines*. Bethesda, MD: National Institute on Aging.

Murray, G., & Costanzo, T. (1999). Usability and the Web: An overview. *Network Notes*, *61*. Information Technology Services, National Library of Canada. Retrieved October 1, 2003, from *http://www.nlc-bnc.ca/9/1/p1-260-e.html*

Nielsen, J. (1993). *Usability engineering*. Cambridge, MA: Academic Press.

Nielsen, J. (1999a). User interface directions for the Web. *Communications of the ACM, 42*(1), 65-72.

Nielsen, J. (1999b). *Designing Web usability: The art of simplicity*. Indianapolis, IN: New Riders Publishing.

Nielsen, J., & Tahir, M. (2002). *Homepage usability 50 Websites deconstructed*. Indianapolis, IN: New Riders Publishing.

Russell, M.C., & Chaparro, B.S. (2002). Reading from a Palm Pilot ™ using RSVP. *Proceedings of the Human Factors and Ergonomic Society 46th Annual Meeting* (pp. 685-689).

Schneiderman, B. (1989). *Designing the user interface: Strategies for effective human-computer interaction*. Boston, MA: Addison-Wesley.

Schneiderman, B. (1998). *Designing the user interface: Strategies for effective human-computer interaction* (3rd ed.). Boston, MA: Addison-Wesley.

Schneiderman, B. (2000). Universal usability. *Communications of the ACM, 43*(5), 85-91.

Schneiderman, B., & Plaisant, C. (2004). *Designing the user interface: Strategies for effective human-computer interaction* (4th ed.). Boston, MA: Addison-Wesley.

Spiegel, R. (2000, January). Report: 70 percent of retailers lack e-commerce strategy. *Ecommerce Times*. Retrieved October 1, 2003, from *http://www.ecommercetimes.com/news/articles2000/000126-1.shtml*

Spool, J., Scanlon, T., Schroeder, W., Snyder, C., & DeAngelo, T. (1999). *Web site usability: A designer's guide*. San Francisco, CA: Morgan Kaufman.

KEY TERMS

Dot Com: A Web site that is intended for business use, though the term is commonly used to represent any kind of Web site. The term evolved from the "com" part of a Web site's address, which represents commercial sites. It came to be associated with Internet companies that failed during the mid 2000s (*www.searchWebservices.com*).

Internationalization: It is the process of making a Web site interoperable in a specific market or locale. In general, interoperability means that the functionality of the site is not dependent on a specific language or culture and is readily adaptable to others.

Localization: It is the process of adapting an internationalized Web site to meet language, culture, religion, and other requirements of a specific market or locale.

Universal Usability: Universal usability can be defined as having more than 90% of all households as successful users of information and communications services at least once a week (Schneiderman, 2000, p. 85).

Usability: The ISO 9241-11 standard states that usability is the "effectiveness, efficiency and satisfaction with which a specified set of users can achieve a specified set of tasks in a particular environment".

Usability Engineering: It is a systematic approach to making software (Web designs) easy to use, thus meet the needs of the targeted users.

Web Accessibility: Web accessibility means that any person, regardless of disabilities, is able to use Web technology without encountering any barriers.

ENDNOTES

[1] The textbook is now in its 4th edition taking into account human factors associated with interactive systems (Schneiderman & Plaisant, 2004).

Web-Based Distance Learning and the Second Digital Divide

Sheryl Burgstahler
University of Washington, USA

INTRODUCTION

In no field have we witnessed a greater impact of emerging technologies than in that of distance learning. Correspondence courses using printed material and postal mail have been replaced by Web-based courses with the potential to make learning available to anyone, anywhere at anytime. This potential cannot be realized, however, unless two digital divides are eliminated. Some people are on the wrong side of the first "digital divide" between the technology "haves" and the technology "have-nots". The benefits of technology are less available to those who are poor, who live in rural areas, who are members of minority racial or ethnic groups, and/or who have disabilities (Kaye, 2000; U.S. Department of Commerce, 1999). Lack of access to new technologies limits their options for taking and teaching technology-based courses. This is true for individuals with disabilities, even though the rapid development of assistive technology makes it possible for an individual with almost any type of disability to operate a computer (2003 Closing the Gap Resource Directory, 2003). Unfortunately, many people with disabilities still do not have access to these empowering tools, putting them on the "have not" side of the first digital divide.

Within the group of "haves" with respect to the first digital divide, however, many people with disabilities face a "second digital divide." This line separates people who can make full use of the technological tools, services, and information to which they have access, from those who cannot. Too often people with disabilities lucky enough to be on the right side of the first digital divide, find themselves on the wrong side of this second digital divide (Waddell, 1999). For example, a person who is blind may use a text-to-speech system that reads aloud text that appears on the screen. Because it cannot interpret graphics, it will simply say "image map" at a place where an image map would be displayed to someone using the full features of a multimedia Web browser. It cannot read aloud information within this and other graphic images. This person cannot access the content presented unless this content is provided in a text-based form.

BACKGROUND

Section 504 of the Rehabilitation Act of 1973 mandated that qualified people with disabilities be provided with access to programs and services that receive federal funds. The Americans with Disabilities Act (ADA) of 1990 reinforced and extended Section 504, requiring that people with disabilities have access to public programs and services, regardless of whether or not they are federally funded. According to these laws, no otherwise qualified individuals with disabilities shall, solely by reason of their disabilities, be excluded from the participation in, be denied the benefits of, or be subjected to discrimination in these programs and services, unless it would pose an undue burden to do so. A United States Department of Justice ruling (ADA Accessibility, 1996) clarified that ADA accessibility requirements apply to programs offered on the Internet by stating, "Covered entities that use the Internet for communications regarding their programs, goods, or services must be prepared to offer those communications through accessible means as well." Clearly, if qualified individuals with disabilities enroll in distance learning courses or are qualified to teach them, these opportunities should be made accessible to them. However, the inaccessible design of most Web-based distance learning courses imposes barriers to people with some types of disabilities (Schmetzke, 2001).

UNIVERSAL DESIGN

If an applicant who is blind is the best candidate to teach a Web-based course which has been developed without text alternatives for critical content displayed using graphics, the course will need to be modified in order for him to teach it. If planning for access was done as the course was being developed, this costly redesign would not be necessary. Simple design decisions could have been made to assure accessibility to potential students and instructors with a wide range of abilities and disabilities. This proactive process is called "universal design". Universal design is defined as "the design of products and environments to be usable by all people, to the greatest extent possible, without the need for adaptation or specialized design" (National Center for Universal Design, 2003, p.1). Applying universal design principles makes products and environments usable by people with a wide variety of characteristics, including gender, height, age, ethnicity, primary language, and level of ability to see, hear, speak, and move.

The concept of universal design was first applied to architecture. It has since been applied to the design of household appliances, instructional learning environments, Web sites and other products and environments (Bar & Galluzzo, 1999; Bowe, 2000; Burgstahler, 2001). When the wide range of characteristics of potential students and instructors is considered, distance learning course designers can create learning environments that are accessible to all participants, just as sidewalks with curbcuts are used by everyone, including those who push delivery carts, baby strollers, and wheelchairs.

For many years, examples of isolated distance learning courses designed to be accessible to individuals with disabilities could be found, including a course co-taught by the author of this article and a professor who is blind (Burgstahler, 2000). However, few distance learning programs have policies and guidelines that specifically address the accessibility of distance learning tools and resources (Burgstahler, 2000; Kessler & Keefe, 1999; Schmetzke, 2001). Comprehensive policies, such as the mandate that distance learning options offered by California Community Colleges must afford students with disabilities maximum access (Distance education: Access guidelines for students with disabilities, 1999), are rare.

EXAMPLES OF ACCESSIBLE DESIGN FEATURES

To create Web pages that are accessible to everyone, developers must either avoid certain types of inaccessible features or formats or create alternative methods for navigating or accessing content provided through inaccessible features or formats (Thompson, Burgstahler, & Comden, 2003). For example, including <alt> attributes with descriptive text makes graphic image content accessible to individuals who are blind. Developers should also assure that all functions at a Web site can be accessed using a keyboard alone, so that those who cannot manipulate a mouse can navigate the pages using the keyboard or a keyboard alternative. Another useful feature is to add a "Skip Navigation" link to the top of each page; otherwise, most speech-to-text systems for individuals who are blind will read through all of the navigation links on a page before reading the content in the body of the page.

Students and instructors who have limited vision may use software that enlarges screen images, but allows them to view only a small portion of the content of a standard screen image at one time. Page layouts that are uncluttered and consistent from page to page can facilitate locating and understanding Web content for people with low vision, as well as for those with some types of learning disabilities. Assuring that content and navigation do not require that a viewer distinguish one color from another makes Web-based distance learning accessible to those who are colorblind.

Internet resources that do not require the ability to hear are accessible to people who are deaf or hard of hearing. However, when Web sites include audio output without providing text captioning or transcription, they cannot access the content. Similarly, distance learning programs should provide audio-descriptions (i.e., aural descriptions) of visual content or text-based descriptions for those who are blind.

Some distance learning programs employ real-time "chat" communication in their courses. In this case, students communicate synchronously (at the same time). Synchronous communication is difficult or impossible to use by someone whose input method is slow. For example, a person with limited hand use who can only type characters slowly or someone with a learning disability who takes a long time to compose his thoughts may not be fully included in the discussion. In contrast, with a synchronous tool such as electronic mail, all students and instructors can fully participate. In addition, since flickers at certain rates (often between 2 to 55 hertz) can induce seizures for people who are susceptible to them, they should be avoided.

Tools, Guidelines, and Standards for Accessibility

The most current version of HTML (Hypertext Markup Language) makes it relatively easy to develop accessible Web sites. Commonly used development tools such as WebCT™(n.d.) and Blackboard™ (n.d.) include accessibility tools as well. Electronic tools that can test Web resources for some accessibility features and training courses and reference materials to help distance learning designers develop skills for making distance learning programs accessible are also widely available (Disabilities, Opportunities, Internetworking, and Technology, n.d.).

Technical guidelines and standards have been developed to provide guidance to organizations that wish to make Web content accessible to students with disabilities. The most widely used are those created by the World Wide Web Consortium and the U.S. federal government.

Table 1. Quick tips to make accessible Web sites

For Complete Guidelines & Checklist: www.w3.org/WAI

- **Images & animations**: Use the alt attribute to describe the function of each visual.
- **Image maps.** Use the client-side map and text for hotspots.
- Multimedia. Provide captioning and transcripts of audio, and descriptions of video.
- **Hypertext links.** Use text that makes sense when read out of context. For example, avoid "click here".
- **Page organization.** Use headings, lists, and consistent structure. Use CSS for layout and style where possible.
- **Graphs & charts.** Summarize or use the longdesc attribute.
- **Scripts, applets, & plug-ins.** Provide alternative content in case active features are inaccessible or unsupported.
- **Frames.** Use the noframes element and meaningful titles.
- **Tables.** Make line-by-line reading sensible. Summarize.
- Check your work. Validate. Use tools, checklist, and guidelines at *http://www.w3.org/TR/WCAG*

Table 2. Steps to creating accessible distance learning programs

- Make sure that all stakeholders, including potential students and instructors with disabilities, are represented as accessibility policies, procedures, and guidelines are being developed.
- Review policies and guidelines that have been created by other organizations, such as the California Community Colleges.
- Develop a policy statement that commits the organization to making programs, services, and resources accessible to people with disabilities.
- Articulate access challenges that may face potential participants with disabilities in the context of the programs, services, and/or resources offered and the tools used for their delivery.
- Consult with legal experts to fully understand the requirements for program, information, and service accessibility mandated by the ADA and other legislation relevant to your organization.
- Develop guidelines for all media, tools and strategies used in the distance learning courses; consider Section 508 standards as a model as appropriate.
- Assign a person or a department within the organization to be responsible for updating disability-related program access policies and guidelines and assuring compliance throughout the organization.
- Disseminate accessibility policy, guidelines, and procedures throughout the organization.
- Provide regular training and support regarding accessibility issues.
- Consider developing a plan to phase in compliance with program accessibility guidelines for previously developed courses, with a date at which all programs will be compliant.
- Regularly evaluate progress toward accessibility.
- Besides taking proactive steps to assure accessibility, develop procedures for responding quickly to requests for disability-related accommodations.

The Web Accessibility Initiative (WAI) of the World Wide Web Consortium developed Web Content Accessibility Guidelines (1999) for designing Web pages that are accessible to people with disabilities. Besides providing comprehensive guidelines, the WAI provides the quick tips for making accessible Web pages listed in Table 1 (World Wide Web Consortium Web Accessibility Initiative, 2001).

Section 508, which was added in 1986 to the Rehabilitation Act of 1973 (Architectural and Transportation Barriers Compliance Board, 2000), requires that electronic and information technologies that federal agencies procure, develop, maintain, and use are accessible to people with disabilities, both employees and members of the public, unless it would pose an undue burden to do so. As mandated in Section 508, the U.S. Architectural and Transportation Barriers Compliance Board (Access Board) developed accessibility standards to which federal agencies must comply (Electronic and Information Technology Accessibility Standards, 2000). Although most distance learning programs are not covered entities under this legislation, they can use the Section 508 standards as guidelines for designing accessible courses. These programs can also benefit from following the leadership of the federal government in being pro-active with respect to the accessibility of information technology (IT). "Use of an 'ad hoc' or 'as needed' approach to IT accessibility will result in barriers for persons with disabilities. A much better approach is to integrate accessibility reviews into the earliest stages of design, development, and procurement of IT." (U.S. Department of Justice, 2002)

FUTURE TRENDS AND STEPS TO AN ACCESSIBLE DISTANCE LEARNING PROGRAM

It is unlikely that distance learning courses in the future will be universally designed unless relevant policies, guidelines, and procedures are in place within distance learning programs. Organizations can begin the process of developing accessibility policies, procedures, and guidelines by addressing issues listed in Table 2, as published in *Educational Technology Review* (Burgstahler, 2002).

CONCLUSION

Well-designed distance learning courses create learning opportunities for everyone and thus do not erect barriers for potential students and instructors with disabilities. Employing universal design principles as distance learning courses are created can make learning opportunities accessible to everyone, everywhere, at any time and thereby eliminate the second digital divide.

ACKNOWLEDGEMENT

This article is based upon work supported by the National Science Foundation (Cooperative agreement # HRD-0227995) and the U.S. Department of Education, Office of Postsecondary Education (grant #P33A020044). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of federal government.

REFERENCES

2003 Closing the Gap Resource Directory. (2003). *Closing the Gap*, *21*(6), 37-195.

ADA accessibility requirements apply to Internet Web pages (1996). *The Law Reporter, 10*(6), 1053-1084.

Americans with Disabilities Act of 1990, 42 U.S.C.A. § 12101 (1990).

Architectural and Transportation Barriers Compliance Board. (2000). Electronic and information technology accessibility standards, *Federal Register*, 36 CFR Part 1194 December 21. Retrieved March 7, 2004, from *http://www.access-board.gov/sec508/508standards.htm*

Bar, L., & Galluzzo, J. (1999). *The accessible school: Universal design for educational settings*. Berkeley, CA: MIG Communications.

Blackboard, Inc. (n.d.). Blackboard accessibility. Retrieved March 7, 2004, from *http://www.blackboard.com/products/access/index.htm*

Bowe, F.G. (2000). *Universal design in education*. Westport, CT: Bergin & Garvey.

Burgstahler, S. (2000). Access to Internet-based instruction for people with disabilities. In L.A. Petrides (Ed.), *Case studies on information technology in higher education* (pp.76-88). Hershey, PA: Idea Group Publishing.

Burgstahler, S. (2001). *Universal design of instruction*. Seattle: DO-IT, University of Washington. Retrieved March 7, 2004, from *http://www.washington.edu/doit/Brochures/Academics/instruction.html*

Burgstahler, S. (2002). Distance learning: Universal design, universal access. *Educational Technology Review*, *10*(1).

Disabilities, Opportunities, Internetworking and Technology. (n.d.). *Technology and universal design*. Retrieved March 7, 2004, from *http://www.washington.edu/doit/Resources/technology.html*

Distance education: Access guidelines for students with disabilities. (1999). California Community Colleges Chancellor's Office. Retrieved March 7, 2004, from *http://www.htctu.fhda.edu/publications/guidelines/distance_ed/disted.htm*

Electronic and Information Technology Accessibility Standards. (December 21, 2000). *The Federal Register, 65*(246), 80499-80528.

Kaye, H.S. (2000). *Disability and the digital divide*. San Francisco: University of California Disability Statistics Center. Retrieved March 7, 2004, from *http://dsc.ucsf.edu/publication.php?pub_id=6*

Kessler, D., & Keefe, B. (1999). Going the distance. *American School and University*, *7*(11), 44-46.

National Center for Universal Design. (2003) What is universal design? Retrieved March 7, 2004, from *http://www.design.ncsu.edu/cud/univ_design/ud.htm*

Schmetzke, A. (2001). Online distance education – "Anytime, anywhere" but not for everyone. *Information Technology and Disability Journal*, *7*(2). Retrieved March 7, 2004, from *http://www.rit.edu/~easi/itd/itdv07n2/axel.htm*

Section 504 of the Rehabilitation Act of 1973. 29 O.S.C. § 794(a) (1973).

Section 508 of the Rehabilitation Act of 1973. 29 U.S.C. § 794(d) (1998).

Technology-Related Assistance of Individuals with Disabilities Act of 1988, 29 U.S.C. 2201 et seq.

Thompson, T., Burgstahler, S., & Comden, D. (2003). Research on Web accessibility in higher education. *Journal of Information Technology and Disabilities, 9*(2).

U.S. Department of Commerce, National Telecommunications and Information Administration. (1999*)*. Falling through the net: Defining the digital divide. Washington. D.C. Retrieved March 7, 2004, from *http://www.ntia.doc.gov/ntiahome/fttn99/*

U.S. Department of Justice (2002). Information technology and people with disabilities: The current state of federal accessibility, Section II, Introduction. Retrieved March 7, 2004, from *http://www.usdoj.gov/crt/508/report/content.htm*

Waddell, C.D. (1999). The growing digital divide in access for people with disabilities: Overcoming barriers to participation in the digital economy. *Understanding the Digital Economy Conference*, May. Retrieved March 7, 2004, from *http://www.icdri.org/CynthiaW/the_digital_divide.htm*

Waddell, C.D., & Urban, M.D. (2001). An overview of law and policy for IT accessibility: A resource for state and municipal IT policy makers. International Center for Disability Resources on the Internet. Retrieved March 7, 2004, from *http://www.icdri.org/CynthiaW/SL508overview.html*

WebCT, Inc. (n.d.). WebCT Standards. Retrieved March 7, 2004, from http://www.webct.com/accessibility/home

World Wide Web Consortium (1999). Web content accessibility guidelines. Retrieved March 7, 2004, from *http://www.w3.org/tr/wai-webcontent*

World Wide Web Consortium Web Accessibility Initiative (2001). Quick tips to make accessible Web sites. Retrieved March 7, 2004, from *http://www.w3.org/WAI/References/QuickTips/*

KEY TERMS

Accessible: A product, information, or environment that is fully usable by a person, with or without assistive technology.

Assistive Technology: "Any item, piece of equipment, or system, whether acquired commercially, modified, or customized, that is commonly used to increase, maintain, or improve functional capabilities of individuals with disabilities." (Technology-Related Assistance, 1988). Examples of assistive technology include wheelchairs, hand controls for automobiles, prostheses, communication aids, hand splints, hearing aids, and alternatives to computer keyboards (Technology-Related Assistance).

Electronic Technology: Encompasses information technology, but also includes any equipment or interconnected system or subsystem of equipment, that is used in the creation, conversion, or duplication of data or information. Electronic technology includes telecommunications products such as telephones and office equipment such as fax machines.

Hypertext Markup Language (HTML): A language used to organize and present content on Web pages. HTML uses tags such as <h1> and </h1> to structure text into headings, paragraphs, lists, hypertext links, and so forth.

Information Technology: "Any equipment or interconnected system or subsystem of equipment, that is used in the automatic acquisition, storage, manipulation, management, movement, control, display, switching, interchange, transmission, or reception of data or information." Information technology includes "computers, ancillary equipment, software, firmware and similar procedures" (Electronic and Information Technology Accessibility Standards, 2000, p.80499).

Person with a Disability: Any "person who (a) has a physical or mental impairment that substantially limits one or more major life activities, (b) has record of such an impairment, or (c) is regarded as having such an impairment. Major life activities include walking, seeing, hearing, speaking, breathing, learning, working, caring for oneself, and performing manual tasks" (Americans with Disabilities Act of 1990).

Universal Design: "The design of products and environments to be usable by all people, to the greatest extent possible, without the need for adaptation or specialized design" (National Center for Universal Design).

Web-Based Supply Chain Strategy

Hanns-Christian L. Hanebeck
GlobeRanger Corporation and University of Dallas, USA

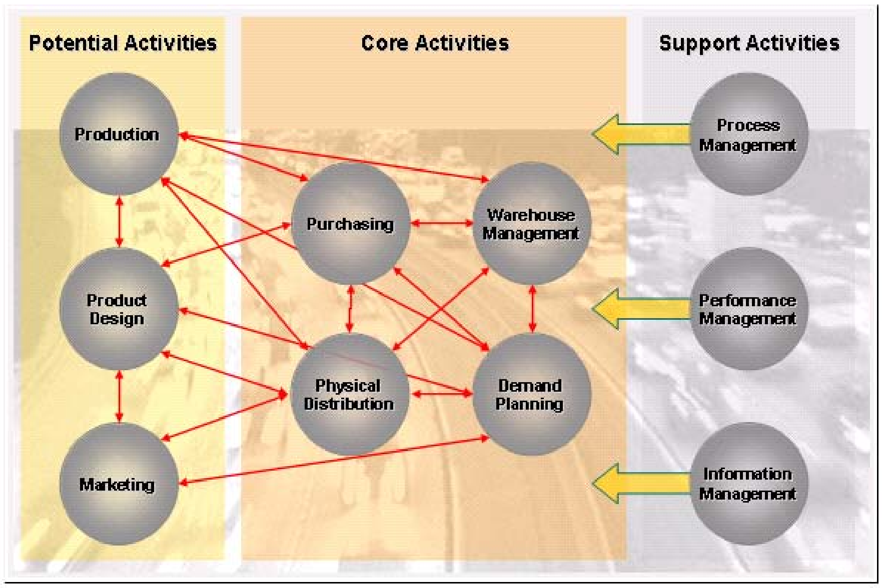
INTRODUCTION: SUPPLY CHAIN MANAGEMENT

Supply chain management is a fairly new creature and one that has evolved out of a variety of different enterprise functions over the past two decades. Traditionally, the majority of activities within supply chain management were related to the physical transport of goods from one point to another. Today, we understand supply chain management to include a broader range of activities such as demand planning, purchasing, warehousing, and transportation at the very least. Many authors will also throw production, product, and package design as well as marketing into the definition (Dornier et al., 1998; Schary & Skjøtt-Larsen, 2001; Taylor, 1997; Gourdin, 2001). For the context of this article, we refer to supply chain management as activities that are related to the planning and fulfillment of market demand. All of the activities within supply chain management can be performed by one or more than one legal entity. We further understand supply chain management as a business process rather than a function or department within a given company. Figure 1 below illustrates the set of core and potential activities as well as supporting ones such as process, performance, and information management.

BACKGROUND: CORPORATE STRATEGY

Strategy refers to the competitive positioning of a company or product in a given market. According to Michael E. Porter, the essence of strategy lies in the activities a company performs – either choosing to perform the same activities differently from rivals or choosing to perform different activities than its rivals (Porter, 1996). These activities can be viewed as distinct activities in some cases or as business processes in others. Ultimately, the execution of activities leads to the occupation of a unique and defensible position from which the firm can outperform its competitors. The degree of defensibility depends on what Porter calls trade-offs: well-defined synergies that tie the activities of a company to one another and create a network for value creation that is harder to imitate than single activities would be. This view of strategy closely ties into the approach of strategic capabilities

Figure 1. Strategic supply chain activities



(Stalk et al., 1992; Hamel & Prahalad, 1990) and also into the more recent approach of the resource-based view (Collis & Montgomery, 1995; Kaplan et al., 2001).

FUTURE TRENDS: SUPPLY CHAIN STRATEGY

Before we venture into the definition of supply chain strategy, it is important to clarify the relationship between strategy and activities a bit further. In its simplest form, strategy is the definition of a desired future state of the firm or business unit or product. Activities are either stand-alone tasks or they are chains of tasks, which are *de facto* business processes. In this sense, strategy drives the definition of processes and processes in turn rely on technologies as much as on other resources in their execution. The explicit dependencies between strategy, process, and technology are illustrated by the framework as shown in Figure 2.

Returning to supply chain management, we can now define supply chain strategy. If corporate strategy is the selection and execution of specific activities in a given firm, then supply chain strategy necessarily is the choice and performance of unique activities within supply and demand management to ultimately achieve higher corporate goals. Further, it is important to differentiate between firms that are primarily driven by supply chain management activities (strategic role) and those where the supply chain is just an enabling function (support role).

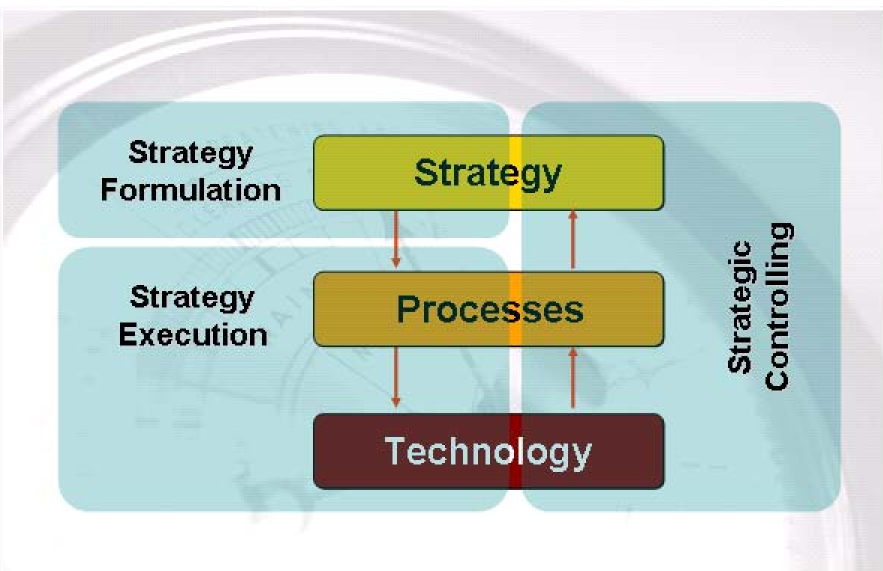
In its support role, supply chain management translates corporate strategy into strategic supply chain objectives. These strategic objectives then drive the definition of processes through the choice of activities. Processes, in turn, are supported by supply chain technologies. As a result, supply chain processes and technologies subsequently become a directly dependent function of corporate strategy. Supply chain managers in this situation have to perform three strategy-relevant tasks:

(1) Definition of supply chain activities in close reliance on the essence and intent of corporate strategy.
(2) Creation (as opposed to depletion) of synergy between all supply chain activities as well as towards all other activities of the firm.
(3) Continuous monitoring of strategy achievement throughout the execution of supply chain activities.

Seen in this light, supply chain management actively contributes to corporate strategy achievement and is as important a corporate function as marketing, finance, or human resources. This view of supply chain as a support function has gained in importance over the past few years and will continue to do so.

In the strategic role, supply chain management drives the definition of corporate strategy and thus will translate its own strategic requirements into business processes and technologies. Supply chain management in the strategic role takes on a life of its own in that the dominant

Figure 2. Strategy framework (Hanebeck 2001)



ingredients of corporate strategy are all ultimately tasks of supply chain management. The core competency of a company often times is based on supply chain management. The relation between supply chain strategy, processes, and technologies is much more immediate, yet the general relationships remain the same as in the support function.

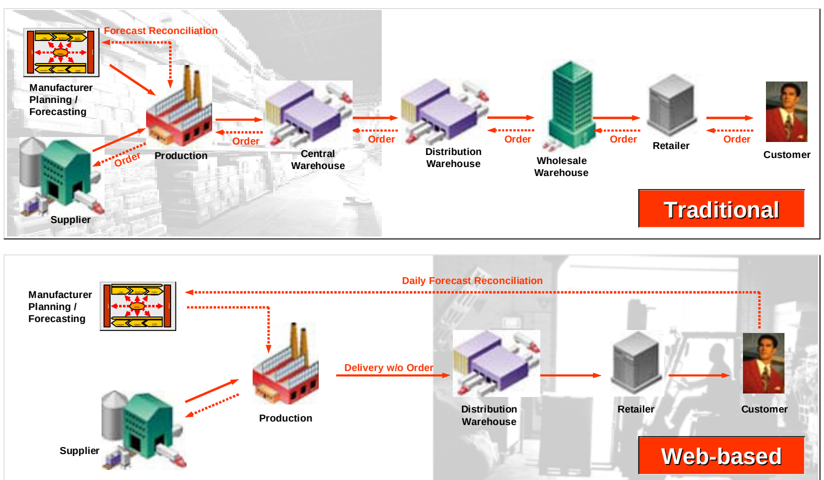
CONCLUSION: WEB-BASED SUPPLY CHAIN STRATEGY

As we had initially stated, supply chain management can be performed by one or more than one legal entity. The latter circumstance is characterized by several firms all executing activities within a single, logically coherent supply chain. This situation poses very specific additional challenges that a single firm would not encounter. On one hand, the coordination of supply chain activities now resides with several firms and thus in the hands of several potentially independent actors. On the other hand, supply chain information has to transcend the traditional boundaries of enterprise systems such as enterprise resource planning (ERP), supply chain planning (SCP), or supply chain execution (SCE) systems. These systems were developed in the 70s to early 90s. As a consequence, they were never intended to enable the close sharing of information across enterprise boundaries and beyond the communities of immediate users. Traditional enterprise systems are thus too inflexible and cumbersome to use when it comes to information sharing between various partners. Yet, this sharing of information is exactly what new theoretical concepts such as supply chain collaboration, event management, or visibility require. In fact, these concepts necessitate a shared platform that allows the exchange and filtering of supply chain relevant data and information in order for it to be used by many diverse and highly involved firms. To make things worse, turning enterprise-level systems inside-out to cater to data and information sharing does not seem very feasible due to their high degree of internal complexity.

A possible way around the re-design of enterprise-level systems is the sharing of supply chain data on a common platform such as a Web-accessible database that allows data and reporting structures to be reconfigured as new needs arise. Such a database can be resident in any of the partners' locations or elsewhere and can provide instrumental support in the coordination of partners as they thrive to accomplish a common goal. Based on our definition of supply chain strategy as the choice and performance of supply chain activities, the partially Web-based execution of supply chain processes constitutes an integral part of the technology layer within the strategy framework illustrated above.

How does this work in a concrete example? Figure 3 illustrates the differences between traditional and Web-based supply chains. What becomes readily apparent is that traditional supply chains rely on slow, error-prone, node-to-node communication that gradually ripples through the system before new information from the point of sale (POS) affects supplier and manufacturing nodes at the opposite end. The occurrence of the bullwhip effect is a good example of the inherent inefficiencies in traditional supply chains (Whang, 1997). Web-based supply chains on the other hand are characterized by a much higher degree of information sharing and also by much faster communication. Naturally, the availability of Web-based technological capabilities as described above allows for a much higher degree of operational effectiveness and strategic efficiency.

Figure 3. Traditional versus Web-based supply chains (Hanebeck, 2001)



The transformation from a traditional to a Web-based supply chain usually takes time and is not easy to accomplish. A key issue for Web-based supply chains lies in the establishment of trust between parties that oftentimes have conflicting goals and need to negotiate their fair share of revenues from a shared financial stream. This latter problem is virtually always a critical factor in supply chain efficiency in that manufacturers, distributors, and wholesalers usually try to reverse-engineer retail revenues and profits based on quantities sold in specific locations. This naturally leads to a lack of trust that can only be overcome by gradual and iterative negotiation. The emergence and proliferation of Web-based supply chain technologies will ultimately contribute to the building of trust as these technologies provide a factual basis for information and pose a strong competitive impetus to grow more efficient over time.

It is interesting to see that the concept of Web-based supply chains resembles the organizational structure of virtual corporations in many ways. Virtual corporations are temporary networks of independent companies - suppliers, customers, even erstwhile rivals - linked by information technology to share skills, costs, and access to one another's markets (Byrne et al., 1993). Virtual corporations are increasingly becoming an acceptable organizational alternative for supply chain managers as they already display many of the same characteristics that their logistical operations have. In virtual organizations, every member focuses on their own core competency (Hamel & Prahalad, 1990) and strives to be the best available provider of that competency. This is true for Web-based supply chains as well. As the above definition illustrates, all members are linked by information technologies in order to communicate and transact. Virtual corporations certainly cannot live up to their often lofty expectations without the same high flexibility and speed that Web-based supply chains offer.

In summary, Web-based supply chains are a very elegant and promising concept in theory that has a lot of potential in practice as we venture into the 21st Century. However, as this article has outlined, every successful endeavor independent of its nature begins with a sound strategy that translates into innovative supply chain processes, which in turn utilize technologies such as Web- and Internet-based systems. Technology in and by itself is of little to no value. The same is true for Web-based supply chains: they too need synergistic and integrated strategies as a foundation on which to drive projects towards implementations that are successful in realizing tangible business benefits. These benefits will drive the company's competitiveness and comparative advantage that ultimately compels others to follow suit until potentially every supply chain, someday, could be a Web-based supply chain.

REFERENCES

Byrne, J.A., Brand, R., & Port, O. (1993, February 8). The virtual corporation. *Business Week*.

Collis, D. J., & Montgomery, C. A. (1995, July-August). Competing on resources: Strategy in the 1990s. *Harvard Business Review*, 118-128.

Dornier et al. (1998). *Global operations and logistics.* New York: John Wiley & Sons.

Gourdin, K. N. (2001). *Global logistics management: A competitive advantage for the new millennium*. Oxford: Blackwell Publishers.

Hamel, G., & Prahalad, C. (1990). Core competencies of the corporation. *Harvard Business Review, 68*(May-June), 79-91.

Hanebeck, H.C.L. (2001). Lecture on global supply chain strategy. University of Dallas, Graduate School of Management, Fall 2001.

Hoffmann, W., Hanebeck, H.C., & Scheer, A.W. (1996). Kooperationsbörse - der Weg zum virtuellen Unternehmen. *Management & Computer*, Gabler Verlag.

Kaplan, S., Schenkel, A., von Krogh, G., & Weber, C. (2001). Knowledge-based theories of the firm in strategic management: A review and extension. Submission to the *Academy of Management Review*.

Porter, M.E. (1996, November-December). What is strategy? *Harvard Business Review, 79*, 61-78.

Schary, P.B., & Skjøtt-Larsen, T. (2001). *Managing the global supply chain (*2nd ed.). Copenhagen: Copenhagen Business School Press.

Stalk, G., Evans, P., & Shulman, L.E. (1992, March-April). Competing on capabilities: The new rules of corporate strategy. *Harvard Business Review*, 57-68.

Taylor, D. (1997). *Global cases in logistics and supply chain management*. London: International Thompson Business Press.

Waters, D. (1999). *Global logistics and distribution planning.* London: Kogan Page.

Whang, S. (1997, Spring). The bullwhip effect in supply chains. *Sloan Management Review*, 93-102.

KEY TERMS

Activities: Single tasks or chains of tasks that form business processes and that allow a firm to differentiate itself in the market place.

Bullwhip Effect: A situation in which ineffective network effects occur because each successive node in the supply chain orders more supplies than the previous one based on wrong assumptions, a lack of communication, and flawed planning processes.

Strategic (Supply Chain) Role: A situation in which supply chain management is the most important corporate function and thus dominates the definition of corporate strategy.

Strategy: The competitive positioning of a company or product in a given market. In its essence strategy means choosing to perform the same activities differently or different activities from rival firms.

Strategy Framework: Strategy drives the definition of processes, which in turn rely on technologies for their flawless execution.

Supporting (Supply Chain) Role: A situation in which supply chain management is driven by other, more important corporate functions and subsequently translates corporate strategy into specific requirements for supply chain management.

Supply Chain Management: All activities that are related to the planning and fulfillment of demand.

Supply Chain Strategy: The choice and performance of unique activities within supply and demand management to ultimately achieve corporate strategy.

Virtual Corporation: A temporary network of independent companies - suppliers, customers, even erstwhile rivals - linked by information technology to share skills, costs, and access to one another's markets (Byrne et al., 1993).

Wireless Ad Hoc Networking

Fazli Erbas
University of Hanover, Germany

INTRODUCTION

Mobile ad hoc networks represent a new form of communication consisting of mobile wireless terminals (e.g., handset, PDAs, notebooks). These type of networks are wireless multi-hop packet networks without any fixed infrastructure and centralized administration, in contrast to today's wireless communications, which is based on a fixed, pre-established infrastructure. The design of wireless ad hoc networks faces many unique challenges. In this article, mobile ad hoc networks and their characteristics are described, and the design issues, applications and future trends of such networks will be discussed.

BACKGROUND

In recent years, widespread availability of wireless communication and handheld devices stimulated the research and development on self-organizing networks that do not require a pre-established infrastructure and any centralized architecture. Those spontaneous networks provide mobile users with ubiquitous communication capability and information access regardless of their location. This type of networking is called mobile ad hoc networks.

The idea of mobile ad hoc networks has been under development from the 1970s and 1980s in the framework of Mobile Packet Radio Technology (PRNET-1973) (Jubin & Tornow, 1987) and Survivable Adaptive Networks (SURAN-1983) (Schacham & Westcott, 1987). These projects supported research on the development of automatic call set up and maintenance in packet radio networks with moderate mobility. However, interest in this area grew rapidly due to the popularity of a large number of portable digital devices, such as laptop and palmtop computers, and the common availability of wireless communication devices.

In the middle of the 1990s, with the definition of standards, commercial radio technologies have begun to appear and the wireless research community identified in ad hoc networks a challenging evolution of wireless networks. The success of a network technology is associated with the development of networking products that can provide wireless network access at a competitive price. A major factor in achieving this goal is the availability of appropriate networking standards. Today's emerging standards and technologies for constructing a mobile ad hoc network are IEEE 802.11, Bluetooth and ETSI Hiperlan/2. The deployment of mobile ad hoc networks opens a wide-range of potential utilisation from military to miscellaneous commercial, private and industrial scenarios (Perkins, 2001).

MOBILE AD HOC NETWORKS AND THEIR CHARACTERISTICS

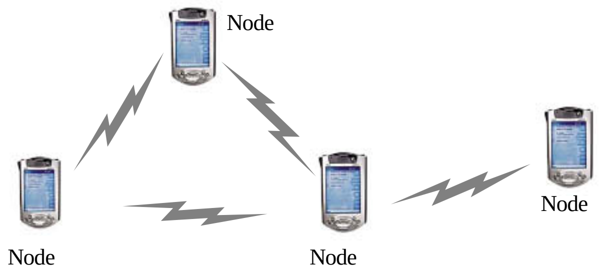
A mobile ad hoc network (MANET) consists of a collection of mobile wireless and autonomous hosts –in this sense simply referred to as "*nodes*"– which spontaneously form a temporary network. The devices may be of various types (e.g., notebook computers, PDAs, cell phones, etc.) and various capacities (e.g., computing power, memory, disk, etc.).

The most important characteristic of such a network is its independence of any fixed infrastructure (e.g., base station or access point) or centralized administration. All networking functions, such as determining the network topology, multiple access, and routing of data over the most appropriate paths, must be performed in a distributed way. These tasks are particularly challenging due to the limited communication bandwidth available in the wireless channel.

Actually, the idea of ad hoc networking is sometimes also called infrastructureless networking (Frodigh, Johansson & Larsson, 2000). An ad hoc network is able to operate autonomously and is completely self-organized and self-configured. Therefore, it can be easily and rapidly installed. In an ad hoc environment people and vehicles can be interworked in areas without a pre-existing communication infrastructure, or when the use of such infrastructure requires wireless extension.

Autonomous nodes may move arbitrarily so that the topology changes frequently without any prior notice. The wireless transmission range of the nodes is also limited; therefore the connection (e.g., wireless link) between the neighboring nodes may break as soon as they move out of range. Consequently, topology of the network and the interconnection patterns among nodes may change dynamically so that links between nodes become

Figure 1. A mobile ad hoc network



unusable. Because of the dynamic nature of ad hoc networks, new routes must be considered and maintained using routing protocols.

Another important property of ad hoc networks is the multi-hop capability. It is given that cellular networks — also called single-hop networks — rely on a fixed wired infrastructure to achieve the task of routing and maintain the connection end-to-end. On the contrary, a mobile node in an ad hoc network that cannot reach the destination directly, because it does not lie within its radio transmission range, will need to relay its information flow through other nodes. This implies the mobile hosts to incorporate routing functionality so that they can act both as routers and hosts.

Other important characteristics of MANET include (Perkins, 2001):

- **Dynamic topologies:** Nodes are free to move arbitrarily; thus, the network topology may change randomly and rapidly at unpredictable times.
- **Bandwith-constrained links:** Caused by the limits of the air interface. Furthermore, multiple access, multipath fading, noise and signal interference decrease the limited capacity available at the allocated frequency rate.
- **Energy-constrained operation:** MANETs inherently imply an underlying reliance on portable, finite power sources.
- **Limited security:** Mobile networks are in general more vulnerable to eavesdropping, spoofing and denial-of-service attacks than fixed-cable networks.

DESIGNING ISSUES

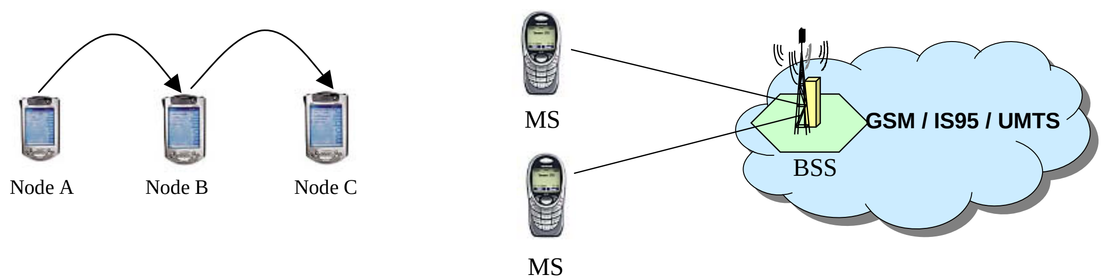
Physical Layer and MAC Layer

A well-designed architecture for mobile ad hoc networks involves all networking layers, ranging from the physical layer to the application layer. Information as node distribution, network density, link failures, and etcetera, must be shared among layers, and the MAC (medium access control) layer and the network layer need to collaborate in order to have a better view of the network topology and to optimise the number of messages in the network (Bruno, Conti, & Gregori, 2001; Kurose, Schwartz, & Yemini, 2000).

The main aspects of designing the physical transmission system are dependent on several characteristics of the radio propagation channel such as path loss, interference and fading. In addition, since mobile terminals usually have limited power resources, the transceiver must be power efficient. These aspects are taken into account while designing the modulation, coding, and power control features in the radio equipment. In principle, the radio equipment in the nodes forming a mobile ad hoc network can use any technology as long as it provides reliable links between neighboring mobile terminals on a common channel. Candidate physical layers that have gained prominence are infrared and spread spectrum radio techniques.

The MAC (medium access control) layer plays the key role in determining the channel usage efficiency by resolving contention amongst a number of unsupervised terminals sharing the common channel. An efficient MAC protocol allows coordinated access to the limited resources. The main goal of a MAC protocol is therefore maximizing the probability of successful transmissions and maintaining fairness amongst all users.

Figure 2. Mobile ad hoc networks (multi-hop networks) in comparison to today's cellular (single-hop) networks



(Node C is reached from node A via node B in multihop way)

Though research on medium access schemes for wired local area networks (LANs) have been done for many years, the same concepts cannot be directly applied to wireless LANs. In a wired medium, the transmitted signals are almost received with the same signal strength at all terminals connected to the same shared medium. Hence a terminal in a LAN can avoid contention by sensing the presence of a carrier to determine if any other terminal is using the channel before it starts a transmission.

However, designing MAC protocols for wireless networks faces a different set of challenges. Propagation path losses in the wireless channel cause the signal power to decline with distance.

Since the strength of the received signal depends on the distance from the transmitter, the same signal is not heard equally well by all terminals. Hence carrier sensing is not very effective in wireless. Typical problems of using carrier sensing to determine the availability of the wireless channel are *hidden node problem* and *exposed node problem* (Nasipuri, 2002). Both the hidden node and the exposed node problems arise due to the fact that carrier sensing is only performed at the transmitter, whereas its effect is determined by the interference power at the receiver, which are usually different due to propagation path loss characteristics.

In order to address the hidden terminal problem, IEEE 802.11 has the option of adding the mechanism of an exchange of *REQUEST TO SEND (RTS)* and *CLEAR TO SEND (CTS)* control packets between a transmitting and receiving nodes before initiating the transmission of a data packet.

Several concerns with the IEEE 802.11 MAC has motivated researchers to explore newer techniques to improve the channel utilization and throughput in mobile ad hoc networks. The basic access method of the IEEE 802.11 MAC protocol is susceptible to inefficiencies due to the hidden and exposed terminal problems. The RTS/CTS option reduces the hidden terminal problem but not the inefficiency caused by the exposed terminal problem.

A wireless transceiver cannot transmit and receive at the same time as the transmitted signal will always be far stronger than any received signal. Hence, a wireless terminal cannot detect if its transmission has been successful. To inform the transmitting node about the successful packet transmission, the receiver sends an ACK (acknowledgement) packet back to the transmitter after it receives a data packet. If the transmitter does not receive an ACK within a fixed period of time, it assumes that the transmitted packet has been lost. Many different schemes have been designed for reducing these problems in wireless channel access.

ROUTING

Movements of nodes in a mobile ad hoc network cause the nodes to move in and out of range from one another. Consequently, topology of the network and the link and connection patterns between nodes may change dynamically so that links between nodes become unusable. As depicted, in contrast to conventional wireless networks, mobile ad hoc networks have no fixed network infrastructure or centralized administrative support for their operations. Because of the dynamic nature of ad hoc networks, new routes must be considered and maintained using routing protocols. Since the network relies on multi-hop transmissions for communication, this imposes major challenges for the network layer to determine the multi-hop route over which data packets can be transmitted between a given pair of source and destination nodes.

Because of this time-varying nature of the topology of mobile ad hoc networks, traditional routing techniques, such as the shortest-path and link-state protocols that are used in fixed networks, cannot be directly applied to ad hoc networks. A fundamental quality of routing protocols for ad hoc networks is that they must dynamically adapt to variations of the network topology. This is implemented by devising techniques for efficiently tracking changes in the network topology and rediscovering new routes when older ones are broken. Since an ad hoc network is infrastructureless, these operations are to be performed in a distributed fashion with the collective cooperation of all nodes in the network (Royer & Toh, 1999; Perkins, 2001). Some of the desirable qualities of dynamic routing protocols for ad hoc networks are:

- **Routing overhead:** Tracking changes of the network topology requires exchange of control packets amongst the mobile nodes. These control packets must carry various types of information, such as node identities, neighbor lists, distance metrics, and etcetera, which consume additional bandwidth for transmission. Since wireless channel bandwidth is at a premium, it is desirable that the routing protocol minimizes the number and size of control packets for tracking the variations of the network.
- **Path optimality:** With constraints on the routing overhead, routing protocols for mobile ad hoc networks are more concerned with avoiding interruptions of communication between source and destination nodes rather than the optimality of the routes. Hence, in order to avoid excess transmission of control packets, the network may be al-

lowed to operate with suboptimal (which are not necessarily the shortest) routes until they break. However, a good routing protocol should minimize overhead as well as the path lengths. Otherwise, it will lead to excessive transmission delays and wastage of power.

- **Loop freedom:** Since the routes are maintained in a distributed fashion, the possibility of loops within a route is a serious concern. The routing protocol must incorporate special features so that the routes remain free of loops.
- **Complexity:** Another problem of distributed routing architectures is the amount of storage space utilized for routing. Ad hoc networks may be applied to small portable devices, such as PDA and handhelds, which are memory and hardware scarce. Hence, it is desirable that the routing protocol be designed to require low storage complexity.
- **Scalability:** Routing protocols should be able to function efficiently even if the size of the network becomes large. This is not very easy to achieve, as determining an unknown route between a pair of mobile nodes becomes more costly in terms of the required time, number of operations, and expended bandwidth when the number of nodes increases.

Because of its many challenges, routing has been a primary focus of researchers in mobile ad hoc networks. The MANET working group within the IETF (IETF, 2004) studies solutions for routing framework and develops IP-based routing protocols for mobile ad hoc networks. (Macker & Corson, 1998; Corson & Macker, 1999). Consequently, a large number of dynamic routing protocols applicable to mobile ad hoc networks have been developed.

APPLICATIONS

The term ad hoc means "as needed for a specific case," "done or set up solely in response to a particular situation or problem without considering wider issues." Both definitions stress out the urgent, specific and short-term need character. Indeed, an ad hoc network will be deployed in an area where a support for mobile communication is not available due to low expected usage and high costs. The network can disappear after its function has been served.

One of the original motivations for the development of this technology lay in satisfying military needs like battlefield survivability. Soldiers must be able to move about freely without any restrictions imposed by wired communication devices. Moreover, the military cannot rely on access to a fixed preplaced communication infrastructure especially during manoeuvres in enemy territories (Perkins, 2001).

Other potential practical utilities of mobile ad hoc networks could include:

- Commercial and industrial scenarios: Associates sharing information during a meeting, participants in a conference exchanging documents or presentations
- Educational scenarios: Students using laptop computers to participate in an interactive lecture
- Emergency coordination in disaster areas, where a hurricane or earthquake have destroyed the communication infrastructure
- Extension of the coverage area of cellular networks
- Communication between smart household appliances in the home environment
- Communication between "wearable" computing devices
- Inter-vehicle communication like within a convoy or in highway
- Sensor Networks: represent a special kind of ad hoc networks. They typically consist of nodes equipped with sensing, processing and communication abilities. Sensor networks are used to monitor remote or inhospitable physical environments

FUTURE TRENDS

Need for wireless ad hoc networking will arises in the context of shared desktop meetings (e.g., cooperate learning, workshops, conferences), disaster recovery, wireless inter-vehicle communication, proprietary local networks for private, commercial, business and governmental, use (universities, hospitals, home environments, office environments) or in various military applications.

New technological standards, regulations (e.g., IEEE 802.11, IEEE 802.15), emerging products (e.g., PDAs, notebooks) and new business models will enable fast deployment of such networking concepts.

CONCLUSION

Wireless ad hoc networking is a new form of communication offering new emerging technologies and applications. Several issues have to be considered which are arising due to characteristics of this type of networks. The critical issues discussed in this article offer many implications and challenges to technology, business and the user community as well.

The successful deployment and a broad acceptance of such networks will depend on the availability of appropriate standards and regulations. Further, the development of new products and business models at a competitive price will play key role.

REFERENCES

Bruno, R., Conti, M., & Gregori, E. (2001). WLAN technologies for mobile ad hoc networks. In *Proceedings of the 34th Hawaii International Conference on System Sciences*.

Corson, S., & Macker, J. (1999, January). *Mobile Ad Hoc Networking* (MANET). IETF RFC 2501. Retrieved from *www.ietf.org/rfc/rfc2501.txt.*

Frodigh, M., Johansson, P., & Larsson, P. (2000). *Wireless ad hoc networking: The art of networking without a network*. Ericsson Review No. 4. Retrieved from *www.ericsson.com/about/publications/review/2000_04/files/2000046.pdf*

Giardano, S. (2000). Mobile ad-hoc networks. In I. Stojmenovic (ed.), *Handbook of Wireless Networks and Mobile Computing*. New York: Wiley (Imprint) Inc.

IETF MANET Working Group. (2004). Retrieved from *http://www.ietf. org/html.charters/manet-charter.html*

Jubin, J., & Tornow, J.D. (1987). The DARPA packet radio network protocols. *Proceedings of the IEEE*, *75* (1), 21–32.

Kurose, J.F., Schwartz, M., & Yemini, Y. (2000). Multiple access protocols and time constraint communications. *ACM Computing Surveys*, *16*, 43-70.

Macker, J., & Corson, S. (1998). Mobile ad hoc networking and the IETF. *ACM Mobile Computing and Communications Review*, *2* (1).

Asis, N. (2004). Mobile ad hoc networks. In Farid Dowla (ed.), *Handbook of RF and Wireless*. Retrieved from *http://www.ece.uncc.edu/~anasipur*

Perkins, C. E. (2001). *Ad hoc networking*. Addison Wesley

Royer, E. M., & Toh, C. K. (1999). A review of current routing protocols for ad hoc mobile wireless networks. *IEEE Personal Communications*, pp. 46–55.

Schacham, N., & Westcott, J. (1987). Future directions in packet radio architectures and protocols. *Proceedings of the IEEE*, *75* (1), 83–99.

Schiller, Jr., J. (2003). *Mobile Ccommunications*. (2nd ed.). Pearson Education.

KEY TERMS

Carrier Sensing: Determination that the medium is not being used by a neighboring transmitter before accessing the channel.

Cellular Networks: A network consisting of several cells served by fixed, pre-established infrastructure to cover a geographic area, for example, GSM, IS-95, UMTS.

Dynamic Topology: Due to the node mobility the network topology of mobile multi-hop ad hoc networks are changing continuously in time.

Exposed Node: This is the reverse problem, where a transmitting or "exposed" node is within range of a sender, but is out of range of the intended destination.

Hidden Node: A node may be hidden or out of range from a sender but within range of its intended receiver.

MANET (Mobile Ad Hoc Networks): A network of wireless mobile nodes formed dynamically, self-organizing and without any central administration.

Multi-Hop: The mobile nodes are cooperating to forward data on behalf of one another node to reach distant stations that would otherwise have been out of range of sending node.

Node: Mobile terminal.

RTS/CTS (Request To Send/Clear To Send): Control packets between the transmitting and receiving nodes before initiating the transmission of a data packet.

Single-Hop: Direct communication between two nodes or entities without any intermediate station or network entity.

W

Wireless Middleware

Kenneth J. MacGregor
University of Cape Town, South Africa

INTRODUCTION

Middleware is not a new concept in distributed computing. It was first accepted as a term used to describe distributed software architecture during the 1990s and has evolved significantly since then with the increase in computer networking and the resulting use of distributed systems. The following is aA definition of middleware that was proposed by Emmerich (2000) and embodies the concepts of middleware: “Middleware is a layer between network operating systems and application components which facilitates communication and coordination of distributed components” (p.120).

Middleware has become widely adopted in industry to simplify the problem of constructing distributed systems as it resolves the heterogeneity between systems and provides higher-level primitives for application engineers to focus on application requirements and not their connectivity.

Wireless middleware, as a specialized subset of middleware, has come into increased significance in the current decade due to the proliferation of wireless networks. Additionally, in the last few years, there has been an upsurge in the development of the mobile devices sector. Personal Digital Assistants (PDAs) and a new generation of cellular phones have the ability to access different types of network services and run many applications. To cope with these innovations, the role of wireless middleware has become increasingly important in providing a reliable channel for data access between mobile devices and servers on a wired network. This provides applications running on the mobile client to synchronize with and participate in the application dialogues that are hosted on the enterprise servers.

Wireless middleware is an intermediate software component that is generally located on a wired network between the wireless device and the application or data residing on a wired network. The purpose of the middleware is to increase performance of applications running across the wireless network by serving as a communication facilitator between components that run on wireless and wired devices (Wireless Nets). Wireless middleware serves as a communication facilitator by addressing the numerous ways in which communication can fail among components in a distributed application. Sources of communication failure, among many others, include a component going off line voluntarily but unexpectedly, a component unexpectedly terminating, a component failing to respond to a request in a reasonable amount of time, and communication being severed in the midst of a request (Sunsted, 1999).

TYPES OF WIRELESS NETWORKS

Currently there are two distinct types of wireless networks that can be categorized by their transmission range and that differ significantly in the latency and bandwidth characteristics of the transmission.

Fixed Wireless Networks

Fixed Wireless Networks use wireless transmitters with a relatively short transmission distance and normally correspond to one of the IEEE 802.11 standards. Networks of this type have many of the characteristics of fixed wire networks, such as low latency, high bandwidth, and relatively good reliability. These networks tend to be used in a specific area such as a Starbucks coffee shop, an airline lounge, or a corporate office. For wireless networks of this type, the standard middleware that is normally employed in wired networks and can be used as such are the de facto standard interprocess communication mechanisms such as XML and SOAP, which are becoming increasingly common.

Highly Mobile Wireless Networks

These wireless networks are used predominately to serve highly mobile clients such as mobile cellular phones or satellite communications. Networks of this type have extremely high latency, relatively low bandwidth, and greater unreliability, as any number of interruptions can occur to disrupt the communication with the client. As this type of network is increasingly becoming used for M-commerce, specialized middleware is required to cater to the additional challenges.

It is this type of wireless network (i.e., high latency, low bandwidth, unreliable, and insecure networks) to which the remainder of this entry refers.

CHALLENGES FOR WIRELESS MIDDLEWARE

Wireless middleware is presented with many challenges by the very nature of the wireless environment. These challenges are not adequately addressed by standard data communication methods that were designed for wired environments. One of the main advantages of middleware is to allow applications to run on multiple platforms without the need to rewrite them. As such, wireless middleware should support heterogeneous devices, permit applications to be ported from one device to another, and, in addition, handle the limitations of mobile devices, among which are the small amount of memory and the lack of processing power. Wireless middleware on mobile devices has some distinct issues not present in traditional fixed line systems.

Disconnected Operation

Wireless communication devices tend to lose and regain network connectivity much more often than non-mobile applications. The middleware must be able to cope with intermittent communication links and implement software layers that ensure the delivery of important data between the server and the wireless device.

Resource Constraints

Mobile applications must be optimized aggressively for small ROM and RAM footprints, as well as for low usage of CPU cycles and battery power.

Multiple Bearers

Internet applications only need to support HTTP or TCP/IP. Wireless applications are written to perform on many different networks. A bearer in a wireless network is a transport mechanism of which wireless bearers could be SMS, GPRS, Infrared, Bluetooth, or HTTP. An application written for one bearer typically needs to undergo substantial modifications in order to run on another bearer.

Heterogeneous Software

There are a number of operating systems that are prevalent in the wireless device market. Because of the memory restrictions placed on these systems, they do not possess all the features of server operating systems. Thus, it is often difficult to ensure that the middleware can effectively run on all the required hardware devices. Often, features that are available on one device are not available on another device with a different operating system. For example, object serialization, which is a standard method of communication for components between clients and server systems, is not available in Windows CE.Net. Thus, programming communications between components where the device is a CE device require the applications developer to interface with a middleware product that cannot serialize objects.

Security

Security is a key concern in a mobile environment. Identification, authentication, and data encryption must be embedded into wireless middleware applications.

Scalability

As a result of the proliferation of wireless devices, scalability of applications can easily grow to hundreds of thousands of client systems. Ensuring the scalability of applications to this number of clients presents a major problem.

Deployment and Management

Deploying applications to many clients, managing them, supporting them, and ensuring their integrity and support when they are not online, present another major challenge to the middleware.

COMMERCIAL SIGNIFICANCE OF WIRELESS MIDDLEWARE

The development of application software such as M-Commerce applications and client/server applications should not be designed so they are independent of the network over which they are deployed. This ensures that applications do not require to be rewritten for new technologies, and the applications developer and business enterprise can concentrate on the business benefits of the application and not the technology on which they reside. For example, a stock ordering application running on a PDA for a manufacturing concern should integrate with the manufacturing schedule in the corporate head office. The speed of taking the orders, scheduling them, and getting them in the production schedule is what is important to the manufacturing company. Wireless middleware provides this network transparency and enables wired applications to be ported to wireless environments with the minimum of software modification.

Wireless Network Requirements

In order to meet the needs of the applications running on them, typical wireless middleware solutions have the following basic requirements:

- **Intelligent restarts**
  The middleware should neither raise communication errors nor lose data when a mobile device loses network coverage. It should incorporate a recovery mechanism that detects when a transmission has been cut, and when the connection is reestablished, the middleware resumes transmission from the break point instead of at the beginning of the transmission.
- **Store-and-forward messaging**
  Message queuing is implemented to ensure that users disconnected from the network will receive their messages once the station comes back online. This, however, can lead to a large number of messages being stored for stolen or broken clients.

Small Footprint

It is particularly important that the messaging client library, stored in the wireless device should have a small memory footprint (ROM and RAM).

Open bearer Models

The middleware should offer the same set of communication abstractions on top of various wireless bearers. Modern wireless communications use GSM, GPRS, and possibly 3G mobile communication protocols. This allows applications to be developed once and operate on top of various bearers. This, however, may lead to the least rich subset of the possible bearers being implemented.

Multi-Platform Language Availability

The API should be compatible with computer languages available on multiple platforms and operating systems. This allows applications to be adapted for various platforms more easily.

Security

Security is essential in a wireless system, particularly for applications such as M-Conmmerce, which are becoming increasingly widespread (Allnet, 2002). As wireless communications cannot be physically secured, the needs of wireless access present a new series of challenges. Wireless access to enterprise systems puts not only the client device, but also the data, well beyond the physical control of the organization. Sniffing of data traffic can be done without any risk of detection over a much wider range of locations. Furthermore, the client device, in the case of a cell phone or PDA, is even easier to steal than a laptop computer with the additional loss of security. Compromise of the wireless client thus poses a double threat to data: the remote access to data that the device enables, and immediate access to the downloaded data which is stored within it. The wireless middleware should incorporate a security mechanism. Thus, access control, identification, authentication, and end-to-end data encryption should be provided by the middleware. Providing these facilities in the middleware and not in the application dramatically simplifies the development of secure mobile solutions.

Scalability

The middleware must support scalability, as many thousands of clients could be connected. It must not become overloaded when repeatedly sending messages, and must handle increasing congestion resulting from disconnect clients. In addition, it must be able to identify and delete messages that can never be delivered perhaps due to a wireless device becoming broken or stolen.

Intelligent restarts are required in order to enhance the robustness of the distributed system that is built using the middleware. A robust distributed system must detect failures, reconfigure the system so that computations may continue, and recover when a link is repaired.

MESSAGING MIDDLEWARE

These basic requirements of wireless middleware can be met by messaging middleware.

Messaging middleware allows general-purpose messages to be exchanged in a client/server system using message queues. Applications communicate over networks by putting messages in queues and getting messages from queues. This messaging and queuing allows clients and servers to communicate across a network without being linked by a private, dedicated, logical connection. The clients and servers can run at different times. All communication is by putting messages on queues and by taking messages from queues. It is used when a distributed application can tolerate a certain level of time-independent responses. For example, nomadic client/server systems can accumulate outgoing transactions in queues and do a bulk upload when a connection can be established with an office server.

Since most of the commercial middleware products are based on the Java Messaging System (JMS) (JMS

Specification, 2002), we will look at how it can be used as a wireless middleware.

JMS, developed by Sun Microsystems, Inc., is part of the Java2 Enterprise Edition (J2EE) platform. JMS is an Application Programming Interface (API) for accessing Messaging applications in Java Programs. The JMS API is not specifically intended for the wireless domain. Traditional JMS was designed to allow Java applications to communicate with existing wireline messaging systems. As a result, a full JMS implementation is too "fat" for wireless devices because low power consumption and a smaller memory footprint are required (Spiritsoft, 2002). However, JMS has become the "de facto" standard as an API for wireless middleware, and most wireless middleware products are based on it.

THE JMS API

The JMS API allows applications to create, send, receive, and read messages which can arrive asynchronously, which means that the client does not have to specifically request messages in order to receive them. In addition, the programmer can specify different levels of reliability, depending on the type of message that is transmitted. Unlike traditional low level network remote procedure call mechanisms, JMS is *loosely coupled*, which means that the sending and receiving applications do not both have to be available at the same time in order to enable communication (Java, 2002).

The JMS API Architecture

The following components are present in a JMS implementation:

- *JMS Provider* – Provides administrative and control features. It is the system that implements JMS interfaces.
- *JMS Clients* – Components or programs, written in Java, that are the producers and consumers of messages, clients, and servers. In the wireless domain, clients may include cellular phones and PDAs.
- *Messages* – The objects that are transmitted between JMS clients.
- *Administered Objects* – These are configured by the administrator and include destinations and connection factories. Together, the administered objects form the Java Naming and Directory Interface (JNDI) API namespace. For one client to establish a logical connection with another through the JMS Provider, it needs to perform a lookup of administered objects, using a standard directory service.

Further details on the JMS communication API can be found on the Java tutorial.

JMS supports the two more common approaches to messaging; namely, *point-to-point* and *publish-subscribe*, which permit messages to be transmitted directly between a client and a server, or for a client to subscribe to specific messages (e.g., sports scores) and servers to broadcast these messages to the specific clients that request them.

NETWORK PROTOCOLS

Contemporary middleware products, whether message oriented or otherwise, have not been designed for mobile devices (Maffeis, 2002). This inappropriateness stems partly from the use of communication protocols that were designed for wired networks. The dominant wired network is the Internet, which has the Transmission Control Protocol (TCP) as its transport protocol. Regular TCP, which is TCP designed for wired networks, could be used in the JMS Connection object; however, this may lead to intolerable inefficiencies. Wireless links do not provide the degree of reliability that hosts expect. The lack of reliability stems from high error rates of wireless links when compared to wired links. Additionally, certain wireless links, such as cellular networks, are subject to intermittent connectivity problems due to handoffs that involve calls being transferred between base transceiver stations in adjacent cells. The root of the problem is that congestion avoidance in the wired TCP algorithms is based on the assumption that most packet losses are due to congestion. As a result, the TCP sender spends excessive amounts of time waiting for acknowledgements that do not arrive.

As a result, most wireless middleware products and JMS implementations use either different transport mechanisms, such as a reliable data protocol similar to that described in RFC 908 (Velten et al., 1984), or a significantly lightweight versions of the standard TCP protocol.

FUTURE TRENDS

As wireless devices become increasingly powerful and more accepted, wireless middleware will have to play an increasing role in ensuring that applications run efficiently in this environment. Currently, the versions of wireless middleware being used are, at best, second-generation versions; however, work is being undertaken in a number of areas.

Network Protocols

Alternative network protocols suitable for wireless communication are being analysed (Farooq & Tassiulas, 2003). This will lead to faster communications between the wired and wireless devices.

Application Communication Paradigms

As wireless technology becomes more available, developers of distributed applications are becoming more interested in how that technology affects the performance of their systems. Thus, the efficiency of different paradigms of communication, such as publish/subscribe, is being investigated (Caporuscio & Carzaniga, 2002); Farooq et al., 2004).

Efficiency of Wired Protocols

As most distributed applications on the Internet are using protocols such as SOAP and XML, work is being undertaken to optimize these protocols for wireless communications (Hanslo, 2004).

Alternative Middleware Products

Currently, Java (JMS, in particular) is effectively the only platform suitable for mobile communications. As the market for wireless applications becomes larger and more profitable, there will be more middleware products vying for a share. Thus, there will be greater competition and more choice for the applications developer. Some of these middleware products could permit applications for wireless ad hoc networks such as peer-to-peer computing, which could easily be developed (Kortuem, 2002).

As a result of this work, wireless applications will be developed, integrated into the enterprise, and executed with the efficiency and reliability that is expected from wired applications.

Conclusion

Wireless middleware has evolved from the traditional middleware background to a product field in its own right. The early attempts of stating that standard wired middleware products are "suitable for wireless" have passed and the wireless middleware products in commercial use today have different requirements, use different transport protocols, and have a different concept of robustness of connection from traditional products. In the future, more advanced products that are specifically catered to the wireless domain will be produced. These products will enable more advanced applications to be implemented and will further increase the usefulness of the mobile devices and PDAs. Finally, with the use of wireless middleware, ubiquitous computing will become a reality.

REFERENCES

Allnetdevices (2002). Mobile payments set to soar. Retrieved September 17, 2003 from *www.allnetdevices.com/wireless/news/2002/02/07/study_mobile.html*

Caporuscio, M., & Carzaniga, A.W. (2002, July). Middleware performance analysis: An experience in evaluating publish/subscribe services in a wireless network. *Proceedings of the Third International Workshop on Software and Performance* in conjunction with the *International Symposium on Software Testing and Analysis (ISSTA)*, Rome, Italy.

Emmerich, W. (2000). Software engineering and middleware: A roadmap. *Proceedings of the Conference on the Future of Software Engineering*, Limerick, Ireland, (pp.117-129). Available online. Retrieved March 17, 2002 from *www.cs.ucl.ac.uk/staff/A.Finkelstein/fose/finalemmerich.pdf*

Farooq, A., & Tassiulas, L. (2003, June). Comparative study of various TCP versions over a wireless link with correlated losses. *IEEE/ACM Transactions on Networking (TON)*, *11*(3), 370-383.

Farooq, U. et al. (2004, January). Performance of publish/subscribe middleware in mobile wireless networks, ACM SIGSOFT software engineering notes. *Proceedings of the Fourth International Workshop on Software and Performance, 29*(1), Redwood Shores, CA.

Hanslo, W., & MacGregor, K.J. (2004, September). The optimization of XML for mobile communications. *Proceeding SAICSIT Conference*, Stellenbosch, South Africa.

Java: Java message service tutorial. Retrieved April 26, 2002 from *http://java.sun.com/products/jms/tutorial/1_3_1-fcs/doc/overview.html*

JMS specification: Java message service specification version 1.1 (2002, April). Retrieved April 17, 2002 from *http:///java.sun.com/products/jms/docs.html*

Kortuem, G. (2002, October). Proem: A middleware platform for mobile peer-to-peer computing: ACM SIGMOBILE. *Mobile Computing and Communications Review*, *6*(4), 62-64.

Maffeis, S. (2002). JMS for mobile applications and wireless communications, chapter 11, p17. Retrieved April 11, 2002 from *www.softwired-inc.com/people/maffeis/ articles/softwired/profjms_ch11.pdf*

Spiritsoft (2002). Go beyond JMS SpritSoft home. Retrieved September 19, 2002 from *www.spiritsoft.com*

Sundsted, T. (1999). Messaging makes its move. Retrieved September 19, 2002 from *http://www.javaworld.com/javaworld/jw-02-1999/jw-02-howto.html*

Velten, D., Hinden, R., & Sax, J. (1984). Reliable data protocol. *RFC-908 BBN Communications Corporation*.

KEY TERMS

Application Synchronization: A specific type of wireless application whereby the data on the wireless device is synchronized with that on the main server.

Handover: In a cellular network, the radio and fixed voice connections are not permanently allocated for the duration of a call. Handover, or handoff as it is called in North America, means switching an ongoing call to a different channel or cell. This often results in loss of connection.

M-Commerce: A wireless application that involves the change of money. This could be through the use of a mobile device to pay for services.

Network Bandwidth: The data capacity through a network.

Network Latency: The time for a message to travel over the network from the sender to the receiver. In cellular networks, this includes the time from the device over the network to the base station, over a landline to the service provider, and out to the application. It could possibly include satellite transmission time from the earth station to the satellite and down to another earth station.

Point to Point: The communication from a single client to a single server.

Publish Subscribe: A method of communication by which clients register to be notified when some event occurs. When the event happens, the server checks though the list of subscribed clients and broadcasts to them the event data. This is often used in wireless applications for information on stock price changes, sports results, and so forth.

Wireless Application: An application running on a wireless device that transmits and receives data over a wireless network.

Wireless Technologies to Enable Electronic Business

Richi Nayak
Queensland University of Technology, Australia

INTRODUCTION

Research and practices in electronic business (e-business) have witnessed an exponential growth in the last few years (Liautand & Hammond, 2001). Wireless technology has also evolved from simple analog products designed for business use to emerging radioactive, signal-based wireless communications (Shafi, 2001). The tremendous potential of mobile computing and e-business has created a new concept of mobile e-business or e-business over wireless devices (m-business).

BACKGROUND

M-business can be defined as the use of mobile technology in exchange of goods, services, information, and knowledge. M-commerce is the execution of transactions done on mobile equipment via mobile networks, which may be wireless or switched public networks. M-business includes the range of online business activities, business-to-business and business-to-consumer, for products and services through wireless devices such as mobile phones with display screens, personal digital assistance (PDA), two-way pagers, and low-end or reduced-size laptops.

Example applications are mobile ticketing and receipting, banking, gaming, e-mail communication, weather forecast, sport scores access, movie database access, stock exchange information, ordering of books, and other daily needs such as food and groceries. With new emerging mobile applications, users only receive selective and real-time information personalized to their interests (Ratsimor, Korolev, Joshi & Finin, 2001). For example by using a positioning system, the advertising information of local services and entertainment can be sent whenever a user is passing by a shopping mall. Another application is "inventory management" that tracks the location of goods, services, and people to determine delivery times. Multiple trucks carry large amounts of inventory that companies could access for just-in-time delivery (Varshney & Vetter, 2002).

Significant benefits of m-business to consumers are convenience, portability, safety, integrating existing mobile phones with mobile computing technology, verifiable receipts, and transaction records that can be made available instantly and permanently on a smartcard. Significant advantages of m-business to service providers and content providers include driving additional revenue and decreasing consumer attrition by offering new m-business services to specific groups of customers.

WIRELESS TECHNOLOGIES TO ENABLE M-BUSINESS

Many wireless technologies exist to enables m-business services (Tsalgatidou, Veijalainen, Markkula, Katasonov & Hadjiefthymiades, 2003). All technologies try to achieve benefits such as being powerful, simple, economical, and secure. Some examples of these techniques follow.

Wireless Application Protocol technology links wireless devices to the Internet by optimizing Internet information so it can be displayed on the small screen of a portable device.[1] Web pages accessed by WAP-enabled mobile portals during m-business transactions must be written in WML.[2] It is not sure how well WAP will be able to proliferate (Tsalgatidou et al., 2000). Developments such as third-generation (3G) mobile communications and XYPOINT WebWirelessNow applications (Wen, 2001) already allow mobile phone consumers to experience the Web services without WAP.

Wireless Internet connecting technologies that offer textual interface such as WAP significantly suffer from the constraints of wireless communication such as having a small display screen. An alternative solution is providing voice access to users. Advances in speech recognition and text-to-speech technologies have made voice-based communication possible between computers and users over the phone.

VoxML[3] technology, based on the W3C XML standard, enables the application interface to be in the form of dialogues. However, there is an extra overhead for content providers to offer the same Web service through different channels, for example, providing a voice-enabled browser for their wireless customers along with the HTML/XML/WML browser. Another overhead is the processing power that speech recognition requires. Also this type of data transfer mode is not appropriate for applications with confidential data where one could be overheard. Overall, the success of this technology depends on public acceptance of mobile phones as data-delivering tools and the type of applications best suited to their use.

The *Bluetooth* technology further enhances the sphere of mobility by conducting m-business without a heavy network infrastructure unlike WAP and VoxML technologies.[4] The Bluetooth technology is designed to allow low-cost, short-range data (asynchronous) and voice (synchronous) radio link (2.4 GHz, 1 Mb/sec) to facilitate protected connections for stationary (homes, buildings, shopping centers, restaurants, cars, etc.) and mobile (phones, PDAs) computing environments. A simple example of a Bluetooth application is to automatically update mobile phone contents such as phone list, e-mails, and memos without any user involvement when the phone comes within the range of the home/office PC. Currently, the Bluetooth networks providing m-business services are limited to 10 meters only. Also, it has too many flaws in terms of security for the services to be trusted. A promising future of Bluetooth technology is its integration with WAP or VoxML.

Based on infrared technology, the *IrDA (Infrared Data Association)* easy-to-use technology provides low-cost, short-range, point-to-point connectivity between devices, interoperable/cross-platform at a wide range of speeds (115.2kb/s to 4Mb/s) with a physical range of 1 meter. IrDA technology is embedded into 40 million new devices each year such as personal computers, laptops, mobile phones, PDAs, digital cameras, pagers, and so forth.[5] The keyword of IrDA advantages is simplicity for ad-hoc, point-to-point exchange. However, the requirement of direct line of sight for devices to communicate is surely a disadvantage for conducting m-business.

IEEE802.11 (Wi-Fi) technology provides a high data rate over different ranges (54Mbps using the 2.4 and 5 GHz ISM band, 11Mbps using the 2.4 GHz ISM band).[6] The single Media Access Control protocol helps to keep the cost down, but interoperability is a problem. The data transmission rate has to be defined before the transmission between devices can start. In terms of the transmission itself, it is based on the well-known TCP/IP protocol. The availability of unlicensed spectrum is a significant enabler for broad acceptance of this technology. However, the technology has some security flaws. Because of the large physical range (100+ meters) and "always-on" connection model, this technology consumes a lot of power, limiting its use in PDAs, phones, and other lightweight mobile devices. The greatest advantage of this technology for conducting m-commerce is its speed.

HiperLAN, a specification substandard of IEEE802.11, is a short-range technology (from 10 to 100 meters) adapted to 3G networks with low power requirements.[7] HiperLAN provides flexible services such as mobility management and quality of service at low cost. The technology has a potential for conducting m-commerce in terms of supporting both ad hoc and client/server networks.

Ultra Wideband technology is a recent RF technology with advantages like large bandwidth, high data transfer rates, and immunity to interference.[8] Still, the technology is in its early stage of development, and there are not many products using this technology yet. However, in the future this network technology may be a very good alternative for conducting m-commerce.

Mobile Agent technology offers a new computing paradigm in which a program, in the form of software agents, is initiated at the host, can suspend its execution on a host computer, launch itself to another agent-enabled host on the network, resume execution on the new host, and return back to its host with the result (Hayzelden & Bigham, 1999). This type of paradigm advocates the client/server model where the client is a mobile portal and the server is a fixed network. The mobile agent performs various optimizations on the server in lieu of its mobile portal to reduce the problems such as C-autonomy, limited bandwidth, and limited computational power. The fixed network offers its services to the agent such as access to local resources and applications, the local exchange of information between agents via message passing, basic security services, creation of new agents, and so froth. Many research papers emphasize that one of the most promising approaches for developing e-business appli-

Figure 1. A typical platform enabling m-business services

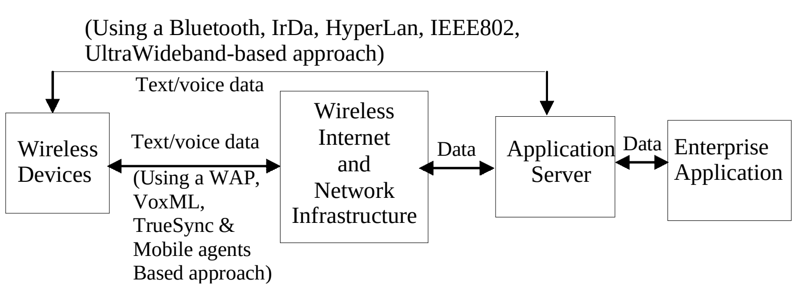
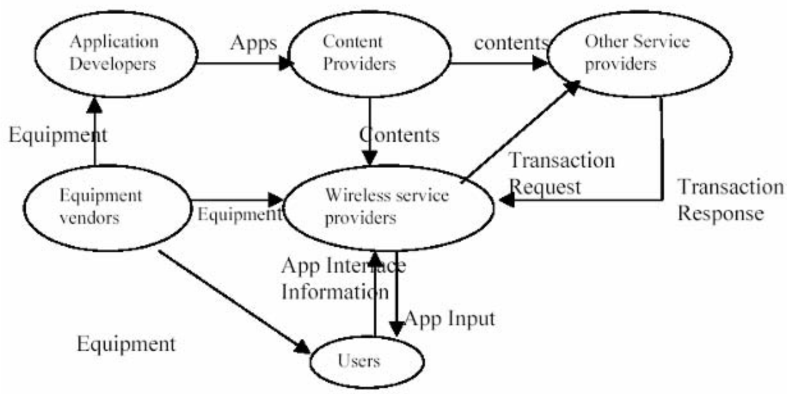


Figure 2. A mobile commerce cycle (Varshey & Vetter, 2002)



cations is mobile agent technology (Dikaiakos & Samaras, 2001; Tsalgatidou et al., 2000).

Figure 1 illustrates a typical platform to enable an m-commerce application. According to the wireless technology, a request from the wireless device is passed on to the application server, where the request is processed and output is returned.

Varshey and Vetter (2002) proposed a mobile commerce cycle that shows the possible interaction between various entities (see Figure 2).

CHALLENGES IN M-BUSINESS

As m-business bears its roots in e-business, many technical, legal, and business perspectives are extrapolated from e-business. But m-business embraces a number of technologies to enable its activities, such as wireless devices, software, and protocols, as well as eliciting a number of issues of its own.

Limited Environment of Mobile Devices

These devices usually have limited display size, input capabilities, data transfer rate, computation power, storage, power usage, and so forth; some may be overcome with the technological developments. For example, low bandwidth problem can be handled by the use of software compression techniques or by only sending the relevant information with the use of positioning systems to reduce the traffic. Such mobile applications should be developed so that a large footprint is not required (one innovation is NTT DoCoMo).[9] However, a low-power, inexpensive, and high-resolution color and larger size display will positively affect the m-business.

Security

Consumers should feel the same sense of security when they shop using a mobile phone as when they shop in the physical world. Connection to a wireless link does not require physical access to the network or devices in an "always-on" mode when connected, as in 3G devices. This makes wireless devices more susceptible to attack. In addition, mobile devices are prone to be lost or stolen. Various encryption and integrity mechanism are required to protect digitally encoded speech and control information in m-business services (Chari, Kermani, Smith & Tassiulas, 2001).

Transactional Issues

Usually transactional models with ACID (Atomicity, Consistency, Isolation, Durability) properties assist an application developer in providing powerful abstraction and semantics to concurrent executions and recovery in transaction management. M-business transactions require these properties to be redefined due to additional requirements such as: (1) the need to transfer money and goods along with data transfer, and (2) increased risk of incomplete transactions as mobile terminals can easily lose network connections.[10] Approaches such as the asymmetric cryptographic algorithms (also called Public Key algorithms) with certification authorities are utilized to fulfill ACID properties of m-business transactions (Veijalainen, 1999; Tsalgatidou et al., 2000).

Interoperability

The m-business market is still in its infancy with many standards, and many standards are competing with each other. This creates doubt for the success of global roaming and for the infrastructure. There is interference to one technology with the use of another technology in terms of broadband spectrum. There have been some improvements in convergence of functions between phones and PDAs. But still, each PDA device comes with its own operating system.

Slow Growth

M-business has not lived up to the promises of the past. Users are not keen to upgrade their mobile phones to take full benefit of m-business applications. Users feel that existing m-business applications do not drive the force to adopt new communications such as 3G. Accordingly, there is delay in implementing system infrastructure such as 3G mobile networks.

FUTURE TRENDS AND CONCLUSION

M-business applications improve the value of the service for the providers, and give an easy and natural interfacing and interacting to the users with mobile portals. Considering it is difficult to provide full convenience due to the limited nature of the wireless devices, the application seems to be able to offer ease of navigation and provide real-time information to users using mobile devices, anytime anywhere.

Many optimists see m-business as a technology that is just one step from everyday use. Many pessimists see many unsolved problems and predict that m-business will not break through in the next few years. As usual, the truth lies somewhere in the middle. Basic techniques are already available. Millions of people are already using mobile portals. Businesses are making profit by moving on to e-business solutions. The potential of m-business is enormous.

So why not integrate them all? Major mobile service providers are taking initiatives (such as MeT,[11] GMCF,[12] WDF,[13] NTT DoCoMo) to envision this technology to flourish. The remaining tasks are rigorous testing and refining of protocols especially suited for m-business applications, resolving related technical and business issues, thus winning the trust of consumers to use m-business services.

REFERENCES

Chari, S., Kermani, P., Smith, S. & Tassiulas, L. (2001). Security issues in m-commerce: A usage-based taxonomy. In J. Liu & Y. Te (Eds.), *E-commerce agents* (pp. 264-282). Berlin: Springer-Verlag (LNAI 2033).

Dikaiakos, M.D. & Samaras, G. (2001). Performance evaluation of mobile agents: Issues and approaches. In R. Dumke et al. (Eds.), *Performance engineering* (pp. 148-166). Berlin: Springer-Verlag (LNCS 2047).

Hayzelden, A. & Bigham, J. (1999). *Software agents for future communication systems*. Berlin, New York: Springer-Verlag.

Liautaud, B. & Hammond, M. (2001). *E-business intelligence: Turning information into knowledge into profit*. New York, London: McGraw-Hill.

Ratsimor, O., Korolev, V., Joshi, A. & Finin, T. (2001). Agents2Go: An infrastructure for location-dependent service discovery in the mobile electronic commerce environment. *ACM Mobile Commerce Workshop*. Retrieved October 5, 2001, from *citeseer.nj.nec.com/455113.html*.

Shafi, M. (2001). *Wireless communication in the 21st century*. John Wiley & Sons/IEEE Press.

Tsalgatidou, A. & Veijalainen, J. (2000). Mobile electronic commerce: Emerging issues. *Proceedings of 1st International Conference on E-Commerce and Web Technologies* (EC-Web), London-Greenwich, UK.

Tsalgatidou, A., Veijalainen, J., Markkula, J., Katasonov, A. & Hadjiefthymiades, S. (2003). Mobile e-commerce and location-based services: Technology and requirements. *ScanGIS 2003*, 1-14.

Varshney, U. & Vetter, R.J. (2002). Mobile commerce: Framework, applications and networking support. *ACM Mobile Networks and Applications, 7*, 185-198.

Veijalainen, J. (1999). Transactions in mobile electronic commerce. In G. Saake, K. Schwarz & C. Trker (Eds.), *Transactions and database dynamics*. Berlin: Springer-Verlag (LNCS 1773).

Wen, H. (2001). Doing the Web without WAP: A discussion with XYPoint's Patrick Carey. *The Wireless Communication Channel*. Retrieved October 1, 2001, from *www.wirelessdevnet.com/channels/lbs/features/xypoint.html*

KEY TERMS

Electronic Business (E-Business): Any type of business transaction or interaction in which the participants operate or transact business or conduct their trade electronically.

Mobile Electronic Business (M-Business): A range of online business activities, business-to-business and business-to-consumer, for products and services through wireless devices such as mobile phones with display screens, personal digital assistance (PDA), two-way pagers, and low-end or reduced-size laptops.

Voice Markup Language (VoxML): Based on the W3C XML standard; designed to support interactive dialogues. VoxML masks the technology behind the voice-to-voice communications by using XML data-tagging structures to link the text-to-speech that generates audio with the speech-recognition software that interprets a user's command.

Wireless Application Protocol (WAP): An approach to link wireless devices to the Internet by optimizing Internet information so it can be displayed on the small screen of a portable device.

ENDNOTES

[1] Edited and retrieved September 28, 2001, from *www.wapforum.org*

[2] Edited and retrieved September 25, 2001, from *www.oasis-open.org/over/wap-wml.html*

[3] *voxml.mot.com*

[4] Edited from "Technology Overview"; retrieved October 1, 2001, from *www.bluetooth.com/v2/document*

[5] *www.irda.org*

[6] *grouper.ieee.org/groups/802/dots.html*

[7] *www.hyperlan.com*

[8] *www.uwb.org*

[9] *www.nttdocomo.com*

[10] Interested readers can refer to Veijalainen (1999) for a detailed study of transaction issues in m-business.

[11] MeT, *www.mobiletransaction.org*, targets to establish a framework for secure mobile transactions, ensuring a consistent user experience independent of device, service, and network.

[12] The Global Mobile Commerce Forum (GMCF) was established in 1997 by a diverse group of companies from around the world to promote the development of mobile commerce services for the benefit of consumers and the companies involved.

[13] The Wireless Data Forum (WDF) was established in 1999 to help the wireless industry to develop new e-business products and services, and to use the Internet to sell products and services.

Women in the IT Profession

Catherine Beise
Salisbury University, USA

Janette Moody
The Citadel, USA

Martha Myers
Kennesaw State University, USA

Amy Woszczynski
Kennesaw State University, USA

INTRODUCTION

Women have been involved with information technology (IT) since the 19th century, when Ada the Countess of Lovelace served as the first programmer of Charles Babbage's analytical engine. Grace Murray Hopper's contributions to COBOL and computing several decades ago are considered so significant that an annual conference is held in her honor (see *www.gracehopper.org*). In fact, the earliest computer programmers tended to be women more often than men (Panteli, Stack & Ramsay, 2001). As the IT field progressed, however, it evolved into what many still view as a male-dominated domain, some say due to its increasing association with power and money (Tapia, Kvasny & Trauth, 2003). Today, women are estimated to make up nearly half of World Wide Web users (Newburger, 2001), but this has apparently not translated into a proportionate participation in IT careers.

IT managers must recruit and retain a skilled and diverse workforce in order to meet the needs of increasingly global enterprises where cross-cultural, heterogeneous workgroups are the norm. However, numerous sources (Computing Research Association, 1999; ITAA, 2003) agree that the proportion of females to males selecting and completing degrees in IT-related fields is declining. Not only are women missing out on career opportunities, the IT profession is also missing potentially valuable alternative perspectives on system design (Woodfield, 2002).

Furthermore, the worldwide digital divide is more extreme for women than men (Hafkin & Taggart, 2001). In many developing countries, women's access to computers is even more limited than men's, and Internet usage statistics, where available, suggest that the majority of users are male. However, IT is an important driver for economic development and should provide women with new opportunities to better their circumstances, provided that a variety of challenges, such as technical education and social and political norms, can be addressed (Hafkin & Taggart, 2001).

Even in more developed countries, females face well-documented (Margolis & Fisher, 2002; Von Hellens & Nielsen, 2001) obstacles all along the pipeline, beginning as early as middle school and continuing through college, graduate school, and industry. Solutions to recruiting and retaining women in IT may serve other underrepresented groups as well, making IT classrooms and IT workplaces more congenial and ultimately more productive environments for everyone.

BACKGROUND

Part of the challenge of recruiting and retaining women in IT stems from the changing nature of IT work. IT is defined for the purpose of this article as an umbrella term that encompasses a variety of job categories that continue to evolve, as hardware, software, and methods continue to increase in sophistication. Traditional professional IT job categories included programmer, systems analyst, system administrator, and software designer. Earlier investigations into women and computing suggested that IT work by its nature was a poor fit for females, seen as solitary and boring, a stereotype that apparently still exists today (AAUW, 2000a). However, as the field has evolved, and as IT has become more integrated into most business organizations and into the work and home lives of many individuals, a wider variety of IT work has emerged. Today, in addition to earlier job titles, IT work also includes software engineer, business analyst, database designer, database administrator, network analyst, network administrator, Web developer, Web engineer, human interface designer, project manager, applications developer, security administrator, and help desk technician.

University curricula have also evolved and expanded to provide formal education, skills, and knowledge suited to these new job types. Entry into the IT field can come from numerous directions, including electrical engineering, computer engineering, computer science (CS), several flavors of information systems (IS), and a more narrowly defined type of IT that focuses on fourth-generation language application development and maintenance. A majority of the data and research studies that point to the precipitous decline in women in IT, however, focuses narrowly on computer science and engineering only (Wardle & Burton, 2002).

THE IT PIPELINE: CHALLENGES AND OPPORTUNITIES

Early Influences

A growing body of educational research documents a variety of factors that influence female attitudes, perceptions, and behaviors toward computers in K-12 (e.g., AAUW, 2000b; Taggart & O'Gara, 2000; Whitley, 1997; Young, 2000). In addition to general socio-cultural trends that appear to dampen girls' enthusiasm and self-efficacy regarding math and sciences around middle-school, girls seem to be influenced by:

- low interest in male-oriented computer games;
- teacher, parental, and peer attitudes that stereotype computing as male oriented;
- lack of access to and experience with computers, leading to lower self-efficacy;
- perceived usefulness, or lack thereof, of computers to themselves and to society;
- lack of IT role models and mentors.

These variables have been grouped into two main categories: environmental and individual. Environmental variables are those which make up the context within which career decisions are made, such as school or work, while individual variables are characteristics of individuals, such as aptitudes and preferences (Woszczynski, Myers & Beise, 2003). Both interact to influence the choices and behaviors of individual girls and women (Trauth, 2002).

College, Graduate School, and the IT Workplace

Some of these same factors apply as women move into college and graduate school. Often due to less previous experience with computers and less preparatory coursework, women continue to experience ambivalence about their interest and abilities in IT, in spite of equal performance regarding computer skills. They often encounter hostile academic environments (Margolis & Fisher, 2002; McGrath-Cohoon, 2001), and their career choices are often influenced by work-family balance concerns.

In the workplace, women are employed at lower levels, make less money, are more likely to leave their organization than men (Baroudi & Igbaria, 1994-1995; Igbaria, Parasuraman & Greenhaus, 1997), and may even be channeled into "softer" positions considered more suited to women, and coincidentally have lower status, are lower paid, and are less visible (Panteli et al., 2001). Some of this may be related to the assumption that women tend to be younger, less experienced, and spend less time in the workplace due to opting out of traditional career paths to raise families, preferring part-time or more flexible options when returning to work. The few studies that have controlled for differences in variables such as age, education, and work experience have shown mixed results (Baroudi & Igbaria, 1994-1995; Igbaria et al., 1997).

Academic institutions and business organizations alike are realizing that they need to supplement earlier efforts that focused on recruiting more women into the field with more attention to retaining them, once there, in school and in the workplace (Tapia et al., 2003). Again, it is expected that both environmental and individual variables will interact to determine outcomes for women.

Interventions and Solutions

A growing body of literature is providing a range of useful approaches to addressing these challenges (e.g., Wardle & Burton, 2002). One way of addressing the problem is to focus on individual factors, that is, change the individuals, by changing attitudes, dispelling stereotypes, improving preparation, and increasing experience with computers. Suggestions aimed at K-12 levels and beyond include:

- Providing more (and more equal) access to computing resources.
- Designing girl-friendly games and applications, to encourage more experience with computers which leads to higher self-efficacy.
- Creating videotapes and other positive media images that demonstrate women in professional IT roles, thus changing attitudes.
- Encouraging girls to take more courses that adequately prepare them for IT-related college majors.

Taken to an extreme, this approach implies that in order to succeed in this male-dominated field, women must become more like men. An alternative perspective, then, is to change the environment, by making it less hostile, less masculine, more family friendly, and more accepting of diversity. Interventions that have been suggested and implemented (e.g., AAUW, 2000b) to address environmental factors, at multiple academic levels, include:

- Train teachers to provide more equal access and to reduce stereotyping of computing as a male domain.
- Provide students with female mentors and role models (female faculty members and mentors from industry).
- Create communities, study groups, clubs, and other social supports for female students.
- Broaden the range of computing activities for younger students.
- Develop programs that do not depend on substantial mathematical and scientific background prior to beginning college. Provide bridge opportunities to increase experience, build competency, and improve self-efficacy.

Finally, many business organizations have implemented a variety of programs to address many of the individual and environmental factors listed above (Bentsen, 2000; McCracken, 2000; Taggart & O'Gara, 2000):

- IT training (intra-organizational and community outreach).
- Mentoring programs, support networks, and general diversity training.
- Visible top management support and an external advisory council to promote cultural change.
- Promotion of work/life balance values and programs for women *and* men.
- Provision of programs that support time management and work-family concerns, such as continuing education, flex-time, day- and elder-care, and concierge services (vs. on-site game tables and sports outings).
- Examination of explicit and implicit reward systems, which may evaluate men differently than women, which may not reward teamwork and communication as valuable leadership skills as much as more traditional definitions of achievement.

FUTURE TRENDS

When IT workers are compared to other categories of workers, two interesting findings emerge: first, persons with a formal education in IT are less likely to pursue work outside of IT than are persons with formal educations in other areas. Second, the IT workforce contains a large number of people without CS or IS credentials, or even necessarily traditional undergraduate educations (Wardle & Burton, 2002). This situation has likely arisen because IT is a relatively new field, because it has grown rapidly, and because there is a growing need to apply IT to other functional areas, from which these employees come.

A related issue is that a frequently cited source of enrollments in computer science programs (Computing Research Association, 1999) focuses solely on research institutions (defined as those that offer a PhD in Computer Science). PhD-granting programs clearly play an important role in the pipeline by providing female IT academics to serve as role models for women enrolled in college. However, a recent study in Georgia (Randall, Price & Reichgelt, 2003) suggests that: 1) more women (all majors) attend non-PhD-granting state universities and colleges than attend research institutions; 2) CS, IS, and IT degree programs at these state universities and colleges attract a higher percentage of women than do the research institutions; and 3) IS and IT programs attract more women than CS programs. The applied nature of IS/IT programs is likely to be part of the reason that these programs are more attractive than CS programs, given women's apparent preferences for work that involves social interaction and social benefit (AAUW, 2000a).

A wider range of data needs to be collected from both academia and industry to gain a better, more up-to-date understanding of the status of women in IT. In addition, students, parents, teachers, and the general public need more education about the range of IT programs and jobs available now and in the future, in the interest of dispelling stereotypes and raising awareness of the variety of career paths and choices in IT.

Past literature on the subject of women in IT has included disparate samples from limited populations. As a result, the picture of why women leave the IT field, or choose not to get into it in the first place, is incomplete and fragmented. One limitation in past research is that, like many academic studies, the subjects are most often students. This is due to convenience as well as the assumption that focusing on students will eventually result in changes to the workplace. However, there is a dearth of research in the IT workplace regarding gender and diversity, and more is needed to understand the problem and apply effective solutions.

Finally, although a number of interventions have been suggested and even implemented, little is known about their effectiveness over time. Researchers need to develop studies that analyze their effects over time. Combined quantitative and qualitative analyses of successful techniques that improve recruitment and retention of females in IS should be provided.

CONCLUSION

Numerous suggestions and guidelines for improving women's representation in the IT workforce with respect to recruitment and retention have been offered. Recurring themes include the lack of self-confidence regarding computing (which can be alleviated with more experience), lack of pre-college preparation, the need for mentors and role models, the importance of community and study groups, and the need to value both family and work priorities. More systemic solutions would transform the "masculinized" IT academic departments and workplaces to become friendlier, more supportive environments for all workers. A summary of recommendations for IT practitioners, IT educators, and IT researchers toward addressing both individual and environmental factors includes:

1. Apply initiatives that work to recruit and retain women in IT education as tools for managers seeking qualified women for IT positions. These include development of mentoring programs, support networks, and general training on respect for diversity and multiculturalism for all workers.
2. Enhance general business initiatives applied to all women to develop more conducive environments for IT women. Rather than on-site games and sports outings, many women find more appealing practical programs that support time management and work-family concerns, such as continuing education, flex-time, day- and elder-care, and concierge services (Bentsen, 2000). Such programs are likely to increasingly appeal to men as well, as youthful IT workers age, marry, and have children.
3. Get involved in post-secondary education to identify, attract, support, and develop well-qualified women for IT positions. Broaden the definition of IT to include not just CS, but also IS and related majors, and contribute to continued curriculum development that balances important theoretical foundations with applied, practical application. Such curricula are more likely to appeal to women, who tend to view computers more as useful tools rather than fun toys.
4. Share knowledge about the evolving nature of IT work, particularly with parents, counselors, teenagers, and other influential groups. Narrow stereotypes need to be replaced by the realities of an exciting and socially fulfilling IT career.

Finally, many organizations are beginning to appreciate the value of a newer, more facilitative leadership style that is often associated with women, that includes teamwork, participatory decision making, and interpersonal communication skills. If such behavior is recognized and rewarded, this could help attract and promote more women into the managerial ranks of IT, eventually transforming the IT workplace, and perhaps leading to broader perspectives in software and interface design.

REFERENCES

AAUW. (2000a). Girls see computer jobs as lonely, boring. *Women in Higher Education, 9*(6), 3.

AAUW. (2000b). *Tech-savvy: Educating girls in the new computer age.* Washington, DC: American Association of University Women Educational Foundation.

Baroudi, J.J. & Igbaria, M. (1994-1995). An examination of gender effects on the career success of information systems employees. *Journal of Management Information Systems*, *11*(3), 181-201.

Bentsen, C. (2000). Why women hate IT. *CIO Magazine*, (September 1). Retrieved August 6, 2001 from *http://www.cio.com/archive/090100_women.html*

Computing Research Association. (1999). *The supply of information technology workers in the United States.* Computing Research Association (CRA). Retrieved June 8, 1999, from *www.cra.org/wits/*

Hafkin, N. & Taggart, N. (2001). *Gender, IT, and developing countries: An analytic study*. Washington, DC: Academy for Educational Development.

Igbaria, J., Parasuraman, J. & Greenhaus, J.H. (1997). Status report on women and men in the workplace. *Information Systems Management, 14*(Summer), 44-53.

ITAA. (2003). *ITAA report of the Blue Ribbon Panel on IT Diversity* (p. 22). Arlington, VA: Information Technology Association of America.

Margolis, J. & Fisher, A. (2002). *Unlocking the clubhouse: Women in computing.* Cambridge, MA: MIT Press.

McCracken, D.M. (2000). Winning the talent war for women: Sometimes it takes a revolution. *Harvard Business Review*, (November/December), 159-167.

McGrath-Cohoon, J. (2001). Toward improving female retention in the computer science major. *Communications of the ACM, 44*(5), 108-115.

Newburger, E. (2001). *Home computers and Internet use in the United States* (p. 12). Washington, DC: U.S Census Bureau.

Panteli, N., Stack, J. & Ramsay, H. (2001). Gendered patterns in computing work in the late 1990s. *New Technology, Work, and Employment, 16*(1), 3-17.

Randall, C., Price, B. & Reichgelt, H. (2003). Women in computing programs: Does the incredible shrinking pipeline apply to all computing programs? *Inroads, SIGCSE Bulletin, 35*(4), 55-59.

Taggart, N. & O'Gara, C. (2000). Training women for leadership and wealth creation in IT. *TechKnowLogia, 2*.

Tapia, A., Kvasny, L. & Trauth, E. (2003). Is there a retention gap for women and minorities? The case for moving in vs. moving up. In M. Igbaria & C. Shayo (Eds.), *Strategies for managing IS/IT personnel* (pp. 143-164). Hershey, PA: Idea Group Publishing.

Trauth, E.M. (2002). Odd girl out: An individual differences perspective on women in the IT profession. *Information Technology and People, 15*(2), 98-118.

Von Hellens, L. & Nielsen, S. (2001). Australian women in IT. *Communications of the ACM, 44*(7), 46-52.

Wardle, C. & Burton, L. (2002). Programmatic efforts encouraging women to enter the information technology workforce. *Inroads, SIGCSE Bulletin, 34*(2), 27-31.

Whitley, B.E. (1997). Gender differences in computer-related attitudes and behavior: A meta-analysis. *Computers in Human Behavior, 13*(1), 1-22.

Woodfield, R. (2002). Woman and information systems development: Not just a pretty (inter)face? *Information Technology and People, 15*(2), 119-138.

Woszczynski, A., Myers, M. & Beise, C. (2003). Women in information technology. In M. Igbaria & C. Shayo (Eds.), *Strategies for managing IS/IT personnel* (pp. 165-193). Hershey, PA: Idea Group Publishing.

Young, B. (2000). Gender differences in student attitudes toward computers. *Journal of Research on Computing in Education, 33*(2), 204-216.

KEY TERMS

Computer Science (CS): A traditional IT curriculum whose focus is technical and theoretical rather than applied, with emphasis on software creation.

Environmental Variables: The context within which career decisions are made, such as the school and work environments.

Fourth-Generation Language: Business application languages and tools such as database and decision supports tools such as SQL, ACCESS, and EXCEL; ERP and other reporting tools; and Web development environments such as Cold Fusion and Frontpage.

Individual Variables: Characteristics of individuals, such as attitudes and preferences.

Information Systems (IS): A curriculum that integrates technical skills and knowledge with applied business and organizational knowledge. Sometimes found in business schools, other times in schools of science, engineering, or in standalone IT academic units. Variations include Business Information Systems (BIS), Computer Information Systems (CIS), and Management Information Systems (MIS).

Information Technology (IT): (a) An umbrella term that encompasses a range of professional positions requiring at least a baccalaureate degree in Computer Science, Information Systems, or closely related majors. (b) A major that focuses mainly on fourth-generation language application development and maintenance.

World Wide Web Search Technologies

Wen-Chen Hu
University of North Dakota, USA

Hung-Jen Yang
University of North Dakota, USA

Jyh-haw Yeh
Boise State University, USA

Chung-wei Lee
Auburn University, USA

INTRODUCTION

The World Wide Web now holds more than six billion pages covering almost all daily issues. The Web's fast growing size and lack of structural style present a new challenge for information retrieval (Lawrence & Giles, 1999a). Traditional search techniques are based on users typing in search keywords which the search services can then use to locate the desired Web pages. However, this approach normally retrieves too many documents, of which only a small fraction are relevant to the users' needs. Furthermore, the most relevant documents do not necessarily appear at the top of the query output list. Numerous search technologies have been applied to Web search engines; however, the dominant search methods have yet to be identified. This article provides an overview of the existing technologies for Web search engines and classifies them into six categories: i) hyperlink exploration, ii) information retrieval, iii) metasearches, iv) SQL approaches, v) content-based multimedia searches, and vi) others. At the end of this article, a comparative study of major commercial and experimental search engines is presented, and some future research directions for Web search engines are suggested. Related Web search technology review can also be found in Arasu, Cho, Garcia-Molina, Paepcke, and Raghavan (2001) and Lawrence and Giles (1999b).

Requirements of Web Search Engines

It is first necessary to examine what kind of features a Web search engine is expected to have in order to conduct effective and efficient Web searches. The requirements for a Web search engine are listed below in order of importance:

1. Effective and efficient location and ranking of Web documents;
2. Thorough Web coverage;
3. Up-to-date Web information;
4. Unbiased access to Web pages;
5. An easy-to-use user interface which also allows users to compose any reasonable query;
6. Expressive and useful search results; and
7. A system that adapts well to user queries.

BACKGROUND

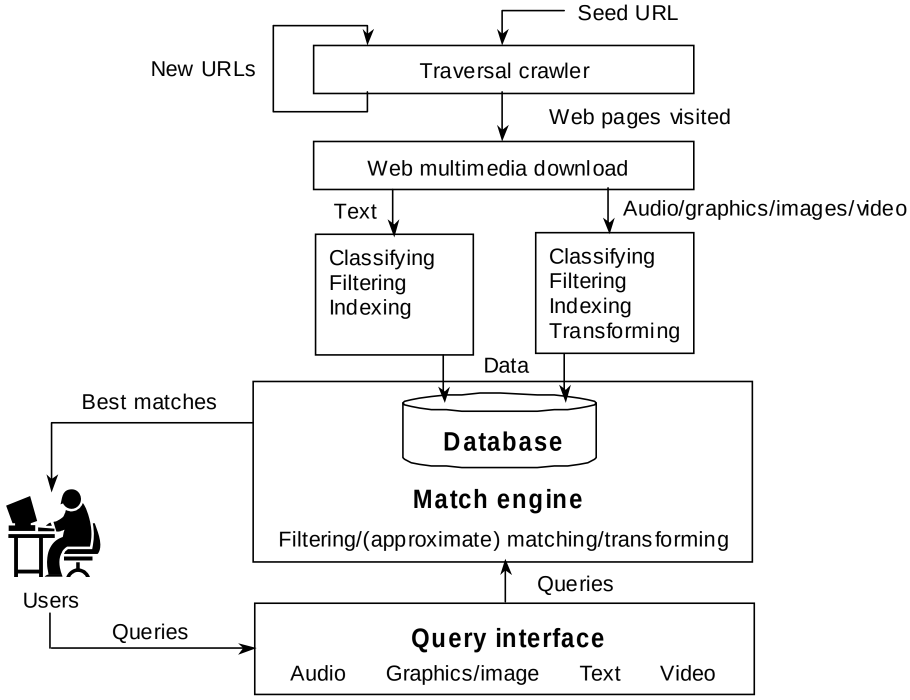
Two different approaches are applied to Web search services: genuine search engines and directories. The difference lies in how listings are compiled.

- Search engines, such as Google, create their listings automatically.
- A directory, such as Yahoo!, depends on humans for its listings.

Some search engines, known as hybrid search engines, maintain an associated directory. Figure 1 shows the system structure of a typical search engine. Search engines traditionally consist of three components: i) the crawler, ii) the indexing software, and iii) the search and ranking software:

- A crawler is a program that automatically scans various Web sites and collects Web documents from them. Two search algorithms, breadth-first searches and depth-first searches, are widely used by crawlers to traverse the Web.
- Automatic indexing is the process of algorithmically examining information items to build a data struc-

Figure 1. System structure of a Web search engine



ture that can be quickly searched. Traditional search engines utilize the following information, provided by HTML scripts, to locate the desired Web pages: i) content, ii) descriptions, iii) hyperlink, iv) hyperlink text, v) keywords, vi) page title, vii) text with a different font, and viii) the first sentence.

- Query processing is the activity of analyzing a query and comparing it to indexes to find relevant items. A user enters a keyword or keywords, along with Boolean modifiers such as "and", "or", or "not", into a search engine, which then scans indexed Web pages for the keywords. To determine in which order to display pages to the user, the engine uses an algorithm to rank pages that contain the keywords.

SEARCH ENGINE TECHNOLOGIES

This section examines the existing technologies for Web search engines and classifies them into six categories: i) hyperlink exploration, ii) information retrieval, iii) metasearches, iv) SQL approaches, v) content-based multimedia searches, and vi) others.

Hyperlink Exploration

Links can be tremendously important sources of information for indexers; the creation of a hyperlink by the author of a Web page represents an implicit endorsement of the page being pointed to. This approach is based on identifying two important types of Web pages for a given topic:

- Authorities, which provide the best source of information on the topic, and
- Hubs, which provide collections of links to authorities.

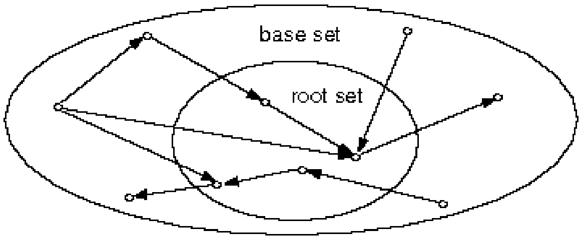
Authorities and hubs are either given top ranking in the search results or used to find related Web pages. A simple method to update a non-negative authority with a weight x_p and a non-negative hub with a weight y_p is given by Chakrabarti et al. (1999). If a page is pointed to by many good hubs, its authority weight is updated by using the following formula:

$$x_p = \sum_{q \text{ such that } q \to p} y_q ,$$

where the notation $q \to p$ indicates that q links to p. Similarly, if a page points to many good authorities, its hub weight is updated via

$$y_p = \sum_{q \text{ such that } p \to q} x_q .$$

Figure 2. Expanding the root set into a base set



Unfortunately, applying the above formulas to the entire Web to find authorities and hubs is impracticable. Ideally, the formulas are applied to a small collection S_σ of pages which contain plenty of relevant documents. The concepts of a root set and a base set have been proposed by (Kleinberg, 1999) to find S_σ. The root set is usually constructed by collecting the t highest-ranked pages for the query σ from a search engine such as Google or Yahoo!. However, the root set may not contain most of the strongest authorities. A base set is therefore built by including any page pointed to by a page in the root set and any page that points to a page in the root set. Figure 2 shows an example of a root set and a base set. The previously mentioned formulas can then be applied to a much smaller set, the base set, instead of the entire Web.

Information Retrieval (IR)

IR techniques are widely used in Web document searches. Among them, relevance feedback and data clustering are two of the most popular techniques used by search engines:

- An initial query is usually a wild guess. Retrieved query results are then used to help construct a more precise query or modify the database indexes (Chang & Hsu, 1999). Two relevance feedback methods, query modification and indexing modification, can be used to improve the search.
- Data clustering is used to improve the search results by dividing the whole data set into data clusters. Each data cluster contains objects of high similarity, and clusters are produced that group documents relevant to the user's query separately from irrelevant ones. For example, the formula below gives a similarity measure:

$$S_{D_i,D_j} = \frac{2\sum_{k=1}^{L}\left(weight_{ik} \cdot weight_{jk}\right)}{\sum_{k=1}^{L} weight_{ik}^2 + \sum_{k=1}^{L} weight_{jk}^2},$$

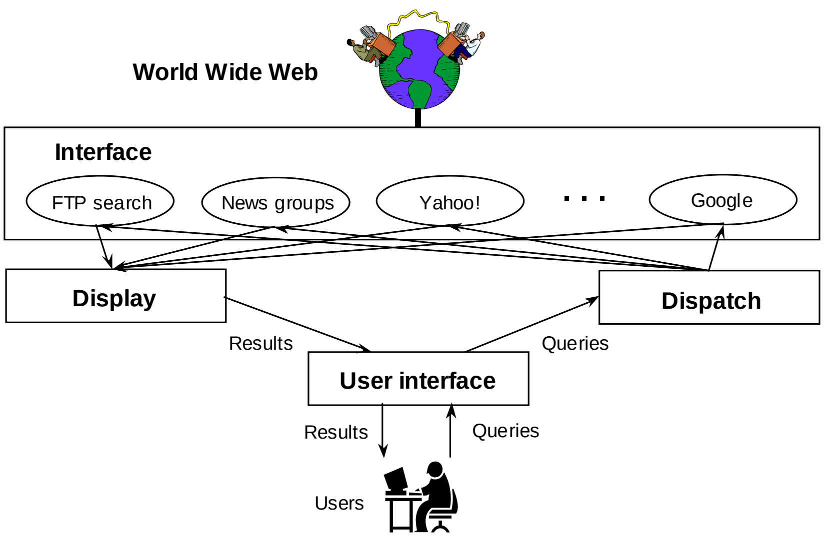
  where $weight_{ik}$ is the weight assigned to $term_k$ in a document D_i (Baeza-Yates, 1992).

Relevance feedback has not so far been applied to any commercial products because it requires some interaction with users, who normally prefer to use a keyword-only interface, whereas data clustering has achieved more success since it does not require any interaction with users to achieve acceptable results.

Metasearches

Metasearch engines (Dreilinger & Howe, 1997) conduct a search using several other search engines simultaneously and then present the results in some sort of integrated format. Figure 3 shows the system structure of a metasearch engine, which consists of three major components:

Figure 3. System structure of a metasearch engine



- Dispatch: Determines to which search engines a specific query is sent. The selection is usually based on network and local computational resources, as well as the long-term performance of search engines on specific query terms.
- Interface: Adapts the user's query format to match the format of a particular search engine, which varies from engine to engine.
- Display: Raw results from the selected search engines are integrated for display to the user. Each search engine also produces different raw results from other search engines and these must be combined to give a uniform format for ease-of-use.

SQL Approaches

Structured Query Language (SQL) approaches view the World Wide Web as a huge database where each record matches a Web page, and use SQL-like languages to support effective and flexible query processing. A typical SQL-like language syntax (Konopnicki & Shmueli, 1998) is

Query:=**select**Attribute_List
from Domain_Specifications
[**where** Search_Conditions];

A query example is given in the following to show the use of the language.

SQL Example. *Find pages in the World Wide Web Consortium (W3C) site where the pages have fewer than 2000 bytes.*
select url **from** http://www.w3c.org/ **where** bytes < 2000;

url is a page's URL and each page has attributes such as bytes, keywords, and text.

Various SQL-like languages have been proposed for Web search engines. The methods introduced previously treat the Web as a graph of discrete objects; another object-oriented approach (Arocena & Mendelzon, 1998) considers the Web as a graph of structured objects. However, neither approach has achieved much success because of its complicated syntax, especially for the latter method.

Content-Based Multimedia Searches

Only a few multimedia search engines are available currently, most of which use name or keyword matching where the keywords are entered by Web reviewers rather than using automatic indexing. The low number of content-based multimedia search engines is mainly due to the difficulty of automated multimedia indexing. Numerous multimedia indexing methods have been suggested in the literature (Yoshitaka & Ichikawa, 1999), yet most do not meet the efficiency requirements of Web multimedia searches, where users expect both a prompt response and the search of a huge volume of Web multimedia data. A few content-based image and video search engines are available online (Lew, 2000). However, a de facto Web image or video search engine is still out of reach because the system's key component—image or video collection and indexing—is either not yet fully automated or not practicable. Similarly, effective Web audio search engines have yet to be constructed since audio information retrieval (Foote, 1999) is considered to be one of the most difficult challenges for multimedia retrieval.

Others

Apart from the above major search techniques, some ad hoc methods worth mentioning include:

- Work aimed at making the components needed for Web searches more efficient and effective, such as better ranking algorithms and more efficient crawlers. Zhang and Dong (2000) propose a ranking algorithm based on a Markov model, which synthesizes the relevance, authority, integrativity, and novelty of each Web resource, and can be computed efficiently through solving a group of linear equations.
- Various enhanced crawlers can be found in the literature (Aggarwal, Al-Garawi, & Yu, 2001). Some crawlers are extensible, personally customized, relocatable, scalable, and Web-site-specific (Heydon & Najork, 1999).
- Artificial Intelligence (AI) can also be used to collect and recommend Web pages. The Webnaut system (Nick & Themis, 2001) learns the user's interests and can adapt as his or her interests change over time. The learning process is driven by user feedback to an intelligent agent's filtered selections.

MAJOR SEARCH ENGINES

Some of the currently available major commercial search engines are listed in Table 1, although many table entries are incomplete as some of the information is classified as confidential due to business considerations (Sullivan, 2003). Most search services are backed up by or are cooperating with several other services. This is because an independent or stand-alone service contains less information and thus tends to lose its users. In the table, the

Table 1. Major commercial Web search engines (SE: Search Engine and AS: Answering Service)

No.	Name	URL	Type	Backup	Method
1	AOL Search	http://search.aol.com/	Hybrid SE	Open Directory	
2	AltaVista	http://www.altavista.com/	SE	LookSmart	
3	Ask Jeeves	http://www.askjeeves.com/	AS		natural language
4	Direct Hit	http://www.directhit.com/	SE	HotBot	hyperlink
5	Excite	http://www.excite.com/	SE	LookSmart	
6	FAST Search	http://www.alltheweb.com/			scalability
7	Google	http://www.google.com/	SE		hyperlink
8	HotBot	http://www.hotbot.com/	Hybrid SE	Direct Hit	
9	Inktomi	http://www.inktomi.com/	SE		
10	LookSmart	http://www.looksmart.com/	Directory	Inktomi	reviewers
11	Lycos	http://www.lycos.com/	Directory	Open Directory	
12	MSN Search	http://search.msn.com/	Directory	LookSmart	
13	Netscape Search	http://search.netscape.com/	SE	Open Directory	
14	Open Directory	http://dmoz.org/	Directory		volunteers
15	Yahoo!	http://www.yahoo.com/	Directory	Google	reviewers

column Backup gives the major backup information provider, and most unfilled methods use keyword matching to locate the desired documents. Most search engines on the list not only provide Web search services but also act as portals, which are Web home bases from which users can access a variety of services, including searches, e-commerce, chat rooms, news, and so forth. Table 2 lists some major experimental search engines, which use advanced search technologies not yet implemented by the commercial search engines. The list in Table 2 is a snapshot of the current situation; the list is highly volatile either because a successful experimental search engine is usually commercialized in a short time or because a prototype system is normally removed after its founders leave the organization. The two tables list major general-purpose search engines; special-purpose search engines including specialty searches, regional searches, kid searches, and so forth, are not considered in this article. They use much smaller databases and therefore give more precise and limited search results.

FUTURE TRENDS

Users of search engines often submit ambiguous queries. Ambiguous queries can be categorized into four types: i) disorderly, ii) incomplete, iii) incorrect, and iv) superfluous queries. Below are examples of perfect and ambiguous queries and the ranked search results from Infoseek, at http://www.infoseek.com/, for the book, *Intelligent multimedia information retrieval*, edited by Mark T. Maybury.

Table 2. Major experimental Web search engines

No.	Name	URL	Method
1	Clever	http://www.almaden.ibm.com/cs/k53/clever.html	hyperlink
2	Grouper	http://longinus.cs.washington.edu/grouper2.html	clustering
3	ImageRover	http://www.cs.bu.edu/groups/ivc/ImageRover/Home.html	image
4	Inquirus	http://www.neci.nj.nec.com/homepages/lawrence/inquirus.html	metasearch
5	Mercator	http://www.ctr.columbia.edu/metaseek/	image
6	MetaSEEk	http://www.research.compaq.com/SRC/mercator/	crawler
7	W3QS	http://www.cs.technion.ac.il/~konop/w3qs.html	SQL
8	WebOQL	http://www.cs.tornoto.edu/~gus/weboql/	Object SQL

- Perfect query: *Intelligent multimedia information retrieval*
  1. Intelligent multimedia information retrieval

- Disorderly query: *Multimedia information intelligent retrieval*
  1. Artificial intelligence, fuzzy logic and neural networks
  2. Intelligent access to information: research in natural language, information retrieval, com puter vision, multimedia and database
  3. Multimedia color PC notebooks
  4. Intelligent multimedia information retrieval

- Incomplete query: *Multimedia information retrieval*
  1. Abstract Stein Mulleller Thiel 95
  2. Corpora Oct 1998 to -: Corpora: TWLT 14: lan guage technology in multimedia information
  3. 3 2.1 Introduction to the workplan

     ...

  6. Intelligent multimedia information retrieval

- Incorrect query: *Intelligent multi-media information retrieval*
  1. Artificial intelligence research laboratory at Iowa State University
  2. Vasant Honavar's home in cyberspace
  3. CIIR multi-media indexing

     ...

  31. Intelligent multimedia information retrieval

- Superfluous query: *Intelligent multimedia infor mation retrieval systems*
  1. Research in multimedia and multimodal parsing and generation
  2. Intelligent multimedia information retrieval

This example shows that even a slight variation in the query produces significant differences among the search results. Users tend to submit ambiguous queries to search engines, most of which use the technology of keyword matching to look for the desired pages. The ambiguity creates undesired search results if keyword matching is used.

CONCLUSION

In less than a decade, the World Wide Web has become one of the three major media, with the other two being print and television. Searching for Web pages is both one of the most common tasks performed on the Web and one of the most frustrating and problematic. This article gave an overview of the current technologies for Web search engines with an emphasis on non-traditional approaches and classified the technologies into six categories. However, apart from the traditional keyword matching techniques, no one method dominates Web search engine technologies. The major reason for this is that the amount of information posted on the World Wide Web is huge and the page formats vary widely.

Another problem with search engines is the skewed search results because of the corporate payments to search engines to boost their rankings. Most search engines list non-commercial pages for free, whereas they charge commercial pages for listing. Take the Google search engine as an example. Other than waiting for their pages being crawled and listed on the Google, the Web content providers may submit their pages via "Add your URL to Google" at http://www.google.com/addurl.html. On the other hand, they may advertise with the Google by using the following two methods:

- AdWords: With AdWords customers create their own ads, choose keywords to tell the Google where to show their ads and pay only when someone clicks on them.
- AdSense: It is for the Web content providers who want to make more revenue from advertising but do not want to serve untargeted ads to their users. AdSense solves this dilemma by delivering text-based AdWords ads that are relevant to what readers see on the pages.

Though the search results are distorted by corporate payments, it is necessary for the search engines to survive and improve their services. Also, the users can normally tell the differences between the paid contents and the ranked search results because they are usually at the different places on the result pages.

REFERENCES

Aggarwal, C.C., Al-Garawi, F., & Yu, P.S. (2001). Intelligent crawling on the World Wide Web with arbitrary predicates. In *Proceedings of the 10th International World Wide Web Conference*, Hong Kong.

Arasu, A., Cho, J., Garcia-Molina, H., Paepcke, A., & Raghavan, S. (2001). Searching the Web. *ACM Transactions on Internet Technology (TOIT)*, *1*(1), 2-43.

Arocena, G.O., & Mendelzon, A.O. (1998). WebOQL: Restructuring documents, databases and Webs. In *Proceedings of the 14th International Conference on Data Engineering*, Orlando, Florida.

Baeza-Yates, R.A. (1992). Introduction to data structures and algorithms related to information intrieval. In W.B. Frakes & R.A. Baeza-Yates (Eds.), *Information retrieval data structures & algorithms*, 13-27. Upper Saddle River, NJ: Prentice Hall.

Chakrabarti, S., Dom, B.E., Kumar, S.R., Raghavan, P., Rajagopalan, S., Tomkins, A., Gibson, D., & Kleinberg, J. (1999). Mining the Web's link structure. *IEEE Computer*, *32*(8), 60-67.

Chang, C., & Hsu, C. (1999). Enabling concept-based relevance feedback for information retrieval on the WWW. *IEEE Transactions on Knowledge and Data Engineering*, *11*(4), 595-609.

Dreilinger, D., & Howe, A.E. (1997). Experiences with selecting search engines using metasearch. *ACM Transactions on Information Systems*, *15*(3), 195-222.

Foote, J. (1999). An overview of audio information retrieval. *Multimedia Systems*, *7*(1), 2-10.

Heydon, A., & Najork, M. (1999). Mercator: A scalable, extensible Web crawler. *World Wide Web*, *2*(4), 219-229.

Kleinberg, J.M. (1999). Authoritative sources in a hyperlinked environment. *JACM*, *46*(5), 604-632.

Konopnicki, D., & Shmueli, O. (1998). Information gathering in the World Wide Web: The W3QL query language and the W3QS system. *ACM Transactions on Database Systems*, *23*(4), 369-410.

Lawrence, S., & Giles, C.L. (1999a). Accessibility of information on the Web. *Nature*, *400*, 107-109.

Lawrence, S., & Giles, C.L. (1999b). Searching the Web: General and scientific information access. *IEEE Communications*, *37*(1), 116-122.

Lew, M. S. (2000). Next generation Web searches for visual content. *IEEE Computer*, *33*(11), 46-53.

Nick, Z.Z., & Themis, P. (2001). Web search using a genetic algorithm. *IEEE Internet Computing*, *5*(2), 18-26.

Sullivan, D. (2003). Major search engines and directories. Search Engine Watch *http://www.searchenginewatch.com/*

Yoshitaka, A., & Ichikawa, T. (1999). A survey on content-based retrieval for multimedia databases. *IEEE Transactions on Knowledge and Data Engineering*, *11*(1), 81-93.

Zhang, D., & Dong, Y. (2000). An efficient algorithm to rank Web resources. In *Proceedings of the 9th International World Wide Web Conference*, Amsterdam, Netherlands.

KEY TERMS

Authority: A Web site that provides the best source of information on a specific topic.

Automatic Indexing: A process that algorithmically examines information items to build a data structure that can be quickly searched.

Crawler/Spider: A program that automatically scans various Web sites and collects Web documents from them. It follows the links on a site to find other relevant pages and is usually used to feed pages to search engines.

Directory: Other than search engines, directories provide another approach of Web searches. A directory is a subject guide, typically organized by major topics and subtopics, which are created based on the submissions from either Webmasters or editors who have reviewed the pages.

Hub: A Web site that provides collections of links to authorities.

Hyperlinks: A hyperlink is a selectable connection from one word, phrase, picture, or information object to another. By clicking on a hyperlink, a Web user can move easily from one Web page to another page. The most common form of hyperlinks is the highlighted word, phrase, or picture.

Metasearch Engine: It conducts a search using several other search engines simultaneously and then presents the results in some sort of integrated format. This lets users see at a glance which particular search engine returned the best results for a query without having to search each one individually.

Search Engine: A software system such as Google and Alta Vista that searches documents on the World Wide Web and USENET newsgroups for specified keywords and returns a list of the documents which are relevant to the keywords.

Structured Query Language (SQL): It is a standard interactive and programming language for accessing and manipulating a database. Its commands include selection, insertion, update, deletion, finding out the location of data, and so forth.

Uniform Resource Locator (URL): It is the address of an object accessible on the Internet. The object could be an HTML document, a text file, an image file, a program such as a common gateway interface application, and so forth. They are mainly used in HTML documents to specify the target of a hyperlink.

XML Schema Integration and E-Commerce

Kalpdrum Passi
Laurentian University, Canada

Louise Lane
Laurentian University, Canada

Sanjay Madria
University of Missouri-Rolla, USA

Mukesh Mohania
IBM India Research Lab, India

INTRODUCTION

XML (eXtensible Markup Language) is used to describe semi-structured data, i.e., irregular or incomplete data whose structure may be subject to unpredictable changes. Unlike traditional semi-structured data, XML documents are self-describing, thus XML provides a platform-independent means to describe data and, therefore, can transport data from one platform to another (Bray, Paoli, & Sperberg-McQueen, 1998). XML documents can be both created and used by applications. The valid content, allowed structure, and metadata properties of XML documents are described by their related schema(s) (Thompson, Beech, Maloney, & Mendelsohn, 2001). An XML document is said to be *valid* if it conforms to its related schema. A schema also gives additional semantic meaning to the data it is used to tag. The schema is provided independently of the data it describes. Any given data set may rely on multiple schemas for validation. Any given schema may itself refer to multiple schemas.

In e-commerce, XML documents can be used to publish everything from product catalogs and airline schedules to stock reports and bank statements. XML forms can be used to place orders, make reservations, and schedule shipments. XML eliminates the need for custom interfaces with every customer and supplier, allowing buyers to compare products across many vendors and catalog formats, and sellers to publish their catalog information once to reach many potential buyers. XML can also enable online businesses to build on one another's published content and services to create innovative virtual companies, markets, and trading communities. With a global view of the Internet-wide shopping directories, a query system can locate all merchants carrying a specific product or service and then query each local schema in parallel to locate the best deals. The query system can sort the offers according to criteria set by the buyers–the cheapest flight, the roomiest aircraft, or some weighted combination. The traditional method used for business-to-business (B2B) information exchange is through Electronic Data Interchange (EDI), which is complex, expensive, and necessitates a custom integration solution between each pair of trading partners. A query-based system that uses XML as the common format to enterprise integration is simpler and more open than traditional EDI, as it eliminates the proprietary message formats used by each company. A complete business integration solution also requires metadata for each commerce community, a means to map each local schema into an integrated global view, and a server for processing XML documents and invoking appropriate applications and services.

RELATED WORK

The problem of schema and integration of heterogeneous and federated databases has been addressed widely. Several approaches to schema integration exist as described in Batini, Lanzerini, and Navathe (1986); Behrens, 2000; Christophides, Cluet, and Simon, (2000); Haas, Miller, Niswanger, Roth, Schwarz, and Wimmers (1999); Miller, Ioannidis, and Ramakrishnan (1993); Parent and Spaccapietra (1998); and Ram and Ramesh (1998). A global schema in the general sense can be viewed as a regular schema, the rules of which encompass the rules of a common data model. A global schema eliminates data model differences and is created by integrating local schemas. The creation of a global schema also helps to eliminate duplication, avoid problems of multiple updates, and thus minimize inconsistencies.

Most schema integration approaches decompose integration into a multi-layered architecture like the one followed in this paper constituting *pre-integration*, *comparison,* and *integration* (Batini et al., 1986; Miller, 1998).

There have been some recent systems (Adali, Candan, Papakonstantinou, & Subramanian, 1996; Papakonstantinou, Garcia-Molina, & Widom, 1995; Tomasic, Raschid, & Valduriez, 1996) that integrate data from multiple sources. Most of these systems provide a set of mediated/global schema(s). Some systems like *Garlic* (Roth & Schwarz, 1997) use wrappers to describe the data from different sources in its repositories and provide a mechanism for a middleware engine to retrieve the data. The *Garlic* system also builds global schema from the individual repositories. The *comparison* and *restructuring* phase of integration is handled in some systems through human interaction using a graphical user interface as in Clio (Hernandez, Miller, Haas, Yan, Ho, & Tian, 2001; Miller, Haas, & Hernandez, 2000; Miller et al., 2001; Yan, Miller, Haas, & Fagin, 2001) and in others semi-automatically through machine learning techniques such as in *Tukwila* data integration system at University of Washington. The *Tukwila* integration system reformulates the user query into a query over the data sources, which are mainly XML documents corresponding to DTD schemas and relational data.

INTEGRATION REQUIREMENTS, ARCHITECTURE AND METHODOLOGY

XML Schema (Thompson et al., 2001) has recently been recommended as the standard schema language to validate XML documents. It has a stronger expressive power than the DTD (Document Type Definition) schema for the purpose of data exchange and integration from various sources of data.

Since here we assume that the schemas to be integrated currently validate a set of existing XML documents, data integrity and continued document delivery are chief concerns of the integration process, thus closely linking XML Schema integration to the theoretical requirements and process of database integration.

Satisfying the condition that the global schema meets the requirements of all the initial schemas is the most difficult part of integration, and is worsened by data model heterogeneity.

We define an object-oriented data model that we call XSDM (XML Schema Data Model) for the purpose of XML Schema integration. We use the three-layered architecture of *pre-integration*, *comparison*, and *integration* to achieve XML Schema integration.

The XML Schema integration process developed in this paper uses a *one shot n-ary* (Batini et al., 1986) strategy. The *one shot n-ary* style integrates all the initial schemas at once. Schema integration should be both *extensible* and *scalable*. It should be easy to add or remove sources of data (i.e., schemas), to manage large numbers of schemas, and to adjust the resulting global schema. With the XML Schema integration approach, multiple schemas can be integrated at one time.

Any global integrated schema must meet the following three criteria: *completeness*, *minimality*, and *understandability*. In order to meet the first criteria of *completeness,* all the elements in the initial schemas should be in the merged schema. The merged schema can be used to validate any of the XML instance documents that were previously validated by one of the initial schema specifications. To satisfy the second criterion, *minimality*, each unique element is defined only once in the schema. Redundancy is eliminated wherever possible through the identification of equivalent elements and attributes, and the subsequent use of substitution groups. Datatypes for terminal elements are expanded through the use of constraint facet redefinition, or unions of incompatible datatypes, only to the point necessary to satisfy boundary conditions. Optionality of elements (i.e., minOccurs and maxOccurs values) is expanded to meet boundary restrictions only. Finally, to comply with the third criterion, *understandability*, in the case of XML Schema integration, the global schema is formulated in a referenced style, rather than an inline style (nested definitions), for ease of reading and assessment.

During *pre-integration*, an analysis of the schemas to be integrated occurs. Priority of integration is determined if the process is not to be *one shot*. Preferences may be given to retaining the entire or certain portions of schemas as whole parts of the global schema. Designer interaction occurs in view integration as assertions of relationships and constraints of elements are discovered.

During the *comparison* stage of integration, correspondences as well as conflicts between elements are identified. There are four *semantic relationships* defined by Batini et al. (1986). The schematic representations can be viewed as *identical*, *equivalent*, *compatible*, or *incompatible*. We identify six types of semantic relationships, which apply to XML Schema elements – *identical*, *equal*, *equivalent*, *subset*, *unique*, and *incompatible*.

The fundamental activity in the *comparison* phase of integration is *conflict resolution*. Conflict identification and resolution is central to successful integration. *Naming conflicts*, *datatype conflicts* & *scale differences*, and *structural conflicts* can occur during XML Schema Integration.

During the *conformance* phase, the *semantic relationships* and *conflicts* identified in the comparison phase are resolved. Initial schemas may be transformed in order to make them more suitable for integration. The XML

Schemas in question are, by definition, *correct*, *complete*, and *minimal* because XML documents exist that have been successfully validated by the schema(s) in question. A well-formed XML Schema, by definition, is *minimal* and *complete*. There is only one root element. All other elements present in the schema are related to the root element as either direct or indirect children. Each element has a unique name. Each element is defined only once. The schema should be viewed as a document that is fixed in nature. The schema is currently used to validate an existing non-empty set of XML documents. Therefore, it is important that these initial schemas are not altered in any fashion.

Transformations that occur to align the initial schemas are accomplished using *restructuring* (renaming, substitution groups, and subsetting) in the global schema only. The datatypes of terminal elements and attributes are expanded in the global schema to meet boundary restrictions through constraint facet redefinition and unions of incompatible datatypes. Optionality (minOccurs, maxOccurs) of child-elements is expanded to satisfy the largest minimum boundaries.

During schema *merging*, initial schemas are superimposed onto each other to result in the merged global schema. The merged schema is *complete* and *minimal*. The global schema is rearranged to ensure the highest level of minimality and understandability.

CONCLUSIONS AND FUTURE WORK

We have defined an object-oriented data model XSDM (XML Schema Data Model) and have defined the requirements, architecture, and a detailed methodology for integrating XML Schemas. In our opinion, this is the first detailed methodology for integrating XML Schema, although a number of papers have appeared on general integration framework as has been cited in the related work. We are developing a semi-automated Java-based XML Schema integration process, which can also be applied to DTD integration.

It may also prove possible to integrate XML Schemas with other schema language specifications using the work contained in this paper as a basis for comparison, conflict resolution, and merging. This possibility has not been explored in this paper; rather, it is presented as a suggestion for further research.

We are currently working on an XML Schema query system, based on the data model and integration strategies discussed in the paper.

REFERENCES

Adali, S., Candan, K., Papakonstantinou, Y., & Subramanian, V.S. (1996). Query caching and optimization in distributed mediator systems. *Proceedings ACM SIGMOD*, Montreal, Canada , 25(2), 137-148.

Batini, C., Lenzerini, M., & Navathe, S.B. (1986). A comparative analysis of methodologies for database schema integration. *ACM Computing Surveys,* 18(4), 323-364.

Behrens, R. (2000). A grammar based model for XML Schema integration. *Proceedings British National Conference on Databases* (BNCOD), Exeter, UK, 172-190.

Bray, T., Paoli, J., & Sperberg-McQueen, C.M. (1998). Extensible markup language (XML) 1.0 - W3C Recommendation. Retrieved September 15, 2000 from *http://www.w3.org/TR/REC-xml.html*

Christophides, V., Cluet, S., & Simon, J. (2000). On wrapping query languages and efficient XML integration. *Proceedings ACM SIGMOD*, Dallas, Texas, 141-152.

Haas, L.H., Miller, R.J., Niswanger, B., Roth, M.T., Schwarz, P.M., & Wimmers, E.L. (1999). Transforming heterogeneous data with database middleware: Beyond integration. *IEEE Data Engineering Bulletin*, 22(1), 31-36.

Hernandez, M.A., Miller, R.J., Haas, L.H., Yan, L., Ho, C.T.H., & Tian, X. (2001). Clio: A semi-automatic tool for schema mapping, system demonstration. *Proceedings ACM SIGMOD*, Santa Barbara, CA, USA, May 21-24, 2001 (p. 607).

Miller, R.J. (1998). Using schematically heterogeneous structures. *Proceedings of the ACM SIGMOD*, International Conference on the Management of Data, Seattle, WA, USA, 27(2),189-200.

Miller, R.J., Haas. L.M., & Hernandez, M. (2000). Schema mapping as query discovery. *Proceedings of the Twenty-Sixth International Conference on Very Large Databases* (VLDB), Cairo, Egypt, 77-88.

Miller, R.J., Hernandez, M., Haas, L.M., Yan, L., Howard Ho, C.T., Fagin, R., &Popa, L. (2001). The Clio project: Managing heterogeneity, *SIGMOD Record*, 30(1), 78-83.

Miller, R. J., Ioannidis, Y.E., & Ramakrishnan, R. (1993). The use of information capacity in schema integration and translation. *Proceedings of the 19th VLDB Conference*, Dublin, Ireland, 120-133.

Papakonstantinou, Y., Garcia-Molina, H., & Widom, J. (1995). Object exchange across heterogeneous information sources. *Proceedings IEEE Conference on Data Engineering*, Taipei, Taiwan, 251-260.

Parent, C. & Spaccapietra, S. (1998). Issues and approaches of database integration. *Communications of the ACM*, 41(5), 166-178.

Ram, S. & Ramesh, V. (1998). Schema integration: Past, present and future. In A. Elmagarmid, M. Rusinkiewicz, & A. Sheth (Eds.), *Management of heterogeneous and autonomous database systems* (pp. 119-155). San Mateo, CA: Morgan-Kaufmann.

Roth, M.T. & Schwarz, P. (1997). Don't scrap it, wrap it! A wrapper architecture for legacy data sources. *Proceedings of the International Conference of Very Large Databases* (VLDB), Athens, Greece, 266-275.

Thompson, H.S., Beech, D., Maloney, M., & Mendelsohn, N. (Eds.). (2001). XML Schema, Part 1: Structures - W3C Recommendation. Retrieved October 15, 2002 from *http://www.w3.org/TR/xmlschema-1/*

Tomasic, A., Raschid, L., & Valduriez, A. (1996). Scaling heterogeneous databases and the design of disco. *Proceedings of the 16th International Conference on Distributed Computing Systems (ICDCS)*, Hong Kong, IEEE Computer Society, 449-457.

Yan, L., Miller, R.J., Haas, L.M., & Fagin, R. (2001). Data-driven understanding and refinement of schema mappings. *Proceedings of the ACM SIGMOD*, Santa Barbara, California, USA, May 21-24, 2001 (pp. 485-496).

KEY TERMS

Business-to-Business (B2B): Used to describe an e-business solution that caters to other businesses. It offers integrated business applications and services that allow small and mid-sized organizations and divisions of large enterprises to connect employees, customers, and suppliers for improved efficiency.

DTD (Document Type Definition): An XML DTD is a mechanism to define the structure of XML documents. An XML DTD lists the various elements and attributes in a document and the context in which they are to be used. It can also list any elements a document cannot contain.

e-Business (e-Mail, e-Commerce, e-Zine, e-Solution, e-Market, e-Service, etc.): Oftentimes used without the hyphen, the "e" originally stood for "electronic," as in "online." Today the term is used rather freely to describe any situation or solution that has made the migration from real world to the Internet.

Electronic Data Interchange (EDI): Exchange between businesses of computer-readable data in a standard format.

eXtensible Markup Language (XML): A simplified meta language, derived from Standard Generalized Markup Language (SGML), emerging as the standard for self-describing data exchange in Internet applications.

Global Schema: Schema obtained after integrating local schemas.

XML Schema: A database-inspired method for specifying constraints on XML documents using an XML-based language. Schemas address deficiencies in DTDs, such as the inability to put constraints on the kinds of data that can occur in a particular field (for example, all numeric). Since schemas are founded on XML, they are hierarchical, so it is easier to create an unambiguous specification, and possible to determine the scope over which a comment is meant to apply.

Index of Key Terms

Symbols

A

N

V

W

Index

F

J

K

T